矩阵分析及其应用

3.1 矩阵序列

定义 3.1 设矩阵序列$\{A^{(k)}\}$,其中 $A^{(k)}=(a_{ij}^{(k)})\in C^{m\times n}$,当 $k\to\infty$,

$a_{ij}^{(k)}\to a_{ij}$ 时，称矩阵序列$\{A^{(k)}\}$收敛，并称矩阵 $A=(a_{ij})$为矩

阵序列$\{A^{(k)}\}$的极限，或称$\{A^{(k)}\}$收敛于 A, 记为

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

不收敛的矩阵序列称为发散的。

由定义，矩阵序列 $A^{(k)}$ 发散的充要条件为存在 ij 使

得数列 $a_{ij}^{(k)}$ 发散。

类似地，我们可以定义矩阵收敛的 Cauchy 定义

定义 3.1' 矩阵序列$\{A^{(k)}\}$收敛的充要条件为

对任给$\varepsilon>0$ 存在 $N(\varepsilon)$, 当 $k, l \ge N(\varepsilon)$ 时有

$$\|A^{(k)}-A^{(l)}\| < \varepsilon$$

其中$\|.\|$为任意的广义矩阵范数。

例 1 $$\mathbf{A}^{(n)} = \begin{pmatrix} 1-\dfrac{1}{n} & \sin(\dfrac{1}{n}) \\ e^{-n} & \displaystyle\sum_{k=1}^{n}\frac{\sin(k)}{k^2} \end{pmatrix}$$

如果直接按定义我们因为求不出 $\mathbf{A}^{(n)}$的极限从而

很难应用定义 3.1 证明收敛。

相反，由于$$\left|\sum_{k=m+1}^{n}\frac{\sin(k)}{k^2}\right| \le \left|\sum_{k=m+1}^{n}\frac{1}{k^2}\right| \le \left|\sum_{k=m+1}^{n}\frac{1}{k(k-1)}\right|$$

$$< 1/m$$

从而只要 l 充分大，则当 $m, n > l$ 时就有

$$\left|\sum_{k=m+1}^{n}\frac{\sin(k)}{k^2}\right| \le \varepsilon$$

这样 $\mathbf{A}^{(l)}$ 收敛。

定理 3.1 $A^{(k)}\to A$ 的充要条件为

$$\|A^{(k)}-A\|\to 0$$

证明：利用广义矩阵范数的等价性定理，仅对∞范数可以证明。

即 $c_1\|A^{(k)}-A\|_\infty \le \|A^{(k)}-A\| \le c_2\|A^{(k)}-A\|_\infty$

性质 0 若 $A^{(k)}\to A$, 则 $\|A^{(k)}\|\to\|A\|$ 成立。

性质 1. 设 $A^{(k)}\to A_{m\times n}$，$B^{(k)}\to B_{m\times n}$，则

$\alpha\cdot A^{(k)}+\beta\cdot B^{(k)}\to\alpha\cdot A+\beta\cdot B$，　$\forall\ \alpha,\beta\in C$

性质 2. 设 $A^{(k)}\to A_{m\times n}$，$B^{(k)}\to B_{n\times l}$，则

$A^{(k)}\cdot B^{(k)}\to A\cdot B$

证明：由于矩阵范数地等价性，我们可以只讨论相容的矩阵范数。

$\|A^{(k)}\cdot B^{(k)}-A\cdot B\|\le\|A^{(k)}\cdot B^{(k)}-A\cdot B^{(k)}\|+\|AB^{(k)}-A\cdot B\|$

$\le\|A^{(k)}-A\|\cdot\|B^{(k)}\|+\|A\|\cdot\|B^{(k)}-B\|$

注意 $\|B^{(k)}\|\to\|B\|$，　则结论可得。

特别地有

性质 2'.　$A^{(k)}\to A$ 的充要条件为

$A^{(k)}x\to Ax$，对任意 x 成立

或者　$y^HA^{(k)}x\to y^HAx$，对任意 x,y 成立.

（在无穷维空间中称为弱收敛，但在有限维空间中和一般收敛性定义是等价的)

对于 Hermite(对称)矩阵我们有如下的定理：

设 $A^{(k)}$，k=1,2,…,和 A 都为 Hermite 矩阵，那么 $A^{(k)}\to A$ 的充要条件为

$x^HA^{(k)}x\to x^HAx$，对任意 $\mathbf{x}$ 成立

推论：$\mathbf{A}^{(k)}$，k=1,2,…，为半正定 Hermite 矩阵，且单调减少，即 $\boldsymbol{A}^{(k)}-\boldsymbol{A}^{(k+1)}$为半正定 Hermite 矩阵，那么 $\boldsymbol{A}^{(k)}$有极限.

性质 3 设 $\boldsymbol{A}^{(k)}$和 $\boldsymbol{A}$ 都为可逆矩阵，且 $\boldsymbol{A}^{(k)}\to\boldsymbol{A}$，则

$(\boldsymbol{A}^{(k)})^{-1}\to\boldsymbol{A}^{-1}$

证明：　因为 $A^{-1}\cdot(A^{(k)})\to\boldsymbol{I}$. 所以存在 K,当 $k>K$ 时有

$\|I-A^{-1}\cdot(A^{(k)})\|<1/2$

我们有$(A^{(k)})^{-1}=A^{-1}+(I-A^{-1}\cdot(A^{(k)}))(A^{(k)})^{-1}$

从而$\|(A^{(k)})^{-1}\|\le\|A^{-1}\|+\|(I-A^{-1}\cdot(A^{(k)}))\|\cdot\|(A^{(k)})^{-1}\|$

当 k>K 时，有

$\|(A^{(k)})^{-1}\|\le\|A^{-1}\|+1/2\cdot\|(A^{(k)})^{-1}\|$

即　$\|(A^{(k)})^{-1}\|\le 2\cdot\|A^{-1}\|$

因为 $A^{-1}-(A^{(k)})^{-1}=A^{-1}(A^{(k)}-A)(A^{(k)})^{-1}$

从而 $\|A^{-1}-(A^{(k)})^{-1}\|\le\|A^{-1}\|\cdot\|A^{(k)}-A\|\cdot\|(A^{(k)})^{-1}\|$

(当 k>K 时)　　$\le\|A^{-1}\|\cdot\|A^{(k)}-A\|\cdot 2\|A^{-1}\|$

(当 $k\to\infty$时)　　$\to 0$

由定理 3.1 有$(A^{(k)})^{-1}\to A^{-1}$

定义 3.2 矩阵序列$\{A^{(k)}\}$称为有界的，如果存在常数 M>0，使得对一切 k 都有

$|a_{ij}^{(k)}|<M$ 或等价的 $\|A^{(k)}\|<M'$

定理：有界的矩阵序列$\{A^{(k)}\}$一定有收敛的子列。

定义 3.3 设 A 为方阵，且当 $k\to\infty$时有 $A^k\to 0$，则称 A 为收敛矩阵。

定理 3.2(迭代法基本定理) $A^k\to 0$ 的充要条件为谱半径

$$\rho(A)<1.$$

证明：必要性：设 $A^k\to 0$，证明$\rho(A)<1$.

对 $\boldsymbol{A}$ 的任意特征值λ和相应的特征向量 $\boldsymbol{x}$ 有

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

这样我们有 $A^k\boldsymbol{x}=\lambda^k\boldsymbol{x}$

从而有$|\lambda|^k\cdot\|x\|=\|A^k x\|\le\|A^k\|\cdot\|x\|$

从而有$|\lambda|^k\le\|A^k\|\to 0$

这样有$|\lambda|<1$， 由于λ为 A 的任意特征值，

所以$\rho(A)<1$, 即必要性得证。

充分性。已知$\rho(A)<1$，证明 $A^k\to 0$.

取$\varepsilon=(1-\rho(A))/2>0$,由定理 2.10 有,存在某种相容的矩阵范数$\|.\|_M$使得 $\|A\|_M<\rho(A)+\varepsilon<1$

从而$\|A^k\|_M\le(\|A\|_M)^k<(\rho(A)+\varepsilon)^k$

所以当 $k\to\infty$有$\|A^k\|_M\to 0$, 从而 $A^k\to 0$.

定理 3.3 $A^k\to 0$ 的充分条件为存在矩阵范数$\|.\|_M$使得 $\|A\|_M<1$

3.2 矩阵级数

定义 3.4 设矩阵序列$\{A^{(k)}\}$,其中 $A^{(k)}=(a_{ij}^{(k)})\in C^{n\times n}$,由它们形成的无穷和 $A^{(0)}+A^{(1)}+\ldots+A^{(k)}+\ldots$称为矩阵级数，

记为 $\sum_{k=0}^{\infty}A^{(k)}$，即有

$$\sum_{k=0}^{\infty}A^{(k)}=A^{(0)}+A^{(1)}+\ldots+A^{(k)}+\ldots$$

定义 3.5 记 $S^{(N)}=\sum_{k=0}^{N}A^{(k)}$,称其为矩阵级数$\sum_{k=0}^{\infty}A^{(k)}$的部分和.

如果矩阵序列$\{S^{(N)}\}$收敛，且有极限 S，即有

$$S^{(N)}\to S$$

那么称矩阵级数$\sum_{k=0}^{\infty}A^{(k)}$收敛，且和为 S，记为

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

不收敛的矩阵级数称为发散的。

显然$\sum_{k=0}^{\infty}A^{(k)}$=S 是指$\sum_{k=0}^{\infty}a_{ij}^{(k)}=s_{ij}$，$\forall\ i,j$

即矩阵级数收敛是指它的每个分量所构成的数项级数收敛。

性质：矩阵级数$\sum_{k=0}^{\infty}A^{(k)}$收敛的充要条件为对任意向量 x,

向量级数$\sum_{k=0}^{\infty}A^{(k)}x$收敛。

定义 3.6 设矩阵级数$\sum_{k=0}^{\infty}A^{(k)}$的每个分量$a_{ij}^{(k)}$所构成的数项级数$\sum_{k=0}^{\infty}a_{ij}^{(k)}$绝对收敛，则称矩阵级数$\sum_{k=0}^{\infty}A^{(k)}$绝对收敛。

关于绝对收敛，我们有如下的定理：

性质 1. 绝对收敛的$\sum_{k=0}^{\infty}A^{(k)}$交换求和次序不改变其绝对收敛性和极限值。

性质 2. 矩阵级数$\sum_{k=0}^{\infty}A^{(k)}$绝对收敛的充要条件为正项级数

$\sum_{k=0}^{\infty}\|A^{(k)}\|$收敛。

性质 3. 如果矩阵级数$\sum_{k=0}^{\infty}A^{(k)}$(绝对)收敛,那么$\sum_{k=0}^{\infty}PA^{(k)}Q$也是(绝对)收敛,且有

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

性质 4. 设 $C^{n\times n}$ 的两个矩阵级数

S_1: $A^{(1)}+A^{(2)}+\ldots+A^{(k)}+\ldots$

S_2: $B^{(1)}+B^{(2)}+\ldots+B^{(k)}+\ldots$

都绝对收敛，其和分别为 A 和 B.则矩阵级数

S_3: $A^{(1)} B^{(1)}+ [A^{(1)} B^{(2)}+ A^{(2)} B^{(1)}]+\ldots$

$+[A^{(1)} B^{(k)}+ A^{(2)} B^{(k-1)} +\ldots+A^{(k)} B^{(1)}]+\ldots$

绝对收敛且和为 AB.

证明：由于 S_1: $A^{(1)}+A^{(2)}+\ldots+A^{(k)}+\ldots$绝对收敛的充要条件为正项级数$\|A^{(1)}\|+\|A^{(2)}\|+\ldots+\|A^{(k)}\|+\ldots$收敛且与排列无关。

我们证明的思路是证明正项级数：

$\|A^{(1)} B^{(1)}\|+ \|A^{(1)} B^{(2)}+ A^{(2)} B^{(1)}\|+\ldots$

$+\|A^{(1)} B^{(k)}+ A^{(2)} B^{(k-1)} +\ldots+A^{(k)} B^{(1)}\|+\ldots$

收敛。引用魏氏定理，我们仅需验证下列正项级数：

$\|A^{(1)}\|\cdot\|B^{(1)}\|+ \{ \|A^{(1)} \|\cdot\|B^{(2)}\|+ \|A^{(2)} \|\cdot\|B^{(1)}\|\}+\ldots$

$+\{\|A^{(1)} \|\cdot\|B^{(k)}\|+ \|A^{(2)} \|\cdot\|B^{(k-1)} \|+\ldots+\|A^{(k)} \|\cdot\|B^{(1)}\|\}+\ldots$

收敛。这由题设

正项级数$\|A^{(1)}\|+\|A^{(2)}\|+\ldots+\|A^{(k)}\|+\ldots$

和正项级数 $\|B^{(1)}\|+\|B^{(2)}\|+\ldots+\|B^{(k)}\|+\ldots$

的收敛性可得。

定理 3.5 幂级数 $I+A+A^2+\ldots+A^k+\ldots$收敛的充要条件为 A 的谱半径$\rho(A)<1$, 收敛时其和为$(I-A)^{-1}$。

若有矩阵范数$\|.\|$使得$\|A\|<1$,则

$\|(I-A)^{-1}- (I+A+A^2+\ldots+A^k)\|\le\|A\|^{k+1}/(1-\|A\|)$

证明： 必要性. 由于 $I+A+A^2+\ldots+A^k+\ldots$收敛，从而 $S^{(k)}= I+A+A^2+\ldots+A^k$ 收敛。记 $T^{(k)}= I+A+A^2+\ldots+A^{k+1}$, $A^{k+1}=T^{(k)}- S^{(k)}$收敛，且 $T^{(k)}- S^{(k)} \to 0$，这样我们有 $A^k \to 0$，从而$\rho(A)<1$.

充分性：设$\rho(A)<1$，$(I-A)^{-1}$存在，由于

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

因 $A^k\to 0$，所以

$I+A+A^2+\ldots+A^k+\ldots\to (I-A)^{-1}$.

又因为

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

从而

$\|(I-A)^{-1} - (I+A+A^2+\ldots+A^k)\|=\| (I-A)^{-1} A^{k+1}\|$

设 $B=(I-A)^{-1}A^{k+1}$,从而$(I-A)B=A^{k+1}$ 即

$B=AB+ A^{k+1}$,从而

$\|B\|\le \|A\|\cdot\|B\|+ \|A^{k+1}\|\le \|A\|\cdot\|B\|+ \|A\|^{k+1}$

因为矩阵范数$\|.\|$使得$\|A\|<1$,所以

$\|B\|\le\|A\|^{k+1}/(1-\|A\|)$成立。

定理 3.6 设幂级数 $f(z)=\sum_{k=0}^{\infty}c_k z^k$ 的收敛半径为 r,

如果方阵 A 满足ρ(A)< r, 则矩阵幂级数

$$f(A)=\sum_{k=0}^{\infty}c_k A^k \text{ 是绝对收敛的；}$$

如果ρ(A) > r, $\sum_{k=0}^{\infty}c_k A^k$ 是发散的。

证明：如果ρ(A)< r,利用绝对收敛的性质可得绝对收敛性。

反之，设 A 的特征值λ满足$|\lambda|=\rho(\boldsymbol{A})$，$\boldsymbol{x}$ 为λ相应的

特征向量 $\sum_{k=0}^{n}c_k(A^k x)=\sum_{k=0}^{n}c_k(\lambda^k x)=(\sum_{k=0}^{n}c_k\lambda^k)x$,

由于ρ(A) > r ,那么 $(\sum_{k=0}^{n}c_k\lambda^k)x$ 发散(注意 x 为非零向量)

从而 $\sum_{k=0}^{n}c_k(A^k x)$ 发散，这样 $\sum_{k=0}^{\infty}c_k A^k$ 发散。

矩阵函数

定义：设一元函数 $f(z)$能展开为 z 的幂级数

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

其中 r>0 表示该幂级数的收敛半径。当 n 阶矩阵 $\boldsymbol{A}$ 的

谱半径ρ($\boldsymbol{A}$) < r 时，把收敛的矩阵幂级数 $\sum_{k=0}^{\infty}c_k A^k$ 的和

为 $f(\boldsymbol{A})$, 即 $f(\boldsymbol{A})=\sum_{k=0}^{\infty}c_k\boldsymbol{A}^k$.

性质 1(代入规则)： 若 $f(z)$能展开为 z 的幂级数,且 $f(z)=g(z)$,
对 $|z|<r$ 成立，则当$\rho(\boldsymbol{A})<r$ 时， $f(\boldsymbol{A})=g(\boldsymbol{A})$.

矩阵函数举例：

$\sin(z)=z-z^3/3!+z^5/5!-\ldots$

则 $\sin(\boldsymbol{A})=\boldsymbol{I}-\boldsymbol{A}^3/3!+\boldsymbol{A}^5/5!-\ldots$

$\cos(z)=1-z^2/2!+z^4/4!-\ldots$

$\cos(\boldsymbol{A})=\mathrm{I}-\boldsymbol{A}^2/2!+\boldsymbol{A}^4/4!-\ldots$

$e^z=1+z+z^2/2!+z^3/3!+\ldots$

$e^{\boldsymbol{A}}=\mathrm{I}+\boldsymbol{A}+\boldsymbol{A}^2/2!+\boldsymbol{A}^3/3!+\ldots$

$\sin^2(z)+\cos^2(z)=1$

可得：$\sin^2(\boldsymbol{A})+\cos^2(\boldsymbol{A})=\boldsymbol{I}$

性质 2 二元函数 $f(x,y)$能展开为 x,y 的幂级数,$f(x,y)=g(x,y)$.

若 $\boldsymbol{AB}=\boldsymbol{BA}$,则 $f(\boldsymbol{A},\boldsymbol{B})=g(\boldsymbol{A},\boldsymbol{B})$ (二元函数的代入规则).

矩阵函数值的求法

1.待定系数法

设 n 阶矩阵 $\boldsymbol{A}$ 的特征多项式$\varphi(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})$. 如果首 1

多项式$\psi(\lambda)=\lambda^m+b_1\lambda^{m-1}+\ldots+b_{m-1}\lambda+b_m$ $(1\le m\le n)$

满足:(1) $\psi(A)=0$;(2) $\psi(\lambda)$整除$\varphi(\lambda)$(矩阵 A 的最小多项式

与特征多项式均满足这些条件). 那么, $\psi(\lambda)$的零点都是

A 的特征值.记 $\psi(\lambda)$的互异零点为

$\lambda_1,\lambda_2,\ldots,\lambda_s$,相应的重数为 $r_1,\ldots,r_s(r_1+r_2+\ldots+r_s=m)$,则有

$\psi^{(l)}(\lambda_i)=0$ $(l=0,1,\ldots,r_i-1;i=1,2,\ldots,s)$

这里, $\psi^{(l)}(\lambda)$表示$\psi(\lambda)$的 l 阶导数(下同).

设 $f(z)=\sum_{k=0}^{\infty}c_kz^k=\psi(z)g(z)+r(z)$. 其中 $r(z)$是次数低于 m 的

多项式，于是可由 $f^{(l)}(\lambda_i)=r^{(l)}(\lambda_i)$ 确定 $r(z)$.

利用 $f(A)=\psi(A)g(A)+r(A)=r(A)$.

因此我们的问题就是给定函数 $f(z)$,由约束条件

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

确定 $r(z)$。

若知道函数 $f(x)$在 x_0 的函数值和直到 n 阶导数值，

则由 Tayler 展开式可得多项式:

$$Tf(x)=f(x_0)+f'(x_0)(x-x_0)+[f''(x_0)/2!](x-x_0)^2$$
$$+[f^{(3)}(x_0)/3!](x-x_0)^3+\ldots+[f^{(n)}(x_0)/n!](x-x_0)^n$$

使得 $Tf^{(l)}(x_0)=f^{(l)}(x_0),l=0,1,2,\ldots,n$ 成立.

相应的若知道函数 $f(x)$在 $x_0,x_1,x_2,\ldots x_n$ 的函数值

则有相应的 Newton 插值公式

$$Nf(x)=f(x_0)+f[x_0,x_1](x-x_0)+f[x_0,x_1,x_2](x-x_0)(x-x_1)+\ldots$$
$$+f[x_0,x_1,\ldots,x_n](x-x_0)(x-x_1)\ldots(x-x_{n-1})$$

使得 $Nf(x_i)=f(x_i)$， $i=0,1,2,\ldots,n$ 成立.

其中 $Nf(x)$的项为均差。

均差的定义和简单性质如下：
均差及性质
定义： $f[x_0,x_k]=(f(x_k)-f(x_0))/(x_k-x_0)$
为 $f(x)$关于点 x_0,x_k 的一阶均差；
$f[x_0,x_1,x_k]=(f[x_0,x_k]-f[x_0,x_1])/(x_k-x_1)$
为二阶均差。一般地，称
$f[x_0,x_1,\ldots,x_k]=(f[x_0,x_1,\ldots,x_{k-2},x_k]-f[x_0,x_1,\ldots,x_{k-1}])/(x_k-x_{k-1})$
为 $f(x)$的 k 阶均差。
性质 1. k 阶均差可表为函数值的线性组合；
性质 2. 均差与节点的排列次序无关；
性质 3(均差与导数) 若 $f(x)$在$[a,b]$上存在 n 阶导数，
且结点 $x_0,x_1,x_2,\ldots x_n\in[a,b]$，则存在$\xi\in[x^0,x^1]$使得
$$f[x_0,x_1,\ldots,x_n]=f^{(n)}(\xi)/n!$$
成立，其中 $x^0=\min\{x_0,x_1,x_2,\ldots x_n\}$, $x^1=\max\{x_0,x_1,x_2,\ldots x_n\}$。
推论:若 $f(x)$在$[a,b]$上存在 n 阶导数,则对任何 $c\in[a,b]$,
$f[c,c,\ldots,c]=f^{(n)}(c)/n!$ 成立。
证明：在性质 3 中令 $x_0,x_1,x_2,\ldots x_n$ 同时趋于 c 可得结论.
由此可见 Newton 插值公式可以看作 Taylor 展开式的推广，Taylor 展开式只不过是插值节点重合的情形，相应的函数值变为导数值而已。

均差的计算：

x_0	$f(x_0)$				
x_1	$f(x_1)$	$f[x_0,x_1]$			
x_2	$f(x_2)$	$f[x_1,x_2]$	$f[x_0,x_1,x_2]$		
x_3	$f(x_3)$	$f[x_2,x_3]$	$f[x_1,x_2,x_3]$	$f[x_0,x_1,x_2,x_3]$	
x_4	$f(x_4)$	$f[x_3,x_4]$	$f[x_2,x_3,x_4]$	$f[x_1,x_2,x_3,x_4]$	$f[x_0,x_1,x_2,x_3,x_4]$

相应地，若有 $x_1=x_2=x_3$，则有

x_0	$f(x_0)$				
x_1	$f(x_1)$	$f[x_0,x_1]$			
x_1	$f(x_1)$	$f'(x_1)$	$f[x_0,x_1,x_1]$		
x_1	$f(x_1)$	$f'(x_1)$	$f''(x_1)/2!$	$f[x_0,x_1,x_1,x_1]$	
x_4	$f(x_4)$	$f[x_1,x_4]$	$f[x_1,x_1,x_4]$	$f[x_1,x_1,x_1,x_4]$	$f[x_0,x_1,x_1,x_1,x_4]$

其它的类似.因此我们可以使用 Newton 插值公式得到满足插值条件的多项式.

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

解　$\varphi(\lambda)=\det(\lambda\mathbf{I}-\mathbf{A})=(\lambda-2)(\lambda-5)(\lambda+1)$

z　$f(z)=e^{zt}$
2　e^{2t}
5　e^{5t}　$(f(5)-f(2))/(5-2)=(e^{5t}-e^{2t})/3$
-1　e^{-t}　$(f(-1)-f(5))/(-1-5)=(e^{5t}-e^{-t})/6$　$(e^{5t}-2e^{2t}+e^{-t})/18$

从而 $f(z)=e^{zt}=\varphi(z)P(z)+r(z)$,
其中 $r(z)=e^{2t}+(z-2)(e^{5t}-e^{2t})/3+(z-2)(z-5)(e^{5t}-2e^{2t}+e^{-t})/18$
$=e^{2t}(1+(z-2)/3-(z-2)(z-5)/9)$
$+e^{5t}((z-2)/3+(z-2)(z-5)/18)$
$+e^{-t}(z-2)(z-5)/18$
$f(\mathbf{A})=e^{\mathbf{A}t}=r(\mathbf{A})=e^{2t}(I+(\mathbf{A}-2\cdot\mathbf{I})/3-(\mathbf{A}-2\cdot\mathbf{I})(\mathbf{A}-5\cdot\mathbf{I})/9)$
$+e^{5t}((\mathbf{A}-2\cdot\mathbf{I})/3+(\mathbf{A}-2\cdot\mathbf{I})(\mathbf{A}-5\cdot\mathbf{I})/18)$
$+e^{-t}(\mathbf{A}-2\cdot\mathbf{I})(\mathbf{A}-5\cdot\mathbf{I})/18$

$$=e^{2t}\begin{bmatrix}1 & 0 & 0\\ 5/9 & 4/9 & 4/3\\ 2/3 & -2/3 & 16/9\end{bmatrix}+e^{5t}\begin{bmatrix}0 & 0 & 0\\ 2/9 & -2/9 & 1/3\\ 1/6 & -/6 & 1/9\end{bmatrix}+$$

$$e^{-t}\begin{bmatrix}0 & 0 & 0\\ -1/9 & 1/9 & -1/3\\ -1/6 & 1/6 & -2/9\end{bmatrix}$$

例 3.5　设 $\mathbf{A}=\begin{bmatrix}2 & 0 & 0\\ 1 & 1 & 1\\ 1 & -1 & 3\end{bmatrix}$，求 $e^{\mathbf{A}t}$

解 $\varphi(\lambda)=\det(\lambda\mathbf{I}-\mathbf{A})=(\lambda-2)^3$
z　$f(z)=e^{zt}$
2　e^{2t}
2　e^{2t}　$f'(2)=te^{2t}$
2　e^{2t}　$f'(2)=te^{2t}$　$f''(2)/2!=t^2e^{2t}/2$
从而 $f(z)=e^{zt}=\varphi(z)P(z)+r(z)$,
其中 $r(z)=e^{2t}+t\,e^{2t}(z-2)+t^2e^{2t}(z-2)^2/2$
$f(\mathbf{A})=e^{\mathbf{A}t}=r(\mathbf{A})=e^{2t}(\mathbf{I}+t(\mathbf{A}-2\cdot\mathbf{I})+t^2(\mathbf{A}-2\cdot\mathbf{I})^2/2)$

$$=e^{2t}\begin{bmatrix}1 & 0 & 0\\ t & 1-t & 1\\ t & -t & 1+t\end{bmatrix}$$

例： $\mathbf{A}=\begin{bmatrix}1 & 2 & -1 & 1\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & -1 & -1 & -1\end{bmatrix}$,

求 $f(\mathbf{A})$, 其中 $f(z)=z^{16}-z^{10}$。

解：由于 $\det(\lambda\mathbf{I}-\mathbf{A})=(\lambda-1)^3(\lambda+1)=\varphi(\lambda)$

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

$$\mathbf{A}^3=\begin{bmatrix}1 & -7 & -4 & 1\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & -1 & -1 & -1\end{bmatrix},$$

求 $r(z)$使得 $f(z)=P(z)\,\varphi(z)+r(z)$, 其中 $f(z)=z^{16}-z^{10}$

z	$f(z)$			
1	0			
1	0	$f'(1)=6$		
1	0	$f'(1)=6$	$f''(1)/2!=75$	
-1	0	0	$(0-6)/(-1-1)=3$	$(3-75)/(-1-1)=36$

因此 $r(z)=0+6(z-1)+75(z-1)^2+36(z-1)^3$

从而 $f(\mathbf{A})=r(\mathbf{A})=6(\mathbf{A}-\mathbf{I})+75(\mathbf{A}-\mathbf{I})^2+36(\mathbf{A}-\mathbf{I})^3$

即 $r(A)=\begin{bmatrix}0 & 9 & -9 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$

2. 数项级数求和法和基于友矩阵的矩阵函数计算。

利用矩阵 A 的零化多项式

$A^m-k_{m-1}A^{m-1}-\ldots-k_2A^2-k_1A-k_0I=0$ （I）

可得

$A^m=k_0I+k_1A+k_2A^2+\ldots+k_{m-1}A^{m-1}$

从而

$A^{m+i}=k_0A^i+k_1A^{i+1}+k_2A^{i+2}+\ldots+k_{m-1}A^{m+i-1}$，i=0,1,2,… (II)

成立。

将矩阵 I, A, A^2，…,A^{m-1},…. 看作向量，定义向量组

$[A^i, A^{i+1}, \ldots, A^{m+i-1}]$,

那么根据(II)式可得

$[A^{i+1},A^{i+2},\dots,A^{m+i}]=[A^i,A^{i+1},\dots,A^{m+i-1}]\boldsymbol{Q}$

其中 $\boldsymbol{Q}=\begin{pmatrix} 0 & 0 & \cdots & 0 & k_0 \\ 1 & 0 & \cdots & 0 & k_1 \\ 0 & 1 & \cdots & 0 & \vdots \\ \vdots & \vdots & \ddots & \vdots & k_{m-2} \\ 0 & 0 & \cdots & 1 & k_{m-1} \end{pmatrix}$

称为 A 的零化多项式(I)的友矩阵，可以证明 Q 的特征多项式为(I)

那么

$[I,A,\dots,A^{m-1}]=[I,A,\dots,A^{m-1}]\,I=[I,A,\dots,A^{m-1}]\boldsymbol{Q}^0$

$[A^1,\dots,A^{m-1},A^m]=[I,A,\dots,A^{m-1}]\boldsymbol{Q}$

$[A^2,\dots,A^m,A^{m+1}]=[A^1,\dots,A^{m-1},A^m]\,\boldsymbol{Q}=[I,A,\dots,A^{m-1}]\boldsymbol{Q}^2$

类推得到

$[A^i,A^{i+1},\dots,A^{m+i-1}]=[I,A,\dots,A^{m-1}]\,\boldsymbol{Q}^i$ (III)

$i=0,1,2,\dots$

在(III)两边同时乘以 m 维单位向量 $\boldsymbol{e}_1=(1,0,0,\dots,0)^T$ 可得

$A^i=[A^i,A^{i+1},\dots,A^{m+i-1}]\,\boldsymbol{e}_1=[I,A,\dots,A^{m-1}]\,\boldsymbol{Q}^i\,\boldsymbol{e}_1$

因此

$A^i=[I,A,\dots,A^{m-1}]\,\boldsymbol{Q}^i\,\boldsymbol{e}_1$

$$f(\boldsymbol{A})=\sum_{i=0}^{\infty}c_i\boldsymbol{A}^i=[I,A,\dots,A^{m-1}]\left(\sum_{i=0}^{\infty}c_i\boldsymbol{Q}^i\right)\boldsymbol{e}_1=[I,A,\dots,A^{m-1}]\,f(\boldsymbol{Q})\boldsymbol{e}_1$$

这时，我们需要计算 $f(\boldsymbol{Q})$,然后得到 $f(\boldsymbol{Q})\boldsymbol{e}_1$，从而得到 $[I,A,\dots,A^{m-1}]$的相应系数。当然这样的计算结果本质和待定系数法得到的结果是一样的。

另外,记 $\mathbf{c}^{(0)}=(c_0,c_1,\dots,c_{m-1})^T$, $\mathbf{c}^{(1)}=(c_m,c_{m+1},\dots,c_{2m-1})^T$, …,

$\mathbf{c}^{(k)}=(c_{km},c_{km+1},\dots,c_{km+m-1})^T$,….,

$\boldsymbol{P}=\boldsymbol{Q}^m$

那么根据(III)式可得

$[A^{km},A^{km+1},\dots,A^{km+m-1}]=[I,A,\dots,A^{m-1}]\,\boldsymbol{Q}^{km}=[I,A,\dots,A^{m-1}]\,\boldsymbol{P}^k$

因此

$$f(\boldsymbol{A})=\sum_{i=0}^{\infty}c_i\boldsymbol{A}^i=\sum_{k=0}^{\infty}[A^{km},A^{km+1},\dots,A^{km+m-1}]\,\mathbf{c}^{(k)}$$

$$=[I,A,\dots,A^{m-1}]\sum_{k=0}^{\infty}\boldsymbol{P}^k\boldsymbol{c}^{(k)}$$

因此，$f(\boldsymbol{A})$归结为求$\sum_{k=0}^{\infty}\boldsymbol{P}^k\boldsymbol{c}^{(k)}$ 的值，这在某些情形下是可以得到。

3. 对角阵法

若 **A** 相似于对角阵，求出它的对角矩阵和相似矩阵

即 $\mathbf{A}=\mathbf{P}\mathbf{D}\mathbf{P}^{-1}$， 则 $f(\mathrm{A})=\mathbf{P}f(\mathrm{D})\mathbf{P}^{-1}$

4.Jordan 标准型法

设 A 的若当标准形为 J,则存在可逆矩阵 P 使得

$$\mathbf{P}^{-1}\mathbf{A}\,\mathbf{P}=\mathbf{J}=\begin{bmatrix} \mathbf{J}_1 & & \\ & \ddots & \\ & & \mathbf{J}_s \end{bmatrix},\ \mathbf{J}_i=\begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{bmatrix}_{m_i\times m_i}$$

$$f(\mathbf{J}_i)=\sum_{k=0}^{\infty} a_k \mathbf{J}_i^k = \sum_{k=0}^{\infty} a_k \begin{bmatrix} \lambda_i^k & C_k^1\lambda_i^{k-1} & \cdots & C_k^{m_i-1}\lambda_i^{k-m_i+1} \\ & \lambda_i^k & \ddots & \vdots \\ & & \ddots & C_k^1\lambda_i^{k-1} \\ & & & \lambda_i^k \end{bmatrix}_{m_i\times m_i}$$

$$=\begin{bmatrix} f(\lambda_i) & \frac{1}{1!}f'(\lambda_i) & \cdots & \frac{1}{(m_i-1)!}f^{(m_i-1)}(\lambda_i) \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & \frac{1}{1!}f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix}_{m_i\times m_i} \qquad (*)$$

$$f(\mathbf{A})=\sum_{k=0}^{\infty} a_k \mathbf{P}\mathbf{J}^k\mathbf{P}^{-1}=\mathbf{P}\Big(\sum_{k=0}^{\infty} a_k \mathbf{J}^k\Big)\mathbf{P}^{-1}=\mathbf{P}\begin{bmatrix} \sum_{k=0}^{\infty} a_k\mathbf{J}_1^k & & \\ & \ddots & \\ & & \sum_{k=0}^{\infty} a_k\mathbf{J}_s^k \end{bmatrix}\mathbf{P}^{-1}$$

$$=\mathbf{P}\begin{bmatrix} f(\mathbf{J}_1) & & \\ & \ddots & \\ & & f(\mathbf{J}_s) \end{bmatrix}\mathbf{P}^{-1} \qquad (**)$$

可见，求 $f(\mathbf{A})$可转化为求 A 的若当标准形 **J** 及变换矩阵 **P** 的问题。

3.3　广义矩阵函数

利用前面的定理 3.6 及其推论的结果，对于可以展开成幂级数形式的解析函数 $f(z)$,它的矩阵函数的定义方式就是在它的幂级数形式中用矩阵 **A** 替代 z,就得到了由矩阵幂级数形式的矩阵函数 f(A)的定义。但是，对于任意给定的函数 f(z)能够展开成收敛幂级数形式的要求条件很强，一般不容易满足，例如 $f(z)=\ln(z)$ 就不满足。借助于前面讨论的矩阵函数若当标准形求法的公式(*)和(**)，　我们可以拓宽矩阵函数的定义如下.

定义 3.8　设 $\mathbf{A}\in C^{n\times n}$ 的 Jordan 标准形为 **J**,即有可逆矩阵 **P**，使得

$$\mathbf{P}^{-1}\mathbf{A}\,\mathbf{P}=\mathbf{J}=\begin{bmatrix}\mathbf{J}_1 & & \\ & \ddots & \\ & & \mathbf{J}_s\end{bmatrix},\ \mathbf{J}_i=\begin{bmatrix}\lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i\end{bmatrix}_{m_i\times m_i}$$

如果函数 $f(z)$在λ_i处具有直到 m_i-1 阶导数($i=1,2,\ldots,s$),令

$$f(\mathbf{J}_i)=\begin{bmatrix}f(\lambda_i) & \frac{1}{1!}f'(\lambda_i) & \cdots & \frac{1}{(m_i-1)!}f^{(m_i-1)}(\lambda_i) \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & \frac{1}{1!}f'(\lambda_i) \\ & & & f(\lambda_i)\end{bmatrix}_{m_i\times m_i}$$

$$f(\mathbf{A})=\mathbf{P}\begin{bmatrix}f(\mathbf{J}_1) & & \\ & \ddots & \\ & & f(\mathbf{J}_s)\end{bmatrix}\mathbf{P}^{-1} \qquad (**)$$

那么，称 $f(\mathbf{A})$为对应于 $f(z)$的**矩阵函数**。

这样定义的矩阵函数在 $f(z)$可展开成幂级数的形式和前面的定义一致，而当 $f(z)$不能展开成幂级数的形式时仍然可以定义，因此确实拓广了矩阵函数的定义。

定义 3.8’　设 $\mathbf{A}\in C^{n\times n}$, $f(z)$解析函数，定义矩阵函数 $f(A)$为

$$f(\mathrm{A})=\int_{\partial\Omega}(z\cdot I-A)^{-1}f(z)dz$$

其中Ω为包含 A 的全部特征根的单连通区域，而∂Ω表示区域的边界。

3.4 矩阵的微分和积分

定义 3.9 如果矩阵 $\mathbf{A}(t)=(a_{ij}(t))_{m\times n}$ 的每一个元素 $a_{ij}(t)$ 是变量 t 的可微函数，则 $\mathbf{A}(t)$关于 t 的导数(微商)定义为

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

性质 3.4.1. $\frac{d}{dt}(\mathbf{A}(t)+\mathbf{B}(t))=\frac{d}{dt}\mathbf{A}(t)+\frac{d}{dt}\mathbf{B}(t)$

性质 3.4.2. $\frac{d}{dt}(\boldsymbol{A}(t)\boldsymbol{B}(t))=\frac{d}{dt}\boldsymbol{A}(t)\cdot\boldsymbol{B}(t)+\boldsymbol{A}(t)\cdot\frac{d}{dt}\boldsymbol{B}(t)$

推论　设 $\mathbf{C}(t)=\boldsymbol{A}(t)\boldsymbol{B}(t)$,那么 $d\boldsymbol{C}=\boldsymbol{dA}\cdot\boldsymbol{B}+\boldsymbol{A}\cdot d\boldsymbol{B}$,

其中 $d\boldsymbol{A}=(\boldsymbol{dA}/dt)\cdot dt$,$d\boldsymbol{B}=(d\boldsymbol{B}/dt)\,dt$

例(逆矩阵的导数): $\frac{d}{dt}(A(t)^{-1})=-A(t)^{-1}\frac{d(A(t))}{dt}A(t)^{-1}$

证明：　根据逆矩阵定义

$A(t)\cdot A(t)^{-1}=\boldsymbol{E}_n$

等式两边求导，根据性质 2 可得

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

因此 $\frac{d}{dt}(A(t)^{-1})=-A(t)^{-1}\frac{d(A(t))}{dt}A(t)^{-1}$

推论，设 $\mathbf{C}(t)=\boldsymbol{A}(t)\boldsymbol{B}(t)$,那么 $dC=\boldsymbol{dA}\cdot\boldsymbol{B}+\boldsymbol{A}\cdot d\boldsymbol{B}$

性质 3.4.3. $\frac{d}{dt}(\alpha(t)\mathbf{A}(t))=\alpha'(t)\mathbf{A}(t)+\alpha(t)\cdot\frac{d}{dt}\mathbf{A}(t)$

性质 3.4.4. 如果 $\mathbf{A}(t)$和 $\frac{d}{dt}\mathbf{A}(t)$可交换，$f(z)$和 t 无关的一元解析函数，

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

特别注意若 $A(t)$和 $\frac{d}{dt}A(t)$不可交换，则上式不一定成立。

证明：设 $\mathbf{A}(t)\in C^{n\times n}$ 的 Jordan 标准形为 $\mathbf{J}(t)$,即有可逆矩阵 $\mathbf{P}(t)$，使得

$$\mathbf{P}^{-1}(t)\mathbf{A}(t)\,\mathbf{P}(t)=\mathbf{J}(t)=\begin{bmatrix}\mathbf{J}_1(t) & & \\ & \ddots & \\ & & \mathbf{J}_s(t)\end{bmatrix},\ \mathbf{J}_i(t)=\begin{bmatrix}\lambda_i(t) & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i(t) & 1 \\ & & & \lambda_i(t)\end{bmatrix}_{m_i\times m_i}$$

$d\,\mathbf{A}(t)/dt=(d\mathbf{P}(t)/dt)\mathbf{J}(t)\,\mathbf{P}^{-1}(t)+\mathbf{P}(t)(d\mathbf{J}(t)/dt)\,\mathbf{P}^{-1}(t)+\mathbf{P}(t)\,\mathbf{J}(t)\,(d\mathbf{P}^{-1}(t)/dt)$

$=(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)\mathbf{A}(t)+\mathbf{P}(t)(d\mathbf{J}(t)/dt)\,\mathbf{P}^{-1}(t)-\mathbf{A}(t)(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)$

记 $\mathbf{B}(t)=d\,\mathbf{A}(t)/dt-\mathbf{P}(t)(d\mathbf{J}(t)/dt)\,\mathbf{P}^{-1}(t)$,那么由假设 $\mathbf{A}(t)$和 $\frac{d}{dt}\mathbf{A}(t)$可交换可得

$\mathbf{A}(t)$和 $\mathbf{B}(t)$可以交换，这主要是因为

$\mathbf{A}(t)\,(d\,\mathbf{A}(t)/dt)=(d\,\mathbf{A}(t)/dt)\,\mathbf{A}(t)$

$\mathbf{A}(t)\,\mathbf{P}(t)(d\mathbf{J}(t)/dt)\,\mathbf{P}^{-1}(t)= \mathbf{P}(t)\mathbf{J}(t)(d\mathbf{J}(t)/dt)\,\mathbf{P}^{-1}(t)= \mathbf{P}(t)\,(d\mathbf{J}(t)/dt)\,\mathbf{J}(t)\mathbf{P}^{-1}(t)$

$= \mathbf{P}(t)\,(d\mathbf{J}(t)/dt)\,\mathbf{P}^{-1}(t)\,\mathbf{P}(t)\mathbf{J}(t)\mathbf{P}^{-1}(t) = \mathbf{P}(t)\,(d\mathbf{J}(t)/dt)\,\mathbf{P}^{-1}(t)\,\mathbf{A}(t)$

所以有 $(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)\mathbf{A}(t) = \mathbf{A}(t)(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t) + \mathbf{B}(t)$

递推可得：$(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)\mathbf{A}^2(t) = \mathbf{A}^2(t)(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t) +2\,\mathbf{B}(t)\,\mathbf{A}(t)$

…

$(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)\mathbf{A}^k(t) = \mathbf{A}^k(t)(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t) +k\,\mathbf{B}(t)\,\mathbf{A}^{k-1}(t)$

代入有 $(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)f(\mathbf{A}(t)) =f(\mathbf{A}(t))(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t) + \mathbf{B}(t)f'(\mathbf{A}(t))$ (*)

由定义函数

$$f(\mathbf{A}(t))= \mathbf{P}(t)\begin{bmatrix} f(\mathbf{J}_1(t)) & & \\ & \ddots & \\ & & f(\mathbf{J}_s(t)) \end{bmatrix}\mathbf{P}^{-1}(t)$$

其中

$$f(\mathbf{J}_i)=\begin{bmatrix} f(\lambda_i(t)) & \frac{1}{1!}f'(\lambda_i(t)) & \cdots & \frac{1}{(m_i-1)!}f^{(m_i-1)}(\lambda_i(t)) \\ & f(\lambda_i(t)) & \ddots & \vdots \\ & & \ddots & \frac{1}{1!}f'(\lambda_i(t)) \\ & & & f(\lambda_i(t)) \end{bmatrix}_{m_i\times m_i}$$

因此 $d\,f(\mathbf{A}(t))/dt =(d\mathbf{P}(t)/dt)\mathbf{P}^{-1}(t)f(\mathbf{A}(t))+ \mathbf{P}(t)(df(\mathbf{J}(t))/dt)\,\mathbf{P}^{-1}(t) - f(\mathbf{A}(t))(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)$

注意到 $\mathbf{J}_i(\mathrm{t})$的特殊性质，我们有 $\mathbf{J}_i(t)(d(\mathbf{J}_i(t))/dt) =(d(\mathbf{J}_i(t))/dt)\mathbf{J}_i(t), i=1,2,\ldots,s$,

从而有

$\mathbf{J}(t)(d(\mathbf{J}(t))/dt) =(d(\mathbf{J}(t))/dt)\mathbf{J}(t)$

进一步得到

$d(f(\mathbf{J}(t)))/dt = (d(\mathbf{J}(t))/dt)\,f'(\mathbf{J}(t))$

代入可得

$d\,f(\mathbf{A}(t))/dt =(d\mathbf{P}(t)/dt)\mathbf{P}^{-1}(t)f(\mathbf{A}(t))+ \mathbf{P}(t)(d(\mathbf{J}(t))/dt)\,f'(\mathbf{J}(t))\mathbf{P}^{-1}(t) - f(\mathbf{A}(t))(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)$

$=(d\mathbf{P}(t)/dt)\mathbf{P}^{-1}(t)f(\mathbf{A}(t)) - f(\mathbf{A}(t))(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t) + \mathbf{P}(t)(d(\mathbf{J}(t))/dt)\mathbf{P}^{-1}(t)\,\mathbf{P}(t)\,f'(\mathbf{J}(t))\,\mathbf{P}^{-1}(t)$

$=(d\mathbf{P}(t)/dt)\mathbf{P}^{-1}(t)f(\mathbf{A}(t)) - f(\mathbf{A}(t))(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t) +\mathbf{P}(t)(d(\mathbf{J}(t))/dt)\mathbf{P}^{-1}(t)\,f'(\mathbf{A}(t))$

$= \mathbf{B}(t)f'(\mathbf{A}(t)) + \mathbf{P}(t)(d(\mathbf{J}(t))/dt)\mathbf{P}^{-1}(t)\,f'(\mathbf{A}(t))$

$= (\mathbf{B}(t)+ \mathbf{P}(t)(d(\mathbf{J}(t))/dt)\mathbf{P}^{-1}(t)\,)\,f'(\mathbf{A}(t))$

$= \mathrm{d}\,\mathbf{A}(t)/dt\cdot f'(\mathbf{A}(t))$

$$$

证明(形式证明，数学上不严格):当然，如果将 $f(z)$展开成幂级数形式 $f(z)=\sum_{k=0}^{\infty}c_k z^k$，

根据假设 $\mathbf{A}(t)$和 $\frac{d}{dt}\mathbf{A}(t)$可交换可得 $d(\mathbf{A}^k(t))/dt =k\,\mathbf{A}^{k-1}(t)\,d(\mathbf{A}(t))/dt$

代入矩阵函数 $d(f(\mathbf{A}(t)))/dt= \sum_{k=1}^{\infty}c_k k\cdot \boldsymbol{A}^{k-1}(t)\frac{d\boldsymbol{A}(t)}{dt} = \mathrm{d}\,\mathbf{A}(t)/dt\cdot f'(\mathbf{A}(t))$ 成立。

$$$

例 $\dfrac{d(\exp(A\cdot t))}{dt}=A\exp(A\cdot t)$

证明: 取 f(z)=exp(z), A(t)= A·t, 应用性质 4 可得结论。

考虑到性质 3.4.4 中，**A**(t)和 $\dfrac{d}{dt}$ **A**(t)可交换的条件比较强，我们可以给出如下的性质。

性质 3.4.5(迹函数求导基本定理). 如果 **A**(t)和 B(t)可交换，f(z)是与 t 无关的一元解析函数，

则有 $\dfrac{d}{dt}\mathrm{tr}(f(\mathbf{A}(t))\,\mathbf{B}(t))=\mathrm{tr}(f'(\mathbf{A}(t))\dfrac{d}{dt}\mathbf{A}(t)\mathbf{B}(t))+\mathrm{tr}(f(\mathbf{A}(t))\dfrac{d}{dt}\mathbf{B}(t))$

证明：首先，我们有

$$\frac{d}{dt}\mathrm{tr}(f(\mathbf{A}(t)\,\mathbf{B}(t))=\mathrm{tr}(\frac{d}{dt}f(\mathbf{A}(t))\,\mathbf{B}(t))+\mathrm{tr}(f(\mathbf{A}(t))\frac{d}{dt}\mathbf{B}(t)\quad)$$

设 $\mathbf{A}(t)\in C^{n\times n}$ 的 Jordan 标准形为 **J**(t),即有可逆矩阵 **P**(t)，使得

$$\mathbf{P}^{-1}(t)\mathbf{A}(t)\,\mathbf{P}(t)=\mathbf{J}(t)=\begin{bmatrix}\mathbf{J}_1(t) & & \\ & \ddots & \\ & & \mathbf{J}_s(t)\end{bmatrix},\ \mathbf{J}_i(t)=\begin{bmatrix}\lambda_i(t) & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i(t) & 1\\ & & & \lambda_i(t)\end{bmatrix}_{m_i\times m_i}$$

$$d\,\mathbf{A}(t)/dt=(d\mathbf{P}(t)/dt)\mathbf{J}(t)\,\mathbf{P}^{-1}(t)+\mathbf{P}(t)(d\mathbf{J}(t)/dt)\,\mathbf{P}^{-1}(t)+\mathbf{P}(t)\,\mathbf{J}(t)\,(d\mathbf{P}^{-1}(t)/dt)$$
$$=(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)\mathbf{A}(t)+\mathbf{P}(t)(d\mathbf{J}(t)/dt)\,\mathbf{P}^{-1}(t)-\mathbf{A}(t)(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)$$

因此可得：

$$\mathbf{P}(t)(d\mathbf{J}(t)/dt)\,\mathbf{P}^{-1}(t)=\mathbf{A}(t)(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)-(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)\mathbf{A}(t)+d\,\mathbf{A}(t)/dt$$

由定义函数

$$f(\mathbf{A}(t))=\mathbf{P}(t)\begin{bmatrix}f(\mathbf{J}_1(t)) & & \\ & \ddots & \\ & & f(\mathbf{J}_s(t))\end{bmatrix}\mathbf{P}^{-1}(t)$$

其中

$$f(\mathbf{J}_i)=\begin{bmatrix}f(\lambda_i(t)) & \frac{1}{1!}f'(\lambda_i(t)) & \cdots & \frac{1}{(m_i-1)!}f^{(m_i-1)}(\lambda_i(t))\\ & f(\lambda_i(t)) & \ddots & \vdots\\ & & \ddots & \frac{1}{1!}f'(\lambda_i(t))\\ & & & f(\lambda_i(t))\end{bmatrix}_{m_i\times m_i}$$

因此 $d\,f(\mathbf{A}(t))/dt=(d\mathbf{P}(t)/dt)\mathbf{P}^{-1}(t)f(\mathbf{A}(t))+\mathbf{P}(t)(df(\mathbf{J}(t))/dt)\,\mathbf{P}^{-1}(t)-f(\mathbf{A}(t))(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)$

注意到 $\mathbf{J}_i$(t)的特殊性质，我们有 $\mathbf{J}_i(t)(d(\mathbf{J}_i(t))/dt)=(d(\mathbf{J}_i(t))/dt)\mathbf{J}_i(t),i=1,2,\ldots,s$,

从而有

$$\mathbf{J}(t)(d(\mathbf{J}(t))/dt)=(d(\mathbf{J}(t))/dt)\mathbf{J}(t)$$

进一步得到

$$d(f(\mathbf{J}(t)))/dt=(d(\mathbf{J}(t))/dt)\,f'(\mathbf{J}(t))$$

代入可得

$d\,f(\mathbf{A}(t))/dt = (d\mathbf{P}(t)/dt)\mathbf{P}^{-1}(t)f(\mathbf{A}(t)) + \mathbf{P}(t)(d(\mathbf{J}(t))/dt)\,f'(\mathbf{J}(t))\mathbf{P}^{-1}(t) - f(\mathbf{A}(t))(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)$

$= (d\mathbf{P}(t)/dt)\mathbf{P}^{-1}(t)f(\mathbf{A}(t)) - f(\mathbf{A}(t))(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t) + \mathbf{P}(t)(d(\mathbf{J}(t))/dt)\mathbf{P}^{-1}(t)\,\mathbf{P}(t)\,f'(\mathbf{J}(t))\,\mathbf{P}^{-1}(t)$

$= (d\mathbf{P}(t)/dt)\mathbf{P}^{-1}(t)f(\mathbf{A}(t)) - f(\mathbf{A}(t))(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t) + \mathbf{P}(t)(d(\mathbf{J}(t))/dt)\mathbf{P}^{-1}(t)\,f'(\mathbf{A}(t))$

代入

$\mathbf{P}(t)(d\mathbf{J}(t)/dt)\,\mathbf{P}^{-1}(t) = \mathbf{A}(t)(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t) - (d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)\mathbf{A}(t) + \mathrm{d}\,\mathbf{A}(t)/\mathrm{dt}$

可得

$d\,f(\mathbf{A}(t))/dt = (d\mathbf{P}(t)/dt)\mathbf{P}^{-1}(t)f(\mathbf{A}(t)) - f(\mathbf{A}(t))(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t) + \mathbf{P}(t)(d(\mathbf{J}(t))/dt)\mathbf{P}^{-1}(t)\,f'(\mathbf{A}(t))$

$= \{\,(d\mathbf{P}(t)/dt)\mathbf{P}^{-1}(t)f(\mathbf{A}(t)) - f(\mathbf{A}(t))(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)\,\}$

$+ \{\,\mathbf{A}(t)(d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t) - (d\mathbf{P}(t)/dt)\,\mathbf{P}^{-1}(t)\mathbf{A}(t)\}\,f'(\mathbf{A}(t)) + (\mathrm{d}\,\mathbf{A}(t)/\mathrm{dt}\,)\,f'(\mathbf{A}(t))$

$= \mathrm{I} + \mathrm{II} + \mathrm{III}$

这样

$$\mathrm{tr}\left(\frac{d}{dt}f(\mathbf{A}(\mathrm{t}))\,\mathbf{B}(\mathrm{t})\right) = \mathrm{tr}(\mathrm{I}\cdot\mathbf{B}(\mathrm{t}) + \mathrm{II}\cdot\mathbf{B}(\mathrm{t}) + \mathrm{III}\cdot\mathbf{B}(\mathrm{t}))$$

由假设条件，A(t)B(t)=B(t)A(t)和迹算子的性质可得 $\mathrm{tr}(\mathrm{I}\cdot\mathbf{B}(\mathrm{t})) = \mathrm{tr}(\mathrm{II}\cdot\mathbf{B}(\mathrm{t})) = 0$

因此 $\mathrm{tr}\left(\frac{d}{dt}f(\mathbf{A}(\mathrm{t}))\,\mathbf{B}(\mathrm{t})\right) = \mathrm{tr}((\mathrm{d}\,\mathbf{A}(t)/\mathrm{dt}\,)\,f'(\mathbf{A}(t))\cdot\mathbf{B}(\mathrm{t}))$

从而可得

$$\frac{d}{dt}\mathrm{tr}(f(\mathbf{A}(\mathrm{t})\,\mathbf{B}(\mathrm{t})) = \mathrm{tr}\left(f'(\mathbf{A}(\mathrm{t}))\frac{d}{dt}\mathbf{A}(\mathrm{t})\,\mathbf{B}(\mathrm{t})\right) + \mathrm{tr}\left(f(\mathbf{A}(\mathrm{t}))\ \frac{d}{dt}\mathbf{B}(\mathrm{t})\right)$$

成立。

$$$

证明(形式证明，数学上不严格):当然，如果将 $f(z)$展开成幂级数形式 $f(z) = \sum_{k=-\infty}^{\infty} c_k z^k$，

根据假设 $\mathbf{A}(t)$和 $\mathbf{B}(t)$可交换，因此 对于 k≥0 的整数：

$$\mathrm{tr}[d(\mathbf{A}^k(t))/dt\,\mathbf{B}(t)] = \mathrm{tr}[\sum_{i=0}^{k-1}\boldsymbol{A}^i(t)\frac{d\boldsymbol{A}(t)}{dt}A^{k-1-i}(t)\boldsymbol{B}(t)]$$

$$= \mathrm{tr}[\sum_{i=0}^{k-1}\boldsymbol{A}^i(t)\frac{d\boldsymbol{A}(t)}{dt}\boldsymbol{B}(t)A^{k-1-i}(t)]\text{(由 }\mathbf{A}(t)\text{和 }\mathbf{B}(t)\text{可交换)}$$

$$= \mathrm{tr}[\sum_{i=0}^{k-1}A^{k-1-i}(t)\cdot\boldsymbol{A}^i(t)\frac{d\boldsymbol{A}(t)}{dt}\boldsymbol{B}(t)]\text{(由迹算子可交换性)}$$

$$= \mathrm{tr}\,[k\,\mathbf{A}^{k-1}(t)\,d(\mathbf{A}(t))/dt\,\mathbf{B}(t)] \qquad \text{(I)}$$

同样

$$\mathrm{tr}[d(\mathbf{A}^{-k}(t))/dt\,\mathbf{B}(t)] = \mathrm{tr}[\sum_{i=0}^{k-1}\boldsymbol{A}^{-i}(t)\frac{d\boldsymbol{A}(t)^{-1}}{dt}A^{-(k-1-i)}(t)\boldsymbol{B}(t)]$$

$$= \mathrm{tr}[-\sum_{i=0}^{k-1}\boldsymbol{A}^{-(i+1)}(t)\frac{d\boldsymbol{A}(t)}{dt}A^{-(k-i)}(t)\boldsymbol{B}(t)]$$

$$= \mathrm{tr}[-\sum_{i=0}^{k-1} \boldsymbol{A}^{-(i+1)}(t)\frac{d\boldsymbol{A}(t)}{dt}\boldsymbol{B}(t)A^{-(k-i)}(t)]$$ (由 **A**(*t*)和 g(**B**(*t*))可交换)

$$= \mathrm{tr}[-\sum_{i=0}^{k-1} A^{-(k-i)}(t)\cdot\boldsymbol{A}^{-(i+1)}(t)\frac{d\boldsymbol{A}(t)}{dt}\boldsymbol{B}(t)]$$ (由迹算子可交换性)

$$=\mathrm{tr}\,[-k\,\mathbf{A}^{-k-1}(t)\,d(\mathbf{A}(t))/dt\,\mathbf{B}(t)] \qquad \text{(II)}$$

综合(I)和(II)可得，对于任给的整数 k∈N，都有

$\mathrm{tr}[d(\mathbf{A}^k(t))/dt\,\mathbf{B}(t)]=\mathrm{tr}\,[k\,\mathbf{A}^{k-1}(t)\,d(\mathbf{A}(t))/dt\,\mathbf{B}(t)]$

代入矩阵函数得

$$\mathrm{tr}[d(f(\mathbf{A}(t)))/dt\,\mathbf{B}(t)]=\mathrm{tr}[\sum_{k=-\infty}^{\infty} c_k k\cdot\boldsymbol{A}^{k-1}(t)\tfrac{d\boldsymbol{A}(t)}{dt}\mathbf{B}(t)]$$

$$=\mathrm{tr}[f'(\mathbf{A}(t))\mathrm{d}\,\mathbf{A}(t)/dt\cdot\mathbf{B}(t)] \qquad (*)$$

成立。

其次，我们有

$$\frac{d}{dt}\mathrm{tr}(f(\mathbf{A}(\mathrm{t})\,\mathbf{B}(\mathrm{t}))=\mathrm{tr}(\frac{d}{dt}f(\mathbf{A}(\mathrm{t}))\,\mathbf{B}(\mathrm{t}))+\mathrm{tr}(f(\mathbf{A}(\mathrm{t}))\frac{d}{dt}\mathbf{B}(\mathrm{t})) \qquad (**)$$

综合以上两式(*)和(**)可得定理成立。

$$

推论 3.4.1. 如果 *f*(*z*)是与 *t* 无关的一元解析函数，

则有 $\frac{d}{dt}\mathrm{tr}(f(\mathbf{A}(\mathrm{t})))=\mathrm{tr}(f'(\mathbf{A}(\mathrm{t}))\frac{d}{dt}\mathbf{A}(\mathrm{t}))$

例 3.4.1：设非奇异矩阵 A(*t*)的对 *t* 可导，则 $d(|\boldsymbol{A}(t)|)/dt=|\boldsymbol{A}|\,\mathrm{tr}(\boldsymbol{A}(t)^{-1}\cdot d\boldsymbol{A}(t)/dt)$

证明：设方阵 A(t)的 n 个特征值为$\lambda_1,\lambda_2,\ldots,\lambda_n$，那么

$|\mathrm{A(t)}|=\lambda_1\cdot\lambda_2\cdots\lambda_n$

因此 $\ln(|\mathrm{A(t)}|)=\ln(\lambda_1)+\ln(\lambda_2\cdot)+\cdots+\ln(\lambda_n)=\mathrm{tr}[\ln(\mathrm{A(t)})]$

根据迹函数求导基本定理:

$$\frac{d\ln(|\boldsymbol{A}(t)|)}{dt}=\frac{d(\mathrm{tr}[\ln(\boldsymbol{A}(t))])}{dt}=\mathrm{tr}[\mathbf{A}(\mathrm{t})^{-1}\frac{d}{dt}\mathbf{A}(\mathrm{t})]$$

因此 $d(|\boldsymbol{A}(t)|)/dt=\mathrm{d}(\exp(\ln|\mathrm{A(t)}|))/\mathrm{dt}=\exp(\ln|\mathrm{A(t)}|)\mathrm{d}(\ln|\mathrm{A(t)}|)/\mathrm{dt}=|\boldsymbol{A}|\,\mathrm{tr}(\boldsymbol{A}(t)^{-1}\cdot d\boldsymbol{A}(t)/dt)$

这个例子的证明我们在下面还需要提到。

定义 3.10 如果矩阵 A(*t*)的每个元素 $a_{ij}(t)$都是区间$[t_0,t_1]$上的连续函数，则定义 A(t)在$[t_0,t_1]$上的积分为分量积分构成的矩阵. 即

$$\int_{t_0}^{t_1}\mathrm{A(t)\,dt}=(\int_{t_0}^{t_1}\mathrm{a_{ij}(t)dt})$$

性质：(积分算子仍为线性算子)

$$\int_{t_0}^{t_1}(A(t)+B(t))dt=\int_{t_0}^{t_1}A(t)dt+\int_{t_0}^{t_1}B(t)dt$$

$$\int_{t_0}^{t_1}P\cdot A(t)\cdot Qdt=P(\int_{t_0}^{t_1}A(t)dt)\,Q$$

当 $a_{ij}(t)$都在$[t_0,t_1]$上连续时，就称 A(t)在$[t_0,t_1]$上连续. 且有

$$\frac{d}{dt}\int_a^t A(s)ds=A(t)$$

当 $a_{ij}(t)$都在[a,b]上连续时

$$\int_a^b A'(s)ds=A(b)-A(a)$$

其它微分概念

I 函数对矩阵的导数(包括向量)

为了简单起见，在本节中我们记 $\boldsymbol{e}_i$ 为第 i 个分量为 1,而其余分量为 0 的单位列向量，**其维数根据上下文确定，**以后不一一说明。

定义 I.1:设 $\mathbf{X}=(\xi_{ij})_{m\times n}$,mn 元函数

$$f(\mathbf{X})=f(\xi_{11},\ldots,\xi_{1n},\xi_{21}\ldots,\xi_{m1,\ldots,}\xi_{mn})$$

定义 $f(\mathbf{X})$对矩阵 $\mathbf{X}$ 的导数为

$$\frac{df}{d\mathbf{X}}=\left(\frac{\partial f}{\partial\xi_{ij}}\right)_{m\times n}=\begin{bmatrix}\frac{\partial f}{\partial\xi_{11}} & \cdots & \frac{\partial f}{\partial\xi_{1n}}\\ \vdots & & \vdots\\ \frac{\partial f}{\partial\xi_{m1}} & \cdots & \frac{\partial f}{\partial\xi_{mn}}\end{bmatrix}\qquad\text{(分量形式)}$$

$$=\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\partial f}{\partial\xi_{ij}}\boldsymbol{e}_i\boldsymbol{e}_j^T\qquad\text{(代数形式)}$$

特别注意在上面的等式中，$\boldsymbol{e}_i$ 为 m 维列向量而 $\boldsymbol{e}_j$ 为 n 维列向量,在本节中我们会经常使用这种约定。

矩阵的元素和矩阵符号的关系

设 $A=(a_{ij})\in C^{m\times n}$，那么

$$\boxed{\boldsymbol{A}=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}\boldsymbol{e}_i\boldsymbol{e}_j^T}$$

$$a_{ij} = \boldsymbol{e}_i^T \boldsymbol{A} \boldsymbol{e}_j$$

$$\mathrm{Vec}(\boldsymbol{A}) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}\boldsymbol{e}_j \otimes \boldsymbol{e}_i = \sum_{i=1}^{m} \boldsymbol{A}^T\boldsymbol{e}_i \otimes \boldsymbol{e}_i = \sum_{j=1}^{n} \boldsymbol{e}_j \otimes \boldsymbol{A}\boldsymbol{e}_j$$

$$a_{ij} = (\boldsymbol{e}_j \otimes \boldsymbol{e}_i)^T vec(A)$$

这个关系式实际上可以为我们把一个关于分量的等式转化为关于矩阵-向量的形式。最后一个式子可以由第二个式子结合 vec(•)向量化算子的定义推得。

定理 I.1 (函数矩阵的微分形式) 设 $\mathbf{X}=(\xi_{ij})_{m\times n}$,mn 元函数 $f(\bullet)$: $C^{m\times n}\to C$：

$f(\mathbf{X})=f(\xi_{11}, \ldots, \xi_{1n}, \xi_{21}\ldots, \xi_{m1,\ldots,} \xi_{mn})$，

那么 $\dfrac{df}{d\mathbf{X}}=\boldsymbol{A}$ 的充要条件为 $\mathrm{d}f=\mathrm{tr}(\boldsymbol{A}^T\cdot \mathrm{d}\boldsymbol{X})=\mathrm{tr}(\boldsymbol{A}\cdot \mathrm{d}\boldsymbol{X}^T)$ (微分形式)

证明:根据定义

$$\mathrm{d}f = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{df}{dx_{ij}}dx_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} e_i^T\frac{df}{d\boldsymbol{X}}e_j\cdot e_i^T d\boldsymbol{X}e_j = \sum_{i=1}^{m}\sum_{j=1}^{n} e_j^T\left(\frac{df}{d\boldsymbol{X}}\right)^T e_i\cdot e_i^T d\boldsymbol{X}e_j$$

$$= \sum_{j=1}^{n} e_j^T\left(\frac{df}{d\boldsymbol{X}}\right)^T\left(\sum_{j=1}^{m} e_i\cdot e_i^T\right)d\boldsymbol{X}e_j = \sum_{j=1}^{n} e_j^T\left(\frac{df}{d\boldsymbol{X}}\right)^T d\boldsymbol{X}e_j = \mathrm{tr}\left[\left(\frac{df}{d\boldsymbol{X}}\right)^T d\boldsymbol{X}\right]$$

在进行进一步的推导之前，我们需要如下的一阶全微分的形式不变性原理：

1.一元函数**复合函数的微分法则和一阶微分的形式不变性**

设 $y=f(u)$及 $u=\varphi(x)$都可导，则复合函数 $y=f[\varphi(x)]$的微分为

$dy=y'_x\,dx=f'(u)\varphi'(x)dx$.

于由$\varphi'(x)dx=du$，所以，复合函数 $y=f[\varphi(x)]$的微分公式也可以写成

$dy=f'(u)du$ 或 $dy=y'_u\,du$.

由此可见，无论 u 是自变量还是另一个变量的可微函数，微分形式 $dy=f'(u)du$ 保持不变．这一性质称为微分形式不变性．这性质表示，当变换自变量时，微分形式 $dy=f'(u)du$ 并不改变.

2. 多元复合函数微分法则和一阶全微分的形式不变性

复合函数的中间变量均为一元函数的情形

如果函数 $u=\varphi(t)$ 及 $v=\psi(t)$ 都在点 t 可导，函数 $z=f(u,v)$ 在对应点 (u,v) 具有连续偏导数，则复合函数 $z=f(\varphi(t),\psi(t))$ 在点 t 可导，且有

$$\frac{\mathrm{d}z}{\mathrm{d}t}=\frac{\partial z}{\partial u}\cdot\frac{\mathrm{d}u}{\mathrm{d}t}+\frac{\partial z}{\partial v}\cdot\frac{\mathrm{d}v}{\mathrm{d}t}$$

因此 $dz=\frac{\partial z}{\partial u}\cdot du+\frac{\partial z}{\partial v}\cdot dv=\frac{dz}{dt}\cdot dt$,

复合函数的中间变量均为多元函数的情形

如果函数 $u=\varphi(x, y)$, $v=\psi(x, y)$都在点(x, y)具有对 x 及 y 的偏导数，函数 $z=f(u, v)$在对应点(u, v)具有连续偏导数，则复合函数 $z=f\,[\varphi(x, y)$，　$(x, y)]$在点(x, y)的两个偏导数存在，则有 $\frac{\partial z}{\partial x}=\frac{\partial z}{\partial u}\cdot\frac{\partial u}{\partial x}+\frac{\partial z}{\partial v}\cdot\frac{\partial v}{\partial x}$,　　$\frac{\partial z}{\partial y}=\frac{\partial z}{\partial u}\cdot\frac{\partial u}{\partial y}+\frac{\partial z}{\partial v}\cdot\frac{\partial v}{\partial y}$

因为 $du=(\partial u/\partial x)dx+(\partial u/\partial y)dy$, $dv=(\partial v/\partial x)dx+(\partial v/\partial y)dy$, ,

因此 $d\,z=(\partial z/\partial x)dx+(\partial z/\partial y)dy=(\partial z/\partial u)du+(\partial z/\partial v)dv$,

全微分形式不变性：无论 z 是自变量 x、y 的函数或中间变量 u 、v 的函数，它的全微分形式是一样的,这个性质叫做**全微分形式不变性**。

类似地，可以推广到多次复合函数的情形，我们称一阶微分的这种性质为一阶微分形式不变性。（高阶微分一般不具有这种性质）

性质 I.1 若 $f(t)=f(\mathbf{x})$, $\mathbf{x}=\mathbf{x}(t)$为向量，则 $df/dt=d\mathbf{x}^T/dt\cdot df/d\mathbf{x}$.

性质 I.2：设 $f(\boldsymbol{X})$是以 $m\times n$ 矩阵 $\boldsymbol{X}$ 的元素为变量的 mn 元函数，

而 $\boldsymbol{X}$ 又是标量 t 的函数，则

$df/dt=\mathrm{tr}(d\mathbf{X}^T/dt\cdot df/d\mathbf{X})$

证明：根据一阶微分不变性可得：

$df=\mathrm{tr}\,[(df/d\mathbf{X})^T\,d\mathbf{X}\,]$, $d\boldsymbol{X}=(d\mathbf{X}/dt)\,dt$,

代入有

$df=\mathrm{tr}\,[(df/d\mathbf{X})^T(d\mathbf{X}/dt)\,dt]=\mathrm{tr}\,[(df/d\mathbf{X})^T(d\mathbf{X}/dt)\,]dt$

$=\mathrm{tr}\,[(d\mathbf{X}/dt)^T(df/d\mathbf{X})]dt$

从而可得结论。

性质 I.3：关于迹函数的导数

$d(\mathrm{tr}(\boldsymbol{X}^T\boldsymbol{A}))/d\boldsymbol{X}=\boldsymbol{A}$.

$d(\mathrm{tr}(\mathbf{A}^T\mathrm{X}))/d\mathbf{X}=\mathbf{A}$.

$d(\mathrm{tr}(\mathbf{A}\mathrm{X}))/d\mathbf{X}=\mathbf{A}^T$.

证明：　设 $y=\mathrm{tr}(\boldsymbol{X}^T\boldsymbol{A})$, 那么 $dy=\mathrm{tr}((d\boldsymbol{X})^T\boldsymbol{A})=\mathrm{tr}(\boldsymbol{A}^Td\boldsymbol{X})$

根据函数矩阵的微分形式可得 $d(\mathrm{tr}(\boldsymbol{X}^T\boldsymbol{A}))/d\boldsymbol{X}=\boldsymbol{A}$.

其余类似。

性质 I.4 设 $\boldsymbol{X}$ 为可逆矩阵，$f(\boldsymbol{X})=\mathrm{tr}(\boldsymbol{A}\boldsymbol{X}^{-1})$，其中 $\boldsymbol{A}$ 为常数矩阵，那么

$d\,f(\boldsymbol{X})/d\boldsymbol{X}=-(\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}^{-1})^T$

证明:令 $\boldsymbol{Y}=\boldsymbol{X}^{-1}$, 那么 $f=\mathrm{tr}[\boldsymbol{A}\boldsymbol{Y}]$,从而 $df=\mathrm{tr}[\boldsymbol{A}\cdot d\boldsymbol{Y}]$

根据 $\boldsymbol{X}\boldsymbol{Y}=\boldsymbol{E}_n$，那么 $d\boldsymbol{X}\cdot\boldsymbol{Y}+\boldsymbol{X}\cdot d\boldsymbol{Y}=0$,因此 $d\boldsymbol{Y}=-\boldsymbol{Y}\cdot d\boldsymbol{X}\cdot\boldsymbol{Y}$

代入有 $df=\mathrm{tr}[\boldsymbol{A}\cdot(-\boldsymbol{Y}\cdot d\boldsymbol{X}\cdot\boldsymbol{Y})]=\mathrm{tr}[-\boldsymbol{Y}\boldsymbol{A}\boldsymbol{Y}\cdot d\boldsymbol{X}]$

根据性质 I.1(函数矩阵的微分形式) 可得结论。

性质 I.5 (迹函数矩阵求导基本定理). 如果 $f(z)$是一元解析函数，

$y=\mathrm{tr}[\,f(\mathbf{A})]$，则 $\frac{dy}{d\boldsymbol{A}}=f'(\boldsymbol{A})^T$

证明：由矩阵导数的代数形式可得

$$\frac{dy}{d\boldsymbol{A}}=\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial y}{\partial a_{ij}}\boldsymbol{e}_i\boldsymbol{e}_j^T=\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial tr[f(A)]}{\partial a_{ij}}\boldsymbol{e}_i\boldsymbol{e}_j^T=\sum_{i=1}^{n}\sum_{j=1}^{n}tr\left[f'(A)\frac{\partial A}{\partial a_{ij}}\right]\boldsymbol{e}_i\boldsymbol{e}_j^T$$

$$=\sum_{i=1}^{n}\sum_{j=1}^{n}tr\left[f'(A)\boldsymbol{e}_i\boldsymbol{e}_j^T\right]\boldsymbol{e}_i\boldsymbol{e}_j^T=\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{e}_i[\boldsymbol{e}_j^T f'(A)\boldsymbol{e}_i]\boldsymbol{e}_j^T=\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{e}_i[\boldsymbol{e}_i^T f'(A)^T\boldsymbol{e}_j]\boldsymbol{e}_j^T$$

$$=\sum_{i=1}^{n}(\boldsymbol{e}_i\boldsymbol{e}_i^T)f'(A)^T\sum_{j=1}^{n}(\boldsymbol{e}_j\boldsymbol{e}_j^T)=f'(A)^T$$

推论 I.1 (迹函数矩阵微分) 如果 $f(z)$是一元解析函数，y=tr[$f(\mathbf{A})$]，则 $dy=tr[f'(\boldsymbol{A})d\boldsymbol{A}]$

证明：由性质 I.5 和迹函数的微分形式显然可得。

例 I.1　$f(\mathbf{x})=\mathbf{x}^T\mathbf{A}\mathbf{x}$, 则 $df/dx=(\mathbf{A}+\mathbf{A}^T)x$,

特别地，若 A 为对称矩阵，$g(x)=\mathbf{x}^T\mathbf{A}\mathbf{x}/2$,

则 $dg/dx=\mathbf{A}\mathbf{x}$.

证明：由于 $df/dx_i=(d\mathbf{x}/dx_i)^T\mathrm{A}x+\mathbf{x}^T\mathrm{A}(d\mathbf{x}/dx_i)=e_i^T\mathbf{A}\mathbf{x}+\mathbf{x}^T\mathbf{A}\mathbf{e}_i=e_i^T(\mathbf{A}+\mathbf{A}^T)\mathbf{x}$

所以 $df/d\mathbf{x}=\sum_{i=1}^{n}\frac{df}{dx_i}e_i=\sum_{i=1}^{n}e_ie_i^T(\mathbf{A}+\mathbf{A}^T)\mathbf{x}=(\mathbf{A}+\mathbf{A}^T)\mathbf{x}$,

另证: $df=(d\mathbf{x})^T\mathbf{A}\mathbf{x}+\mathbf{x}^T\mathbf{A}(d\mathbf{x})=\mathbf{x}^T\mathbf{A}^T(d\mathbf{x})+\mathbf{x}^T\mathbf{A}(d\mathbf{x})=\mathrm{tr}[(\mathbf{A}\mathbf{x}+\mathbf{A}^T\mathbf{x})^Td\mathbf{x}]$

根据函数矩阵的微分形式可得 $df/dx=(\mathbf{A}+\mathbf{A}^T)x$.

例 I.2 设 $f(\boldsymbol{X})=\mathrm{tr}((\boldsymbol{A}\boldsymbol{X}^T\boldsymbol{B}+\boldsymbol{C})^{-1}\boldsymbol{D})$，证明

$$d\,f(\boldsymbol{X})/d\boldsymbol{X}=-B(AX^TB+C)^{-1}D(AX^TB+C)^{-1}A$$

其中 **A,B,C,D 为与 X 无关的常数矩阵。**

证明：设 $\boldsymbol{Y}=\boldsymbol{A}\boldsymbol{X}^T\boldsymbol{B}+\boldsymbol{C}$, 那么 $d\boldsymbol{Y}=\boldsymbol{A}\cdot d\boldsymbol{X}^T\cdot\boldsymbol{B}$;

根据性质 I.4 可得 $df=\mathrm{tr}[-\boldsymbol{Y}^{-1}\boldsymbol{D}\boldsymbol{Y}^{-1}\cdot d\boldsymbol{Y}]$,

因此，$df=\mathrm{tr}[-\boldsymbol{Y}^{-1}\boldsymbol{D}\boldsymbol{Y}^{-1}\boldsymbol{A}\cdot d\boldsymbol{X}^T\cdot\boldsymbol{B}]=\mathrm{tr}[-\boldsymbol{B}\boldsymbol{Y}^{-1}\boldsymbol{D}\boldsymbol{Y}^{-1}\boldsymbol{A}\cdot d\boldsymbol{X}^T]$

根据性质 I.1(函数矩阵的微分形式) 可得结论。

例 I.3　1)设非奇异矩阵 $\mathrm{A}(t)$的对 t 可导，则 $d(|\boldsymbol{A}|)/dt=|\boldsymbol{A}|\,\mathrm{tr}(\boldsymbol{A}^{-1}\cdot d\boldsymbol{A}/dt)$

2)对于任给非奇异矩阵 **A**, $d(\ln|\boldsymbol{A}|)/d\boldsymbol{A}=(\boldsymbol{A}^T)^{-1}$

3)对于任给 $\mathbf{A}\in\mathrm{R}^{m\times n}$, $d(\ln|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}|)/d\mathbf{A}=2\mathbf{A}(\delta\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}$,其中$\delta$为和矩阵 **A** 无关的常数。

证明：1）由于 $d|\boldsymbol{A}|/dt=\sum_{i=1}^{n}|(a_1,a_2,...,\frac{da_i}{dt},...,a_n)|$

$$=|\boldsymbol{A}|\sum_{i=1}^{n}\frac{|(\boldsymbol{a}_1,\boldsymbol{a}_2,...,\frac{d\boldsymbol{a}_i}{dt},...,\boldsymbol{a}_n)|}{|\boldsymbol{A}|}$$

考虑线性方程组 $\boldsymbol{A}\cdot\boldsymbol{x}=d\mathbf{a}_i/dt$, 其中 $\boldsymbol{a}_i$ 为矩阵 $\boldsymbol{A}$ 的第 i 列。
根据 Gramer 法则可得

$$x_i=|(\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\frac{d\boldsymbol{a}_i}{dt},\ldots,\boldsymbol{a}_n)|/|\boldsymbol{A}|$$

另一方面，我们知道 $\boldsymbol{x}=\boldsymbol{A}^{-1}\cdot d\mathbf{a}_i/dt$, 所以有

$$x_i=|(\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\frac{d\boldsymbol{a}_i}{dt},\ldots,\boldsymbol{a}_n)|/|\boldsymbol{A}|=\boldsymbol{e}_i^T\boldsymbol{A}^{-1}\frac{d\boldsymbol{a}_i}{dt}$$

因此有 $d|\boldsymbol{A}|/dt \;=\; |\boldsymbol{A}|\sum_{i=1}^{n}e_i^T A^{-1}\frac{da_i}{dt}$

$$=|\boldsymbol{A}|\sum_{i=1}^{n}e_i^T A^{-1}\frac{dA}{dt}e_i$$

$$=|\boldsymbol{A}|\,\mathrm{tr}(\boldsymbol{A}^{-1}\cdot d\boldsymbol{A}/dt)$$

1) $d(|\boldsymbol{A}|)/dt=|\boldsymbol{A}|\,\mathrm{tr}[\boldsymbol{A}^{-1}\cdot d\boldsymbol{A}/\mathrm{d}t]$

2). 由于 $d(|\boldsymbol{A}|)/d\boldsymbol{A}=\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{d|\mathbf{A}|}{da_{ij}}e_ie_j^T$

根据 1)的结果我们知道 $d(|\boldsymbol{A}|)/da_{ij}=|\boldsymbol{A}|\,\mathrm{tr}[\boldsymbol{A}^{-1}\cdot d\boldsymbol{A}/da_{ij}]=|\boldsymbol{A}|\,\mathrm{tr}[\boldsymbol{A}^{-1}\cdot e_i e_j^T]=|\boldsymbol{A}|\,e_j^T\boldsymbol{A}^{-1}e_i$

因此 $d(|\boldsymbol{A}|)/d\boldsymbol{A}=\sum_{i=1}^{n}\sum_{j=1}^{n}|A|e_j^T\mathbf{A}^{-1}e_i\cdot e_ie_j^T=\sum_{i=1}^{n}\sum_{j=1}^{n}|A|e_ie_i^T(\mathbf{A}^{-1})^Te_je_j^T=|\mathbf{A}|(\mathbf{A}^{-1})^T$

从而 $d(\ln|\boldsymbol{A}|)/d\boldsymbol{A}=(1/|\boldsymbol{A}|)\cdot d(|\boldsymbol{A}|)/d\boldsymbol{A}=(\boldsymbol{A}^{\mathrm{T}})^{-1}$

$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$

或者 $\det(\boldsymbol{A})=\sum_{j=1}^{n}a_{ij}\boldsymbol{A}_{ij}$，其中 A_{ij} 为在矩阵 $\boldsymbol{A}$ 中

去掉第 i 行第 j 列的代数余子式。排列 A_{ij} 得如下矩阵

$$\tilde{\boldsymbol{A}}=\begin{bmatrix}A_{11}&A_{12}&\cdots&A_{1n}\\A_{21}&A_{22}&\cdots&A_{2n}\\\vdots&\vdots&\ddots&\vdots\\A_{n1}&A_{n2}&\cdots&A_{nn}\end{bmatrix}$$

则根据代数余子式的定义可得

$$\boldsymbol{A}(\tilde{\boldsymbol{A}})^{\mathrm{T}}=|\boldsymbol{A}|\,\boldsymbol{I}$$

因此 $\tilde{\boldsymbol{A}}=|\boldsymbol{A}|\,(\boldsymbol{A}^{\mathrm{T}})^{-1}$.

注意 A_{ij} 显然和 a_{ij} 无关，因此我们有 $\mathrm{d}(|\boldsymbol{A}|)/\mathrm{d}\boldsymbol{A}=\tilde{\boldsymbol{A}}=|\boldsymbol{A}|\,(\boldsymbol{A}^{T})^{-1}$

$$

3) 由于 $d(|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}|)/d\boldsymbol{A}=\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{d\,|\mathbf{A}|}{da_{ij}}e_ie_j^T$

根据 1)的结果我们知道

$d(|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}\,|)/da_{ij}=|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}|\ \mathrm{tr}[(\delta\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}\cdot\ d(\delta\mathbf{I}+\mathbf{A}^T\mathbf{A})/da_{ij}]$

$=|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}|\ \mathrm{tr}[(\delta\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}\cdot\ (e_j e_i^T\ \mathbf{A}+\mathbf{A}^T e_i e_j^T\ )]$

$=|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}|\{\ \mathrm{tr}[(\delta\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}\cdot e_j e_i^T\ \mathbf{A}]+\ \mathrm{tr}[(\delta\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T e_i e_j^T\ )]\}$

$=|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}|\{\ [\ e_i^T\ \mathbf{A}(\delta\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}\cdot e_j+e_j^T\ (\delta\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T e_i)]\}$

$=2\cdot|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}|\ \ e_i^T\ \mathbf{A}(\delta\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}e_j$

因此 $d(|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}\,|)/d\boldsymbol{A}=2\cdot|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}|\sum_{i=1}^{m}\sum_{j=1}^{n}e_i^T\mathbf{A}(\delta\cdot\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}e_j\cdot e_ie_j^T$

$=2\cdot|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}|\sum_{i=1}^{m}\sum_{j=1}^{n}e_ie_i^T\mathbf{A}(\delta\cdot\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}e_je_j^T$

$=2\cdot|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}|(\sum_{i=1}^{m}e_ie_i^T)\mathbf{A}(\delta\cdot\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}\sum_{j=1}^{n}e_je_j^T$

$=2\cdot|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}|\ \cdot\mathbf{A}(\delta\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}$

从而 $d(\ln|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}|)/d\boldsymbol{A}=(1/|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}\,|)\cdot d(|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}\,|)/d\boldsymbol{A}=2\cdot\mathbf{A}(\delta\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}$

另证：记 $\mathbf{B}=\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}$, $y=\log|\delta\mathbf{I}+\mathbf{A}^T\mathbf{A}|=\ln|\mathbf{B}|$,

因为 $\mathbf{dy}/d\mathbf{B}=(\mathbf{B}^T)^{-1}$，因此 $dy=\mathrm{tr}[(\mathbf{B})^{-1}d\mathbf{B}]$. **考虑到** $d\mathbf{B}=(d\mathbf{A})^T\mathbf{A}+\mathbf{A}^Td\mathbf{A}$,代入有

$dy=\mathrm{tr}[(\mathbf{B})^{-1}(d\mathbf{A})^T\mathbf{A}]+\mathrm{tr}[(\mathbf{B})^{-1}\mathbf{A}^Td\mathbf{A}]=\mathrm{tr}[\mathbf{A}(\mathbf{B})^{-1}(d\mathbf{A})^T]+\mathrm{tr}[(\mathbf{B})^{-1}\mathbf{A}^Td\mathbf{A}]$

$=\mathrm{tr}[2(\mathbf{B})^{-1}\mathbf{A}^Td\mathbf{A}]$,根据函数矩阵的微分形式得

$dy/d\mathbf{A}=2\mathbf{A}(\mathbf{B})^{-1}=2\cdot\mathbf{A}(\delta\mathbf{I}+\mathbf{A}^T\mathbf{A})^{-1}$.从而结论得证。

前面讨论的函数对矩阵的微分定义没有考虑特殊矩阵的性质，所以有时在公式表达时针对某些特殊矩阵并不能得到很简洁的表达性质。

举例来说，对于对称矩阵 **X** 来说，我们考虑函数 $f(\mathbf{X})=\mathrm{tr}(\mathbf{A}^T\mathbf{X})$, 其中 **A** 为任意 n 阶矩阵，如果仍然使用前面的关于函数对矩阵的微分定义我们有

$$df/d\mathbf{X}=\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\partial f}{\partial \xi_{ij}}\boldsymbol{e}_i\boldsymbol{e}_j^T=\mathbf{A}+\mathbf{A}^T-\mathbf{D}$$

其中 **D** 表示 **A** 的对角部分构成的对角矩阵。这种表达形式看起来不是很简洁。
这是因为对于对称矩阵 **A** 来说,我们有如下关系式

$$\boldsymbol{A}=\sum_{i=1}^{m}a_{ii}\boldsymbol{e}_i\boldsymbol{e}_i^T+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}a_{ij}(\boldsymbol{e}_i\boldsymbol{e}_j^T+\boldsymbol{e}_j\boldsymbol{e}_i^T)$$

$$a_{ij}=\boldsymbol{e}_i^T\boldsymbol{A}\boldsymbol{e}_j=\boldsymbol{e}_j^T\boldsymbol{A}\boldsymbol{e}_i$$

由前面的讨论，在很多情形下，如果 **X** 为 $n\times n$ 对称矩阵，$f(\mathbf{X})$为 $n(n-1)$个变元的函数

$$f(\mathbf{X})=f(\xi_{11},\ \ldots,\ \xi_{1n},\ \xi_{22}\ldots,\ \xi_{2n,\ldots,}\ \xi_{nn})$$

那么定义

$$\frac{df}{d\mathbf{X}}=\begin{bmatrix}\dfrac{\partial f}{\partial \xi_{11}} & \cdots & \dfrac{1}{2}\dfrac{\partial f}{\partial \xi_{1n}}\\ \vdots & & \vdots\\ \dfrac{1}{2}\dfrac{\partial f}{\partial \xi_{m1}} & \cdots & \dfrac{\partial f}{\partial \xi_{mn}}\end{bmatrix}==\sum_{i=1}^{n}\frac{\partial f}{\partial \xi_{ii}}\boldsymbol{e}_i\boldsymbol{e}_i^T+\sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{2}\frac{\partial f}{\partial \xi_{ij}}\boldsymbol{e}_i\boldsymbol{e}_j^T$$

这时对于对称矩阵的函数 $f(\mathbf{X})=\mathrm{tr}(\mathbf{A}^T\mathbf{X})$，我们有 $df/d\mathbf{X}=(\mathbf{A}+\mathbf{A}^T)/2$,特别地，当 A 也为对称矩阵时有 $d\,(\mathrm{tr}(\mathbf{AX}))/d\mathbf{X}=\mathbf{A}$。

II 矩阵函数对矩阵的导数

(主要考虑向量值函数对向量的导数)

为了进行下面的讨论我们引入 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的直积定义.

定义 5.9 设 $\boldsymbol{A}=(a_{ij})\in C^{m\times n}$,$\boldsymbol{B}=(b_{ij})\in C^{p\times q}$,则称

如下的分块矩阵

$$\boldsymbol{A}\otimes\boldsymbol{B}=\begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B}\cdots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B}\cdots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B}\cdots & a_{mn}\boldsymbol{B} \end{bmatrix}\in C^{mp\times nq} \quad (5.4.1)$$

为 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的直积（Kronecker 积）。

性质 II.1：如果 **A** 为 $m\times1$ 的向量和 **B** 为 $n\times1$ 的向量时，那么

$$\mathbf{A}\otimes\mathbf{B}^T=\mathbf{A}\,\mathbf{B}^T$$

定义 3.13 设 $X=(\xi_{ij})_{m\times n}$ 的 mn 元函数

$$f_{ij}(\boldsymbol{X})=f_{ij}(\xi_{11},\ldots,\xi_{1n},\xi_{21}\ldots,\xi_{m1},\ldots,\xi_{mn})$$

定义矩阵

$$\boldsymbol{F}(\boldsymbol{X})=\begin{bmatrix} f_{11} & \cdots & f_{1s} \\ \vdots & & \vdots \\ f_{r1} & \cdots & f_{rs} \end{bmatrix}$$

对矩阵 $\boldsymbol{X}$ 的导数如下：

$$\frac{dF}{dX}=\begin{bmatrix} \frac{\partial F}{\partial\xi_{11}} & \cdots & \frac{\partial F}{\partial\xi_{1n}} \\ \vdots & & \vdots \\ \frac{\partial F}{\partial\xi_{m1}} & \cdots & \frac{\partial F}{\partial\xi_{mn}} \end{bmatrix}=\sum_{i=1}^{m}\sum_{j=1}^{n}(\boldsymbol{e}_i\boldsymbol{e}_j^T)\otimes\frac{\partial\boldsymbol{F}}{\partial\xi_{ij}}$$

其中

$$\frac{\partial F}{\partial\xi_{ij}}=\begin{bmatrix} \frac{\partial f_{11}}{\partial\xi_{ij}} & \cdots & \frac{\partial f_{1s}}{\partial\xi_{ij}} \\ \vdots & & \vdots \\ \frac{\partial f_{r1}}{\partial\xi_{ij}} & \cdots & \frac{\partial f_{rs}}{\partial\xi_{ij}} \end{bmatrix}=\sum_{k=1}^{r}\sum_{l=1}^{s}\frac{\partial f_{kl}}{\partial\xi_{ij}}\boldsymbol{e}_k\boldsymbol{e}_l^T$$

特别的，当 **F** 为 $m\times1$ 的向量和 **X** 为 $n\times1$ 的向量时，那么

$$\frac{dF}{dX^T}=\begin{bmatrix}\frac{\partial f_1}{\partial \xi_1} & \cdots & \frac{\partial f_1}{\partial \xi_n}\\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial \xi_1} & \cdots & \frac{\partial f_m}{\partial \xi_n}\end{bmatrix}$$

对于前面的定义我们可以进行形式化的表示为：

$$\frac{dF}{dX}=\left(\frac{1}{dX}\right)\otimes dF$$

$$

类似地，我们可以定义 $\frac{dF}{dX}=dF\otimes\left(\frac{1}{dX}\right)$

当然只是形式上的表示而已.

$$

性质 II.1： $\frac{dF}{dX}=\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{r}\sum_{l=1}^{s}\frac{df_{kl}}{d\xi_{ij}}(\boldsymbol{e}_i\boldsymbol{e}_j^T)\otimes(\boldsymbol{e}_k\boldsymbol{e}_l^T)$

证明：根据 $\frac{\partial F}{\partial \xi_{ij}}=\sum_{k=1}^{r}\sum_{l=1}^{s}\frac{\partial f_{kl}}{\partial \xi_{ij}}\boldsymbol{e}_k\boldsymbol{e}_l^T$ 代入 $\frac{dF}{dX}$ 的定义可得。

实际上有所谓的转换定理

转换定理：设 **X**=(x_{ij})和 **Y**=(y_{ij})分别是 $m\times n$ 和 $p\times q$ 的矩阵，
A,**B**,**C** 和 **D** 分别为 $p\times m, n\times q, p\times n$ 和 $m\times q$ 的矩阵，它们可以是
X，**Y** 的函数，则如下两条是等价的

(1) $\frac{\partial Y}{\partial x_{ij}}=AE_{ij}(m\times n)B+CE_{ij}^T(m\times n)D$， $i=1,...,m,\ j=1,...,n$

(2) $\frac{\partial y_{ij}}{\partial X}=A^TE_{ij}(p\times q)B^T+DE_{ij}^T(p\times q)C$， $i=1,...,p,\ j=1,...,q$

其中 $\boldsymbol{E}_{ij}(m\times n)$表示第 i 行第 j 列元素为 1，而其余元素为 0 的 $m\times n$ 矩阵.
证明：为了简单起见，我们记 $\boldsymbol{e}_i$ 为第 i 个分量为 1,而其余分量为 0 的
单位列向量，其维数根据上下文确定。
首先证明(1)$\to$(2), 实际上

$$\begin{aligned}\frac{\partial y_{kl}}{\partial x_{ij}}&=e_k^T\frac{\partial Y}{\partial x_{ij}}e_l\\ &=e_k^TAe_ie_j^TBe_l+e_k^TCe_je_i^TDe_l\\ &=a_{ki}b_{jl}+c_{kj}d_{il}\end{aligned}$$

由于

$$\frac{\partial y_{ij}}{\partial X}=\sum_{s=1}^{m}\sum_{t=1}^{n}\frac{\partial y_{ij}}{\partial x_{st}}e_s e_t^T$$

$$=\sum_{s=1}^{m}\sum_{t=1}^{n}(a_{is}b_{tj}+c_{it}d_{sj})e_s e_t^T$$

$$=\sum_{s=1}^{m}\sum_{t=1}^{n}(e_s^T A^T e_i e_j^T B^T e_t+e_s^T D e_j e_i^T C e_t)e_s e_t^T$$

$$=\sum_{s=1}^{m}\sum_{t=1}^{n}e_s e_s^T A^T e_i e_j^T B^T e_t e_t^T+\sum_{s=1}^{m}\sum_{t=1}^{n}e_s e_s^T D e_j e_i^T C e_t e_t^T$$

$$=\sum_{s=1}^{m}e_s e_s^T(A^T e_i e_j^T B^T)\sum_{t=1}^{n}e_t e_t^T+\sum_{s=1}^{m}e_s e_s^T(D e_j e_i^T C)\sum_{t=1}^{n}e_t e_t^T$$

$$=A^T E_{ij}(p\times q)B^T+DE_{ij}^T(p\times q)C$$

而证明(2)→(1)是一个类似的过程。

性质 II.2： $\mathbf{y}=\mathbf{W}\mathbf{x}$, 函数 $f(\mathbf{y})$是向量 $\mathbf{y}$ 的函数，其中 $\mathbf{W}\in C^{m\times n}$ 和 $\mathbf{x}\in C^n$ 无关

则 $d\mathbf{y}^T/d\mathbf{x}=\mathbf{W}^T$, $d(f(\mathbf{y}))/d\mathbf{W}=d(f(\mathbf{y}))/d\mathbf{y}\cdot \mathbf{x}^T$

性质 II.3：设 $f(\mathbf{x})$是向量 $\mathbf{x}$ 的函数，而 $\mathbf{x}$ 又是 $\mathbf{u}$ 的函数，则

$df/d\mathbf{u}=d\mathbf{x}^T/d\mathbf{u}\cdot df/d\mathbf{x}$

根据定义 $d\mathbf{x}^T/d\mathbf{u}=(d\mathbf{x}^T/du_1, d\mathbf{x}^T/du_2,.., d\mathbf{x}^T/du_n)^T$

其中 $d\mathbf{x}^T/du_i=(dx_1/du_i,\ldots, dx_n/du_i)$

类似地，设 $f(\mathbf{x})$是向量 $\mathbf{x}$ 的函数，而 $\mathbf{x}$ 又是向量 $\mathbf{u}$ 的

的向量值函数，$\mathbf{u}$ 是 $\mathbf{v}$ 的向量值函数，则

$df/d\mathbf{v}=d\mathbf{u}^T/d\mathbf{v}\cdot d\mathbf{x}^T/d\mathbf{u}\cdot df/d\mathbf{x}$

证明：由题设 $f(\mathbf{x})$是向量 $\mathbf{u}$ 的函数，所以

$df/d\mathbf{v}= d\mathbf{u}^T/d\mathbf{v}\cdot df/d\mathbf{u}$

而 $df/d\mathbf{u}= d\mathbf{x}^T/d\mathbf{u}\cdot df/d\mathbf{x}$

从而 $df/d\mathbf{v}=d\mathbf{u}^T/d\mathbf{v}\cdot d\mathbf{x}^T/d\mathbf{u}\cdot df/d\mathbf{x}$ 成立。

性质 II.4(向量函数关于向量导数的微分形式)

设向量 $\boldsymbol{y}=F(\boldsymbol{x})$是向量 $\boldsymbol{x}$ 的向量值函数，那么

$d\boldsymbol{y}=\boldsymbol{A}d\boldsymbol{x}$ 的充要条件为 $\boldsymbol{A}=d\boldsymbol{y}/d\boldsymbol{x}^T$

证明：根据定义可得结论。

应用：考虑如下前馈神经网络模型模型的学习问题：

$\boldsymbol{u}=\boldsymbol{W}\boldsymbol{x}$

$\boldsymbol{v}=\varphi(\boldsymbol{u})$ (即 $v_i=\varphi(u_i)$, 其中, $\varphi(\cdot)$为一元可微函数)

$\boldsymbol{p}=\boldsymbol{U}\boldsymbol{v}$

$\boldsymbol{q}=\phi(\boldsymbol{p})$ (即 $q_i=\phi(p_i)$, 其中, $\phi(\cdot)$为一元可微函数)

$\boldsymbol{o} = \boldsymbol{Vq}$

$e=(1/2)\cdot \|\boldsymbol{y}-\boldsymbol{o}\|^2$

其中：$\boldsymbol{x}$, $\boldsymbol{u},\boldsymbol{v},\boldsymbol{p},\boldsymbol{q},\boldsymbol{o}$ 为向量,$\boldsymbol{W},\boldsymbol{U},\boldsymbol{V}$ 为矩阵.

求 $de/d\boldsymbol{V}$, $de/d\boldsymbol{U}$,$de/d\boldsymbol{W}$.

解：$de/d\boldsymbol{o}=\mathbf{o}-\mathbf{y}$，$de/d\mathbf{V}=de/d\boldsymbol{o}\cdot\boldsymbol{q}^T$

$de/d\mathbf{U}=de/d\boldsymbol{p}\cdot\boldsymbol{v}^{\mathrm{T}}=(d\boldsymbol{q}^{\mathrm{T}}/d\mathbf{p})(d\boldsymbol{o}^T/d\boldsymbol{q})\cdot de/d\boldsymbol{o}\cdot\boldsymbol{v}^T$

$=\mathrm{diag}(\phi'(p_i))\boldsymbol{V}^T de/d\boldsymbol{o}\cdot\boldsymbol{v}^T$

其中 $\mathrm{diag}(\phi'(p_i))=\mathrm{diag}((\phi'(p_1), \phi'(p_2),\ldots, \phi'(p_n))$为对角矩阵,后面类似.

$de/d\boldsymbol{W} =de/d\boldsymbol{u}\cdot\boldsymbol{x}^T$

$= (d\mathbf{v}^T/d\mathbf{u})(d\mathbf{p}^T/d\mathbf{v})\cdot de/d\mathbf{p}\cdot\boldsymbol{x}^T$

$=\mathrm{diag}(\varphi'(u_i))\boldsymbol{U}^{\mathrm{T}} de/d\mathbf{p}\cdot\boldsymbol{x}^T$

定义$\boldsymbol{\delta}_3= \mathbf{o}-\mathbf{y}$, $\boldsymbol{\delta}_2 = de/d\boldsymbol{p}$, $\boldsymbol{\delta}_1 = de/d\boldsymbol{u}$, 那么有

$\boldsymbol{\delta}_2 = \mathrm{diag}(\phi'(p_i))\boldsymbol{V}^T\boldsymbol{\delta}_3$, $\boldsymbol{\delta}_1=\mathrm{diag}(\varphi'(u_i))\boldsymbol{U}^{\mathrm{T}}\boldsymbol{\delta}_2$,

这就是在前馈神经网络模型中著名的误差反向传播算法。

其中，$\boldsymbol{x}$ 为网络输入，$\boldsymbol{y}$ 为网络的期望输出，而 $\boldsymbol{o}$ 则为网络实际输出，$\boldsymbol{\delta}_3$ 为误差.