矩 阵 微 分 法

在现代控制理论中，经常会遇到矩阵的微分（导数），如对表达式 $\dfrac{d\boldsymbol{A}}{d\boldsymbol{B}}$ 来说，由于 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 都可能是数量、向量或矩阵，可代表九种不同的导数。除数量函数对数量变量的导数外，还剩下八种。下面分别介绍八种导数的定义和运算公式。

一、　相对于数量变量的微分（自变量是数量变量，如时间 t ）

定义 1　对于 n 维向量函数

$$\boldsymbol{a}(t)=\begin{bmatrix}a_1(t) & a_2(t) & \ldots\ldots & a_n(t)\end{bmatrix}^T$$

定义它对 t 的导数为

$$\frac{d\,\boldsymbol{a}(t)}{dt}\triangleq\begin{bmatrix}\dfrac{d\,a_1(t)}{dt} & \dfrac{d\,a_2(t)}{dt} & \cdots & \dfrac{d\,a_n(t)}{dt}\end{bmatrix}^T \qquad \cdots\cdots\cdots（1-1）$$

定义 2　对于 n × m 维矩阵函数

$$\boldsymbol{A}(t)=\begin{bmatrix}a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t)\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1}(t) & a_{n2}(t) & \cdots & a_{nn}(t)\end{bmatrix}=\begin{bmatrix}a_{ij}(t)\end{bmatrix}_{nm}$$

定义它对 t 的导数为

$$\frac{d\boldsymbol{A}(t)}{dt}\triangleq\begin{bmatrix}\dfrac{da_{11}(t)}{dt} & \dfrac{da_{12}(t)}{dt} & \cdots & \dfrac{da_{1n}(t)}{dt}\\ \vdots & \vdots & \ddots & \vdots\\ \dfrac{da_{n1}(t)}{dt} & \dfrac{da_{n2}(t)}{dt} & \cdots & \dfrac{da_{nn}(t)}{dt}\end{bmatrix}^T=\begin{bmatrix}\dfrac{da_{ij}(t)}{dt}\end{bmatrix}_{nm} \qquad \cdots\cdots\cdots（1-2）$$

我们不难看出，上述两个定义是一致的。当矩阵 $\boldsymbol{A(t)}$ 退化为向量 $\boldsymbol{a(t)}$时，定义 2 就变为定义 1。再退一步讲，当向量 $\boldsymbol{a(t)}$ 退化为数量函数 $a(t)$时，定义 1 就变为一般的导数定义。这说明这样定义是合理的，是统一的。

根据上述的两个定义，我们还可以推出下列的运算公式

$$\frac{d}{dt}\left\{\boldsymbol{A}(t)\pm\boldsymbol{B}(t)\right\}=\frac{d\boldsymbol{A}(t)}{dt}\pm\frac{d\boldsymbol{B}(t)}{dt} \qquad \cdots\cdots\cdots（1-3）$$

$$\frac{d}{dt}\left\{\boldsymbol{\lambda}(t)\cdot\boldsymbol{A}(t)\right\}=\frac{d\boldsymbol{\lambda}(t)}{dt}\cdot\boldsymbol{A}(t)+\boldsymbol{\lambda}(t)\cdot\frac{d\boldsymbol{A}(t)}{dt} \qquad \cdots\cdots\cdots（1-4）$$

$\boldsymbol{\lambda}(t)$——为变量 t 的数量函数

$$\frac{d}{dt}\left\{\boldsymbol{A}(t)\cdot\boldsymbol{B}(t)\right\}=\frac{d\boldsymbol{A}(t)}{dt}\cdot\boldsymbol{B}(t)+\boldsymbol{A}(t)\cdot\frac{d\boldsymbol{B}(t)}{dt} \qquad \cdots\cdots\cdots（1-5）$$

这些公式都很容易证明，现证明最后一式（1－5），设矩阵 ***A(t)*** 和 ***B(t)*** 分别为 n×m 和 m×l 矩阵

证：

$$\boldsymbol{A}(t)=\begin{bmatrix} a_{11}(t) & a_{12}(t) & \cdots & a_{1n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}(t) & a_{n2}(t) & \cdots & a_{nm}(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1^T(t) \\ \vdots \\ \boldsymbol{a}_n^T(t) \end{bmatrix}$$

$$\boldsymbol{B}(t)=\begin{bmatrix} b_{11}(t) & b_{12}(t) & \cdots & b_{1\ell}(t) \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1}(t) & b_{m2}(t) & \cdots & b_{m\ell}(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_1(t) & \boldsymbol{b}_2(t) & \cdots & \boldsymbol{b}_\ell(t) \end{bmatrix}$$

$$\boldsymbol{A}(t)\cdot\boldsymbol{B}(t)=\begin{bmatrix} \boldsymbol{a}_1^T(t)\boldsymbol{b}_1(t) & \cdots & \boldsymbol{a}_1^T(t)\boldsymbol{b}_\ell(t) \\ \vdots & \ddots & \vdots \\ \boldsymbol{a}_n^T(t)\boldsymbol{b}_1(t) & \cdots & \boldsymbol{a}_n^T(t)\boldsymbol{b}_\ell(t) \end{bmatrix} = \left[\boldsymbol{a}_i^T(t)\cdot\boldsymbol{b}_j(t)\right]_{n\ell}$$

从而根据矩阵导数定义 2，有

$$\begin{aligned}\frac{d}{dt}\left[\boldsymbol{A}(t)\cdot\boldsymbol{B}(t)\right] &= \frac{d}{dt}\left[\boldsymbol{a}_i^T(t)\cdot\boldsymbol{b}_j(t)\right]_{n\ell} \\ &= \left[\frac{d\boldsymbol{a}_i^T(t)}{dt}\cdot\boldsymbol{b}_j(t)+\boldsymbol{a}_i^T(t)\cdot\frac{d\boldsymbol{b}_j(t)}{dt}\right]_{n\ell} = \frac{d\boldsymbol{A}(t)}{dt}\cdot\boldsymbol{B}(t)+\boldsymbol{A}(t)\cdot\frac{d\boldsymbol{B}(t)}{dt}\end{aligned}$$

证毕

例 1：求 $\boldsymbol{X}^T\boldsymbol{A}\,\boldsymbol{X}$ 对 t 的导数，其中

$$\boldsymbol{X}=\begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \qquad \boldsymbol{A}=\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}$$ —— 对称常系数矩阵

解

$$\begin{aligned}\frac{d}{dt}[X^T\cdot A\cdot X] &= \frac{dX^T}{dt}A\cdot X+X^T\cdot\frac{d(A\cdot X)}{dt} \\ &= \frac{dX^T}{dt}A\cdot X+X^T(\frac{dA}{dt}\cdot X+A\cdot\frac{dX}{dt}) \\ &= \dot{X}^T\cdot A\cdot X+X^T\cdot A\cdot\dot{X}=(\dot{X}^T\cdot A\cdot X)^T+X^T\cdot A\cdot\dot{X} \\ &= X^TA\dot{X}+X^TA\dot{X}=2X^TA\dot{X}\end{aligned}$$

即 $$\frac{d}{dt}(\boldsymbol{X}^T\boldsymbol{A}\,\boldsymbol{X})=2\boldsymbol{X}^T\boldsymbol{A}\,\dot{\boldsymbol{X}} \qquad \cdots\cdots\cdots（1-6）$$

注：$\dot{\boldsymbol{X}}^T\boldsymbol{A}\,\boldsymbol{X}$ 和 $\boldsymbol{X}^T\boldsymbol{A}\,\dot{\boldsymbol{X}}$ 都是数量函数且 $\boldsymbol{A}$ 为对称阵，它们等于自己的转置。

习题

1．若 $A=\begin{bmatrix}1 & 2\\ 2 & 1\end{bmatrix}$ $X=\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}$ 证明上式。

2．若 $A=\begin{bmatrix}1 & 1\\ 2 & 1\end{bmatrix}$ $X=\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}$ 证明上式。

3．若 $A=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$ $X=\begin{bmatrix}x_1(t)\\ x_2(t)\end{bmatrix}$ 求 $\frac{d}{dt}[X^T A X]$

二、 相对于向量的微分（自变量是向量 $\boldsymbol{X}$ ）

1、数量函数的导数

设函数 $f(\boldsymbol{X})=f(x_1, x_2, \cdots, x_n)$ 是以向量 $\boldsymbol{X}$ 为自变量的数量函数，即以 n 个变量 x_i 为自变量的数量函数。

定义 3

我们将列向量 $\begin{bmatrix}\frac{\partial f}{\partial x_1}\\ \vdots \\ \frac{\partial f}{\partial x_n}\end{bmatrix}$ 叫做数量函数 f 对列向量 $\boldsymbol{X}$ 的导数，

记作

$$\frac{df}{d\boldsymbol{X}} = \begin{bmatrix}\frac{\partial f}{\partial x_1}\\ \vdots \\ \frac{\partial f}{\partial x_n}\end{bmatrix} \triangleq \boldsymbol{grad}\, f \triangleq \nabla f \qquad \frac{df}{d\boldsymbol{X}^T} = \begin{bmatrix}\frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n}\end{bmatrix}$$

例 2．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：根据定义

$$\frac{df}{d\boldsymbol{X}}=\begin{bmatrix}\frac{\partial f}{\partial x_1}\\ \vdots \\ \frac{\partial f}{\partial x_n}\end{bmatrix}=\begin{bmatrix}2x_1\\ \vdots \\ 2x_n\end{bmatrix}=2\begin{bmatrix}x_1\\ \vdots \\ x_n\end{bmatrix}=2\boldsymbol{X} \qquad 即 \quad \frac{d(\boldsymbol{X}^T\boldsymbol{X})}{d\boldsymbol{X}}=2\boldsymbol{X} \qquad \cdots\cdots\cdots（1-7）$$

2、向量函数的导数

设函数 $\boldsymbol{a}(\boldsymbol{X})=\begin{bmatrix}a_1(\boldsymbol{X})\\ \vdots \\ a_m(\boldsymbol{X})\end{bmatrix}$ ， $\boldsymbol{X}=(x_1,x_2,\cdots,x_n)^T$

定义 4 n×m 阶矩阵函数

$$\begin{bmatrix}\frac{\partial a_1(\boldsymbol{X})}{\partial x_1} & \cdots & \frac{\partial a_m(\boldsymbol{X})}{\partial x_1}\\ \vdots & \ddots & \vdots \\ \frac{\partial a_1(\boldsymbol{X})}{\partial x_n} & \cdots & \frac{\partial a_m(\boldsymbol{X})}{\partial x_n}\end{bmatrix}=\left[\frac{\partial a_j}{\partial x_i}\right]_{nm}=\frac{d\boldsymbol{a}^T(\boldsymbol{X})}{d\boldsymbol{X}} \qquad \cdots\cdots\cdots（1-8）$$

称之为 m 维向量函数 $\boldsymbol{a}^T(\boldsymbol{X})$ 对 n 维列向量 $\boldsymbol{X}$ 的导数。

m×n 阶矩阵函数

$$\begin{bmatrix}\frac{\partial a_1(\boldsymbol{X})}{\partial x_1} & \cdots & \frac{\partial a_m(\boldsymbol{X})}{\partial x_n}\\ \vdots & \ddots & \vdots \\ \frac{\partial a_1(\boldsymbol{X})}{\partial x_1} & \cdots & \frac{\partial a_m(\boldsymbol{X})}{\partial x_n}\end{bmatrix}=\left[\frac{\partial a_i}{\partial x_j}\right]_{m\,n}=\frac{d\boldsymbol{a}(\boldsymbol{X})}{d\boldsymbol{X}^T} \qquad \cdots\cdots\cdots（1-9）$$

称之为 m 维向量函数 $\boldsymbol{a}(\boldsymbol{X})$ 对 n 维横向量 $\boldsymbol{X}^{\mathrm{T}}$ 的导数。

从定义可看出 $\frac{d\boldsymbol{a}^T(\boldsymbol{X})}{d\boldsymbol{X}}\neq\frac{d\boldsymbol{a}(\boldsymbol{X})}{d\boldsymbol{X}^T}$ $\qquad$ $\frac{d\boldsymbol{a}^T(\boldsymbol{X})}{d\boldsymbol{X}}=\left[\frac{d\boldsymbol{a}(\boldsymbol{X})}{d\boldsymbol{X}^T}\right]^T$ ……（1−10）

若 $\boldsymbol{a}(\boldsymbol{X})$ 和 $\boldsymbol{b}(\boldsymbol{X})$ 是 m 维列向量函数，$\lambda(\boldsymbol{X})$ 是数量函数，$\boldsymbol{X}$ 是 m 维列向量，有以下 3 个运算公式

$$\frac{d[\boldsymbol{a}^T(\boldsymbol{X})\pm\boldsymbol{b}^T(\boldsymbol{X})]}{d\boldsymbol{X}}=\frac{d\boldsymbol{a}^T(\boldsymbol{X})}{d\boldsymbol{X}}\pm\frac{d\boldsymbol{b}^T(\boldsymbol{X})}{d\boldsymbol{X}} \qquad 加法运算公式\cdots\cdots（1-11）$$

$$\frac{d[\lambda(\boldsymbol{X})\boldsymbol{a}^T(\boldsymbol{X})]}{d\boldsymbol{X}} = \frac{d\lambda(\boldsymbol{X})}{d\boldsymbol{X}}\cdot\boldsymbol{a}^T(\boldsymbol{X})+\lambda(\boldsymbol{X})\cdot\frac{d\boldsymbol{a}^T(\boldsymbol{X})}{d\boldsymbol{X}}$$ 数乘运算公式……（1—12）

$$\frac{d[\boldsymbol{a}^T(\boldsymbol{X})\cdot\boldsymbol{b}(\boldsymbol{X})]}{d\boldsymbol{X}}=\frac{d\boldsymbol{a}^T(\boldsymbol{X})}{d\boldsymbol{X}}\cdot\boldsymbol{b}(\boldsymbol{X})+\frac{d\boldsymbol{b}^T(\boldsymbol{X})}{d\boldsymbol{X}}\cdot\boldsymbol{a}(\boldsymbol{X})$$ 乘法运算公式……（1—13）

证明最后一个公式，前两个公式请同学们根据定义去证明。

证：为简明起见隐去 $\boldsymbol{X}$

$$\frac{d}{d\boldsymbol{X}}[\boldsymbol{a}^T(\boldsymbol{X})\cdot\boldsymbol{b}(\boldsymbol{X})] = \begin{pmatrix}\frac{\partial}{\partial x_1}(\boldsymbol{a}^T\boldsymbol{b})\\ \vdots\\ \frac{\partial}{\partial x_i}(\boldsymbol{a}^T\boldsymbol{b})\\ \vdots\\ \frac{\partial}{\partial x_n}(\boldsymbol{a}^T\boldsymbol{b})\end{pmatrix} = \begin{pmatrix}\frac{\partial\boldsymbol{a}^T}{\partial x_1}\cdot\boldsymbol{b}+\boldsymbol{a}^T\cdot\frac{\partial\boldsymbol{b}}{\partial x_1}\\ \vdots\\ \frac{\partial\boldsymbol{a}^T}{\partial x_i}\cdot\boldsymbol{b}+\boldsymbol{a}^T\cdot\frac{\partial\boldsymbol{b}}{\partial x_i}\\ \vdots\\ \frac{\partial\boldsymbol{a}^T}{\partial x_n}\cdot\boldsymbol{b}+\boldsymbol{a}^T\cdot\frac{\partial\boldsymbol{b}}{\partial x_n}\end{pmatrix}$$

$$= \begin{pmatrix}\frac{\partial\boldsymbol{a}^T}{\partial x_1}\cdot\boldsymbol{b}+\boldsymbol{b}^T\cdot\frac{\partial\boldsymbol{a}}{\partial x_1}\\ \vdots\\ \frac{\partial\boldsymbol{a}^T}{\partial x_i}\cdot\boldsymbol{b}+\boldsymbol{b}^T\cdot\frac{\partial\boldsymbol{a}}{\partial x_i}\\ \vdots\\ \frac{\partial\boldsymbol{a}^T}{\partial x_n}\cdot\boldsymbol{b}+\boldsymbol{b}^T\cdot\frac{\partial\boldsymbol{a}}{\partial x_n}\end{pmatrix} = \frac{\partial\boldsymbol{a}^T(\boldsymbol{X})}{\partial\boldsymbol{X}}\cdot\boldsymbol{b}(\boldsymbol{X})+\boldsymbol{b}^T(\boldsymbol{X})\cdot\frac{\partial\boldsymbol{a}(\boldsymbol{X})}{\partial\boldsymbol{X}}$$

证毕

例 3： 求 $\frac{d\,\boldsymbol{X}}{d\boldsymbol{X}^T}=?$ $\frac{d\,\boldsymbol{X}^T}{d\boldsymbol{X}}=?$ 其中 $\boldsymbol{X}$ 为 n 维列向量

解：根据定义 4

$$\boldsymbol{a}(\boldsymbol{X})=\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}=\begin{bmatrix}\boldsymbol{a}_1(\boldsymbol{X})\\ \vdots\\ \boldsymbol{a}_m(\boldsymbol{X})\end{bmatrix}$$

$$\frac{d\boldsymbol{a}(\boldsymbol{X})}{d\boldsymbol{X}^T}=\begin{bmatrix}\frac{\partial a_1(\boldsymbol{X})}{\partial x_1}\cdots\frac{\partial a_1(\boldsymbol{X})}{\partial x_n}\\ \vdots\quad\ddots\quad\vdots\\ \frac{\partial a_n(\boldsymbol{X})}{\partial x_1}\cdots\frac{\partial a_n(\boldsymbol{X})}{\partial x_n}\end{bmatrix}=\begin{bmatrix}\frac{\partial x_1}{\partial x_1} & \cdots & \frac{\partial x_1}{\partial x_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial x_n}{\partial x_1} & \cdots & \frac{\partial x_n}{\partial x_n}\end{bmatrix}=\begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & 1 & \cdots & 0\\ \vdots & & \ddots & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}=\boldsymbol{I}$$

即 $$\frac{d\boldsymbol{X}}{d\boldsymbol{X}^T}=\boldsymbol{I} \qquad \cdots\cdots（1-14）$$

同理 $$\frac{d\boldsymbol{X}^T}{d\boldsymbol{X}}=\boldsymbol{I}$$ 注意：移乘作除要加转置 ……（1－15）

例 4：求$\frac{d}{d\boldsymbol{X}}(\boldsymbol{X}^T\boldsymbol{A})$　　$\boldsymbol{X}$—— n 维列向量，$\boldsymbol{A}$—— n×m 维常数阵

解：设$\boldsymbol{A}=\left[\boldsymbol{a}_1\ \ \boldsymbol{a}_2\ \ \cdots\ \ \boldsymbol{a}_m\right]$，$\boldsymbol{a}_i=[a_{1i}, a_{2i}, \cdots, a_{ni}]^{\mathrm{T}}$ 为 n×1 列向量

因此 $$\boldsymbol{X}^T\boldsymbol{A}=\left[\boldsymbol{X}^T\boldsymbol{a}_1\quad \boldsymbol{X}^T\boldsymbol{a}_2\quad \cdots\quad \boldsymbol{X}^T\boldsymbol{a}_m\right]$$

根据定义 $$\frac{d}{d\boldsymbol{X}}\left(\boldsymbol{X}^T\boldsymbol{A}\right)=\left[\frac{d}{d\boldsymbol{X}}\left(\boldsymbol{X}^T\boldsymbol{a}_1\right)\quad \frac{d}{d\boldsymbol{X}}\left(\boldsymbol{X}^T\boldsymbol{a}_2\right)\quad \cdots\quad \frac{d}{d\boldsymbol{X}}\left(\boldsymbol{X}^T\boldsymbol{a}_m\right)\right]$$

其中每一个列向量 $$\frac{d}{d\boldsymbol{X}}\left(\boldsymbol{X}^T\boldsymbol{a}_i\right)=\frac{d\boldsymbol{X}^T}{d\boldsymbol{X}}\cdot\boldsymbol{a}_i+\frac{d\boldsymbol{a}_i^T}{d\boldsymbol{X}}\cdot\boldsymbol{X}=\boldsymbol{a}_i$$

因此有 $$\frac{d}{d\boldsymbol{X}}\left(\boldsymbol{X}^T\boldsymbol{A}\right)=\left[\boldsymbol{a}_1\ \ \boldsymbol{a}_2\ \ \cdots\ \ \boldsymbol{a}_m\right]=\boldsymbol{A} \qquad \cdots\cdots（1-16）$$

推论：若$\boldsymbol{A}$为 n×n 方阵，有 $$\frac{d}{d\boldsymbol{X}}\left(\boldsymbol{X}^T\boldsymbol{A}^T\right)=\boldsymbol{A}^T \qquad \cdots\cdots（1-17）$$

例 5：求$\frac{d}{d\boldsymbol{X}^T}\left(\boldsymbol{B}\boldsymbol{X}\right)$　　$\boldsymbol{X}$—— n 维列向量，　$\boldsymbol{B}$—— m×n 矩阵

解：设 $$\boldsymbol{B}=\begin{bmatrix}\boldsymbol{b}_1^T\\ \boldsymbol{b}_2^T\\ \vdots\\ \boldsymbol{b}_m^T\end{bmatrix}$$ 则 $$\boldsymbol{B}\boldsymbol{X}=\begin{bmatrix}\boldsymbol{b}_1^T\boldsymbol{X}\\ \boldsymbol{b}_2^T\boldsymbol{X}\\ \vdots\\ \boldsymbol{b}_m^T\boldsymbol{X}\end{bmatrix}$$

类似可得：$$\frac{d}{d\boldsymbol{X}^T}\left(\boldsymbol{B}\boldsymbol{X}\right)=\boldsymbol{B} \qquad \cdots\cdots（1-18）$$

例 6：求二次型$\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}$对$\boldsymbol{X}$的导数，$\boldsymbol{A}$为对称方阵

解：根据乘法运算公式（1－13）

$$\frac{d}{d\boldsymbol{X}}\left(\boldsymbol{X}^T\boldsymbol{A}\,\boldsymbol{X}\right)=\frac{d\boldsymbol{X}^T}{d\boldsymbol{X}}\left(\boldsymbol{A}\,\boldsymbol{X}\right)+\frac{d\left(\boldsymbol{A}\,\boldsymbol{X}\right)^T}{d\boldsymbol{X}}\ \boldsymbol{X}$$

$$=\boldsymbol{A}\,\boldsymbol{X}+\frac{d\left(\boldsymbol{X}^T\boldsymbol{A}^T\right)}{d\boldsymbol{X}}\cdot\boldsymbol{X}=\boldsymbol{A}\,\boldsymbol{X}+\boldsymbol{A}^T\boldsymbol{X}$$

$$=\left(\boldsymbol{A}+\boldsymbol{A}^T\right)\boldsymbol{X}=2\boldsymbol{A}\,\boldsymbol{X}$$

即 $$\frac{d}{d\boldsymbol{X}}\left(\boldsymbol{X}^T\boldsymbol{A}\,\boldsymbol{X}\right)=2\,\boldsymbol{A}\,\boldsymbol{X}\qquad\cdots\cdots（1—19）$$

根据（1—10）式 $\frac{d\boldsymbol{a}^T(\boldsymbol{X})}{d\boldsymbol{X}}=\left[\frac{d\boldsymbol{a}(\boldsymbol{X})}{d\boldsymbol{X}^T}\right]^T \to\to \frac{d\boldsymbol{a}(\boldsymbol{X})}{d\boldsymbol{X}^T}=\left[\frac{d\boldsymbol{a}^T(\boldsymbol{X})}{d\boldsymbol{X}}\right]^T$ （两边同取转置）

有 $$\frac{d}{d\boldsymbol{X}^T}\left(\boldsymbol{X}^T\boldsymbol{A}\,\boldsymbol{X}\right)=\left[\frac{d}{d\boldsymbol{X}}(\boldsymbol{X}^T\boldsymbol{A}\,\boldsymbol{X})\right]^T=[2\,\boldsymbol{A}\,\boldsymbol{X}]^T=2\boldsymbol{X}^T\boldsymbol{A}$$

例 7：求函数 $\boldsymbol{\lambda}^T\boldsymbol{AX}$ 对 $\boldsymbol{X}$ 的导数， 其中 $\boldsymbol{\lambda}^T$—— 1×n 行向量，

$\boldsymbol{A}$—— n×n 常数阵， $\boldsymbol{X}$—— n 维列向量

解：

$$\boldsymbol{\lambda}^T\boldsymbol{AX}=\left(\boldsymbol{\lambda}^T\boldsymbol{AX}\right)^T=\boldsymbol{X}^T\boldsymbol{A}^T\boldsymbol{\lambda}$$

因为 $\boldsymbol{\lambda}^T\boldsymbol{AX}$ 是标量，所以它与它的转置相等

$$\frac{d}{d\boldsymbol{X}}\left(\boldsymbol{\lambda}^T\boldsymbol{AX}\right)=\frac{d}{d\boldsymbol{X}}\left(\boldsymbol{X}^T\boldsymbol{A}^T\boldsymbol{\lambda}\right)=\boldsymbol{A}^T\boldsymbol{\lambda}$$

例 8：求方程 $\boldsymbol{AX}=\boldsymbol{b}$ 的最小范数的平方解，其中 $\boldsymbol{A}$ 是 m×n 阶常数矩阵，其秩为 m(m<n), $\boldsymbol{b}$ 为 m×1 常数列向量。

解：这实际上就是求数量函数 $f\left(x\right)=\boldsymbol{X}^T\cdot\boldsymbol{X}=\|\boldsymbol{X}\|^2$，在约束条件 $\boldsymbol{AX}=\boldsymbol{b}$ 的条件极小值，

采用拉格朗日乘数法，作函数

$$F\left(\boldsymbol{X}\right)=\boldsymbol{X}^T\cdot\boldsymbol{X}+\boldsymbol{\lambda}^T\left(\boldsymbol{AX}-\boldsymbol{b}\right)$$

$\frac{d\boldsymbol{F}\left(\boldsymbol{X}\right)}{d\boldsymbol{X}}=2\boldsymbol{X}+\boldsymbol{A}^T\boldsymbol{\lambda}=\boldsymbol{0}$　解出 $\boldsymbol{X}=-\frac{1}{2}\boldsymbol{A}^T\boldsymbol{\lambda}$　代入约束方程

$-\frac{1}{2}\boldsymbol{AA}^T\boldsymbol{\lambda}=\boldsymbol{b}$　其中 $\boldsymbol{AA}^T$ 是 m×m 常数矩阵，根据给定条件，秩为 m，其逆存在

因而有：$\boldsymbol{\lambda}=-2\left(\boldsymbol{AA}^T\right)^{-1}\boldsymbol{b}$　代入 $\boldsymbol{X}$ 的表达式

$$\boldsymbol{X}=\boldsymbol{A}^T\left(\boldsymbol{AA}^T\right)^{-1}\boldsymbol{b}$$

再由$\dfrac{d}{d\boldsymbol{X}^T}\left[\dfrac{d\boldsymbol{F}(\boldsymbol{X})}{d\boldsymbol{X}}\right]=\dfrac{d}{d\boldsymbol{X}^T}\left(2\boldsymbol{X}+\boldsymbol{A}^T\boldsymbol{\lambda}\right)=2\boldsymbol{I}>\boldsymbol{0}$

可知所得的解是最小范数解。

三、 相对于矩阵的微分（自变量是矩阵 ）

1、数量函数的导数

设函数 $f=f(\boldsymbol{A})$ 是以 P×m 矩阵 $\boldsymbol{A}$ 的 P×m 元素 a_{ij} 为自变量的数量函数，简称以矩阵 $\boldsymbol{A}$ 为自变量的数量函数。例如

$$f=a_{11}^3+(1+a_{12})a_{11}^2+(a_{21}+a_{22}+a_{23})a_{11}+a_{21}+a_{22}$$

$$=\begin{bmatrix}a_{11} & 1\end{bmatrix}\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}\begin{pmatrix}a_{11}\\ 1\end{pmatrix}=\boldsymbol{a}^T\boldsymbol{A}\,\boldsymbol{a}=f(\boldsymbol{A}) \qquad \boldsymbol{A}=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}$$

定义：P×m 矩阵

$$\begin{pmatrix}\dfrac{\partial f}{\partial a_{11}} & \cdots & \dfrac{\partial f}{\partial a_{1m}}\\ \vdots & \ddots & \vdots\\ \dfrac{\partial f}{\partial a_{p1}} & \cdots & \dfrac{\partial f}{\partial a_{pm}}\end{pmatrix}=\left[\dfrac{\partial f}{\partial a_{ij}}\right]_{pm}=\dfrac{d\,f(\boldsymbol{A})}{d\boldsymbol{A}} \qquad \cdots\cdots（1-20）$$

称为数量函数 f 对矩阵 $\boldsymbol{A}$ 的导数，记作 $\dfrac{d\,f(\boldsymbol{A})}{d\boldsymbol{A}}$

例 9：求 $f(\boldsymbol{A})=\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X}$ 对矩阵 $\boldsymbol{A}$ 的导数，其中向量 $\boldsymbol{X}$ 是定常的， $\boldsymbol{A}$ 是对称的。

解： $f(\boldsymbol{A})=\begin{bmatrix}x_1 & x_2\end{bmatrix}\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=x_1^2a_{11}+x_1x_2a_{12}+x_1x_2a_{21}+x_2^2a_{22}$

根据定义有

$$\frac{df(\boldsymbol{A})}{d\boldsymbol{A}}=\begin{pmatrix}\dfrac{\partial f}{\partial a_{11}} & \dfrac{\partial f}{\partial a_{12}}\\ \dfrac{\partial f}{\partial a_{21}} & \dfrac{\partial f}{\partial a_{22}}\end{pmatrix}=\begin{bmatrix}x_1^2 & x_1x_2\\ x_1x_2 & x_2^2\end{bmatrix}=\begin{bmatrix}x_1 & x_2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}=\boldsymbol{X}\boldsymbol{X}^T$$

即 $$\frac{d}{d\boldsymbol{A}}(\boldsymbol{X}^T\boldsymbol{A}\,\boldsymbol{X})=\boldsymbol{X}\,\boldsymbol{X}^T \qquad \cdots\cdots（1-21）$$

2. 向量函数的导数

设函数

$$\boldsymbol{Z}(\boldsymbol{A})=\begin{bmatrix} Z_1(\boldsymbol{A}) \\ Z_2(\boldsymbol{A}) \\ \vdots \\ Z_n(\boldsymbol{A}) \end{bmatrix}$$ 是以矩阵 $\boldsymbol{A}$ 为自变量的 n 维列向量函数，$\boldsymbol{A}=\begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pm} \end{pmatrix}$

定义：$$\frac{d\boldsymbol{Z}(\boldsymbol{A})}{d\boldsymbol{A}}=\begin{pmatrix} \dfrac{\partial \boldsymbol{Z}}{\partial a_{11}} & \cdots & \dfrac{\partial \boldsymbol{Z}}{\partial a_{1m}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial \boldsymbol{Z}}{\partial a_{p1}} & \cdots & \dfrac{\partial \boldsymbol{Z}}{\partial a_{pm}} \end{pmatrix}=\left(\frac{\partial \boldsymbol{Z}}{\partial a_{ij}}\right)_{pm} \qquad \cdots\cdots（1－22）$$

其中 $$\frac{\partial \boldsymbol{Z}}{\partial a_{ij}}=\begin{bmatrix} \dfrac{\partial z_1(\boldsymbol{A})}{\partial a_{ij}} \\ \vdots \\ \dfrac{\partial z_n(\boldsymbol{A})}{\partial a_{ij}} \end{bmatrix}_{n\times 1}$$

3． 矩阵函数的导数

设函数

$$\boldsymbol{F}(\boldsymbol{A})=\begin{pmatrix} f_{11}(\boldsymbol{A}) & \cdots & f_{1\ell}(\boldsymbol{A}) \\ \vdots & \ddots & \vdots \\ f_{n1}(\boldsymbol{A}) & \cdots & f_{n\ell}(\boldsymbol{A}) \end{pmatrix} \qquad \boldsymbol{A}=\begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{p1} & \cdots & a_{pm} \end{pmatrix}$$

定义：

$$\frac{d\boldsymbol{F}(\boldsymbol{A})}{d\boldsymbol{A}}=\begin{pmatrix} \dfrac{\partial \boldsymbol{F}(\boldsymbol{A})}{\partial a_{11}} & \cdots & \dfrac{\partial \boldsymbol{F}(\boldsymbol{A})}{\partial a_{1m}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial \boldsymbol{F}(\boldsymbol{A})}{\partial a_{p1}} & \cdots & \dfrac{\partial \boldsymbol{F}(\boldsymbol{A})}{\partial a_{pm}} \end{pmatrix} \qquad \cdots\cdots（1－23）$$

其中每个分块矩阵

$$\left[\frac{\partial \boldsymbol{F}(\boldsymbol{A})}{\partial a_{ij}}\right]=\begin{pmatrix} \dfrac{\partial f_{11}(\boldsymbol{A})}{\partial a_{ij}} & \cdots & \dfrac{\partial f_{1\ell}(\boldsymbol{A})}{\partial a_{ij}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_{n1}(\boldsymbol{A})}{\partial a_{ij}} & \cdots & \dfrac{\partial f_{n\ell}(\boldsymbol{A})}{\partial a_{ij}} \end{pmatrix}$$

例 10：$\boldsymbol{X}$ 是 n 维列向量，$\boldsymbol{Y}$ 是 m 维列向量，$\boldsymbol{A}$ 是 n×m 矩阵，求：$\dfrac{\partial \boldsymbol{X}^T\boldsymbol{A}\boldsymbol{Y}}{\partial \boldsymbol{A}}=?$

解：根据矩阵乘法

$$\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{Y}=\sum_{j=1}^{m}\sum_{i=1}^{n}a_{ij}\,x_i\,y_j \quad => \quad \frac{\partial \boldsymbol{X}^T\boldsymbol{A}\boldsymbol{Y}}{\partial a_{ij}}=x_i\,y_j$$

根据数量函数导数的定义

$$\frac{\partial \boldsymbol{X}^T\boldsymbol{A}\boldsymbol{Y}}{\partial \boldsymbol{A}}=\left[x_i\,y_j\right]_{nm}=\boldsymbol{X}\cdot\boldsymbol{Y}^T \qquad \cdots\cdots（1－24）$$

顺便说一下，$\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{Y}$ 是一个数量函数，与它的转置相等，即 $\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{Y}=\left[\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{Y}\right]^T=\boldsymbol{Y}^T\boldsymbol{A}^T\boldsymbol{X}$

所以又有 $$\frac{\partial\left(\boldsymbol{Y}^T\boldsymbol{A}^T\boldsymbol{X}\right)}{\partial \boldsymbol{A}}=\boldsymbol{X}\cdot\boldsymbol{Y}^T \qquad \cdots\cdots（1－25）$$

四、　复合函数的微分

公式 1　设 $f=f(\boldsymbol{Y})$，$\boldsymbol{Y}=\boldsymbol{Y}(\boldsymbol{X})$，则

$$\begin{cases}\dfrac{df}{d\boldsymbol{X}}=\dfrac{d\boldsymbol{Y}^T}{d\boldsymbol{X}}\,\dfrac{df}{d\boldsymbol{Y}}\\[2ex]\dfrac{df}{d\boldsymbol{X}^T}=\dfrac{df}{d\boldsymbol{Y}^T}\,\dfrac{d\boldsymbol{Y}}{d\boldsymbol{X}^T}\end{cases} \qquad \cdots\cdots（1－26）$$

证明：由给定条件有

$df=\dfrac{df}{d\boldsymbol{Y}^T}\cdot d\boldsymbol{Y}$ 和 $d\boldsymbol{Y}=\dfrac{d\boldsymbol{Y}}{d\boldsymbol{X}^T}\cdot d\boldsymbol{X}$

将上式结合起来

$$df=\frac{df}{d\boldsymbol{Y}^T}\cdot\frac{d\boldsymbol{Y}}{d\boldsymbol{X}^T}\cdot d\boldsymbol{X} \quad => \quad \frac{df}{d\boldsymbol{X}^T}=\frac{df}{d\boldsymbol{Y}^T}\cdot\frac{d\boldsymbol{Y}}{d\boldsymbol{X}^T}$$

公式 2　设 $f=f(\boldsymbol{X},\boldsymbol{Y})$，$\boldsymbol{Y}=\boldsymbol{Y}(\boldsymbol{X})$，则

$$\begin{cases}\dfrac{df}{d\boldsymbol{X}}=\dfrac{\partial f}{\partial \boldsymbol{X}}+\dfrac{d\boldsymbol{Y}^T}{d\boldsymbol{X}}\dfrac{\partial f}{\partial \boldsymbol{Y}}\\[2ex]\dfrac{df}{d\boldsymbol{X}^T}=\dfrac{df}{d\boldsymbol{X}^T}+\dfrac{\partial f}{\partial \boldsymbol{Y}^T}\dfrac{d\boldsymbol{Y}}{d\boldsymbol{X}^T}\end{cases} \qquad \cdots\cdots（1－27）$$