矩 阵 论

第 1 章　矩阵的相似变换

§1.1　线性代数回顾

- $m \times n$ 矩阵: $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}_{m\times n} = (a_{ij})_{m\times n}$

- 实矩阵: $a_{ij} \in \mathbb{R}$

  $\mathbb{R}^{m\times n} = m \times n$ 阶实矩阵的全体

- 复矩阵: $a_{ij} \in \mathbb{C}$

  $\mathbb{C}^{m\times n} = m \times n$ 阶复矩阵的全体

- 转置矩阵: $A^T = A' = (a_{ji})_{n\times m}$
- 共轭矩阵: $\overline{A} = (\overline{a_{ij}})_{m\times n}$
- 共轭转置矩阵: $A^H = \overline{A}^T = (\overline{a_{ji}})_{n\times m}$
- n 阶对角矩阵: $\mathrm{diag}(a_{11}, a_{22}, \cdots, a_{nn}) = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix}$
- 单位矩阵: $I_n = \mathrm{diag}(1, 1, \cdots, 1)$
- 数量矩阵: $\mathrm{diag}(a, a, \cdots, a) = aI_n$

- n 维列向量: $n \times 1$ 矩阵, 即 $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$

  $\mathbb{R}^n = n$ 维实列向量的全体

  $\mathbb{C}^n = n$ 维复列向量的全体

- n 维行向量: $1 \times n$ 矩阵, 即 $(a_1, a_2, \cdots, a_n)$
- n 阶方阵: $m = n$
- 方阵的行列式: $\det A = |A|$
  - $|AB| = |A||B| = |BA|$
  - n 个方程的 n 元齐次线性方程组有非零解 $\Longleftrightarrow$ 它的系数矩阵的行列式等于零.

矩阵的秩

设矩阵 $A=(a_{ij})_{m\times n}$, 则 $\mathrm{rank}(A)=$ 矩阵 A 的秩

$=$ 行 (或列) 向量组的极大线性无关组的向量个数

$=$ 不为零的子式的最高阶数.

- $\mathrm{rank}(AB)\leqslant\min\{\mathrm{rank}(A),\mathrm{rank}(B)\}$
- $\mathrm{rank}(A)=\mathrm{rank}(A^T)$
- 设 A 是实矩阵, 则 $\mathrm{rank}(AA^T)=\mathrm{rank}(A^TA)=\mathrm{rank}(A)$

可逆矩阵

- 矩阵 A 是可逆的 $\Longleftrightarrow$ 存在矩阵 B 使得 $AB = BA = I_n$. A^{-1} 表为 A 的逆矩阵
- 矩阵 A可逆 $\Longleftrightarrow |A| \neq 0$
- $A^* =$ 矩阵 A 的伴随矩阵. $A^*A = |A|I_n$
- 若 A 可逆, 则 $A^{-1} = \dfrac{1}{|A|}A^*$
- $(AB)^{-1} = B^{-1}A^{-1}$

矩阵的初等变换

矩阵 $A=(a_{ij})_{m\times n}$ 的初等行 (列) 变换:

(1) 交换两行 (列): 交换 i,j 两行 (列), 记作 $r_i \longleftrightarrow r_j$ $(c_i \longleftrightarrow c_j)$.

(2) 数 $k(k\neq 0)$ 乘某行 (列) 的所有元素: 第 i 行 (列) 乘 k, 记作 $r_i\times k$ $(c_i\times k)$.

(3) 把某一行 (列) 所有元素的 k(任意) 倍加到另一行 (列) 对应的元素上去, 第 j 行 (列) 的 k 倍加到第 i 行 (列) 上, 记作 r_i+kr_j (c_i+kc_j).

初等矩阵

单位矩阵 I_n 做一次初等变换所得矩阵称为初等矩阵. 则有如下三种初等矩阵:

$$(1)\ P_{ij} = \begin{bmatrix} 1 &&&&&&&&&& \\ & \ddots &&&&&&&&& \\ && 1 &&&&&&&& \\ &&& 0 &&&& 1 &&& \\ &&&& 1 &&&&&& \\ &&&&& \ddots &&&&& \\ &&&&&& 1 &&&& \\ &&& 1 &&&& 0 &&& \\ &&&&&&&& 1 && \\ &&&&&&&&& \ddots & \\ &&&&&&&&&& 1 \end{bmatrix},\ (2)\ D_i(k) = \begin{bmatrix} 1 &&&&&&&& \\ & \ddots &&&&&&& \\ && \ddots &&&&&& \\ &&& 1 &&&&& \\ &&&& k &&&& \\ &&&&& 1 &&& \\ &&&&&& \ddots && \\ &&&&&&& \ddots & \\ &&&&&&&& 1 \end{bmatrix} \quad (k \neq 0)$$

$$(3)\ \ T_{ij}(k) = \begin{bmatrix} 1 &&&&&& \\ & \ddots &&&&& \\ && 1 && k && \\ &&& \ddots &&& \\ &&&& 1 && \\ &&&&& \ddots & \\ &&&&&& 1 \end{bmatrix} \quad (k\ \text{任意})$$

定理 1.1

初等矩阵都是可逆的, 且其逆仍然为初等矩阵.

$$P_{ij}^{-1} = P_{ij},\ D_i(k) = D_i(\frac{1}{k}),\ T_{ij}(k)^{-1} = T_{ij}(-k).$$

容易验证, 对矩阵 A 做三种初等行 (列) 变换, 相当于在矩阵 A 左 (右) 乘相应的初等矩阵 ($\rightsquigarrow$ 左行右列).

由线性代数知任何矩阵 A 总可以通过初等行变换和交换两列的初等变换化为:

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & c_{1,r+1} & \cdots & c_{1n} \\ 0 & 1 & \cdots & 0 & c_{2,r+1} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & c_{r,r+1} & \cdots & c_{rn} \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

定理 1.2

任意矩阵 $A = (a_{ij})_{m\times n}$ 总可以通过初等变换将 A 化为如下形式的一个矩阵:

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

称之为矩阵 A 的标准型. 这里 $r = \operatorname{rank} A$.

矩阵的等价分类

定义 1.3

设 A, B 是 $m \times n$ 矩阵, 如果 A 经过有限次初等变换变成 B, 则称 A 与 B 是等价的.

两个 $m \times n$ 矩阵 A 与 B 等价 $\Longleftrightarrow$ 存在 m 阶可逆矩阵 P 与 n 阶可逆矩阵 Q, 使得 $B = PAQ$.

定理 1.4

(1) 设一个 $m \times n$ 矩阵 A 的秩为 $r(\neq 0)$, 则 A 等价于矩阵 $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$. 即存在 m 阶可逆矩阵 P 与 n 阶可逆矩阵 Q, 使得

$$PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

(2) 矩阵 A 与 B 等价 $\Longleftrightarrow \operatorname{rank} A = \operatorname{rank} B$.

分块矩阵

设矩阵 $A = (a_{ij})_{m\times}$, 在 A 的行列之间加一些线, 把矩阵分为若干小块, 例如:

$$A = \left[\begin{array}{c|cc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{array}\right]$$

此时 A 可以简单地记为

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

我们比较关心的是矩阵的乘法, 已知矩阵 $A=(a_{ij})_{m\times n}, B=(b_{ij})_{n\times s}$, 则 $AB=C=(c_{ij})_{m\times s}$. 分别对 A,B 分块, 例如:

$$A=\left[\begin{array}{cc|c} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{array}\right]=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B=\left[\begin{array}{cc} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \hline b_{31} & b_{32} \end{array}\right]=\begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}$$

则

$$AB=\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}=\begin{bmatrix} A_{11}B_{11}+A_{12}B_{21} \\ A_{21}B_{11}+A_{22}B_{21} \end{bmatrix}=\begin{bmatrix} C_{11} \\ C_{21} \end{bmatrix}$$

注记 1.5

A,B 矩阵可以做分块乘法的原则是: A 的列分法和 B 的行分法要一致.

§1.2 特征值和特征向量

定义 1.6

设 A 是 n 阶复方阵, 如果对复数 λ_0, 存在 n 维非零复列向量 α, 使得

$$A\alpha = \lambda_0\alpha$$

则称 λ_0 是矩阵 A 的一个**特征值**, α 称为 A 的属于特征值 λ_0 的一个**特征向量**.

注记 1.7

(1) 若 α 是矩阵 A 的属于特征值 λ_0 的特征向量, 则 $k\alpha$ $(k \neq 0)$ 仍是 A 的属于 λ_0 的特征向量.

(2) 矩阵 A 属于特征值 λ_0 的特征向量的线性组合仍是 A 的属于 λ_0 的特征向量.

特征矩阵和特征多项式

$\lambda_0(\in \mathbb{C})$ 是 A 的一个特征值, $\alpha(\in \mathbb{C}^n)$ 是 A 的属于 λ_0 的一个特征向量 $\Longleftrightarrow A\alpha = \lambda_0\alpha,\ \alpha \neq 0 \Longleftrightarrow (\lambda_0 I - A)\alpha = 0,\ \alpha \neq 0\ \Longleftrightarrow \alpha$ 是齐次线性方程组 $(\lambda_0 I - A)X = 0$ 的一个非零解 $\Longleftrightarrow$ 系数矩阵 $|\lambda_0 I - A| = 0$
$\Longleftrightarrow\ \lambda_0$ 是 $|\lambda I - A| = 0$ 一个根.

定义 1.8

设 A 是 n 阶方阵, 称 $\lambda I - A$ 为 A 的特征矩阵, $|\lambda I - A|$ 为 A 的特征多项式.

定理 1.9

设 A 是 n 阶复方阵, 则 (1): λ_0 是 A 的一个特征值 $\Longleftrightarrow \lambda_0$ 是 A 的特征多项式的一个根. 从而, A 有 n 个特征值 (计算重数).
(2): α 是 A 的属于特征值 λ_0 的一个特征向量 $\Longleftrightarrow \alpha$ 是 $(\lambda_0 I - A)X = 0$ 的一个非零解.

特征值和特征向量的求法

求 n 阶方阵 A 的特征值和特征向量的方法如下:

第一步: 计算行列式 $|\lambda I - A|$.

第二步: 求出 $|\lambda I - A|$ 的全部根, 即 A 的全部特征值.

第三步: 对于 A 的每一特征值 λ_i, 求 $(\lambda_i I - A)X = 0$ 的一个基础解系 $\eta_1, \eta_2 \cdots, \eta_t$. 则 A 的属于 λ_i 的全部特征向量是 $k_1\eta_1 + k_2\eta_2 + \cdots + k_t\eta_t$, 其中 $k_1, k_2, \cdots, k_t$ 是不全为零的复数.

例 1.1

求下列矩阵的特征值和特征向量.

$$(1)\ A=\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & 1 & 2 \end{bmatrix},\ (2)\ A=\begin{bmatrix} -6 & -4 & -24 \\ 1 & 2 & 3 \\ 2 & 1 & 8 \end{bmatrix},$$

$$(3)\ A=\begin{bmatrix} 3 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 2 \end{bmatrix}.$$

(1) $A = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & 1 & 2 \end{bmatrix}, \varphi(\lambda) = (\lambda - 1)(\lambda - 2)(\lambda - 3),$

$\lambda_1 = 1, p_1 = (-1, 0, 1)^T.$ $\lambda_2 = 2, p_2 = (-1, 1, 1)^T.$ $\lambda_3 = 3, p_3 = (-1, 1, 0)^T.$

(2) $A = \begin{bmatrix} -6 & -4 & -24 \\ 1 & 2 & 3 \\ 2 & 1 & 8 \end{bmatrix}, \varphi(\lambda) = (\lambda - 2)(\lambda - 1)^2,$

$\lambda_1 = 2, p_1 = (-3, 0, 1)^T, \lambda_2 = \lambda_3 = 1, p_2 = (-4, 1, 1)^T.$

(3) $A = \begin{bmatrix} 3 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 2 \end{bmatrix}, \varphi(\lambda) = (\lambda - 2)^3,$

$\lambda_1 = \lambda_2 = \lambda_3 = 2,\ p_1 = (0, 0, 1)^T, p_2 = (-1, 1, 0)^T.$

矩阵多项式

定义 1.10

设 $f(\lambda)$ 是 λ 的多项式 $f(\lambda) = a_s\lambda^s + a_{s-1}\lambda^{s-1} + \cdots + a_1\lambda + a_0$. 对 $A \in \mathbb{C}^{n\times n}$, 规定

$$f(A) = a_sA^s + a_{s-1}A^{s-1} + \cdots + a_1A + a_0I.$$

称 $f(A)$ 为矩阵 A 的多项式.

定理 1.11

1. 设 n 阶方阵 A 的 n 个特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 对应的特征向量分别为 $x_1, x_2, \cdots, x_n$. 若 $f(\lambda)$ 是多项式, 则 $f(A)$ 的特征值为 $f(\lambda_1), \cdots, f(\lambda_n)$, 对应的特征向量仍为 $x_1, x_2, \cdots, x_n$.
2. 设 $f(\lambda)$ 是一多项式. 若 $f(A) = 0$, 则 A 的任一特征值 λ_i 都满足 $f(\lambda_i) = 0$.

特征值和特征向量的性质

定理 1.12

设 n 阶方阵 A 的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$.

(1) $\sum\limits_{i=1}^{n} \lambda_i = a_{11} + a_{22} + \cdots + a_{nn}$.

(2) $\prod\limits_{i=1}^{n} \lambda_i = |A|$. 特别地, 0 是 A 的特征值 $\Longleftrightarrow |A| = 0$.

(3) A^T 的特征值是 $\lambda_1, \lambda_2, \cdots, \lambda_n$.

(4) A^H 的特征值为 $\overline{\lambda_1}, \overline{\lambda_2}, \cdots, \overline{\lambda_n}$.

方阵的迹

定义 1.13

n 阶方阵 $A=(a_{ij})$ 的主对角元素之和称为 A 的迹, 记作 $\mathrm{tr}(A)$, 即

$$\mathrm{tr}(A) := a_{11}+a_{22}+\cdots+a_{nn}.$$

- $\mathrm{tr}(A+B)=\mathrm{tr}(A)+\mathrm{tr}(B)$
- $\mathrm{tr}(kA)=k\,\mathrm{tr}(A)$
- $\mathrm{tr}(A^T)=\mathrm{tr}(A)$
- $\mathrm{tr}(A^H)=\overline{\mathrm{tr}(A)}$
- $\mathrm{tr}(AB)=\mathrm{tr}(BA)$

特征值和特征向量的性质 (续)

定理 1.14

(1) 设 λ_i 是 $A \in \mathbb{C}^{n\times n}$ 的 r_i 重特征值 (称 r_i 为特征值 λ_i 的代数重数), 对应 λ_i 有 t_i 个线性无关的特征向量 (称 t_i 为特征值 λ_i 的几何重数), 则 $1 \leqslant t_i \leqslant r_i$.

(2) 设 λ_1, λ_2 是 $A \in \mathbb{C}^{n\times n}$ 的不同的特征值, 且 $x_1, \cdots, x_s$ 是 λ_1 的线性无关的特征向量, $y_1, \cdots, y_r$ 是 λ_2 的线性无关的特征向量, 则 $x_1, \cdots, x_s, y_1, \cdots, y_r$ 线性无关.

(3) 设 $\lambda_1, \cdots, \lambda_m$ 是方阵 A 的 m 个互不相同的特征值, $x_{i1}, \cdots, x_{ir_i}$ 是 $\lambda_i (i = 1, 2, \cdots, m)$ 对应的线性无关的特征向量, 则下列向量组仍然线性无关

$$\underbrace{x_{11}, \cdots, x_{1r_1}}_{\lambda_1}, \underbrace{x_{21}, \cdots, x_{2r_2}}_{\lambda_2}, \cdots, \underbrace{x_{m1}, \cdots, x_{mr_m}}_{\lambda_m}.$$

习题 §1.2

1. 若 n 阶矩阵 A 满足 $A^3 = A$, 证明 A 的特征值只能是 ______ .
2. 设 A 是 n 阶可逆矩阵, 若 λ 为 A 的一个特征值, 则 $\lambda \neq 0$, 且 $\dfrac{1}{\lambda}$ 是 A^{-1} 的一个特征值.

§1.3 相似对角化

定义 1.15

设 A 与 B 是 n 阶复方阵, 如果存在可逆 n 阶复方阵 P, 使得 $P^{-1}AP = B$, 则称 A 与 B **相似**, 记作 $A \sim B$, 称 P 为把 A 变成 B 的相似变换矩阵.

- 反身性: $A \sim A$.
- 对称性: 若 $A \sim B$, 则 $B \sim A$.
- 传递性: 若 $A \sim B$ 和 $B \sim C$, 则 $A \sim C$.
- 若 $A \sim B$, 则 $\mathrm{rank}(A) = \mathrm{rank}(B)$, $|\lambda I - A| = |\lambda I - B|$, 即特征多项式相同, 从而特征值相同, 继而有 $\mathrm{tr}(A) = \mathrm{tr}(B)$, $|A| = |B|$.

可对角化的矩阵

定义 1.16

若 n 阶方阵 A 相似于一个对角矩阵 D, 则称 A 可对角化, 称 D 为 A 的相似标准形.

定理 1.17 (可对角化的充分必要条件)

n 阶方阵 A 可对角化 $\iff$ A 有 n 个线性无关的特征向量.

设 $D = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$, 则 $A \sim D = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$

$\iff$ 存在可逆矩阵 $P = (p_1, p_2, \cdots, p_n)$, 使得 $P^{-1}AP = D \Longrightarrow AP = PD$,

$$\text{即 } A(p_1, p_2, \cdots, p_n) = (p_1, p_2, \cdots, p_n)D$$

$$\text{即 } (Ap_1, Ap_2, \cdots, Ap_n) = (\lambda_1 p_1, \lambda_2 p_2, \cdots, \lambda_n p_n)$$

$\iff$ 存在 n 个线性无关的列向量 $p_i, i = 1, 2, \cdots, n$, 使得 $Ap_i = \lambda_i p_i$.

方阵可对角化的条件

推论 1.18

(1) 若 n 阶方阵 A 有 n 个互不相同特征值, 则 A 可对角化.

(2) n 阶方阵 A 可对角化 $\Longleftrightarrow$ A 的每个特征值的代数重数与几何重数相等.

例 1.2

下列哪个矩阵可对角化, 哪个不可对角化? 如果可对角化, 求出相似变换矩阵和相应的对角矩阵.

$$(1)\ A=\begin{bmatrix}2&-1&1\\-1&2&-1\\1&1&2\end{bmatrix}\rightsquigarrow\begin{bmatrix}1\\2\\3\end{bmatrix},$$

$$(2)\ A=\begin{bmatrix}-6&-4&-24\\1&2&3\\2&1&8\end{bmatrix}\rightsquigarrow\begin{bmatrix}1\\1\\2\end{bmatrix}\rightsquigarrow 1\ \text{有}\ 1\ \text{个无关特征向量},$$

$$(3)\ A=\begin{bmatrix}3&1&0\\-1&1&0\\-1&-1&2\end{bmatrix}\rightsquigarrow\begin{bmatrix}2\\2\\2\end{bmatrix}\rightsquigarrow 2\ \text{有两个无关特征向量}.$$

例 1.3

设 $A=\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & 1 & 2 \end{bmatrix}$, 求 A^{100}. $\rightsquigarrow P=\begin{bmatrix} -1 & -1 & -1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

例 1.4

求解一阶线性常系数微分方程组

$$\begin{cases} \dfrac{\mathrm{d}\,x_1}{\mathrm{d}\,t} = 2x_1 - x_2 + x_3 \\ \dfrac{\mathrm{d}\,x_2}{\mathrm{d}\,t} = -x_1 + 2x_2 - x_3 \\ \dfrac{\mathrm{d}\,x_3}{\mathrm{d}\,t} = x_1 + x_2 + 2x_3 \end{cases} \rightsquigarrow \frac{dx}{dt} = Ax, x = Py, \frac{d(Py)}{dt} = P\frac{dy}{dt}.$$

第 2 章　矩阵的 Jordan 标准型

§2.1　Jordan 标准形的定义

定义 2.1

形如 $J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}_{r_i \times r_i}$ 的矩阵称为 r_i 阶 Jordan 块, 由若干个 Jordan 块构成的分块对角矩阵

$$\begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{bmatrix}$$

称为 Jordan 矩阵.

例 2.1

$$
A=\left[\begin{array}{ccc|c|cc|cc}
-1 & 1 & & & & & & \\
 & -1 & 1 & & & & & \\
 & & -1 & {\color{red}0} & & & & \\
\hline
 & & & 2 & {\color{red}0} & & & \\
\hline
 & & & & i & 1 & & \\
 & & & & & i & {\color{red}0} & \\
\hline
 & & & & & & 0 & 1 \\
 & & & & & & & 0
\end{array}\right]
=\begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & J_3 & \\ & & & J_4 \end{bmatrix}
$$

注记 2.2

Jordan 标准型的此对角线上的元素只能是 0,1.

Jordan 标准形

定理 2.3

n 阶复方阵 A 与一个 Jordan 矩阵 J 相似, 即存在可逆方阵 P, 使得 $P^{-1}AP = J$. 矩阵 J 除 Jordan 块的排列次序外由 A 唯一确定, 称 J 为 A 的Jordan 标准形.

注记 2.4

以后 Jordan 标准型只写一种, 不考虑块的顺序.

问题 2.5

对于任意 A, 如何寻找 P 和 J?

定理 2.6

设 A 是 n 阶方阵. 若 λ_i 是 A 的 r_i 重特征值, 对应有 $s_i(1 \leqslant s_i \leqslant r_i)$ 个线性无关的特征向量, 则 A 的 Jordan 标准型中含有 s_i 个以 λ_i 为对角元的 Jordan 块, 且这些 Jordan 块的阶数之和恰好为 r_i.

证明: 由于 $P^{-1}AP = J$, 即相似矩阵有相同的特征值, 则 J 中以 λ_i 为对角元的 Jordan 块的阶数之和恰好为 r_i, 且 $\mathrm{rank}(\lambda_i I - A) = \mathrm{rank}(\lambda_i I - J)$. 又因为 λ_i 对应有 s_i 个线性无关的特征向量, 则 $n - \mathrm{rank}(\lambda_i I - A) = s_i \Rightarrow\ n - \mathrm{rank}(\lambda_i I - J) = s_i$.

对 Jordan 块 $J_t = \begin{bmatrix} \lambda_t & 1 & & \\ & \lambda_t & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_t \end{bmatrix}_{k_t \times k_t} \quad \Rightarrow \quad \mathrm{rank}(\lambda_i I - J_t) = \begin{cases} k_t - 1, & 若\lambda_t = \lambda_i \\ k_t, & 若\lambda_t \neq \lambda_i \end{cases}$

所以 $\mathrm{rank}(\lambda_i I - J) = n - \{以\lambda_i为对角元的\ \mathrm{Jordan}\ 块的个数\}$, 即

$$s_i = n - \mathrm{rank}(\lambda_i I - J) = \{以\lambda_i为对角元的\ \mathrm{Jordan}\ 块的个数\}. \quad \square$$

注记 2.7

- 单根对应一阶的 Jordan 块.
- 只有重根时才有可能有二阶及以上的 Jordan 块出现.
- 3 重及以下的特征值对应的 Jordan 块, 用试探法由特征向的个数唯一确定.
- 但是阶数 $\geqslant 4$ 时就会出现问题. 比如:

设 $A \in \mathbb{C}^{n\times n}$, $\lambda = 2$ 是 A 的 4 重特征值, 则对应 $\lambda = 2$ 的 Jordan 块可能有如下几种情况:

- 对应 $\lambda = 3$ 有 4 个线性无关的特征向量时 (可对角化), Jordan 块为

$$\begin{bmatrix} 2 & & & \\ & 2 & & \\ & & 2 & \\ & & & 2 \end{bmatrix}.$$

注记 2.8 (续)

- 对应 $\lambda = 2$ 有 3 个线性无关的特征向量时, Jordan 块为

$$\begin{bmatrix} 2 & & & \\ & 2 & & \\ & & 2 & 1 \\ & & & 2 \end{bmatrix}.$$

- 对应 $\lambda = 2$ 有 2 个线性无关的特征向量时, Jordan 块为

$$\begin{bmatrix} 2 & & & \\ & 2 & 1 & \\ & & 2 & 1 \\ & & & 2 \end{bmatrix} \text{或} \begin{bmatrix} 2 & 1 & & \\ & 2 & & \\ & & 2 & 1 \\ & & & 2 \end{bmatrix}.$$

注记 2.9 (续)

- 对应 $\lambda = 2$ 有 1 个线性无关的特征向量时, Jordan 块为

$$\begin{bmatrix} 2 & 1 & & \\ & 2 & 1 & \\ & & 2 & 1 \\ & & & 2 \end{bmatrix}.$$

例 2.2

求矩阵 $A = \begin{bmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$ 的 Jordan 标准型.

解: 特征多项式为 $\varphi(\lambda) = (\lambda - 2)(\lambda - 1)^2$, 而特征值 1 只有一个线性无关的特征向量, 故 A 的 Jordan 标准型为

$$\begin{bmatrix} 2 & & \\ & 1 & 1 \\ & & 1 \end{bmatrix}$$

□

练习 2.1

求下列矩阵的 Jordan 标准形.

$$(1)\ A=\begin{bmatrix} -6 & -4 & -24 \\ 1 & 2 & 3 \\ 2 & 1 & 8 \end{bmatrix};\ (2)\ A=\begin{bmatrix} 3 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 2 \end{bmatrix}.$$

§2.3 求 Jordan 标准形的方法—初等变换法

设 λ 是未定元, 称形如 $f(\lambda)=a_n\lambda^n+a_{n-1}\lambda^{n-1}+\cdots+a_1\lambda+a_0,\ a_n\neq 0$ 的式子是关于 λ 的多项式, a_i 称为系数, $n(\geqslant 0)$ 称为 $f(\lambda)$ 的次数, 记为 $\deg f(\lambda)=n$. 若 $a_n=1$, 称 $f(\lambda)$ 为首 1多项式. 另外, 规定 0 的次数是 $-\infty$.

定理 2.10 (带余除法)

设 $f(\lambda), g(\lambda)$ 是两个多项式, $g(\lambda)\neq 0$, 则存在唯一多项式 $q(\lambda), r(\lambda)$ 使得

$$f(\lambda)=q(\lambda)g(\lambda)+r(\lambda),$$

其中 $r(\lambda)=0$ 或 $\deg r(\lambda)<\deg g(\lambda)$.

如果 $r(\lambda)=0$, 则称 $g(\lambda)$ 整除 $f(\lambda)$, 记为 $g(\lambda)|f(\lambda)$.

例 2.3

$(\lambda - 1)^2 \mid (\lambda - 1)^3(\lambda + 1)$

例 2.4

$\lambda^3 + 4\lambda^2 - 3\lambda + 7 = (\lambda^2 + 6\lambda + 9)(\lambda - 2) + 25$

λ-矩阵的初等行 (列) 变换

定义 2.11

设 $A(\lambda)=(a_{ij}(\lambda))_{m\times n}$, 其中 $a_{ij}(\lambda)$ 都是 λ 的多项式, 则称 $A(\lambda)$ 是 λ- 矩阵或多项式矩阵.

例 2.5

$$A(\lambda)=\begin{bmatrix}-\lambda+1 & 2\lambda-1 & \lambda \\ \lambda & \lambda^2 & -\lambda \\ \lambda^2+1 & \lambda^2+\lambda-1 & -\lambda^2\end{bmatrix}$$ 为一个 λ-矩阵.

定义 2.12

λ-矩阵 $A(\lambda)$ 中不恒等于 0 的子式的最高阶数 r 称为 λ-矩阵的秩, 记为 $\operatorname{rank} A(\lambda) = r$.

例 2.6

$$A(\lambda) = \begin{bmatrix} 1 & \lambda \\ \lambda & 0 \end{bmatrix}, \quad \det A(\lambda) = -\lambda^2 \neq 0, \operatorname{rank} A(\lambda) = 2.$$

显然, 特征矩阵 $\lambda I - A$ 为我们常见的一类 λ-矩阵, 且 $\operatorname{rank} \lambda I - A = n$, 这里 n 是 A 的阶数.

λ-矩阵的初等行 (列) 变换:

1. 交换两行 (列): 交换 i, j 两行 (列), 记作 $r_i \longleftrightarrow r_j$ ($c_i \longleftrightarrow c_j$).
2. 数 $k(k \neq 0)$ 乘某行 (列) 的所有元素: 第 i 行 (列) 乘 k, 记作 $r_i \times k$ ($c_i \times k$).
3. 把某一行 (列) 所有元素的 $\varphi(\lambda)$ 倍加到另一行 (列) 对应的元素上去, 其中 $\varphi(\lambda)$ 是一个多项式: 第 j 行 (列) 的 $\varphi(\lambda)$ 倍加到第 i 行 (列) 上, 记作 $r_i + \varphi(\lambda) r_j$ ($c_i + \varphi(\lambda) c_j$).

定义 2.13

若 λ-矩阵 $A(\lambda)$ 经过有限次初等变换变成 λ-矩阵 $B(\lambda)$, 则称 $A(\lambda)$ 与 $b(\lambda)$ 等价, 记为 $A(\lambda) \sim B(\lambda)$.

引理 2.14

若 λ-矩阵 $A(\lambda)$ 中 $a_{11}(\lambda)\neq 0$, 且 $A(\lambda)$ 中至少有一个元素不能被它除尽, 则必可以找到与之等价的矩阵 $B(\lambda)$，使得 $b_{11}(\lambda)\neq 0$, 且 $\deg a_{11}(\lambda)>\deg b_{11}(\lambda)$.

证明思路: 分三种情况, 情况 1, 第一列中 $a_{i1}(\lambda)$ 不能被 $a_{11}(\lambda)$ 整除, 则有 $a_{i1}(\lambda)=q(\lambda)a_{11}(\lambda)+r(\lambda)$,

$$A(\lambda)=\begin{bmatrix} a_{11}(\lambda) & \vdots \\ \vdots & \vdots \\ a_{i1}(\lambda) & \vdots \\ \vdots & \vdots \end{bmatrix}\to\begin{bmatrix} a_{11}(\lambda) & \vdots \\ \vdots & \vdots \\ r(\lambda) & \vdots \\ \vdots & \vdots \end{bmatrix}\to\begin{bmatrix} r(\lambda) & \vdots \\ \vdots & \vdots \\ a_{11}(\lambda) & \vdots \\ \vdots & \vdots \end{bmatrix}=B(\lambda)$$

情况 2, 第一行中有一个不能被 $a_{11}(\lambda)$ 整除, 与情况 1 类似. 情况 3, 第一行和第一列的其他元素都能被 $a_{11}(\lambda)$ 整除, 但是存在 $a_{ij}(\lambda)(i,j>1)$ 不能被 $a_{11}(\lambda)$ 整除. 设 $a_{i1}=h(\lambda)a_{11}(\lambda)$, 则

$$
\begin{bmatrix} a_{11}(\lambda) & \cdots & a_{1j}(\lambda) & \cdots \\ \vdots & & \vdots & \\ a_{i1}(\lambda) & \cdots & a_{ij}(\lambda) & \cdots \\ \vdots & & & \ddots \end{bmatrix} \to \begin{bmatrix} a_{11}(\lambda) & \cdots & a_{1j}(\lambda) & \cdots \\ 0 & & \vdots & \\ \vdots & \cdots & a_{ij}(\lambda) - a_{1j}h(\lambda) & \cdots \\ 0 & & & \ddots \end{bmatrix}
$$

$$
\to \begin{bmatrix} a_{11}(\lambda) & \cdots & a_{ij}(\lambda) + a_{1j}(\lambda)(1 - h(\lambda)) & \cdots \\ 0 & & \vdots & \\ \vdots & \cdots & a_{ij}(\lambda) - a_{1j}h(\lambda) & \cdots \\ 0 & & & \ddots \end{bmatrix} \xrightarrow{\text{由情况2}} \cdots \to B(\lambda).
$$

例 2.7

找 $A(\lambda)=\begin{bmatrix}\lambda-1 & 0\\ 0 & (\lambda-2)^2\end{bmatrix}$ 的一个等价标准型.

由带余除法知, $(\lambda-2)^2=(\lambda-3)(\lambda-1)+1$,

$$\begin{bmatrix}\lambda-1 & 0\\ 0 & (\lambda-3)(\lambda-1)+1\end{bmatrix}\to\begin{bmatrix}\lambda-1 & \lambda-1\\ 0 & (\lambda-3)(\lambda-1)+1\end{bmatrix}\to$$

$$\begin{bmatrix}\lambda-1 & \lambda-1\\ -(\lambda-3)(\lambda-1) & 1\end{bmatrix}\to\begin{bmatrix}1 & -(\lambda-3)(\lambda-1)\\ \lambda-1 & \lambda-1\end{bmatrix}=B(\lambda).$$

不变因子和初等因子

定理 2.15

秩为 r 的 λ-矩阵 $A(\lambda)$ 可通过初等变换化为如下形式

$$S(\lambda) = \begin{bmatrix} d_1(\lambda) & & & & \\ & d_2(\lambda) & & & \mathbf{0} \\ & & \ddots & & \\ & & & d_r(\lambda) & \\ & \mathbf{0} & & & \mathbf{0} \end{bmatrix}$$

其中 $d_i(\lambda)(i=1,2,\cdots,r)$ 都是首 1 多项式, 且

$$d_i(\lambda)|d_{i+1}(\lambda) \quad (i=1,2,\cdots,r-1),$$

称为 $A(\lambda)$ 的不变因子. $S(\lambda)$ 由 $A(\lambda)$ 唯一确定, 称为 $A(\lambda)$ 的 Smith 标准形.

例 2.8

求 λ-矩阵 $A(\lambda)=\begin{bmatrix}\lambda-1 & 0\\ 0 & (\lambda-2)^2\end{bmatrix}$ 的 Smith 标准形, 不变因子.

$$A(\lambda)\to\cdots\to\begin{bmatrix}1 & -(\lambda-3)(\lambda-1)\\ \lambda-1 & \lambda-1\end{bmatrix}\to\begin{bmatrix}1 & -(\lambda-3)(\lambda-1)\\ 0 & (\lambda-1)(\lambda-2)^2\end{bmatrix}$$

$$\to\begin{bmatrix}1 & 0\\ 0 & (\lambda-1)(\lambda-2)^2\end{bmatrix}$$

不变因子由 $d_1=1, d_2=(\lambda-1)(\lambda-2)^2$.

注记 2.16

把 $A(\lambda)$ 的每个次数大于零的不变因子分解为互不相同的一次因子方幂的乘积, 所有这些一次因子方幂 (相同的必须按出现的次数计算) 称为 $A(\lambda)$ 的初等因子.

例 2.9

求 $\lambda I - A$ 的 Smith 标准形, 不变因子和初等因子. 其中

$$(1)\ A = \begin{bmatrix} -6 & -4 & -24 \\ 1 & 2 & 3 \\ 2 & 1 & 8 \end{bmatrix};\ (2)\ A = \begin{bmatrix} 3 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 2 \end{bmatrix}.$$

(1)

$$\lambda I - A = \begin{bmatrix} \lambda+6 & 4 & 24 \\ -1 & \lambda-2 & -3 \\ -2 & -1 & \lambda-8 \end{bmatrix} \to \begin{bmatrix} -1 & \lambda-2 & -3 \\ \lambda+6 & 4 & 24 \\ -2 & -1 & \lambda-8 \end{bmatrix}$$

$$\to \begin{bmatrix} -1 & \lambda-2 & -3 \\ 0 & \lambda^2+4\lambda-8 & -3\lambda+6 \\ 0 & -2\lambda+3 & \lambda-2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & (\lambda-1)^2 & 0 \\ 0 & -2\lambda+3 & \lambda-2 \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 0 \\ 0 & (\lambda-1)^2 & 0 \\ 0 & -1 & \lambda-2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & (\lambda-2)(\lambda-1)^2 \\ 0 & -1 & \lambda-2 \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (\lambda-2)(\lambda-1)^2 \end{bmatrix}$$

(2)

$$\lambda I - A = \begin{bmatrix} \lambda-3 & -1 & 0 \\ 1 & \lambda-1 & 0 \\ 1 & 1 & \lambda-2 \end{bmatrix} \to \begin{bmatrix} 1 & \lambda-1 & 0 \\ \lambda-3 & -1 & 0 \\ 1 & 1 & \lambda-2 \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & \lambda-1 & 0 \\ 0 & -\lambda^2+4\lambda-4 & 0 \\ 0 & 2-\lambda & \lambda-2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & -(\lambda-2)^2 & 0 \\ 0 & 2-\lambda & \lambda-2 \end{bmatrix}$$

$$\to \begin{bmatrix} 1 & 0 & 0 \\ 0 & -(\lambda-2)^2 & 0 \\ 0 & 0 & \lambda-2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda-2 & 0 \\ 0 & & (\lambda-2)^2 \end{bmatrix}.$$

初等变换法

设 A 是 n 阶复方阵.

(1) 用初等变换求特征矩阵 $\lambda I - A$ 的 Smith 标准形

(2) 求出不变因子和初等因子

(3) 写出每个初等因子 $(\lambda - \lambda_i)^{r_i}$ 对应的 Jordan 块

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}_{r_i \times r_i}$$

(4) 得到 Jordan 矩阵 $J = \mathrm{diag}(J_1, J_2, \cdots, J_s)$.

例 2.10

求下列矩阵的 Jordan 标准形.

$$(1)\ A=\begin{bmatrix} -6 & -4 & -24 \\ 1 & 2 & 3 \\ 2 & 1 & 8 \end{bmatrix};\ (2)\ A=\begin{bmatrix} 3 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 2 \end{bmatrix}.$$

例 2.11

已知一个 12 阶矩阵 A 的不变因子是

$$\underbrace{1,1,\cdots,1}_{9},(\lambda-1)^2,(\lambda-1)^2(\lambda+2),(\lambda-1)^2(\lambda+2)(\lambda^2+3)^2$$

求 A 的 Jordan 标准形.

§2.4 求 Jordan 标准形的方法—行列式因子法

定义 2.17

如果多项式 $d(\lambda)$ 满足 $d(\lambda)|f(\lambda)$, $d(\lambda)|g(\lambda)$, 则称 $d(\lambda)$ 为 $f(\lambda), g(\lambda)$ 的一个公因子. $f(\lambda), g(\lambda)$ 的公因子中次数最高的首 1多项式的称为最大公因子, 记为 $d(\lambda) = (f(\lambda), g(\lambda))$.

同样可以定义多个多项式 $f_1(\lambda), f_2(\lambda), \cdots, f_n(\lambda)$ 的最大公因子.

例 2.12

$\pm 1, \pm(\lambda-1), \pm(\lambda-1)(\lambda+1)$ 都是多项式 $f(\lambda) = (\lambda-1)^2(\lambda+1)$, $g(\lambda) = (\lambda-1)(\lambda+1)^2$ 的公因子, 但是 $(f(\lambda), g(\lambda)) = (\lambda-1)(\lambda+1)$.

例 2.13

多项式 $2(\lambda-1)^3$ 与 $6(\lambda-1)^2$ 的公因子有 $2, (\lambda-1), 2(\lambda-1), \cdots$, 但是 $(2(\lambda-1)^3, 6(\lambda-1)^2) = (\lambda-1)^2$.

定义 2.18

设 λ-矩阵 $A(\lambda)$ 的秩为 r. 对于正整数 $k(1 \leqslant k \leqslant r)$, $A(\lambda)$ 的全部 k 阶子式的最大公因式 $D_k(\lambda)$ 称为 $A(\lambda)$ 的 k 阶行列式因子.

命题 2.19

若 λ-矩阵 $A(\lambda)$ 经过初等变换变成 $B(\lambda)$, 则 $A(\lambda)$ 与 $B(\lambda)$ 的行列式因子相同.

定理 2.20

设 $A(\lambda)$ 是秩为 r 的 $m \times n$ 矩阵, 则 $A(\lambda)$ 的行列式因子 $D_k(\lambda)$ 为

$$D_k(\lambda) = d_1(\lambda)d_2(\lambda)\cdots d_k(\lambda), \quad (k = 1, 2, \cdots, r),$$

其中 $d_k(\lambda)(k = 1, 2, \cdots, r)$ 是 $A(\lambda)$ 的不变因子. 于是

$$d_1(\lambda) = D_1(\lambda), d_2(\lambda) = \frac{D_2(\lambda)}{D_1(\lambda)}, \cdots, d_r(\lambda) = \frac{D_r(\lambda)}{D_{r-1}(\lambda)}.$$

例 2.14

求 $A(\lambda)=\begin{bmatrix}\lambda-1 & 0\\ 0 & (\lambda-2)^2\end{bmatrix}$ 的 Smith 标准型.

显然 $D_1(\lambda)=1, D_2(\lambda)=(\lambda-1)(\lambda-2)^2$, 则

$d_1=1, d_2=\dfrac{D_2(\lambda)}{D_1(\lambda)}=(\lambda-1)(\lambda-2)^2$, 所以 Smith 标准型为

$$\begin{bmatrix}1 & 0\\ 0 & (\lambda-1)(\lambda-2)^2\end{bmatrix}$$

行列式因子法

(1) 求特征矩阵 $\lambda I - A$ 的各阶行列式因子

(2) 求出不变因子、初等因子

(3) 写出对应的 Jordan 块, 得到 Jordan 矩阵

例 2.15

求下列矩阵的 Jordan 标准形.

$$(1)\ A = \begin{bmatrix} -6 & -4 & -24 \\ 1 & 2 & 3 \\ 2 & 1 & 8 \end{bmatrix};\ (2)\ A = \begin{bmatrix} 3 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 2 \end{bmatrix}.$$

注记 2.21

因为初等变换不改变行列式因子, 可结合初等变换变成类似上三角或者对角的形式再找行列式因子.

分块对角矩阵

命题 2.22

若 λ-矩阵 $A(\lambda)$ 等价于分块对角矩阵

$$\begin{bmatrix} A_1(\lambda) & & & \\ & A_2(\lambda) & & \\ & & \ddots & \\ & & & A_r(\lambda) \end{bmatrix}$$

则 $A_1(\lambda), A_2(\lambda), \cdots, A_r(\lambda)$ 的初等因子全体构成 $A(\lambda)$ 的初等因子组.

练习 2.2

设 $A_1 = \begin{bmatrix} 1 & -4 & -1 \\ 1 & -3 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix},$

求 (1) λ-矩阵 $\lambda I - A$ 的 Smith 标准型, 不变因子和初等因子.

(2) 矩阵 A 的 Jordan 标准型.

§2.5 相似变换矩阵的计算

例 2.16

求下列矩阵的 Jordan 标准形和所用的相似变换矩阵.

$$(1)\ A=\begin{bmatrix} -6 & -4 & -24 \\ 1 & 2 & 3 \\ 2 & 1 & 8 \end{bmatrix};\ (2)\ A=\begin{bmatrix} 3 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 2 \end{bmatrix}.$$

注记 2.23

当一个重特征值对应2 个或 2 个以上的 Jordan 块时, 可能需要这样处理.

§2.6 应用案例一、Jordan 标准型的应用举例

定理 2.24

对于 r 阶 Jordan 块 $J = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}_{r\times r}$ 有

$$J^k = \begin{bmatrix} \lambda^k & C_k^1\lambda^{k-1} & C_k^2\lambda^{k-2} & \cdots & C_k^{r-1}\lambda^{k-r+1} \\ & \lambda^k & C_k^1\lambda^{k-1} & \ddots & \vdots \\ & & \ddots & \ddots & C_k^2\lambda^{k-2} \\ & & & \lambda^k & C_k^1\lambda^{k-1} \\ & & & & \lambda^k \end{bmatrix}$$

若 $k < r-1$, 规定 $C_k^r = 0$.

例 2.17

已知 $A=\begin{bmatrix} 3 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & -1 & 2 \end{bmatrix}$, 求 A^k. $\rightsquigarrow P=\begin{bmatrix} 1 & 1 & 0 \\ -1 & -1 & 1 \\ 0 & -1 & 0 \end{bmatrix}$

例 2.18

求解一阶线性常系数微分方程组

$$\begin{cases} \dfrac{\mathrm{d}\,x_1}{\mathrm{d}\,t} = 3x_1 + x_2 \\ \dfrac{\mathrm{d}\,x_2}{\mathrm{d}\,t} = -x_1 + x_2 \\ \dfrac{\mathrm{d}\,x_3}{\mathrm{d}\,t} = -x_1 - x_2 + 2x_3 \end{cases} \rightsquigarrow x = Py, P\frac{dy}{dt} = APy, \frac{dy}{dt} = Jy.$$

第 3 章 Hamilton-Cayley 定理

§3.1 Hamilton-Cayley 定理

定理 3.1

设 A 是 n 阶方阵, $\varphi(\lambda)=|\lambda I-A|$, 则 $\varphi(A)=0$.

命题 3.2

如果 λ_0 是多项式 $\varphi(\lambda)$ 的 m 重根, 则

$$\varphi(\lambda_0)=\varphi'(\lambda_0)=\cdots=\varphi^{(m-1)}(\lambda_0)=0$$

例 3.1

已知矩阵 $A=\begin{bmatrix} 3 & -3 & 2 \\ -1 & 5 & -2 \\ -1 & 3 & 0 \end{bmatrix}$, 试计算

(1) $A^6+A^5+6A-4I$; (2) A^{100}. $\rightsquigarrow$ $\varphi(\lambda)=|\lambda I-A|=(\lambda-2)^2(\lambda-4)$.

§3.2 最小多项式

定义 3.3

设 A 是 n 阶方阵, 如果有多项式 $f(\lambda)$ 使得 $f(A) = 0$, 则称 $f(\lambda)$ 为 A 的零化多项式.

例 3.2

例如: $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 的一个零化多项式为 $f(\lambda) = (\lambda - 1)^2$, 但是 $g(\lambda) = \lambda - 1$ 也是一个零化多项式.

定义 3.4

设 A 是 n 阶方阵, 在 A 的零化多项式中, 次数最低的首一多项式称为 A 的最小多项式, 记为 $m_A(\lambda)$.

定理 3.5

设 A 是 n 阶方阵, 则 (1). A 的最小多项式 $m_A(\lambda)$ 整除 A 的任一零化多项式, 而且最小多项式是唯一的.

(2). λ_0 是 A 的特征值 $\iff$ λ_0 是 $m_A(\lambda)$ 零点. 从而最小多项式是特征多项式的因子且包含所有的特征值作为零点.

例 3.3 (试探法)

求下列矩阵的最小多项式.

$$A = \begin{bmatrix} 3 & -3 & 2 \\ -1 & 5 & -2 \\ -1 & 3 & 0 \end{bmatrix} \rightsquigarrow \varphi(\lambda) = (\lambda-2)^2(\lambda-4)$$

例 3.4 (试探法)

求 $A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}_{n\times n}$ 的最小多项式. $\rightsquigarrow \varphi(\lambda) = \lambda^n$

引理 3.6

r 阶 Jordan 块 $J = \begin{bmatrix} \lambda_0 & 1 & & \\ & \lambda_0 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_0 \end{bmatrix}_{r\times r}$ 的最小多项式和特征多项式相等, 为 $(\lambda - \lambda_0)^r$.

定理 3.7

1. 相似的矩阵有相同的最小多项式.
2. 设 A 是 n 阶方阵, $\lambda_1, \lambda_2, \cdots, \lambda_t$ 是 A 的所有互不相同的特征值, 则

$$m_A(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_t)^{m_t},$$

   其中 m_i 是 A 的 Jordan 标准形 J 中含 λ_i 的 Jordan 块的最高阶数. (即相同特征值的次数最高的初等因子的乘积.)
3. 矩阵可对角化 $\Longleftrightarrow$ 最小多项式无重根.

例 3.5

求下列矩阵的最小多项式.

(1) $$A=\begin{bmatrix}2&1&&&&&&\\&2&&&&&&\\&&2&1&&&&\\&&&2&1&&&\\&&&&2&&&\\&&&&&3&&\\&&&&&&3&1\\&&&&&&&3\end{bmatrix},\ m(\lambda)=(\lambda-2)^3(\lambda-3)^2.$$

(2) $$A=\begin{bmatrix}3&-3&2\\-1&5&-2\\-1&3&0\end{bmatrix}\rightsquigarrow J=\begin{bmatrix}2&&\\&2&\\&&4\end{bmatrix}.$$

设 A 是 n 阶方阵, D_{n-1} 是 $\lambda I - A$ 的 $n-1$ 阶行列式因子, 则

$$m_A(\lambda) = \frac{|\lambda I - A|}{D_{n-1}(\lambda)} = \text{次数最高的不变因子 } d_n(\lambda).$$

例 3.6

求下列矩阵的最小多项式.

$$(1)\ A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad (2)\ A = \begin{bmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}.$$

最小多项式的简化应用

例 3.7

已知矩阵 $A=\begin{bmatrix} 3 & -3 & 2 \\ -1 & 5 & -2 \\ -1 & 3 & 0 \end{bmatrix}$, 试计算

(1) $A^6+A^5+6A-4I$;

(2) A^{100}.

第 4 章 酉矩阵的定义及性质

§4.1 酉矩阵的定义及性质

定义 4.1

设 $\alpha = (x_1, \cdots, x_n)^T \in \mathbb{C}^n$, $\beta = (y_1, \cdots, y_n)^T \in \mathbb{C}^n$, 称

$$(\alpha, \beta) = \sum_{k=1}^{n} x_k \overline{y_k} = \beta^H \alpha,$$

为向量 α 与 β 的**内积**.

定理 4.2 (内积的性质)

设 $\alpha, \beta, \gamma \in \mathbb{C}^n$, $\lambda \in \mathbb{C}$, 则

(1) $(\alpha, \beta) = \overline{(\beta, \alpha)}$;

(2) $(\lambda\alpha, \beta) = \lambda(\alpha, \beta)$, $(\alpha, \lambda\beta) = \overline{\lambda}(\alpha, \beta)$;

(3) $(\alpha + \beta, \gamma) = (\alpha, \gamma) + (\beta, \gamma)$;

(4) $(\alpha, \alpha) \geqslant 0$, $(\alpha, \alpha) = 0$ 当且仅当 $\alpha = 0$;

(5) Cauchy-Schwarz 不等式:

$$(\alpha, \beta)(\beta, \alpha) \leqslant (\alpha, \alpha)(\beta, \beta) \text{ 或 } |(\alpha, \beta)|^2 \leqslant (\alpha, \alpha)(\beta, \beta)$$

由 Cauchy-Schwarz 不等式, 有

$$\left(\sum_{k=1}^{n} |x_k||y_k|\right)^2 \leqslant \left(\sum_{k=1}^{n} |x_k|^2\right)\left(\sum_{k=1}^{n} |y_k|^2\right).$$

向量的长度

定义 4.3

设 $\alpha=(x_1,x_2,\cdots,x_n)^T\in\mathbb{C}^n$, 称

$$\|\alpha\|_2=\sqrt{(\alpha,\alpha)}=\sqrt{\sum_{k=1}^{n}|x_k|^2},$$

为向量 α 的长度或 2 范数.

定理 4.4

设 $\alpha,\beta\in\mathbb{C}^n$, $\lambda\in\mathbb{C}$, 则

(1) 非负性: $\|\alpha\|_2\geqslant 0$; $\|\alpha\|_2=0\Longleftrightarrow\alpha=0$

(2) 齐次性: $\|\lambda\alpha\|_2=|\lambda|\|\alpha\|_2$

(3) 三角形不等式: $\|\alpha+\beta\|_2\leqslant\|\alpha\|_2+\|\beta\|_2$

正交向量

定义 4.5

设 $\alpha, \beta \in \mathbb{C}^n$.

- 若 $\|\alpha\|_2 = 1$, 则称 α 是**单位向量**.
- 当 $\alpha \neq 0$ 时, $\dfrac{\alpha}{\|\alpha\|}$ 是单位向量, 称之为将向量 α **单位化** 或**规范化**.
- 当 $(\alpha, \beta) = 0$ 时, 称向量 α 与 β **正交**.

定理 4.6

设 $\alpha_1, \alpha_2, \cdots, \alpha_s \in \mathbb{C}^n$ 是一组两两正交的非零向量, 则 $\alpha_1, \alpha_2, \cdots, \alpha_s \in \mathbb{C}^n$ 线性无关.

Gram-Schmidt 正交化

设 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 是一组线性无关的向量.

- 令 $\beta_1 = \alpha_1$
- 令 $\beta_2 = \alpha_2 - \dfrac{(\alpha_2, \beta_1)}{(\beta_1, \beta_1)}\beta_1$
- 依次类似作出 $\beta_3, \beta_4, \cdots, \beta_s$, 其中

$$\beta_k = \alpha_k - \sum_{j=1}^{k-1} \frac{(\alpha_k, \beta_j)}{(\beta_j, \beta_j)}\beta_j$$

- 把两两正交的向量 $\beta_1, \beta_2, \cdots, \beta_s$ 分别单位化, 即令

$$\gamma_k = \frac{\beta_k}{\|\beta_k\|_2}$$

向量组 $\beta_1, \beta_2, \cdots, \beta_s$ 与向量组 $\alpha_1, \alpha_2, \cdots, \alpha_s$ 等价.

例 4.1

用 Gram-Schmidt 正交化方法把下面这组向量

$$\alpha_1 = \begin{bmatrix} 1 \\ \mathrm{i} \\ 0 \end{bmatrix}, \alpha_2 = \begin{bmatrix} 1 \\ 0 \\ \mathrm{i} \end{bmatrix}, \alpha_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

正交单位化.

酉矩阵

定义 4.7

设 A 是 n 阶复方阵. 若 A 满足

$$A^H A = I \quad \text{或} \quad A^{-1} = A^H,$$

则称 A 为**酉矩阵**.

酉矩阵的性质

定理 4.8

设 A 和 B 是 n 阶复方阵.

(1) 若 A 是酉矩阵, 则 A^H 和 A^{-1} 也是酉矩阵.

(2) 若 A, B 是酉矩阵, 则 AB 与 $\begin{bmatrix} A & \\ & B \end{bmatrix}$ 也是酉矩阵.

(3) 若 A 是酉矩阵, 则对任意 $x \in \mathbb{C}^n$ 有 $\|Ax\|_2 = \|x\|_2$. (长度的酉不变性)

(4) 若 A 是酉矩阵, 则 $|\det(A)| = 1$. (这里 $|\cdot|$ 表示模长)

(5) A 酉矩阵 $\Longleftrightarrow$ A 的 n 个列向量是两两正交的单位向量.

(6) A 酉矩阵 $\Longleftrightarrow$ A 的 n 个行向量是两两正交的单位向量.

例 4.2

矩阵 $A=\begin{bmatrix}1 & 1 & 0\\ \mathrm{i} & 0 & 0\\ 0 & \mathrm{i} & 1\end{bmatrix}$ 是否是酉矩阵, 若不是, 试利用其列向量构造一个酉矩阵.

§4.2 Schur 分解定理

定义 4.9

设 $A, B \in \mathbb{C}^{n\times n}$. 若存在酉矩阵 U 使得

$$U^{-1}AU = B$$

则称 A 与 B 酉相似.

引理 4.10

对任意 n 元单位复向量 u_1, 存在酉矩阵 U, 使得 u_1 为其列向量.

定理 4.11 (Schur)

设 $A \in \mathbb{C}^{n\times n}$, 则 A 可酉相似于上三角矩阵 T, 即存在 n 阶酉矩阵 U, 使得

$$U^{-1}AU = U^H AU = T.$$

酉相似于对角阵的矩阵

什么样的矩阵可酉相似于对角阵?

定理 4.12

设 $A \in \mathbb{C}^{n\times n}$, 则 A 酉相似于对角矩阵 $\iff A^HA = AA^H$.

定义 4.13

设 $A \in \mathbb{C}^{n\times n}$. 若 A 满足

$$A^HA = AA^H,$$

则称 A 为正规矩阵.

对角矩阵, 正交矩阵 ($AA^T = I$), 实对阵矩阵, 实反对称矩阵, 酉矩阵, Hermite 矩阵 ($A = A^H$), 反 Hermite 矩阵 ($A^H = -A$) $\cdots$.

正规矩阵的基本性质

命题 4.14

- 设 $A \in \mathbb{C}^{n\times n}$ 是正规矩阵, λ 是 A 的特征值, α 是对应 λ 的特征向量, 则 $\overline{\lambda}$ 是 A^H 的特征值, 对应 $\overline{\lambda}$ 的特征向量仍为 α.
- 设 $A \in \mathbb{C}^{n\times n}$ 是正规矩阵, λ, μ 是 A 的特征值, α, β 是分别对应 λ, μ 的特征向量. 若 $\lambda \neq \mu$, 则 α 与 β 正交.

推论 4.15

- Hermite 矩阵的特征值均为实数, 反 Hermite 矩阵的特征值为零或纯虚数.
- 实对称矩阵的特征值均为实数, 实反对称矩阵的特征值为零或纯虚数.

正规矩阵酉相似于对角阵的计算

- 求 A 的特征值和特征向量;
- 将属于重特征值的线性无关特征向量正交化、单位化;
- 将特征值排成对角矩阵 Λ, 相应的特征向量排成矩阵 U, 则有 $U^H AU = \Lambda$.

例 4.3

矩阵 $\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$ 是否正规? 若是, 试将其酉相似对角化.

Hermite (半) 正定矩阵

定义 4.16

设 $A \in \mathbb{C}^{n\times n}$ 是 Hermite 矩阵. 如果对任意 $0 \neq \alpha \in \mathbb{C}^n$ 都有

$$\alpha^H A\alpha > 0 \qquad (\alpha^H A\alpha \geqslant 0)$$

则称 A 是 Hermite 正定矩阵 (半正定矩阵).

Hermite 正定矩阵的等价条件

定理 4.17

设 $A \in \mathbb{C}^{n\times n}$ 是 Hermite 矩阵, TFAE

(1) A 是 Hermite 正定矩阵.

(2) A 的特征值全为正实数.

(3) 存在 n 阶满秩方阵 P 使得 $A = P^H P$.

推论 4.18

Hermite 正定矩阵的行列式大于 0.

Hermite 半正定矩阵的判定及应用

定理 4.19

设 $A \in \mathbb{C}^{n\times n}$ 是 Hermite 矩阵, 下列等价

(1) A 是 Hermite 半正定矩阵.

(2) A 的特征值全为非负实数.

(3) 存在 n 阶方阵 P 使得 $A = P^H P$.

顺序主子式判定正定的方法

定理 4.20

设 $A=(a_{ij})$ 是 n 阶复方阵. 分别称

$$A_k=\begin{bmatrix} a_{11} & \cdots & a_{1k} \\ a_{21} & \cdots & a_{2k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{bmatrix}, \quad \Delta_k=|A_k|.$$

为 A 的 k 阶顺序主子阵和顺序主子式. 则 A 是 Hermite 正定矩阵 $\iff \Delta_k>0(k=1,2,\cdots,n)$.

注意: 所有顺序主子式非负不能保证 Hermite 矩阵的半正定性. 例如 $\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$.

第 5 章 矩阵分解

§5.1 矩阵的 LU 分解

定义 5.1

设 $A \in \mathbb{C}^{n\times n}$. 若存在下三角矩阵 $L \in \mathbb{C}^{n\times n}$ 和上三角矩阵 $U \in \mathbb{C}^{n\times n}$, 使得

$$A = LU,$$

则称 A 可以作三角分解.

注记 5.2

注意矩阵未必有三角分解, 比如可逆矩阵 $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, 即使有也未必唯一,

$$A = AI_2 = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1/2 \end{bmatrix}.$$

注记 5.3

三种特殊的三角分解

单位下三角矩阵: 对角元素为 1 的下三角矩阵.

单位上三角矩阵: 对角元素为 1 的上三角矩阵.

定义 5.4

设 $A \in \mathbb{C}^{n\times n}$.

- Doolittle 分解: $A = LU$, 其中 L 是单位下三角矩阵, U 是上三角矩阵.
- Crout 分解: $A = LU$, 其中 L 是下三角矩阵, U 是单位上三角矩阵.
- LDU 分解: $A = LDU$, 其中 L 是单位下三角矩阵, D 是对角矩阵, U 是单位上三角矩阵.

三角分解的唯一性

定理 5.5

设 A 是 n 阶满秩复方阵, 则 A 有唯一 LDU 分解当且仅当 $\Delta_k \neq 0 (k = 1, \cdots, n)$
此时对角阵 $D = \mathrm{diag}(d_1, \cdots, d_n)$ 的元素满足

$$d_1 = \Delta_1,\ d_k = \frac{\Delta_k}{\Delta_{k-1}} \quad (k = 2, 3, \cdots, n).$$

推论 5.6

设 A 是 n 阶满秩复方阵, 则 A 有唯一 Doolittle 分解或 Crout 分解当且仅当 $\Delta_k \neq 0 (k = 1, 2, \cdots, n)$.

注记 5.7

我们也可以用初等变换求矩阵的 LU 分解.

定理 5.8

设 A 是 n 阶 Hermite 正定矩阵, 则存在下三角矩阵 G 使得 $A = GG^H$. 称之为 A 的 Cholesky 分解. 如果要求 G 的主对角线上元素都正, 则 G 唯一.

Doolittle 分解的算法

设 A 是 n 阶满秩复方阵且可作三角分解. 由 A 的 Doolittle 分解 $A = LU$, 有

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} 1 & & & \\ l_{21} & 1 & & \\ \vdots & \ddots & \ddots & \\ l_{n1} & \cdots & l_{n,n-1} & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}$$

$$\begin{cases} r_{1j} = a_{1j} \ (j = 1, 2, \cdots, n), \\ l_{i1} = \dfrac{a_{i1}}{r_{11}} \ (i = 2, 3, \cdots, n), \\ r_{kj} = a_{kj} - \sum\limits_{t=1}^{k-1} l_{kt} r_{tj} \ (k \leqslant j \leqslant n; k = 2, \cdots, n), \\ l_{ik} = \frac{1}{r_{kk}}(a_{ik} - \sum\limits_{t=1}^{k-1} l_{it} r_{tk}) \ (k \leqslant i \leqslant n; k = 2, \cdots, n). \end{cases}$$

Crout 分解的算法

设 A 是 n 阶满秩复方阵且可作三角分解. 由 A 的 Doolittle 分解 $A = LU$, 有

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & & & \\ l_{21} & l_{22} & & \\ \vdots & \ddots & \ddots & \\ l_{n1} & \cdots & l_{n,n-1} & l_{nn} \end{bmatrix} \begin{bmatrix} 1 & r_{12} & \cdots & r_{1n} \\ & 1 & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix}$$

$$\begin{cases} l_{i1} = a_{i1} \ (i = 1, 2, \cdots, n), \\ r_{1j} = \dfrac{a_{1j}}{l_{11}} \ (j = 2, 3, \cdots, n), \\ l_{ik} = a_{ik} - \displaystyle\sum_{t=1}^{k-1} l_{it} r_{tk} \ (k \leqslant i \leqslant n; k = 2, \cdots, n), \\ r_{kj} = \frac{1}{l_{kk}} (a_{kj} - \displaystyle\sum_{t=1}^{k-1} l_{kt} r_{tj}) \ (k \leqslant j \leqslant n; k = 2, \cdots, n). \end{cases}$$

Cholesky 分解的算法

设 A 是 n 阶正定 Hermite 矩阵, 存在 n 阶下三角矩阵 $G=(g_{ij})_{n\times n}$ $(g_{ij}=0, i<j)$, 使得

$$A = GG^H = \begin{bmatrix} g_{11} & & & \\ g_{21} & g_{22} & & \\ \vdots & \vdots & \ddots & \\ g_{n1} & g_{n2} & \cdots & g_{nn} \end{bmatrix} \begin{bmatrix} \overline{g_{11}} & \overline{g_{21}} & \cdots & \overline{g_{n1}} \\ & \overline{g_{22}} & \cdots & \overline{g_{n2}} \\ & & \ddots & \vdots \\ & & & \overline{g_{nn}} \end{bmatrix},$$

对角元素 $g_{ii}>0$. 当 A 为实矩阵时, G 也可以取成实矩阵.

$$\begin{cases} g_{11} = \sqrt{a_{11}}, \\ g_{i1} = \dfrac{a_{i1}}{g_{11}} \ (i=2,3,\cdots,n), \\ g_{kk} = \sqrt{a_{kk} - \displaystyle\sum_{t=1}^{k-1} |g_{kt}|^2} \ (k=2,3,\cdots,n), \\ g_{ik} = \frac{1}{g_{kk}}(a_{ik} - \displaystyle\sum_{t=1}^{k-1} g_{it}\overline{g}_{kt}) \ (k\leqslant i\leqslant n; k=2,\cdots,n). \end{cases}$$

例 5.1

求矩阵 $A = \begin{bmatrix} 2 & -1 & 3 \\ 1 & 2 & 1 \\ 2 & 4 & 3 \end{bmatrix}$ 的 Doolittle 分解和 Crout 分解.

例 5.2

求矩阵 $A = \begin{bmatrix} 5 & -2 & 0 \\ -2 & 3 & -1 \\ 0 & -1 & 1 \end{bmatrix}$ 的 Cholesky 分解.

LU 分解的应用

考虑线性方程组 $Ax=b$ 且系数矩阵 $A=LU$, 则方程组等价于两个以三角矩阵系数为系数的线性方程组

$$Ly=b, \quad Ux=y$$

例 5.3

求解线性方程组 $Ax=b$, 其中 $A=\begin{bmatrix}1 & -1 & 2\\ 2 & 0 & 5\\ 3 & 5 & 13\end{bmatrix}, b=\begin{bmatrix}2\\ 8\\ 28\end{bmatrix}$.

$$\begin{bmatrix} 1 & -1 & 2 \\ 2 & 0 & 5 \\ 3 & 5 & 13 \end{bmatrix} \xrightarrow[E_1=\begin{bmatrix} 1 & & \\ -2 & 1 & \\ & & 1 \end{bmatrix}]{E_2=\begin{bmatrix} 1 & & \\ & 1 & \\ -3 & & 1 \end{bmatrix}} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & 1 \\ 0 & 8 & 7 \end{bmatrix} \xrightarrow{E_3=\begin{bmatrix} 1 & & \\ & 1 & \\ & -4 & 1 \end{bmatrix}}$$

$$\begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix} = U$$

$$L = (E_3E_2E_1)^{-1} = \begin{bmatrix} 1 & & \\ -2 & 1 & \\ 5 & -4 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{bmatrix}, \quad A = LU.$$

解线性方程组 $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 4 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \\ 28 \end{bmatrix} \Longrightarrow \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$,

再解线性方程组 $\begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} \Longrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$.

注记 5.9

为什么不在消元的过程中解方程组呢？为什么先做 LU 分解再化成两个线性方程组求解呢？原因是实际中 $Ax = b$, 当 A 不变而 b 经常发生变化时, 这样先分解再求解会避免重复使用消元法, 节省计算量提高效率.

§5.2 初等反射与初等旋转矩阵

定义 5.10

设 $u \in \mathbb{C}^n$ 是单位列向量, 即 $u^H u = 1$, 称

$$H = I - 2uu^H$$

为 Householder 矩阵或者初等反射矩阵.

由 Householder 矩阵 H 确定的 $\mathbb{C}^n$ 上的线性变换

$$y = Hx$$

称为 Householder 变换 或者 初等反射变换.

Householder 矩阵具有的性质如下:

定理 5.11

设 $H \in \mathbb{C}^n$ 是 Householder 矩阵, 则

(1) $H^H = H$ (Hermite 矩阵);

(2) $H^2 = I$ (对合矩阵);

(3) $H^H H = I$ (酉矩阵);

(4) $H^{-1} = H$ (自逆矩阵);

(5) $\begin{bmatrix} I_r & 0 \\ 0 & H \end{bmatrix}$ 是 $n + r$ 阶 Householder 矩阵;

(6) $\det H = -1$. $|\lambda I_m - AB| = \lambda^{m-n}|\lambda I_n - BA|$.

Householder 矩阵的应用主要基于如下的结果:

定理 5.12

设 $z \in \mathbb{C}^n$ 是单位向量, 则对任意 $x \in \mathbb{C}^n$, 存在 Householder 矩阵 H, 使得 $Hx = \lambda z$, 其中 $|\lambda| = \|x\|_2$, 且 $\lambda x^H z$ 为实数.

推论 5.13

对任意 $x \in \mathbb{C}^n$, 存在 Householder 矩阵 $H = I - 2uu^H$, 使得 $Hx = \lambda e_1$, 其中 $|\lambda| = \|x\|_2$, 且 $\lambda x^H e_1 = \lambda \bar{\xi}_1$ 为实数, 这里 ξ_1 为 x 第一个分量.

推论 5.14

对任意 $x \in \mathbb{R}^n$, 存在 Householder 矩阵

$$H = I - 2uu^T \quad (u \in \mathbb{R}^n \text{ 且 } u^T u = 1),$$

使得 $Hx = \lambda e_1$, 其中 $\lambda = \pm\|x\|_2$.

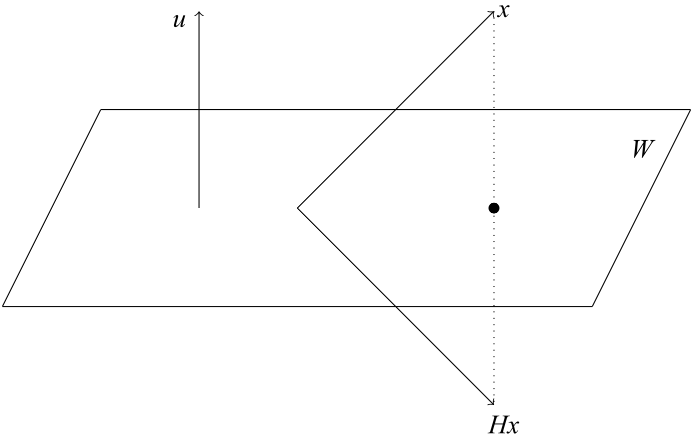
以上两个推论称为用 Householder 变换化向量 x 与 e_1 共线.

例 5.4

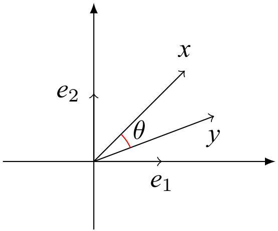
用 Householder 变换化下列向量与 e_1 共线;

$$(1)\ x = (1, 2, 2)^T; \qquad (2)\ x = (-2\,\mathrm{i}; \mathrm{i}, 2)^T.$$

Householder (初等反射) 矩阵几何解释



Givens 矩阵



在 $\mathbb{R}^2$ 中, 向量 $x=(\xi_1,\xi_2)$ 按照顺时针方向旋转 θ 变成 $y=(\eta_1,\eta_2)$, 则他们长度相等, 且坐标满足

$$\begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix}=\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}$$

称 $T=\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}$ 为平面旋转矩阵, 它是正交矩阵, 推广到 $\mathbb{C}^n$ 上即得如下定义:

定义 5.15

设 $c, s \in \mathbb{C}$, 且满足 $|c|^2 + |s|^2 = 1$, 称 n 阶方阵

$$T_{pq} = \begin{bmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & \bar{c} & & & \bar{s} & & \\ & & & & 1 & & & & \\ & & & & & \ddots & & & \\ & & & & & & 1 & & \\ & & & -s & & & c & & \\ & & & & & & & 1 & \\ & & & & & & & & \ddots \\ & & & & & & & & & 1 \end{bmatrix}$$

为 Givens 矩阵或初等旋转矩阵. T_{pq} 是酉矩阵且 $|T_{pq}| = 1$.

由 Givens 矩阵 T_{pq} 确定的 $\mathbb{C}^n$ 上的线性变换

$$y = T_{pq} x$$

称为 Givens 变换或初等旋转变换.

Givens 矩阵的基本性质

Givens 矩阵的应用主要基于如下:

定理 5.16

对任意 $x = (\xi_1, \xi_2, \cdots, \xi_n)^T \in \mathbb{C}^n$, 存在 Givens 矩阵 T_{pq}, 使得 $T_{pq}x$ 的第 p 个分量为 $\sqrt{|\xi_p|^2 + |\xi_q|^2}$, 第 q 个分量为 0, 其他分量不变.

推论 5.17

对任意 $x = (\xi_1, \xi_2, \cdots, \xi_n)^T \in \mathbb{C}^n$, 存在一列 Givens 矩阵 $T_{12}, T_{13}, \cdots, T_{1n}$, 使得

$$T_{1n} \cdots T_{13} T_{12} x = \|x\|_2 e_1.$$

称之为用 Givens 变换化向量 x 与 e_1 同方向.

例 5.5

用 Givens 变换化下列向量与 e_1 同方向:

(1) $x = (1, 2, 2)^T$; (2) $x = (-2\,\mathrm{i}, \mathrm{i}, 2)^T$.

§5.3 矩 阵 的 QR 分解

定义 5.18

如果存在 n 阶酉矩阵 Q 和 n 阶上三角矩阵 R，使得

$$A = QR$$

则称之为 A 的 QR 分解或者酉三角分解. 当 $A \in \mathbb{R}^{n\times n}$, 称之为 A 的正交三角分解.

定理 5.19

任意 $A \in \mathbb{C}^{n\times n}$ 都可以作 QR 分解.

例 5.6

求矩阵 $A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$ 的 QR 分解.

(1): 用 Household 变换

令 $\alpha_1 = (0,1,0,1)^T$, 则 $\|\alpha_1\|_2 = \sqrt{2}$, 取 $\lambda_1 = \sqrt{2}$,

$$\mu_1 = \frac{\alpha_1 - \lambda_1 e_1}{\|\alpha_1 - \lambda_1 e_1\|_2} = \frac{1}{2}\begin{bmatrix} -\sqrt{2} \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

$$H_1 = I_4 - 2\mu_1\mu_1^H = \begin{bmatrix} 0 & \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

$$H_1 A = \begin{bmatrix} \sqrt{2} & 0 & \sqrt{2} & 0 \\ 0 & \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 & 1 \\ 0 & \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{bmatrix}$$

令 $\beta_2=(\frac{\sqrt{2}}{2},1,\frac{\sqrt{2}}{2})^T$, 则 $\|\beta_2\|_2=\sqrt{2}$. 取 $\lambda_2=\sqrt{2}$,

$$\widetilde{\mu}_2=\frac{\beta_2-\lambda_2\tilde{e}_1}{\|\beta_2-\lambda_2\tilde{e}_1\|_2}=\frac{1}{\sqrt{2}}\begin{bmatrix}-\frac{\sqrt{2}}{2}\\ 1\\ \frac{\sqrt{2}}{2}\end{bmatrix}$$

$$\widetilde{H}_2=I_3-2\widetilde{\mu}_2\widetilde{\mu}_2^H=\begin{bmatrix}\frac{1}{2} & \frac{\sqrt{2}}{2} & \frac{1}{2}\\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2}\\ \frac{1}{2} & -\frac{\sqrt{2}}{2} & \frac{1}{2}\end{bmatrix}$$

$$H_2=\begin{bmatrix}1 & \\ & \widetilde{H}_2\end{bmatrix},\ H_2H_1A=\begin{bmatrix}\sqrt{2} & 0 & \sqrt{2} & 0\\ 0 & \sqrt{2} & 0 & \sqrt{2}\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}=R,$$

$$A = H_1H_2\begin{bmatrix} \sqrt{2} & 0 & \sqrt{2} & 0 \\ 0 & \sqrt{2} & 0 & \sqrt{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},\text{令}$$

$$Q = H_1H_2 = \begin{bmatrix} 0 & \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} & 0 \\ 0 & \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} & 0 \end{bmatrix}.$$

则

$$A = \begin{bmatrix} 0 & \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} & 0 \\ 0 & \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & 0 & -\frac{\sqrt{2}}{2} & 0 \end{bmatrix}\begin{bmatrix} \sqrt{2} & 0 & \sqrt{2} & 0 \\ 0 & \sqrt{2} & 0 & \sqrt{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

(2): 用 Givens 变换

$$T_{12}=\begin{bmatrix}0&1&0&0\\-1&0&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix},\ T_{12}A=\begin{bmatrix}1&0&1&0\\0&-1&0&-1\\0&1&0&1\\1&0&1&0\end{bmatrix},$$

$$T_{14}=\begin{bmatrix}\frac{1}{\sqrt{2}}&0&0&\frac{1}{\sqrt{2}}\\0&1&0&0\\0&0&1&0\\-\frac{1}{\sqrt{2}}&0&0&\frac{1}{\sqrt{2}}\end{bmatrix},\ T_{14}T_{12}A=\begin{bmatrix}\sqrt{2}&0&\sqrt{2}&0\\0&-1&0&-1\\0&1&0&1\\0&0&0&0\end{bmatrix},$$

$$T_{23}=\begin{bmatrix}1&0&0&0\\0&-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}&0\\0&-\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}&0\\0&0&0&1\end{bmatrix},\ T_{23}T_{14}T_{12}A=\begin{bmatrix}\sqrt{2}&0&\sqrt{2}&0\\0&\sqrt{2}&0&\sqrt{2}\\0&0&0&0\\0&0&0&0\end{bmatrix}=R$$

$$A = \underbrace{T_{12}^T T_{14}^T T_{23}^T}_{Q} R = \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} \sqrt{2} & 0 & \sqrt{2} & 0 \\ 0 & \sqrt{2} & 0 & \sqrt{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

练习 5.1

用 Givens 变换求矩阵 $A=\begin{bmatrix}4&5&0&1\\0&4&0&1\\0&0&2&1\\3&0&0&2\end{bmatrix}$ 的 QR 分解.

解:

取 $T_{14}=\dfrac{1}{5}\begin{bmatrix}4&0&0&3\\0&5&0&0\\0&0&5&0\\-3&0&0&4\end{bmatrix}$, $T_{14}A=\begin{bmatrix}5&4&0&2\\0&4&0&1\\0&0&2&1\\0&-3&0&1\end{bmatrix}$,

取 $T_{24}=\dfrac{1}{5}\begin{bmatrix}5&0&0&0\\0&4&0&-3\\0&0&5&0\\0&3&0&4\end{bmatrix}$, $T_{24}T_{14}A=\begin{bmatrix}5&4&0&2\\0&5&0&\frac{1}{5}\\0&0&2&1\\0&0&0&\frac{7}{5}\end{bmatrix}=R$,

取 $Q = T_{14}^H T_{24}^H = \dfrac{1}{25}\begin{bmatrix} 4 & 0 & 0 & -3 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 3 & 0 & 0 & 4 \end{bmatrix}\begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 4 & 0 & 3 \\ 0 & 0 & 5 & 0 \\ 0 & -3 & 0 & 4 \end{bmatrix} =$
$\dfrac{1}{25}\begin{bmatrix} 20 & 9 & 0 & -12 \\ 0 & 20 & 0 & 15 \\ 0 & 0 & 25 & 0 \\ 15 & -12 & 0 & 16 \end{bmatrix}$,

$$A = QR = (T_{14}^H T_{24}^H)R = \left(\frac{1}{25}\begin{bmatrix} 20 & 9 & 0 & -12 \\ 0 & 20 & 0 & 15 \\ 0 & 0 & 25 & 0 \\ 15 & -12 & 0 & 16 \end{bmatrix}\right)\begin{bmatrix} 5 & 4 & 0 & 2 \\ 0 & 5 & 0 & \frac{1}{5} \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & \frac{7}{5} \end{bmatrix}.$$

练习 5.2

用 Householder 变换求矩阵 $A = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 1 & 0 & 3 & 1 \\ 1 & 0 & 3 & 2 \\ 1 & 2 & 0 & 2 \end{bmatrix}$ 的 QR 分解.

长矩阵的 QR 分解

若 $A \in \mathbb{C}^{m\times n}$, 同样可以做 QR 分解, 证明方法类似, 则有如下结论

定理 5.20

设 $A \in \mathbb{C}^{m\times n}$, 则存在酉矩阵 $Q \in \mathbb{C}^{m\times m}$, 使得 $A = QR$, 其中 $R \in \mathbb{C}^{m\times n}$ 是阶梯型矩阵.

例 5.7

求矩阵 $\begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$ 的 QR 分解.

用 Household 变换求出 Q 和前面一样, 只是 $R = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

§5.4 矩阵的奇异值分解

定理 5.21

设 $A \in \mathbb{C}^{m\times n}$, 则

(1) A^HA 和 AA^H 的特征值均为非负实数.

(2) A^HA 和 AA^H 的非零特征值相同.

(3) $\mathrm{rank}(A^HA) = \mathrm{rank}(AA^H) = \mathrm{rank}\, A = \mathrm{rank}\, A^H$.

奇异值的定义

定义 5.22

设 $A \in \mathbb{C}^{m\times n}$ 是秩为 $r(\,r>0)$ 的矩阵, A^HA 的特征值为

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0,$$

则称 $\sigma_i = \sqrt{\lambda_i}\ (i = 1, 2, \cdots, n)$ 为 A 的奇异值.

注记 5.23

由 $\operatorname{rank}(A^HA) = \operatorname{rank} A$, 且 A^HA 与 AA^H 有相同的非零特征值, 从而 A 非零奇异值个数等于 $\operatorname{rank} A$, 且 A 与 A^H 有相同的非零奇异值.

酉等价定义和性质

定义 5.24

设 $A, B \in \mathbb{C}^{m\times n}$. 若存在 m 阶酉矩阵 U 和 n 阶酉矩阵 V 使得

$$U^H A V = B,$$

则称 A 与 B 酉等价.

定理 5.25

酉等价矩阵有相同的奇异值.

下面是本节的主要定理.

定理 5.26

设 $A \in \mathbb{C}^{m\times n}$ 是秩为 $r(\,r>0)$ 的矩阵, 则存在 m 阶酉矩阵 U 和 n 阶酉矩阵 V, 使得

$$U^H A V = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix},$$

其中 $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$, 而 $\sigma_i (i = 1, 2, \cdots, r)$ 为 A 的非零奇异值. 或者将上式改写为

$$A = U \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^H,$$

称之为 A 的奇异值分解.

推论 5.27

设 $A \in \mathbb{C}^{n\times n}$ 是满秩方阵, 则存在 n 阶酉矩阵 U 和 V, 使得

$$U^H AV = \mathrm{diag}\,(\sigma_1, \sigma_2, \cdots, , \sigma_n),$$

其中 $\sigma_i > 0(\ i = 1, 2, \cdots, n)$ 为 A 的奇异值.

例 5.8

求矩阵 $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ 的奇异值分解.

推论 5.28

设 $A \in \mathbb{C}^{m\times n}$ 是秩为 $r(\, r > 0)$ 的矩阵, 则 A 的奇异值分解 $A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^H$ 中, V 的列向量是 $A^H A$ 的单位正交特征向量, 而 U 的列向量是 AA^H 的单位正交特征向量.

注记 5.29

这就给出了求 U 和 V 的方法, 但结果需要检验.

§5.5 矩阵的满秩分解

定义 5.30 (Hermtie 标准形)

设 H 是秩为 $r\ (r>0)$ 的 $m\times n$ 矩阵, 如果 H 满足:

(1) 前 r 行中每行至少含有一个非零元素, 且第一个非零元素为 1, 而后 $m-r$ 行只含有零元素;

(2) 各列中, 如果含某行的第一个非零元 1, 则其余各元素均为 0;

(3) 若第 i 行的第一个非零元 1 位于第 j_i 列 $(i=1,\cdots,r)$ 则

$$j_1<j_2<\cdots<j_r$$

则称 H 为 Hermite 标准形或者行最简形式.

Hermite 标准型的基本形式

秩为 r 的 Hermite 标准型或者行最简形式具有如下形式:

$$
H = \left[\begin{array}{ccc|cccc|ccc|cccc}
0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & 0 & * & \cdots & * \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots & 0 & * & \cdots & * \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & * & \cdots & * \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & 0 & * & \cdots & * \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & 1 & * & \cdots & * \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \cdots & \vdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0
\end{array}\right]
$$

关于标准型的基本结论

定理 5.31

任意秩为 $r\ (r > 0)$ 的 $m \times n$ 矩阵 A 都可以通过初等行变换化为 Hermite 标准形 H, 或者采用矩阵的说法为: 存在 m 阶可逆矩阵 S 使得 $SA = H$.

注记 5.32

- 矩阵 A 的 Hermite 标准形是唯一的, 但是变换矩阵 S 不一定唯一; 且从标准形中可以直接读出矩阵 A 的秩;
- 具体求法: 构造 $m \times (m + n)$ 矩阵 (A, I_m), 由

$$S(A, I_m) = (SA, S) = (H, S)$$

例 5.9

求矩阵 $A=\begin{bmatrix}2&4&1&1\\1&2&-1&2\\-1&-2&-2&1\end{bmatrix}$ 的 Hermite 标准形 H 和变换矩阵 S.

置换矩阵的定义和基本性质

定义 5.33

以 n 阶单位矩阵 I_n 的 n 个列向量 $e_1, e_2, \cdots, e_n$ 为列构成的 n 阶方阵

$$P = (e_{i_1}, e_{i_2}, \cdots, e_{i_n})$$

称为 n 阶置换矩阵, $\{i_1, \cdots, i_n\}$ 是 $1, \cdots, n$ 的一个全排列.

定理 5.34

设 $P = (e_{i_1}, e_{i_2}, \cdots, e_{i_n})$ 是置换矩阵, 则

- P 是正交矩阵;
- 对任意 $A \in \mathbb{C}^{m\times n}$, AP 是将 A 的列按 $i_1, i_2, \cdots, i_n$ 的次序重新排列所得矩阵.

等价标准型

定理 5.35

设 $m \times n$ 矩阵 A 的秩为 $r(> 0)$, 则存在 m 阶可逆矩阵 S 与 n 阶置换矩阵 P, 使得 $SAP = \begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix}$.

如果对上面的矩阵继续进行初等列变换, 则有如下结果:

定理 5.36

设 $m \times n$ 矩阵 A 的秩为 $r(> 0)$, 则存在 m 阶可逆矩阵 S 与 n 阶可逆矩阵 T, 使得

$$SAT = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{(称之为 } A \text{ 的等价标准形)}.$$

具体求出 S 和 T 的方法:

- 方法 1:

$$(A, I_m) \xrightarrow{\text{初等行变换}} (H, S), \quad \begin{pmatrix} H \\ I_n \end{pmatrix} \xrightarrow{\text{初等列变换}} \begin{pmatrix} \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} \\ T \end{pmatrix},$$

则 $SAT = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$;

- 方法 2: $\begin{pmatrix} A & I_m \\ I_n & 0 \end{pmatrix} \xrightarrow[\text{前}n\text{列做初等列变换}]{\text{前}m\text{行做初等行变换}} \begin{pmatrix} \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} & S \\ T & 0 \end{pmatrix}.$

例 5.10

已知矩阵 $A = \begin{bmatrix} 2 & 4 & 1 & 1 \\ 1 & 2 & -1 & 2 \\ -1 & -2 & -2 & 1 \end{bmatrix}$

(1): 求 A 的 Hermite 标准型 H 和所用的变换矩阵 S;

(2): 求置换矩阵 P 使之成为 $\begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix}$ 的形式;

(3): 求 A 的等价标准型和所用的矩阵 S 和 T.(保留)

矩阵的满秩分解

定义 5.37

设 $A \in \mathbb{C}^{m\times n}$ 是秩为 r 的矩阵, 如果存在秩为 r 的矩阵 $F \in \mathbb{C}^{m\times r}$ 和 $G \in \mathbb{C}^{r\times n}$ 使得

$$A = FG,$$

则称之为 A 的满秩分解.

定理 5.38

设 $A \in \mathbb{C}^{m\times n}$ 是秩为 r 的矩阵, 则 A 的满秩分解总是存在的.

例 5.11

求矩阵 $A=\begin{bmatrix} 2 & 4 & 1 & 1 \\ 1 & 2 & -1 & 2 \\ -1 & -2 & -2 & 1 \end{bmatrix}$ 的满秩分解.

满秩分解的另一方法

定理 5.39

设 $A \in \mathbb{C}^{m\times n}$ 是秩为 $r\,(r>0)$ 的矩阵, 且 A 的 Hermtie 标准形为 H, 则在 A 的满秩分解 $A = FG$ 中, 可以取 F 为 A 的 $j_1, j_2, \cdots, j_r$ 列（前 r 个非零行中第 1 个非零元 1 所在的列）所构成的 $m\times r$ 矩阵, G 为 H 的前 r 行构成的 $r\times n$ 矩阵.

例 5.12

求矩阵 $A = \begin{bmatrix} 2 & 4 & 1 & 1 \\ 1 & 2 & -1 & 2 \\ -1 & -2 & -2 & 1 \end{bmatrix}$ 的满秩分解.

§5.6 应用案例二、QR 分解与最小二乘问题

1. **QR 迭代算法**: 求解一般方阵全部特征值与特征向量的最有效的方法之一, 基本步骤是: 设 $A \in \mathbb{C}^{n\times n}$, 记 $A_1 = A$, 求 A_1 的 QR 分解 $A_1 = Q_1R_1$, 令

$A_2 = R_1Q_1$, 求 A_2 的 QR 分解 $A_2 = Q_2R_2$;

$$\cdots, A_k = Q_kR_k, \text{令} A_{k+1} = R_kQ_k (k = 1, 2, \cdots)$$

其中 Q_k 是酉矩阵, R_k 是上三角矩阵, 如此迭代下去, 得到矩阵序列

$A_1, A_2, A_3, \cdots$, 在一定条件下, 当 $k \to \infty$ 时, A_k 接近于上三角矩阵 (其实为 A 的 Schur 分解标准型), 其对角线上的元素为 A 的特征值.

由于 $A_{k+1} = R_kQ_k = Q_k^HQ_kR_kQ_k = Q_k^HA_kQ_k$ 知 $A_{k+1} = Q_k^H\cdots Q_1^HA_kQ_1\cdots Q_k$, 令 $Q = Q_1\cdots Q_k \Longrightarrow A_{k+1} = Q^HAQ$. 即 A_{k+1} 酉相似于 A, 所以当 k 充分大时, A_k 充分接近于 A 的 Schur 标准型.

例 5.13

用 QR 迭代算法求矩阵 $A=\begin{bmatrix}2 & 3\\3 & 6\end{bmatrix}$ 的 Schur 分解.

用数学软件模拟数值计算

$$A_1=\begin{bmatrix}2 & 3\\3 & 6\end{bmatrix}\Longrightarrow A_2=\begin{bmatrix}7.5385 & 0.6923\\0.6923 & 0.4615\end{bmatrix}\Longrightarrow A_3=\begin{bmatrix}7.6054 & -0.0362\\-0.0362 & 0.3946\end{bmatrix}\Longrightarrow$$
$$A_4=\begin{bmatrix}7.6056 & 0.0019\\0.0019 & 0.3944\end{bmatrix}\Longrightarrow A_5=\begin{bmatrix}7.6056 & -0.0001\\-0.0001 & 0.3944\end{bmatrix}\Longrightarrow A_6=$$
$$\begin{bmatrix}7.6056 & 0.0000\\0.0000 & 0.3944\end{bmatrix},$$

而真实的特征值为 $\lambda_1=\dfrac{8-\sqrt{52}}{2}\approx 0.39455,\lambda_2=\dfrac{8+\sqrt{52}}{2}\approx 7.60555.$

1.2 QR 分解与最小二乘问题

设 $A \in \mathbb{C}^{m\times n}, \mathbf{b} \in \mathbb{C}^m$. 当线性方程组 $Ax = b$ 有解时, 也称为方程组相容, 当线性方程组 $A\mathbf{x} = \mathbf{b}$ 无解时, 也称方程组矛盾或不相容.

当 $Ax = b$ 无解时, 我们希望找一个向量 $\mathbf{x_0}$ 使得 $\|A\mathbf{x} - \mathbf{b}\|_2$ 达到最小, 即

$$\|A\mathbf{x_0} - \mathbf{b}\|_2 = \min_{\mathbf{x}\in\mathbb{C}^n} \|A\mathbf{x} - \mathbf{b}\|_2.$$

称 $\mathbf{x_0}$ 为矛盾方程组 $A\mathbf{x} = \mathbf{b}$ 的最小二乘解.

线性方程组 $Ax=b$, 设 $A=QR, Q$ 为酉矩阵, 若 x_0 满足

$$\|Ax_0-b\|_2\text{最小} \Longleftrightarrow \|Q^H(Ax_0-b)\|_2\text{最小} \Longleftrightarrow \|Rx_0-Q^Hb\|_2\text{最小}$$

即 $Ax=b$ 的最小二乘解与 $Rx=Q^Hb$ 的最小二乘解相同.

例 5.14

设 $A=\begin{bmatrix}\sqrt{2} & 1\\ 0 & 1\\ 0 & 1\end{bmatrix}, b=\begin{bmatrix}2\\1\\2\end{bmatrix}$, 求 $Ax=b$ 的最小二乘解.

解: 用 Householder 变换知 $A = QR = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 1 \\ 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix}$

考虑 $Rx = Q^H b = \begin{bmatrix} 2 \\ \frac{3}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$, 原方程组化为 $\begin{cases} \sqrt{2}x_1 + x_2 & = 2 \\ \sqrt{2}x_2 & = \frac{3}{\sqrt{2}} \\ 0 & = -\frac{1}{\sqrt{2}} \end{cases}$,

则最小二乘解为 $x_0 = \begin{bmatrix} \frac{\sqrt{2}}{4} \\ 1.5 \end{bmatrix}$, 误差为 $\|Ax_0 - b\|_2 = \frac{\sqrt{2}}{2}$. □

1.3 奇异值分解与最小二乘问题

定理 5.40

设 $A \in \mathbb{C}^{m\times n}$ 是秩为 $r(\ r > 0)$ 的矩阵，$b \in \mathbb{C}^m$, 且 A 的奇异值分解为

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^H,$$

则 $x_0 = V \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^H b$ 是矛盾方程组 $A\mathbf{x} = \mathbf{b}$ 的最小二乘解.

证明思路: 根据 2-范数的酉不变性有

$$\|Ax - b\|_2 = \left\| U \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} V^H x - b \right\|_2 = \left\| \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} V^H x - U^H b \right\|_2$$

$Ax=b$ 的最小二乘解即是 $\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} V^H x = U^H b$ 的最小二乘解.

令 $y = V^H x, c = U^H b$, 则 $\begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} y = c$ 等价于

$$\begin{cases} \sigma_1 y_1 = c_1 \\ \sigma_2 y_2 = c_2 \\ \vdots \\ \sigma_r y_r = c_r \\ 0 = c_{r+1} \\ \vdots \\ 0 = c_m \end{cases} \Longrightarrow \begin{cases} y_1 = c_1/\sigma_1 \\ \vdots \\ y_r = c_r/\sigma_r \\ y_{r+1} = 0 \\ \vdots \\ y_n = 0 \end{cases} \Longrightarrow y = \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^H b,$$

这 $y_{r+1}, \cdots, y_n$ 为自由变量. 替代回去得 $x_0 = V \begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{pmatrix} U^H b$ 为最小二乘解.

注记 5.41

从推导过程看出, 取不同得 $y_{r+1}, \cdots, y_n$ 得到的都是最小二乘解, 不过 x_0 是最小二乘解中范数最小的.

如果最小二乘解不唯一, 在所有最小二乘解中, 2-范数最小的 x_0 称为**极小范数最小二乘解**或**最佳逼近解**.

例 5.15

设 $A = \begin{bmatrix} \sqrt{2} & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, b = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$, 求 $Ax = b$ 的最小二乘解.

解: 先求出 A 的奇异值分解

$$A = U\begin{pmatrix}\Sigma\\0\end{pmatrix}V^H = \begin{bmatrix}\frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{3}} & 0\\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}2 & 0\\0 & 1\\0 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}}\\ -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}}\end{bmatrix}$$

则有最小二乘解 $x_0 = V\begin{pmatrix}\Sigma^{-1}\\0\end{pmatrix}U^H b = \begin{pmatrix}\frac{\sqrt{2}}{4}\\ \frac{3}{2}\end{pmatrix}$ □

第 6 章 广义逆矩阵

§6.1 广义逆矩阵与线性方程组

定义 6.1

设 $A \in \mathbb{C}^{m\times n}$. 如果存在 $X \in \mathbb{C}^{n\times m}$ 满足四个 Penrose 方程

(1) $AXA = A$,

(2) $XAX = X$,

(3) $(AX)^H = AX$,

(4) $(XA)^H = XA$

中的某几个或者全部, 则称 X 为 A 的广义逆矩阵. 满足全部四个方程的广义逆矩阵 X 称为 A 的 Moore-Penrose 逆.

注记 6.2

- 当 A 可逆时, 取 $X = A^{-1}$ 满足四个 Penrose 方程.
- 按照定义, 可以分为满足一个, 二个, 三个或四个 Penrose 方程的广义逆矩阵共有

$$C_4^1 + C_4^2 + C_4^3 + C_4^4 = 15 \text{ 类}.$$

Moore-Penrose 逆的存在唯一性

定理 6.3

设 $A \in \mathbb{C}^{m\times n}$, 则 A 的 Moore-Penrose 逆存在且唯一.

设 X, Y 都满足 MP 方程, 则

$$\begin{aligned} X &= XAX = X(AX)^H = XX^HA^H = XX^H(AYA)^H \\ &= X(AYAX)^H = X(AX)^H(AY)^H = XAXAY = XAY \\ &= XAYAY = (XA)^H(YA)^HY = (YAXA)^HY \\ &= (YA)^HY = YAY = Y. \end{aligned}$$

注: 只要 A 不是可逆矩阵, 则除了 Moore-Penrose 逆以外的其他 14 类广义逆矩阵都不是唯一的.

逆的相关定义

定义 6.4

设 $A \in \mathbb{C}^{m\times n}$. 若 $X \in \mathbb{C}^{n\times m}$ 满足 Penrose 方程中的第 $(i), (j), \cdots, (l)$ 等方程, 则称 X 为 A 的 $\{i, j, \cdots, l\}$ 逆, 记为 $A^{(i,j,\cdots,l)}$, 其全体记为 $A\{i, j, \cdots, l\}$. A 的唯一的 Moore-Penrose 逆记为 A^{+}, 也称之为 A 的加号逆.

注记 6.5

后面应用较多的是一下 2 类:

$$A\{1\},\ A^{+}.$$

由于 $\{1\}$ 基本, A^{+} 唯一, 所以下面主要讨论这两类广义逆矩阵的性质.

$\{1\}$ 逆

定理 6.6

设 $A \in \mathbb{C}^{m\times n}$ 是秩为 $r(\, r > 0)$ 的矩阵, 且有行满秩矩阵 $S \in \mathbb{C}^{m\times m}$ 和 n 阶置换矩阵 P 使得

$$SAP = \begin{bmatrix} I_r & K \\ 0 & 0 \end{bmatrix} \quad (K \in \mathbb{C}^{r\times(n-r)}),$$

则对任意 $L \in \mathbb{C}^{(n-r)\times(m-r)}$, $n \times m$ 矩阵

$$X = P \begin{bmatrix} I_r & 0 \\ 0 & L \end{bmatrix} S \tag{6.1}$$

是 A 的 $\{1\}$ 逆.

注: (6.1) 中 L 可任意变换, 但所得矩阵 X 只是 $A\{1\}$ 的一子集.

例 6.1

已知矩阵 $A = \begin{bmatrix} 1 & 2 & -1 & 2 \\ 2 & 4 & 1 & 1 \\ -1 & -2 & -2 & 1 \end{bmatrix}$, 求 $A^{(1)}$.

问题: 如何求出所有的 $\{1\}$ 逆呢?

定理 6.7

设 $A \in \mathbb{C}^{m\times n}$ 是秩为 $r(\ r>0)$ 的矩阵, 且有 m 阶可逆方阵 S 和 n 阶可逆方阵 T, 使得

$$SAT = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix},$$

则

$$A\{1\} = \left\{ T \begin{bmatrix} I_r & L_{12} \\ L_{21} & L_{22} \end{bmatrix} S \middle| L_{12} \in \mathbb{C}^{r\times(m-r)}, L_{21} \in \mathbb{C}^{(n-r)\times r},\ L_{22} \in \mathbb{C}^{(n-r)\times(m-r)} \right\}.$$

推论 6.8

设 $A \in \mathbb{C}^{m\times n}$, 则 A 有唯一 $\{1\}$ 逆的充分必要条件是 $m=n$, 且 $\operatorname{rank} A = n$, 即 A 可逆. 此时唯一的 $\{1\}$ 逆为 A^{-1}.

$\{1\}$ 逆的性质

命题 6.9

设 $A \in \mathbb{C}^{m\times n}$, $A^{(1)} \in A\{1\}$, 则

(1) $(A^{(1)})^H \in A^H\{1\}$, $(A^{(1)})^T \in A^T\{1\}$;

(2) $\lambda^+ A^{(1)} \in (\lambda A)\{1\}$, 其中 $\lambda \in \mathbb{C}$, 且

$$\lambda^+ = \begin{cases} \lambda^{-1}, & \lambda \neq 0, \\ 0, & \lambda = 0; \end{cases}$$

(3) 当 $S \in \mathbb{C}_m^{m\times m}$, $T \in \mathbb{C}_n^{n\times n}$, 有

$$T^{-1}A^{(1)}S^{-1} \in (SAT)\{1\};$$

(4) $\operatorname{rank}(A^{(1)}) \geqslant \operatorname{rank} A$;

(5) $\operatorname{rank}(AA^{(1)}) = \operatorname{rank}(A^{(1)}A) = \operatorname{rank} A$;

(6) $AA^{(1)} = I_m \iff \operatorname{rank} A = m$;

(7) $A^{(1)}A = I_n \iff \operatorname{rank} A = n$.

$\{1\}$ 逆的应用

利用 $\{1\}$ 逆可以求解矩阵方程及线性方程组.

引理 6.10

设 $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{p\times q}$, $D \in \mathbb{C}^{m\times q}$, 则矩阵方程

$$AXB = D$$

有解的充分必要条件是

$$AA^{(1)}DB^{(1)}B = D,$$

其中 $A^{(1)} \in A\{1\}$, $B^{(1)} \in B\{1\}$. 当矩阵方程有解时, 其通解为

$$X = A^{(1)}DB^{(1)} + Y - A^{(1)}AYBB^{(1)} \quad (Y \in \mathbb{C}^{n\times p} \text{ 任意}).$$

推论 6.11

设 $A \in \mathbb{C}^{m\times n}$, $A^{(1)} \in A\{1\}$, 则有

$$A\{1\} = \{A^{(1)} + Z - A^{(1)}AZAA^{(1)} \mid Z \in \mathbb{C}^{n\times m} \text{ 任意}\}.$$

说明用某个给定的 $A^{(1)}$ 就能刻画集合 $A\{1\}$ 的全部元素.

定理 6.12

设 $A \in \mathbb{C}^{m\times n}$, $b \in \mathbb{C}^m$, 则线性方程组 $Ax = b$ 有解的充分必要条件是

$$AA^{(1)}b = b,$$

其中 $A^{(1)} \in A\{1\}$. 如果 $Ax = b$ 有解, 其通解为

$$x = A^{(1)}b + (I - A^{(1)}A)y \quad (y \in \mathbb{C}^n \text{ 任意}).$$

例 6.2

用广义逆矩阵方法求解线性方程组

$$\begin{cases} x_1 + 2x_2 - x_3 + 2x_4 = 1 \\ 2x_1 + 4x_2 + x_3 + x_4 = 5 \\ -x_1 - 2x_2 - 2x_3 + x_4 = -4 \end{cases}$$

§6.2 Moore-Penrose 逆 A^+ 及 应 用

之前用奇异值分解给出了求 A^+ 的方法, 这里给出用满秩分解求 A^+ 的方法较为简便.

定理 6.13

设 $A \in \mathbb{C}^{m\times n}$ 是秩为 $r(\, r > 0)$ 的矩阵, 且 A 的满秩分解为

$$A = FG \qquad (F \in \mathbb{C}^{m\times r},\ G \in \mathbb{C}^{r\times n} \text{ 秩都为 } r),$$

则

$$A^+ = G^H(GG^H)^{-1}(F^HF)^{-1}F^H.$$

推论 6.14

设 $A \in \mathbb{C}^{m\times n}$, 则当 $\operatorname{rank} A = m$ 时, 有

$$A^+ = A^H(AA^H)^{-1},$$

而当 $\operatorname{rank} A = n$ 时, 有

$$A^+ = (A^HA)^{-1}A^H.$$

例 6.3

求下列矩阵的 Moore-Penrose 逆:

$$(1)\ A = \begin{bmatrix} 1 & 2 & -1 & 2 \\ 2 & 4 & 1 & 1 \\ -1 & -2 & -2 & 1 \end{bmatrix}; \quad (2)\ A = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

A^+ 的性质

定理 6.15

设 $A \in \mathbb{C}^{m\times n}$, 则

(1) $(A^+)^+ = A$; $(A^+)^H = (A^H)^+$, $(A^+)^T = (A^T)^+$;

(2) $(\lambda A)^+ = \lambda^+ A^+$, 其中 $\lambda \in \mathbb{C}$, 且 λ^+ 如前所述.

(3) $\operatorname{rank} A^+ = \operatorname{rank} A$; $\operatorname{rank}(AA^+) = \operatorname{rank}(A^+A) = \operatorname{rank} A$;

(4) $A^+ = (A^HA)^+A^H = A^H(AA^H)^+$;

(5) $(A^HA)^+ = A^+(A^H)^+$, $(AA^H)^+ = (A^H)^+A^+$;

(6) 当 U, V 分别是 m 阶和 n 阶酉矩阵时, 有 $(UAV)^+ = V^HA^+U^H$;

(7) $AA^+ = I_m \Longleftrightarrow \operatorname{rank} A = m$;

(8) $A^+A = I_n \Longleftrightarrow \operatorname{rank} A = n$.

注记 6.16 (与逆矩阵不一样的性质举例)

- 设 $A = (1\ 0)$, $B = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, 于是 $AB = (1)$, 而

$$(AB)^+ = (1), \quad A^+ = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad B^+ = (1\ 0),$$

  所以

$$B^+A^+ = (\tfrac{1}{2}) \neq (AB)^+.$$

- 设 $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, 则 $A = FG = \binom{1}{0}(1\ 1)$, 所以 $A^+ = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$, 但是

$$AA^+ = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \neq \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = A^+A.$$

A^+ 在解线性方程组中的应用

利用 $\{1\}$ 逆已经解决了判定线性方程组是否有解及通解的问题. 由于 A^+ 是特殊的 $\{1\}$ 逆, 所以相应地有如下定理:

定理 6.17

设 $A \in \mathbb{C}^{m\times n}$, $b \in \mathbb{C}^m$, 则线性方程组 $Ax = b$ 有解的充分必要条件是

$$AA^+b = b,$$

且通解为

$$x = A^+b + (I - A^+A)y \quad (y \in \mathbb{C}^n \text{ 任意}).$$

当且仅当 $A^+A = I\,(\iff \operatorname{rank} A = n)$ 时, 有唯一解.
但是在实际问题中, 当线性方程组的无穷多个解时, 常常需求出 2-范数最小解, 即

$$\|x_0\|_2 = \min_{Ax=b} \|x\|_2.$$

称 x_0 为线性方程组 $Ax = b$ 极小范数解.

定理 6.18

设 $A \in \mathbb{C}^{m\times n}$, $b \in \mathbb{C}^m$, 且 $Ax = b$ 有解, 则它唯一极小范数解为 $x_0 = A^+b$.

当线性方程组无解时, 即对任意的 $x \in \mathbb{C}^n$ 都有 $Ax - b \neq 0$. 往往希望求出这样的 $z \in \mathbb{C}$, 使得 $\|Az - b\|_2$ 为最小, 即

$$\|Az - b\|_2 = \min_{x\in\mathbb{C}^n} \|Ax - b\|_2,$$

即它的最小二乘解 (具体参见 3.5.3). 利用 Moore-Penrose 逆可以解决这一问题.

定理 6.19

设 $A \in \mathbb{C}^{m\times n}$, $b \in \mathbb{C}^m$, 矛盾方程组 $Ax = b$ 的全部最小二乘解为

$$z = A^+b + (I - A^+A)y \quad (y \in \mathbb{C}^n \text{ 任意.})$$

推论 6.20

设 $A \in \mathbb{C}^{m\times n}, b \in \mathbb{C}^m$, 则 $z \in \mathbb{C}^n$ 是矛盾方程组 $Ax = b$ 的最小二乘解的充要条件是: z 是相容方程组 $Ax = AA^+b$ 的解.

推论 6.21

设 $A \in \mathbb{C}^{m\times n}, b \in \mathbb{C}^m$, 则 $z \in \mathbb{C}^n$ 是矛盾方程组 $Ax = b$ 最小二乘解的充要条件是: z 是相容方程组 $A^H Ax = A^H b$ 的解.

注记 6.22

最小二乘解 $\{z = A^+ b + (I - A^+ A)y \quad (y \in \mathbb{C}^n \text{ (任意)})\}$ 知道, 矛盾方程组 $Ax = b$ 有唯一解的充分必要条件是 $A^+ A = I$ ($i.e.$ $\operatorname{rank} A = n$). 最小二乘解一般不唯一, 在所有的最小二乘解中 2 范数最小的解称为矛盾方程组 $Ax = b$ 的**极小范数最小二乘解**或者**最佳逼近解**.

定理 6.23

设 $A \in \mathbb{C}^{m\times n}, b \in \mathbb{C}^m$, 则矛盾方程组 $Ax = b$ 的唯一极小范数最小二乘解为 $x_0 = A^+ b$.

利用 Moore-Penrose 逆 A^+ 求解线性方程组 $Ax = b$ 的解, 有如下总结:

- $Ax = b$ 有解 (或者相容) $\Longleftrightarrow AA^+ b = b$;
- $x = A^+ b + (I - A^+ A)y \quad (y \in \mathbb{C}^n$ 任意) 是相容方程组 $Ax = b$ 的通解, 或者是矛盾方程组 $Ax = b$ 的全部最小二乘解;
- $x_0 = A^+ b$ 是相容方程组 $Ax = b$ 的唯一极小范数解, 或是矛盾方程组 $Ax = b$ 的唯一极小范数最小二乘解.

例 6.4

设 $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$, $b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

(1) 求 A 的满秩分解;

(2) 求 A^+;

(3) 判断 $Ax = b$ 是否有解;

(4) 有解时求 $Ax = b$ 的极小范数解, 无解时求 $Ax = b$ 的极小范数最小二乘解.

练习 6.1

已知矩阵 $A = \begin{bmatrix} 2 & -2 & 0 \\ -1 & 1 & 1 \\ -1 & 1 & 0 \end{bmatrix}$, $b = \begin{bmatrix} -4 \\ 2 \\ 2 \end{bmatrix}$.

(1) 求 A 的满秩分解.

(2) 求 A^{+}.

(3) 用广义逆矩阵方法判断线性方程组 $Ax = b$ 是否有解.

(4) 求线性方程组 $Ax = b$ 极小范数解或极小范数最小二乘解, 并说明是哪种解.

第 7 章 范数理论

§7.1 向量范数

定义 7.1

设 $x=(\xi_1,\xi_2,\cdots,\xi_n)^T\in\mathbb{C}^n$, 称 $\|x\|_2=\sqrt{(x,x)}=\sqrt{\sum\limits_{k=1}^{n}|\xi_k|^2}$, 为向量 x 的长度或 2 范数.

定理 7.2

设 $x,y\in\mathbb{C}^n$, $\lambda\in\mathbb{C}$, 则

(1) 非负性: $\|x\|_2\geqslant 0$; $\|x\|_2=0\Longleftrightarrow x=0$.

(2) 齐次性: $\|\lambda x\|_2=|\lambda|\|x\|_2$, $\lambda\in\mathbb{C}$.

(3) 三角形不等式: $\|x+y\|_2\leqslant\|x\|_2+\|y\|_2$.

向量范数

定义 7.3

若函数 $\|\cdot\|: \mathbb{C}^n \to \mathbb{R}$ 满足如下三个条件 (向量范数三公理):

(1) 非负性: $\|x\| \geqslant 0$; $\|x\| = 0 \iff x = 0$.

(2) 齐次性: $\|\lambda x\| = |\lambda|\|x\|$, $\lambda \in \mathbb{C}$.

(3) 三角形不等式: $\|x+y\| \leqslant \|x\| + \|y\|$.

则称 $\|x\|$ 为 $\mathbb{C}^n$ 上向量 x 的范数, 简称向量范数.

引理 7.4

对任意 $x, y \in \mathbb{C}^n$, 有

(1) $\|-x\| = \|x\|$;

(2) $\big|\|x\| - \|y\|\big| \leqslant \|x - y\|$. 两边之差小于第三边

2 范数

例 7.1

设 $x=(\xi_1,\xi_2,\cdots,\xi_n)^T\in\mathbb{C}^n$, 称

$$\|x\|_2=\sqrt{\sum_{k=1}^{n}|\xi_k|^2}=\sqrt{x^Hx},$$

为向量 2 范数.

注: 向量 2 范数具有酉不变性: 设 U 是 n 阶酉矩阵, 则

$$\|Ux\|_2=\|x\|_2.$$

1 范数和 ∞ 范数

例 7.2

设 $x = (\xi_1, \xi_2, \cdots, \xi_n)^T \in \mathbb{C}^n$, 称

$$\|x\|_1 = \sum_{k=1}^{n} |\xi_k|,$$

为向量 1 范数.

例 7.3

设 $x = (\xi_1, \xi_2, \cdots, \xi_n)^T \in \mathbb{C}^n$, 称

$$\|x\|_\infty = \max_k |\xi_k|,$$

为向量 ∞ 范数.

两个重要不等式

引理 7.5

Young 不等式: 对任意实数 $a, b \geqslant 0$, 有

$$ab \leqslant \frac{a^p}{p} + \frac{b^q}{q},$$

其中 $p > 1, q > 1$, 且 $\frac{1}{p} + \frac{1}{q} = 1$.

提示: $ab = 0$ 显然, 假设 $a, b > 0$. 取 $\varphi(t) = \frac{t^p}{p} + \frac{t^{-q}}{q}, \quad (0 < t < +\infty)$. 则 $\varphi'(t) = \frac{t^{p+q} - 1}{t^{q+1}}$, 得 $\varphi(t) \geqslant \varphi(1) = 1$, 取 $t = a^{\frac{1}{q}} b^{-\frac{1}{p}}$ 得

$$1 \leqslant \frac{1}{p}(a^{\frac{1}{q}} b^{-\frac{1}{p}})^p + \frac{1}{q}(a^{\frac{1}{q}} b^{-\frac{1}{p}})^{-q} = \frac{1}{ab}(\frac{a^p}{p} + \frac{b^q}{q}).$$

定理 7.6

Hölder 不等式: 对任意 $\xi_k, \eta_k \in \mathbb{C}(k = 1, 2, \cdots, n)$, 有

$$\sum_{k=1}^{n} |\xi_k||\eta_k| \leqslant (\sum_{k=1}^{n} |\xi_k|^p)^{\frac{1}{p}} (\sum_{k=1}^{n} |\eta_k|^q)^{\frac{1}{q}},$$

其中 $p > 1, q > 1$, 且 $\dfrac{1}{p} + \dfrac{1}{q} = 1$.

提示: $\xi_k = 0$ 或 $\eta_k = 0(k = 1, 2, \cdots, n)$ 成立, 假设 ξ_k, η_k 都不全为零. 由 Young 不等式得

$$\begin{aligned}\frac{\sum |\xi_k||\eta_k|}{(\sum |\xi_k|^p)^{\frac{1}{p}} (\sum |\eta_k|^q)^{\frac{1}{q}}} &= \sum \left[\frac{|\xi_k|}{(\sum |\xi_k|^p)^{\frac{1}{p}}}\right] \left[\frac{|\eta_k|}{(\sum |\eta_k|^q)^{\frac{1}{q}}}\right] \\ &\leqslant \sum \left[\frac{|\xi_k|^p}{p(\sum |\xi_k|^p)} + \frac{|\eta_k|^q}{q(\sum |\eta_k|^q)}\right] = \frac{1}{p} + \frac{1}{q} = 1.\end{aligned}$$

向量 p 范数

例 7.4

设 $x = (\xi_1, \xi_2, \cdots, \xi_n)^T \in \mathbb{C}^n$, $1 \leqslant p < +\infty$, 称

$$\|x\|_p = \left(\sum_{k=1}^{n} |\xi_k|^p\right)^{\frac{1}{p}},$$

为向量 p **范数**.

提示: 只证三角性. $p=1$ 显然成立. 设 $p>1$,

$$
\begin{aligned}
\|x+y\|_p^p &= \sum|\xi_k+\eta_k|^p = \sum|\xi_k+\eta_k||\xi_k+\eta_k|^{p-1} \\
&\leqslant \sum|\xi_k||\xi_k+\eta_k|^{p-1} + \sum|\eta_k||\xi_k+\eta_k|^{p-1} \\
&\leqslant \left(\sum|\xi_k|^p\right)^{\frac{1}{p}}\left(\sum|\xi_k+\eta_k|^{(p-1)q}\right)^{\frac{1}{q}} \\
&\quad + \left(\sum|\eta_k|^p\right)^{\frac{1}{p}}\left(\sum|\xi_k+\eta_k|^{(p-1)q}\right)^{\frac{1}{q}} \quad ((p-1)q=p) \\
&= (\|x\|_p+\|y\|_p)\times\|x+y\|_p^{\frac{p}{q}} \quad \left(p-\frac{p}{q}=1\right).
\end{aligned}
$$

定理 7.7

设 $x=(\xi_1,\xi_2,\cdots,\xi_n)^T\in\mathbb{C}^n$, 则

$$\lim_{p\to+\infty}\|x\|_p=\|x\|_\infty.$$

构造新范数的方法

定理 7.8

设 A 是列满秩的 $m \times n$ 复矩阵, $\|\cdot\|_a$ 是 $\mathbb{C}^m$ 上的一种向量范数. 对任意 $x \in \mathbb{C}^n$, 规定

$$\|x\|_b = \|Ax\|_a,$$

则 $\|x\|_b$ 是 $\mathbb{C}^n$ 中的向量范数.

例 7.5

取

$$A = \mathrm{diag}\,(1, 2, \cdots, n),$$

对于任意 $x = (\xi_1, \xi_2, \cdots, \xi_n)^T \in \mathbb{C}^n$, 由 $\mathbb{C}^n$ 上向量的 1 范数和 2 范数可得

$$\|x\|_a = \|Ax\|_1 = \sum_{k=1}^{n} k|\xi_k|,$$

$$\|x\|_b = \|Ax\|_2 = \sqrt{\sum_{k=1}^{n} k^2|\xi_k|^2}.$$

例 7.6 (椭圆范数)

设 A 是 n 阶 Hermite 正定矩阵, 则

$$\|x\|_A = \sqrt{x^H A x},$$

是向量范数.

虽然可以定义出不同的范数, 且同一向量按不同范数算出的值一般不等, 例如向量 $e = (1, 1, \cdots, 1)^T \in \mathbb{C}^n$, 有

$$\|e\|_1 = n, \qquad \|e\|_2 = \sqrt{n}, \qquad \|e\|_\infty = 1$$

下面考察不同范数间的关系.

各种范数之间的关系

易知有如下不等式

$$\|x\|_2 \leqslant \|x\|_1 \leqslant \sqrt{n}\|x\|_2$$

$$\|x\|_\infty \leqslant \|x\|_1 \leqslant n\|x\|_\infty$$

$$\|x\|_\infty \leqslant \|x\|_2 \leqslant \sqrt{n}\|x\|_\infty$$

定义 7.9

设 $\|\cdot\|_a$ 和 $\|\cdot\|_b$ 都是 $\mathbb{C}^n$ 上的向量范数. 若存在正数 c_1 和 c_2, 使得对任意 $x \in \mathbb{C}^n$ 都有

$$c_1\|x\|_b \leqslant \|x\|_a \leqslant c_2\|x\|_b,$$

则称向量范数 $\|\cdot\|_a$ 和 $\|\cdot\|_b$ 等价.

向量范数的等价

向量范数的等价有下述三个性质:

(1) 自反性: $\|\cdot\|_a$ 和 $\|\cdot\|_a$ 等价.

(2) 对称性: $\|\cdot\|_a$ 和 $\|\cdot\|_b$ 等价, 则 $\|\cdot\|_b$ 和 $\|\cdot\|_a$ 等价.

(3) 传递性: $\|\cdot\|_a$ 和 $\|\cdot\|_b$ 等价, $\|\cdot\|_b$ 和 $\|\cdot\|_c$ 等价, 则 $\|\cdot\|_a$ 和 $\|\cdot\|_c$ 等价.

定理 7.10

$\mathbb{C}^n$ 上的所有向量范数都等价. (证明略)

向量序列的收敛性

定义 7.11

给定 $\mathbb{C}^n$ 中的向量序列 $\{x^{(k)}\}$, 其中

$$x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \cdots, x_n^{(k)})^T, \quad (k = 0, 1, 2, \cdots).$$

如果

$$\lim_{k\to+\infty} x_j^{(k)} = a_j, \quad (j = 1, 2, \cdots, n),$$

则称向量序列 $\{x^{(k)}\}$ 收敛于 $\boldsymbol{a} = (a_1, a_2, \cdots, a_n)^T$. 简称 $\{x^{(k)}\}$ 收敛, 记为

$$\lim_{k\to+\infty} x^{(k)} = \boldsymbol{a} \quad 或 \quad x^{(k)} \to \boldsymbol{a} \ (k \to +\infty).$$

不收敛的向量序列称为是发散的.

向量序列收敛性的判别法

例 7.7

向量序列 $x^{(k)} = (-2+\frac{1}{k}, (1+\frac{1}{k})^k, 2)^T \stackrel{k\to\infty}{\longrightarrow} (-2, e, 2)$. 但是序列 $x^{(k)} = (1-\frac{1}{k}, \sin k)^T$ 是发散的.

定理 7.12

$\mathbb{C}^n$ 中的向量序列 $\{x^{(k)}\}$ 收敛于 $\boldsymbol{a}$ 的充要条件是对于 $\mathbb{C}^n$ 上的任意一种向量范数 $\|\cdot\|$, 都有

$$\lim_{k\to+\infty} \|x^{(k)} - \boldsymbol{a}\| = 0.$$

先试试 $\|\cdot\|_\infty$.

§7.2 矩阵范数

定义 7.13

若函数 $\|\cdot\|:\mathbb{C}^{n\times n}\to\mathbb{R}$ 满足如下四个条件:

(1) 非负性: $\|A\|\geqslant 0$; $\|A\|=0\iff A=0$.

(2) 齐次性: $\|\lambda A\|=|\lambda|\|A\|$, $\lambda\in\mathbb{C}$.

(3) 三角形不等式: $\|A+B\|\leqslant\|A\|+\|B\|$.

(4) 相容性: $\|AB\|\leqslant\|A\|\|B\|$.

则称 $\|A\|$ 为 $\mathbb{C}^{n\times n}$ 上矩阵 A 的范数, 简称矩阵范数.

与向量范数所具有的性质类似: $\mathbb{C}^{n\times n}$ 上的任意两个矩阵范数都等价以及
$\|-A\|=\|A\|$, $\big|\|A\|-\|B\|\big|\leqslant\|A-B\|$

m_1 范数与 m_∞ 范数

例 7.8

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

$$\|A\|_{m_1}=\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|$$

是 $\mathbb{C}^{n\times n}$ 上的一种矩阵范数, 称之为矩阵的 m_1 范数.

注记 7.14

对于 $A=(a_{ij})_{n\times n}\in\mathbb{C}^{n\times n}$, 如果将向量的 ∞ 范数直接推广到矩阵上, 则前三条公里成立, 但相容性公理一般不成立.

例如取 $A=B=\begin{bmatrix}1&1\\1&1\end{bmatrix}$, 则 $\|A\|=\|B\|=1$, 而 $AB=\begin{bmatrix}2&2\\2&2\end{bmatrix}$, 显然 $\|AB\|=2$, 不满足 $\|AB\|\leqslant\|A\|\|B\|$.

例 7.9

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

$$\|A\|_{m_\infty}=n\max_{i,j}|a_{ij}|$$

是 $\mathbb{C}^{n\times n}$ 上的一种矩阵范数, 称之为矩阵的 m_∞ 范数.

F 范数

例 7.10

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

$$\|A\|_F=\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2}=\sqrt{\mathrm{tr}(A^HA)}$$

是 $\mathbb{C}^{n\times n}$ 上的一种矩阵范数, 称之为矩阵的Frobenius 范数, 简称为 F 范数.

定理 7.15 (酉不变性)

设 $A\in\mathbb{C}^{n\times n}$, 则对任意 n 阶酉矩阵 U 和 V, 恒有

$$\|AU\|_F=\|VA\|_F=\|VAU\|_F=\|A\|_F.$$

方阵范数与向量范数的相容性

定义 7.16

设 $\|\cdot\|_m$ 是 $\mathbb{C}^{n\times n}$ 上的矩阵范数, $\|\cdot\|_v$ 是 $\mathbb{C}^n$ 上的向量范数. 如果对任意 $A\in\mathbb{C}^{n\times n}$ 和 $x\in\mathbb{C}^n$ 都有

$$\|Ax\|_v \leqslant \|A\|_m\|x\|_v,$$

则称矩阵范数 $\|\cdot\|_m$ 与向量范数 $\|\cdot\|_v$ 是相容的.

例 7.11

$\mathbb{C}^{n\times n}$ 上方阵的 m_∞ 范数分别与 $\mathbb{C}^n$ 上向量的 1，2，∞ 范数相容.

与一个矩阵范数相容的向量范数未必唯一, 那么对 $\mathbb{C}^{n\times n}$ 上任意给定的矩阵范数, 是否存在与之相容的向量范数呢?

定理 7.17

设 $\|\cdot\|_m$ 是 $\mathbb{C}^{n\times n}$ 上的一种矩阵范数, 则在 $\mathbb{C}^n$ 上必存在与它相容的向量范数.

I_n 相当于矩阵乘法中的单位元, 而

$$\|I_n\|_{m_1} = n, \quad \|I_n\|_F = \sqrt{n}, \quad \|I_n\|_{m_\infty} = n,$$

对于一般的矩阵范数, 总有 $\|I_n\| \geqslant 1$.

$$\|x\|_v = \|I_n x\|_v \leqslant \|I_n\| \|x\|_v \;\Rightarrow\; \|I_n\| \geqslant 1 (\text{其中} \|\cdot\|_v \text{与} \|\cdot\| \text{相容}).$$

能否构造出使得 $\|I_n\| = 1$ 的范数矩阵呢?

定理 7.18

已知 $\mathbb{C}^n$ 上的向量范数 $\|\cdot\|_v$, 对任意 n 阶方阵 A, 规定

$$\|A\| = \max_{x\neq 0}\frac{\|Ax\|_v}{\|x\|_v} \quad \left(\|A\| = \max_{\|x\|_v=1}\|Ax\|_v\right),$$

则 $\|\cdot\|$ 是 $\mathbb{C}^{n\times n}$ 上与向量范数 $\|\cdot\|_v$ 相容的矩阵范数, 且 $\|I_n\| = 1$. 称之为由向量范数 $\|\cdot\|_v$ 导出的矩阵范数或从属于向量范数 $\|\cdot\|_v$ 的矩阵范数, 简称为导出范数或从属范数.

矩阵的从属范数的计算归结为求函数的最大值, 虽然连续函数在有界闭集上可以达到最大值, 但不容易计算, 下面给出部分导出范数的性质.

定理 7.19

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

(1) $\|A\|_1=\max\limits_{j}\sum\limits_{i=1}^{n}|a_{ij}|$;

(2) $\|A\|_\infty=\max\limits_{i}\sum\limits_{j=1}^{n}|a_{ij}|$;

(3) $\|A\|_2=\sqrt{\lambda_1}$, λ_1 为 A^HA 的最大特征值.

通常 $\|A\|_1$, $\|A\|_2$ 和 $\|A\|_\infty$ 依次称为矩阵的1 范数、2 范数和 ∞ 范数, 或称为列和范数、谱范数和行和范数.

定理 7.20

设 $A \in \mathbb{C}^{n\times n}$, U 和 V 为 n 阶酉矩阵, 则

(1) $\|A^H\|_2 = \|A\|_2$;

(2) $\|AU\|_2 = \|VA\|_2 = \|VAU\|_2 = \|A\|_2$;

(3) 若 A 是正规矩阵, 且 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 是 A 的 n 个特征值, 则 $\|A\|_2 = \max\limits_k |\lambda_k|$.

例 7.12

已知矩阵

$$A = \begin{bmatrix} -1 & 0 & 2 & 1+\mathrm{i} \\ 3 & 5 & -1 & 0 \\ 1 & 2 & 0 & -1 \\ -\mathrm{i} & 1 & 2 & -4 \end{bmatrix},$$

求 $m_1-, F-, m_\infty-, 1-, \infty-$ 范数.

长方阵的范数

相容性修改为对于任意 $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{n\times l}$, 都有

$$\|AB\| \leqslant \|A\|\|B\|.$$

它们应该取同类型的范数, 比如说 F 范数.

其次, 与向量范数相容性, 对任意 $A \in \mathbb{C}^{m\times n}$, $x \in \mathbb{C}^n$, 都有

$$\|Ax\|_v \leqslant \|A\|_m \|x\|_v,$$

两边分别取自 $\mathbb{C}^m, \mathbb{C}^n$ 上的同类的向量范数.

最后从属范数中, 对任意 $A \in \mathbb{C}^{m\times n}$, 有

$$\|A\| = \max_{x\neq 0} \frac{\|Ax\|_v}{\|x\|_v},$$

分子分母取同类的向量范数.

常用的长方阵范数

设 $A=(a_{ij})_{m\times n}\in\mathbb{C}^{m\times n}$, 则有如下常用范数:

- m_1 范数: $\|A\|_{m_1}=\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}|a_{ij}|$
- F 范数: $\|A\|_F=\sqrt{\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}|a_{ij}|^2}=\sqrt{\mathrm{tr}(A^HA)}$
- M 范数 (最大范数): $\|A\|_M=\max\{m,n\}\max\limits_{i,j}|a_{ij}|$
- G 范数 (几何平均范数): $\|A\|_G=\sqrt{mn}\max\limits_{i,j}|a_{ij}|$
- 1 范数 (列和范数): $\|A\|_1=\max\limits_{j}\sum\limits_{i=1}^{m}|a_{ij}|$
- 2 范数 (谱范数): $\|A\|_2=\sqrt{A^HA\text{的最大特征值}}$
- ∞ 范数 (行和范数): $\|A\|_\infty=\max\limits_{i}\sum\limits_{j=1}^{n}|a_{ij}|$

§7.3 范数应用举例

定义 7.21

设 A 是 n 阶复方阵, $\lambda_1, \lambda_2, \cdots, \lambda_n$ 为 A 的 n 个特征值, 称

$$\rho(A) = \max_j |\lambda_j|$$

为矩阵 A 的**谱半径**.

定理 7.22

设 A 是 n 阶复方阵, 则

(1) $\rho(A^k) = (\rho(A))^k, \quad k \in \mathbb{N}$.

(2) $\rho(A^H A) = \rho(AA^H) = \|A\|_2^2$.

(3) 若 A 是正规矩阵, 则 $\rho(A) = \|A\|_2$.

谱半径的估计

定理 7.23

设 A 是 n 阶复方阵, 则对 $\mathbb{C}^{n\times n}$ 上的任一矩阵范数 $\|\cdot\|$ 都有

$$\rho(A) \leqslant \|A\|.$$

例 7.13

试估算矩阵 $A = \begin{bmatrix} 0 & 0.2 & 0.1 \\ -0.2 & 0 & 0.2 \\ -0.1 & -0.2 & 0 \end{bmatrix}$ 的特征值的范围.

实际上 A 的特征值为 $0, -0.3\,\mathrm{i}, 0.3\,\mathrm{i}$.

定理 7.24

设 A 是 n 阶复方阵, 对任意给定的正数 ε, 存在某一矩阵范数 $\|\cdot\|_m$, 使得

$$\|A\|_m \leqslant \rho(A) + \varepsilon.$$

存在可逆矩阵 P, 使 $P^{-1}AP = J = \begin{bmatrix} \lambda_1 & * & & \\ & \lambda_2 & \ddots & \\ & & \ddots & * \\ & & & \lambda_n \end{bmatrix}$, $* = 0$ 或 1.

对于 $\varepsilon > 0$, 取 $D = \mathrm{diag}(1, \varepsilon, \cdots, \varepsilon^{n-1})$, 易知

$$\|D^{-1}P^{-1}APD\|_\infty \leqslant \rho(A) + \varepsilon.$$

范数在数值分析中的应用

1. 扰动的影响

在实际工程中, 问题中的原始数据或参数有微小的扰动或误差, 对问题的解会产生什么样的影响, 下面以矩阵求逆和线性方程组求解举例说明.

问题 7.25 (矩阵的逆)

当矩阵 A 可逆时, 对 A 附加一个扰动 δA 后, 矩阵 $A+\delta A$ 的情况如何? 是否可逆, 若可逆, 则 $(A+\delta A)^{-1}$ 与 A^{-1} 相差多少?

例 7.14

$$A = \begin{bmatrix} 5 & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix},\ A^{-1} = \begin{bmatrix} 68 & -41 & -17 & 10 \\ -41 & 25 & 10 & -6 \\ -17 & 10 & 5 & -3 \\ 10 & -6 & -3 & 2 \end{bmatrix},\ |A| = 1$$

例 7.15

设 $\delta A = \begin{bmatrix} \varepsilon & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$，则 $A + \delta A = \begin{bmatrix} 5+\varepsilon & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix}$

可以求出 $|A + \delta A| = 1 + 68\varepsilon$. 当 $\varepsilon = -\frac{1}{68} \approx -0.015$ 时, $|A + \delta A| = 0$.

例 7.16

取 $\varepsilon = -0.01$, 则有 $A + \delta A = \begin{bmatrix} 4.99 & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix}$.

此时有 $(A + \delta A)^{-1} = \begin{bmatrix} 204.82 & -128.12 & -53.12 & 31.25 \\ -128.12 & 77.53 & 31.78 & -18.81 \\ -53.12 & 31.78 & 14.03 & -8.31 \\ 31.25 & -18.81 & 8.31 & 5.12 \end{bmatrix}$

注记 7.26

微小变化可以使矩阵发生质的变化, 或逆矩阵的差别很大, 说明这类矩阵的可逆性和逆矩阵对于微小扰动很敏感, 称该矩阵关于求逆是病态.

问题 7.27 (线性方程组求解问题)

线性方程组 $Ax = b$. 如果 A 可逆, 则解唯一. 但若系数矩阵 A 有微小扰动 δA, 或 b 有扰动 δb, 或者两者都有扰动, 相应的解 x 也会有扰动 δx. 即为

$$(A + \delta A)(x + \delta x) = b$$

$$A(x + \delta x) = b + \delta b$$

$$(A + \delta A)(x + \delta x) = b + \delta b$$

其解的误差程度如何, 若有误差如何估算?

例 7.17

考察 4 阶实对称线性方程组

$$\begin{bmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 32 \\ 23 \\ 33 \\ 31 \end{bmatrix}, \quad \text{其解为} \rightsquigarrow \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

$$\text{考虑右边微小扰动 } \delta b = \begin{bmatrix} 0.1 \\ -0.1 \\ 0.1 \\ -0.1 \end{bmatrix}, \text{其解变为} \rightsquigarrow \begin{bmatrix} 9.2 \\ -12.6 \\ 4.5 \\ -1.1 \end{bmatrix}$$

扰动定理

2. 近似逆矩阵的误差估计 (扰动定理)

引理 7.28

设 A 是 n 阶复方阵. 若对 $\mathbb{C}^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$ 有 $\|A\|<1$, 则 I_n-A 可逆, 特别地, $\rho(A)<1 \rightsquigarrow I_n-A$ 可逆.

定理 7.29

设 A 是 n 阶复方阵且 A 可逆. 若对 $\mathbb{C}^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$ 有 $\|A^{-1}\delta A\|<1$, 则

(1) $A+\delta A$ 可逆;

(2) $\|(A+\delta A)^{-1}\| \leqslant \dfrac{\|A^{-1}\|}{1-\|A^{-1}\delta A\|}$;

(3) $\dfrac{\|A^{-1}-(A+\delta A)^{-1}\|}{\|A^{-1}\|} \leqslant \dfrac{\|A^{-1}\delta A\|}{1-\|A^{-1}\delta A\|}$.

推论 7.30

设 A 是 n 阶复方阵且 A 可逆. 若对 $\mathbb{C}^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$ 有 $\|A^{-1}\|\|\delta A\| < 1$, 则

$$\frac{\|A^{-1}-(A+\delta A)^{-1}\|}{\|A^{-1}\|} \leqslant \frac{\|A^{-1}\|\|\delta A\|}{1-\|A^{-1}\|\|\delta A\|} = \frac{\|A\|\|A^{-1}\|\frac{\|\delta A\|}{\|A\|}}{1-\|A\|\|A^{-1}\|\frac{\|\delta A\|}{\|A\|}}.$$

3. 线性方程组解的误差估计

定理 7.31

设 $A, \delta A$ 是 n 阶复方阵且 A 可逆, $b, \delta b \in \mathbb{C}^n$. 若对 $\mathbb{C}^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$ 有 $\|A^{-1}\|\|\delta A\| < 1$, 则非齐次线性方程组 $Ax = b$ 与 $(A+\delta A)(x+\delta x) = b+\delta b$ 的解满足

$$\frac{\|\delta x\|_v}{\|x\|_v} \leqslant \frac{\|A\|\|A^{-1}\|}{1 - \|A\|\|A^{-1}\|\frac{\|\delta A\|}{\|A\|}}\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|_v}{\|b\|_v}\right),$$

其中 $\|\cdot\|_v$ 是 $\mathbb{C}^n$ 上与矩阵范数 $\|\cdot\|$ 相容的向量范数.

矩阵的条件数

定义 7.32

设 A 是 n 阶可逆复方阵, $\|\cdot\|$ 是 $\mathbb{C}^{n\times n}$ 上的矩阵范数, 称

$$\mathrm{cond}(A) = \|A\|\|A^{-1}\|$$

为矩阵 A (关于求逆或求解线性方程组) 的条件数.

一般地, 若矩阵 A 的条件数大就称 A 对于求逆矩阵或求解线性方程组是病态的, 或坏条件的; 否则, 则称为 良态的或好条件的.

常用的条件数

设 A 可逆, 由于条件数与所取范数有关, 有

$$\mathrm{cond}_\infty(A) = \|A\|_\infty \|A^{-1}\|_\infty$$

$$\mathrm{cond}_2(A) = \|A\|_2 \|A^{-1}\|_2 = \sqrt{\frac{\mu_1}{\mu_2}}$$

其中 μ_1, μ_2 分别是 $A^H A$ 的最大和最小特征值.

当 A 是正规矩阵时, 有

$$\mathrm{cond}_2(A) = \|A\|_2 \|A^{-1}\|_2 = \left|\frac{\lambda_1}{\lambda_n}\right|$$

其中 λ_1, λ_n 分别是 A 的按模最大和最小的特征值.

第 8 章　矩阵分析

§8.1　矩阵序列

定义 8.1

设有 $\mathbb{C}^{m\times n}$ 中的矩阵序列 $\{A^{(k)}\}$, 其中 $A^{(k)}=(a_{ij}^{(k)})$. 若

$$\lim_{k\to+\infty} a_{ij}^{(k)}=a_{ij},\quad (i=1,2,\cdots,m;\ j=1,2,\cdots,n),$$

则称矩阵序列 $\{A^{(k)}\}$ 收敛于 $A=(a_{ij})_{m\times n}$, 或称 A 为矩阵序列 $\{A^{(k)}\}$ 的极限, 记为

$$\lim_{k\to+\infty} A^{(k)}=A \quad 或 \quad A^{(k)}\to A\ (k\to+\infty).$$

不收敛的矩阵序列称为是发散的.

例 8.1

(1) $A^{(k)} = \begin{bmatrix} 1+\frac{1}{k} & 2 \\ \frac{\sin k}{k} & \frac{2k+3}{4k-5} \end{bmatrix}$; (2) $B^{(k)} = \begin{bmatrix} (-1)^k & \frac{1}{k+1} \\ \frac{2k+3}{4k-5} & 2 \end{bmatrix}$

注记 8.2

矩阵序列收敛相当于 mn 个数列极限的收敛. 用初等的方法研究 mn 个数列的极限很繁琐, 与向量序列一样, 可以用矩阵范数来研究矩阵序列的极限.

矩阵序列收敛性的判别法

回忆矩阵 $A \in \mathbb{C}^{m\times n}$ 的 G-范数 (几何平均范数):

$$\|A\|_G = \sqrt{mn} \max_{i,j} |a_{ij}|.$$

定理 8.3

$\mathbb{C}^{m\times n}$ 中的矩阵序列 $\{A^{(k)}\}$ 收敛于 A 当且仅当对于 $\mathbb{C}^{m\times n}$ 上的任意一种矩阵范数 $\|\cdot\|$, 都有

$$\lim_{k\to+\infty} \|A^{(k)} - A\| = 0.$$

推论 8.4

若 $\mathbb{C}^{m\times n}$ 中的矩阵序列 $\{A^{(k)}\}$ 收敛于 A, 则对于 $\mathbb{C}^{m\times n}$ 上的任意一种矩阵范数 $\|\cdot\|$, 都有

$$\lim_{k\to+\infty}\|A^{(k)}\| = \|A\| \ \rightsquigarrow\ \|A^{(k)}\| \text{ 有界}.$$

注: 此推论的逆命题不成立. 例如矩阵序列

$$A^{(k)} = \begin{bmatrix} (-1)^k & \frac{1}{k+1} \\ 1 & 2 \end{bmatrix}$$

不收敛，但是

$$\lim_{k\to+\infty}\|A^{(k)}\|_F = \lim_{k\to+\infty}\sqrt{6+\frac{1}{(k+1)^2}} = \sqrt{6}.$$

收敛的矩阵序列的性质

定理 8.5

设 $\lim\limits_{k\to+\infty} A^{(k)} = A$, $\lim\limits_{k\to+\infty} B^{(k)} = B$, 其中 $A^{(k)}, B^{(k)}, A,B$ 为适当阶的矩阵, $a, b \in \mathbb{C}$, 则

(1) $\lim\limits_{k\to+\infty} (aA^{(k)} + bB^{(k)}) = aA + bB$;

(2) $\lim\limits_{k\to+\infty} A^{(k)}B^{(k)} = AB$;

(3) 当 $A^{(k)}$ 与 A 均可逆时, $\lim\limits_{k\to+\infty} (A^{(k)})^{-1} = A^{-1}$.

上述定理中的当 $A^{(k)}$ 与 A 均可逆是不可少的, 因为即使所有的 $A^{(k)}$ 可逆也不能保证 A 一定可逆. 例如

$$A^{(k)} = \begin{bmatrix} 1+\frac{1}{k} & 1 \\ 1 & 1 \end{bmatrix}$$

对每个 $A^{(k)}$ 都有逆矩阵 $(A^{(k)})^{-1} = \begin{bmatrix} k & -k \\ -k & k+1 \end{bmatrix}$, 但是

$$\lim_{k\to+\infty} A^{(k)} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

不可逆.

收敛矩阵

定义 8.6

设 A 是 n 阶复方阵. 若 $\lim_{k\to+\infty} A^k = 0$, 则称 A 为收敛矩阵.

定理 8.7

设 A 是 n 阶复方阵, 则 A 是收敛矩阵的充分必要条件是 $\rho(A) < 1$.

推论 8.8

设 A 是 n 阶复方阵. 若对 $\mathbb{C}^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$ 有 $\|A\| < 1$, 则 A 为收敛矩阵.

例 8.2

判断下列矩阵是否为收敛矩阵:

(1) $A = \begin{bmatrix} \frac{1}{6} & -\frac{4}{3} \\ -\frac{1}{3} & \frac{1}{6} \end{bmatrix} \rightsquigarrow$ 特征值为$\lambda_1 = 5/6,\ \lambda_2 = -1/2.$

(2) $A = \begin{bmatrix} 0.2 & 0.1 & 0.2 \\ 0.5 & 0.5 & 0.4 \\ 0.1 & 0.2 & 0.3 \end{bmatrix} \rightsquigarrow \|A\|_1 = 0.9 < 1.$

§8.2　矩 阵 级 数

定义 8.9

由 $\mathbb{C}^{m\times n}$ 中的矩阵序列 $\{A^{(k)}\}$ 构成的无穷和

$$A^{(0)}+A^{(1)}+A^{(2)}+\cdots+A^{(k)}+\cdots$$

称为矩阵级数, 记为 $\sum\limits_{k=0}^{\infty} A^{(k)}$.

$S^{(N)}=\sum\limits_{k=0}^{N} A^{(k)}$ 称为矩阵级数的部分和. 若部分和序列 $\{S^{(N)}\}$ 收敛且极限为 S, 则称矩阵级数 $\sum\limits_{k=0}^{\infty} A^{(k)}$ 收敛, 并称 S 为 矩阵级数的和, 记为 $S=\sum\limits_{k=0}^{\infty} A^{(k)}$. 不收敛的矩阵级数称为发散的.

注记 8.10

如果 $A^{(k)}=(a_{ij}^{(k)})_{m\times n}$, $S=(s_{ij})_{m\times n}$, 显然 $S=\sum\limits_{k=0}^{+\infty}A^{(k)}$ 相当于

$$\sum_{k=0}^{+\infty}a_{ij}^{(k)}=s_{ij}\ \ (i=1,2,\cdots,m;\ j=1,2,\cdots,n),$$

即 mn 个数项级数都收敛.

例 8.3

已知

$$A^{(k)} = \begin{bmatrix} \frac{1}{2^k} & \frac{\pi}{4^k} \\ 0 & \frac{1}{(k+1)(k+2)} \end{bmatrix} \quad (k = 0, 1, 2, \cdots)$$

研究矩阵级数 $\sum\limits_{k=0}^{+\infty} A^{(k)}$ 的敛散性.

$$S^N = \sum_{k=0}^{N} A^{(k)} = \begin{bmatrix} 2 - \frac{1}{2^N} & \frac{\pi}{3}\left(4 - \frac{1}{4^N}\right) \\ 0 & 1 - \frac{1}{N+2} \end{bmatrix} \xrightarrow{N \to +\infty} \begin{bmatrix} 2 & \frac{4\pi}{3} \\ 0 & 1 \end{bmatrix}.$$

矩阵级数的绝对收敛性

定义 8.11

设 $A^{(k)} = (a_{ij}^{(k)})_{m\times n}$ $(k = 0, 1, 2, \cdots)$. 如果 mn 个数项级数

$$\sum_{k=0}^{\infty} a_{ij}^{(k)} \quad (i = 1, 2, \cdots, m;\ j = 1, 2, \cdots n)$$

都绝对收敛, 即 $\sum\limits_{k=0}^{+\infty} |a_{ij}^{(k)}|$ 都收敛, 则称矩阵级数 $\sum\limits_{k=0}^{\infty} A^{(k)}$ 绝对收敛.

定理 8.12

设 $A^{(k)} = (a_{ij}^{(k)})_{m\times n}$ $(k = 0, 1, 2, \cdots)$, 则矩阵级数 $\sum\limits_{k=0}^{\infty} A^{(k)}$ 绝对收敛当且仅当正项级数 $\sum\limits_{k=0}^{\infty} \|A^{(k)}\|$ 收敛, 其中 $\|\cdot\|$ 是 $\mathbb{C}^{m\times n}$ 上任一矩阵范数.

定理 8.13 (比较原理)

设 $\sum u_n$ 和 $\sum v_n$ 是两个正项级数, 若存在某个正数 N, 对一切 $n > N$ 都有 $u_n \leqslant v_n$, 则

(1): 若 $\sum v_n$ 收敛, 则 $\sum u_n$ 收敛.

(2): 若 $\sum u_n$ 发散, 则 $\sum v_n$ 发散.

(3): 正项级数 $\sum u_n$ 收敛 $\iff$ 部分和数列 $\{S_n\}$ 有界.

定理 8.14 (矩阵级数收敛的性质)

设 $\sum\limits_{k=0}^{\infty} A^{(k)} = A$, $\sum\limits_{k=0}^{\infty} B^{(k)} = B$, 其中 $A^{(k)}, B^{(k)}, A, B$ 是适当阶的矩阵, 则

(1) 对 $\lambda, \mu \in \mathbb{C}$, 有 $\sum\limits_{k=0}^{\infty}(\lambda A^{(k)} + \mu B^{(k)}) = \lambda A + \mu B$.

(2) 绝对收敛的矩阵级数必收敛, 并且任意调换其项的顺序所得的矩阵级数仍收敛, 且其和不变.

(3) 若矩阵级数 $\sum\limits_{k=0}^{\infty} A^{(k)}$ 收敛 (或绝对收敛), 则矩阵级数 $\sum\limits_{k=0}^{\infty} PA^{(k)}Q$ 也收敛 (或绝对收敛), 并且有

$$\sum_{k=0}^{\infty} PA^{(k)}Q = P\Big(\sum_{k=0}^{\infty} A^{(k)}\Big)Q.$$

矩阵级数收敛的性质

(4) 若 $\sum_{k=0}^{\infty} A^{(k)}$ 和 $\sum_{k=0}^{\infty} B^{(k)}$ 均绝对收敛, 则它们按项相乘所得的矩阵级数

$$A^{(0)}B^{(0)} + (A^{(0)}B^{(1)} + A^{(1)}B^{(0)}) + \cdots +$$

$$(A^{(0)}B^{(k)} + A^{(1)}B^{(k-1)} + \cdots + A^{(k)}B^{(0)}) + \cdots$$

也绝对收敛, 且其和为 AB.

矩阵幂级数

定义 8.15

设 A 是 n 阶复方阵, $c_k \in \mathbb{C}$ $(k = 0, 1, 2, \cdots)$, 称矩阵级数 $\sum\limits_{k=0}^{\infty} c_k A^k$ 为矩阵 A 的 幂级数.

复幂级数: 幂级数 $\sum\limits_{k=0}^{+\infty} a_k z^k (a_k \in \mathbb{C})$ 的收敛范围是包含原点的一个圆域 $|z| < r$, 且在此圆域内绝对收敛, 其中 r 是收敛半径, 它由 $\lim\limits_{n\to+\infty} |\frac{a_{n+1}}{a_n}| = \rho$ 所确定:
若 $\rho \neq 0$, 则 $r = 1/\rho$. 若 $\rho = 0$, 则 $r = +\infty$, 若 $\rho = +\infty$, 则 $r = 0$.
区间端点需要另外判别其敛散性.

矩阵幂级数的收敛性

定理 8.16

设复幂级数 $\sum_{k=0}^{\infty} c_k z^k$ 的收敛半径为 r, A 是 n 阶复方阵, 则

(1) 当 $\rho(A) < r$ 时, 矩阵幂级数 $\sum_{k=0}^{\infty} c_k A^k$ 绝对收敛;

(2) 当 $\rho(A) > r$ 时, 矩阵幂级数 $\sum_{k=0}^{\infty} c_k A^k$ 发散.

推论 8.17

设复幂级数 $\sum_{k=0}^{\infty} c_k z^k$ 的收敛半径为 r, A 是 n 阶复方阵. 若存在 $\mathbb{C}^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$ 使得 $\|A\| < r$, 则矩阵幂级数 $\sum_{k=0}^{\infty} c_k A^k$ 绝对收敛.

例 8.4

判断矩阵幂级数 $\sum_{k=0}^{\infty} \frac{k}{6^k} \begin{bmatrix} 1 & -8 \\ -2 & 1 \end{bmatrix}^k$ 的敛散性.

Neumann 级数

引理 8.18 (Neumann 引理)

设 A 是 n 阶复方阵, Neumann 级数 $\sum\limits_{k=0}^{\infty} A^k$ 收敛当且仅当 $\rho(A) < 1$, 并且收敛时, 有

$$\sum_{k=0}^{\infty} A^k = (I_n - A)^{-1}.$$

推论 8.19

设 A 是 n 阶复方阵, 若存在 $\mathbb{C}^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$ 使得 $\|A\| < 1$, Neumann 级数 $\sum\limits_{k=0}^{\infty} A^k$ 收敛.

例 8.5

已知 $A = \begin{bmatrix} 0.2 & 0.1 & 0.2 \\ 0.5 & 0.5 & 0.4 \\ 0.1 & 0.3 & 0.2 \end{bmatrix}$, 判断矩阵幂级数 $\sum_{k=0}^{\infty} A^k$ 的敛散性. 若收敛, 试求其和.

$$\sum_{k=0}^{+\infty} A^k = (I - A)^{-1} = \frac{1}{14} \begin{bmatrix} 28 & 14 & 14 \\ 44 & 62 & 42 \\ 20 & 25 & 35 \end{bmatrix}.$$

例 8.6

已知 $A = \begin{bmatrix} \frac{1}{6} & -\frac{4}{3} \\ -\frac{1}{3} & \frac{1}{6} \end{bmatrix}$, 则 $\sum_{k=0}^{+\infty} A^k$ 收敛由于 $\rho(A) < \lambda_1 = 5/6$.

§8.3 矩阵函数

定义 8.20

设幂级数 $\sum\limits_{k=0}^{+\infty} a_k z^k$ 的收敛半径为 r, 且当 $|z| < r$ 时, 幂级数收敛于函数 $f(z)$, 即

$$f(z) = \sum_{k=0}^{+\infty} a_k z^k \quad (|z| < r).$$

如果 n 阶复方阵 A 满足 $\rho(A) < r$, 则称收敛的矩阵幂级数 $\sum\limits_{k=0}^{+\infty} a_k A^k$ 和为**矩阵函数**, 记为 $f(A)$, 即

$$f(A) = \sum_{k=0}^{+\infty} a_k A^k.$$

初等函数的泰勒展式

$$e^z = \sum_{k=0}^{+\infty} \frac{z^k}{k!} \qquad (r = +\infty)$$

$$\sin z = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k+1)!} z^{2k+1} \qquad (r = +\infty)$$

$$\cos z = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k)!} z^{2k} \qquad (r = +\infty)$$

$$(1-z)^{-1} = \sum_{k=0}^{+\infty} z^k \qquad (r = 1)$$

$$\ln(1+z) = \sum_{k=0}^{+\infty} \frac{(-1)^k}{k+1} z^{k+1} \qquad (r = 1)$$

常用矩阵函数

$$e^A = \sum_{k=0}^{+\infty} \frac{A^k}{k!} \qquad (\forall\, A \in \mathbb{C}^{n\times n})$$

$$\sin A = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k+1)!} A^{2k+1} \qquad (\forall\, A \in \mathbb{C}^{n\times n})$$

$$\cos A = \sum_{k=0}^{+\infty} \frac{(-1)^k}{(2k)!} A^{2k} \qquad (\forall\, A \in \mathbb{C}^{n\times n})$$

$$(I-A)^{-1} = \sum_{k=0}^{+\infty} A^k \qquad (\rho(A) < 1)$$

$$\ln(I+A) = \sum_{k=0}^{+\infty} \frac{(-1)^k}{k+1} A^{k+1} \qquad (\rho(A) < 1)$$

带参数 t 的矩阵函数

如果把矩阵函数 $f(A)$ 变元换成 At(t 为参数), 则得到

$$f(At)=\sum_{k=0}^{+\infty}a_k(At)^k=\sum_{k=0}^{+\infty}a_kA^kt^k.$$

例如:

$$e^{At}=\sum_{k=0}^{+\infty}\frac{A^kt^k}{k!} \qquad (\forall A\in\mathbb{C}^{n\times n})$$

$$\sin At=\sum_{k=0}^{+\infty}\frac{(-1)^k}{(2k+1)!}A^{2k+1}t^{2k+1} \qquad (\forall A\in\mathbb{C}^{n\times n})$$

$$\cos At=\sum_{k=0}^{+\infty}\frac{(-1)^k}{(2k)!}A^{2k}t^{2k} \qquad (\forall A\in\mathbb{C}^{n\times n})$$

常用矩阵函数的性质

定理 8.21

设 A 是 n 阶复方阵, $\lambda_1, \lambda_2, \cdots, \lambda_n$ 是 A 的 n 个特征值, 则矩阵函数 $f(A)$ 的特征值为 $f(\lambda_1), f(\lambda_2), \cdots, f(\lambda_n)$.

定理 8.22

对任意 n 阶复方阵 A, 有

(1) $\sin(-A) = -\sin A, \ \cos(-A) = \cos A;$

(2) $e^{\mathrm{i}A} = \cos A + \mathrm{i}\sin A, \ e^{-\mathrm{i}A} = \cos A - \mathrm{i}\sin A;$

(3) $\cos A = \dfrac{e^{\mathrm{i}A} + e^{-\mathrm{i}A}}{2}, \ \sin A = \dfrac{e^{\mathrm{i}A} - e^{-\mathrm{i}A}}{2\mathrm{i}};$

定理 8.23

设 A 和 B 是 n 阶复方阵, 且 $AB = BA$, 则

(1) $e^{A+B} = e^A e^B = e^B e^A$;

(2) $\cos(A+B) = \cos A \cos B - \sin A \sin B$;

(3) $\sin(A+B) = \sin A \cos B + \cos A \sin B$.

推论 8.24

对任意 n 阶复方阵 A, 有

$$\sin(2A) = 2\sin A \cos A, \quad \cos(2A) = \cos^2 A - \sin^2 A.$$

注记 8.25

当 $AB \neq BA$ 时, $e^{A+B} = e^A e^B$ 或 $e^{A+B} = e^B e^A$ 不成立. 比如取

$$A = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

则

$$e^A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, e^B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, e^A e^B = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \neq e^B e^A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

又

$$A + B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \rightsquigarrow e^{A+B} = \frac{1}{2} \begin{bmatrix} e + e^{-1} & e - e^{-1} \\ e - e^{-1} & e + e^{-1} \end{bmatrix}.$$

命题 8.26

$\det(e^A) = e^{\mathrm{tr}(A)} \rightsquigarrow e^A$ 可逆, $(e^A)^{-1} = e^{-A}$.

注记 8.27

不管 A 是否可逆, e^A 总是可逆, 而当 A 可逆时, $\sin A$, $\cos A$ 未必可逆. 例如若 π 为 A 的一个特征值, 则 $\det(\sin A) = \sin \pi \times \cdots = 0$

若 $\pi/2$ 为 A 的一个特征值, 则 $\det(\cos A) = \cos \frac{\pi}{2} \times \cdots = 0$.

矩阵函数值的计算

含参数的矩阵函数: t 为参数

$$f(At) = \sum_{k=0}^{+\infty} a_k (At)^k$$

计算矩阵函数值的方法:

- 利用矩阵本身的性质
- 利用相似对角化 (略)
- 公式法 (利用 Jordan 标准形)
- 待定系数法 (零化多项式: 如特征多项式, 最小多项式)

利用矩阵本身的性质

例 8.7

已知 $A=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$, 求 e^A, e^{At}, $\sin A$, $\cos At$.

$$e^{At}=\begin{bmatrix}\cos t & \sin t\\ -\sin t & \cos t\end{bmatrix},\ e^{A}=\begin{bmatrix}\cos 1 & \sin 1\\ -\sin 1 & \cos 1\end{bmatrix},$$

$$\cos At=\begin{bmatrix}\frac{e^t+e^{-t}}{2} & 0\\ 0 & \frac{e^t+e^{-t}}{2}\end{bmatrix},\ \sin A=\begin{bmatrix}0 & \frac{e-e^{-1}}{2}\\ \frac{e-e^{-1}}{2} & 0\end{bmatrix}.$$

公式法—利用 Jordan 标准形

r_i 阶 Jordan 块 $J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}_{r_i \times r_i}$, $f(\lambda t) = \sum_{k=0}^{+\infty} a_k \lambda^k t^k$ 为幂级数, 且 $|\lambda_i| < r$, r 为收敛半径, 有

$$f(J_i t) = \begin{bmatrix} f(\lambda t) & \frac{1}{1!}f'(\lambda t) & \frac{1}{2!}f''(\lambda t) & \cdots & \frac{1}{(r_i-1)!}f^{(r_i-1)}(\lambda t) \\ & f(\lambda t) & \frac{1}{1!}f'(\lambda t) & \cdots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & f(\lambda t) & \frac{1}{1!}f'(\lambda t) \\ & & & & f(\lambda t) \end{bmatrix}_{\lambda=\lambda_i}.$$

令 $u=\lambda t$, 则有$f(u)=\sum\limits_{k=0}^{+\infty} a_k u^k$ 且

$$f(J_i t)=\begin{bmatrix} f(u) & \frac{t}{1!}f'(u) & \frac{t^2}{2!}f''(u) & \cdots & \frac{t^{r_i-1}}{(r_i-1)!}f^{(r_i-1)}(u) \\ & f(u) & \frac{t}{1!}f'(u) & \cdots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & f(u) & \frac{t}{1!}f'(u) \\ & & & & f(u) \end{bmatrix}_{u=\lambda_i t}.$$

例 8.8

已知 $A=\begin{bmatrix}2&1&0&0\\0&2&1&0\\0&0&2&1\\0&0&0&2\end{bmatrix}$, 求 e^A, e^{At}, $\sin At$, $\cos A$.

$$e^{At}=\begin{bmatrix}e^{2t}&te^{2t}&\frac{t^2}{2}e^{2t}&\frac{t^3}{3!}e^{2t}\\0&e^{2t}&te^{2t}&\frac{t^2}{2}e^{2t}\\0&0&e^{2t}&te^{2t}\\0&0&0&e^{2t}\end{bmatrix}\rightsquigarrow e^A=\begin{bmatrix}e^2&e^2&\frac{1}{2}e^2&\frac{1}{3!}e^2\\0&e^2&e^2&\frac{1}{2}e^2\\0&0&e^2&e^2\\0&0&0&e^2\end{bmatrix}$$

$$\sin At=\begin{bmatrix}\sin 2t&t\cos 2t&-\frac{t^2}{2}\sin 2t&-\frac{t^3}{3!}\cos 2t\\0&\sin 2t&t\cos 2t&-\frac{t^2}{2}\sin 2t\\0&0&\sin 2t&t\cos 2t\\0&0&0&\sin 2t\end{bmatrix}$$

公式法—利用 Jordan 标准形

设 A 是 n 阶复方阵, 则存在可逆矩阵 P 使得

$$P^{-1}AP = \Lambda = \mathrm{diag}(J_1, J_2, \cdots, J_s),$$

其中 J_i 是特征值 λ_i 对应的阶为 r_i 的 Jordan 块. 则有

$$f(At) = P\,\mathrm{diag}(f(J_1t), f(J_2t), \cdots, f(J_st))P^{-1}.$$

例 8.9

已知 $A = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 2 & 0 \\ -4 & 0 & 3 \end{bmatrix}$, 求 e^{At}, $\sin At$, $\cos At$.

第一步: 求得 Jordan 标准形和相应的相似变换矩阵

$$P = \begin{bmatrix} 1 & 0 & 0 \\ -1 & -1 & 1 \\ 2 & 1 & 0 \end{bmatrix}, \quad P^{-1}AP = \Lambda = \begin{bmatrix} 1 & 1 & \\ & 1 & \\ & & 2 \end{bmatrix}$$

第二步: 计算矩阵函数值

待定系数法—零化多项式

设 A 是 n 阶复方阵, 其特征多项式为

$$\varphi(\lambda) = \det(\lambda I - A) = (\lambda - \lambda_1)^{r_1}(\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_s)^{r_s}$$

其中 $\lambda_1, \cdots, \lambda_s$ 是 A 的互异特征值, $r_1 + \cdots + r_s = n$.

由 Hamilton-Caylay 定理知 $\varphi(A) = 0$.

计算矩阵函数 $f(A)=\sum\limits_{k=0}^{+\infty} a_k A^k$ 的方法:

设 $f(\lambda)=\sum\limits_{k=0}^{+\infty} a_k\lambda^k$, 由带余除法

$$f(\lambda)=q(\lambda)\varphi(\lambda)+r(\lambda),\ \ r(\lambda)=b_{n-1}\lambda^{n-1}+\cdots+b_1\lambda+b_0,$$

则有 $f(A)=r(A)$. 如何计算 $b_0,b_1,\cdots,b_{n-1}$.

注意到 $\varphi^{(l)}(\lambda_i)=0(l=0,1,\cdots,r_i-1;i=1,2,\cdots,s)$ 有

$$r^{(l)}(\lambda_i)=f^{(l)}(\lambda_i)\quad(l=0,1,\cdots,r_i-1;i=1,2,\cdots,s)$$

列方程组求解 $b_0,b_1,\cdots,b_{n-1}$;

从而 $f(A)=r(A)=b_{n-1}A^{n-1}+\cdots+b_1A+b_0I$.

注记 8.28

用最小多项式 $m(\lambda)$ 效果更好.

计算矩阵函数 $f(At)=\sum_{k=0}^{+\infty} a_k(At)^k=\sum_{k=0}^{+\infty}(a_k t^k)A^k$ 的方法:

$$\text{设} \qquad f(\lambda t)=\sum_{k=0}^{+\infty}(a_k t^k)\lambda^k=\sum_{k=0}^{+\infty}a_k(t)\lambda^k.$$

由带余除法

$$f(\lambda t)=q_t(\lambda)\varphi(\lambda)+r_t(\lambda),\ r_t(\lambda)=b_0(t)+b_1(t)\lambda+\cdots+b_{n-1}(t)\lambda^{n-1}.$$

同样对 λ 求导, 由 $\varphi^{(l)}(\lambda_i)=0$ 知

$$\frac{\mathrm{d}^l}{\mathrm{d}\,\lambda^l}f(\lambda t)\Big|_{\lambda=\lambda_i}=\frac{\mathrm{d}^l}{\mathrm{d}\,\lambda^l}r_t(\lambda)\Big|_{\lambda=\lambda_i}\quad (l=0,\cdots,r_i-1;i=1,\cdots,s)$$

例 8.10

已知 $A = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 2 & 0 \\ -4 & 0 & 3 \end{bmatrix}$, 求 e^{At}, $\cos A$.

第一步: 求出特征多项式

$$\det(\lambda I - A) = (\lambda - 1)^2(\lambda - 2)$$

第二步: 利用待定系数法计算矩阵函数值

利用最小多项式

例 8.11

已知 $A = \begin{bmatrix} 3 & 1 & -1 \\ -2 & 0 & 2 \\ -1 & -1 & 3 \end{bmatrix}$, 求 e^{At}, $\sin A$.

第一步: 求出最小多项式

$$\text{Jordan 标准形 } \Lambda = \begin{bmatrix} 2 & & \\ & 2 & 1 \\ & & 2 \end{bmatrix} \rightsquigarrow m_A(\lambda) = (\lambda - 2)^2.$$

第二步: 利用待定系数法计算矩阵函数值.

§8.4 矩阵的微分和积分

定义 8.29

以变量 t 的函数为元素的矩阵 $A(t)=(a_{ij}(t))_{m\times n}$ 称为函数矩阵, 其中 $a_{ij}(t)(i=1,2,\cdots,m;j=1,2,\cdots,n)$ 都是变量 t 的函数. 若 $t\in[a,b]$, 则称 $A(t)$ 是定义在 $[a,b]$ 上的; 又若每个 $a_{ij}(t)$ 在 $[a,b]$ 上连续、可微、可积, 则称函数矩阵 $A(t)$ 在 $[a,b]$ 上是连续、可微、可积的.

例 8.12

$A(t)=\begin{bmatrix} t^2+1 & 1 & e^t \\ 0 & \sin t & 2 \\ 1 & 0 & \ln t \end{bmatrix}$ 是一矩阵函数.

定义 8.30

当 $A(t)$ 可微时, 规定其 **导数**为

$$A'(t) = \left(a'_{ij}(t)\right)_{m\times n} \quad \text{或} \quad \frac{\mathrm{d}}{\mathrm{d}\,t}A(t) = \left(\frac{\mathrm{d}}{\mathrm{d}\,t}a_{ij}(t)\right)_{m\times n},$$

而当 $A(t)$ 在 $[a,b]$ 上可积时, 规定 $A(t)$ 在 $[a,b]$ 上的**积分**为

$$\int_a^b A(t)\,\mathrm{d}\,t = \left(\int_a^b a_{ij}(t)\,\mathrm{d}\,t\right)_{m\times n}.$$

微分的性质

定理 8.31

设 $A(t)$ 与 $B(t)$ 是适当阶的可微矩阵, 则

(1) $\big(A(t)+B(t)\big)' = A'(t)+B'(t)$

(2) 当 $\lambda(t)$ 为可微函数时, 有

$$(\lambda(t)A(t))' = \lambda'(t)A(t)+\lambda(t)A'(t)$$

(3) $\big(A(t)B(t)\big)' = A'(t)B(t)+A(t)B'(t)$

注记 8.32

一般地, $\frac{\mathrm{d}}{\mathrm{d}\,t}(A(t))^m \neq m(A(t))^{m-1}\frac{\mathrm{d}}{\mathrm{d}\,t}(A(t))$, 取 $m=2$ 验证即可. 等式成立的条件为

$$A'(t)A(t) = A(t)A'(t)$$

微分的性质 (续)

定理 8.33

设 $A(t)$ 与 $B(t)$ 是适当阶的可微矩阵, 则

(4) 当 $u=f(t)$ 关于 t 可微时, 有 $\frac{\mathrm{d}}{\mathrm{d}\,t}A(u)=f'(t)\frac{\mathrm{d}}{\mathrm{d}\,u}A(u)$

(5) 当 $A^{-1}(t)$ 是可微矩阵时, 有 $\frac{\mathrm{d}}{\mathrm{d}\,t}(A^{-1}(t))=-A^{-1}(t)(\frac{\mathrm{d}}{\mathrm{d}\,t}A(t))A^{-1}(t)$

定理 8.34

设 A 是 n 阶方阵, 则有

(1) $\frac{\mathrm{d}}{\mathrm{d}\,t}e^{At}=Ae^{At}=e^{At}A;$

(2) $\frac{\mathrm{d}}{\mathrm{d}\,t}\sin(At)=A\cos(At)=\cos(At)A;$

(3) $\frac{\mathrm{d}}{\mathrm{d}\,t}\cos(At)=-A\sin(At)=-\sin(At)A.$

积分的性质

定理 8.35

设 $A(t)$ 与 $B(t)$ 是区间 $[a,b]$ 上适当阶的可积矩阵, A,B 是适当阶的常数矩阵, $\lambda\in\mathbb{C}$, 则

(1) $\int_a^b (A(t)+B(t))\,\mathrm{d}\,t=\int_a^b A(t)\,\mathrm{d}\,t+\int_a^b B(t)\,\mathrm{d}\,t$,

(2) $\int_a^b \lambda A(t)\,\mathrm{d}\,t=\lambda\int_a^b A(t)\,\mathrm{d}\,t$,

(3) $\int_a^b A(t)B\,\mathrm{d}\,t=\Big(\int_a^b A(t)\,\mathrm{d}\,t\Big)B$, (区分矩阵乘法)

$\int_a^b AB(t)\,\mathrm{d}\,t=A\int_a^b B(t)\,\mathrm{d}\,t$,

微积分基本定理

定理 8.36

设 $A(t)$ 与 $B(t)$ 是区间 $[a,b]$ 上适当阶的可积矩阵, A,B 是适当阶的常数矩阵, 则

(4) 当 $A(t)$ 在 $[a,b]$ 上连续时, 对任意 $t\in(a,b)$, 有

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\left(\int_a^t A(\tau)\,\mathrm{d}\,\tau\right)=A(t)$$

(5) 当 $A(t)$ 在 $[a,b]$ 上连续可微时, 有

$$\int_a^b A'(t)\,\mathrm{d}\,t=A(b)-A(a)$$

数量函数对矩阵变量的导数

定义 8.37

设 $f(X)$ 是以矩阵 $X=(x_{ij})_{m\times n}$ 为自变量的 mn 元函数, 且 $\dfrac{\partial f}{\partial x_{ij}}(i=1,2,\cdots,m;$ $j=1,2,\cdots,n)$ 都存在, 规定 f 对矩阵变量 X 的导数 $\dfrac{\mathrm{d}f}{\mathrm{d}X}$ 为

$$\frac{\mathrm{d}f}{\mathrm{d}X}=\left(\frac{\partial f}{\partial x_{ij}}\right)_{m\times n}=\begin{bmatrix}\dfrac{\partial f}{\partial x_{11}} & \cdots & \dfrac{\partial f}{\partial x_{1n}}\\ \vdots & & \vdots\\ \dfrac{\partial f}{\partial x_{m1}} & \cdots & \dfrac{\partial f}{\partial x_{mn}}\end{bmatrix}$$

特别地, 以 $\mathbf{x}=(x_1,x_2,\cdots,x_n)^T$ 为自变量的函数 $f(\mathbf{x})$ 的导数

$$\mathrm{grad}\, f=\frac{\mathrm{d}\, f}{\mathrm{d}\,\mathbf{x}}=(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\cdots,\frac{\partial f}{\partial x_n})^T$$

称为数量函数对向量变量的导数, 即函数 f 的梯度向量.

例 8.13

给定向量 $\mathbf{a}=(a_1,a_2,\cdots,a_n)^T$. 设 $\mathbf{x}=(x_1,x_2,\cdots,x_n)^T$ 是向量变量, 且 $f(\mathbf{x})=\mathbf{a}^T\mathbf{x}=\mathbf{x}^T\mathbf{a}$, 求 $\frac{\mathrm{d}\, f}{\mathrm{d}\,\mathbf{x}}$. $(=\mathbf{a})$

例 8.14

给定矩阵 $A=(a_{ij})_{n\times n}$. 设 $\mathbf{x}=(x_1,x_2,\cdots,x_n)^T$ 是向量变量, 且 $f(\mathbf{x})=\mathbf{x}^TA\mathbf{x}$, 求 $\frac{\mathrm{d}\, f}{\mathrm{d}\,\mathbf{x}}$. $\left(=(A+A^T)\mathbf{x}\right)$

最小二乘问题

设 $A \in \mathbb{C}^{m\times n}, \mathbf{b} \in \mathbb{C}^m$. 当线性方程组 $A\mathbf{x} = \mathbf{b}$ 无解时, 我们希望找一个向量 $\mathbf{x}_0$ 使得 $\|A\mathbf{x} - \mathbf{b}\|_2$ 达到最小, 即

$$\|A\mathbf{x}_0 - \mathbf{b}\|_2 = \min_{\mathbf{x}\in\mathbb{C}^n} \|A\mathbf{x} - \mathbf{b}\|_2.$$

称 $\mathbf{x}_0$ 为矛盾方程组 $A\mathbf{x} = \mathbf{b}$ 的最小二乘解.

定理 8.38

设 $A \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$. 若 $\mathbf{x}_0 \in \mathbb{R}^n$ 是 $A\mathbf{x} = \mathbf{b}$ 的最小二乘解, 则 $\mathbf{x}_0$ 是方程组 (normal equations)

$$A^T A\mathbf{x} = A^T\mathbf{b}$$

的解.

例 8.15

给定矩阵 $A=(a_{ij})_{m\times n}$. 设 $X=(x_{ij})_{n\times m}$ 是矩阵变量, 且 $f(X)=\mathrm{tr}(AX)$, 求 $\dfrac{\mathrm{d}f}{\mathrm{d}X}$. $(=A^T)$

例 8.16

设 $X=(x_{ij})_{n\times n}$ 是矩阵变量, 且 $\det X\neq 0$, 证明

$$\frac{\mathrm{d}}{\mathrm{d}X}\det X=(\det X)(X^{-1})^T.$$

矩阵值函数对矩阵变量的导数

定义 8.39

设矩阵 $F(X)=(f_{ij})_{s\times t}$ 的元素 $f_{ij}(X)$ $(i=1,2,\cdots,s; j=1,2,\cdots,t)$ 都是矩阵变量 $X=(x_{ij})_{m\times n}$ 的函数, 则称 $F(X)$ 为矩阵值函数, 规定 $F(X)$ 对矩阵变量 X 的导数 $\dfrac{\mathrm{d}F}{\mathrm{d}X}$ 为

$$\frac{\mathrm{d}F}{\mathrm{d}X}=\begin{bmatrix} \dfrac{\partial F}{\partial x_{11}} & \cdots & \dfrac{\partial F}{\partial x_{1n}} \\ \vdots & & \vdots \\ \dfrac{\partial F}{\partial x_{m1}} & \cdots & \dfrac{\partial F}{\partial x_{mn}} \end{bmatrix}_{ms\times nt}, \text{其中 } \frac{\partial F}{\partial x_{ij}}=\begin{bmatrix} \dfrac{\partial f_{11}}{\partial x_{ij}} & \cdots & \dfrac{\partial f_{1t}}{\partial x_{ij}} \\ \vdots & & \vdots \\ \dfrac{\partial f_{s1}}{\partial x_{ij}} & \cdots & \dfrac{\partial f_{st}}{\partial x_{ij}} \end{bmatrix}$$

例 8.17

已知 $A=(a_{ij})_{m\times n}$, $\mathbf{x}=(x_1,x_2,\cdots,x_m)^T$, 且 $F(\mathbf{x})=\mathbf{x}^TA$, 求 $\dfrac{\mathrm{d}\,F}{\mathrm{d}\,\mathbf{x}}$.

由于 $F(\mathbf{x})=\mathbf{x}^TA=(\sum\limits_{k=1}^{m}x_ka_{k1},\sum\limits_{k=1}^{m}x_ka_{k2},\cdots,\sum\limits_{k=1}^{m}x_ka_{kn})$, 因为

$$\frac{\partial F}{\partial x_i}=(a_{i1},a_{i2},\cdots,a_{in}),$$

所以

$$\frac{\mathrm{d}\,F}{\mathrm{d}\,\mathbf{x}}=\begin{bmatrix}\frac{\partial F}{\partial x_1}\\ \vdots\\ \vdots\\ \frac{\partial F}{\partial x_m}\end{bmatrix}=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\ a_{21}&a_{22}&\cdots&a_{2n}\\ \vdots&\vdots&\vdots&\\ a_{m1}&a_{m2}&\cdots&a_{mn}\end{bmatrix}=A.$$

§8.5 矩阵分析应用举例

求一阶常系数微分方程组

$$\begin{cases} \dfrac{\mathrm{d}\,x_1(t)}{\mathrm{d}\,t} = a_{11}x_1(t) + a_{12}x_2 t + \cdots + a_{1n}x_n(t) + f_1(t) \\ \dfrac{\mathrm{d}\,x_2(t)}{\mathrm{d}\,t} = a_{21}x_1(t) + a_{22}x_2 t + \cdots + a_{2n}x_n(t) + f_2(t) \\ \qquad\qquad \cdots\cdots\cdots\cdots \\ \dfrac{\mathrm{d}\,x_n(t)}{\mathrm{d}\,t} = a_{n1}x_1(t) + a_{n2}x_2 t + \cdots + a_{nn}x_n(t) + f_n(t) \end{cases}$$

满足初始条件 $x_i(t_0) = c_i \quad (i = 1, 2, \cdots, n)$ 的解.

令 $A = (a_{ij})_{n\times n}$, $\mathbf{x}(t) = (x_1(t), \cdots, x_n(t))^T$, $\mathbf{f}(t) = (f_1(t), \cdots, f_n(t))^T$, $\mathbf{c} = (c_1, \cdots, c_n)^T$,

$$\begin{cases} \dfrac{\mathrm{d}\,\mathbf{x}(t)}{\mathrm{d}\,t} = A\mathbf{x}(t) + \mathbf{f}(t) \\ \mathbf{x}(t_0) = \mathbf{c} \qquad \rightsquigarrow \text{初始条件} \end{cases}$$

如果 $\mathbf{f}(t) = (f_1(t), f_2(t), \cdots, f_n(t))^T = 0$, 则称该方程组是齐次的, 否则称为非齐次的.

将方程组改写为 $\dfrac{\mathrm{d}\,\mathbf{x}(t)}{\mathrm{d}\,t} - A\mathbf{x}(t) = \mathbf{f}(t)$, 两边左乘 e^{-At} 得

$$e^{-At}\frac{\mathrm{d}\,\mathbf{x}(t)}{\mathrm{d}\,t} - e^{-At}A\mathbf{x}(t) = e^{-At}\mathbf{f}(t)$$

整理得

$$\frac{\mathrm{d}}{\mathrm{d}\,t}(e^{-At}\mathbf{x}(t)) = e^{-At}\mathbf{f}(t)$$

在 $[t_0, t]$ 上积分得

$$e^{-At}\mathbf{x}(t) - e^{-At_0}\mathbf{x}(t_0) = \int_{t_0}^{t} e^{-A\tau}\mathbf{f}(\tau)\,\mathrm{d}\,\tau$$

则微分方程组的解为

$$\mathbf{x}(t) = e^{A(t-t_0)}\mathbf{c} + e^{At}\int_{t_0}^{t} e^{-A\tau}\mathbf{f}(\tau)\,\mathrm{d}\,\tau.$$

例 8.18

用矩阵函数方法求解微分方程组

$$
\begin{cases}
\dfrac{\mathrm{d}\,x_1}{\mathrm{d}\,t} = 2x_1 + x_2 + 4x_3 + 4t \\
\dfrac{\mathrm{d}\,x_2}{\mathrm{d}\,t} = 2x_2 \\
\dfrac{\mathrm{d}\,x_3}{\mathrm{d}\,t} = 3x_2 + x_3 \\
x_1(0) = 2,\ x_2(0) = 1,\ x_3(0) = 3
\end{cases}
$$

$$
A = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 2 & 0 \\ 0 & 3 & 1 \end{bmatrix},\ \mathbf{f}(t) = \begin{bmatrix} 4t \\ 0 \\ 0 \end{bmatrix},\ \mathbf{c} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}.
$$

求解 Lyapunov(李雅普洛夫) 方程

Lyapunov(李雅普诺夫) 方程: $AX + XB = F$, 其中 A, B 均为方阵.

定理 8.40

给定李雅普诺夫方程 $AX + XB = F$. 若 A 和 B 的所有特征值具有负实部 (即稳定矩阵), 则该矩阵方程有唯一解

$$X = -\int_0^{+\infty} e^{At} F e^{Bt} \, \mathrm{d}\, t.$$

推论 8.41

矩阵微分方程 $\begin{cases} \dfrac{\mathrm{d}\,X(t)}{\mathrm{d}\,t} = AX(t) + X(t)B \\ X(0) = F \end{cases}$ 的解为

$$X(t) = e^{At}Fe^{Bt}.$$

推论 8.42

设方阵 A 的所有特征值均具有负实部, 则矩阵方程 $A^HX + XA = -F$ 的唯一解为

$$X = \int_0^{+\infty} e^{A^Ht}Fe^{At}\,\mathrm{d}\,t.$$

如果 F 是 Hermite 正定矩阵, 则解矩阵 X 也是 Hermite 正定矩阵.

第 9 章 特征值的估计与表示

§9.1 特征值界的估计

引理 9.1

设 $A \in \mathbb{C}^{n\times n}, y \in \mathbb{C}^n$ 满足 $\|y\|_2 = 1$, 则 $|y^H A y| \leqslant \|A\|_{m_\infty}$.

定理 9.2

设 $A \in \mathbb{C}^{n\times n}, B = \dfrac{1}{2}(A + A^H), C = \dfrac{1}{2}(A - A^H)$, 则 A 的任一特征值 λ 满足

$$|\lambda| \leqslant \|A\|_{m_\infty},$$

$$|\operatorname{Res}(\lambda)| \leqslant \|B\|_{m_\infty}, |\operatorname{Im}(\lambda)| \leqslant \|C\|_{m_\infty}.$$

推论 9.3

Hermite 矩阵的特征值都是实数, 反 Hermite 矩阵的特征值为零或纯虚数.

实矩阵特征值虚部的界

定理 9.4

设 $A \in \mathbb{R}^{n\times n}$, $C = \dfrac{1}{2}(A - A^T)$, 则 A 的任一特征值 λ 满足

$$|\operatorname{Im}(\lambda)| \leqslant \sqrt{\frac{n-1}{2n}}\,\|C\|_{m_\infty}.$$

例 9.1

设 $A = \begin{bmatrix} 0 & 2 & 1 \\ -2 & 0 & 2 \\ -1 & -2 & 0 \end{bmatrix}$, 估计 A 的特征值的界.

解: 显然, $B = \dfrac{1}{2}(A + A^T) = 0,\ C = \dfrac{1}{2}(A - A^T) = A$, 所以

$$\|A\|_{m_\infty} = 6, \quad \|B\|_{m_\infty} = 0, \quad \|C\|_{m_\infty} = 6.$$

由定理9.2知

$$|\lambda| \leqslant 6, \quad |\,\mathsf{Res}(\lambda)| = 0, \quad |\,\mathsf{Im}(\lambda)| \leqslant 6.$$

但是由定理9.4知

$$|\,\mathsf{Im}(\lambda)| \leqslant \sqrt{\frac{3-1}{2\times 3}}\|C\|_{m_\infty} \approx 3.4641.$$

A 的特征值为 $\lambda_1 = 0, \lambda_2 = 3\,\mathrm{i}, \lambda_3 = -3\,\mathrm{i}$.

特征值模的平方和的估计

定理 9.5 (Schur)

设 A 的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$, 则

$$|\lambda_1|^2 + |\lambda_2|^2 + \cdots + |\lambda_n|^2 \leqslant \|A\|_F^2,$$

且等号成立当且仅当 $AA^H = A^HA$, 即为正规矩阵.

提示: 由 Schur 引理知存在酉矩阵 U 使得 $U^HAU = T$, 这里 T 是上三角矩阵. 于是

$$\sum_{k=1}^{n} |\lambda_k|^2 = \sum_{k=1}^{n} |t_{kk}|^2 \leqslant \|T\|_F^2 = \|U^HAU\|_F^2 = \|A\|_F^2.$$

等号成立 $\iff$ T 为对角矩阵 $\iff$ A 为正规矩阵.

§9.2 特征值的包含区域

定义 9.6

设 $A=(a_{ij})_{n\times n}$, 记

$$R_i=\sum_{j=1,j\neq i}^{n}|a_{ij}|\quad (i=1,2,\cdots,n),$$

称复平面上的圆域

$$G_i=G_i(A)=\{z \mid |z-a_{ii}|\leqslant R_i,\ z\in\mathbb{C}\}\quad (i=1,2,\cdots,n)$$

为矩阵 A 的第 i 个 Gerschgorin discs(盖尔圆), R_i 为 G_i 的半径.

定理 9.7 (Gerschgorin I)

矩阵 $A \in \mathbb{C}^{n\times n}$ 的全体特征值都在它的 n 个盖尔圆构成的并集之中.

注记 9.8

A^T 与 A 的特征值相同, 所以 A 的特征值都在 A^T 的 n 个盖尔圆的并集之中, 称 A^T 的盖尔圆为 A 的列盖尔圆.

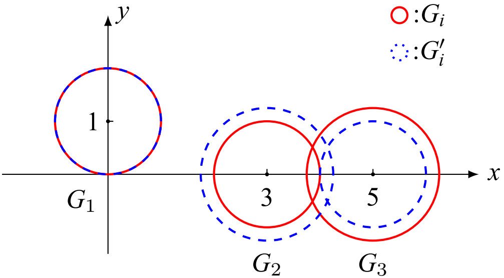
例 9.2

估计矩阵

$$A = \begin{bmatrix} \mathrm{i} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 3 & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{4} & 5 \end{bmatrix}$$

的特征值范围.

A 的 3 个行盖尔圆为 $G_1 : |z-\mathrm{i}| \leqslant 1$, $G_2 : |z-3| \leqslant 1$, $G_3 : |z-5| \leqslant \frac{5}{4}$,
A 的 3 个列盖尔圆为 $G_1' : |z-\mathrm{i}| \leqslant 1$, $G_2' : |z-3| \leqslant \frac{5}{4}$, $G_3' : |z-5| \leqslant 1$.



例 9.3

估计矩阵

$$A=\begin{bmatrix} 2 & -1 & -2 & 0 \\ -1 & 3 & 2\,\mathrm{i} & 0 \\ 0 & -\mathrm{i} & 10 & \mathrm{i} \\ -2 & 0 & 0 & 6\,\mathrm{i} \end{bmatrix}$$

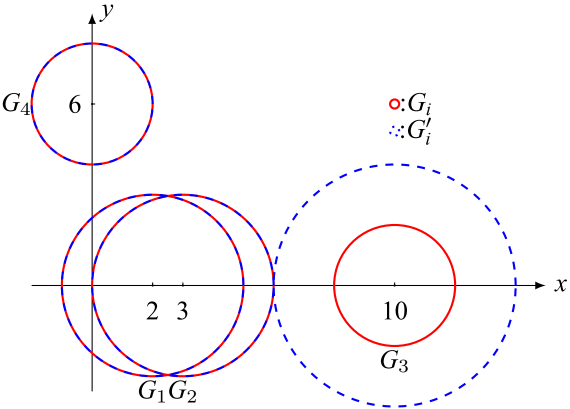
特征值分布范围.

提示: 矩阵 A 的四个行盖尔圆为

$$G_1:|z-2|\leqslant 3,\quad G_2:|z-3|\leqslant 3,$$
$$G_3:|z-10|\leqslant 2,\quad G_4:|z-6\,\mathrm{i}\,|\leqslant 2$$

故 A 的 4 个特征值都在 $\cup_{i=1}^{4}G_i$ 中. 矩阵 A' 的四个行盖尔圆为

$$G_1':|z-2|\leqslant 3,\quad G_2':|z-3|\leqslant 3,$$
$$G_3':|z-10|\leqslant 4,\quad G_4':|z-6\,\mathrm{i}\,|\leqslant 1,$$



问题：每个盖尔圆中具体有多少个特征值?

定义 9.9

矩阵 A 盖尔圆中, 相交在一起的盖尔圆构成的最大连通区域称为一个连通分支. 孤立的一个盖尔圆也是一个连通部分.

定理 9.10 (Gerschgorin II)

若矩阵 A 的某一连通部分由 A 的 k 个盖尔圆构成, 则其中有且仅有 A 的 k 个特征值 (盖尔圆相重时重复计数, 特征值相同时也重复计数).

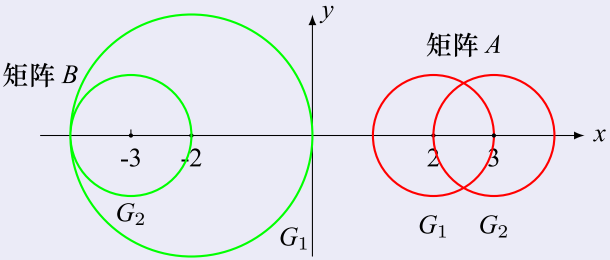
注记 9.11

由两个或两个以上的盖尔圆构成的连通分支, 特征值不一定是平均的, 即可能在其中的某个盖尔圆中有几个特征值, 而另外的一些盖尔圆中没有特征值.

例 9.4

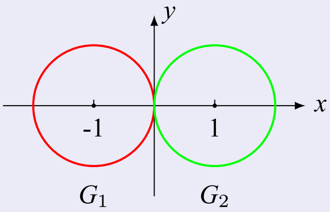
比如说, $A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$, 两个盖尔圆为 $G_1 : |z-2| \leqslant 1$, $G_2 : |z-3| \leqslant 1$. 很容易求出 A 的特征值为 $\lambda_1 = \frac{5+\sqrt{5}}{2} \approx 3.6$, $\lambda_2 = \frac{5-\sqrt{5}}{2} \approx 1.38$.

而对 $B = \begin{bmatrix} -2 & 2 \\ -1 & -3 \end{bmatrix}$, 两个盖尔圆为 $G_1 : |z+2| \leqslant 2$, $G_2 : |z+3| \leqslant 1$. B 的特征值为 $\lambda_1 = \frac{-5+\sqrt{7}i}{2}$, $\lambda_2 = \frac{-5-\sqrt{7}i}{2}$, 且 $\mathrm{Im}(\lambda_{1,2}) = \frac{\sqrt{7}}{2} \approx 1.3$.



例 9.5

$C=\begin{bmatrix}-1 & -1\\ 1 & 1\end{bmatrix}$, 两个盖尔圆为 $G_1:|z+1|\leqslant 1$, $G_2:|z-1|\leqslant 1$. 很容易求出 C 的特征值为 $\lambda_1=\lambda_2=0$. 做图可见, 他们在两个盖尔圆的公共部分.



推论 9.12

若 $A \in \mathbb{C}^{n\times n}$ 的 n 个盖尔圆互不相交, 则 A 相似与对角阵.

推论 9.13

若 $A \in \mathbb{R}^{n\times n}$ 的 n 个盖尔圆互不相交, 则 A 的特征值均为实数.

推论 9.14

若 $A = (a_{ij}) \in \mathbb{C}^{n\times n}$ 是按行 (列) 严格对角占优的, 即

$$|a_{kk}| > R_k (k = 1, 2, \cdots, n),$$

则 $\det A \neq 0$.

隔离定理

应用盖尔圆定理估计矩阵的特征值时, 往往希望每个盖尔圆中只含有一个特征值. 当 A 的若干个盖尔圆相交时, 通常采用如下两种方法隔离其特征值.

一、考虑 A 的列盖尔圆. 由于 A^T 与 A 有相同的特征值, 所以 A 的所有特征值包含在 n 个列盖尔圆

$$G_j(A^T) = \{z \mid |z - a_{jj}| \leqslant \sum_{k=1,k\neq j}^{n} |a_{kj}|\} \quad (j = 1, 2, \cdots, n)$$

的并集之中. 若有 k 个列盖尔圆构成一个孤立与其他圆盘的连通部分, 则其中恰有 A 的 k 个特征值.

定理 9.15

$A=(a_{ij})\in\mathbb{C}^{n\times n}$ 的所有特征值都位于复平面的区域

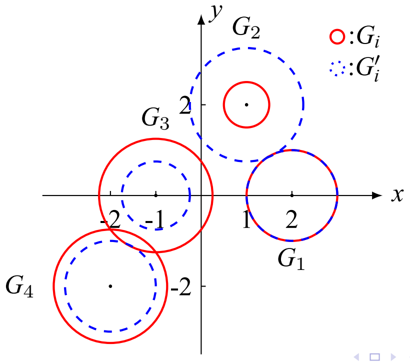
$$(G_1(A)\cup G_2(A)\cup\cdots\cup G_n(A))\cap(G_1(A^T)\cup G_2(A^T)\cup\cdots\cup G_n(A^T)).$$

例 9.6

估算矩阵

$$A = \begin{bmatrix} 2 & -\frac{1}{2} & \frac{1}{4} & -\frac{1}{4} \\ \frac{1}{4} & 1+2\,\mathrm{i} & 0 & \frac{1}{4} \\ -\frac{1}{2} & \frac{1}{4} & -1 & \frac{1}{2} \\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{2} & -2-2\,\mathrm{i} \end{bmatrix}$$

的特征值范围并作图表示.



二、利用相似性 (注意与课本有差别). 选取正数 $d_1, \cdots, d_n$, 并设对角矩阵

$$D = \mathrm{diag}\{d_1, d_2, \cdots, d_n\},$$

构造与 A 相似的矩阵 $B = D^{-1}AD = (a_{ij}\frac{d_j}{d_i})_{n\times n}$. 则 A, B 有相同的特征值. 适当选取正数 $d_1, \cdots, d_n$, 可能使 B 的每个盖尔圆包含 A 的一个特征值.

定理 9.16

设 $A = (a_{ij}) \in \mathbb{C}^{n\times n}$, $d_1, \cdots, d_n$ 为任一组正数, 则 A 的所有特征值包含在 n 个盖尔圆

$$\widetilde{G}_k(A) = \{z \mid |z - a_{kk}| \leqslant \sum_{j=1, j\neq k}^{n} \frac{d_j}{d_k}|a_{kj}|\} \quad (k = 1, 2, \cdots, n)$$

的并集中.

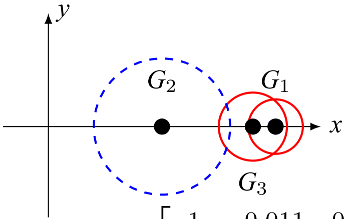
注记 9.17

- 取 $B = D^{-1}AD$, 而课本取的是 $B = DAD^{-1}$, 本质无差别, 具体用时需注意.
- 选 $d_1, d_2, \cdots, d_n$ 的一般原则是: 欲使 A 的第 i 个盖尔圆缩小, 可取 $d_i > 1$, 其余取 1, 此时 B 的其余盖尔圆适量放大 (相对于 A 的同序号盖尔圆); 反之, 欲使 A 的第 i 个盖尔圆放大, 可取 $d_i < 1$, 其余取 1, 此时 B 的其余盖尔圆适量缩小 (相对于 A 的同序号盖尔圆). 这里 d_i 与课本恰好是倒数关系.
- 并不是任意具有互异特征值的矩阵都能用这种方法隔离特征值, 比如主对角线上有相同元素的矩阵.

例 9.7

试估计矩阵 $A = \begin{bmatrix} 1 & 0.11 & 0.02 \\ 0.02 & 0.5 & 0.01 \\ 0.01 & 0.14 & 0.9 \end{bmatrix}$ 的特征值分布范围, 并适当选择一组正数, 使得 A 的三个盖尔圆互不相交.

A 的三个盖尔圆为 $G_1 : |z-1| \leqslant 0.12,\ G_2 : |z-0.5| \leqslant 0.03,\ G_3 : |z-0.9| \leqslant 0.15.$

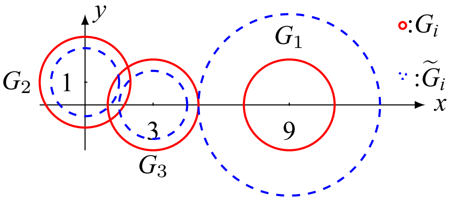


取 $D = \mathrm{diag}\{1, 0.1, 1\}$, $B = D^{-1}AD = \begin{bmatrix} 1 & 0.011 & 0.02 \\ 0.2 & 0.5 & 0.1 \\ 0.01 & 0.014 & 0.9 \end{bmatrix}$, B 的三个盖尔圆 $\widetilde{G}_1, \widetilde{G}_2, \widetilde{G}_3$, 在 $\widetilde{G}_1, G_2, \widetilde{G}_3$ 中各有 1 个特征值.

例 9.8

试分离矩阵 $A=\begin{bmatrix}9&1&1\\1&\mathrm{i}&1\\1&1&3\end{bmatrix}$ 的特征值.

A 的三个盖尔圆为 $G_1:|z-9|\leqslant 2,\ G_2:|z-\mathrm{i}\,|\leqslant 2,\ G_3:|z-3|\leqslant 2.$

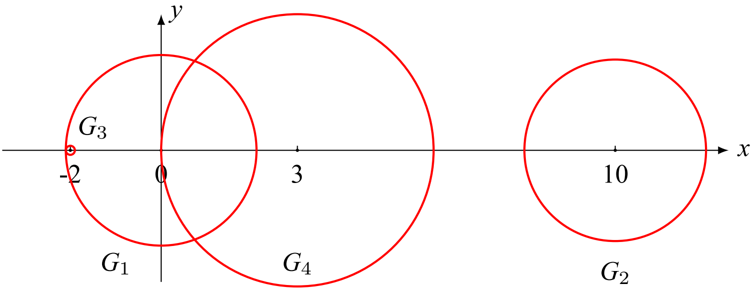


取 $D=\mathrm{diag}\{0.5,1,1\}$, 则 $B=D^{-1}AD=\begin{bmatrix}9&2&2\\0.5&\mathrm{i}&1\\0.5&1&3\end{bmatrix}$, B 的三个盖尔圆 $\widetilde{G}_1,\widetilde{G}_2,\widetilde{G}_3$, 在 $G_1,\widetilde{G_2},\widetilde{G}_3$ 中各有 1 个特征值.

例 9.9

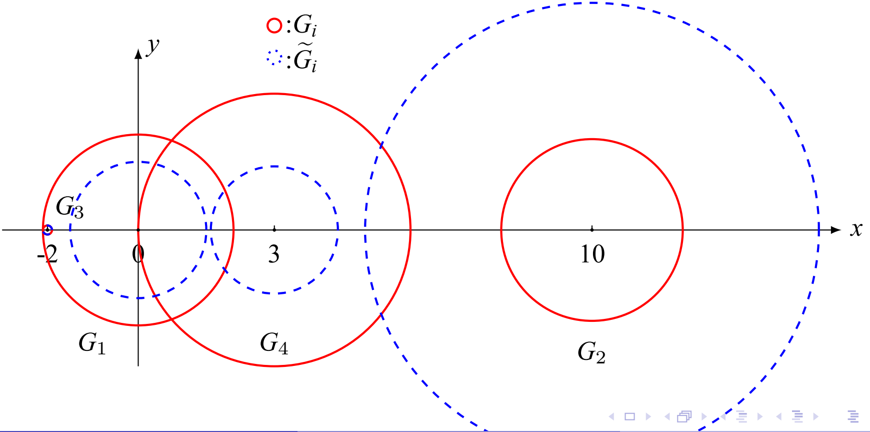
应用盖尔定理隔离矩阵 $A = \begin{bmatrix} 0 & 1 & 0.1 & 1 \\ 1 & 10 & 0 & 1 \\ 0.1 & 0 & -2 & 0 \\ 1 & 1 & 0 & 3 \end{bmatrix}$ 的特征值 (要求画图表示), 并利用实矩阵特征值的性质改进所得结果.

A 的盖尔圆为 $G_1 : |z| \leqslant 2.1,\ G_2 : |z-10| \leqslant 2,\ G_3 : |z+2| \leqslant 0.1,\ G_4 : |z-3| \leqslant 3.$



取 $D=\text{diag}\{1,0.4,1,1\}$, 则 $B=D^{-1}AD=\begin{bmatrix} 0 & 0.4 & 0.1 & 1 \\ 2.5 & 10 & 0 & 2.5 \\ 0.1 & 0 & -2 & 0 \\ 1 & 0.4 & 0 & 3 \end{bmatrix}$, B 的盖尔圆

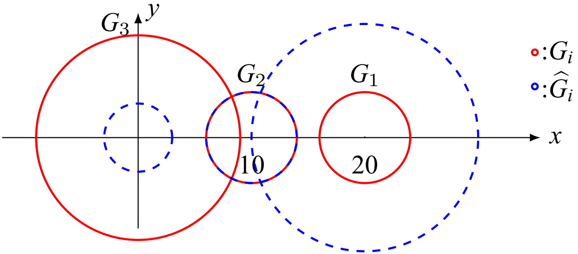
$\widetilde{G}_1: |z| \leqslant 1.5, \widetilde{G}_2: |z-10| \leqslant 5, \widetilde{G}_3: |z+2| \leqslant 0.1, \widetilde{G}_4: |z-3| \leqslant 1.4$, 故在 $\widetilde{G_1}, G_2, G_3, \widetilde{G}_4$ 中各有 1 个特征值, 于是在 $[-1.5, 1.5]$,$[8,12]$,$[-2.1, -1.9]$, $[1.6, 4.4]$ 中各有 1 个特征值.



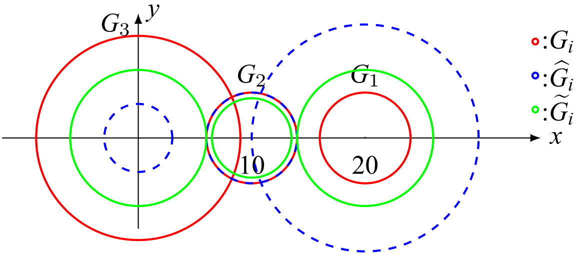
例 9.10

试分离矩阵 $A=\begin{bmatrix}20 & 3 & 1\\ 2 & 10 & 2\\ 8 & 1 & 0\end{bmatrix}$ 的特征值, 并在平面上画图.

A 的三个盖尔圆为 $G_1:|z-20|\leqslant 4,\ G_2:|z-10|\leqslant 4,\ G_3:|z|\leqslant 9.$



应取 $d_1<1, d_2=d_3=1$, 但 G_1, G_2 太近, 无法分开. 考虑 $A^T=\begin{bmatrix}20 & 2 & 8\\ 3 & 10 & 1\\ 1 & 2 & 0\end{bmatrix}$, 此时盖尔圆分别为 $\widehat{G}_1:|z-20|\leqslant 10, \widehat{G}_2:|z-10|\leqslant 4, \widehat{G}_3:|z|\leqslant 3.$



取 $D = \mathrm{diag}\{1, 1, 0.5\}$, 则 $B = D^{-1}AD = \begin{bmatrix} 20 & 2 & 4 \\ 3 & 10 & 0.5 \\ 2 & 4 & 0 \end{bmatrix}$, B 的三个盖尔圆

$\widetilde{G}_1 : |z-20| \leqslant 6, \widetilde{G}_2 : |z-10| \leqslant 3.5, \widetilde{G}_3 : |z| \leqslant 6$, 它们已经分离, 故在 $G_1, \widetilde{G_2}, \widehat{G}_3$ 中各有 1 个特征值.

Ostrowski 定理 *

定理 9.18 (Ostrowski I)

设 $A=(a_{ij})\in\mathbb{C}^{n\times n}$, $0\leqslant\alpha\leqslant 1$, 则 A 的所有特征值包含在 n 个圆盘

$$O_k(A)=\left\{z \mid |z-a_{kk}|\leqslant R_k^{\alpha}\widetilde{R}_k^{1-\alpha}\right\}\quad (k=1,2,\cdots,n)$$

的并集之中, 其中

$$R_k=\sum_{j=1,j\neq k}^{n}|a_{kj}|,\ \widetilde{R}_k=\sum_{j=1,j\neq k}^{n}|a_{jk}|\quad (k=1,2,\cdots,n)$$

注记 9.19

当 $\alpha=1$, 即为盖尔圆定理, 当 $\alpha=0$, 即为列盖尔圆定理, 适当选择 α 可以估计的更精确.

改进的结果

引理 9.20

设 σ, τ 是任意两个非负实数, $0 \leqslant \alpha \leqslant 1$, 则有

$$\tau^{\alpha}\sigma^{1-\alpha} \leqslant \alpha\tau + (1-\alpha)\sigma.$$

推论 9.21

设 $A = (a_{ij}) \in \mathbb{C}^{n\times n}$, $0 \leqslant \alpha \leqslant 1$, 则 A 的所有特征值包含在 n 个圆盘

$$\widetilde{O}_k(A) = \left\{ z \mid |z - a_{kk}| \leqslant \alpha R_k + (1-\alpha)\widetilde{R}_k \right\} \quad (k = 1, 2, \cdots, n)$$

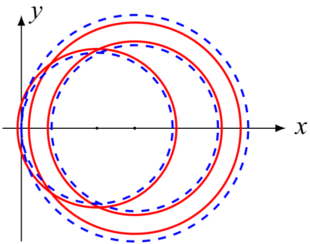
的并集之中, $R_k, \widetilde{R}_k$ 同上.

例 9.11

试估计矩阵 $A = \begin{bmatrix} 2 & 1.1 & 1 \\ -0.8 & 3 & 2 \\ 1.2 & 1.1 & 3 \end{bmatrix}$ 的特征值分布区域.

A 的盖尔圆为 $G_1 : |z-2| \leqslant 2.1,\ G_2 : |z-3| \leqslant 2.8,\ G_3 : |z-3| \leqslant 2.3.$

A 的列盖尔圆为 $\widetilde{G}_1 : |z-2| \leqslant 2,\ \widetilde{G}_2 : |z-3| \leqslant 2.2,\ \widetilde{G}_3 : |z-3| \leqslant 3.$



而 A 的 3 个 Ostrowski 圆为 (取 $\alpha = 1/2$):

$$O_1 : |z-2| \leqslant \sqrt{2.1} \cdot \sqrt{2} \approx 2.049,$$
$$O_2 : |z-3| \leqslant \sqrt{2.8} \cdot \sqrt{2.2} \approx 2.48,$$
$$O_3 : |z-3| \leqslant \sqrt{2.3} \cdot \sqrt{3} \approx 2.63.$$

利用推论9.21可得 A 的特征值在 3 个圆盘 (取 $\alpha = 1/2$):

$$\widetilde{O}_1 : |z-2| \leqslant \tfrac{1}{2} \cdot 2.1 + (1 - \tfrac{1}{2}) \cdot 2 = 2.05,$$
$$\widetilde{O}_2 : |z-3| \leqslant \tfrac{1}{2} \cdot 2.8 + (1 - \tfrac{1}{2}) \cdot 2.2 = 2.5,$$
$$\widetilde{O}_3 : |z-3| \leqslant \tfrac{1}{2} \cdot 2.3 + (1 - \tfrac{1}{2}) \cdot 3 = 2.65.$$

定理 9.22 (Ostrowski II)

设 $A=(a_{ij})\in\mathbb{C}^{n\times n}$, 则 A 的所有特征值包含在 $\dfrac{n(n-1)}{2}$ 个 Cassini 卵形

$$O_{ij}(A)=\left\{z \mid |z-a_{ii}||z-a_{jj}|\leqslant R_iR_j\right\}\ (i<j;i,j=1,2,\cdots,n)$$

的并集之中. (称$O_{ij}(A)$ 为 A 的 Cassini 卵形.)

例 9.12

设 $A=(a_{ij})\in\mathbb{C}^{n\times n}(n\geqslant 2)$, 则 $\cup_{i<j}O_{ij}\subset\cup_{i=1}^{n}G_i$.

例 9.13

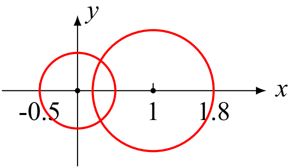
试估计矩阵 $A=\begin{bmatrix}1 & -0.8\\ 0.5 & 0\end{bmatrix}$ 的特征值分布区域.

A 的两个盖尔圆为 $G_1:|z-1|\leqslant 0.8,\ G_2:|z|\leqslant 0.5$; A 的一个 Cassini 卵形为: $O_{12}:|z-1||z|\leqslant 0.8\times 0.5=0.4.$

令 $z=x+\mathrm{i}y$, 由 $|z-1||z|=0.4$ 得 Cassini 卵形线的标准方程

$$[(x-\tfrac{1}{2})^2+y^2]-2(\tfrac{1}{2})^2[(x-\tfrac{1}{2})^2-y^2]=(0.63)^4-(\tfrac{1}{2})^4$$

(中心在 0.5, 左端点 $(-0.3,0)$, 右端点 $(1.3,0)$, 最高点 $(0.2,\pm 0.4),(0.8,\pm 0.4)$, 最低点 $(0.5,\pm 0.38)$.)



例 9.14

若 $A=(a_{ij})\in\mathbb{C}^{n\times n}(n\geqslant 2)$ 按行广义严格对角占优, 即

$$|a_{ii}||a_{jj}|>R_iR_j(i<j,i,j=1,2,\cdots,n),$$

则 $\det A\neq 0$.

§9.3 Hermite 矩阵特征值的表示

定义 9.23

设 $A=(a_{ij})\in\mathbb{C}^{n\times n}$ 为 Hermite 矩阵, $0\neq x\in\mathbb{C}^n$, 称

$$R(x)=\frac{x^HAx}{x^Hx}$$

为矩阵 A 的 Rayleigh 商.

设 $A=(a_{ij})\in\mathbb{C}^{n\times n}$ 为 Hermite 矩阵, 其特征值都是实数, 且从小到大排列为

$$\lambda\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n.$$

令 $\Lambda=\mathrm{diag}\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$, 则存在酉矩阵 Q 使得

$$Q^HAQ=\Lambda.$$

将 Q 按列分块为 $q=(q_1,q_2,\cdots,q_n)$, 则 $Aq_j=\lambda_jq_j$, 且 $q_1,q_2,\cdots,q_n$ 是 A 两两正交的单位特征向量.

定理 9.24

设 $A=(a_{ij})\in\mathbb{C}^{n\times n}$ 为 Hermite 矩阵, 则其最大最小特征值为

$$\lambda_1=\max_{x\neq 0}R(x),\qquad \lambda_n=\max_{x\neq 0}R(x).$$

注记 9.25

Hermite 矩阵 A 的最大特征值与最小特征值就是它的 Rayleigh 商 $R(x)$ 在 $\mathbb{C}^n$ 上的最大值与最小值.

问题 9.26

其他的特征值是不是也可以转化为极值问题呢?

回答是肯定的.

如果对应于 λ_1, λ_n 的单位特征向量 q_1, q_n 已经求出, 那么齐次线性方程组 $q_1^H x = 0$ 与 $q_n^H x = 0$ 的解都可以分别被 $\{q_2, q_3, \cdots, q_n\}$ 与 $\{q_1, a_2, \cdots, q_{n-1}\}$ 线性组合出来. 记

$$
\begin{aligned}
&\operatorname{span}\{q_2, q_3, \cdots, q_n\} \\
&\qquad = \{k_2 q_2 + k_3 q_3 + \cdots + k_n q_n \mid k_2, k_3, \cdots, k_n \in \mathbb{C}\} \\
&\operatorname{span}\{q_1, q_2, \cdots, q_{n-1}\} \\
&\qquad = \{l_1 q_1 + l_2 q_2 + \cdots + l_{n-1} q_{n-1} \mid l_1, l_2, \cdots, l_{n-1} \in \mathbb{C}\}
\end{aligned}
$$

定理 9.27

设 $A=(a_{ij})\in\mathbb{C}^{n\times n}$ 为 Hermite 矩阵, 其特征值如前所述, 则

$$\begin{aligned}\lambda_2 &= \max\{R(x)\mid x\neq 0, q_1^H x=0\}\\ &= \max\{R(x)\mid x\neq 0, x\in\operatorname{span}\{q_2,q_3,\cdots,q_n\}\\ \lambda_{n-1} &= \min\{R(x)\mid x\neq 0, q_n^H x=0\}\\ &= \min\{R(x)\mid x\neq 0, x\in\operatorname{span}\{q_1,q_2,\cdots,q_{n-1}\}.\end{aligned}$$

更一般地, 如果 $q_1,\cdots,q_{s-1}$ 与 $q_{r+1},\cdots,q_n$ 已经求出来, 构造矩阵

$$P_1=(q_1,\cdots,q_{s-1}),\quad P_2=(q_{r+1},\cdots,q_n).$$

定理 9.28

设 $A = (a_{ij}) \in \mathbb{C}^{n\times n}$ 为 Hermite 矩阵, 其特征值都是实数, 且从小到大排列为

$$\lambda \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n.$$

则有

$$\lambda_s = \min_{P_1 \in \mathbb{C}^{n\times(s-1)}} \max\{R(x) \mid x \neq 0, P_1^H x = 0\} \cdots\cdots (1)$$

$$\lambda_r = \max_{P_2 \in \mathbb{C}^{n\times(n-r)}} \min\{R(x) \mid x \neq 0, P_2^H x = 0\} \cdots\cdots (2),$$

其中, $1 < s \leqslant n, 1 \leqslant r < n$.

式子 (1) 称为特征值的极大极小原理, (2) 称为特征值的极大极小原理.