§5.4 矩阵的微分

矩阵微分的引进，为微分方程组的研究带来方便。因此，它是系统工程与控制论的基础，为了叙述简明起见，略去矩阵微分的范数形式的定义，而直接采用公式形式给出，并对不同情况的矩阵微分叙述其计算方法。

一、相对于数量变量的微分法

对于 n 维函数向量

$$\boldsymbol{a}(t)=(a_1(t),a_2(t),\cdots,a_n(t))^T$$

定义它对数量变量 t 的导数为

$$\frac{\mathrm{d}\boldsymbol{a}(t)}{\mathrm{d}t}\triangleq\left(\frac{\mathrm{d}a_1(t)}{\mathrm{d}t},\frac{\mathrm{d}a_2(t)}{\mathrm{d}t},\cdots,\frac{\mathrm{d}a_n(t)}{\mathrm{d}t}\right)^T$$

对于 $m\times n$ 阶函数矩阵

$$A(t)=(a_{ij}(t))_{m\times n}$$

定义它对数量变量 t 的导数为

$$\frac{\mathrm{d}A(t)}{\mathrm{d}t}\triangleq\left(\frac{\mathrm{d}a_{ij}(t)}{\mathrm{d}t}\right)_{m\times n}$$

显然，若将 n 维函数向量 $\boldsymbol{a}(t)$ 作为 $n\times 1$ 阶函数矩阵看待，它们对数量变量 t 的导数的定义是一致的。因此，下面仅就函数矩阵，讨论其对数量变量的导数的计算法则。

定理1 设 $A(t)$、$B(t)$、$A^{-1}(t)$ 及数量函数 $\lambda(t)$ 对数量变量 t 均可导，则

a) $\dfrac{\mathrm{d}(A\pm B)}{\mathrm{d}t}=\dfrac{\mathrm{d}A}{\mathrm{d}t}\pm\dfrac{\mathrm{d}B}{\mathrm{d}t}\qquad\forall\ A、B\in F^{m\times n}$

b) $\dfrac{\mathrm{d}\lambda A}{\mathrm{d}t}=\dfrac{\mathrm{d}\lambda}{\mathrm{d}t}A+\lambda\dfrac{\mathrm{d}A}{\mathrm{d}t}\qquad\forall\ A\in F^{m\times n}$

c) $\dfrac{\mathrm{d}AB}{\mathrm{d}t}=\dfrac{\mathrm{d}A}{\mathrm{d}t}B+A\dfrac{\mathrm{d}B}{\mathrm{d}t}\qquad\forall\ A\in F^{m\times n},B\in F^{n\times p}$

d) $\dfrac{\mathrm{d}A^{-1}}{\mathrm{d}t}=-A^{-1}\dfrac{\mathrm{d}A}{\mathrm{d}t}A^{-1}\qquad\forall\ A\in F_n^{n\times n}$

证 a)、b) 由导数定义即得；

c) 设 $\boldsymbol{a}(t)=(\alpha_1(t),\alpha_2(t),\cdots,\alpha_n(t))^T$，$\boldsymbol{b}(t)=(\beta_1(t),\beta_2(t),\cdots,\beta_n(t))^T$，则

$$\frac{\mathrm{d}\boldsymbol{a}^T\boldsymbol{b}}{\mathrm{d}t}=\sum_{i=1}^{n}\frac{\mathrm{d}\alpha_i\beta_i}{\mathrm{d}t}=\sum_{i=1}^{n}\left(\frac{\mathrm{d}\alpha_i}{\mathrm{d}t}\beta_i+\alpha_i\frac{\mathrm{d}\beta_i}{\mathrm{d}t}\right)$$
$$=\frac{\mathrm{d}\boldsymbol{a}^T}{\mathrm{d}t}\boldsymbol{b}+\boldsymbol{a}^T\frac{\mathrm{d}\boldsymbol{b}}{\mathrm{d}t}$$

设

$$A(t)=\begin{bmatrix}\boldsymbol{a}_1(t)^T\\ \boldsymbol{a}_2(t)^T\\ \vdots\\ \boldsymbol{a}_m(t)^T\end{bmatrix}$$

$$B(t)=(\boldsymbol{b}_1(t),\boldsymbol{b}_2(t),\cdots,\boldsymbol{b}_p(t))$$

其中 $\boldsymbol{a}_i(t)^T=(a_{i1}(t),a_{i2}(t),\cdots,a_{in}(t))\quad i\in\underline{m}$

$\boldsymbol{b}_j(t)=(b_{1j}(t),b_{2j}(t),\cdots,b_{nj}(t))^T\quad j\in\underline{p}$

则由

$$AB=(\boldsymbol{a}_i{}^T\boldsymbol{b}_j)_{m\times p}$$

知

$$\frac{\mathrm{d}AB}{\mathrm{d}t}=\left(\frac{\mathrm{d}\boldsymbol{a}_i{}^T\boldsymbol{b}_j}{\mathrm{d}t}\right)_{m\times p}$$
$$=\left(\frac{\mathrm{d}\boldsymbol{a}_i{}^T}{\mathrm{d}t}\boldsymbol{b}_i+\boldsymbol{a}_i{}^T\frac{\mathrm{d}\boldsymbol{b}_j}{\mathrm{d}t}\right)_{m\times p}$$
$$=\frac{\mathrm{d}A}{\mathrm{d}t}B+A\frac{\mathrm{d}B}{\mathrm{d}t}$$

d) 由于 $AA^{-1}=I_n$，故

$$\frac{\mathrm{d}AA^{-1}}{\mathrm{d}t}=\frac{\mathrm{d}A}{\mathrm{d}t}A^{-1}+A\frac{\mathrm{d}A^{-1}}{\mathrm{d}t}=O$$

于是

$$\frac{dA^{-1}}{dt}=-A^{-1}\frac{dA}{dt}A^{-1}$$

例1 设 $A\in F^{n\times n}(\|\cdot\|)$ 是常数矩阵，则

a) $\dfrac{de^{At}}{dt}=Ae^{At}$

b) $\dfrac{d}{dt}\sin At=A\cos At$

c) $\dfrac{d}{dt}\cos At=-A\sin At$

这只要注意到 e^{At}、$\sin At$、$\cos At$ 的级数表达式及函数矩阵的导数计算法则，即知上述结果是成立的。

例2 设 $A\in F^{n\times n}(\|\cdot\|)$ 是常数矩阵，$x(t)=(x_1(t),x_2(t),\cdots,x_n(t))^T$，则

a) e^{At} 是微分方程

$$\dot{x}=Ax\qquad\left(\dot{x}=\frac{dx}{dt}\right)$$

的基本解矩阵，故 $x=e^{At}$ 为此微分方程的解矩阵；

b) 若 A 可逆，则

$$\ddot{x}+A^2x=\theta$$

的通解为

$$x=(\sin At)c_1+(\cos At)c_2\quad\forall\, c_1,c_2\in F^n$$

证 a) 容易验证 $e^{At}c$ 是 $\dot{x}=Ax$ 的解向量，若 $x(t_0)=x_0$ 即为初值条件；显然，$e^{A(t-t_0)}x_0$ 为 $\dot{x}=Ax$ 满足 $x(t_0)=x_0$ 的特解，故 e^{At} 为 $\dot{x}=Ax$ 的基本解矩阵；

b) 容易验证 $(\sin At)c_1$、$(\cos At)c_2$ 为 $\ddot{x}+A^2x=\theta$ 的解向量，故 $x=(\sin At)c_1+(\cos At)c_2$ 为其通解；因为，只须取 $c_1=A^{-1}x_1$、$c_2=x_0$，即可得 $\ddot{x}+A^2x=\theta$ 满足初值条件

$$x(0)=x_0,\quad \dot{x}(0)=x_1$$

的特解。

例3 求二次型 x^TAx 对 t 的导数。

其中 $x=x(t)$ 是 n 维函数向量；

$A=A^T\in R^{n\times n}$ 是数字矩阵。

解
$$\begin{aligned}\frac{dx^TAx}{dt}&=\frac{dx^T}{dt}Ax+x^T\frac{dAx}{dt}\\&=\frac{dx^T}{dt}Ax+x^T\left(\frac{dA}{dt}x+A\frac{dx}{dt}\right)\\&=\frac{dx^T}{dt}Ax+x^TA\frac{dx}{dt}\\&=2\frac{dx^T}{dt}Ax\left(=2x^TA\frac{dx}{dt}\right)\end{aligned}$$

特别，当 $A=I_n$ 时，

$$\frac{dx^Tx}{dt}=2x^T\frac{dx}{dt}=2\frac{dx^T}{dt}x$$

二、相对于向量变量的微分法

1. 数量函数的导数

设

$$f(x)=g(x_1,x_2,\cdots,x_n)$$

是以向量 $x=(x_1,x_2,\cdots,x_n)^T$ 为自变量的数量函数，定义

$$\frac{df(x)}{dx}\triangleq\begin{bmatrix}\dfrac{\partial g}{\partial x_1}\\ \dfrac{\partial g}{\partial x_2}\\ \vdots\\ \dfrac{\partial g}{\partial x_n}\end{bmatrix}=\left(\frac{\partial}{\partial x_i}\right)_{n\times 1}\otimes g$$

显然，它是 R^3 中数量场 $u=u(x,y,z)$ 的梯度

$$\mathrm{grad}\,u=\nabla u=\left(\frac{\partial u}{\partial x},\frac{\partial u}{\partial y},\frac{\partial u}{\partial z}\right)^T$$

的概念的推广，故也可记为

$$\frac{df(x)}{d\boldsymbol{x}} = \text{grad } g(x_1,x_2,\cdots,x_n) = \nabla g(x_1,x_2,\cdots,x_n)$$

同理可定义

$$\frac{df(x)}{d\boldsymbol{x}^T} \triangleq \left[\frac{\partial g}{\partial x_1},\frac{\partial g}{\partial x_2},\cdots,\frac{\partial g}{\partial x_n}\right]$$

显然，若还有向量 $\boldsymbol{x}$ 的数量函数

$$h(\boldsymbol{x}) = k(x_1,x_2,\cdots,x_n)$$

则下列导数法则成立：

$$\frac{d[f(\boldsymbol{x}) \pm h(\boldsymbol{x})]}{d\boldsymbol{x}} = \frac{df(\boldsymbol{x})}{d\boldsymbol{x}} \pm \frac{dh(\boldsymbol{x})}{d\boldsymbol{x}}$$

$$\frac{df(\boldsymbol{x})h(\boldsymbol{x})}{d\boldsymbol{x}} = \frac{df(\boldsymbol{x})}{d\boldsymbol{x}}h(\boldsymbol{x}) + f(\boldsymbol{x})\frac{dh(\boldsymbol{x})}{d\boldsymbol{x}}$$

例4 求函数

$$f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{x} = x_1^2 + x_2^2 + \cdots + x_n^2$$

对 $\boldsymbol{x}$ 的导数。

解 $\frac{df(\boldsymbol{x})}{d\boldsymbol{x}} = (2x_1,2x_2,\cdots,2x_n)^T = 2\boldsymbol{x}$

2. 函数向量的导数

设 $\boldsymbol{x} = (x_1,x_2,\cdots,x_n)^T$，且

$$a(x) = (\alpha_1(x),\alpha_2(x),\cdots,\alpha_m(x))^T$$

是向量 $\boldsymbol{x}$ 的函数向量，定义

$$\frac{da(\boldsymbol{x})}{d\boldsymbol{x}^T} \triangleq \begin{bmatrix} \frac{\partial\alpha_1}{\partial x_1} & \frac{\partial\alpha_1}{\partial x_2}\cdots & \frac{\partial\alpha_1}{\partial x_n} \\ \frac{\partial\alpha_2}{\partial x_1} & \frac{\partial\alpha_2}{\partial x_2}\cdots & \frac{\partial\alpha_2}{\partial x_n} \\ \cdots & \cdots\ \cdots & \cdots \\ \frac{\partial\alpha_m}{\partial x_1} & \frac{\partial\alpha_m}{\partial x_2}\cdots & \frac{\partial\alpha_m}{\partial x_n} \end{bmatrix}_{m\times n}$$

$$\frac{da(\boldsymbol{x})^T}{d\boldsymbol{x}} \triangleq \left[\frac{da(\boldsymbol{x})}{d\boldsymbol{x}^T}\right]^T$$

定理2 设 $a(\boldsymbol{x}),b(\boldsymbol{x}) \in F^m, \boldsymbol{x} \in F^n, \lambda(\boldsymbol{x}) \in F$，则

a) $\frac{d(\boldsymbol{a}\pm\boldsymbol{b})}{d\boldsymbol{x}^T} = \frac{d\boldsymbol{a}}{d\boldsymbol{x}^T} \pm \frac{d\boldsymbol{b}}{d\boldsymbol{x}^T}$

b) $\frac{d(\lambda\boldsymbol{a})}{d\boldsymbol{x}^T} = \frac{d\lambda}{d\boldsymbol{x}^T}\boldsymbol{a} + \lambda\frac{d\boldsymbol{a}}{d\boldsymbol{x}^T}$

c) $\frac{d(\boldsymbol{a}^T\boldsymbol{b})}{d\boldsymbol{x}^T} = \boldsymbol{b}^T\frac{d\boldsymbol{a}}{d\boldsymbol{x}^T} + \boldsymbol{a}^T\frac{d\boldsymbol{b}}{d\boldsymbol{x}^T}$

证 a)、b)由定义即知；

c)
$$\begin{aligned}\frac{d(\boldsymbol{a}^T\boldsymbol{b})}{d\boldsymbol{x}^T} &= \left[\frac{\partial(\boldsymbol{a}^T\boldsymbol{b})}{\partial x_1},\frac{\partial(\boldsymbol{a}^T\boldsymbol{b})}{\partial x_2},\cdots,\frac{\partial(\boldsymbol{a}^T\boldsymbol{b})}{\partial x_n}\right] \\ &= \left[\frac{\partial\boldsymbol{a}^T}{\partial x_1}\boldsymbol{b}+\boldsymbol{a}^T\frac{\partial\boldsymbol{b}}{\partial x_1},\cdots,\frac{\partial\boldsymbol{a}^T}{\partial x_n}\boldsymbol{b}+\boldsymbol{a}^T\frac{\partial\boldsymbol{b}}{\partial x_n}\right] \\ &= \boldsymbol{b}^T\frac{d\boldsymbol{a}}{dx^T}+\boldsymbol{a}^T\frac{d\boldsymbol{b}}{dx^T}\end{aligned}$$

特别

$$\frac{d\boldsymbol{x}}{d\boldsymbol{x}^T} = \frac{d\boldsymbol{x}^T}{d\boldsymbol{x}} = I_m$$

将定理2中各公式转置，即可得另一组对向量 $\boldsymbol{x}$ 的导数公式。

例5 a) 求行向量 $\boldsymbol{x}^T A$ 对 $\boldsymbol{x}$ 的导数；

b) 求列向量 $B\boldsymbol{x}$ 对 $\boldsymbol{x}^T$ 的导数。

其中 $A \in F^{n\times m}$、$B \in F^{m\times n}$ 均为数字矩阵，$\boldsymbol{x} \in F^n$。

解 a) 设 $A = (\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m), \boldsymbol{a}_i \in F^n, i \in \underline{m}$，

则

$$\boldsymbol{x}^T A = (\boldsymbol{x}^T\boldsymbol{a}_1,\boldsymbol{x}^T\boldsymbol{a}_2,\cdots,\boldsymbol{x}^T\boldsymbol{a}_m)$$

于是

$$\frac{d(\boldsymbol{x}^T A)}{d\boldsymbol{x}} = \left[\frac{d\boldsymbol{x}^T\boldsymbol{a}_1}{d\boldsymbol{x}},\frac{d\boldsymbol{x}^T\boldsymbol{a}_2}{d\boldsymbol{x}},\cdots,\frac{d\boldsymbol{x}^T\boldsymbol{a}_m}{d\boldsymbol{x}}\right]$$

$$=\left[\frac{dx^T}{dx}a_1+\frac{da_1^T}{dx}x,\cdots,\frac{dx^T}{dx}a_m+\frac{da_m^T}{dx}x\right]$$

$$=(a_1,a_2,\cdots,a_m)=A$$

b) 设

$$B=\begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_m^T \end{bmatrix}$$

则

$$Bx=\begin{bmatrix} b_1^Tx \\ b_2^Tx \\ \vdots \\ b_m^Tx \end{bmatrix}$$

于是

$$\frac{d(Bx)}{dx^T}=\begin{bmatrix} \dfrac{db_1^Tx}{dx^T} \\ \dfrac{db_2^Tx}{dx^T} \\ \vdots \\ \dfrac{db_m^Tx}{dx^T} \end{bmatrix}$$

$$=\begin{bmatrix} x^T\dfrac{db_1}{dx^T}+b_1^T\dfrac{dx}{dx^T} \\ \vdots \\ \vdots \\ x^T\dfrac{db_m}{dx^T}+b_m^T\dfrac{dx}{dx^T} \end{bmatrix}=\begin{bmatrix} b_1^T \\ \vdots \\ \vdots \\ \vdots \\ b_m^T \end{bmatrix}=B$$

例6 求二次型 x^TAx 对 x 的导数。

解 $\dfrac{d(x^TAx)}{dx}=\dfrac{dx^T}{dx}Ax+\dfrac{d(Ax)^T}{dx}x$

$$=Ax+\frac{dx^TA^T}{dx}x=(A+A^T)x$$

若 $A=A^T$,则

$$\frac{dx^TAx}{dx}=2Ax$$

例7 求数量函数 p^TAx 对 x 的导数。

其中 $p\in F^n, A\in F^{n\times n}$,它们的元均为常数。

解 $\dfrac{d(p^TAx)}{dx}=\dfrac{dp^T}{dx}Ax+\dfrac{d(Ax)^T}{dx}p=A^Tp$

3. 函数矩阵的导数

设矩阵 $A=(a_{ij})\in F^{m\times l}$的元 a_{ij}是向量 $x\in F^m$ 的函数,即

$$A(x)=(a_{ij}(x))_{m\times l}$$

定义

$$\frac{d(Ax)}{dx}\triangleq\begin{bmatrix} \dfrac{\partial A}{\partial x_1} \\ \dfrac{\partial A}{\partial x_2} \\ \vdots \\ \dfrac{\partial A}{\partial x_n} \end{bmatrix}_{mn\times l}$$

$$\frac{d(Ax)}{dx^T}\triangleq\left[\frac{\partial A}{\partial x_1},\frac{\partial A}{\partial x_2},\cdots,\frac{\partial A}{\partial x_n}\right]_{m\times ln}$$

函数矩阵对向量的微分法则如下。

定理3

a) $\dfrac{d[A(x)\pm B(x)]}{dx}=\dfrac{dA(x)}{dx}\pm\dfrac{dB(x)}{dx}$

b) $\dfrac{d[\lambda(x)A(x)]}{dx}=\dfrac{d\lambda(x)}{dx}\otimes A+\lambda(x)\dfrac{dA(x)}{dx}$

其中 $\lambda(\boldsymbol{x})$是向量 $\boldsymbol{x}$ 的数量函数。

c) $\dfrac{d[A(\boldsymbol{x})B(\boldsymbol{x})]}{d\boldsymbol{x}}=\dfrac{dA(\boldsymbol{x})}{d\boldsymbol{x}}B(\boldsymbol{x})+[I_n\otimes A]\dfrac{dB(\boldsymbol{x})}{d\boldsymbol{x}}$

证 a) 由定义即知；

b) $$\frac{d[\lambda(\boldsymbol{x})A(\boldsymbol{x})]}{d\boldsymbol{x}}=\begin{bmatrix}\dfrac{\partial\lambda A}{\partial x_1}\\ \dfrac{\partial\lambda A}{\partial x_2}\\ \vdots\\ \dfrac{\partial\lambda A}{\partial x_n}\end{bmatrix}=\begin{bmatrix}\dfrac{\partial\lambda}{\partial x_1}A+\lambda\dfrac{\partial A}{\partial x_1}\\ \dfrac{\partial\lambda}{\partial x_2}A+\lambda\dfrac{\partial A}{\partial x_2}\\ \vdots\\ \dfrac{\partial\lambda}{\partial x_n}A+\lambda\dfrac{\partial A}{\partial x_n}\end{bmatrix}$$

$$=\begin{bmatrix}\dfrac{\partial\lambda}{\partial x_1}A\\ \dfrac{\partial\lambda}{\partial x_2}A\\ \vdots\\ \dfrac{\partial\lambda}{\partial x_n}A\end{bmatrix}+\begin{bmatrix}\lambda\dfrac{\partial A}{\partial x_1}\\ \lambda\dfrac{\partial A}{\partial x_2}\\ \vdots\\ \lambda\dfrac{\partial A}{\partial x_n}\end{bmatrix}=\frac{d\lambda(\boldsymbol{x})}{d\boldsymbol{x}}\otimes A+\lambda(\boldsymbol{x})\frac{dA(\boldsymbol{x})}{d\boldsymbol{x}}$$

c) $$\frac{d[A(\boldsymbol{x})B(\boldsymbol{x})]}{d\boldsymbol{x}}=\begin{bmatrix}\dfrac{\partial AB}{\partial x_1}\\ \vdots\\ \dfrac{\partial AB}{\partial x_n}\end{bmatrix}$$

$$=\begin{bmatrix}\dfrac{\partial A}{\partial x_1}B+A\dfrac{\partial B}{\partial x_1}\\ \vdots\\ \dfrac{\partial A}{\partial x_n}B+A\dfrac{\partial B}{\partial x_n}\end{bmatrix}$$

$$=\frac{dA}{d\boldsymbol{x}}B+[I_n\otimes A]\frac{dB}{d\boldsymbol{x}}$$

例8 求 $\boldsymbol{x}^TA$ 对向量 $\boldsymbol{x}$ 的导数。

其中 A 为常数矩阵。

解 $\dfrac{d\boldsymbol{x}^TA}{d\boldsymbol{x}}=\dfrac{d\boldsymbol{x}^T}{d\boldsymbol{x}}A+(I_n\otimes\boldsymbol{x}^T)\dfrac{dA}{d\boldsymbol{x}}=A$

例9 求 $\boldsymbol{p}^TA\boldsymbol{q}$ 对向量 $\boldsymbol{x}$ 的导数。

其中 $\boldsymbol{p}$、$\boldsymbol{q}$ 均为常向量；

A 为向量 $\boldsymbol{x}$ 的函数矩阵。

解 $$\frac{d[\boldsymbol{p}^TA\boldsymbol{q}]}{d\boldsymbol{x}}=\frac{d\boldsymbol{p}^T}{d\boldsymbol{x}}A\boldsymbol{q}+\frac{d[A\boldsymbol{q}]^T}{d\boldsymbol{x}}\boldsymbol{p}$$

$$=\frac{d\boldsymbol{q}^T}{d\boldsymbol{x}}A^T\boldsymbol{p}+[I_n\otimes\boldsymbol{q}^T]\frac{dA^T}{d\boldsymbol{x}}\boldsymbol{p}$$

$$=[I_n\otimes\boldsymbol{q}^T]\frac{dA^T}{d\boldsymbol{x}}\boldsymbol{p}$$

三、相对于矩阵的微分法

1. 数量矩阵函数的导数

设 $A\in F^{m\times l}$，$f(A)$为矩阵 A 的数量函数，定义

$$\frac{df(A)}{dA}\triangleq\left[\frac{\partial f}{\partial a_{ij}}\right]_{m\times l}\qquad A=(a_{ij})_{m\times l}$$

显然，由此有

$$\frac{d[f(A)\pm g(A)]}{dA}=\frac{df(A)}{dA}\pm\frac{dg(A)}{dA}$$

$$\frac{df(A)g(A)}{dA}=\frac{df(A)}{dA}g(A)+f(A)\frac{dg(A)}{dA}$$

例10 求二次型 $\boldsymbol{x}^TA\boldsymbol{x}$ 对 A 的导数。

其中 $A=(a_{ij})\in F^{n\times n}$，$\boldsymbol{x}\in F^n$ 与 A 无关。

解 $$\frac{d\boldsymbol{x}^TA\boldsymbol{x}}{dA}=\left[\frac{\partial\boldsymbol{x}^TA\boldsymbol{x}}{\partial a_{ij}}\right]_{n\times n}=(x_ix_j)_{n\times n}=\boldsymbol{x}\boldsymbol{x}^T$$

这里用到 $\boldsymbol{x}^TA\boldsymbol{x}=\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}a_{ij}x_ix_j$。

2. 向量矩阵函数的导数

设 $\boldsymbol{z}(A)=(z_1(A),z_2(A),\cdots,z_n(A))^T$，$A=(a_{ij})_{m\times l}$，定义

$$\frac{\mathrm{d}z(A)}{\mathrm{d}A} \triangleq \left[\frac{\partial z(A)}{\partial a_{ij}}\right]_{mn\times l}$$

其中、$\frac{\partial z(A)}{\partial a_{ij}}=\begin{bmatrix}\frac{\partial z_1(A)}{\partial a_{ij}}\\ \vdots \\ \frac{\partial z_n(A)}{\partial a_{ij}}\end{bmatrix}$

仿此，可定义 $z(A)^T=(z_1(A),z_2(A),\cdots,z_n(A))$ 对矩阵 A 的导数

$$\frac{\mathrm{d}z(A)^T}{\mathrm{d}A} \triangleq \left[\frac{\partial z(A)^T}{\partial a_{ij}}\right]_{m\times nl}$$

其中 $\frac{\partial z(A)^T}{\partial a_{ij}}=\left[\frac{\partial z_1(A)}{\partial a_{ij}},\cdots,\frac{\partial z_n(A)}{\partial a_{ij}}\right]$

3. 矩阵函数的导数

设矩阵 G 是矩阵 A 的函数，即

$$G(A)=\begin{bmatrix}g_{11}(A) & g_{12}(A)\cdots g_{1q}(A)\\ g_{21}(A) & g_{22}(A)\cdots g_{2q}(A)\\ \cdots\ \cdots & \cdots\ \cdots\ \cdots \\ g_{p1}(A) & g_{p2}(A)\cdots g_{pq}(A)\end{bmatrix}$$

是以 $A\in F^{m\times n}$ 为自变量的 $p\times q$ 阶矩阵，定义

$$\frac{\mathrm{d}G(A)}{\mathrm{d}A} \triangleq \left[\frac{\partial G}{\partial a_{ij}}\right]_{pm\times qn}=\nabla\otimes G$$

其中 $$\frac{\partial G}{\partial a_{ij}}=\begin{bmatrix}\frac{\partial g_{11}}{\partial a_{ij}}\cdots\frac{\partial g_{1q}}{\partial a_{ij}}\\ \cdots\ \ \cdots\ \ \cdots \\ \frac{\partial g_{p1}}{\partial a_{ij}}\cdots\frac{\partial g_{pq}}{\partial a_{ij}}\end{bmatrix}$$

与 $$\nabla=\begin{bmatrix}\frac{\partial}{\partial a_{11}}\cdots\frac{\partial}{\partial a_{1n}}\\ \cdots\ \ \cdots\ \ \cdots \\ \frac{\partial}{\partial a_{m1}}\cdots\frac{\partial}{\partial a_{mn}}\end{bmatrix}$$

为哈米尔顿(Hamilton)算子矩阵。

定理4 a) $\frac{\mathrm{d}[G(A)\pm D(A)]}{\mathrm{d}A}=\frac{\mathrm{d}G(A)}{\mathrm{d}A}\pm\frac{\mathrm{d}D(A)}{\mathrm{d}A}$

b) 设 $\lambda(A)\in F$，则

$$\frac{\mathrm{d}[\lambda(A)G(A)]}{\mathrm{d}A}=\frac{\mathrm{d}\lambda(A)}{\mathrm{d}A}\otimes G(A)+\lambda(A)\frac{\mathrm{d}G(A)}{\mathrm{d}A}$$

c) $$\frac{\mathrm{d}[G(A)H(A)]}{\mathrm{d}A}=\frac{\mathrm{d}G(A)}{\mathrm{d}A}[I_n\otimes H(A)]+[I_m\otimes G(A)]\frac{\mathrm{d}H(A)}{\mathrm{d}A}$$

其中 $A\in F^{m\times n}$。

证明留给读者练习。

例11 求 y^TA^TAy 对 A 的导数。

其中 $A\in F^{m\times n}$、$y\in F^m$ 是常向量。

解 $$\frac{\mathrm{d}[y^TA^TAy]}{\mathrm{d}A}=\left[\frac{\partial[y^TA^TAy]}{\partial a_{ij}}\right]$$
$$=\left[\frac{\partial[y^TA^T]}{\partial a_{ij}}Ay+y^TA^T\frac{\partial[Ay]}{\partial a_{ij}}\right]$$
$$=\left[y^T\frac{\partial A^T}{\partial a_{ij}}Ay+y^TA^T\frac{\partial A}{\partial a_{ij}}y\right]$$
$$=\left[\sum_{k=1}^n a_{ik}y_ky_j\right]+\left[\sum_{k=1}^n a_{ki}y_ky_j\right]$$
$$=Ayy^T+Ayy^T=2Ayy^T$$

由此可得

$$\frac{\mathrm{d}(Ay-x)^T(Ay-x)}{\mathrm{d}A}=2(Ay-x)y^T$$

事实上

$$(Ay-x)^T(Ay-x)=y^TA^TAy-x^TAy-y^TA^Tx+x^Tx$$

由例10可知

$$\frac{\mathrm{d}x^TAy}{\mathrm{d}A}=xy^T$$

又因 $y^TA^Tx=(x^TAy)^T$，故 $\frac{\mathrm{d}x^TAy}{\mathrm{d}A}=xy^T$，因此有上述结果。

例12 求 $\mathrm{tr}(Ay-x)(Ay-x)^T$、$\mathrm{tr}AB$、$\mathrm{tr}ACA^T$ 对 A 的导数。

其中 B、C 是常矩阵；

x、y 是常向量。

解 由于 $\mathrm{tr}(Ay-x)(Ay-x)^T=\mathrm{tr}(Ay-x)^T(Ay-x)$，故由例11知

$$\frac{\mathrm{dtr}(Ay-x)(Ay-x)^T}{\mathrm{d}A}=2(Ay-x)y^T$$

又

$$\mathrm{tr}AB=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}b_{ji}$$

故

$$\frac{\mathrm{dtr}AB}{\mathrm{d}A}=\left(\frac{\partial\,\mathrm{tr}AB}{\partial a_{kl}}\right)=(b_{lk})=B^T$$

其中 $A\in F^{n\times m}$、$B\in F^{m\times n}$。

由矩阵迹的性质有

$$\mathrm{tr}AB=\mathrm{tr}BA=\mathrm{tr}A^TB^T=\mathrm{tr}B^TA^T$$

故

$$\frac{\mathrm{dtr}AB}{\mathrm{d}A}=\frac{\mathrm{dtr}BA}{\mathrm{d}A}=\frac{\mathrm{dtr}A^TB^T}{\mathrm{d}A}=\frac{\mathrm{dtr}B^TA^T}{\mathrm{d}A}=B^T$$

又

$$\frac{\mathrm{dtr}ACA^T}{\mathrm{d}A}=\frac{\mathrm{dtr}AB_1}{\mathrm{d}A}+\frac{\mathrm{dtr}B_2A^T}{\mathrm{d}A}=B_1{}^T+B_2=A(C+C^T)$$

其中 $B_1=CA^T$、$B_2=AC$ 在求导时均看作常数矩阵。

例13 求 $f=(x-a-Bz)^T(x-a-Bz)$ 对矩阵 B 的导数。其中 x、a、z 均为常向量。

解 由例11知

$$\frac{\mathrm{d}(x-a-Bz)^T(x-a-Bz)}{\mathrm{d}B}$$

$$=\frac{\mathrm{d}(Bz-x+a)^T(Bz-x+a)}{\mathrm{d}B}=2(Bz-x+a)z^T$$

四、复合函数微分法

1. 数量函数的求导公式

设 $f=f(y)$，$y=y(t)$，若 f、t 为数量变量，y 为向量变量，则

$$\frac{\mathrm{d}f}{\mathrm{d}t}=\frac{\mathrm{d}f}{\mathrm{d}y^T}\frac{\mathrm{d}y}{\mathrm{d}t}=\frac{\mathrm{d}y^T}{\mathrm{d}t}\frac{\mathrm{d}f}{\mathrm{d}y}\qquad(1)$$

设 $f=f(y)$，$y=y(x)$，若 f 为数量变量，x、y 为向量变量，则

$$\frac{\mathrm{d}f}{\mathrm{d}x}=\frac{\mathrm{d}y^T}{\mathrm{d}x}\frac{\mathrm{d}f}{\mathrm{d}y}\qquad(2)$$

$$\frac{\mathrm{d}f}{\mathrm{d}x^T}=\frac{\mathrm{d}f}{\mathrm{d}y^T}\frac{\mathrm{d}y}{\mathrm{d}x^T}\qquad(3)$$

这只要注意到

$$\mathrm{d}f=\frac{\mathrm{d}f}{\mathrm{d}y^T}\mathrm{d}y,\ \mathrm{d}y=\frac{\mathrm{d}y}{\mathrm{d}x^T}\mathrm{d}x$$

即可。

又设 $f=f(x,y)$，$y=y(x)$，若 f 是常量变量，x、y 是向量

变量，则

$$\frac{\mathrm{d}f}{\mathrm{d}x}=\frac{\partial f}{\partial x}+\frac{\mathrm{d}y^T}{\mathrm{d}x}\frac{\partial f}{\partial y} \tag{4}$$

$$\frac{\mathrm{d}f}{\mathrm{d}x^T}=\frac{\partial f}{\partial x^T}+\frac{\partial f}{\partial y^T}\frac{\mathrm{d}y}{\mathrm{d}x^T} \tag{5}$$

以上公式的证明略去。下面举例说明其应用。

例14 求 $f=x^TAx$ 对 x 的导数。

解 设 $y=Ax$，于是由公式(4)

$$\frac{\mathrm{d}f}{\mathrm{d}x}=y+A^Tx=(A+A^T)x$$

例15 求 $f=(x-a-Bz)^T(x-a-Bz)$ 对 a 的导数。

解 令 $y=x-a-Bz$，则 $f=y^Ty$，于是由公式(2)有

$$\frac{\mathrm{d}f}{\mathrm{d}a}=\frac{\mathrm{d}y^T}{\mathrm{d}a}\frac{\mathrm{d}f}{\mathrm{d}y}=-I(2y)=-2(x-a-Bz)$$

2. 向量函数的求导公式

设 $z=z(y)$，$y=y(t)$，若 t 为数量变量，y、z 为向量变量，则

$$\frac{\mathrm{d}z}{\mathrm{d}t}=\frac{\mathrm{d}z}{\mathrm{d}y^T}\frac{\mathrm{d}y}{\mathrm{d}t} \tag{6}$$

设 $z=(zy)$，$y=y(x)$，若 x、y、z 均为向量变量，则

$$\frac{\mathrm{d}z^T}{\mathrm{d}x}=\frac{\mathrm{d}y^T}{\mathrm{d}x}\frac{\mathrm{d}z^T}{\mathrm{d}y} \tag{7}$$

$$\frac{\mathrm{d}z}{\mathrm{d}x^T}=\frac{\mathrm{d}z}{\mathrm{d}y^T}\frac{\mathrm{d}y}{\mathrm{d}x^T} \tag{8}$$

设 $z=f(x,y)$，$y=y(x)$，若 x、y、z 均为向量变量，则

$$\frac{\mathrm{d}z^T}{\mathrm{d}x}=\frac{\partial f}{\partial x}+\frac{\mathrm{d}y^T}{\mathrm{d}x}\frac{\partial f}{\partial y} \tag{9}$$

$$\frac{\mathrm{d}z}{\mathrm{d}x^T}=\frac{\partial f}{\partial x^T}+\frac{\partial f}{\partial y^T}\frac{\mathrm{d}y}{\mathrm{d}x^T} \tag{10}$$

例16 求向量函数 $[\sin(c^Tx)]x^T$ 对 x 的导数。其中 c 为常向量。

解 令 $c^Tx=y$，则 $z^T=(\sin y)x^T$，由公式(9)有

$$\begin{aligned}\frac{\mathrm{d}z^T}{\mathrm{d}x}&=\frac{\partial f}{\partial x}+\frac{\mathrm{d}y^T}{\mathrm{d}x}\frac{\partial f}{\partial y}=\sin y\frac{\mathrm{d}x^T}{\mathrm{d}x}+\frac{\mathrm{d}x^T}{\mathrm{d}x}c(\cos y)x^T\\&=(\sin y)I+c\cdot(\cos y)x^T\\&=[\sin(c^Tx)]I+[\cos(c^Tx)]cx^T\end{aligned}$$

例17 求 $f=(Ax-b)^TR(Ax-b)$ 对 x 的导数。其中 A、R 为常数矩阵，b 是常数向量。

解 令 $y=Ax-b$，则 $f=y^TRy$，由公式(4)有

$$\begin{aligned}\frac{\mathrm{d}f}{\mathrm{d}x}&=\frac{\mathrm{d}y^T}{\mathrm{d}x}\frac{\mathrm{d}f}{\mathrm{d}y}=A^T(Ry+R^Ty)\\&=A^T(R+R^T)(Ax-b)\end{aligned}$$

§ 5.5 矩阵的积分

设 $A(t)=(a_{ij}(t))_{m\times n}$，若 $a_{ij}(t)$ 在 $[t_1,t_2]$ 上可积，定义

$$\int_{t_1}^{t_2}A(t)\mathrm{d}t\triangleq\left(\int_{t_1}^{t_2}a_{ij}(t)\mathrm{d}t\right)_{m\times n}$$

$$\mathrm{d}A(t)\triangleq(\mathrm{d}a_{ij}(t))_{m\times n}$$

于是

$$\int_{t_1}^{t_2}[A(t)\pm B(t)]\mathrm{d}t=\int_{t_1}^{t_2}A(t)\mathrm{d}t\pm\int_{t_1}^{t_2}B(t)\mathrm{d}t$$

$$\int_{t_1}^{t_2}kA(t)\mathrm{d}t=k\int_{t_1}^{t_2}A(t)\mathrm{d}t\qquad k=\text{const}$$