矩阵分析基础

1. 矩阵序列

定义1 $\{\mathbf{A}_k\}_{k=1}^{\infty}$为 $\mathbf{C}^{m\times n}$中的矩阵序列，其中 $\mathbf{A}_k=\left(a_{ij}^{(k)}\right)$ 又 $\mathbf{A}=\left(a_{ij}\right)\in\mathbf{C}^{m\times n}$。如果 $\lim\limits_{k\to\infty}a_{ij}^{(k)}=a_{ij}$ 对 $i=1,2,\ldots,m$，$j=1,2,\ldots,n$ 均成立，则称矩阵序列 $\{\mathbf{A}_k\}_{k=1}^{\infty}$ 收敛，而A称为矩阵序列 $\{\mathbf{A}_k\}_{k=1}^{\infty}$ 的极限，记为 $\lim\limits_{k\to\infty}\mathbf{A}_k=\mathbf{A}$。不收敛的矩阵序列称为发散的。

矩阵序列收敛 $\Leftrightarrow$ 元素收敛

例1 讨论矩阵序列 $\{\mathbf{A}_k\}_{k=1}^{\infty}$ 的收敛性。其中

解：根据定义，只须求出它的每一个元素的极限即可，因此它的极限为：

$$\lim_{k\to\infty} A_k = \begin{pmatrix} \lim\limits_{k\to\infty}\left(1+\dfrac{1}{k}\right)^k & \lim\limits_{k\to\infty}\dfrac{\sin k}{k} \\ \lim\limits_{k\to\infty} 1 & \lim\limits_{k\to\infty} e^{-k} \\ \lim\limits_{k\to\infty}\dfrac{2+k}{k} & \lim\limits_{k\to\infty}\sqrt[k]{k} \end{pmatrix} = \begin{pmatrix} e & 0 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} = A$$

由矩阵序列极限的定义可以看出，矩阵序列收敛的性质和数列收敛性质相似。

由定义可见，$\mathbf{C}^{m\times n}$ 中的矩阵序列的收敛相当于mn个数列同时收敛。因此可以用初等分析的方法来研究它。

同时研究mn个数列的极限未免繁琐，我们可以利用**矩阵范数**来研究矩阵序列的极限。

矩阵收敛 $\Leftrightarrow$ 元素收敛

元素收敛 $\Rightarrow$ **范数收敛**

定理1 设 $\{\mathbf{A}_k\}_{k=1}^{\infty}$ 为 $\mathbf{C}^{m\times n}$ 中的矩阵序列，$\|\cdot\|$ 为 $\mathbf{C}^{m\times n}$ 中的一种矩阵范数，则矩阵序列 $\{\mathbf{A}_k\}_{k=1}^{\infty}$ 收敛于矩阵A的充要条件是 $\|\mathbf{A}_k-\mathbf{A}\|$ 收敛于零。

证：首先，利用范数的等价性知，对于$\mathbf{C}^{m\times n}$ 中的任意两个矩阵范数 $\|\cdot\|_t$ 和 $\|\cdot\|_s$，存在常数 $c_1\geq c_2>0$，使得

$$c_2\cdot\|\mathbf{A}_k-\mathbf{A}\|_t\leq\|\mathbf{A}_k-\mathbf{A}\|_s\leq c_1\cdot\|\mathbf{A}_k-\mathbf{A}\|_t$$

即有

$$\lim_{k\to\infty}\|\mathbf{A}_k-\mathbf{A}\|_t=0=\lim_{k\to\infty}\|\mathbf{A}_k-\mathbf{A}\|_s$$

即收敛于零是一致的。

因此，只需证明定理对一种特定的矩阵范数成立即可。

我们选取∞-范数加以证明。根据 ∞ -范数的定义，对于 $1 \le i \le m$， $1 \le j \le n$，均有

$$\left| a_{ij}^{(k)} - a_{ij} \right| \le \max_{1\le i\le m}\left\{ \sum_{j=1}^{n} \left| a_{ij}^{(k)} - a_{ij} \right| \right\} = \| \boldsymbol{A}_k - \boldsymbol{A} \|_\infty$$

$$\le \sum_{i=1}^{m}\sum_{j=1}^{n} \left| a_{ij}^{(k)} - a_{ij} \right|$$

因此，

$$\lim_{k\to\infty} A_k = A \quad \Longleftrightarrow \quad \lim_{k\to\infty} \|\mathbf{A}_k - \mathbf{A}\|_\infty = 0$$

证毕

推论 设 $\{\mathbf{A}_k\}_{k=1}^{\infty}$，$A\in\mathbf{C}^{m\times n}$，并且 $\lim\limits_{k\to\infty}\mathbf{A}_k=\mathbf{A}$，则 $\lim\limits_{k\to\infty}\|A_k\|=\|A\|$

证：由 $\left|\|A_k\|-\|A\|\right|\le\|A_k-A\|$，即结论成立。

需要指出的是，此结论只是充分条件，反过来不一定成立。

给定矩阵序列 $A_k=\begin{pmatrix}(-1)^k & \frac{1}{k+1}\\ 1 & 2\end{pmatrix}$ 和矩阵 $A=\begin{pmatrix}1 & 0\\ 1 & 2\end{pmatrix}$

显然有 $\lim\limits_{k\to\infty}\|A_k\|_F=\lim\limits_{k\to\infty}\sqrt{|-1|^k+1^2+2^2+\frac{1}{(k+1)^2}}=\sqrt{6}=\|A\|_F$

但是矩阵序列A_k不收敛，故更不收敛于矩阵 $A=\begin{pmatrix}1 & 0\\ 1 & 2\end{pmatrix}$。

性质1 设 $\{\mathbf{A}_k\}_{k=1}^{\infty}$ 和 $\{\mathbf{B}_k\}_{k=1}^{\infty}$ 为 $C^{m\times n}$ 中的矩阵序列，并且

$$\lim_{k\to\infty}\mathbf{A}_k=\mathbf{A}, \quad \lim_{k\to\infty}\mathbf{B}_k=\mathbf{B}$$

则

$$\lim_{k\to\infty}\left(\alpha\mathbf{A}_k+\beta\mathbf{B}_k\right)=\alpha\mathbf{A}+\beta\mathbf{B},\ \ \forall\alpha,\beta\in\mathbf{C}$$

证 **由** $\left\|(\alpha A_k+\beta B_k)-(\alpha A+\beta B)\right\|=\left\|\alpha\left(A_k-A\right)-\beta\left(B_k-B\right)\right\|$

$$\le|\alpha|\cdot\|A_k-A\|+|\beta|\cdot\|B_k-B\|$$

由定理1，即结论成立。

性质2 设 $\{\mathbf{A}_k\}_{k=1}^{\infty}$ 和 $\{\mathbf{B}_k\}_{k=1}^{\infty}$ 分别为 $\mathbf{C}^{m\times n}$ 和 $\mathbf{C}^{n\times l}$ 中的矩阵序列，

并且

$$\lim_{k\to\infty}\mathbf{A}_k=\mathbf{A}\ ,\qquad \lim_{k\to\infty}\mathbf{B}_k=\mathbf{B}$$

则

$$\lim_{k\to\infty}\mathbf{A}_k\mathbf{B}_k=\mathbf{AB}$$

证 由

$$\|A_kB_k-AB\|=\|A_kB_k-A_kB+A_kB-AB\|$$

$$\le\|B\|\cdot\|A_k-A\|+\|A_k\|\cdot\|B_k-B\|$$

由定理1和推论可知，结论成立。

性质3 设 $\{\mathbf{A}_k\}_{k=1}^{\infty}\in\mathbf{C}^{n\times n}$中的矩阵序列，$\lim\limits_{k\to\infty}\mathbf{A}_k=\mathbf{A}$ 并且 $\mathbf{A}_k\,(k=1,2,\cdots)$ 和 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$均为可逆矩阵，则

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

证 因为 $A_kA_k^{-1}=I$， 由性质**3.2**

$$\lim_{k\to\infty}\left(\boldsymbol{A}_k\boldsymbol{A}_k^{-1}\right)=\boldsymbol{A}\lim_{k\to\infty}\boldsymbol{A}_k^{-1}=\boldsymbol{I}$$

$$\Longrightarrow\quad\lim_{k\to\infty}\mathbf{A}_k^{-1}=\mathbf{A}^{-1}$$

注意， 性质3中条件 $\mathbf{A}_k(k=1,2,\cdots)$ 和$\boldsymbol{A}$均为可逆的是不可少的。

因为即使 $\mathbf{A}_k(k=1,2,\cdots)$ 可逆也不能保证$\boldsymbol{A}$一定可逆。

例如，$\boldsymbol{A}_k=\begin{pmatrix}1+\dfrac{1}{k} & 1\\ 1 & 1\end{pmatrix}$，对于 $\mathbf{A}_k(k=1,2,\cdots)$ 都有 $\left(A_k\right)^{-1}=\begin{pmatrix}k & k\\ -k & k+1\end{pmatrix}$

但是 $\lim\limits_{k\to\infty}\boldsymbol{A}_k=\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}=\boldsymbol{A}$ 不可逆。

在矩阵序列中，最常见的是由一个方阵的幂构成的序列。关于这样的矩阵序列有以下的概念和收敛定理。

定义 设 $A\in \mathbf{C}^{n\times n}$，若 $\lim\limits_{k\to\infty}\mathbf{A}^k=\boldsymbol{0}$ ，则称A为**收敛矩阵**。

例 2 设 $A\in \mathbf{C}^{n\times n}$，证明 $\lim\limits_{k\to\infty}\mathbf{A}^k=\boldsymbol{0}$ 的充分必要条件是 $\rho(\mathbf{A})<1$。

证 必要性 由定理1知 $\lim\limits_{k\to\infty}\mathbf{A}^k=0$ 的充分必要条件是对任意一种矩阵范数 $\|\cdot\|$ 均有$\lim\limits_{k\to\infty}\|\mathbf{A}^k\|=0$。因此对充分大的$k$, 必有$\|A^k\|<1$

因此得

$$\left(\rho(\mathbf{A})\right)^k=\rho(\mathbf{A}^k)\le\|\mathbf{A}^k\|<1$$

利用矩阵谱半径的定义以及相容矩阵范数的性质有：

$$\rho(\mathbf{A})<1$$

充分性 根据定理2.8，对于 $\varepsilon=\frac{1}{2}(1-\rho(\mathbf{A}))>0$ 一定存在一种相容的矩阵范数 $\|\cdot\|$，使得 $\|\mathbf{A}\|\le\rho(\mathbf{A})+\varepsilon$。

又根据相容矩阵范数的性质有， 再注意到上述关系式中

$$\rho(\mathbf{A})+\varepsilon=\frac{1}{2}\big(1+\rho(\mathbf{A})\big)<1$$

那么

$$\left\|\mathbf{A}^k\right\|\le\|\mathbf{A}\|^k\le\big(\rho(\mathbf{A})+\varepsilon\big)^k\le q^k<1\qquad(0<q<1)$$

于是，$\lim\limits_{k\to\infty}\left\|\mathbf{A}^k\right\|=0$ 根据定理1 即知 $\lim\limits_{k\to\infty}\mathbf{A}^k=0$。

推论 设 $A\in\mathbf{C}^{n\times n}$，$\lim\limits_{k\to\infty}\mathbf{A}^k=\boldsymbol{0}$ 的充分必要条件是

存在$\mathbf{C}^{n\times n}$上的某种范数 $\|\cdot\|$，使得 $\|A\|<1$

$$\boxed{\rho(\mathbf{A})<1}\Longleftrightarrow\boxed{\lim_{k\to\infty}\mathbf{A}^k=\boldsymbol{0}}\ \overset{\Longleftarrow}{\not\Longrightarrow}\ \boxed{\|A\|<1}$$

练习题 判断对下列矩阵是否有 $\lim\limits_{k\to\infty}\mathbf{A}^k=0$

（1）$A=\dfrac{1}{6}\begin{pmatrix}1 & -8\\ -2 & 1\end{pmatrix}$，（2）$A=\begin{pmatrix}0.2 & 0.1 & 0.2\\ 0.5 & 0.5 & 0.4\\ 0.1 & 0.3 & 0.2\end{pmatrix}$

解：（1）取 $B=\begin{pmatrix}1 & -8\\ -2 & 1\end{pmatrix}$， 则 $\lambda(A)=\dfrac{1}{6}\lambda(B)$ ，令

$$\det(\lambda I-B)=\begin{vmatrix}\lambda-1 & 8\\ 2 & \lambda-1\end{vmatrix}=(\lambda-1)^2-16=(\lambda-5)(\lambda+3)=0$$

得 $\lambda_1(B)=5$，$\lambda_2(B)=-3$，进而得 $\lambda_1(A)=\dfrac{5}{6}$，$\lambda_2(A)=-\dfrac{1}{2}$。

于是，$\rho(A)=\dfrac{5}{6}<1$ 故 $\lim\limits_{k\to\infty}\mathbf{A}^k=\boldsymbol{0}$。

（2）因为 $\|A\|_1=0.9<1$，由推论，故 $\lim\limits_{k\to\infty}A^k=\boldsymbol{0}$

2 矩阵级数

定义2 设 $\{\mathbf{A}_k\}_{k=1}^{\infty}$ 为$\mathbf{C}^{m\times n}$中的矩阵序列，称

$$\mathbf{A}_1+\mathbf{A}_2+\cdots+\mathbf{A}_k+\cdots$$

为由矩阵序列 $\{\mathbf{A}_k\}_{k=1}^{\infty}$ 构成的矩阵级数，记为 $\sum_{k=1}^{\infty}\mathbf{A}_k$。

定义3 记 $\mathbf{S}_k=\sum_{i=1}^{k}\mathbf{A}_i$，称之为矩阵级数 $\sum_{k=1}^{\infty}\mathbf{A}_k$ 的前k项部分和。若矩阵序列 $\{\mathbf{S}_k\}_{k=1}^{\infty}$ 收敛且 $\lim_{k\to\infty}\mathbf{S}_k=\mathbf{S}$，则称矩阵级数 $\sum_{k=1}^{\infty}\mathbf{A}_k$ 收敛，而矩阵S称为矩阵级数的和矩阵，记为 $\mathbf{S}=\sum_{k=1}^{\infty}\mathbf{A}_k$。不收敛的矩阵级数称为发散的。

显然，和 $\sum_{k=1}^{\infty}A_k=\mathbf{S}=(s_{ij})$ 的意义指的是：$\sum_{k=1}^{\infty}a_{ij}^{(k)}=s_{ij}$ $(i=1,2,\cdots,m,\quad j=1,2,\cdots,n)$ 即$m\times n$个数项级数 $\sum_{k=1}^{\infty}a_{ij}^{(k)}$ 均为收敛的。

例 研究矩阵级数 $\sum_{k=1}^{\infty}\mathbf{A}_k$ 的收敛性，其中

解：因为

$$\boldsymbol{S}_N=\sum_{k=0}^{N}\boldsymbol{A}_k=\begin{pmatrix}\sum_{k=0}^{N}\frac{1}{(k+1)(k+2)} & \sum_{k=0}^{N}\frac{1}{2^k}\\ 0 & \frac{\pi}{3}\left(\sum_{k=0}^{N}\frac{1}{4^k}\right)\end{pmatrix},\quad k=1,2,\cdots,$$

于是

$$\boldsymbol{S}=\lim_{N\to\infty}\boldsymbol{S}_N=\begin{pmatrix}1 & 2\\ 0 & \frac{4}{3}\pi\end{pmatrix}$$

故矩阵级数 $\sum_{k=1}^{\infty}\mathbf{A}_k$ 收敛，且和为S。

例3 设A为n阶方阵，证明：矩阵级数 $\mathbf{I}_n+\mathbf{A}+\mathbf{A}^2+\cdots+\mathbf{A}^k+\cdots$ 收敛($A_0=I_n$)的充要条件是 $\rho(\mathbf{A})<1$，而且矩阵级数收敛时有

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

证 必要性 若矩阵级数$\sum_{k=1}^{\infty}\mathbf{A}_k$收敛，则有 $S=\lim_{k\to\infty}S_k$。又级数 $\mathbf{I}_n+\mathbf{A}+\mathbf{A}^2+\cdots+\mathbf{A}^k+\cdots$ 的前k项部分和与前k+1项部分和分别为

$$\mathbf{S}_k = I_n + A + A^2 + \cdots + A^{k-1},\ S_{k+1} = I_n + A + A^2 + \cdots + A^k$$

因此 $A^k = S_{k+1} - S_k$ ，利用极限运算法则有

$$\lim_{k\to\infty}\mathbf{A}^k = \lim_{k\to\infty}\left[\mathbf{S}_{k+1} - \mathbf{S}_k\right] = 0$$

根据例2，$\rho(\mathbf{A})<1$。

★ 级数收敛的必要条件是通项的极限为0.

充分性 由 $\mathbf{A}\mathbf{S}_k = \mathbf{A}\left(I+\mathbf{A}+\cdots+\mathbf{A}^{k-1}\right) = \mathbf{A}+\mathbf{A}^2+\cdots+\mathbf{A}^k$

则有 $S_k - AS_k = I - A^k$，$(I-A)S_k = I - A^k$

由 $\rho(\mathbf{A})<1$，则存在某种范数 $\|\cdot\|$，使得 $\|A\|<1$，且（$I-A$）可逆

又有 $\lim\limits_{k\to\infty}\mathbf{A}^k=0$，根据矩阵序列极限法则，有

$$\sum_{k=0}^{\infty}A^k=\lim_{k\to\infty}\mathbf{S}_k = \lim_{k\to\infty}\left[\left(\mathbf{I}-\mathbf{A}\right)^{-1}(\mathbf{I}-\mathbf{A}^k)\right]=\left(\mathbf{I}-\mathbf{A}\right)^{-1}\lim_{k\to\infty}(\mathbf{I}-\mathbf{A}^k)=\left(\mathbf{I}-\mathbf{A}\right)^{-1}。$$

计算 $\sum\limits_{k=0}^{\infty}A^k$，其中 $A=\begin{pmatrix}0.1 & 0.7\\ 0.3 & 0.6\end{pmatrix}$。由于 $\|A\|_\infty=0.9<1$，故 $\rho(\mathbf{A})<1$，从而 $\sum\limits_{k=0}^{\infty}A^k$ 收敛，且有

$$\sum_{k=0}^{\infty}\mathbf{A}^k=(\mathbf{I}-\mathbf{A})^{-1}=\begin{pmatrix}0.9 & -0.7\\ -0.3 & 0.4\end{pmatrix}^{-1}=\frac{2}{3}\begin{pmatrix}4 & 7\\ 3 & 9\end{pmatrix}$$

矩阵级数收敛的定义与数项级数的定义没有本质的区别，我们有一些类似于数项级数的概念和结论。

定义4 设 $\sum_{k=1}^{\infty}\mathbf{A}_k$ 为$\mathbf{C}^{m\times n}$中的矩阵级数，其中 $\mathbf{A}_k=\left(a_{ij}^{(k)}\right)$。如果 $\sum_{k=1}^{\infty}a_{ij}^{(k)}$ 对任意的$1\leqslant i\leqslant m$，$1\leq j\leq n$均为绝对收敛的，则称矩阵级数 $\sum_{k=1}^{\infty}\mathbf{A}_k$ 绝对收敛。

对比矩阵级数绝对收敛的定义以及高等数学中的数项级数的绝对收敛的定义可以得出矩阵级数收敛的一些性质。

性质6 矩阵级数 $\sum_{k=1}^{\infty}\mathbf{A}_k$ 为绝对收敛的充分必要条件是正项级数 $\sum_{k=1}^{\infty}\|\mathbf{A}_k\|$ 收敛。

利用矩阵范数的等价性，只需证明对于∞−范数定理成立即可。

证 必要性 如果 $\sum_{k=1}^{\infty}\mathbf{A}_k$ 是绝对收敛的，由定义即对任意的 $1\leqslant i\leqslant m$，$1\leqslant j\leqslant n$，$\sum_{k=1}^{\infty}a_{ij}^{(k)}$ 均绝对收敛，即存在充分大的 N 和一个与N无关的正数M，使得

$$\sum_{k=1}^{N}\left|a_{ij}^{(k)}\right|< M\ ,\ (i=1,2,\cdots,m,\quad j=1,2,\cdots,n)$$

从而有

$$\sum_{k=1}^{N}\|A_k\|_{\infty}=\sum_{k=1}^{N}\left(\max_i\sum_{j=1}^{n}\left|a_{ij}^{(k)}\right|\right)=\max_i\left\{\sum_{j=1}^{n}\sum_{k=1}^{N}\left|a_{ij}^{(k)}\right|\right\}\leq\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{N}\left|a_{ij}^{(k)}\right|< nm\cdot M$$

因此 $\sum_{k=1}^{\infty}\|\mathbf{A}_k\|_{\infty}$ 为收敛的正项级数。

充分性 如果 $\sum_{k=1}^{\infty}\|\mathbf{A}_k\|_\infty$ 为收敛的正项级数，那么有

$$\sum_{k=1}^{N}\left|a_{ij}^{(k)}\right| \le \sum_{k=1}^{N}\|\mathbf{A}_k\|_\infty < M \quad i=1,2,\cdots,m, \quad j=1,2,\cdots,n$$

可知$\boldsymbol{m}\times\boldsymbol{n}$个级数 $\sum_{k=1}^{\infty}a_{ij}^{(k)}$ 均为绝对收敛的，利用定义4可知矩阵级数 $\sum_{k=1}^{\infty}\mathbf{A}_k$ 是绝对收敛的。

性质 5 若矩阵级数 $\sum_{k=1}^{\infty}\mathbf{A}_k$ 是绝对收敛，则它一定是收敛的，并且任意调换各项的顺序所得到的级数还是收敛的，且级数和不变。

绝对收敛 $\rightleftarrows$ 收敛

性质7 设 $\sum_{k=1}^{\infty}\mathbf{A}_k$ 为$\mathbf{C}^{m\times n}$中的绝对收敛的矩阵级数，$\sum_{k=1}^{\infty}\mathbf{B}_k$ 为 $\mathbf{C}^{n\times l}$中的绝对收敛的级数，并且 $\mathbf{A}=\sum_{k=1}^{\infty}\mathbf{A}_k$，$\mathbf{B}=\sum_{k=1}^{\infty}\mathbf{B}_k$，则 $\sum_{k=1}^{\infty}\mathbf{A}_k\bullet\sum_{k=1}^{\infty}\mathbf{B}_k$ 按任何方式排列得到的级数也是绝对收敛的，且和为$\boldsymbol{AB}$.

$$\sum_{k=1}^{\infty} A_k \cdot \sum_{k=1}^{\infty} B_k = (A_1 + A_2 + \cdots A_p + \cdots)(B_1 + B_2 + \cdots + B_p + \cdots)$$

$$= A_1B_1 + A_1B_2 + A_1B_3 + \cdots + A_1B_p + \cdots$$

$$+ A_2B_1 + A_2B_2 + A_2B_3 + \cdots + A_2B_p + \cdots$$

$$+ A_3B_1 + A_3B_2 + A_3B_3 + \cdots + A_3B_p + \cdots$$

$$+ \cdots$$

$$+ A_pB_1 + A_pB_2 + A_pB_3 + \cdots + A_pB_p + \cdots$$

$$+ \cdots$$

$$= \sum_{k=1}^{\infty} C_k, \quad C_k = \sum_{i,j} A_i B_j$$

有限项和

证 因为矩阵级数 $\sum_{k=1}^{\infty}\mathbf{A}_k$ 和 $\sum_{k=1}^{\infty}\mathbf{B}_k$ 均为绝对收敛的，故存在正数 M_{A}、M_{B}使得对任意的正整数p，均有，

$$\sum_{k=1}^{p}\|\mathbf{A}_k\|_\infty \le M_{\mathbf{A}} \qquad \sum_{k=1}^{p}\|\mathbf{B}_k\|_\infty \le M_{\mathbf{B}}$$

按任意排列方式得到的矩阵级数部分和 $\sum_{k=1}^{p}C_k$ ，$C_k=\sum_{i,j}A_iB_j$

均满足 $\sum_{k=1}^{p}\|\mathbf{C}_k\|_\infty \le \sum_{k=1}^{p}\sum_{i,j}\|\mathbf{A}_i\mathbf{B}_j\|_\infty \le \sum_{k=1}^{N_i}\|\mathbf{A}_k\|_\infty \cdot \sum_{k=1}^{N_j}\|\mathbf{B}_k\|_\infty \le \mathrm{M}_{\mathbf{A}}\cdot \mathrm{M}_{\mathbf{B}}$

其中, 设 N_i、N_j是构成 $\mathbf{C}_k(k=1, 2, \ldots, p)$的$A_i$、$B_j$角标的最大者。因此 $\sum_{k=1}^{\infty}\|\mathbf{C}_k\|_\infty$ 是一个有上界的正项级数，故收敛。

$$\sum_{k=1}^{\infty} \boldsymbol{C}_k = \sum_{k=1}^{\infty} \boldsymbol{A}_k \cdot \sum_{k=1}^{\infty} \boldsymbol{B}_k =$$

$$\begin{aligned}
&= \boldsymbol{A}_1\boldsymbol{B}_1 + \boldsymbol{A}_1\boldsymbol{B}_2 + \boldsymbol{A}_1\boldsymbol{B}_3 + \cdots + \boldsymbol{A}_1\boldsymbol{B}_p + \cdots \\
&+ \boldsymbol{A}_2\boldsymbol{B}_1 + \boldsymbol{A}_2\boldsymbol{B}_2 + \boldsymbol{A}_2\boldsymbol{B}_3 + \cdots + \boldsymbol{A}_2\boldsymbol{B}_p + \cdots \\
&+ \boldsymbol{A}_3\boldsymbol{B}_1 + \boldsymbol{A}_3\boldsymbol{B}_2 + \boldsymbol{A}_3\boldsymbol{B}_3 + \cdots + \boldsymbol{A}_3\boldsymbol{B}_p + \cdots \\
&+ \cdots \\
&+ \boldsymbol{A}_p\boldsymbol{B}_1 + \boldsymbol{A}_p\boldsymbol{B}_2 + \boldsymbol{A}_p\boldsymbol{B}_3 + \cdots + \boldsymbol{A}_p\boldsymbol{B}_p + \cdots \\
&+ \cdots
\end{aligned}$$

柯西积: $\boldsymbol{C}_1 = \boldsymbol{A}_1\boldsymbol{B}_1, \quad \boldsymbol{C}_2 = \boldsymbol{A}_1\boldsymbol{B}_2 + \boldsymbol{A}_2\boldsymbol{B}_1,$

$\boldsymbol{C}_3 = \boldsymbol{A}_1\boldsymbol{B}_3 + \boldsymbol{A}_2\boldsymbol{B}_2 + \boldsymbol{A}_1\boldsymbol{B}_3, \quad \boldsymbol{C}_k = \sum\limits_{i+j=k+1} \boldsymbol{A}_i\boldsymbol{B}_j,$

$$\sum_{k=1}^{\infty} C_k = \sum_{k=1}^{\infty} A_k \cdot \sum_{k=1}^{\infty} B_k =$$

$$\begin{aligned}
&= A_1B_1 + A_1B_2 + A_1B_3 + \cdots + A_1B_p + \cdots \\
&+ A_2B_1 + A_2B_2 + A_2B_3 + \cdots + A_2B_p + \cdots \\
&+ A_3B_1 + A_3B_2 + A_3B_3 + \cdots + A_3B_p + \cdots \\
&+ \cdots \\
&+ A_pB_1 + A_pB_2 + A_pB_3 + \cdots + A_pB_p + \cdots \\
&+ \cdots
\end{aligned}$$

$$C_1 = A_1B_1, \quad C_2 = A_1B_2 + A_2B_1 + A_2B_2,$$

$$C_3 = A_3(B_1 + B_2 + B_3) + (A_1 + A_2)B_3, \quad C_p = A_p\sum_{i=1}^{p} B_i + \sum_{i=1}^{p-1} A_iB_p,$$

记 $\boldsymbol{C}_k = \boldsymbol{A}_k \sum_{i=1}^{k} \boldsymbol{B}_i + \sum_{i=1}^{k-1} \boldsymbol{A}_i \boldsymbol{B}_k, \quad (k = 1,2,\cdots)$

级数 $\sum_{k=1}^{\infty} \mathrm{C}_k$ 的前$\boldsymbol{p}$项的部分和为：

$$\sum_{k=1}^{p} \boldsymbol{C}_k = \sum_{k=1}^{p} \left(\boldsymbol{A}_k \sum_{i=1}^{k} \boldsymbol{B}_i + \boldsymbol{B}_k \sum_{i=1}^{k-1} \boldsymbol{A}_i \right)$$

$$= \sum_{k=1}^{p} \boldsymbol{A}_k \left(\boldsymbol{B}_1 + \boldsymbol{B}_2 + \cdots + \boldsymbol{B}_k \right) + \sum_{k=1}^{p} \boldsymbol{B}_k \left(\boldsymbol{A}_1 + \boldsymbol{A}_2 + \cdots + \boldsymbol{A}_{k-1} \right)$$

$$= \left(\boldsymbol{A}_1 + \boldsymbol{A}_2 + \cdots + \boldsymbol{A}_p \right) \left(\boldsymbol{B}_1 + \boldsymbol{B}_2 + \cdots + \boldsymbol{B}_p \right) = \sum_{k=1}^{p} \boldsymbol{A}_k \cdot \sum_{k=1}^{p} \boldsymbol{B}_k$$

利用极限的运算法则有 $\sum_{k=1}^{\infty} \mathrm{C}_k = \boldsymbol{AB}$

性质8 设 $\mathbf{P}\in\mathbf{C}^{p\times m}$和 $\mathbf{Q}\in\mathbf{C}^{n\times q}$ 为给定矩阵，如果$\boldsymbol{m}\times\boldsymbol{n}$型矩阵级数 $\sum_{k=0}^{\infty}\boldsymbol{A}_k$ 收敛（或绝对收敛），则 $p\times q$ 矩阵级数 $\sum_{k=0}^{\infty}\boldsymbol{P}\boldsymbol{A}_k\mathbf{Q}$ 也收敛（或绝对收敛），且有等式

$$\sum_{k=0}^{\infty}\boldsymbol{P}\boldsymbol{A}_k\boldsymbol{Q} = \left(\sum_{k=0}^{\infty}\boldsymbol{A}_k\right)$$

证 设 $\sum_{k=0}^{\infty}\mathbf{A}_k$收敛于矩阵$\boldsymbol{S}$，即 $\boldsymbol{S}=\sum_{k=0}^{\infty}\boldsymbol{A}_k=\lim_{n\to\infty}\sum_{k=0}^{n}\boldsymbol{A}_k$，而由等式 $\sum_{k=0}^{n}\boldsymbol{P}\boldsymbol{A}_k\boldsymbol{Q}=\boldsymbol{P}\left(\sum_{k=0}^{n}\boldsymbol{A}_k\right)\boldsymbol{Q}$ 取极限即得：

$$\sum_{k=0}^{\infty}\boldsymbol{P}\boldsymbol{A}_k\boldsymbol{Q}=\lim_{n\to\infty}\sum_{k=0}^{n}\boldsymbol{P}\boldsymbol{A}_k\boldsymbol{Q}=\boldsymbol{P}\left(\lim_{n\to\infty}\sum_{k=0}^{n}\boldsymbol{A}_k\right)\boldsymbol{Q}=\boldsymbol{P}\boldsymbol{S}\boldsymbol{Q}=\boldsymbol{P}\left(\sum_{k=0}^{\infty}\mathbf{A}_k\right)\boldsymbol{Q}$$

即 $\sum_{k=0}^{\infty}\mathbf{P}\mathbf{A}_k\mathbf{Q}$ 收敛，且有 $\sum_{k=0}^{\infty}\mathbf{P}\mathbf{A}_k\mathbf{Q}=\boldsymbol{P}\left(\sum_{k=0}^{\infty}\mathbf{A}_k\right)\boldsymbol{Q}$

现设$\sum_{k=0}^{\infty}\mathbf{A}_k$绝对收敛， 由矩阵级数性质6知，$\sum_{k=0}^{\infty}\|\mathbf{A}_k\|$收敛， 又

$$\|\mathbf{P}\mathbf{A}_k\mathbf{Q}\| \le \|\mathbf{P}\|\,\|\mathbf{A}_k\|\,\|\mathbf{Q}\|$$

利用比较判别法， 即知级数$\sum_{k=0}^{\infty}\|\mathbf{P}\mathbf{A}_k\mathbf{Q}\|$收敛，再利用矩阵级数性质5，便知矩阵级数 $\sum_{k=0}^{\infty}\mathbf{P}\mathbf{A}_k\mathbf{Q}$ 绝对收敛。

绝对收敛

$$\sum_{k=0}^{\infty}\mathbf{P}\mathbf{A}_k\mathbf{Q} = \boldsymbol{P}\left(\sum_{k=0}^{\infty}\mathbf{A}_k\right)\boldsymbol{Q}$$

乘积运算和无穷和运算可交换

二、矩阵幂级数

定理2 设 $\sum_{k=0}^{\infty} a_k t^k$ 为收敛半径为r的幂级数，A为n阶方阵，则

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

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

证 （1）如果 $\rho(\mathbf{A})<r$ ，根据矩阵范数的性质，对于 $\varepsilon=\frac{1}{2}[r-\rho(\mathbf{A})]>0$, 一定存在一种相容的矩阵范数 $\|\cdot\|$ ，使得

$$\|\mathbf{A}\| \le \rho(\mathbf{A})+\varepsilon = \rho(A)+\frac{1}{2}r-\frac{1}{2}\rho(A)=\frac{r+\rho(A)}{2} < r$$

因此，数项级数 $\sum_{k=0}^{\infty}|a_k|\|A\|^k$ 收敛，又

$$\|a_k A^k\| = |a_k|\|A^k\| \le |a_k|\|A\|^k$$

再由数项级数比较判别法可知，$\sum_{k=0}^{\infty}\|a_k\mathbf{A}^k\|$ 收敛。

再利用矩阵级数性质3.6知，$\sum_{k=0}^{\infty}a_k\mathbf{A}^k$ 绝对收敛。

（2）如果 $\rho(A)>r$，设 $\boldsymbol{A}\boldsymbol{x}=\lambda_i\boldsymbol{x}$，其中 $|\lambda_i|=\rho(A)$， 且 $\boldsymbol{x}$ 为单位长度特征向量。下面用反证法证明矩阵幂级数 $\sum_{k=0}^{\infty}a_k\mathbf{A}^k$ 发散。如果它是收敛的，则利用矩阵收敛的性质8知，数项级数

$$\infty > \boldsymbol{x}^H\left(\sum_{k=0}^{\infty}a_k\mathbf{A}^k\right)\boldsymbol{x}=\sum_{k=0}^{\infty}a_k(\boldsymbol{x}^H\boldsymbol{A}^k\boldsymbol{x})=\sum_{k=0}^{\infty}a_k\lambda_i^k$$

也收敛。但 $|\lambda_i|=\rho(\mathbf{A})>r$，数项级数 $\sum_{k=0}^{\infty}a_kz^k$ 在收敛圆外是发散的。故 $\sum_{k=0}^{\infty}a_k\lambda_i^k$ 应该是发散的，因此矛盾，故结论（2）成立。

经过简单的变换便可得到如下推论:

推论3 设 $\sum_{k=0}^{\infty} a_k (z-z_0)^k$ 为收敛半径为$\boldsymbol{r}$的幂级数，$\boldsymbol{A}$为$\boldsymbol{n}$阶方阵，如果$\boldsymbol{A}$的特征值均落在收敛圆内，即 $|\lambda - z_0| < r$ ，其中λ 为**A**的任意特征值，则矩阵幂级数 $\sum_{k=0}^{\infty} a_k (\boldsymbol{A} - z_0 \boldsymbol{I})^k$ 绝对收敛； 若有某个 λ_{i_0} 使得$|\lambda_{i_0} - z_0| > r$，则幂级数 $\sum_{k=0}^{\infty} a_k (\boldsymbol{A} - z_0 \boldsymbol{I})^k$ 发散。

根据幂级数性质，幂级数的和函数是收敛圆内的解析函数(任意次可微，在任一点处均可展成Taylor级数），而一个圆内解析的函数可以展开成收敛的幂级数。于是，如果 $f(z)$ 是 $|z-z_0|<r$ 内的解析函数，其展成绝对收敛的幂级数为

$$f(z)=\sum_{k=0}^{\infty}a_k(z-z_0)^k$$

则当矩阵$\mathbf{A}\in\mathbf{C}^{n\times n}$的特征值落在收敛圆 $|z-z_0|<r$ 内时， 定义

$$f(A)\overset{\Delta}{=}\sum_{k=0}^{\infty}a_k(A-z_0I)^k$$

并称之为A关于解析函数$f(z)$的矩阵函数。

例如，解析函数 $$e^{z}=1+z+\frac{z^{2}}{2!}+\frac{z^{3}}{3!}+\cdots\quad(|z|<+\infty)$$

$$\sin z=z-\frac{z^{3}}{3!}+\frac{z^{5}}{5!}-\cdots\quad(|z|<+\infty)\qquad \cos z=1-\frac{z^{2}}{2!}+\frac{z^{4}}{4!}-\cdots\quad(|z|<+\infty)$$

$$(1-z)^{-1}=1+z+z^{2}+z^{3}+\cdots\,(|z|<1)\qquad \ln(1+z)=z-\frac{z^{2}}{2}+\frac{z^{3}}{3}-\cdots\quad(|z|<1)$$

矩阵指数函数为： $$e^{\mathbf{A}}=\mathbf{I}+\mathbf{A}+\frac{\mathbf{A}^{2}}{2!}+\frac{\mathbf{A}^{3}}{3!}+\cdots\quad(\rho(A)<+\infty)$$

矩阵的三角函数为：

$$\cos\mathbf{A}=\mathbf{I}-\frac{\mathbf{A}^{2}}{2!}+\frac{\mathbf{A}^{4}}{4!}-\cdots\quad(\rho(A)<+\infty)\qquad \sin\mathbf{A}=\mathbf{A}-\frac{\mathbf{A}^{3}}{3!}+\frac{\mathbf{A}^{5}}{5!}-\cdots\quad(\rho(A)<+\infty)$$

以及

$$(\mathbf{I}-\mathbf{A})^{-1}=\mathbf{I}+\mathbf{A}+\mathbf{A}^{2}+\mathbf{A}^{3}+\cdots(\rho(A)<1)\qquad \ln(\mathbf{I}+\mathbf{A})=\mathbf{A}-\frac{\mathbf{A}^{2}}{2}+\frac{\mathbf{A}^{3}}{3}-\cdots(\rho(A)<1)$$

现在利用矩阵的**Jordan**分解写出矩阵函数 $f(A)$ 的具体表达式

首先介绍一个引理

引理1 设 $f(z)=\sum_{k=0}^{\infty}a_k z^k$ 是收敛半径为r的幂级数，J是特征值为λ的n阶Jordan块阵，且 $|\lambda|<r$ ，则

$$f(J)=\begin{pmatrix} f(\lambda) & f'(\lambda) & \cdots & \dfrac{f^{(n-1)}(\lambda)}{(n-1)!} \\ & f(\lambda) & \ddots & \vdots \\ & & \ddots & f'(\lambda) \\ & & & f(\lambda) \end{pmatrix} \qquad (1)$$

证 根据定理2，$\rho(\boldsymbol{J})=|\lambda_{\max}|<r$ 故，矩阵级数 $\sum_{k=0}^{\infty}a_k\boldsymbol{J}^k$是收敛的，记其和函数为 $f(\boldsymbol{J})$。先考虑它的前m+1项部分和 $\mathbf{S}_{m+1}=\sum_{k=0}^{m}a_k\boldsymbol{J}^k$

因为 $\boldsymbol{J}=\lambda\boldsymbol{I}+\boldsymbol{N}$ 其中 $\boldsymbol{N}=\begin{pmatrix}0 & 1 & & \boldsymbol{0}\\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & 0\end{pmatrix}$，$\boldsymbol{N}^0=\boldsymbol{I}$ 且

k+1列

$$\boldsymbol{N}^k=\begin{pmatrix}0 & \cdots & 0 & 1 & & \boldsymbol{0}\\ & \ddots & & \ddots & \ddots & \\ & & \ddots & & \ddots & 1\\ & & & \ddots & & 0\\ & & & & \ddots & \vdots\\ & & & & & 0\end{pmatrix}$$

$n-k$行

$k\le n-1$，或 $\boldsymbol{N}^k=\mathbf{0}$，$k\ge n$

注意到：$(\lambda\boldsymbol{I})\boldsymbol{N}=\boldsymbol{N}(\lambda\boldsymbol{I})$，$\boldsymbol{J}^k=(\lambda\boldsymbol{I}+\boldsymbol{N})^k$

因此

$$\mathbf{J}^k=\sum_{i=0}^{k}C_k^i\lambda^{k-i}\mathbf{N}^i=\lambda^k\mathbf{I}_n+C_k^1\lambda^{k-1}\mathbf{N}+\cdots+C_k^i\lambda^{k-i}\mathbf{N}^i+\cdots$$

$$=\begin{pmatrix}
\lambda^k & C_k^1\lambda^{k-1} & \cdots & C_k^i\lambda^{k-i} & \cdots & C_k^{n-1}\lambda^{k-n+1} \\
 & \lambda^k & C_k^1\lambda^{k-1} & & \ddots & C_k^i\lambda^{k-i} \\
 & & \ddots & \ddots & & \vdots \\
 & & & \lambda^k & & C_k^1\lambda^{k-1} \\
 & & & & & \lambda^k
\end{pmatrix}$$

其中 $i\le\min\{k,n-1\}$， $C_k^i=\dfrac{k(k-1)\cdots(k-i+1)}{i!}$

进一步，有

$$\mathbf{S}_{m+1}=\sum_{k=0}^{m}a_k\boldsymbol{J}^k=\begin{pmatrix}\sum\limits_{k=0}^{m}a_k\lambda^k & \sum\limits_{k=0}^{m}a_kC_k^1\lambda^{k-1} & \cdots & \sum\limits_{k=0}^{m}a_kC_k^i\lambda^{k-i} & \sum\limits_{k=0}^{m}a_kC_k^{n-1}\lambda^{k-n+1}\\ & \sum\limits_{k=0}^{m}a_k\lambda^k & \sum\limits_{k=0}^{m}a_kC_k^1\lambda^{k-1} & \cdots & \sum\limits_{k=0}^{m}a_kC_k^i\lambda^{k-i}\\ & & \ddots & \ddots & \vdots\\ & & & \sum\limits_{k=0}^{m}a_k\lambda^k & \sum\limits_{k=0}^{m}a_kC_k^1\lambda^{k-1}\\ & & & & \sum\limits_{k=0}^{m}a_k\lambda^k\end{pmatrix}$$

注意到 $C_k^i=\dfrac{k(k-1)\cdots(k-i+1)}{i!}$，记 $S_{m+1}(\lambda)=\sum\limits_{k=0}^{m}a_k\lambda^k$，则

$$\sum_{k=0}^{m} a_k C_k^i \lambda^{k-i} = \sum_{k=i}^{m} a_k C_k^i \lambda^{k-i} = \frac{1}{i!}\sum_{k=i}^{m} a_k k(k-1)\cdots(k-i+1)\lambda^{k-i}$$

$$= \frac{1}{i!}\left[\sum_{k=i}^{m} a_k \left(\lambda^k\right)^{(i)}\right] = \frac{S_{m+1}^{(i)}(\lambda)}{i!}$$

其中 $S_{m+1}^{(i)}(\lambda)$ 表示 $S_{m+1}(\lambda)$ 对λ的 i阶导数。

$$S'_{m+1}(\lambda) = \sum_{k=i}^{m} a_k \cdot \left(\lambda^k\right)' = \sum_{k=i}^{m} a_k \cdot k \cdot \lambda^{k-1}$$

$$S''_{m+1}(\lambda) = \sum_{k=i}^{m} a_k \cdot k \cdot (k-1)\lambda^{k-2}$$

$$S_{m+1}^{(i)}(\lambda) = \sum_{k=i}^{m} a_k k(k-1)\cdots(k-i+1)\lambda^{k-i}$$ ，故

$$f(\boldsymbol{J})=\lim_{m\to\infty}\mathbf{S}_{m+1}=\begin{pmatrix} S_{m+1}(\lambda) & S'_{m+1}(\lambda) & \cdots & \dfrac{S^{(i)}_{m+1}(\lambda)}{i!} & \cdots & \dfrac{S^{(n-1)}_{m+1}(\lambda)}{(n-1)!} \\ & S_{m+1}(\lambda) & S'_{m+1}(\lambda) & & \ddots & \\ & & \ddots & \ddots & & \dfrac{S^{(i)}_{m+1}(\lambda)}{i!} \\ & & & \ddots & \ddots & \vdots \\ & & & & \ddots & S'_{m+1}(\lambda) \\ & & & & & S_{m+1}(\lambda) \end{pmatrix}$$

根据幂级数性质知，有 $\lim\limits_{m\to\infty} S^{(i)}_{m+1}(\lambda)=f^{(i)}(\lambda)$ ，因此有

推论4 设 $f(z)=\sum_{k=0}^{\infty}a_k(z-z_0)^k$是收敛半径为$r$的幂级数，$\boldsymbol{J}$是特征值为$\lambda$ 的Jordan块，且 $|\lambda-z_0|<r$ ，则

$$f(\mathbf{J})=\begin{pmatrix} f(\lambda) & f'(\lambda) & \cdots & \dfrac{f^{(n-1)}(\lambda)}{(n-1)!} \\ & f(\lambda) & \ddots & \vdots \\ & & \ddots & f'(\lambda) \\ & & & f(\lambda) \end{pmatrix}$$

推论5 设 $f(z)=\sum_{k=0}^{\infty}a_k z^k$ 是收敛半径为***r***的幂级数，***J***是特征值为 λ 的***n***阶**Jordan**块阵，且 $|t\lambda|<r$，则

$$f(t\mathbf{J})=\begin{pmatrix} f(t\lambda) & t f'(t\lambda) & \cdots & \dfrac{t^{n-1}f^{(n-1)}(t\lambda)}{(n-1)!} \\ & f(t\lambda) & \ddots & \vdots \\ & & \ddots & t f'(t\lambda) \\ & & & f(t\lambda) \end{pmatrix}$$

根据上面的引理和矩阵级数的性质，有

定理3 设 $f(z)=\sum_{k=0}^{\infty}a_k(z-z_0)^k$为收敛半径为$r$的幂级数，$A$为$n$阶方阵，$A=TJT^{-1}$为其Jordan分解，$J=\mathrm{diag}(J_1, J_2, \cdots, J_s)$。当$A$的特征值均落在收敛圆内时，即 $|\lambda - z_0|< r$ ，其中λ为A的任意特征值，则矩阵幂级数 $\sum_{k=0}^{\infty}a_k(A-z_0I)^k$ 绝对收敛， 并且和矩阵为

$$f(A)= T\ \mathrm{diag}\left(f(J_1), f(J_2), \cdots, f(J_s)\right) T^{-1}$$

其中$f(J_i)$的定义如表达式（1）。

例1 设 $\mathbf{A}=\begin{pmatrix}3 & 0 & 8\\ 3 & -1 & 6\\ -2 & 0 & -5\end{pmatrix}$，求 $\sin A$

解 根据矩阵A的Jordan分解

$$\mathbf{A}=TJT^{-1}=\begin{pmatrix}0 & 4 & 1\\ 1 & 3 & 0\\ 0 & -2 & 0\end{pmatrix}\begin{pmatrix}-1 & & \\ & -1 & 1\\ & & -1\end{pmatrix}\begin{pmatrix}0 & 1 & \frac{3}{2}\\ 0 & 0 & -\frac{1}{2}\\ 1 & 0 & 2\end{pmatrix}$$

因此，由

$$\sin A=f(A)=T\ \mathbf{diag}\Big(f(J_1),\quad f(J_2)\Big)\ T^{-1}$$

$$=\begin{pmatrix}0 & 4 & 1\\ 1 & 3 & 0\\ 0 & -2 & 0\end{pmatrix}\begin{pmatrix}-\sin 1 & & \\ & -\sin 1 & \cos 1\\ & & -\sin 1\end{pmatrix}\begin{pmatrix}0 & 1 & \frac{3}{2}\\ 0 & 0 & -\frac{1}{2}\\ 1 & 0 & 2\end{pmatrix}=\begin{pmatrix}4\cos 1-\sin 1 & 0 & 8\cos 1\\ 3\cos 1 & -\sin 1 & 6\cos 1\\ -2\cos 1 & 0 & -\sin 1-4\cos 1\end{pmatrix}$$

例2 设 $\mathbf{A}=\begin{pmatrix}3&1&-1\\1&2&-1\\2&1&0\end{pmatrix}$，求 e^{At}

解 根据矩阵的**Jordan** 分解

$$\mathbf{A}=\begin{pmatrix}0&1&1\\1&0&1\\1&1&1\end{pmatrix}\begin{pmatrix}1&&\\&2&1\\&&2\end{pmatrix}\begin{pmatrix}-1&0&1\\0&-1&1\\1&1&-1\end{pmatrix}$$ 则

$$e^{At}=f(At)=T\ \mathbf{diag}\big(f(tJ_1),\quad f(tJ_2)\big)\ T^{-1}$$

$$e^{At}=\begin{pmatrix}0&1&1\\1&0&1\\1&1&1\end{pmatrix}\begin{pmatrix}e^t&&\\&e^{2t}&te^{2t}\\&&e^{2t}\end{pmatrix}\begin{pmatrix}-1&0&1\\0&-1&1\\1&1&-1\end{pmatrix}=\begin{pmatrix}(1+t)e^{2t}&te^{2t}&-te^{2t}\\e^{2t}-e^t&e^{2t}&-e^{2t}+e^t\\(1+t)e^{2t}-e^t&te^{2t}&e^t-te^{2t}\end{pmatrix}$$

有限待定系数法

思路　设 $\psi(\lambda)=\det(\lambda I-A)$ 为A 的特征多项式，若

$$f(\lambda)=p(\lambda)\psi(\lambda)+q(\lambda),\qquad 其中\deg(q)\le n-1$$

则 $f(A)=p(A)\psi(A)+q(A)=q(A)$

$f(At)$ 的求法类似，这时$f(\lambda t)$改写为

$$f(\lambda t)=p(\lambda,t)\psi(\lambda)+q(\lambda,t),$$

其中$p(\lambda,t)$为含参数t的幂级数，$q(\lambda,t)$是含参数t且次数不超过$n-1$的多项式，则

$$f(At)=p(A,t)\psi(A)+q(A,t)=q(A,t)$$

求解的关键是求出**q(A)**或**q(A,t)**，可采用待定系数法求。

$q(\lambda)$和$q(\lambda,t)$的求法

设 $\psi(\lambda)=\det(\lambda \boldsymbol{I}-\boldsymbol{A})=(\lambda-\lambda_1)^{m_1}(\lambda-\lambda_2)^{m_2}\cdots(\lambda-\lambda_s)^{m_s}$

$$q(\lambda,t)=b_{n-1}(t)\lambda^{n-1}+b_{n-2}(t)\lambda^{n-2}+\cdots+b_0(t)$$

由于$\psi^{(j)}(\lambda_i)=0, j=0,1,\ldots,m_i-1; i=1,2,\ldots,s,$ 则

$$\left.\frac{\mathrm{d}^j}{\mathrm{d}\lambda^j}f(\lambda t)\right|_{\lambda=\lambda_i}=\left.\frac{\mathrm{d}^j}{\mathrm{d}\lambda^j}q(\lambda,t)\right|_{\lambda=\lambda_i}$$

即

$$\left.t^j\frac{\mathrm{d}^j}{\mathrm{d}u^j}f(u)\right|_{u=\lambda_i t}=\left.\frac{\mathrm{d}^j}{\mathrm{d}\lambda^j}q(\lambda,t)\right|_{\lambda=\lambda_i}$$

通过求解方程组，可立即得到$b_0(t),b_1(t),\cdots,b_{n-1}(t)$

计算矩阵函数$f(A)$ 和$f(At)$的步骤

(1) 求矩阵A的特征多项式

(2) 用待定系数法求$q(\lambda)$ 或$q(\lambda, t)$

(3) 通过 $f(A) = p(A)(f(At) = p(A, t))$ 计算 $f(A)(f(At))$.

例3 用待定系数法求 (1) $\sin(A)$ (2) e^{At} 其中

$$\mathbf{A}=\begin{pmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & -5 \end{pmatrix}.$$

解 先求特征多项式

$$\det(\lambda \boldsymbol{I}-\boldsymbol{A})=\begin{vmatrix} \lambda-3 & 0 & -8 \\ -3 & \lambda+1 & -6 \\ 2 & 0 & \lambda+5 \end{vmatrix}=(\lambda+1)^3$$

(1) 设$q(\lambda)=b_2\lambda^2+b_1\lambda+b_0, f(\lambda)=\sin(\lambda)$. 则有

$$\begin{cases} q(-1)=b_2-b_1+b_0=-\sin 1=f(-1) \\ q'(-1)=-2b_2+b_1=\cos 1=f'(-1) \\ q''(-1)=2b_2=\sin 1=f''(-1) \end{cases}$$

解得

$$\begin{cases} b_1=\cos 1-\dfrac{1}{2}\sin 1 \\ b_1=\sin 1+\cos 1 \\ b_2=\dfrac{1}{2}\sin 1 \end{cases}$$

$$\sin A = b_2 \boldsymbol{A}^2 + b_1 \boldsymbol{A} + b_0 \boldsymbol{I}$$

$$= \frac{\sin 1}{2} A^2 + (\sin 1 + \cos 1) A + (\cos 1 - \tfrac{1}{2}\sin 1) I$$

$$= \begin{pmatrix} 4\cos 1 - \sin 1 & 0 & 8\cos 1 \\ 3\cos 1 & -\sin 1 & 6\cos 1 \\ -2\cos 1 & 0 & -\sin 1 - 4\cos 1 \end{pmatrix}$$

(2) 设$q(\lambda) = b_2\lambda^2 + b_1\lambda + b_0, f(\lambda t) = e^{\lambda t}$. 则有

$$
\begin{cases}
q(-1) = b_2 - b_1 + b_0 = e^{-t} = f(-t) \\
q'(-1) = -2b_2 + b_1 = te^{-t} = tf'(-t) \\
q''(-1) = 2b_2 = t^2 e^{-t} = t^2 f''(-t)
\end{cases}
$$

解得

$$
\begin{cases}
b_0 = \left(1 + t + \dfrac{t^2}{2}\right) e^{-t} \\
b_1 = te^{-t} + t^2 e^{-t} \\
b_2 = \dfrac{t^2}{2} e^{-t}
\end{cases}
$$

$$e^{At}=b_2\boldsymbol{A}^2+b_1\boldsymbol{A}+b_0\boldsymbol{I}$$

$$=\frac{t^2}{2}e^{-t}A^2+\left(te^{-t}+t^2e^{-t}\right)A+\left(1+t\ +\frac{t^2}{2}\right)e^{-t}I$$

$$=\begin{pmatrix}(1+4t)e^{-t} & 0 & 8te^{-t}\\ 3te^{-t} & e^{-t} & 6te^{-t}\\ -2te^{-t} & 0 & (1-4t)e^{-t}\end{pmatrix}$$

我们还可以证明

（Ⅰ）$\forall \mathbf{A}\in \mathbf{C}^{n\times n}$，总有

（1）$\sin(-\mathbf{A})=-\sin\mathbf{A}$， $\cos(-\mathbf{A})=\cos\mathbf{A}$

（2）$e^{i\mathbf{A}}=\cos\mathbf{A}+i\sin\mathbf{A}$，$\cos\mathbf{A}=\frac{1}{2}(e^{i\mathbf{A}}+e^{-i\mathbf{A}})$， $\sin\mathbf{A}=\frac{1}{2}(e^{i\mathbf{A}}-e^{-i\mathbf{A}})$

（Ⅱ）$\mathbf{A},\mathbf{B}\in \mathbf{C}^{n\times n}$， 且 $\mathbf{AB}=\mathbf{BA}$ ，则

（0） $e^{A}e^{B}=e^{B}e^{A}=e^{A+B}$

（1） $\sin(\mathbf{A}+\mathbf{B})=\sin\mathbf{A}\cos\mathbf{B}+\cos\mathbf{A}\sin\mathbf{B}$

（2） $\cos(\mathbf{A}+\mathbf{B})=\cos\mathbf{A}\cos\mathbf{B}-\sin\mathbf{A}\sin\mathbf{B}$

若A=*B*，则 $\cos 2\mathbf{A}=\cos^2\mathbf{A}-\sin^2\mathbf{A}$， $\sin 2\mathbf{A}=2\sin\mathbf{A}\cos\mathbf{A}$

需要指出的是，对任何n阶方阵A，e^A总是可逆矩阵。

$$e^{A}e^{-A}=e^{-A}e^{A}=e^{O}=I$$

$\sin A$和$\cos A$不一定可逆. 例如 $A=\begin{pmatrix}\pi & 0\\ 0 & \dfrac{\pi}{2}\end{pmatrix}$

$$\sin A=\begin{pmatrix}\sin\pi & 0\\ 0 & \sin\dfrac{\pi}{2}\end{pmatrix}=\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}$$

$$\cos A=\begin{pmatrix}\cos\pi & 0\\ 0 & \cos\dfrac{\pi}{2}\end{pmatrix}=\begin{pmatrix}-1 & 0\\ 0 & 0\end{pmatrix}$$

当 $AB \neq BA$ 时，$e^{A+B}=e^Ae^B$ 或 $e^{A+B}=e^Be^A$ 不成立。

$$\boldsymbol{A}=\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix} \quad \boldsymbol{B}=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix} \qquad A+B=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$$

则

$$AB=\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\neq\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}=BA$$

容易计算

$$e^B=\begin{pmatrix}e^0 & \left.\left(\dfrac{\mathbf{d}e^t}{\mathbf{dt}}\right)\right|_{t=0}\\ 0 & e^0\end{pmatrix}=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, \quad e^A=e^{B^T}=(e^B)^T=\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}$$

显然

$$e^{A}e^{B}=\begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix}\neq\begin{pmatrix}2 & 1\\ 1 & 1\end{pmatrix}=e^{B}e^{A}$$

另外，

$$\boldsymbol{A}+\boldsymbol{B}=\boldsymbol{T}\boldsymbol{J}\boldsymbol{T}^{-1}=\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2}\end{pmatrix}\begin{pmatrix}-1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}$$

$$e^{\boldsymbol{A}+\boldsymbol{B}}=Te^{J}T^{-1}=\frac{1}{2}\begin{pmatrix}e+e^{-1} & e-e^{-1}\\ e-e^{-1} & e+e^{-1}\end{pmatrix}$$

因此 $e^{\boldsymbol{A}+\boldsymbol{B}}\neq e^{\boldsymbol{A}}e^{\boldsymbol{B}}$， $e^{\boldsymbol{A}+\boldsymbol{B}}\neq e^{\boldsymbol{B}}e^{\boldsymbol{A}}$

三、矩阵的微积分

在研究微分方程组时，为了简化对问题的表达及求解过程，需要考虑以函数为元素的矩阵的微分和积分；在研究优化等问题时，则要碰到数量函数对向量变量或矩阵变量的导数，以及向量值或矩阵值函数对向量变量或矩阵变量的导数。

1 相对于数量变量的微分和积分

定义5 如果矩阵 $\mathbf{A}(t)=\left(a_{ij}(t)\right)_{m\times n}$ 的每一个元素 $a_{ij}(t)$ $i=1,2,\cdots,m;$ $j=1,2,\cdots,n$ 在$[a,b]$上均为变量t的可微函数， 则称$A(t)$可微，且导数定义为

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

例如

$$\mathbf{A}(t)=\begin{pmatrix} t+e^t & \sin t \\ t & 4 \end{pmatrix},\quad \text{则}\quad A'(t)=\begin{pmatrix} 1+e^t & \cos t \\ 1 & 0 \end{pmatrix}$$

由定义 5 可以验证矩阵导数的如下运算性质.

定理4 设$A(t)$、$B(t)$是可进行运算的两个可微矩阵，则以下的运算规则成立

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

（2）$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{A}(t)\mathbf{B}(t))=(\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{A}(t))\cdot\mathbf{B}(t)+\mathbf{A}(t)\cdot(\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{B}(t))$

（3）$\frac{\mathrm{d}}{\mathrm{d}t}(\alpha\mathbf{A}(t))=\alpha\cdot\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{A}(t)$，其中$\alpha$为任意常数

（4）当$u=f(t)$关于t可微时，有

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

（5）当$A^{-1}(t)$为可微矩阵时，有

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{A}^{-1}(t))=-\mathbf{A}^{-1}(t)(\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{A}(t))\mathbf{A}^{-1}(t)$$

由于 $\frac{\mathrm{d}}{\mathrm{d}t}(A(t))$ 仍是函数矩阵，如果它仍是可导函数矩阵，则可定义其二阶导数。不难给出函数矩阵的高阶导数：

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

证：（2）设 $\mathbf{A}(t)=\left(a_{ij}(t)\right)_{m\times n}$，$\boldsymbol{B}(t)=\left(b_{ij}(t)\right)_{n\times p}$ 则

$$\frac{d}{dt}\left(\boldsymbol{A}(t)\boldsymbol{B}(t)\right)=\frac{d}{dt}\left(\sum_{k=1}^{n}a_{ik}(t)b_{kj}(t)\right)_{m\times p}=\left(\sum_{k=1}^{n}\left[\frac{d}{dt}\left(a_{ik}(t)b_{kj}(t)\right)\right]\right)_{m\times p}$$

$$=\left[\sum_{k=1}^{n}\left[\frac{d}{dt}\left(a_{ik}(t)\right)\cdot b_{kj}(t)+a_{ik}(t)\cdot\frac{d}{dt}\left(b_{kj}(t)\right)\right]\right]_{m\times p}$$

$$=\left[\sum_{k=1}^{n}\left(\frac{d}{dt}\left(a_{ik}(t)\right)\right)\cdot b_{kj}(t)\right]_{m\times p}+\left[\sum_{k=1}^{n}a_{ik}(t)\cdot\left(\frac{d}{dt}\left(b_{kj}(t)\right)\right)\right]_{m\times p}$$

$$=\frac{d}{dt}\left(\boldsymbol{A}(t)\right)\boldsymbol{B}(t)+\boldsymbol{A}(t)\frac{d}{dt}\boldsymbol{B}(t)$$

（5）由于$A(t)^{-1}A(t)=I$，两端对t求导得

从而 $$\frac{d}{dt}\left(A^{-1}(t)\right)A(t)+A^{-1}(t)\frac{d}{dt}\left(A(t)\right)=O,\quad \frac{d}{dt}\left(A^{-1}(t)\right)=-A^{-1}(t)\frac{d}{dt}\left(A(t)\right)A^{-1}(t)$$

注： $\frac{d}{dt}\left(A^{m}(t)\right)=m\,A^{m-1}(t)\,\frac{d}{dt}\left(A(t)\right)$ 不一定成立。

例如，$m=2$

取 $A(t)=\begin{pmatrix} t^2 & t \\ 0 & t \end{pmatrix}$，$A^2(t)=\begin{pmatrix} t^4 & t^3+t^2 \\ 0 & t^2 \end{pmatrix}$，则 $\frac{d}{dt}\left(A(t)\right)=\begin{pmatrix} 2t & 1 \\ 0 & 1 \end{pmatrix}$，

$$\frac{d}{dt}\left(A^2(t)\right)=\begin{pmatrix} 4t^3 & 3t^2+2t \\ 0 & 2t \end{pmatrix},\quad 2A(t)\frac{d}{dt}\left(A(t)\right)=2\begin{pmatrix} t^2 & t \\ 0 & t \end{pmatrix}\begin{pmatrix} 2t & 1 \\ 0 & 1 \end{pmatrix}=\begin{pmatrix} 4t^3 & 2t^2+2t \\ 0 & 2t \end{pmatrix}$$

故$\frac{d}{dt}\left(A^2(t)\right)\neq 2A(t)\frac{d}{dt}\left(A(t)\right)$。只有当$A(t)\frac{d}{dt}\left(A(t)\right)=\frac{d}{dt}\left(A(t)\right)A(t)$时成立。

定理5 设n阶方阵A与t无关，则有

（1） $$\frac{\mathrm{d}}{\mathrm{d}t}e^{t\mathbf{A}} = \mathbf{A}e^{t\mathbf{A}} = e^{t\mathbf{A}}\mathbf{A}$$

（2） $$\frac{\mathrm{d}}{\mathrm{d}t}\sin(t\mathbf{A}) = \mathbf{A}\cdot\cos(t\mathbf{A}) = \cos(t\mathbf{A})\cdot\mathbf{A}$$

（3） $$\frac{\mathrm{d}}{\mathrm{d}t}\cos(t\mathbf{A}) = -\mathbf{A}\cdot\sin(t\mathbf{A}) = -\sin(t\mathbf{A})\cdot\mathbf{A}$$

证， 只证（1），（2）和（3）的证明与（1）类似。

由 $e^{tA} = \sum_{k=0}^{\infty}\frac{t^k}{k!}A^k$ 并利用绝对收敛的级数可以逐项求导的性质得

$$\frac{d\left(e^{tA}\right)}{dt} = \frac{d}{dt}\left(\sum_{k=0}^{\infty}\frac{t^k}{k!}A^k\right) = \left(\sum_{k=0}^{\infty}\frac{d}{dt}\left(\frac{t^k}{k!}\right)A^k\right) = \sum_{k=1}^{\infty}\frac{t^{k-1}}{(k-1)!}A^k$$

$$= A\left(\sum_{k=1}^{\infty}\frac{t^{k-1}}{(k-1)!}A^{k-1}\right) = \left(\sum_{k=1}^{\infty}\frac{t^{k-1}}{(k-1)!}A^{k-1}\right)A \Rightarrow Ae^{tA} = e^{tA}A$$

定义6 如果矩阵 $\mathbf{A}(t)=\left(a_{ij}(t)\right)_{m\times n}$ 的每一个元素$a_{ij}(t)$都是区间$[t_0, t_1]$上的可积函数，则定义 $A(t)$ 在区间$[t_0, t_1]$上的积分为

$$\int_{t_0}^{t_1}\mathbf{A}(t)dt=\left(\int_{t_0}^{t_1}a_{ij}(t)dt\right)_{m\times n}$$

容易验证如下运算法则成立

（1）$\int_{t_0}^{t_1}\left(\alpha A(t)+\beta B(t)\right)dt=\alpha\int_{t_0}^{t_1}A(t)\,dt+\beta\int_{t_0}^{t_1}B(t)\,dt \quad \forall\alpha,\beta\in\mathbf{C}$

（2）$\int_{t_0}^{t_1}\left(A(t)\,B\right)dt=\int_{t_0}^{t_1}A(t)\,dt\;B$， 其中$B$为常数矩阵；

$\int_{t_0}^{t_1}\left(A\;B(t)\right)dt=A\int_{t_0}^{t_1}B(t)\,dt$， 其中$A$为常数矩阵；

（3）当$A(t)$在$[a，b]$上连续可微时，对任意 $t\in(a,b)$，有

$$\frac{d}{dt}\left(\int_a^t A(\tau)d\tau\right)=A(t)$$

（4）当$A(t)$在$[a，b]$上连续可微时，对任意 $t\in(a,b)$，有

$$\int_a^b \frac{d\left(A(t)\right)}{dt}dt= A(b)-A(a)$$

2 相对于矩阵变量的微分

定义7 设 $X=\left(x_{ij}\right)_{m\times n}$，函数 $f(\mathbf{X})=f(x_{11},x_{12},\cdots x_{1n},x_{21},\cdots,x_{mn})$ 为mn元的多元函数，且 $\frac{\partial f}{\partial x_{ij}}$ $(i=1,2,\cdots,m;\ \ j=1,2,\cdots,n)$ 都存在，定义 $f(X)$ 对矩阵X的导数为

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

例2 设 $\boldsymbol{x}=(\xi_1,\xi_2,\cdots\xi_n)^T$, n元函数 $f(\boldsymbol{x})=f(\xi_1,\xi_2,\cdots\xi_n)$，求

$\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}^T}$, $\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}$.

解 根据定义有

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}^T}=\left(\frac{\partial f}{\partial\xi_1},\frac{\partial f}{\partial\xi_2},\cdots,\frac{\partial f}{\partial\xi_n}\right)$$

特别地，以$\boldsymbol{x}$为自变量的函数的导数，一般记为：

$$\nabla f(\boldsymbol{x})=\frac{\mathbf{d}f}{\mathbf{d}\boldsymbol{x}}=\left(\frac{\partial f}{\partial\xi_1},\frac{\partial f}{\partial\xi_2},\cdots,\frac{\partial f}{\partial\xi_n}\right)^T$$

称为数量函数对向量变量的导数，即为高等数学学过的函数的梯度向量，也记为**grad** f

例3 设 $\boldsymbol{x}=(\xi_1,\xi_2,\cdots\xi_n)^T$, $\mathbf{A}=\left(a_{ij}\right)_{n\times n}$, n元函数 $f(\boldsymbol{x})=\boldsymbol{x}^T\mathbf{A}\boldsymbol{x}$，

求 $\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}$。

解 因

$$f(\boldsymbol{x})=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}\xi_i\,\xi_j=\xi_1\sum_{j=1}^{n}a_{1j}\xi_j+\cdots+\xi_k\sum_{j=1}^{n}a_{kj}\xi_j+\cdots+\ \xi_n\sum_{j=1}^{n}a_{nj}\xi_j$$

所以

$$\frac{\partial f(\boldsymbol{x})}{\partial\xi_k}=\ \xi_1a_{1k}+\ldots+\xi_{k-1}a_{k-1,k}+\left(\sum_{j=1}^{n}a_{kj}\xi_j+\xi_ka_{kk}\right)+\xi_{k+1}a_{k+1,k}+\ldots+\xi_na_{nk}$$

$$=\sum_{i=1}^{n}a_{ik}\xi_i+\sum_{j=1}^{n}a_{kj}\xi_j,\qquad k=1,2,\cdots,n$$

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=\begin{pmatrix}\frac{\partial f}{\partial \xi_1}\\ \frac{\partial f}{\partial \xi_2}\\ \vdots \\ \frac{\partial f}{\partial \xi_n}\end{pmatrix}=\begin{pmatrix}\sum_{j=1}^{n} a_{1j}\xi_j\\ \sum_{j=1}^{n} a_{2j}\xi_j\\ \vdots\\ \sum_{j=1}^{n} a_{nj}\xi_j\end{pmatrix}+\begin{pmatrix}\sum_{i=1}^{n} a_{i1}\xi_i\\ \sum_{i=1}^{n} a_{i2}\xi_i\\ \vdots\\ \sum_{i=1}^{n} a_{in}\xi_i\end{pmatrix}$$

$$=\mathbf{A}\boldsymbol{x}+\mathbf{A}^T\boldsymbol{x}=(\mathbf{A}+\mathbf{A}^T)\boldsymbol{x}$$

特别地，当A时对称矩阵时，

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=2Ax$$

例 设 $\boldsymbol{a}=(a_1,a_2,\cdots,a_n)^T$ 为常向量，$\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n)^T$ 为向量变量，且 $f(\boldsymbol{x})=(\boldsymbol{x},\boldsymbol{a})=\boldsymbol{a}^T\boldsymbol{x}=\boldsymbol{x}^T\boldsymbol{a}$。求 $\dfrac{\partial f}{\partial \boldsymbol{x}}$。

解，由于

$$f(\boldsymbol{x})=\sum_{i=1}^{n}a_i\xi_i\,,\quad \frac{\partial f}{\partial \xi_j}=a_j\,,\quad (j=1,2,\cdots,n)$$ 所以

$$\frac{\partial f}{\partial \boldsymbol{x}}=\begin{pmatrix}\dfrac{\partial f}{\partial \xi_1}\\ \dfrac{\partial f}{\partial \xi_2}\\ \vdots\\ \dfrac{\partial f}{\partial \xi_n}\end{pmatrix}=\begin{pmatrix}a_1\\ a_2\\ \vdots\\ a_n\end{pmatrix}=\boldsymbol{a}$$

例 设 $\boldsymbol{A}=(a_{ij})_{m\times n}$ 为常矩阵，$\boldsymbol{X}=(x_{ij})_{n\times m}$ 为矩阵变量，且 $f(\boldsymbol{X})=\mathbf{tr}(\boldsymbol{AX})$，求 $\dfrac{\partial f}{\partial \boldsymbol{X}}$。

解，由于 $\boldsymbol{AX}=\left(\sum_{k=1}^{n} a_{ik}x_{kj}\right)_{m\times m}$，所以 $f(\boldsymbol{X})=\mathbf{tr}(\boldsymbol{AX})=\sum_{s=1}^{m}\sum_{k=1}^{n} a_{sk}x_{ks}$

而

$$\left(\frac{\partial f}{\partial x_{ij}}\right)_{n\times m}=(a_{ji})_{n\times m}\quad (i=1,2,\cdots,n\quad j=1,2,\cdots,m),$$

故

$$\frac{\partial f}{\partial \boldsymbol{X}}=\left(\frac{\partial f}{\partial x_{ij}}\right)_{n\times m}=(a_{ji})_{n\times m}=\boldsymbol{A}^T$$

例 设 $A\in\mathbf{R}^{m\times n}$，$\boldsymbol{b}\in\mathbf{R}^{m}$， $\boldsymbol{x}\in\mathbf{R}^{n}$， $f(\boldsymbol{x})=\|A\boldsymbol{x}-\boldsymbol{b}\|_2^2$，试求 $\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}$

解：因为

$$f(\boldsymbol{x})=\|A\boldsymbol{x}-\boldsymbol{b}\|_2^2=(A\boldsymbol{x}-\boldsymbol{b}，A\boldsymbol{x}-\boldsymbol{b})=(A\boldsymbol{x}-\boldsymbol{b})^T(A\boldsymbol{x}-\boldsymbol{b})$$

$$=(\boldsymbol{x}^TA^T-\boldsymbol{b}^T)(A\boldsymbol{x}-\boldsymbol{b})$$

$$=\boldsymbol{x}^TA^TA\boldsymbol{x}-\boldsymbol{b}^TA\boldsymbol{x}-\boldsymbol{x}^TA^T\boldsymbol{b}+\boldsymbol{b}^T\boldsymbol{b}$$

$$=\boldsymbol{x}^T\left(A^TA\right)\boldsymbol{x}-\left(A^T\boldsymbol{b}\right)^T\boldsymbol{x}-\boldsymbol{x}^T\left(A^T\boldsymbol{b}\right)+\boldsymbol{b}^T\boldsymbol{b}$$

从而由前面的结论可得

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=2A^TA\boldsymbol{x}-A^T\boldsymbol{b}-A^T\boldsymbol{b}=2\left(A^TA\boldsymbol{x}-A^T\boldsymbol{b}\right)$$

第三章结束

3　矩阵函数在微分方程中的应用

在线性控制系统中，常常涉及求解线性微分方程组的问题。矩阵函数在其中有重要的应用。

我们首先讨论一阶线性常系数齐次微分方程组的定解问：

$$\begin{cases} \dfrac{\mathbf{d}x_1(t)}{\mathbf{d}t} = a_{11}x_1(t) + a_{12}x_2(t) + \cdots + a_{1n}x_n(t) \\ \dfrac{\mathbf{d}x_2(t)}{\mathbf{d}t} = a_{21}x_1(t) + a_{22}x_2(t) + \cdots + a_{2n}x_n(t) \\ \qquad\qquad\vdots \\ \dfrac{\mathbf{d}x_n(t)}{\mathbf{d}t} = a_{n1}x_1(t) + a_{n2}x_2(t) + \cdots + a_{nn}x_n(t) \end{cases}$$

给定初始条件：$x_i(0)$，$(i=1,2,\cdots,n)$，记　$\mathbf{A} = (a_{ij}) \in \mathbf{C}^{n\times n}$，

$$\boldsymbol{X}(0) = \left(x_1(0), x_2(0), \cdots, x_n(0)\right)^T$$

$$\boldsymbol{X}(t) = (x_1(t), x_2(t), \ldots, x_n(t))^T,$$

则上述微分方程组可写成：

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{X}(t) & (2) \\ \boldsymbol{X}(0) = (x_1(0), x_2(0), \ldots, x_n(0))^T & (3) \end{cases}$$

利用矩阵微分的性质有

$$\frac{\mathrm{d}(e^{-\mathbf{A}t}\boldsymbol{X}(t))}{\mathrm{d}t} = \frac{\mathrm{d}e^{-\mathbf{A}t}}{\mathrm{d}t} \cdot \boldsymbol{X}(t) + e^{-\mathbf{A}t}\frac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t}$$

$$= -e^{-\mathbf{A}t}\mathbf{A} \cdot \boldsymbol{X}(t) + e^{-\mathbf{A}t}\frac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t} = e^{-\mathbf{A}t}\left(\frac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t} - \mathbf{A}\boldsymbol{X}(t)\right)$$

方程（2）意味着

$$\frac{\mathrm{d}(e^{-\mathbf{A}t}\boldsymbol{X}(t))}{\mathrm{d}t} = \boldsymbol{0}$$

因此**X**=e^{At}C， 其中C为常数向量，由初始条件（3），$\boldsymbol{C} = \boldsymbol{X}(0)$

$$\boldsymbol{X}(t) = e^{\mathbf{A}t}\boldsymbol{X}(0) \qquad (4)$$

下面说明解的唯一性。

如果定解问题（2）和(3)有两个解$\boldsymbol{X}_1(t)$，$\boldsymbol{X}_2(t)$，则令$\boldsymbol{Y}(t)=\boldsymbol{X}_1(t)-\boldsymbol{X}_2(t)$，显然满足

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{Y}(t)}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{Y}(t) \\ \boldsymbol{Y}(0)=\boldsymbol{X}_1(0)-\boldsymbol{X}_2(0)=\boldsymbol{0} \end{cases}$$

由上述推导可知，$\boldsymbol{Y}(t)=e^{\mathrm{A}t}\boldsymbol{Y}(0)=\boldsymbol{0}$，即$\boldsymbol{X}_1(t)=\boldsymbol{X}_2(t)$。

综上所述，

定理6 一阶线性常系数齐次微分方程组的定解问题（2）**(3)** 有唯一解$\boldsymbol{X}(t)=\mathrm{e}^{At}\boldsymbol{X}(0)$。

最后我们考虑一阶线性常系数非齐次微分方程组的定解问题

$$\begin{cases} \dfrac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{X}(t) + \boldsymbol{F}(t) & (4) \\ \boldsymbol{X}(t_0) = (x_1(t_0), x_2(t_0), \ldots, x_n(t_0))^T & (5) \end{cases}$$

这里 $\boldsymbol{F}(t) = (f_1(t), f_2(t), \cdots, f_n(t))^T$ 是已知向量函数，$\boldsymbol{A}$和$\boldsymbol{X}$意义同前。

改写方程为 并以$\boldsymbol{e}^{-At}$左乘方程两边，即

$$e^{-\mathbf{A}t}\left(\frac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t} - \boldsymbol{A}\boldsymbol{X}(t)\right) = e^{-\mathbf{A}t}\boldsymbol{F}(t)$$

即 $\dfrac{\mathrm{d}(e^{-\mathbf{A}t}\boldsymbol{X}(t))}{\mathrm{d}t} = e^{-\mathbf{A}t}\boldsymbol{F}(t)$ 对此方程在$[\boldsymbol{t_0}, \boldsymbol{t}]$上进行积分，可得

$$e^{-\mathbf{A}t}\boldsymbol{X}(t) - e^{-\mathbf{A}t_0}\boldsymbol{X}(t_0) = \int_{t_0}^{t} e^{-\mathbf{A}\tau}\boldsymbol{F}(\tau)\mathrm{d}\tau$$

$$e^{-At}\boldsymbol{X}(t) = e^{-At_0}\boldsymbol{X}(t_0) + \int_{t_0}^{t} e^{-A\tau}\boldsymbol{F}(\tau)\mathbf{d}\tau$$

$$\boldsymbol{X}(t) = e^{\mathbf{A}(t-t_0)}\boldsymbol{X}(t_0) + \int_{t_0}^{t} e^{\mathbf{A}(t-\tau)}\boldsymbol{F}(\tau)\mathrm{d}\tau$$ 就是上述定解问题的解。

例8 求定解问题 $\begin{cases} \dfrac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t} = \mathbf{A}\boldsymbol{X}(t) \\ \boldsymbol{X}(0) = (1,1,1)^T \end{cases}$ 其中 $\mathbf{A} = \begin{pmatrix} 3 & -1 & 1 \\ 2 & 0 & -1 \\ 1 & -1 & 2 \end{pmatrix}$

解 $\det(\lambda\mathbf{I} - \mathbf{A}) = \begin{vmatrix} \lambda-3 & 1 & -1 \\ -2 & \lambda & 1 \\ -1 & 1 & \lambda-2 \end{vmatrix} = \lambda(\lambda-2)(\lambda-3),$

故A有三个不同的特征根，**A**可与对角形矩阵相似。 与特征根 $\boldsymbol{\lambda}_1=0,\boldsymbol{\lambda}_2=2,\boldsymbol{\lambda}_3=3$相应的三个线性无关的特征向量分别为：

$$\boldsymbol{X}_1 = (1,5,2)^T, \quad \boldsymbol{X}_2 = (1,1,0)^T, \quad \boldsymbol{X}_3 = (2,1,1)^T$$

进一步得 $\boldsymbol{T} = \begin{pmatrix} 1 & 1 & 2 \\ 5 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{T}^{-1} = -\dfrac{1}{6}\begin{pmatrix} 1 & -1 & -1 \\ -3 & -3 & 9 \\ -2 & 2 & -4 \end{pmatrix}$

由定理6可得所求的解为

$$\boldsymbol{X} = e^{\mathbf{A}t}\boldsymbol{X}(0) = \boldsymbol{T}\begin{pmatrix} 1 & & \\ & e^{2t} & \\ & & e^{3t} \end{pmatrix}\boldsymbol{T}^{-1}\boldsymbol{X}(0) = \begin{pmatrix} 1 & 1 & 2 \\ 5 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & & \\ & e^{2t} & \\ & & e^{3t} \end{pmatrix}\left(-\frac{1}{6}\right)\begin{pmatrix} 1 & -1 & -1 \\ -3 & -3 & 9 \\ -2 & 2 & -4 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\boldsymbol{X}(t) = -\frac{1}{6}\begin{pmatrix} -1+3e^{2t}-8e^{3t} \\ -5+3e^{2t}-4e^{3t} \\ -2-4e^{3t} \end{pmatrix}$$

例9 求定解问题 $\begin{cases} \dfrac{\mathrm{d}\boldsymbol{X}(t)}{\mathrm{d}t} = \mathbf{A}\boldsymbol{X}(t)+\boldsymbol{F}(t) \\ \boldsymbol{X}(0)=(1,1,1)^T \end{cases}$ 的解, 其中矩阵**A**见例**8**。 $\boldsymbol{F}(t)=(0,\,0,\,e^{2t})^T$

解 由前面讨论，该问题的解为 $\boldsymbol{X}(t)=e^{\mathbf{A}t}\boldsymbol{X}(0)+\int_0^t e^{\mathbf{A}(t-\tau)}\boldsymbol{F}(\tau)\mathrm{d}\tau$

下面计算 $t=\int_0^t e^{\mathbf{A}(t-\tau)}\boldsymbol{F}(\tau)\mathrm{d}\tau$, 由 $e^{\mathbf{A}(t-\tau)}\boldsymbol{F}(\tau)=\boldsymbol{T}e^{[\boldsymbol{J}(t-\tau)]}\boldsymbol{T}^{-1}\boldsymbol{F}(\tau)$

$$=\begin{pmatrix}1&1&2\\5&1&1\\2&0&1\end{pmatrix}\begin{pmatrix}1&&\\&e^{2(t-\tau)}&\\&&e^{3(t-\tau)}\end{pmatrix}\left(-\frac{1}{6}\right)\begin{pmatrix}1&-1&-1\\-3&-3&9\\-2&2&-4\end{pmatrix}\begin{pmatrix}0\\0\\e^{2\tau}\end{pmatrix}=\left(\frac{-1}{6}\right)\begin{pmatrix}1&1&2\\5&1&1\\2&0&1\end{pmatrix}\begin{pmatrix}-e^{2\tau}\\9e^{2t}\\-4e^{3t-\tau}\end{pmatrix}$$

$$=-\frac{1}{6}\begin{pmatrix}-e^{2\tau}+9e^{2t}-8e^{3t-\tau}\\-5e^{2\tau}+9e^{2t}-4e^{3t-\tau}\\-2e^{2\tau}-4e^{3t-\tau}\end{pmatrix}$$

。将这一结果对变量τ从0到t进行积分，即得 $\boldsymbol{P}=-\dfrac{1}{6}\begin{pmatrix}\frac{1}{2}+(9t+\frac{15}{2})e^{2t}-8e^{3t}\\ \frac{5}{2}+(9t+\frac{3}{2})e^{2t}-4e^{3t}\\ 1+3e^{2t}-4e^{3t}\end{pmatrix}$

因此 $\boldsymbol{X}(t)=\mathrm{e}^{At}\boldsymbol{X}(0)+\boldsymbol{t}$

$$\boldsymbol{X}(t)=-\frac{1}{6}\begin{pmatrix}-\frac{1}{2}+(9t+\frac{21}{2})e^{2t}-16e^{3t}\\ -\frac{5}{2}+(9t+\frac{9}{2})e^{2t}-8e^{3t}\\ -1+3e^{2t}-8e^{3t}\end{pmatrix}$$