数值分析

—— 科学与工程计算基础

任课教师：黄忠亿

清华大学数学科学系

目录

引言

许多工程计算问题都需要计算偏微分方程特征值问题, 离散后往往转化为一些代数特征值问题.

例 5.1 (考虑描述弦的振动的波动方程:)

$$\frac{\partial^2 u(x,t)}{\partial t^2} = c^2 \frac{\partial^2 u(x,t)}{\partial x^2},$$

这里 u − −表示位移, c − −表示所谓声速.

假设弦长为1，两端固定, 即 $u(0,t) = u(1,t) = 0$.

可以取 $u(x,t) = v(x)e^{\mathrm{i}\omega t}$ 形式的谐波解, 代入上述方程, 有

$$\begin{cases} -v''(x) = \lambda v, & x \in (0,1) \\ v(0) = v(1) = 0. \end{cases}$$

其中 $\lambda = \omega^2/c^2$.

引言

显然, $v_m(x) = \sin m\pi x$ 就是上述特征值问题的解, 相应的特征值为 $\lambda_m = m^2\pi^2, m = 1, 2, \cdots$

如果我们将 $[0, 1]$ 离散为 $x_j = jh, h = \frac{1}{n+1}, j = 0, 1, \cdots, n+1$.

将二阶导数近似为 $v''(x_j) \approx \frac{1}{h^2}[v(x_{j+1}) - 2v(x_j) + v(x_{j-1})]$, 代入上式

$$\begin{cases} \frac{1}{h^2}(-v_{j-1} + 2v_j - v_{j+1}) = \lambda v_j, & j = 1, \cdots, n \\ v_0 = v_{n+1} = 0, \end{cases}$$

记

$$A = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & & \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{pmatrix}, V = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix},$$ 问题写为 $AV = \lambda V$.

这就是一个代数特征值问题.

特征值问题

定义 5.1

设 $A = (a_{ij}) \in \mathbb{C}^{n\times n}$, 称

$$\varphi(\lambda) = \det(\lambda I - A) \equiv \lambda^n - tr(A)\lambda^{n-1} + \cdots + (-1)^n \det A$$

为 A 的特征多项式. 方程 $\varphi(\lambda) = 0$ 称为 A 的特征方程.

特征多项式的 n 个复根称为 A 的特征值.

一般记 $\Lambda(A)$—A 的所有特征值之集.

$(\lambda I - A)\mathbf{x} = 0$ 的非零解 $\mathbf{x} \in \mathbb{C}^n$ 称为 A 对应于 λ 的特征向量,

即 $A\mathbf{x} = \lambda\mathbf{x}$.

回顾一下关于矩阵特征值的常用性质.

关于特征值的性质

定理 5.1

设 $A\mathbf{x} = \lambda\mathbf{x}$ $(\mathbf{x} \neq 0)$, 则

1. $\forall 0 \neq c \in \mathbb{C}$, $c\lambda$ 为 cA 的特征值;
2. $\lambda - p$ 为 $A - pI$ 的特征值, 且 $(A - pI)\mathbf{x} = (\lambda - p)\mathbf{x}$, $\forall p \in \mathbb{C}$;
3. $\forall k \in \mathbb{N}$, λ^k 为 A^k 的特征值;
4. 若 A 可逆, 则 λ^{-1} 为 A^{-1} 的特征值, $A^{-1}\mathbf{x} = \frac{1}{\lambda}\mathbf{x}$.

一般来讲特征值越集中越好算,
而这一步相当于一个平移

定理 5.2

设 λ_i 为 A 的特征值, $i = 1, \cdots, n$, 则

1) $\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii} \equiv tr(A)$; 2) $\det A = \prod_{i=1}^{n} \lambda_i$.

关于特征值的性质

定理 5.3

$\forall B \in \mathbb{C}^{n\times n}$, $\Lambda(B^T) = \Lambda(B)$, 且 $\Lambda(\bar{B}) = \overline{\Lambda(B)}$.

定理 5.4

设 A 是分块上(下)三角阵: $A = \begin{pmatrix} A_{11} & \cdots & A_{1m} \\ & \ddots & \vdots \\ & & A_{mm} \end{pmatrix}$,

其中 A_{ii} 为方阵. 则 $\Lambda(A) = \bigcup_{i=1}^{m} \Lambda(A_{ii})$

定理 5.5

设A与B相似, 即 $\exists P$非奇异, s.t. $A = P^{-1}BP$, 则

1) $\Lambda(A) = \Lambda(B)$; 2) 若 $A\mathbf{y} = \lambda\mathbf{y}$, 则 $B(P\mathbf{y}) = \lambda(P\mathbf{y})$.

关于特征值的估计—圆盘定理

1. 圆盘定理

定义 5.2 (Gerschgorin 圆盘)

设 $A=(a_{ij})\in\mathbb{C}^{n\times n}$, 对 $i=1,\cdots,n$ 令

- $r_i=\sum\limits_{j\neq i}|a_{ij}|,\ r_i^*=\sum\limits_{j\neq i}|a_{ji}|$;
- $D_i=\{z\in\mathbb{C}\,|\,|z-a_{ii}|\leq r_i\},\ D_i^*=\{z\in\mathbb{C}\,|\,|z-a_{ii}|\leq r_i^*\}$.

称圆盘 D_i (D_i^*) 为A的 *Gerschgorin* 圆盘.

转置后矩阵特征值不变，因此考虑行与考虑列基本没有区别

关于特征值的估计—圆盘定理

定理 5.6 (Gerschgorin 圆盘定理)

对 $A=(a_{ij})_{n\times n}$, 有 由该定理发现对角占优的矩阵特征值情况

1. A 的每一个特征值都必在上面定义的某个 *Gerschgorin* 圆盘中, 即 $\Lambda(A)\subset\bigcup_{i=1}^n D_i \bigcap \bigcup_{i=1}^n D_i^*$; 注意不是Di的交再并, 而是整体的并的交
2. 若 A 的 m 个 D_i 组成一个连通集S, 且 S 与其余 $n-m$ 个圆盘 D_i 不相交, 则 S 内恰好有 A 的 m 个特征值.

$\lhd$ 1)设 $A\mathbf{x}=\lambda\mathbf{x}$, 且 $\|\mathbf{x}\|_\infty=1$. 即若 $\mathbf{x}=(x_1,\cdots,x_n)^T\in\mathbb{C}^n$,

设其中 $|x_j|=\|\mathbf{x}\|_\infty=1$. 那么 **在特征向量的最大模分量所**对应的指标上

$$|\lambda-a_{jj}|=|(\lambda-a_{jj})x_j|\overset{A\mathbf{x}=\lambda\mathbf{x}}{=}\left|\sum_{k\neq j}a_{jk}x_k\right|\le\sum_{k\neq j}|a_{jk}|=r_j.$$

因此有 $\lambda\in D_j$. 因为 $\Lambda(A^T)=\Lambda(A)$, 因此也有 $\lambda\in D_j^*$.

这样 $\Lambda(A)\subset\bigcup_{i=1}^n D_i\bigcap\bigcup_{i=1}^n D_i^*$.

关于特征值的估计—圆盘定理

2) 记 $D=\mathrm{diag}(a_{ii})\equiv\begin{pmatrix}a_{11} & & \\ & \ddots & \\ & & a_{nn}\end{pmatrix}$, 令 $C=A-D$.

考虑矩阵族 $\{\Phi(\theta)=D+\theta C\}_{0\le\theta\le1}$, 显然 $\Phi(0)=D$; $\Phi(1)=A$. 易见 $\Phi(\theta)$ 的特征多项式为 θ 的多项式 (自然也是 θ 的连续函数). 由第一部分, 我们已经证明 $\Phi(\theta)$ 的特征值均在 $\bigcup_{i=1}^{n} D_i(\theta)$ 中, 其中

$$D_i(\theta)=\left\{z\in\mathbb{C}\,\middle|\,|z-a_{ii}|\le\theta\cdot r_i\right\},\quad i=1,\cdots,n.$$

而且显然 $D_i(\theta)\subset D_i$, 当 $0\le\theta\le1$ 时.

无妨设 A 的前 m 个圆盘之并是一个连通集 $S=\bigcup_{i=1}^{m} D_i$, 与其他 $n-m$ 个圆盘 D_j 不相交, 即 $S\bigcap D_j=\emptyset$, $j=m+1,\cdots,n$.

关于特征值的估计—圆盘定理

由于 $D_i(\theta) \subset D_i$, 自然 $S(\theta) \equiv \bigcup_{i=1}^m D_i(\theta) \subset S, \forall \theta \in [0,1]$.

而且 $S(\theta) \bigcap D_j(\theta) = \emptyset, j = m+1, \cdots, n, \forall \theta \in [0,1]$.

$\theta = 0$ 时, $\Phi(0) = D$, 其特征值就是 a_{ii}, 每个圆盘 $D_i(\theta)$ 的圆心.

我们记 $\Phi(\theta)$ 的特征值为 $\lambda_i(\theta)$, 即有 $\lambda_i(0) = a_{ii}, i = 1, \cdots, n$.

我们已说过, $\Phi(\theta)$ 的特征值 $\lambda_i(\theta)$ 是 θ 的连续函数,

$\theta = 0$ 时, $\lambda_i(\theta) \in S, i = 1, \cdots, m$ 都成立.

当 θ 从 0 连续变化到 1, 每个 $\lambda_i(\theta)$ 的变化轨迹都是连通集,

它不可能穿过 S 的边界跳跃到某个圆盘 $D_j(\theta)$ $(j \geq m+1)$ 中.

同样那些特征值 $\lambda_j(\theta)$ $(j \geq m+1)$ 也不会跳到 S 中.

因而 S 中应该恰好有 $A = \Phi(1)$ 的 m 个特征值. $\triangleright$

关于特征值的估计—圆盘定理

例 5.2

设 $A=\begin{pmatrix} 4 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & -4 \end{pmatrix}$, 其特征值为 $\lambda_{1,2,3}\approx 4.2, -0.44, -3.76.$

A 的圆盘为: $D_1 : |\lambda-4|\le 1,\ \ D_2 : |\lambda|\le 2,\ \ D_3 : |\lambda+4|\le 2;$

$D_1^* : |\lambda-4|\le 2,\ \ D_2^* : |\lambda|\le 2,\ \ D_3^* : |\lambda+4|\le 1.$

若取对角阵 $P=\mathrm{diag}(1,1,0.9)$, 有 $PAP^{-1}=\begin{pmatrix} 4 & 1 & 0 \\ 1 & 0 & -\frac{10}{9} \\ 0.9 & 0.9 & -4 \end{pmatrix}$,

此时三个圆盘都不相交: **Ei互不相交或者Ei*互不相交**

$E_1 : |\lambda-4|\le 1,\quad E_2 : |\lambda|\le \frac{19}{9},\quad E_3 : |\lambda+4|\le 1.8;$

$E_1^* : |\lambda-4|\le 1.9,\ \ E_2^* : |\lambda|\le 1.9,\ \ E_3^* : |\lambda+4|\le \frac{10}{9}.$

关于特征值的估计—Rayleigh 商

设 $A \in \mathbb{C}^{n\times n}$ 为 Hermite 阵, 其特征值为实数 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

定义 5.3

$\forall 0 \neq \mathbf{x} \in \mathbb{C}^n$, 称 $R(\mathbf{x}) = \dfrac{(A\mathbf{x}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})} \in \mathbb{R}$ 为 A 相应于 $\mathbf{x}$ 的 Rayleigh 商.

定理 5.7

设 A 为 Hermite 阵, 特征值为 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, 相应特征向量 $\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(n)}$ 为正交规范基, 则有

1) $\forall 0 \neq \mathbf{x}$, $\lambda_1 \geq R(\mathbf{x}) \geq \lambda_n$; *2)* $\lambda_1 = \max\limits_{\mathbf{x}\neq 0} R(\mathbf{x})$, $\lambda_n = \min\limits_{\mathbf{x}\neq 0} R(\mathbf{x})$;

3)
$$\lambda_i = \max_{dim(W)=i} \min_{0\neq\mathbf{x}\in W} R(\mathbf{x}) = \min_{0\neq\mathbf{x}\in\mathrm{span}\{\mathbf{x}^{(1)},\cdots,\mathbf{x}^{(i)}\}} R(\mathbf{x})$$
$$= \min_{dim(W)=n+1-i} \max_{0\neq\mathbf{x}\in W} R(\mathbf{x}) = \max_{0\neq\mathbf{x}\in\mathrm{span}\{\mathbf{x}^{(i)},\cdots,\mathbf{x}^{(n)}\}} R(\mathbf{x})$$

关于特征值的估计—Rayleigh 商

$\triangleleft$ 1)、2)是第 3) 条的特例, 且 3) 中两行是类似地, 因此我们下面只证3)中的第二行: 先证 $\lambda_i = \max\limits_{0\neq\mathbf{x}\in\operatorname{span}\{\mathbf{x}^{(i)},\cdots,\mathbf{x}^{(n)}\}} R(\mathbf{x})$

任取 $0 \neq \mathbf{x} \in \operatorname{span}\{\mathbf{x}^{(i)},\cdots,\mathbf{x}^{(n)}\}$, 由 $\mathbf{x}^{(1)},\cdots,\mathbf{x}^{(n)}$ 为正交规范基可知, $(\mathbf{x},\mathbf{x}^{(k)}) = 0$, $k = 1,\cdots,i-1$.

也就是说 $\mathbf{x} = \sum_{k=i}^{n}(\mathbf{x},\mathbf{x}^{(k)})\mathbf{x}^{(k)}$, 因而 $(\mathbf{x},\mathbf{x}) = \sum_{k=i}^{n}|(\mathbf{x},\mathbf{x}^{(k)})|^2$.

那么有 $A\mathbf{x} = \sum_{k=i}^{n}(\mathbf{x},\mathbf{x}^{(k)})\lambda_k\mathbf{x}^{(k)}$, 还是利用 $\mathbf{x}^{(k)}$ 的正交性, 有

关于特征值的估计—Rayleigh 商

◁ 1)、2)是第 3) 条的特例, 且 3) 中两行是类似地, 因此我们下面只证3)中的第二行: 先证 $\lambda_i = \max\limits_{0 \neq \mathbf{x} \in \text{span}\{\mathbf{x}^{(i)},\cdots,\mathbf{x}^{(n)}\}} R(\mathbf{x})$

任取 $0 \neq \mathbf{x} \in \text{span}\{\mathbf{x}^{(i)}, \cdots, \mathbf{x}^{(n)}\}$, 由 $\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(n)}$ 为正交规范基可知, $(\mathbf{x}, \mathbf{x}^{(k)}) = 0$, $k = 1, \cdots, i-1$.

也就是说 $\mathbf{x} = \sum_{k=i}^{n} (\mathbf{x}, \mathbf{x}^{(k)}) \mathbf{x}^{(k)}$, 因而 $(\mathbf{x}, \mathbf{x}) = \sum_{k=i}^{n} |(\mathbf{x}, \mathbf{x}^{(k)})|^2$.

那么有 $A\mathbf{x} = \sum_{k=i}^{n} (\mathbf{x}, \mathbf{x}^{(k)}) \lambda_k \mathbf{x}^{(k)}$, 还是利用 $\mathbf{x}^{(k)}$ 的正交性, 有

$$(A\mathbf{x}, \mathbf{x}) = \sum_{k=i}^{n} \lambda_k |(\mathbf{x}, \mathbf{x}^{(k)})|^2 \leq \lambda_i \sum_{k=i}^{n} |(\mathbf{x}, \mathbf{x}^{(k)})|^2 = \lambda_i (\mathbf{x}, \mathbf{x})$$

$\Longrightarrow \lambda_i \geq \sup\limits_{0 \neq \mathbf{x} \in \text{span}\{\mathbf{x}^{(i)},\cdots,\mathbf{x}^{(n)}\}} \dfrac{(A\mathbf{x}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})}$. 特别取 $\mathbf{x} = \mathbf{x}^{(i)}$

$\Longrightarrow (A\mathbf{x}^{(i)}, \mathbf{x}^{(i)}) = \lambda_i (\mathbf{x}^{(i)}, \mathbf{x}^{(i)})$. 即 $\lambda_i = \max\limits_{0 \neq \mathbf{x} \in \text{span}\{\mathbf{x}^{(i)},\cdots,\mathbf{x}^{(n)}\}} R(\mathbf{x})$.

关于特征值的估计—Rayleigh 商

再来证 $\lambda_i = \min\limits_{dim(W)=n+1-i} \max\limits_{0\neq \mathbf{x}\in W} R(\mathbf{x})$.

规范正交基

任给一个维数是 $n+1-i$ 的子空间 W, 设其一组基向量

为$\mathbf{z}_1, \cdots, \mathbf{z}_{n+1-i}$. $\forall \mathbf{x}\in W$, 有 $\mathbf{x} = \sum\limits_{j=1}^{n+1-i} (\mathbf{x}, \mathbf{z}_j)\mathbf{z}_j \equiv \sum\limits_{j=1}^{n+1-i} a_j \mathbf{z}_j$,

要想让 $(\mathbf{x}, \mathbf{x}^{(k)}) = 0$, $k = i+1, \cdots, n \Longrightarrow \sum\limits_{j=1}^{n+1-i} a_j(\mathbf{z}_j, \mathbf{x}^{(k)}) = 0$.

一共 $n-i$ 个方程 (约束), 但是有 $n+1-i$ 个自由度(系数), 因此必有非零解 $\{a_j\}_{j=1}^{n+1-i}$. 因此必存在 $0\neq \mathbf{x}\in W$, *s.t.* $(\mathbf{x}, \mathbf{x}^{(k)}) = 0$,

$k =$ [redacted] $\cdots, n$. 因而由 $\mathbf{x}^{(k)}$ 之正交性, 有 $\mathbf{x} = \sum\limits_{k=1}^{i} (\mathbf{x}, \mathbf{x}^{(k)})\mathbf{x}^{(k)}$.

i+1

关于特征值的估计—Rayleigh 商

这样 $A\mathbf{x} = \sum_{k=1}^{i} (\mathbf{x}, \mathbf{x}^{(k)})\lambda_k \mathbf{x}^{(k)}$

$$\Longrightarrow (A\mathbf{x}, \mathbf{x}) = \sum_{k=1}^{i} \lambda_k |(\mathbf{x}, \mathbf{x}^{(k)})|^2 \geq \lambda_i \sum_{k=1}^{i} |(\mathbf{x}, \mathbf{x}^{(k)})|^2 = \lambda_i(\mathbf{x}, \mathbf{x})$$

$$\Longrightarrow \max_{0\neq\mathbf{x}\in W} \frac{(A\mathbf{x}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})} \geq \lambda_i.$$

另外, 如果令 $W = \mathrm{span}\{\mathbf{x}^{(i)}, \cdots, \mathbf{x}^{(n)}\}$, 有 $dimW = n + 1 - i$,

且前面已证 $\lambda_i = \max_{0\neq\mathbf{x}\in W} \frac{(A\mathbf{x}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})}$. 故 3) 得证. ▷

关于特征值的估计—Rayleigh 商

推论 5.1

设 A 和 B 是两个 *Hermite* 阵, 特征值分别为 $\lambda_1^A \geq \lambda_2^A \geq \cdots \geq \lambda_n^A$ 和 $\lambda_1^B \geq \lambda_2^B \geq \cdots \geq \lambda_n^B$, 则任取算子范数 $\|\cdot\|$ 都有

$$\left|\lambda_i^A - \lambda_i^B\right| \leq \|A - B\|, \quad i = 1, \cdots, n.$$

◁ 由 Cauchy-Schwartz 不等式有

$$((A-B)\mathbf{x}, \mathbf{x}) \leq \|(A-B)\mathbf{x}\|_2 \cdot \|\mathbf{x}\| \leq \|A-B\|_2 \cdot \|\mathbf{x}\|_2^2.$$

即 $(A\mathbf{x}, \mathbf{x}) \leq (B\mathbf{x}, \mathbf{x}) + \|A-B\|_2 \|\mathbf{x}\|_2^2 \Longrightarrow R^A(\mathbf{x}) \leq R^B(\mathbf{x}) + \|A-B\|_2$. 类似可以得到 $R^B(\mathbf{x}) \leq R^A(\mathbf{x}) + \|B-A\|_2$. 利用前面定理结论有

$$\lambda_i^A \leq \max_{\mathbf{x} \in U_i^B} R^A(\mathbf{x}) \leq \max_{\mathbf{x} \in U_i^B} R^B(\mathbf{x}) + \|A-B\|_2 = \lambda_i^B + \|A-B\|_2$$

以及类似的 $\lambda_i^B \leq \lambda_i^A + \|B-A\|_2 \Longrightarrow \left\|\lambda_i^A - \lambda_i^B\right\| \leq \|A-B\|_2$.

关于特征值的估计—Rayleigh 商

因为 A, B 为 Hermite 阵, 因此对任意一种算子范数都有

$$\|A-B\|_2 = \rho(A-B) \le \|A-B\|.$$

该推论结论得证. $\triangleright$

推论 5.2

设 A 是 *Hermite* 阵, 特征值为 $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$. 假设 $\{\tilde{a}_{ii}\}_{i=1}^n$ 为 $\{a_{ii}\}_{i=1}^n$ 按照从大到小重排的结果, 即有 $\tilde{a}_{11} \ge \tilde{a}_{22} \ge \cdots \ge \tilde{a}_{nn}$. 则有

$$\|\lambda_i - \tilde{a}_{ii}\| \le \sqrt{\sum_{j\ne k} |a_{jk}|^2}, \quad i = 1, \cdots, n.$$

$\triangleleft$ 在前面推论中取 $B = \operatorname{diag}(a_{ii})$, 取上面的范数为 $\|\cdot\|_2$, 再利用

$$\|A\|_2^2 \le \sum_{j=1}^n \sum_{k=1}^n |a_{jk}|^2, \quad \|A-B\|_2 \le \sqrt{\sum_{j\ne k} |a_{jk}|^2}$$

即得本推论结果. $\triangleright$

目录

模最大特征值的计算—幂法

如果只是求 A 的某一个特征值及相应特征向量, 用幂法和反幂法是很方便的算法.

严格大，可对角化

设 $A \in \mathbb{C}^{n\times n}$ 的特征值为 $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$, 相应的特征向量 $\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(n)}$ 构成 $\mathbb{C}^n$ 的一组基. 这说明 A 可对角化, 即存在 P 非奇异, 使得 $P^{-1}AP = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$.

这样 $\forall \mathbf{v} \in \mathbb{C}^n$, 有 $\mathbf{v} = \sum_{j=1}^{n} \alpha_j \mathbf{x}^{(j)}$.

因而 $$A^k\mathbf{v} = \sum_{j=1}^{n} \lambda_j^k \alpha_j \mathbf{x}^{(j)} = \lambda_1^k \left[\alpha_1 \mathbf{x}^{(1)} + \sum_{j=2}^{n} \left(\frac{\lambda_j}{\lambda_1}\right)^k \alpha_j \mathbf{x}^{(j)}\right]$$

由于 $|\frac{\lambda_j}{\lambda_1}| < 1, j = 2, \cdots, n \Longrightarrow k \to +\infty$ 时, $A^k\mathbf{v} \sim \lambda_1^k \alpha_1 \mathbf{x}^{(1)}$.

这样, 如果令 $\mathbf{v}_k = A^k\mathbf{v}$, 就有 $\mathbf{v}_{k+1} \sim \lambda_1^{k+1}\alpha_1\mathbf{x}^{(1)} \sim \lambda_1 \mathbf{v}_k$

模最大特征值的计算—幂法

如果取其一个非零分量, 就有 $\frac{(\mathbf{v}_{k+1})_i}{(\mathbf{v}_k)_i} \sim \lambda_1$. 这样可以得到以下幂法迭代算法:

算法 5.1 (幂法)

取 $\mathbf{v}_0 \neq 0$ (例如可取 $v_0 = (1, \cdots, 1)^T$), 令 $\mathbf{u}_0 = \mathbf{v}_0$. 对 $k = 1, 2, \cdots$

- $\mathbf{v}_k = A\mathbf{u}_{k-1}$. 取 $m_k = (\mathbf{v}_k)_{i_0}$ s.t. $|m_k| = \|\mathbf{v}_k\|_\infty$ (模最大的元素), 令 $\mathbf{u}_k = \frac{\mathbf{v}_k}{m_k}$ (归一化一下, 以免浮点计算时溢出)
- 判断: $|m_k - m_{k-1}| < \varepsilon$ 则停止迭代, 输出 $\mathbf{x}^{(1)} \approx \mathbf{u}_k$, $\lambda_1 \approx m_k$.

因为关注的是方向而不是长度

- 否则 $k = k + 1$, 继续.

一般来说, 如果 $\frac{|\lambda_2|}{|\lambda_1|} \ll 1$ 则收敛较快; 否则收敛较慢.

由于舍入误差的影响, 一般总会有 $\alpha_1 \neq 0$.

模最大特征值的计算—幂法

Aitken 加速: 类似地, 这里我们也可以采用 Aitken 加速技巧. 即如果我们得到了 m_k, m_{k+1}, m_{k+2}, 可以令

$$\widetilde{\mathbf{m}}_k = \frac{m_k m_{k+2} - m_{k+1}^2}{m_{k+2} - 2m_{k+1} + m_k}$$ 会更接近于 λ_1.

例 5.3

取 $A = \begin{pmatrix} -4 & 14 & 0 \\ -5 & 13 & 0 \\ -1 & 0 & 2 \end{pmatrix}$, 特征值为 6, 3, 2. 取 $\mathbf{v}_0 = (1, 1, 1)^T$,

计算有 $m_1 = 10,\ m_2 = 7.2,\ \cdots, m_{12} = 6.000837$

加速后则得 $\widetilde{m}_1 = 6.266\cdots, \widetilde{m}_2 = 6.062\cdots, \widetilde{m}_{10} = 6.0000009\cdots$

特征向量可以类似(按分量)加速.

模最小特征值的计算—反幂法

设 A 非奇异, 特征值满足 $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_2| > |\lambda_1| > 0$, 对 A^{-1} 用幂法迭代, 可以计算 λ_n^{-1} 及 $\mathbf{x}^{(n)}$:

算法 5.2 (反幂法)

取 $\mathbf{u}_0 = \mathbf{v}_0 \neq 0$ (例如可取 $v_0 = (1, \cdots, 1)^T$), 对 $k = 1, 2, \cdots$

- 求 $\mathbf{v}_k$ s.t. $A\mathbf{v}_k = \mathbf{u}_{k-1}$. 取 $m_k = (\mathbf{v}_k)_{i_0}$ s.t. $|m_k| = \|\mathbf{v}_k\|_\infty$ (模最大的元素), 令 $\mathbf{u}_k = \frac{\mathbf{v}_k}{m_k}$ (归一化一下, 以免浮点计算时溢出)
- 判断: $|m_k - m_{k-1}| < \varepsilon$ 则停止迭代, 输出 $\mathbf{x}^{(n)} \approx \mathbf{u}_k$, $\lambda_n \approx m_k^{-1}$.
- 否则 $k = k + 1$, 继续.

模最小特征值的计算—反幂法

特别，如果想求某个值 p 附近的特征值 λ^*, 可以对 $(A-pI)^{-1}$ 用幂法: $(A-pI)^{-1}$ 的特征值为 $\frac{1}{\lambda_1 - p}, \cdots, \frac{1}{\lambda_n - p}$.

(当λ^*不是重根时)一般有 $\frac{1}{\lambda^* - p}$ 为模最大的, 且 p 越靠近 λ^*, 会收敛越快!

例 5.4

对上例的 A, 若取 $p = 5.5$, 对 $B = (A - pI)$ 用反幂法, 迭代6次就得到 $m_6 = 1.99956\cdots \longrightarrow \frac{1}{\lambda^* - p} \Longrightarrow \lambda^* \approx 6.0001097$.

再用 Aitken 加速可得 $\bar{\lambda} = 6.00000010\cdots$ 有7位有效数字了.

目录

Householder变换与QR分解

对任何 $A \in \mathbb{C}^{n\times n}$, 存在酉矩阵 Q 和上三角阵 R s.t. $A = QR$. 该分解可以利用 Householder 变换实现.

定义 5.4 (Householder 变换)

对 $\mathbf{v} \in \mathbb{C}^n$, $\mathbf{v}^*\mathbf{v} = 1$ (单位长度向量). 令 $H = I - 2\mathbf{v}\cdot\mathbf{v}^* \in \mathbb{C}^{n\times n}$ 称为 Householder 矩阵.

引理 5.1

Householder 矩阵为酉矩阵, 且 $H = H^*$.

◁ 直接计算有 $H^* = I - 2(\mathbf{v}\cdot\mathbf{v}^*)^* = I - 2\mathbf{v}\cdot\mathbf{v}^* = H$.

$HH^* = H^2 = (I - 2\mathbf{v}\cdot\mathbf{v}^*)(I - 2\mathbf{v}\cdot\mathbf{v}^*) = I - 4\mathbf{v}\mathbf{v}^* + 4\mathbf{v}(\mathbf{v}^*\mathbf{v})\mathbf{v}^* = I$

▷

Householder变换与QR分解

Householder 变换为一个镜面反射：$H = I - 2\mathbf{v}\cdot\mathbf{v}^*$

$H\mathbf{x}$ 相当于把 $\mathbf{x}$ 关于与 $\mathbf{v}$ 垂直的平面做了一个镜面反射:

令 $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 \equiv \mathbf{x}_1 + (\mathbf{v}^*\mathbf{x})\mathbf{v}$, 即 $\mathbf{x}_2 = (\mathbf{v}^*\mathbf{x})\mathbf{v}$ 表示向量 $\mathbf{x}$ 在 $\mathbf{v}$ 上的投影, 而 $\mathbf{x}_1 \perp \mathbf{v}$, 即 $\mathbf{v}^*\mathbf{x}_1 = 0$.

直接计算有

$$H\mathbf{x} = (I - 2\mathbf{v}\cdot\mathbf{v})\mathbf{x} = \mathbf{x} - 2\mathbf{v}\mathbf{v}^*\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2 - 2\mathbf{v}\mathbf{v}^*(\mathbf{v}^*\mathbf{x})\mathbf{v} = \mathbf{x}_1 - \mathbf{x}_2$$

而且 $(H\mathbf{x}, H\mathbf{x}) = \mathbf{x}_1^*\mathbf{x}_1 + \mathbf{x}_2^*\mathbf{x}_2 = (\mathbf{x}, \mathbf{x})$, 即 Householder 变换保持向量长度不变.

Householder变换与QR分解

下面我们看如何利用一系列Householder变换把 A 变成一个上三角阵. 记 $A = (\vec{a}_1 \cdots \vec{a}_n)$,

并设 $\vec{a}_1 = \begin{pmatrix} a_{11} \\ \vdots \\ a_{1n} \end{pmatrix}$. 欲构造 H_1 使得 $H_1 \vec{a}_1 = \mu_1 \vec{e}_1 = \mu_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$.

因保持长度不变, 有 $|\mu_1|^2 = \vec{a}_1^{\,*} \vec{a}_1 = \sum_{i=1}^{n} |a_{i1}|^2 \equiv \sigma_1^2$, 即 $\sigma_1 = |\vec{a}_1|$.

由镜面反射的提示, 欲构造 $H_1 = I - 2\mathbf{v} \cdot \mathbf{v}^*$ 使得 $H_1 \vec{a}_1 = \mu_1 \vec{e}_1$,

则应该取 $\mathbf{v} = \dfrac{\vec{a}_1 - \mu_1 \vec{e}_1}{\left\| \vec{a}_1 - \mu_1 \vec{e}_1 \right\|_2}$.

Householder变换与QR分解

令 $\mathbf{u}_1 = \vec{a}_1 - \mu_1 \vec{e}_1$, 即 $\mathbf{v}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|_2}$. 为了减少舍入误差影响, 自然希望 $\|\mathbf{u}_1\|_2$ 越大越好. 直接计算有

$$\begin{aligned}
\mathbf{u}_1^*\mathbf{u}_1 &= \left(\vec{a}_1^{\,*} - \mu_1^* \vec{e}_1^{\,*}\right)\left(\vec{a}_1 - \mu_1 \vec{e}_1\right) \\
&= \vec{a}_1^{\,*}\vec{a}_1 + \mu_1^*\mu_1 \vec{e}_1^{\,*}\vec{e}_1 - (\mu_1^* a_{11} + \mu_1 a_{11}^*) \\
&= 2\sigma_1^2 - (\mu_1^* a_{11} + \mu_1 a_{11}^*).
\end{aligned}$$

这样当 $a_{11} \neq 0$ 时, 可取 $\mu_1 = -\frac{a_{11}}{|a_{11}|}\sigma_1$ 有 $\mathbf{u}_1^*\mathbf{u}_1 = 2\sigma_1(\sigma_1 + |a_{11}|)$.
当 $a_{11} = 0$ 时, 可取 $\mu_1 = \sigma_1$ 即可, 有 $\mathbf{u}_1^*\mathbf{u}_1 = 2\sigma_1^2$.

这样 $H_1 = I - 2\dfrac{\left(\vec{a}_1 - \mu_1 \vec{e}_1\right)\left(\vec{a}_1^{\,*} - \mu_1^* \vec{e}_1^{\,*}\right)}{2\sigma_1(\sigma_1 + |a_{11}|)} \Longrightarrow H_1 \vec{a}_1 = \mu_1 \vec{e}_1$.

Householder变换与QR分解

假如类似地进行了 m 次变换后, 有

$$H_m H_{m-1}\cdots H_1 A = A^{(m)} = \begin{pmatrix} R_m & * \\ 0 & \widetilde{A}_{n-m} \end{pmatrix}, \quad R_m\text{为 } m \text{ 阶上三角阵}.$$

可以构造 $n-m$ 阶 Householder 矩阵 $\widetilde{H}_{n-m}$ s.t. $\widetilde{H}_{n-m}\widetilde{A}_{n-m}$ 会把 $\widetilde{A}_{n-m}$ 的第一列变换成 $\tilde{\sigma}\tilde{\mathbf{e}}_1$ 的形式. 这样令

$$H_{m+1} = \begin{pmatrix} I_m & 0 \\ 0 & \widetilde{H}_{n-m} \end{pmatrix} \text{即可}.$$

归纳就有 $H_{n-1}\cdots H_1 A = R$ 为一个上三角阵.

令 $Q = H_1\cdots H_{n-1}$, 注意 $H_i^2 = H_i H_i^* = I$, 即 $Q^*Q = I$, 因此我们得到了 A 的“QR分解式” $A = QR$.

Givens 变换

定义 5.5 (Givens 变换)

$\forall\theta\in\mathbb{R}$, 记 $s=\sin\theta$, $c=\cos\theta$, $J=\begin{pmatrix} c & s \\ -s & c\end{pmatrix}$ 是一个正交阵.

$\forall\mathbf{x}\in\mathbb{R}^2$, $J\mathbf{x}$ 表示将向量 $\mathbf{x}$ 旋转 θ 角 (弧度). 推广到 n 维情形, 令

$$J(j,k;\theta)=\begin{pmatrix} 1 &&&&&&&&& \\ & \ddots &&&&&&&& \\ && 1 &&&&&&& \\ &&& c &&&& s && \\ &&&& 1 &&&&& \\ &&&&& \ddots &&&& \\ &&&&&& 1 &&& \\ &&& -s &&&& c && \\ &&&&&&&& 1 && \\ &&&&&&&&& \ddots & \\ &&&&&&&&&& 1 \end{pmatrix} \text{— Givens变换.}$$

Givens 变换

显然上面定义的 $J(j,k;\theta)$ 是一个正交阵. 对 $\mathbf{x}\in\mathbb{R}^n$, 如果 $\mathbf{y}=J(j,k;\theta)\mathbf{x}$, 相当于在 $(\mathbf{e}_j,\mathbf{e}_k)$ 张成的平面上做了一个旋转, 即 $\mathbf{y}$ 的分量为

$$\begin{cases} y_j = cx_j + sx_k, & y_k = -sx_j + cx_k, \\ y_i = x_i, & \text{对其他 } i \neq j,k. \end{cases}$$

如果想让 $\mathbf{y}$ 的分量 $y_k=0$ (自然我们假设 $x_k\neq 0$), 只需选择 θ *s.t.*

$$c=\cos\theta=\frac{x_j}{\sqrt{x_j^2+x_k^2}},\qquad s=\sin\theta=\frac{x_k}{\sqrt{x_j^2+x_k^2}}$$

我们可以看到:

$J(j,k;\theta)A$ 只改变 A 的 j,k 行, $AJ^T(j,k;\theta)$ 只改变 A 的 j,k 列, $J(j,k;\theta)AJ^T(j,k;\theta)$ 只改变 A 的第 j,k 行和 j,k 列 (为相似变换).

Shur 分解

引理 5.2

对 $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{p\times p}$, $X \in \mathbb{C}^{n\times p}$, s.t.

$$AX = XB, \quad \mathrm{rank}(X) = p \quad (\text{即有} p \le n)$$

则存在酉矩阵 $U \in \mathbb{C}^{n\times n}$ s.t.

$$U^*AU = T = \begin{array}{cc} \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix} & \begin{array}{l} p \\ n-p \end{array} \end{array}$$

其中 $T_{11} \in \mathbb{C}^{p\times p}$ 的特征值为 A 和 B 的特征值集合之交, 即 $\Lambda(T_{11}) = \Lambda(A) \bigcap \Lambda(B)$.

$\lhd$ 对 $T = \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix}$, 设 $T \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} = \lambda \begin{array}{cc} \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} & \begin{array}{l} p \\ n-p \end{array} \end{array}$

Shur 分解

易见 $\Lambda(T)=\Lambda(T_{11})\bigcup\Lambda(T_{22})$:

$$\left.\begin{array}{l}\ulcorner 若\ \mathbf{x}_2=0, 则\ T_{11}\mathbf{x}_1=\lambda\mathbf{x}_1 \Longrightarrow \lambda\in\Lambda(T_{11}) \\ \ 若\ \mathbf{x}_2\neq 0, 则\ T_{22}\mathbf{x}_2=\lambda\mathbf{x}_2 \Longrightarrow \lambda\in\Lambda(T_{22})\end{array}\right\}$$

$\Longrightarrow \Lambda(T)\subset\Lambda(T_{11})\bigcup\Lambda(T_{22})$

反之, 若 $T_{11}\mathbf{x}_1=\lambda\mathbf{x}_1 \Longrightarrow T\begin{pmatrix}\mathbf{x}_1\\0\end{pmatrix}=\lambda\begin{pmatrix}\mathbf{x}_1\\0\end{pmatrix}$

若 $T_{22}\mathbf{x}_2=\lambda\mathbf{x}_2$ 且 λ 不是 T_{11} 的特征值.

$\Longrightarrow \exists\mathbf{x}_1$ *s.t.* $(T_{11}-\lambda)\mathbf{x}_1=-T_{12}\mathbf{x}_2$

$\Longrightarrow T\begin{pmatrix}\mathbf{x}_1\\\mathbf{x}_2\end{pmatrix}=\lambda\begin{pmatrix}\mathbf{x}_1\\\mathbf{x}_2\end{pmatrix} \Longrightarrow \Lambda(T_{11})\bigcup\Lambda(T_{22})\subset\Lambda(T).\lrcorner$

现对 X 做 QR 分解, 即有酉矩阵 $Q\in\mathbb{C}^{n\times n}$ *s.t.* $X=Q\begin{pmatrix}R\\0\end{pmatrix}\begin{array}{l}p\\n-p\end{array}$

Shur 分解

代入 $AX = XB$ 中 $AQ\begin{pmatrix} R \\ 0 \end{pmatrix} = Q\begin{pmatrix} R \\ 0 \end{pmatrix}B.$

设 $Q^*AQ = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \Longrightarrow \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix}\begin{pmatrix} R \\ 0 \end{pmatrix} = \begin{pmatrix} R \\ 0 \end{pmatrix}B$

由条件$rank(X) = p \Longrightarrow R$ 非奇异. 而 $T_{21}R = 0 \Longrightarrow T_{21} = 0.$

又 $T_{11}R = RB \Longrightarrow T_{11} = RBR^{-1} \Longrightarrow \Lambda(T_{11}) = \Lambda(B).$

前面已证 $\Lambda(A) = \Lambda(T) = \Lambda(T_{11}) \bigcup \Lambda(T_{22})$, 结合上式

$\Longrightarrow \Lambda(T_{11}) = \Lambda(A) \bigcap \Lambda(B).$ $\triangleright$

这样我们可以证明如下 Shur 分解定理

Shur 分解

定理 5.8 (Shur 分解)

对任何 $A \in \mathbb{C}^{n\times n}$, 存在酉矩阵 U s.t. $U^*AU = R$ 为上三角阵.

$\triangleleft$ 可用归纳法证明. $n = 1$ 显然, 可取 $U = I$ 即可.

设 $n-1$ 阶情形成立. 对n 阶方阵 A, 设 $A\mathbf{x} = \lambda\mathbf{x}$, 令 $B = \lambda$,

由上述引理, 即存在酉矩阵 P s.t.

$$P^*AP = \begin{array}{c} \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix} \end{array} \begin{array}{l} 1 \\ n-1 \end{array}$$

对 $n-1$ 阶方阵 T_{22} 用归纳假设,

即有酉矩阵 $Q \in \mathbb{C}^{(n-1)\times(n-1)}$ s.t. $Q^*T_{22}Q = \widetilde{R}$ 为上三角阵.

这样令 $U = P\begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix}$, 便有 $U^*AU = \begin{pmatrix} \lambda & T_{12}Q \\ 0 & \widetilde{R} \end{pmatrix}$ 为上三角阵.

$\triangleright$

实 Shur 分解

对于实矩阵 $A \in \mathbb{R}^{n\times n}$, 有如下实 Shur 分解:

定理 5.9

对任何 $A \in \mathbb{R}^{n\times n}$, 存在正交矩阵 $Q \in \mathbb{R}^{n\times n}$ s.t.

$$Q^T A Q = \begin{pmatrix} R_{11} & \cdots & R_{1m} \\ & \ddots & \vdots \\ 0 & & R_{mm} \end{pmatrix} \text{为分块上三角阵.}$$

其中 R_{ii} 为 1×1 或者 2×2 的方阵.

一个 2×2 的子块会对应一对互为共轭的复特征值.

目录

Rayleigh商加速

对于Hermite(实对称)矩阵, 我们知道其特征值均为实数. 即存在酉(正交)矩阵 U s.t.

$$U^*AU = \mathrm{diag}(\lambda_1, \cdots, \lambda_n).$$

此外我们有关于 A 的 Frobenius 范数的性质为

$$\|A\|_F^2 = \sum_{i,j=1}^{n} |a_{ij}|^2 = tr(A^*A) = \sum_{k=1}^{n} |\lambda_k|^2.$$

假设 $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|$, 则对 A 可用幂法求 λ_1 及其对应的特征向量 $\mathbf{x}^{(1)}$. 但是由前面分析可知, $m_k = \lambda_1[1 + \mathcal{O}(|\frac{\lambda_2}{\lambda_1}|^k)]$, 收敛得不快.

对于Hermite(实对称)矩阵我们可以使用 Rayleigh 商加速.

Rayleigh商加速

设 $A\mathbf{x}^{(i)} = \lambda_i \mathbf{x}^{(i)}$, 任取 $\mathbf{v}^{(0)} = \sum_{i=1}^{n} \alpha_i \mathbf{x}^{(i)}$. 因为 A 为 Hermite (实对称) 阵, 我们已证 $\mathbf{x}^{(i)}$ 为互相正交向量, 故可无妨设 $\{\mathbf{x}^{(i)}\}_{i=1}^{n}$ 为正交规范基. 这样对应于 $\mathbf{v}^{(k)} = A^k \mathbf{v}^{(0)}$ 的 Rayleigh 商为

$$
\begin{aligned}
R(\mathbf{v}^{(k)}) &= \frac{(A\mathbf{v}^{(k)}, \mathbf{v}^{(k)})}{(\mathbf{v}^{(k)}, \mathbf{v}^{(k)})} = \frac{(A^{k+1}\mathbf{v}^{(0)}, A^k\mathbf{v}^{(0)})}{(A^k\mathbf{v}^{(0)}, A^k\mathbf{v}^{(0)})} = \frac{\sum_{j=1}^{n} |\alpha_j|^2 \lambda_j |\lambda_j|^{2k}}{\sum_{j=1}^{n} |\alpha_j|^2 |\lambda_j|^{2k}} \\
&= \lambda_1 \left[1 + \mathcal{O}\left(\left| \frac{\lambda_2}{\lambda_1} \right|^{2k} \right) \right]
\end{aligned}
$$

收敛速度加快了一倍.

Rayleigh商加速

另一方面, 如果 μ 是 λ_1 的一个好的近似值时, 可对 $(A-\mu I)$ 用反幂法, 这样可以快速迭代求出 $\mathbf{x}^{(1)}$ 的一个好近似. 即得以下算法

算法 5.3

取 $\mathbf{v}^{(0)}$ *s.t.* $\|\mathbf{v}^{(0)}\|_2 = 1$. 对 $k = 0, 1, \cdots$

1. $\mu_k = R(\mathbf{v}^{(k)}) = \dfrac{(A\mathbf{v}^{(k)}, \mathbf{v}^{(k)})}{(\mathbf{v}^{(k)}, \mathbf{v}^{(k)})}$;
2. 求解 $(A-\mu_k I)\mathbf{y}^{(k+1)} = \mathbf{v}^{(k)}$, 得到 $\mathbf{y}^{(k)}$;
3. $\mathbf{v}^{(k+1)} = \mathbf{y}^{(k)} / \|\mathbf{y}^{(k)}\|_2$.

可以验证, 该算法至少会二次收敛. 如果 A 为 Hermite 阵可以达到三阶收敛.

Jacobi 方法

对 Hermite 阵 A, 希望通过一系列相似变换, 让其收敛于对角阵, 从而得到其所有特征值. 所用正交变换即 Givens 变换.

记 A 的非对角线元素的平方和为 $N(A) = \sum_{i \neq j} |a_{ij}|^2$.

希望通过相似变换 $JAJ^{-1} = B$ 减少非对角元 $(N(B) < N(A))$:

$$
\begin{cases}
b_{ij} = a_{ij}, \quad (i \neq k, l;\ j \neq k, l;) \\
b_{kk} = a_{kk}\cos^2\theta + a_{kl}\sin 2\theta + a_{ll}\sin^2\theta \\
b_{ll} = a_{kk}\sin^2\theta - a_{kl}\sin 2\theta + a_{kk}\cos^2\theta \\
b_{kl} = b_{lk} = a_{kl}\cos 2\theta + \frac{1}{2}(a_{ll} - a_{kk})\sin 2\theta \\
b_{ik} = b_{ki} = a_{ik}\cos\theta + a_{il}\sin\theta \quad (i \neq k, l) \\
b_{il} = b_{li} = -a_{ik}\sin\theta + a_{il}\cos\theta \quad (i \neq k, l)
\end{cases}
$$

Jacobi 方法

我们先证明一个引理:

引理 5.3

$\forall A \in \mathbb{C}^{n\times n}$, Frobenius 范数 $\|A\|_F = \left(\sum_{i,j=1}^{n} |a_{ij}|^2\right)^{1/2}$ 在酉相似变换下保持不变.

$\triangleleft$ 首先易见 $tr(AB) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{n} b_{ji}a_{ij} = tr(BA)$.

特别 $tr(AA^*) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}a_{ij}^* = \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2 = \|A\|_F^2$.

这样任取酉矩阵 Q ($Q^*Q = I$): $\|Q^*AQ\|_F^2 = tr(Q^*AQ \cdot Q^*A^*Q)$

$= tr(Q^*A \cdot A^*Q) = tr(A^*Q \cdot Q^*A) = tr(A^*A) = \|A\|_F^2$. $\triangleright$

Jacobi 方法

利用上述引理, 以及Givens矩阵 $J(k,l;\theta)$ 为正交(酉)阵, 因此 $\|B\|_F^2 = \|A\|_F^2 \Longrightarrow$

$$N(B) = \|B\|_F^2 - \sum_{i=1}^{n} |b_{ii}|^2 = \|A\|_F^2 - \sum_{i\neq k,l} |a_{ii}|^2 - |b_{kk}|^2 - |b_{ll}|^2$$

$$= N(A) + |a_{kk}|^2 + |a_{ll}|^2 - |b_{kk}|^2 - |b_{ll}|^2 = N(A) + 2|b_{kl}|^2 - 2|a_{kl}|^2$$

要想让 $N(B)$ 尽可能小, 即让 $b_{kl} = 0$. 那么

$$\begin{cases} \text{当 } a_{kk} \neq a_{ll} \text{ 时}, & \tan 2\theta = \dfrac{2a_{kl}}{a_{kk} - a_{ll}} \\ \text{当 } a_{kk} = a_{ll} \text{ 时}, & \cos 2\theta = 0 \Longrightarrow \theta = \frac{\pi}{4} \end{cases}$$

有了 $\tan 2\theta$ 便可利用三角公式给出 $\cos 2\theta, \sin\theta, \cos\theta$.

Jacobi 经典算法

这样, 当 $a_{kl} \neq 0$ 时, 经过一次 Givens 相似变换, 就使得

$$N(B) = N(A) - 2|a_{kl}|^2, \quad 且\ b_{kl} = b_{lk} = 0.$$

这样如果扫描 A 找到绝对值最大的非对角元 a_{kl}, 确定 $J(k, l; \theta)$ 做相似变换 $B = JAJ^{-1}$, 就可以使非对角平方和尽可能的小!

算法 5.4 (经典Jacobi方法)

记 $A^{(1)} = A$. 确定一个阈值 $\varepsilon > 0$. 对 $m = 1, 2, \cdots$

1. 扫描一遍 $a_{ij}^{(m)}$ $(i \neq j)$, 找出绝对值最大非对角元 $|a_{kl}^{(m)}| = \max_{i \neq j} |a_{ij}^{(m)}|$;
2. 如果 $|a_{kl}^{(m)}| < \varepsilon$ 则停止变换, 否则继续;
3. 确定 $J(k, l; \theta)$, 做变换 $A^{(m+1)} = JA^{(m)}J^{-1}$.

Jacobi 经典算法

定理 5.10 (Jacobi 经典算法)

上述经典*Jacobi*算法收敛, 即 $A^{(m)}$ 会收敛到对角阵.

$\triangleleft$ 只需证明 $N(A^{(m)}) \to 0$: (因为每次 $|a_{kl}|$ 为最大非对角元)

$$\Longrightarrow N(A) \leq (n^2-n)\max_{i\neq j}|a_{ij}|^2 = (n^2-n)|a_{kl}|^2$$

$$即\ |a_{kl}|^2 = \max_{i\neq j}|a_{ij}|^2 \geq \frac{N(A)}{n^2-n}$$

由前面经典Jacobi算法, 对 $B = JAJ^T$,

$$\Longrightarrow N(B) = N(A) - 2|a_{kl}|^2 \leq q \cdot N(A)$$

其中 $q = 1 - \frac{2}{n(n-1)} \in [0,1)$,　当$n \geq 2$时 ($n=1$ 当然就不用算了).

这样就有 $N(A^{(m+1)}) \leq q^m N(A) \to 0$. $\triangleright$

Jacobi 过关法

显然，上面经典算法每次扫描 A 找到绝对值最大的非对角元 a_{kl} 后, 只做一次相似变换. 这样搜索上花的时间过多. 为提高效率, 可以用如下过关法:

算法 5.5 (Jacobi过关法)

记 $A^{(1)} = A$. 确定一个阈值 $\delta_1 > 0$, 如 $\delta_1 = \sqrt{N(A)}/n$. 取 $m = 1$ 及选取一个 $\varepsilon > 0$.

1. 扫描 a_{ij} $(i \neq j)$, 只要非对角元 $|a_{kl}| > \delta_m$ 就确定 $J(k, l; \theta)$, 做相似变换 JAJ^{-1}.
2. 扫描完一遍 A 的非对角元后, 令 $\delta_{m+1} = \delta_m/n$, $m = m + 1$.
3. 如果 $\delta_m > \varepsilon$ 就继续, 否则停止迭代.

Jacobi 方法举例

例 5.5

设 $A = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$, 显然 $a_{12} = a_{21} = a_{23} = a_{32} = -1$ 为绝对值最大非对角元. 取 $(k,l) = (1,2)$, 因为 $a_{11} = a_{22}$, 故 $\theta = \frac{\pi}{4}$, 即 $\cos 2\theta = 0$, $\sin\theta = \cos\theta = \frac{\sqrt{2}}{2}$. 即

$$J_1 = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \Longrightarrow A^{(2)} = J_1 A J_1^T = \begin{pmatrix} 1 & 0 & -\frac{\sqrt{2}}{2} \\ 0 & 3 & -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 2 \end{pmatrix}.$$

这样再取 $(k,l) = (1,3)$,

$$\tan 2\theta = \frac{2a_{kl}}{a_{kk}-a_{ll}} = \frac{-\sqrt{2}}{1-2} = \sqrt{2} \Longrightarrow \cos\theta = \sqrt{\frac{3+\sqrt{3}}{6}}, \sin\theta = \sqrt{\frac{3-\sqrt{3}}{6}}.$$

Jacobi 方法举例

即有 $J_2 = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \Longrightarrow A^{(3)} = J_2 A^{(2)} J_2^T \cdots\cdots$

直到 $A^{(10)} = \begin{pmatrix} 0.58578 & 0.00000 & 0.00000 \\ 0.00000 & 2.00000 & 0.00000 \\ 0.00000 & 0.00000 & 3.41421 \end{pmatrix}$. 得到三个特征值.

且记 $Q = J_1^T \cdots J_9^T = \begin{pmatrix} 0.50000 & 0.70710 & 0.50000 \\ 0.70710 & 0.00000 & -0.70710 \\ 0.50000 & -0.70710 & 0.50000 \end{pmatrix}$

$\equiv \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{pmatrix}$ 为相应特征向量. □

目录

试图用酉相似变换化为Shur分解形式

假如要计算所有特征值, 按照前面(反)幂法的思想则计算太慢.

我们希望通过一系列酉相似变换将一般矩阵化为 Shur 分解形式, 自然就得到了所有特征值及相应特征向量. 对于实矩阵可以通过正交相似变换化为实 Shur 分解的形式, 也可以得到实或复共轭对特征值及相应特征向量.

为了达到上述目的, 也为了计算过程中节约计算量, 我们先想法把一般矩阵化为一种特殊形式 (这种形式会在上述酉变换过程中保持不变).

这种特殊矩阵叫上 Hessenberg 矩阵.

上 Hessenberg 矩阵

定义 5.6 (上 Hessenberg 矩阵)

称 $A=(a_{ij})\in\mathbb{C}^{n\times n}$ 为上 *Hessenberg* 矩阵, 是指如果有

$$a_{jk}=0,\quad \forall 1\le k\le j-2,\ j=3,4,\cdots,n$$

也就是说A的严格下三角部分中除了次对角线元素外, 其余都是*0:*

$$A=\begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & \cdots & a_{2n} \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{n,n-1} & a_{nn} \end{pmatrix}.$$

如果 A 的下次对角线元素均不为零, 则称为不可约上 *Hessenberg* 矩阵.

上 Hessenberg 矩阵

利用 Householder 变换, 可把矩阵化为 Hessenberg 矩阵形式.

对于 $A = \begin{pmatrix} a_{11} & * \\ \mathbf{u}_1 & A_{22} \end{pmatrix}$, 这里 $\mathbf{u}_1 \in \mathbb{C}^{n-1}$, $A_{22} \in \mathbb{C}^{(n-1)\times(n-1)}$.

用前面方法求得Householder矩阵 $\widetilde{H}_1 = I - 2\mathbf{v}_1\mathbf{v}_1^* \in \mathbb{C}^{(n-1)\times(n-1)}$,

s.t. $\widetilde{H}_1\mathbf{u}_1 = \mu_1\widetilde{\mathbf{e}}_1 = \mu_1(1, 0, \cdots, 0)^T \in \mathbb{C}^{n-1}$, 其中 $|\mu_1|^2 = \|\mathbf{u}_1\|_2^2$.

这样如果令 $H_1 = \begin{pmatrix} 1 & 0 \\ 0 & \widetilde{H}_1 \end{pmatrix}$, 有 $AH_1^* = \begin{pmatrix} a_{11} & * \\ \mathbf{u}_1 & A_{22}\widetilde{H}_1^* \end{pmatrix}$,

即 $H_1AH_1^* = \begin{pmatrix} a_{11} & * \\ \mu_1\widetilde{\mathbf{e}}_1 & \widetilde{H}_1A_{22}\widetilde{H}_1^* \end{pmatrix} = \left(\begin{array}{cc} a_{11} & * \\ \begin{pmatrix} \mu_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} & \widetilde{H}_1A_{22}\widetilde{H}_1^* \end{array}\right)$.

上 Hessenberg 矩阵

重复上面的做法, 就有

$$H_{n-2}\cdots H_1 A H_1^* \cdots H_{n-2}^* = \begin{pmatrix} * & & & & \\ \mu_1 & * & & * & \\ 0 & \mu_2 & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \cdots & 0 & \mu_{n-2} & * \end{pmatrix} \equiv H$$

为上 Hessenberg 矩阵. 令酉矩阵 $Q = H_{n-2}\cdots H_1$, 有 $QAQ^* = H$.

QR 算法 (求所有特征值)

由前面 QR 分解定理, 我们知道 $\forall A \in \mathbb{C}^{n\times n}$, A 可以分解为一个酉矩阵 Q 与一个上三角阵之积: $A = QR$.

如果令 $B = RQ$, 即 $B = Q^*AQ$, 表明 $\Lambda(B) = \Lambda(A)$, 它们特征值一样. 这样可以定义如下 QR 算法:

算法 5.6 (QR算法)

令 $A_1 = A$. 对 $k = 1, 2, \cdots$

$$\begin{cases} \text{先做 QR 分解:} & A_k = Q_k R_k, \\ \text{然后交换位置相乘:} & A_{k+1} = R_k Q_k. \end{cases}$$

由QR算法有 $A_{k+1} = Q_k^* A_k Q_k = Q_k^* Q_{k-1}^* A_{k-1} Q_{k-1} Q_k = \widetilde{Q}_k^* A \widetilde{Q}_k$,
其中 $\widetilde{Q}_k = Q_1 \cdots Q_k$ 仍为酉矩阵.

QR 算法 (求所有特征值)

所以说由QR算法产生的矩阵序列总有 $\Lambda(A_k) = \Lambda(A)$.

另外如果还令 $\widetilde{R}_k = R_k R_{k-1} \cdots R_1$, 计算有

$$\begin{aligned}
\widetilde{Q}_k \widetilde{R}_k &= \widetilde{Q}_{k-1} Q_k R_k \widetilde{R}_{k-1} = \widetilde{Q}_{k-1} A_k \widetilde{R}_{k-1} \\
&\overset{A_k = \widetilde{Q}_{k-1}^* A \widetilde{Q}_{k-1}}{=} \widetilde{Q}_{k-1} (\widetilde{Q}_{k-1}^* A \widetilde{Q}_{k-1}) \widetilde{R}_{k-1} = A \widetilde{Q}_{k-1} \widetilde{R}_{k-1} \\
&= \cdots = A^{k-1} Q_1 R_1 = A^k.
\end{aligned}$$

如果上面产生的矩阵序列基本上收敛到Shur分解形式, 即(块)上三角阵, 则可以很容易求出特征值. 然后再用反幂法可以很快求出特征向量.

所以关键是什么时候 A_k 会收敛到(块)上三角阵?

QR 算法 (求所有特征值)

容易验证, 如果 A 是上 Hessenberg 矩阵，则此性质在QR算法中保持不变, 即所有 A_k 均为上 Hessenberg 阵. 由于可用酉相似变换将一般矩阵 A 化为上 Hessenberg 矩阵, 所以我们下面可无妨设 A 是上 Hessenberg 矩阵.

对于实矩阵情形, 我们引入"基本收敛"的概念:

定义 5.7 (基本收敛)

对矩阵序列 $\{A_k\}_{k=1}^{+\infty} \in \mathbb{R}^{n\times n}$, 若其对角元均收敛, 且严格下三角部分收敛到 0, 则称 $\{A_k\}_{k=1}^{+\infty}$ 基本收敛到上三角阵.

关于基本收敛性质，我们有以下引理: 因为只关心对角线元素

QR 算法 (求所有特征值)

引理 5.4

设 $M_k = Q_k R_k$, 其中 Q_k 为正交阵, R_k 为正对角元的上三角阵. 若 $M_k \to I$, 则有 $Q_k \to I$, $R_k \to I$. 这样的分解唯一

$\lhd$ 由 $M_k^T \cdot M_k = R_k^T Q_k^T Q_k R_k = R_k^T R_k$, 我们有

$$M_k \to I \Longrightarrow R_k^T R_k \to I.$$

设 $R_k = (r_{ij}^{(k)})$, 简单计算有 $R_k^T R_k$ 的第一行是

$$(r_{11}^{(k)} r_{11}^{(k)}, r_{11}^{(k)} r_{12}^{(k)}, \cdots, r_{11}^{(k)} r_{1n}^{(k)}),$$

即 $(r_{11}^{(k)})^2 \to 1$, 考虑到 R_k 对角元为正, 即 $r_{11}^{(k)} \to 1$, 且有 $r_{1j}^{(k)} \to 0$. 依次考虑 $R_k^T R_k$ 的第二行、第三行$\cdots$ 可以得到 $R_k \to I$, 即也有 $R_k^{-1} \to I$. 这样 $Q_k = M_k \cdot R_k^{-1} \to I$. $\rhd$

QR 算法 (求所有特征值)

最后我们可得关于QR算法收敛性的定理:

定理 5.11

设$A \in \mathbb{R}^{n\times n}$, 其特征值满足 $|\lambda_1| > \cdots > |\lambda_n| > 0$, 对应特征向量为 $\mathbf{x}^{(i)}, i = 1, \cdots, n$. 记 $X = [\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(n)}]$ 并设 $X^{-1} = LU$, 其中 L 为单位下三角阵, U 为上三角阵. 则 QR 算法产生的序列$\{A_k\}_{k=1}^{+\infty}$ 基本收敛到上三角阵: $\lim\limits_{k\to+\infty} a_{ii}^{(k)} = \lambda_i$, $i = 1, \cdots, n$.

$\lhd$ 根据假设, A 的特征值按模分离, 自然都是单根, 且都是实根. 这样 A 可以对角化, 即所有 $\mathbf{x}^{(i)}$ 是线性无关的. 自然 X^{-1} 存在.

由前面已证 $A_{k+1} = \widetilde{Q}_k^T A \widetilde{Q}_k$, 其中 $\widetilde{Q}_k = Q_1 \cdots Q_k$.

我们先来分析 $\widetilde{Q}_k$ 的性质, 然后再看 A_{k+1} 的收敛性.

QR 算法 (求所有特征值)

我们还证明了 $A^k = \widetilde{Q}_k \widetilde{R}_k$, 其中 $\widetilde{R}_k = R_k R_{k-1} \cdots R_1$ 仍是上三角阵. 我们下面将通过分析 X 的性质得到 A^k 的另一个 QR 分解形式, 以得到 $\widetilde{Q}_k$ 的性质.

对于 $X = [\mathbf{x}^{(1)} \cdots \mathbf{x}^{(n)}]$ 为A的特征向量组成的矩阵, 显然有

$$A = X\Lambda X^{-1}, \Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_n).$$

由 $X^{-1} = LU$: $\quad A^k = X\Lambda^k X^{-1} = X\Lambda^k LU = X(\Lambda^k L\Lambda^{-k})\Lambda^k U.$

设 $L = (l_{ij})$, 有 $\Lambda^k L \Lambda^{-k}$ 的元素为

L 为单位下三角阵

$$(\Lambda^k L\Lambda^{-k})_{ij} = \begin{cases} 0, & i < j; \\ 1, & i = j; \\ (\frac{\lambda_i}{\lambda_j})^k l_{ij}, & i > j \end{cases}$$

QR 算法 (求所有特征值)

由假设 $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n| > 0$, 因此 $|\frac{\lambda_i}{\lambda_j}| < 1$, 当 $i > j$.
即 $\lim\limits_{k\to+\infty} (\Lambda^k L \Lambda^{-k}) = I$. 记 $\Lambda^k L \Lambda^{-k} = I + E_k$, 有 $\lim\limits_{k\to+\infty} E_k = 0$.
且 $(E_k)_{ij} = \mathcal{O}(|\frac{\lambda_i}{\lambda_j}|^k)$, $i > j$; $(E_k)_{ij} = 0$, $i \leq j$.

即 $\|E_k\|_\infty \leq C \max\limits_{1\leq i\leq n-1} \left|\dfrac{\lambda_{i+1}}{\lambda_i}\right|^k$.

另外, 由 X 可逆, 可做 QR 分解: $X = QR$, 其中 Q 为正交阵, R 是对角元为正的上三角阵. 这样

$$A^k = QR(I + E_k)\Lambda^k U = Q(I + RE_kR^{-1})R\Lambda^k U.$$

因 $E_k \to 0$, 故 k 充分大时, $\|RE_kR^{-1}\|_\infty < 1$, 即 $I + RE_kR^{-1}$ 可逆.
于是也可以做唯一 QR 分解: $I + RE_kR^{-1} = Q^{(k)}R^{(k)} \to I$.

QR 算法 (求所有特征值)

由引理有 $Q^{(k)} \to I$, $R^{(k)} \to I$.

这样 $A^k = (QQ^{(k)}) \cdot (R^{(k)} R \Lambda^k U)$ 也是其一种 QR 分解.

当然此时还不能保证上三角阵部分对角元为正. 如果令

$$D_1 = \mathrm{diag}(\frac{\lambda_1}{|\lambda_1|}, \cdots, \frac{\lambda_n}{|\lambda_n|}), \quad D_2 = \mathrm{diag}(\frac{u_{11}}{|u_{11}|}, \cdots, \frac{u_{nn}}{|u_{nn}|}).$$

有 $A^k = (QQ^{(k)} D_2 D_1^k) \cdot (D_1^{-k} D_2 R^{(k)} R \Lambda^k U)$, 此时就可以保证上三角阵部分对角元为正了.

这样对照前面的QR算法中已证的 $A^k = \widetilde{Q}_k \widetilde{R}_k$ (可以规定 R_k 对角元为正), 由这种QR分解的唯一性, 我们有

$$\widetilde{Q}_k = QQ^{(k)} D_2 D_1^k, \quad \widetilde{R}_k = D_1^{-k} D_2 R^{(k)} R \Lambda^k U.$$

QR 算法 (求所有特征值)

再用上 $A = X\Lambda X^{-1} = QR\Lambda R^{-1}Q^{-1} \Longrightarrow$

$$\begin{aligned} A_{k+1} &= \widetilde{Q}_k^T A \widetilde{Q}_k \\ &= (D_2D_1^k)^T (Q^{(k)})^T Q^T \cdot (QR\Lambda R^{-1}Q^{-1}) \cdot QQ^{(k)} D_2 D_1^k \\ &= (D_2D_1^k)^T [(Q^{(k)})^T R\Lambda R^{-1} Q^{(k)}](D_2D_1^k). \end{aligned}$$

由 $Q^{(k)} \to I$, R 为上三角阵, 有

$$B_k \equiv (Q^{(k)})^T R\Lambda R^{-1} Q^{(k)} \to R\Lambda R^{-1} = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

由 $D_2D_1^k$ 为对角元是 ± 1 的对角阵, 因此 A_{k+1} 严格下三角阵部分收敛到 0, 且对角元 $(A_{k+1})_{ii} = (B_k)_{ii} \to (R\Lambda R^{-1})_{ii} = \lambda_i$, $i = 1, \cdots, n$. $\triangleright$

QR 算法 (求所有特征值)

对于复值情形, 在 A 可对角化的前提下, 假设

$$|\lambda_1| > \cdots > |\lambda_n|, \quad \mathrm{span}\{\mathbf{e}_1, \cdots, \mathbf{e}_m\} \cap \mathrm{span}\{\mathbf{x}_{m+1}, \cdots, \mathbf{x}_n\} = 0.$$

那么有 $A_{k+1} \to$ “$\Lambda +$ 严格上三角阵”, 即:

对角元收敛到 $\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$, 严格下三角部分收敛到零.

一般情形则很复杂. 对于有等模长特征值情形, 若 X^{-1} 仍然有 LU 分解, 则可以证明 A_{k+1} 基本收敛于块上三角阵. 每一块对应一个 A 的等模特征值 (p 个等模特征值对应一个 p 阶子矩阵).

此外还可以用类似于前面反幂法时用过的平移技巧进行加速.

分块三角阵由于阶数小计算起来也算方便