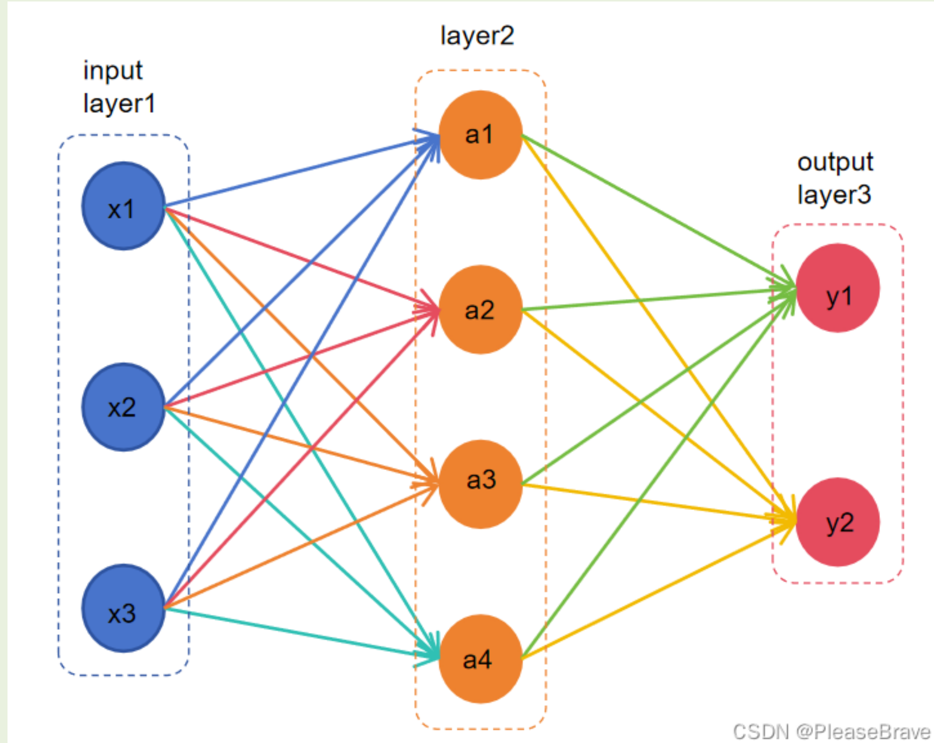
第六章

矩阵函数

引导问题

神经网络及反向传播



$$\begin{bmatrix} a_1^{[1]} \\ a_2^{[1]} \\ a_3^{[1]} \\ a_4^{[1]} \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{32} \\ w_{41} & w_{42} & w_{43} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \Rightarrow \boldsymbol{a}^{[1]} = W\boldsymbol{x} \Rightarrow \frac{\partial \boldsymbol{a}^{[1]}}{\partial W} ?$$

已知

提纲

- 1. 矩阵序列
- 2. 矩阵级数
- 3. 常见矩阵函数
- 4. 矩阵微积分
- 5. 广义逆矩阵
- 6. 广义特征值

【定义】 对于n阶方阵序列 $\{A_k\}$, 若

$$\lim_{k\to\infty}\|A_k - A\| = 0$$

则称**方阵序列$\{A_k\}$收敛于n阶方阵A.**

- 通常表示成

$$\lim_{k\to\infty} A_k = A$$

➢ 矩阵序列收敛的充分必要条件

【定理】 设$A_k=\left[a_{ij}^{(k)}\right]\in C^{n\times n}$，$n$阶方阵序列 $\{A_k\}$收敛于n阶方阵$A=[a_{ij}]$的充要条件是

$$\lim_{k\to\infty} a_{ij}^{(k)}=a_{ij},\quad (i,j=1:n)$$

- 矩阵序列的收敛可归结为对应**元素序列的收敛**.

矩阵序列极限的性质

(1) 一个收敛矩阵序列的极限是唯一的

(2) 设 $\lim\limits_{k\to\infty} A_k = A,\ \lim\limits_{k\to\infty} B_k = B,$ 则

$$\lim_{k\to\infty}(aA_k + bB_k) = aA + bB \qquad a, b \in C$$

(3) 设 $\lim\limits_{k\to\infty} A_k = A,\ \lim\limits_{k\to\infty} B_k = B,$ 其中$A, B \in C^{n\times n}$则

$$\lim_{k\to\infty}(A_k B_k) = AB$$

(4) 设 $\lim\limits_{k\to\infty} A_k = A,\ P, Q \in C^{n\times n},$则

$$\lim_{k\to\infty}(PA_kQ) = PAQ$$

(5) 设 $\lim\limits_{k\to\infty} A_k = A,\ A_k, A$均可逆,则$\{A_k^{-1}\}$也收敛，且

$$\lim_{k\to\infty} A_k^{-1} = A^{-1}$$

【定理】 设A是任意n阶方阵, 由A的各次幂所组成的矩阵序列

$$I, A, A^2, \ldots, A^k, \ldots$$

若对矩阵A的某一种范数$\|A\|<1$, 则

$$\boxed{\lim_{k\to\infty} A^k = 0}$$

收敛矩阵

直接根据收敛的定义证明

$$\boxed{\left\|A^k - 0\right\| \le \left\|A\right\|^k \to 0 \;\; (k\to\infty)}$$

$$D=\begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \mathbf{0} \\ & & \ddots & \ddots & \\ & \mathbf{0} & & 0 & 1 \\ & & & & 0 \end{bmatrix}_n$$

求D^k?

D的幂次每增加1, 主对角线上方一排1就向右上方平移1排,

$$D^2=\begin{bmatrix} 0 & 0 & 1 & & \mathbf{0} \\ & 0 & 0 & \ddots & \\ & & \ddots & \ddots & 1 \\ & \mathbf{0} & & 0 & 0 \\ & & & & 0 \end{bmatrix}_n,$$

$$D^{n-1}=\begin{bmatrix} & & 1 \\ & \mathbf{0} & \\ & & \end{bmatrix}_n$$

$$D^k=0,\quad (k\ge n).$$

设 $J = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & \mathbf{0} & \\ & & \ddots & \ddots & \\ & \mathbf{0} & & \lambda & 1 \\ & & & & \lambda \end{bmatrix}_n$ 则 $\lim\limits_{k\to\infty} J^k = 0 \Leftrightarrow |\lambda| < 1$

$$J = \lambda I + D$$

$$J^k = (\lambda I + D)^k$$
$$= \lambda^k I + C_k^1 \lambda^{k-1} D + C_k^2 \lambda^{k-2} D^2 + \cdots + C_k^{n-1} \lambda^{k-n+1} D^{n-1}$$

$$J^k = \begin{bmatrix} \lambda^k & C_k^1\lambda^{k-1} & C_k^2\lambda^{k-2} & \cdots & C_k^{n-1}\lambda^{k-n+1} \\ & \lambda^k & C_k^1\lambda^{k-1} & \ddots & \vdots \\ & & \ddots & \ddots & C_k^2\lambda^{k-2} \\ & \mathbf{0} & & \lambda^k & C_k^1\lambda^{k-1} \\ & & & & \lambda^k \end{bmatrix}_n$$

$$\begin{cases} C_k^l = \dfrac{k(k-1)\cdots(k-l+1)}{l!} & (l \le k) \\ C_k^l = 0 & (l > k) \end{cases}$$

$$\boxed{\lim_{k\to\infty} J^k = 0} \Leftrightarrow \lim_{k\to\infty} \lambda^k = 0 \boxed{\Leftrightarrow |\lambda| < 1}$$

【定理】： 设A是任意n阶方阵，

$$\boxed{\lim_{k\to\infty} A^k = 0 \Leftrightarrow \rho(A) < 1}$$

证明 设J为A的Jordan标准形，即

$$A = PJP^{-1}, \quad J = \mathrm{diag}\big(J_1(\lambda_1), J_2(\lambda_2), \cdots, J_r(\lambda_r)\big)$$

则 $A^k = P\mathrm{diag}\big(J_1^k(\lambda_1), J_2^k(\lambda_2), \cdots, J_r^k(\lambda_r)\big)P^{-1}$

$$\lim_{k\to\infty} A^k = 0 \Leftrightarrow \lim_{k\to\infty} J_i^k(\lambda_i) = 0, (i = 1:r)$$

$$\Leftrightarrow |\lambda_i| < 1, (i = 1:r)$$

因此 $\boxed{\lim_{k\to\infty} A^k = 0 \Leftrightarrow \rho(A) < 1}$

【例】：已知 $A=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$ 求 $\lim\limits_{k\to\infty} A^k$

$$A^k=\begin{bmatrix}1 & k\\ 0 & 1\end{bmatrix}$$

发散

【例】：已知 $A=\begin{bmatrix}1 & 0 & 0\\ 0 & 0.9 & 1\\ 0 & 0 & 0.9\end{bmatrix}$ 求 $\lim\limits_{k\to\infty} A^k =$

令$A=\begin{bmatrix}A_1 & 0\\ 0 & A_2\end{bmatrix}$

$A_1=[1]$，$A_2=\begin{bmatrix}0.9 & 1\\ 0 & 0.9\end{bmatrix}$

$$\lim_{k\to\infty} A^k =\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}$$

【例】：已知$A=\frac{1}{5}\begin{bmatrix}1&2&2\\2&1&2\\2&2&1\end{bmatrix}$ 求$\lim\limits_{k\to\infty}A^k=$

$$A=P\begin{bmatrix}1&0&0\\0&-1/5&0\\0&0&-1/5\end{bmatrix}P^{-1},\quad P=\begin{bmatrix}1&-1&-1\\1&-1&0\\1&2&1\end{bmatrix}$$

$$P^{-1}=\frac{1}{3}\begin{bmatrix}1&1&1\\1&-2&1\\-3&3&0\end{bmatrix},\quad A^k=P\begin{bmatrix}1&0&0\\0&(-0.2)^k&0\\0&0&(-0.2)^k\end{bmatrix}P^{-1}$$

$$\lim_{k\to\infty}A^k=P\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix}P^{-1}=\frac{1}{3}\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}$$

提纲

- 1. 矩阵序列
- 2. 矩阵级数
- 3. 常见矩阵函数
- 4. 矩阵微积分
- 5. 广义逆矩阵
- 6. 广义特征值

【定义】：设 $A_k=[a_{ij}^{(k)}]\in C^{n\times n}\ (k=1:\infty)$,

若 $\sum_{k=1}^{\infty} a_{ij}^{(k)},\ (i,j=1:n)$ 都收敛, 则称矩阵级数

$$\sum_{k=1}^{\infty} A_k = A_1 + A_2 + \cdots + A_k + \cdots$$ 收敛.

若 $\sum_{k=1}^{\infty} a_{ij}^{(k)},\ (i,j=1:n)$ 都绝对收敛, 则称矩阵级数

$$\sum_{k=1}^{\infty} A_k = A_1 + A_2 + \cdots + A_k + \cdots$$ 绝对收敛

【定理】： 若级数 $\sum_{k=1}^{\infty} A_k$ 绝对收敛, 则一定是收敛的

$$\text{设} A_k = \left[a_{ij}^{(k)}\right], \quad \sum_{k=1}^{\infty} A_k \text{收敛} \Leftrightarrow \sum_{k=1}^{\infty} a_{ij}^{(k)} \text{收敛} \ (i, j = 1 : n)$$

$$\Uparrow$$

$$\sum_{k=1}^{\infty} A_k \text{绝对收敛} \Leftrightarrow \sum_{k=1}^{\infty} a_{ij}^{(k)} \text{绝对收敛} \ (i, j = 1 : n)$$

【例】:设
$$A_k=[a_{ij}^{(k)}]=\begin{bmatrix}\dfrac{1}{k(k+1)} & \dfrac{1}{k^3}\\ \dfrac{\pi}{2^k} & \sin\dfrac{\pi}{2^k}\end{bmatrix}\in C^{2\times 2},$$
讨论级数 $\sum_{k=1}^{n}A_k$ 的敛散性

$$\sum_{k=1}^{\infty}a_{11}^{(k)}=\sum_{k=1}^{\infty}\frac{1}{k(k+1)},$$
绝对收敛

$$\sum_{k=1}^{\infty}a_{12}^{(k)}=\sum_{k=1}^{\infty}\frac{1}{k^3},$$
绝对收敛

$$\sum_{k=1}^{\infty}a_{21}^{(k)}=\sum_{k=1}^{\infty}\frac{\pi}{2^k},$$
绝对收敛

$$\sum_{k=1}^{\infty}a_{22}^{(k)}=\sum_{k=1}^{\infty}\sin\frac{\pi}{2^k}.$$
绝对收敛

故矩阵级数 $\sum_{k=1}^{\infty}A_k=A_1+A_2+\cdots+A_k+\cdots$ 绝对收敛

【定理】： 设 $A_k=[a_{ij}^{(k)}]\in C^{n\times n}$，则矩阵级数

$$\sum_{k=1}^{\infty}A_k=A_1+A_2+\cdots+A_k+\cdots$$

绝对收敛的充分必要条件是正项级数

$$\sum_{k=1}^{\infty}\|A_k\|=\|A_1\|+\|A_2\|+\cdots+\|A_k\|+\cdots$$

收敛, 其中$\|A\|$ 为任意一种矩阵范数.

证明要点 $A_k=[a_{ij}^{(k)}]\in C^{n\times n}$，取矩阵范数 $\|A_k\|=\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}^{(k)}|$

$\sum_{k=1}^{\infty}A_k$ 绝对收敛 $\overset{\text{定义}}{\Longleftrightarrow}$ $\sum_{k=1}^{\infty}|a_{ij}^{(k)}|$ 收敛, $(i,j=1:n)$

$\Updownarrow$

$\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{\infty}|a_{ij}^{(k)}|$ 收敛,

$\Updownarrow$

$\sum_{k=1}^{\infty}\|A_k\|$ 收敛, $\Longleftrightarrow$ $\sum_{k=1}^{\infty}\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}^{(k)}|$ 收敛,

$\|A_k\|$

范数等价定理 ✓

【定义】(矩阵幂级数)设 $A=[a_{ij}]\in C^{n\times n}$，称形如

$$\sum_{k=0}^{\infty} c_k A^k = c_0 I + c_1 A + c_2 A^2 + \cdots + c_k A^k + \cdots$$

的矩阵级数为矩阵幂级数.

研究矩阵幂级数收敛的条件

- 对Jordan 块矩阵的幂级数给出收敛性条件
- 把一般方阵的幂级数收敛性转化为Jordan块矩阵幂级数的收敛性

定理

- 设幂级数 $f(x)=\sum_{k=0}^{\infty}c_k x^k$ 的收敛半径为 R,
- J 为 n 阶 Jordan 块.
- 若 $|\lambda|<R$, 则矩阵幂级数 $\sum_{k=0}^{\infty}c_k J^k$ 绝对收敛,
- 若 $|\lambda|>R$, 则矩阵幂级数 $\sum_{k=0}^{\infty}c_k J^k$ 发散.

$$J=\begin{bmatrix} \lambda & 1 & & & \mathbf{0} \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & \mathbf{0} & & \lambda & 1 \\ & & & & \lambda \end{bmatrix}$$

证

$$J^k=\begin{bmatrix}\lambda^k & C_k^1\lambda^{k-1} & C_k^2\lambda^{k-2} & \cdots & C_k^{n-1}\lambda^{k-n+1}\\ & \lambda^k & C_k^1\lambda^{k-1} & \ddots & \vdots\\ & & \ddots & \ddots & C_k^2\lambda^{k-2}\\ & \mathbf{0} & & \lambda^k & C_k^1\lambda^{k-1}\\ & & & & \lambda^k\end{bmatrix}_n$$

其中

$$\begin{cases}C_k^l=\dfrac{k(k-1)\cdots(k-l+1)}{l!} & (l\le k)\\ C_k^l=0 & (l>k)\end{cases}$$

故

$$\sum_{k=0}^{\infty} c_k J^k(\lambda) = \begin{bmatrix} \sum_{k=0}^{\infty} c_k \lambda^k & \sum_{k=0}^{\infty} c_k C_k^1 \lambda^{k-1} & \cdots & \sum_{k=0}^{\infty} c_k C_k^{n-1} \lambda^{k-n+1} \\ & \sum_{k=0}^{\infty} c_k \lambda^k & \ddots & \vdots \\ & & \ddots & \sum_{k=0}^{\infty} c_k C_k^1 \lambda^{k-1} \\ & \mathbf{0} & & \sum_{k=0}^{\infty} c_k \lambda^k \end{bmatrix}_n$$

幂级数 $f(x) = \sum_{k=0}^{\infty} c_k x^k$ 的收敛半径为R

当$|\lambda| < R \Rightarrow \sum_{k=0}^{\infty} c_k \lambda^k, \ \sum_{k=0}^{\infty} c_k C_k^1 \lambda^{k-1}, \quad \ldots, \ \sum_{k=0}^{\infty} c_k C_k^{n-1} \lambda^{k-n+1}$ 均

绝对收敛 $\Rightarrow \sum_{k=0}^{\infty} c_k J^k$ 绝对收敛.

故

$$\sum_{k=0}^{\infty} c_k J^k(\lambda) = \begin{bmatrix} \sum_{k=0}^{\infty} c_k \lambda^k & \sum_{k=0}^{\infty} c_k C_k^1 \lambda^{k-1} & \cdots & \sum_{k=0}^{\infty} c_k C_k^{n-1} \lambda^{k-n+1} \\ & \sum_{k=0}^{\infty} c_k \lambda^k & \ddots & \vdots \\ & & \ddots & \sum_{k=0}^{\infty} c_k C_k^1 \lambda^{k-1} \\ & \mathbf{0} & & \sum_{k=0}^{\infty} c_k \lambda^k \end{bmatrix}_n$$

幂级数 $f(x) = \sum_{k=0}^{\infty} c_k x^k$ 的收敛半径为R

当$|\lambda| > R$时，$\sum_{k=0}^{\infty} c_k \lambda^k$发散， 因此$\sum_{k=0}^{\infty} c_k J^k$发散.

矩阵函数

$$f(J)=\sum_{k=0}^{\infty} c_k J^k=\begin{bmatrix} \sum_{k=0}^{\infty} c_k\lambda^k & \sum_{k=0}^{\infty} c_k C_k^1\lambda^{k-1} & \cdots & \sum_{k=0}^{\infty} c_k C_k^{n-1}\lambda^{k-n+1} \\ & \sum_{k=0}^{\infty} c_k\lambda^k & \ddots & \vdots \\ & & \ddots & \sum_{k=0}^{\infty} c_k C_k^1\lambda^{k-1} \\ & & & \sum_{k=0}^{\infty} c_k\lambda^k \end{bmatrix}_n$$

$$=\begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} & \cdots & \frac{f^{(n-1)}(\lambda)}{(n-1)!} \\ & f(\lambda) & f'(\lambda) & \ddots & \vdots \\ & & \ddots & \ddots & \frac{f''(\lambda)}{2!} \\ & & & & f'(\lambda) \\ & & & & f(\lambda) \end{bmatrix}$$

【定理】 设P, Q为非奇异矩阵，则矩阵级数

$$\sum_{k=1}^{\infty} A_k \text{收敛(或绝对收敛)} \Leftrightarrow \sum_{k=1}^{\infty} PA_kQ \text{收敛(或绝对收敛)}$$

并且 $$\sum_{k=1}^{\infty} PA_kQ = P\left(\sum_{k=1}^{\infty} A_k\right)Q$$

有了Jordan块矩阵幂级数收敛性的结论，
可把对一般方阵幂级数的收敛性讨论转化
为对它的Jordan标准形矩阵的收敛性讨论

定理

- 设幂级数 $f(x)=\sum_{k=0}^{\infty}c_k x^k$ 的收敛半径为 R,
- A 为 n 阶矩阵.
- 若 $\rho(A)<R$，则矩阵幂级数 $\sum_{k=0}^{\infty}c_k A^k$ 绝对收敛;
- 若 $\rho(A)>R$，则矩阵幂级数 $\sum_{k=0}^{\infty}c_k A^k$ 发散.

总原则：把关于矩阵A的幂级数的收敛性问题转化为关于Jordan块的幂级数的收敛性问题

定理给出了A的矩阵幂级数绝对收敛与发散的充分条件,

A的谱半径<幂级数的收敛半径，绝对收敛

A的谱半径>幂级数的收敛半径，发散

当A的谱半径与幂级数的收敛半径相等时,

通过A的Jordan标准形幂级数的敛散性来得到A的幂级数的敛散性.

定理：矩阵幂级数

$$I+A+A^2+\ldots+A^n+\ldots$$

绝对收敛的充分必要条件是$\rho(A)<1$. 且其和为

$$(I-A)^{-1}=I+A+A^2+A^3+\cdots+A^k+\cdots$$

令$A=PJP^{-1}$

$$I+A+A^2+\ldots+A^n+\ldots$$

$$=P\begin{bmatrix} \sum_{k=0}^{\infty}\lambda_1{}^k & \cdots & & \\ 0 & \sum_{k=0}^{\infty}\lambda_2{}^k & \cdots & \\ & 0 & \ddots & \\ & & 0 & \sum_{k=0}^{\infty}\lambda_m{}^k \end{bmatrix}P^{-1}$$

$$(I-A)(I+A+A^2+\ldots+A^n+\ldots)=I$$

定理 若$\|A\|<1$, 则$I\pm A$都是非奇异阵, 且

$$\frac{1}{1+\|A\|} \le \left\|(I \pm A)^{-1}\right\| \le \frac{1}{1-\|A\|}$$

其中范数$\|A\|$是矩阵A的算子范数.

$\|A\|<1$, ± 1都不是A的特征值，$|I\pm A|\neq 0$，

所以 $I\pm A$都是非奇异阵

$$I = (I - A)(I - A)^{-1}$$

$$1 = \|I\| = \left\|(I - A)(I - A)^{-1}\right\| \le \|I - A\|\left\|(I - A)^{-1}\right\|$$

$$\le (1+\|A\|)\left\|(I - A)^{-1}\right\|$$

$$\frac{1}{1+\|A\|} \le \left\|(I - A)^{-1}\right\|$$

定理 若$\|A\|<1$, 则$I\pm A$都是非奇异阵, 且

$$\frac{1}{1+\|A\|}\le\left\|(I\pm A)^{-1}\right\|\le\frac{1}{1-\|A\|}$$

其中范数$\|A\|$是矩阵A的算子范数.

$$(I-A)^{-1}=I+A+A^2+A^3+\cdots+A^k+\cdots$$

$$\left\|(I-A)^{-1}\right\|=\left\|I+A+A^2+A^3+\cdots+A^k+\cdots\right\|$$

$$\le 1+\|A\|+\|A\|^2+\|A\|^3+\cdots+\|A\|^k+\cdots$$

$$=\frac{1}{1-\|A\|}$$

例 (1) 求下面级数的收敛半径

$$\sum_{k=1}^{\infty}\frac{x^k}{k\cdot 2^k}=\frac{x}{1\cdot 2}+\frac{x^2}{2\cdot 2^2}+\cdots+\frac{x^k}{k\cdot 2^k}+\cdots$$

$R=2.$

(2) 设

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

判断矩阵幂级数 $\sum_{k=1}^{\infty}\frac{A^k}{k\cdot 2^k}$ 的敛散性.

解:

$$\lim_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right|=\lim_{k\to\infty}\frac{\frac{1}{(k+1)2^{k+1}}}{\frac{1}{k2^k}}=\lim_{k\to\infty}\frac{k}{2(k+1)}=\frac{1}{2}$$

此级数的收敛半径为2.

例 (1) 求下面级数的收敛半径

$$\sum_{k=1}^{\infty}\frac{x^k}{k\cdot 2^k}=\frac{x}{1\cdot 2}+\frac{x^2}{2\cdot 2^2}+\cdots+\frac{x^k}{k\cdot 2^k}+\cdots \quad R=2.$$

(2) 设

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

判断矩阵幂级数 $\sum_{k=1}^{\infty}\frac{A^k}{k\cdot 2^k}$ 的敛散性.

解 (2) $\rho(A)=1$,

矩阵幂级数 $$\sum_{k=1}^{\infty}\frac{A^k}{2^k\cdot k}=\frac{A}{2\cdot 1}+\frac{A^2}{2^2\cdot 2}+\frac{A^3}{2^3\cdot 3}+\cdots+\frac{A^k}{2^k\cdot k}+\cdots$$

绝对收敛

例 已知 $A=\begin{bmatrix}-2 & 1\\ -1 & 0\end{bmatrix}$

分别讨论矩阵幂级数 $\sum_{k=1}^{\infty}\frac{1}{k}A^k$ 与 $\sum_{k=1}^{\infty}\frac{1}{k^2}A^k$ 的敛散性.

幂级数 $\sum_{k=1}^{\infty}\frac{1}{k}x^k$ 与 $\sum_{k=1}^{\infty}\frac{1}{k^2}x^k$ 的收敛半径均为1,

矩阵A的谱半径也是1,

要通过A的Jordan标准形矩阵幂级数的敛散性来得到A的幂级数的敛散性

【例】：已知 $A=\begin{bmatrix}-2 & 1\\ -1 & 0\end{bmatrix}$

发散

分别讨论矩阵幂级数 $\sum_{k=1}^{\infty}\frac{1}{k}A^k$ 与 $\sum_{k=1}^{\infty}\frac{1}{k^2}A^k$ 的敛散性.

$$P^{-1}AP=J=\begin{bmatrix}-1 & 1\\ 0 & -1\end{bmatrix}\qquad J^k=\begin{bmatrix}(-1)^k & k(-1)^{k-1}\\ 0 & (-1)^k\end{bmatrix}$$

$$\sum_{k=0}^{\infty}\frac{1}{k}J^k=\begin{bmatrix}\sum_{k=0}^{\infty}(-1)^k/k & \sum_{k=0}^{\infty}(-1)^{k-1}\\ 0 & \sum_{k=0}^{\infty}(-1)^k/k\end{bmatrix}$$

【例】 已知 $A=\begin{bmatrix}-2 & 1\\ -1 & 0\end{bmatrix}$ $P^{-1}AP=J=\begin{bmatrix}-1 & 1\\ 0 & -1\end{bmatrix}$

分别讨论矩阵幂级数 $\sum_{k=1}^{\infty}\frac{1}{k}A^k$ 与 $\sum_{k=1}^{\infty}\frac{1}{k^2}A^k$ 的敛散性.

发散

收敛

条件收敛

$$J^k=\begin{bmatrix}(-1)^k & k(-1)^{k-1}\\ 0 & (-1)^k\end{bmatrix}$$

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\ldots=\ln 2$$

$$\sum_{k=0}^{\infty}\frac{1}{k^2}J^k=\begin{bmatrix}\sum_{k=0}^{\infty}(-1)^k/k^2 & \sum_{k=0}^{\infty}(-1)^{k-1}/k\\ 0 & \sum_{k=0}^{\infty}(-1)^k/k^2\end{bmatrix}$$

$$1-\frac{1}{2^2}+\frac{1}{3^2}-\frac{1}{4^2}+\ldots=\frac{\pi^2}{12}$$

提纲

- 1. 矩阵序列
- 2. 矩阵级数
- **3. 常见矩阵函数**
- 4. 矩阵微积分
- 5. 广义逆矩阵
- 6. 广义特征值

【定义】：设$A\in C^{n\times n}$，一元函数$f(x)$能够展开成关于x的幂级数

$$f(x)=\sum_{k=0}^{\infty}c_k x^k$$

并且该幂级数的收敛半径为R. 当矩阵A 的谱半径$\rho(A)<R$ 时，将收敛矩阵幂级数

$$\sum_{k=0}^{\infty}c_k A^k$$

的和定义为矩阵函数 $f(A)$，即

$$f(A)=\sum_{k=0}^{\infty}c_k A^k$$

当$|x|<+\infty$ 时，

$$e^x = 1 + x + \frac{1}{2!}x^2 + \cdots + \frac{1}{k!}x^k + \cdots$$

$$\sin x = x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \cdots + \frac{(-1)^k}{(2k+1)!}x^{2k+1} + \cdots$$

$$\cos x = 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 - \cdots + (-1)^n \frac{1}{(2k)!}x^{2k} + \cdots$$

当$|x|<1$ 时，

$$(1-x)^{-1} = 1 + x + x^2 + x^3 + \cdots + x^k + \cdots$$

当$-1<x \le 1$ 时，

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots + \frac{(-1)^{k-1}}{k}x^k + \cdots$$

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

对任意的矩阵$A\in C^{n\times n}$有

$$e^{A}=I+A+\frac{1}{2!}A^{2}+\cdots+\frac{1}{k!}A^{k}+\cdots$$

$$\sin A=A-\frac{1}{3!}A^{3}+\frac{1}{5!}A^{5}-\cdots+\frac{(-1)^{k}}{(2k+1)!}A^{2k+1}+\cdots$$

$$\cos A=I-\frac{1}{2!}A^{2}+\frac{1}{4!}A^{4}-\cdots+\frac{(-1)^{k}}{(2k)!}A^{2k}+\cdots$$

当$\rho(A)<1$时，有

$$(I-A)^{-1}=I+A+A^{2}+A^{3}+\cdots+A^{k}+\cdots$$

$$\ln(I+A)=A-\frac{A^{2}}{2}+\frac{A^{3}}{3}-\cdots+\frac{(-1)^{k-1}}{k}A^{k}+\cdots$$

【例】： 已知 $A=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$，求$e^{At}$.

$$e^{At}=I+tA+\frac{t^2}{2!}A^2+\frac{t^3}{3!}A^3+\cdots$$

A的特征方程为 $\lambda^2+1=0$

由Cayley-Hamilton定理知 $A^2+I=0$

$$A^2=-I,\quad A^3=-A,\quad A^4=I,\quad A^5=A,\cdots$$

从而有

$$e^{At}=I+tA-\frac{t^2}{2!}I-\frac{t^3}{3!}A+\frac{t^4}{4!}I+\frac{t^5}{5!}A-\frac{t^6}{6!}I-\frac{t^7}{7!}A+\cdots$$

【例】：已知 $A=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$，求 e^{At}.

$$e^{At}=I(1-\frac{t^2}{2!}+\frac{t^4}{4!}-\cdots)+A(t-\frac{t^3}{3!}+\frac{t^5}{5!}-\cdots)$$

$$=I\cos t+A\sin t=\begin{bmatrix}\cos t & \sin t\\ -\sin t & \cos t\end{bmatrix}$$

$$e^{At}=I+tA-\frac{t^2}{2!}I-\frac{t^3}{3!}A+\frac{t^4}{4!}I+\frac{t^5}{5!}A-\frac{t^6}{6!}I-\frac{t^7}{7!}A+\cdots$$

矩阵函数$f(A)$的Jordan表示式

设A为n阶方阵，由Jordan定理，存在可逆阵P，使得

$$P^{-1}AP = J = \mathrm{diag}\left[J_1(\lambda_1), J_2(\lambda_2), \cdots, J_r(\lambda_r)\right]$$

即

$$\boxed{A = PJP^{-1}}$$

当$f(x)$具有足够多阶的导数值，使得

$$f(\lambda_i), f'(\lambda_i), f''(\lambda_i), \cdots, f^{(d_i-1)}(\lambda_i) \quad (i = 1:r)$$

有意义. 定义

$$\boxed{f(A) = Pf(J)P^{-1}}$$

$$f(J)=f\left(\begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_r \end{bmatrix}\right)=\begin{bmatrix} f(J_1) & & & \\ & f(J_2) & & \\ & & \ddots & \\ & & & f(J_r) \end{bmatrix}$$

其中

$$f(J_i)=\begin{bmatrix} f(\lambda_i) & f'(\lambda_i) & \cdots & \dfrac{f^{(d_i-1)}(\lambda_i)}{(d_i-1)!} \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix}_{d_i\times d_i}$$

例:设
$$A=\begin{bmatrix}17 & 0 & -25\\ 0 & 1 & 0\\ 9 & 0 & -13\end{bmatrix}$$
计算e^A, e^{tA}, $\cos(A)$, $\sin(\frac{3\pi}{2}A)$

$$A=PJP^{-1},\quad J=\begin{bmatrix}1 & 0 & 0\\ 0 & 2 & 1\\ 0 & 0 & 2\end{bmatrix},\quad P=\begin{bmatrix}0 & 5 & 2\\ 1 & 0 & 0\\ 0 & 3 & 1\end{bmatrix},$$

$$P^{-1}=\begin{bmatrix}0 & 1 & 0\\ -1 & 0 & 2\\ 3 & 0 & -5\end{bmatrix},\quad f(J)=\begin{bmatrix}f(1) & 0 & 0\\ 0 & f(2) & f'(2)\\ 0 & 0 & f(2)\end{bmatrix}$$

$$f(A)=Pf(J)P^{-1}$$

$$f(A)=Pf(J)P^{-1}$$

e^A, e^{tA}, $\cos(A)$, $\sin(\frac{3\pi}{2}A)$

$$=\begin{bmatrix}0&5&2\\1&0&0\\0&3&1\end{bmatrix}\begin{bmatrix}f(1)&0&0\\0&f(2)&f'(2)\\0&0&f(2)\end{bmatrix}\begin{bmatrix}0&1&0\\-1&0&2\\3&0&-5\end{bmatrix},$$

$$f(x)=e^x,\ f'(x)=e^x$$

$$e^A=\begin{bmatrix}0&5&2\\1&0&0\\0&3&1\end{bmatrix}\begin{bmatrix}e&0&0\\0&e^2&e^2\\0&0&e^2\end{bmatrix}\begin{bmatrix}0&1&0\\-1&0&2\\3&0&-5\end{bmatrix},$$

$$=\begin{bmatrix}16e^2&0&-25e^2\\0&e&0\\9e^2&0&-14e^2\end{bmatrix}$$

$$f(A)=Pf(J)P^{-1} \qquad e^{A},\ e^{tA},\ \cos(A),\ \sin(\frac{3\pi}{2}A)$$

$$=\begin{bmatrix}0&5&2\\1&0&0\\0&3&1\end{bmatrix}\begin{bmatrix}f(1)&0&0\\0&f(2)&f'(2)\\0&0&f(2)\end{bmatrix}\begin{bmatrix}0&1&0\\-1&0&2\\3&0&-5\end{bmatrix},$$

$$f(x)=e^{xt},\ f'(x)=te^{xt}$$

$$e^{At}=\begin{bmatrix}0&5&2\\1&0&0\\0&3&1\end{bmatrix}\begin{bmatrix}e^{t}&0&0\\0&e^{2t}&te^{2t}\\0&0&e^{2t}\end{bmatrix}\begin{bmatrix}0&1&0\\-1&0&2\\3&0&-5\end{bmatrix},$$

$$=\begin{bmatrix}(1+15t)e^{2t}&0&-25te^{2t}\\0&e^{t}&0\\9te^{2t}&0&(1-15t)e^{2t}\end{bmatrix}$$

$$f(A)=Pf(J)P^{-1} \qquad e^{A},\ e^{tA},\ \cos(A),\ \sin(\frac{3\pi}{2}A)$$

$$=\begin{bmatrix}0&5&2\\1&0&0\\0&3&1\end{bmatrix}\begin{bmatrix}f(1)&0&0\\0&f(2)&f'(2)\\0&0&f(2)\end{bmatrix}\begin{bmatrix}0&1&0\\-1&0&2\\3&0&-5\end{bmatrix},$$

$$f(x)=\cos(x),\ f'(x)=-\sin(x)$$

$$\cos(A)=\begin{bmatrix}0&5&2\\1&0&0\\0&3&1\end{bmatrix}\begin{bmatrix}\cos 1&0&0\\0&\cos 2&-\sin 2\\0&0&\cos 2\end{bmatrix}\begin{bmatrix}0&1&0\\-1&0&2\\3&0&-5\end{bmatrix},$$

$$=\begin{bmatrix}\cos 2-15\sin 2&0&25\sin 2\\0&\cos 1&0\\-9\sin 2&0&\cos 2+15\sin 2\end{bmatrix}$$

$$f(A)=Pf(J)P^{-1}$$

e^{A}, e^{tA}, $\cos(A)$, $\sin(\frac{3\pi}{2}A)$

$$=\begin{bmatrix}0&5&2\\1&0&0\\0&3&1\end{bmatrix}\begin{bmatrix}f(1)&0&0\\0&f(2)&f'(2)\\0&0&f(2)\end{bmatrix}\begin{bmatrix}0&1&0\\-1&0&2\\3&0&-5\end{bmatrix},$$

$$f(x)=\sin\left(\frac{3\pi x}{2}\right),\ f'(x)=\frac{3\pi}{2}\cos\left(\frac{3\pi x}{2}\right)$$

$$\sin\left(\frac{3\pi A}{2}\right)=\begin{bmatrix}0&5&2\\1&0&0\\0&3&1\end{bmatrix}\begin{bmatrix}-1&0&0\\0&0&-3\pi/2\\0&0&0\end{bmatrix}\begin{bmatrix}0&1&0\\-1&0&2\\3&0&-5\end{bmatrix},$$

$$=\begin{bmatrix}-45\pi/2&0&75\pi/2\\0&-1&0\\-27\pi/2&0&45\pi/2\end{bmatrix}$$

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

【例】：已知$A=\begin{bmatrix}2&0&0\\1&1&1\\1&-1&3\end{bmatrix}$，求矩阵幂级数$\sum_{k=0}^{\infty}\frac{k+1}{10^{k+1}}A^k$

解：令$f(x)=\sum_{k=0}^{\infty}\frac{k+1}{10^{k+1}}x^k=\sum_{k=0}^{\infty}\frac{k+1}{10}(\frac{x}{10})^k=\sum_{k=0}^{\infty}[(\frac{x}{10})^{k+1}]'$

$$=[\sum_{k=0}^{\infty}(\frac{x}{10})^{k+1}]'=[\sum_{k=1}^{\infty}(\frac{x}{10})^{k}]'=[\sum_{k=0}^{\infty}(\frac{x}{10})^{k}]'$$

$|x|<10$

$$f(x)=\left[\frac{1}{1-\frac{x}{10}}\right]'=\frac{1}{10}(1-\frac{x}{10})^{-2}$$

$\rho(A)=2<10$

$$f(A)=\frac{1}{10}(1-\frac{A}{10})^{-2}=\frac{5}{128}\begin{bmatrix}4&0&0\\1&3&1\\1&-1&5\end{bmatrix}$$

两种特殊矩阵函数的性质

(1) $$e^{At}=\sum_{k=0}^{\infty}\frac{1}{k!}A^k t^k$$

(2) $$\sin At=\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)!}A^{2k+1}t^{2k+1}$$

(3) $$\cos At=\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k)!}A^{2k}t^{2k}$$

由矩阵函数幂级数可以得到一些简单的推论：

(1) $e^{O_{n\times n}} = I_{n\times n}$

(2) $e^{iA} = \cos A + i\sin A$

(3) $\cos A = \dfrac{1}{2}(e^{iA} + e^{-iA})$

(4) $\sin A = \dfrac{1}{2i}(e^{iA} - e^{-iA})$

(5) $\sin(-A) = -\sin A$

(6) $\cos(-A) = \cos A$

【定理】：设$A, B \in C^{n\times n}$，则当**$AB=BA$**时，有

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

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

(3) $\sin(2A)=2\sin A\cos A$

(4) $\cos(A+B)=\cos A\cos B-\sin A\sin B$

(5) $\cos(2A)=\cos^{2}A-\sin^{2}A$

(6) $e^{A}e^{-A}=e^{-A}e^{A}=I$

(7) $\sin^{2}A+\cos^{2}A=I$

注意：矩阵A与B的交换性条件是必不可少的.

提纲

- 1. 矩阵序列
- 2. 矩阵级数
- 3. 常见矩阵函数
- 4. 矩阵微积分
- 5. 广义逆矩阵
- 6. 广义特征值

【定义】

•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。

•若每个$a_{ij}(t)$是定义在$[a, b]$上的连续、可微、可积函数，则称$A(t)$在$[a, b]$上是连续、可微、可积的。

导数 $A'(t)=\left(a_{ij}'(t)\right)_{m\times n}$ 或 $\dfrac{\mathrm{d}}{\mathrm{d}t}A(t)=\left(\dfrac{\mathrm{d}}{\mathrm{d}t}a_{ij}(t)\right)_{m\times n}$

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

【例】：求函数矩阵$A(t)$的导数

$$A(t)=\begin{pmatrix} \sin t & \cos t & t \\ 2^t & \mathrm{e}^t & t^2 \\ 0 & 1 & t^3 \end{pmatrix}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)=\begin{pmatrix} \cos t & -sin t & 1 \\ 2^t \ln 2 & \mathrm{e}^t & 2t \\ 0 & 0 & 3t^2 \end{pmatrix}$$

运算法则

【定理】设$A(t)$与$B(t)$是适当阶可微的矩阵，

(1) $$\frac{\mathrm{d}}{\mathrm{d}t}(A(t)+B(t))=\frac{\mathrm{d}}{\mathrm{d}t}A(t)+\frac{\mathrm{d}}{\mathrm{d}t}B(t)$$

(2)当$\lambda(t)$为可微函数时有

$$\frac{\mathrm{d}}{\mathrm{d}t}(\lambda(t)A(t))=\left(\frac{\mathrm{d}}{\mathrm{d}t}\lambda(t)\right)A(t)+\lambda(t)\frac{\mathrm{d}}{\mathrm{d}t}A(t)$$

(3) $$\frac{\mathrm{d}}{\mathrm{d}t}(A(t)B(t))=\left(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right)B(t)+A(t)\frac{\mathrm{d}}{\mathrm{d}t}B(t)$$

(4)当$u=f(t)$关于t可微时，有

$$\frac{\mathbf{d}}{\mathbf{d}t}A(u)=f'(t)\frac{\mathbf{d}}{\mathbf{d}u}A(u)$$

(5)当$A^{-1}(t)$是可微矩阵时，有

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(A^{-1}(t)\right)=-A^{-1}(t)\left(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right)A^{-1}(t).$$

【定理】设$A(t)$与$B(t)$是适当阶可微的矩阵，

(5) $$\frac{\mathrm{d}}{\mathrm{d}t}\left(A^{-1}(t)\right)=-A^{-1}(t)\left(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right)A^{-1}(t).$$

运算法则

$$A(t)A^{-1}(t)=I$$

两边求导

$$\frac{d}{dt}A(t)\,A^{-1}(t)+A(t)\frac{d}{dt}A^{-1}(t)=0$$

【定理】：

（1） $\dfrac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{At}=A\mathrm{e}^{At}=\mathrm{e}^{At}A$

（2） $\dfrac{\mathrm{d}}{\mathrm{d}t}\sin At=A\cos At=(\cos At)A$

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

运算法则

【定理】设$A(t)$与$B(t)$是$[a, b]$上适当阶的可积矩阵，C和D是常数矩阵，λ是常数，那么

（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）$$\int_a^b A(t)C\mathrm{d}t = \left(\int_a^b A(t)\mathrm{d}t\right)C$$

$$\int_a^b DB(t)\mathrm{d}t = D\int_a^b B(t)\mathrm{d}t$$

（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

$$\int_a^b A'(t)\mathrm{d}t = A(b) - A(a)$$

矩阵作变量的微积分

【定义】设$f(X)$是以矩阵$X=(x_{ij})_{m\times n}$为自变量的mn元函数，且如下数量函数的导数都存在

$$\frac{\partial f}{\partial x_{ij}}(i=1,2,\cdots,m;\ j=1,2,\cdots,n)$$

规定f对矩阵变量X的导数$\frac{df}{dX}$为

$$\frac{\mathrm{d}f}{\mathrm{d}X}=\left(\frac{\partial f}{\partial x_{ij}}\right)_{m\times n}=\begin{pmatrix}\frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1n}}\\ \vdots & & \vdots\\ \frac{\partial f}{\partial x_{m1}} & \cdots & \frac{\partial f}{\partial x_{mn}}\end{pmatrix}$$

【定义】当$\boldsymbol{x}$是向量 $\boldsymbol{x}=(x_1,x_2,\ldots,x_n)^{\mathrm{T}}$，那么以其作为自变量的函数$f(X)$的导数

$$\frac{df}{d\boldsymbol{x}}=\left(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\cdots,\frac{\partial f}{\partial x_n}\right)^{\mathrm{T}}$$

称为数量函数对**向量变量的导数**。

$\boxed{\text{函数的梯度向量 grad } f}$

【例】：设$\alpha=\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$是给定的向量，$f(X)$是以向量$X=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$作为变量的函数：$f(X)=\alpha^T X=X^T\alpha$，求$\frac{df}{dX}$。

$$\frac{df}{dX}=\begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}=\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}=\alpha$$

【例】：设$A=[a_{ij}]_{n\times n}$是给定的矩阵，$X=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$向量变量，且$f(X)=X^TAX$，求$\frac{df}{dX}$。

$$\frac{df}{dX}=A^TX+AX=(A^T+A)X$$

特别地当A是对称矩阵的时候，$\frac{df}{dX}=2AX$

【例】： 设$X=[x_{ij}]_{n\times n}$是矩阵变量，并且X非奇异，令$f(X)=\det(X)$，求$\frac{\mathrm{d}}{\mathrm{d}X}f(X)$。

设x_{ij}的代数余子式X_{ij}

$$\frac{\partial}{\partial x_{ij}}f(X)=X_{ij}$$

$$\frac{\mathrm{d}}{\mathrm{d}X}f(X)=\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nn} \end{bmatrix}$$

$$=\det(X)(X^{-1})^T$$

用途

最小二乘问题：当线性方程组$Ax=b$无解时，对任意向量x都有$Ax-b\neq 0$，那么使得函数$\|Ax-b\|_2$最小的向量x_0称为最小二乘解，即

$$\|Ax_0-b\|_2=\min\|Ax-b\|_2$$

$$\|Ax-b\|_2^2=(Ax-b)^{\mathrm{T}}(Ax-b)$$

$$=x^{\mathrm{T}}A^{\mathrm{T}}Ax-x^{\mathrm{T}}A^{\mathrm{T}}b-b^{\mathrm{T}}Ax+b^{\mathrm{T}}b$$

$$\frac{\partial\|Ax-b\|_2^2}{\partial x}=2(A^TA)x-2A^Tb=0$$

$$(A^TA)x=A^Tb$$

提纲

广义逆

设方阵$A\in M_n$且 $\det(A)\neq 0$，那么存在逆矩阵A^{-1}，使得$AA^{-1}=A^{-1}A=E$。

推广到长方阵

【定义】设$A\in M_{m\times n}$, $X\in M_{n\times m}$满足下列等式中的某几个或全部

（1）$AXA=A$ （2）$XAX=X$ **Penrose方程**

（3）$(AX)^*=AX$ （4）$(XA)^*=XA$

则称X为A的广义逆矩阵。

满足全部四个等式称为Moore-Penrose逆。

广义逆

(1) $AXA = A$ (2) $XAX = X$

(3) $(AX)^* = AX$ (4) $(XA)^* = XA$

Penrose方程

X满足Penrose方程中的第$(i_1)\cdots(i_k)$等方程，则称X为A的$\{i_1,\cdots,i_k\}$逆，记为$A^{(i1,\ldots,ik)}$。

$A^{(1,2,3,4)}$：Moore-Penrose逆，也记为A^+

- **【定理】**设 $A \in M_{m\times n}$ ，则A的Moore-Penrose逆A^+存在且唯一。

证明：（存在性）

$$X = V\begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^*$$

（唯一性）:根据定义验证两个必须相等

如果A不是可逆矩阵，则除Moore-Penrose逆以外的广义逆矩阵都不是唯一的。

Moore-Penrose广义逆

- A^+的计算

【定理】设$A\in M_{m\times n}$的秩为r，且A的满秩分解为 $A=BC$，那么

$A^+=C^H(CC^H)^{-1}(B^HB)^{-1}B^H$

$$\mathrm{rank}(A)=m \;\Rightarrow A^+=A^*(AA^*)^{-1}$$

$$\mathrm{rank}(A)=n \;\Rightarrow A^+=(A^*A)^{-1}A^*$$

Moore-Penrose广义逆

- A^{+}的计算

【例】：求下列矩阵的Moore-Penrose逆

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

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

Moore-Penrose广义逆

- 【性质】设 $A\in M_{m\times n}$，则

（1）$(A^+)^+ = A$

（2）$(A^+)^* = (A^*)^+$

（3）$(\lambda A)^+ = \lambda^+ A^+$，其中 $\lambda = 0, \lambda^+ = 0; \lambda \neq 0, \lambda^+ = 1/\lambda$

（4）$\operatorname{rank}(A^+) = \operatorname{rank}(A)$

（5）$AA^+ = I_m$ 的充分必要条件是 $\operatorname{rank}(A) = m$

（6）$A^+A = I_n$ 的充分必要条件是 $\operatorname{rank}(A) = n$

Moore-Penrose广义逆

【用途】线性方程组求解

定理：设 $A\in M_{m\times n}$，$\boldsymbol{b}\in F^m$，则线性方程组 $A\boldsymbol{x}=\boldsymbol{b}$ 有解的充分必要条件是

$$AA^+\boldsymbol{b}=\boldsymbol{b}$$

且通解为

$$\boldsymbol{x}=A^+\boldsymbol{b}+(I-A^+A)\boldsymbol{y}$$

其中 $\boldsymbol{y}\in F^n$ 是任意向量。

Moore-Penrose广义逆

【用途】线性方程组求解

唯一解：当且仅当 $A^{+}A=I$，即 $\operatorname{rank}(A)=n$

极小范数解：线性方程组的无穷多个解中2范数最小的解，即

$$\|\boldsymbol{x}_0\|_2=\min_{A\boldsymbol{x}=\boldsymbol{b}}\|\boldsymbol{x}\|_2$$

最小二乘解：线性方程组不存在严格意义的解时的最优解

Moore-Penrose广义逆

【用途】线性方程组求解

定理：设 $A\in M_{m\times n}$，$\boldsymbol{b}\in F^{m}$，且 $A\boldsymbol{x}=\boldsymbol{b}$ 有解，则它的唯一极小范数解为

$$\boldsymbol{x}_0 = A^{+}\boldsymbol{b}$$

极小范数解：

$$\begin{aligned}\|\boldsymbol{x}\|_2^2 &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=[A^{+}\boldsymbol{b}+(I-A^{+}A)\boldsymbol{y}]^{\mathrm{T}}[A^{+}\boldsymbol{b}+(I-A^{+}A)\boldsymbol{y}]\\ &=\|\boldsymbol{A}^{+}\boldsymbol{b}\|_2^2+\|(I-A^{+}A)\boldsymbol{y}\|_2^2+\boldsymbol{b}^{\mathrm{T}}(A^{+})^{\mathrm{T}}(I-A^{+}A)\boldsymbol{y}\\ &\quad+\boldsymbol{y}^{\mathrm{T}}(I-A^{+}A)^{\mathrm{T}}A^{+}\boldsymbol{b}\\ &=\|\boldsymbol{A}^{+}\boldsymbol{b}\|_2^2+\|(I-A^{+}A)\boldsymbol{y}\|_2^2\geq\|\boldsymbol{A}^{+}\boldsymbol{b}\|_2^2\end{aligned}$$

Moore-Penrose广义逆

【用途】线性方程组求解

定理：设 $A \in M_{m\times n}$，$\boldsymbol{b} \in F^m$，且 $A\boldsymbol{x} = \boldsymbol{b}$ 有解，则它的唯一极小范数解为

$$\boldsymbol{x}_0 = A^+\boldsymbol{b}$$

极小范数解的唯一性：设x_0是极小范数解，则

$$\| \boldsymbol{x}_0 \|_2 = \| \boldsymbol{A}^+\boldsymbol{b} \|_2$$

存在y_0使得 $\boldsymbol{x}_0 = \boldsymbol{A}^+\boldsymbol{b} + (I - A^+A)\boldsymbol{y}_0$，从而

$$\| (I - A^+A)\boldsymbol{y}_0 \|_2 = 0 \Rightarrow \boldsymbol{x}_0 = \boldsymbol{A}^+\boldsymbol{b}$$

Moore-Penrose广义逆

【用途】线性方程组求解

定理： 设 $A \in M_{m\times n}$，$\boldsymbol{b} \in F^m$，方程组 $A\boldsymbol{x} = \boldsymbol{b}$ 的全部最小二乘解为

$$\boldsymbol{z} = A^{+}\boldsymbol{b} + (I - A^{+}A)\boldsymbol{y}$$

其中 $\boldsymbol{y} \in F^n$ 是任意向量。

提纲

- 1. 矩阵序列
- 2. 矩阵级数
- 3. 常见矩阵函数
- 4. 矩阵微积分
- 5. 广义逆矩阵
- **6. 广义特征值**

【定义】

- 设 $A, B \in M_n$是Hermite矩阵，且B是Hermite正定矩阵。若存在 $\lambda \in F$ 和向量$\theta \neq \mathbf{x} \in F^n$满足

$$A\mathbf{x} = \lambda B\mathbf{x}$$

则称λ为A相对于B的广义特征值，$\mathbf{x}$为属于λ的广义特征向量。

- 广义特征值与广义特征向量

$$A\boldsymbol{x} = \lambda B\boldsymbol{x}$$

$$\Longleftrightarrow \quad (\lambda B - A)\boldsymbol{x} = \boldsymbol{0}$$

广义特征值满足：

$$\det(\lambda B - A) = 0$$

- 广义特征值与广义特征向量

例：已知下列矩阵*A*和*B*，求A关于B的广义特征值问题

$$A=\begin{bmatrix}5 & 1\\ 1 & 1\end{bmatrix} \qquad B=\begin{bmatrix}18 & 2\\ 2 & 2\end{bmatrix}$$

$$\det(\lambda B-A)=(2\lambda-1)(16\lambda-4)$$

$$\lambda_1=\frac{1}{2},\ \lambda_2=\frac{1}{4} \qquad k_1\begin{pmatrix}0\\ 1\end{pmatrix},\quad k_2\begin{pmatrix}1\\ -1\end{pmatrix}$$

【性质】

- 【定义】设$B\in M_n$是Hermite正定矩阵。若列向量组$\boldsymbol{x}_1$、$\boldsymbol{x}_2$、…、$\boldsymbol{x}_m$满足

$$\boldsymbol{x}_i^* B\boldsymbol{x}_j = 0 \qquad (i \neq j; i, j = 1, 2, \ldots, m)$$

则称向量组$\boldsymbol{x}_1$、$\boldsymbol{x}_2$、…、$\boldsymbol{x}_m$按B正交；若还有

$$\boldsymbol{x}_i^* B\boldsymbol{x}_i = 1 \qquad (i = 1, 2, \ldots, m)$$

则称按B标准正交。

【性质】

【性质1】设B是Hermite正定矩阵，则按B正交的非零向量组线性无关。

【性质2】广义特征值都是实数，且有n个按B正交的广义特征向量，构成F^n的一组基。

【性质3】对应于不同广义特征值的广义特征向量按B正交。