附录一　矩阵微分法

在现代控制理论中，常遇到矩阵微分法，就表达式

$$\frac{dA}{dB}$$

变量、向量、矩阵
variable
变量函数 variable function

来说，由于A和B都能分别是数量、向量或矩阵，而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。这八种导数的定义并不统一，这里采用的是多数的讲法。

下面先分别介绍八种导数的定义和运算公式，然后再统一起来。

我们采用符号X、a、b 代表列向量，A、B 代表矩阵。上标T代表转置。如不特别指明，所用的向量都假定是n 维的。

第一节　相对于数量变量的微分法

定义 1　对于n 维向量函数　向量函数 vector function

$$a(t)=[a_1(t)\quad a_2(t)\quad \cdots a_n(t)]^T$$

定义它对 t 的导数为

$$\frac{da(t)}{dt}\triangleq\left[\frac{da_1(t)}{dt}\quad\frac{da_2(t)}{dt}\cdots\frac{da_n(t)}{dt}\right]^T \qquad (1-1)$$

定义 2　对于$n\times m$ 的矩阵函数　矩阵函数　matrix function

$$A(t)=\begin{bmatrix} a_{11}(t)\cdots & a_{1m}(t)\\ \vdots & \vdots\\ a_{n1}(t)\cdots & a_{nm}(t)\end{bmatrix}=[a_{ij}(t)]_{nm}$$

定义它对 t 的导数为

$$\frac{dA(t)}{dt}=\begin{pmatrix}\frac{da_{11}(t)}{dt} & \cdots & \frac{da_{1m}(t)}{dt}\\ \vdots & & \vdots\\ \frac{da_{n1}(t)}{dt} & \cdots & \frac{da_{nm}(t)}{dt}\end{pmatrix}=\left[\frac{da_{ij}(t)}{dt}\right]_{nm} \quad (1-2)$$

不难看出，上述两个定义是一致的．当矩阵$A(t)$蜕化为向量 $a(t)$ 时，定义 2 就变成定义 1．就是说，矩阵 （包括向量）函数对数量变量的导数等于它的各个元素的导数，结果是个同阶的矩阵．

根据上述定义，可以推出下列的运算公式．

运算公式 1　在下列公式中，A、B 都是变量 t 的矩阵函数．但它们也可以代表向量函数（当其列数为一时），或数量函数（行数、列数都为一时）．其中 λ 是变量 t 的数量函数．

（1）加法运算公式

$$\frac{d}{dt}(A\pm B)=\frac{dA}{dt}\pm\frac{dB}{dt} \quad (1-3)$$

（2）数乘运算公式

$$\frac{d}{dt}(\lambda A)=\frac{d\lambda}{dt}A+\lambda\frac{dA}{dt} \quad (1-4)$$

（3）乘法运算公式

$$\frac{d}{dt}(AB)=\frac{dA}{dt}B+A\frac{dB}{dt} \quad (1-5)$$

这些公式都容易证明．现在来证明后一个公式．

证明　分两步证明．先证A、B均为向量的情况，即证

$$\frac{d}{dt}(a^Tb)=\frac{da^T}{dt}b+a^T\frac{db}{dt} \quad (1-6)$$

其中a^T 和b 分别为行向量和列向量；

$$a^{T}=[a_1\ \ a_2\cdots a_m],\quad b=\begin{pmatrix}b_1\\ b_2\\ \vdots\\ b_m\end{pmatrix}$$

将a^T 和b 代入方程(1—6)，直接求导并加以整理，可得

$$\frac{d}{dt}[a^Tb]=\sum_{k=1}^{m}\left(\frac{da_k}{dt}b_k+a_k\frac{db_k}{dt}\right)$$

$$=\frac{da^T}{dt}b+a^T\frac{db}{dt}$$

因此，当A、B 为向量时，公式（1—5）成立。

现设A、B 分别为$n\times m$ 和$m\times l$ 矩阵，可以写成

$$A=\begin{pmatrix}a_{11}\cdots a_{1m}\\ \vdots\qquad\vdots\\ a_{n1}\cdots a_{nm}\end{pmatrix}=\begin{pmatrix}a_1^T\\ \vdots\\ a_n^T\end{pmatrix},\quad B=\begin{pmatrix}b_{11}\cdots b_{1l}\\ \vdots\qquad\vdots\\ b_{m1}\cdots b_{ml}\end{pmatrix}=[b_1\cdots b_l]$$

其中a_i^T 代表A的第 i 行行向量；b_j 代表B的第 j列列向量，依此有

$$AB=\begin{pmatrix}a_1^Tb_1\cdots a_1^Tb_l\\ \vdots\\ a_n^Tb_1\cdots a_n^Tb_l\end{pmatrix}=[a_i^Tb_j]_{nl}$$

从而得

$$\frac{d}{dt}(AB)=\left[\frac{d}{dt}a_i^T\,b_j\right]_{nl}$$

$$=\left[\frac{da_i^T}{dt}b_j+a_i^T\frac{db_j}{dt}\right]_{nl}$$

$$=\frac{dA}{dt}B+A\frac{dB}{dt}$$

例1　求X^TAX 对 t 的导数。其中X 是 t 的n 维向量函数，A是$n\times n$ 对称常数矩阵。

解　利用公式（1—5），有

$$\frac{d}{dt}(X^T AX) = \frac{dX^T}{dt}AX + X^T\frac{d}{dt}(AX)$$

$$= \frac{dX^T}{dt}AX + X^T\left(\frac{dA}{dt}X + A\frac{dX}{dt}\right)$$

$$= 2X^T A\dot{X}$$

在写后一等式时，我们利用了 $\dot{X}^T AX$ 和 $X^T A\dot{X}$ 都是数量函数，且A为对称阵，它们等于自己的转置：

$$[\dot{X}^T AX]^T = X^T A\dot{X}$$

因而两者相等的事实。当A为单位阵时，此结果变为

$$\frac{d|X|^2}{dt} = 2X^T\dot{X} = 2\dot{X}^T X$$

上式的数量形式是

$$\frac{d}{dt}(x_1^2 + x_2^2 + \cdots + x_n^2) = 2(x_1\dot{x}_1 + x_2\dot{x}_2 + \cdots + x_n\dot{x}_n)$$

第二节　相对于向量的微分法

1. 数量函数的导数

设函数

$$f(X) = f(x_1, \cdots, x_n)$$

是以向量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

$$\begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{pmatrix}$$

叫做数量函数f对列向量X的导数，记作

$$\frac{df}{dX}=\begin{bmatrix}\frac{\partial f}{\partial x_1}\\ \vdots\\ \frac{\partial f}{\partial x_n}\end{bmatrix} \tag{2—1}$$

上述导数习惯上叫做函数 f 的梯度，它是三维空间中梯度概念的推广，记作 $\mathrm{grad}f$ 或 ∇f

例 2　求函数

$$f(X)=X^TX=x_1^2+x_2^2+\cdots+x_n^2$$

对 X 的导数。

解　根据定义

$$\frac{df}{dX}=\left[\frac{\partial f}{\partial x_1}\ \frac{\partial f}{\partial x_2}\cdots\frac{\partial f}{\partial x_n}\right]^T$$

$$=[2x_1\ 2x_2\cdots 2x_n]^T$$

$$=2X$$

上式的数量形式是

$$\frac{\partial f}{\partial x_1}=2x_1$$

$$\frac{\partial f}{\partial x_2}=2x_2$$

$$\vdots$$

$$\frac{\partial f}{\partial x_n}=2x_n$$

下列的运算公式是明显的：

运算公式 2　在下列公式中，f 和 g 都是 X 的数量函数。

（1）加法运算公式

$$\frac{d}{dX}(f\pm g)=\frac{df}{dX}\pm\frac{dg}{dX} \tag{2—2}$$

（2）乘法运算公式

$$\frac{d}{dX}(fg)=\frac{df}{dX}g+f\frac{dg}{dX} \tag{2—3}$$

我们把函数f对行向量X^T的导数定义为如下的行向量，记作

$$\frac{df}{dX^T}=\left[\frac{\partial f}{\partial x_1}\quad\frac{\partial f}{\partial x_2}\cdots\frac{\partial f}{\partial x_n}\right]$$

2．向量函数的导数

设函数

$$a(X)=\begin{pmatrix}a_1(X)\\ \vdots\\ a_m(X)\end{pmatrix}$$

是X的m维列向量函数。

定义 4　$n\times m$阶矩阵函数

$$\begin{pmatrix}\frac{\partial a_1}{\partial x_1}&\cdots&\frac{\partial a_m}{\partial x_1}\\ \vdots& &\vdots\\ \frac{\partial a_1}{\partial x_n}&\cdots&\frac{\partial a_m}{\partial x_n}\end{pmatrix}=\left[\frac{\partial a_j}{\partial x_i}\right]_{nm} \tag{2—4}$$

称为m维行向量函数$a^T(X)$对n维列向量X的导数。$m\times n$阶矩阵函数

$$\begin{pmatrix}\frac{\partial a_1}{\partial x_1}&\cdots&\frac{\partial a_1}{\partial x_n}\\ \vdots& &\vdots\\ \frac{\partial a_m}{\partial x_1}&\cdots&\frac{\partial a_m}{\partial x_n}\end{pmatrix}=\left[\frac{\partial a_i}{\partial x_j}\right]_{mn} \tag{2—5}$$

称为m维列向量函数$a(X)$对n维行向量X^T的导数。分别记作

$$\frac{da^T(X)}{dX}\quad 和\quad \frac{da(X)}{dX^T}$$

根据定义显然可见，导数矩阵的行数等于不加转置的向量的行数；列数等于加以转置的向量的列数。据此有

$$\frac{da^T}{dX}=\left(\frac{da}{dx^T}\right)^T$$

在此情况下，存在下列的运算公式：

运算公式 3　在以下的公式中，$a(X)$ 和$b(X)$ 是m 维的列向量函数，$\lambda(X)$ 是数量函数.

（1）加法运算公式

$$\frac{d}{dX}(a^T\pm b^T)=\frac{da^T}{dX}\pm\frac{db^T}{dX} \qquad (2-6)$$

（2）数乘运算公式

$$\frac{d}{dX}(\lambda a^T)=\frac{d\lambda}{dX}a^T+\lambda\frac{da^T}{dX} \qquad (2-7)$$

（3）乘法运算公式

$$\frac{d}{dX}(a^Tb)=\frac{da^T}{dX}b+\frac{db^T}{dX}a \qquad (2-8)$$

头两个公式，两端都是$n\times m$ 矩阵，证明不难。后一个公式，两端都是$n\times 1$ 向量，证明如下。

证明　根据定义并利用公式（1—6），有

$$\frac{d}{dX}(a^Tb)=\begin{pmatrix}\dfrac{\partial}{\partial x_1}(a^Tb)\\ \vdots\\ \dfrac{\partial}{\partial x_i}(a^Tb)\\ \vdots\\ \dfrac{\partial}{\partial x_n}(a^Tb)\end{pmatrix}=\begin{pmatrix}\dfrac{\partial a^T}{\partial x_1}b+a^T\dfrac{\partial b}{\partial x_1}\\ \vdots\\ \dfrac{\partial a^T}{\partial x_i}b+a^T\dfrac{\partial b}{\partial x_i}\\ \vdots\\ \dfrac{\partial a^T}{\partial x_n}b+a^T\dfrac{\partial b}{\partial x_n}\end{pmatrix}$$

$$
=\begin{bmatrix} \dfrac{\partial a^T}{\partial x_1}b+\dfrac{\partial b^T}{\partial x_1}a \\ \vdots \\ \dfrac{\partial a^T}{\partial x_i}b+\dfrac{\partial b^T}{\partial x_i}a \\ \vdots \\ \dfrac{\partial a^T}{\partial x_n}b+\dfrac{\partial b^T}{\partial x_n}a \end{bmatrix}=\frac{da^T}{dX}b+\frac{db^T}{dX}a
$$

关于对行向量 X^T 求导的运算公式，与此相仿，就不重复了。留给读者自行写出，当作练习。

需要指出，根据定义我们可以直接验证两个有用的等式：

$$
\frac{dX}{dX^T}=I \quad 和 \quad \frac{dX^T}{dX}=I \qquad (2-9)
$$

其中 I 是 $n\times n$ 单位矩阵。但要注意，从以上两个等式可以看出下列运算是错误的，拿头个等式说，不能把左端的分母移乘到右端得出

$$
dX=dX^T
$$

如果要把 dX （或 dX^T ）移乘作除或移除作乘的话，必须加以转置。例如，将等式

$$
dX=dX
$$

右端的 dX 移到左端作除时，应加以转置，得

$$
\frac{dX}{dX^T}=I
$$

例 3　（1）求行向量 X^TA 对 X 的导数；

（2）求列向量 BX 对 X^T 的导数，其中 A、B 是常数矩阵，但不一定是方阵。

解（1）设 X 为 n 维列向量，A 为 $n\times m$ 矩阵，可以写成

$$
A=[a_1\ a_2\cdots a_m]
$$

其中a_i 是$n\times 1$ 列向量。因此可将行向量X^TA 表为

$$X^TA=[X^Ta_1\ X^Ta_2\cdots X^Ta_m]$$

根据定义则得

$$\frac{d}{dX}(X^TA)=\left[\frac{d}{dX}(X^Ta_1)\ \frac{d}{dX}(X^Ta_2)\cdots\frac{d}{dX}(X^Ta_m)\right]$$

其中的每个列向量，由公式（2—8）可得

$$\frac{d}{dX}(X^Ta_i)=\frac{dX^T}{dX}a_i+\frac{da_i^T}{dX}X$$

$$=a_i$$

因此有

$$\frac{d}{dX}(X^TA)=[a_1\ \ a_2\cdots a_m]=A$$

（2）设B 为$m\times n$ 矩阵，将它写成

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

其中b_i^T 是$1\times n$ 行向量。因此可将列向量BX 表示为

$$BX=\begin{bmatrix}b_1^TX\\ b_2^TX\\ \vdots\\ b_m^TX\end{bmatrix}$$

根据定义则得

$$\frac{d}{dX^T}(BX)=\begin{bmatrix}\frac{d}{dX^T}(b_1^TX)\\ \frac{d}{dX^T}(b_2^TX)\\ \vdots\\ \frac{d}{dX^T}(b_m^TX)\end{bmatrix}$$

其中的每项可算出为

$$\frac{d}{dX^T}(b_i^T X) = X^T \frac{db_i}{dX^T} + b_i^T \frac{dX}{dX^T}$$

$$= b_i^T$$

因此有

$$\frac{d}{dX^T}(BX) = B$$

从以上结果可以看出，在此情况下的常数矩阵可以当成常数，提出导数号外，再进行求导。

例 4　求二次型$X^T AX$ 对X 的导数

解　利用公式（2—8），将列向量AX 看成b，得到

$$\frac{d}{dX}(X^T AX) = \frac{dX^T}{dX}AX + \frac{d(AX)^T}{dX}X$$

$$= AX + A^T X$$

$$= (A + A^T)X$$

它是个列向量。当A为2×2矩阵时，上式的数量形式是：

$$\frac{\partial}{\partial x_1}[a_{11}x_1^2 + (a_{12} + a_{21})x_1x_2 + a_{22}x_2^2] = 2a_{11}x_1 + (a_{12} + a_{21})x_2$$

$$\frac{\partial}{\partial x_2}[a_{11}x_1^2 + (a_{12} + a_{21})x_1x_2 + a_{22}x_2^2] = (a_{12} + a_{21})x_1 + 2a_{22}x_2$$

当A为对称阵时，上式变为

$$\frac{d}{dX}(X^T AX) = 2AX$$

同理可得

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

当A为对称阵时，上式变为

$$\frac{d}{dX^T}(X^T AX) = 2X^T A$$

它是个行向量。以上结果可视为数量函数 cx^2 导数公式的推广。

例 5　求数量函数$p^T AX$ 对X 的导数，其中p^T是$1\times n$ 行向量，A是$n\times n$ 矩阵，都是常量。X是$n\times 1$ 列向量。

解　因为$p^T AX$ 是数量函数，它等于自己的转置，即

$$p^T AX = X^T A^T p$$

考虑到$A^T p$ 是常数向量，可以提出导数符号外，便得

$$\frac{d}{dX}(p^T AX) = A^T p$$

例 6　求方程

$$AX = b$$

的最小范数解。其中A是$m\times n$ 常数矩阵，其秩为 $m(m<n)$，b为$m\times 1$ 常数列向量。

解　这实际上就是求数量函数

$$f(X) = X^T X$$

在条件

$$AX = b$$

下的条件极小值。采用拉格朗日乘子法，作函数

$$F(X) = X^T X + p^T(AX - b)$$

其中m 维行向量p^T 代表乘子，求上式对X 的导数，并令它等于零，得到

$$\frac{dF}{dX} = 2X + A^T p = 0$$

由此解出

$$X = -\frac{1}{2}A^T p$$

将它代入约束方程，可得

$$-\frac{1}{2}AA^T p = b$$

其中AA^T 是$m\times m$ 常数矩阵，根据给定条件可知它的秩为 m，其逆存在。因而有

$$p=-2(AA^T)^{-1}b$$

将它代入X的表达式，便得

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

再由

$$\frac{d}{dX^T}\frac{dF}{dX}=\frac{d^2F}{dX^2}=2I>0$$

可知所得到的解是最小范数解。这也就是第三章第一节中定理2给出的结果。

需要指出，根据定义我们可以直接验证下列等式

$$\frac{da^T}{dX}=\left[\frac{da_1}{dX}\ \frac{da_2}{dX}\cdots\frac{da_m}{dX}\right]=\begin{bmatrix}\frac{\partial a^T}{\partial x_1}\\ \frac{\partial a^T}{\partial x_2}\\ \vdots\\ \frac{\partial a^T}{\partial x_n}\end{bmatrix}$$

$$\frac{da}{dX^T}=\begin{bmatrix}\frac{da_1}{dX^T}\\ \frac{da_2}{dX^T}\\ \vdots\\ \frac{da_m}{dX^T}\end{bmatrix}=\left[\frac{\partial a}{\partial x_1}\ \frac{\partial a}{\partial x_2}\cdots\frac{\partial a}{\partial x_n}\right]$$

3．矩阵函数的导数

设函数

$$A(X)=\begin{bmatrix}a_{11}(X)\cdots a_{1l}(X)\\ \vdots\qquad\quad\vdots\\ a_{m1}(X)\cdots a_{ml}(X)\end{bmatrix}$$

是X的 $m\times l$ 矩阵函数，即其中的每个元素都是X的函数。

定义 5　$nm\times l$ 的矩阵函数

$$\begin{pmatrix} \dfrac{\partial A(X)}{\partial x_1} \\ \dfrac{\partial A(X)}{\partial x_2} \\ \vdots \\ \dfrac{\partial A(X)}{\partial x_n} \end{pmatrix}$$

称为$m\times l$ 矩阵函数$A(X)$ 对列向量X的导数。$m\times ln$ 的矩阵函数

$$\left[\frac{\partial A(X)}{\partial x_1} \quad \frac{\partial A(X)}{\partial x_2} \cdots \frac{\partial A(X)}{\partial x_n}\right]$$

称为$m\times l$ 矩阵函数$A(X)$ 对行向量 X^T 的导数。其中的每个分块矩阵是矩阵函数$A(X)$ 对变量 x_i 的导数矩阵，仍是一个$m\times l$矩阵：

$$\left[\frac{\partial A(X)}{\partial x_i}\right]=\begin{pmatrix} \dfrac{\partial a_{11}(X)}{\partial x_i} & \cdots & \dfrac{\partial a_{1l}(X)}{\partial x_i} \\ \vdots & & \vdots \\ \dfrac{\partial a_{m1}(X)}{\partial x_i} & \cdots & \dfrac{\partial a_{ml}(X)}{\partial x_i} \end{pmatrix}$$

我们将以上两种导数分别记作

$$\frac{dA}{dX} \text{ 和 } \frac{dA}{dX^T}$$

一般地说，上述两种导数不存在互为转置的关系，即

$$\frac{dA}{dX}\neq\left[\frac{dA}{dX^T}\right]^T \quad \left(\text{但有}\frac{dA^T}{dX}=\left[\frac{dA}{dX^T}\right]^T\right)$$

不难看出，当$A(X)$ 蜕化为向量时，上述定义就与定义 4 相同了。

可以证明存在下列的运算公式。

运算公式 4　在以下的公式中，A 和 C 都是 $p\times m$ 矩阵，B 是 $m\times l$ 矩阵，λ 是 X 的数量函数。

（1）加法运算公式

$$\frac{d}{dX}(A\pm C)=\frac{dA}{dX}\pm\frac{dC}{dX} \tag{2—10}$$

（2）数乘运算公式

$$\frac{d}{dX}(\lambda A)=\left[\frac{d\lambda}{dX}\right]A+\lambda\frac{dA}{dX} \tag{2—11}$$

其中右端第一项的含义是，把 A 看成一个数量那样与向量 $\frac{d\lambda}{dX}$ 相乘，即

$$\left[\frac{d\lambda}{dX}\right]A\triangleq\begin{bmatrix}\frac{\partial\lambda}{\partial x_1}A\\ \frac{\partial\lambda}{\partial x_2}A\\ \vdots\\ \frac{\partial\lambda}{\partial x_n}A\end{bmatrix}$$

其中的每个分块都是 $p\times m$ 矩阵，所以它本身是个 $np\times m$ 矩阵。

（3）乘法运算公式

$$\frac{d}{dX}(AB)=\frac{dA}{dX}B+A\left[\frac{dB}{dX}\right]$$

其中上式右端第二项的含义是

$$A\left[\frac{dB}{dX}\right]\triangleq\begin{bmatrix}A\dfrac{\partial B}{\partial x_1}\\A\dfrac{\partial B}{\partial x_2}\\\vdots\\A\dfrac{\partial B}{\partial x_n}\end{bmatrix}$$

其中每个分块都是 $p\times l$ 矩阵，所以它本身是个 $np\times l$ 矩阵。现在我们给出这个公式的证明。

证明　将矩阵A和B分别写成

$$A=\begin{bmatrix}a_1^T\\a_2^T\\\vdots\\a_p^T\end{bmatrix},\quad B=[b_1\ \ b_2\cdots b_l]$$

其中a_i^T 是A的第 i 行行向量，b_i 是 B 的第 i 列列向量，依此有

$$AB=\begin{bmatrix}a_1^Tb_1&\cdots a_1^Tb_l\\\vdots&\vdots\\a_p^Tb_1&\cdots a_p^Tb_l\end{bmatrix}=[a_i^Tb_j]_{pl}$$

因而

$$\frac{\partial}{\partial x_k}(AB)=\left[\frac{\partial}{\partial x_k}(a_i^Tb_j)\right]_{pl}$$

$$=\left[\frac{\partial a_i^T}{\partial x_k}b_j+a_i^T\frac{\partial b_j}{\partial x_k}\right]_{pl}$$

$$=\frac{\partial A}{\partial x_k}B+A\frac{\partial B}{\partial x_k}$$

这实际就是公式（1—5），由此得

$$\frac{d}{dX}(AB)=\begin{bmatrix}\frac{\partial}{\partial x_1}(AB)\\ \frac{\partial}{\partial x_2}(AB)\\ \vdots\\ \frac{\partial}{\partial x_n}(AB)\end{bmatrix}=\begin{bmatrix}\frac{\partial A}{\partial x_1}B+A\frac{\partial B}{\partial x_1}\\ \frac{\partial A}{\partial x_2}B+A\frac{\partial B}{\partial x_2}\\ \vdots\\ \frac{\partial A}{\partial x_n}B+A\frac{\partial B}{\partial x_n}\end{bmatrix}$$

$$=\frac{dA}{dX}B+A\left[\frac{dB}{dX}\right] \qquad (2-12)$$

这里有个问题要解释一下．上式右端第一项中的$\frac{dA}{dX}$是个 $np\times m$ 矩阵，B是个 $m\times l$ 矩阵，所以

$$\begin{bmatrix}\frac{\partial A}{\partial x_1}B\\ \frac{\partial A}{\partial x_2}B\\ \vdots\\ \frac{\partial A}{\partial x_n}B\end{bmatrix}=\frac{dA}{dX}B$$

而第二项中的A是个 $p\times m$ 矩阵，$\frac{dB}{dX}$是个 $nm\times l$ 矩阵．两者不能相乘，自然不能等于矩阵

$$\begin{bmatrix}A\frac{\partial B}{\partial x_1}\\ A\frac{\partial B}{\partial x_2}\\ \vdots\\ A\frac{\partial B}{\partial x_n}\end{bmatrix}$$

所以我们把它写成公式中的形式。

不难看出，当A和B蜕化为向量时，上述运算公式就与运算公式3一致了。

例7　求行向量X^TA对X的导数。

解　利用公式（2—12），有

$$\frac{d}{dX}(X^TA)=\frac{dX^T}{dX}A+X^T\left[\frac{dA}{dX}\right]=A+X^T\left[\frac{dA}{dX}\right]$$

当A是常数阵时，得

$$\frac{d(X^TA)}{dX}=A$$

比例3的算法简单，

第三节　相对于矩阵的微分法

1．数量函数的导数

设函数

$$f=f(A)$$

是以$p\times m$矩阵A的$p\times m$个元a_{ij}为自变量的数量函数，简称以矩阵A为自变量的数量函数。例如函数

$$f=a_{11}^2+2a_{12}a_{21}+5a_{12}a_{22}$$

就是以

$$A=\begin{pmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}$$

为自变量的函数。

定义6　$p\times m$矩阵

$$\begin{pmatrix} \frac{\partial f}{\partial a_{11}} & \cdots & \frac{\partial f}{\partial a_{1m}} \\ \vdots & & \vdots \\ \frac{\partial f}{\partial a_{p1}} & \cdots & \frac{\partial f}{\partial a_{pm}} \end{pmatrix} = \left[\frac{\partial f}{\partial a_{ij}}\right]_{pm} \tag{3—1}$$

称为数量函数 f 对矩阵 A 的导数，记作

$$\frac{df}{dA}$$

例 8　求 $f=X^{T}AX$ 对矩阵 A 的导数，其中向量 X 是定常的，A 是对称的。

解　我们先研究 A 为二阶矩阵的情况，这时数量函数

$$f=[x_1\quad x_2]\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$=x_1^2a_{11}+x_1x_2a_{12}+x_1x_2a_{21}+x_2^2a_{22}$$

根据定义有

$$\frac{df}{dA}=\begin{pmatrix} \frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} \\ \frac{\partial f}{\partial a_{21}} & \frac{\partial f}{\partial a_{22}} \end{pmatrix}=\begin{pmatrix} x_1^2 & x_1x_2 \\ x_1x_2 & x_2^2 \end{pmatrix}$$

$$=\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}[x_1x_2]=XX^T$$

对于一般的情况，函数

$$f=X^TAX=\sum_{i=1}^{n}\sum_{j=1}^{n}x_ix_ja_{ij}$$

根据定义直接可以算出

$$\frac{df}{dA}=XX^T$$

2. 向量函数的导数

设函数

$$Z(A)=[z_1(A)\,z_2(A)\cdots z_n(A)]^T$$

是以矩阵A为自变量的n 维列向量函数。

定义 7　$np\times m$ 矩阵

$$\begin{bmatrix} \frac{\partial Z}{\partial a_{11}} & \cdots & \frac{\partial Z}{\partial a_{1m}} \\ \vdots & & \vdots \\ \frac{\partial Z}{\partial a_{p1}} & \cdots & \frac{\partial Z}{\partial a_{pm}} \end{bmatrix}=\left[\frac{\partial Z}{\partial a_{ij}}\right]_{pm} \qquad (3-2)$$

称为列向量函数 $Z(A)$ 对 $p\times m$ 矩阵A的导数，其中每个分块矩阵是个 $n\times 1$ 矩阵：

$$\frac{\partial Z}{\partial a_{ij}}=\begin{bmatrix} \frac{\partial z_1}{\partial a_{ij}} \\ \frac{\partial z_2}{\partial a_{ij}} \\ \vdots \\ \frac{\partial z_n}{\partial a_{ij}} \end{bmatrix} \qquad (3-3)$$

同样我们可以定义行向量函数$Z^T(A)$对A的导数，它是个 $p\times mn$ 矩阵，分别记为 $\frac{dZ}{dA}$ 和$\frac{dZ^T}{dA}$。

3. 矩阵函数的导数

设函数

$$F(A)=\begin{bmatrix} f_{11}(A)\cdots f_{1l}(A) \\ \vdots \qquad\qquad \vdots \\ f_{n1}(A)\cdots f_{nl}(A) \end{bmatrix}$$

是以 $p\times m$ 矩阵A为自变量的 $n\times l$ 矩阵函数。

定义 8　$np\times ml$ 矩阵

$$\begin{pmatrix} \frac{\partial F}{\partial a_{11}} & \cdots & \frac{\partial F}{\partial a_{1m}} \\ \vdots & & \vdots \\ \frac{\partial F}{\partial a_{p1}} & \cdots & \frac{\partial F}{\partial a_{pm}} \end{pmatrix} \qquad (3-4)$$

称为矩阵函数 $F(A)$ 对矩阵 A 的导数，其中每个分块矩阵是个 $n\times l$ 矩阵：

$$\left[\frac{\partial F}{\partial a_{ij}}\right] = \begin{pmatrix} \frac{\partial f_{11}}{\partial a_{ij}} & \cdots & \frac{\partial f_{1l}}{\partial a_{ij}} \\ \vdots & & \vdots \\ \frac{\partial f_{n1}}{\partial a_{ij}} & \cdots & \frac{\partial f_{nl}}{\partial a_{ij}} \end{pmatrix} \qquad (3-5)$$

此一导数的记号是：$\frac{dF}{dA}$

相对于矩阵求导的运算公式比较复杂，宜于具体情况具体处理，现把几个常用的公式汇列如下，以供查阅。

在下列公式中，X 是 n 维列向量，Y 是 m 维列向量，A 是 $n\times m$ 矩阵。

$$(1)\ \frac{\partial X^T AY}{\partial A} = XY^T \qquad (3-6)$$

证明　根据矩阵乘法，有

$$X^T AY = \sum_{j=1}^{m}\sum_{i=1}^{n} a_{ij}x_i y_j$$

由此可得

$$\frac{\partial X^T AY}{\partial a_{ij}} = x_i y_j$$

利用定义 6，便得

$$\frac{\partial X^T AY}{\partial A} = [x_i y_j]_{nm} = XY^T$$

由于X^TAY 是个数量函数，与它的转置相等，即

$$Y^TA^TX = X^TAY$$

利用此一结果，我们又得到所论公式的第二形式

$$\frac{\partial Y^TA^TX}{\partial A} = XY^T$$

(2) $$\frac{\partial Y^TA^TAY}{\partial A} = 2AYY^T \qquad (3-7)$$

证明　在上式左端含有矩阵的乘积A^TA. 根据数量函数乘积的求导公式，可以想到：上式的求导也应包括两项，一项只对A^T求导；一项只对A求导。就是说，不难验证下式的正确性

$$\frac{\partial Y^TA^TAY}{\partial A} = \frac{\partial Y^TA^TX}{\partial A} + \frac{\partial X^TAY}{\partial A} \quad (X = AY)$$

在以上求导过程中，X和Y都被看作常数向量. 依此，利用公式（3—6）就得到

$$\frac{\partial Y^TA^TAY}{\partial A} = XY^T + XY^T = 2AYY^T$$

(3) $$\frac{\partial (AY-X)^T(AY-X)}{\partial A} = 2(AY-X)Y^T \quad (3-8)$$

证明　我们有

$$\begin{aligned}&(AY-X)^T(AY-X)\\&= Y^TA^TAY - X^TAY - Y^TA^TX + X^TX\end{aligned}$$

求上式对A的导数，利用公式（3—6）和（3—7），则得

$$\begin{aligned}\frac{\partial (AY-X)^T(AY-X)}{\partial A} &= 2AYY^T - 2XY^T\\&= 2(AY-X)Y^T\end{aligned}$$

由于

$$(AY-X)^T(AY-X) = \mathrm{Tr}(AY-X)(AY-X)^T$$

利用公式（3—8），又可得

$$（4）\frac{\partial \mathrm{Tr}(AY-X)(AY-X)^T}{\partial A}=2(AY-X)Y^T \quad （3—9）$$

在下列公式中，A是$n\times m$ 变元矩阵，B 是 $m\times n$ 常数矩阵，C 是$m\times m$ 常数矩阵。

$$（5）\frac{\partial \mathrm{Tr}AB}{\partial A}=\frac{\partial \mathrm{Tr}BA}{\partial A}=\frac{\partial \mathrm{Tr}A^TB^T}{\partial A}=\frac{\partial \mathrm{Tr}B^TA^T}{\partial A}=B^T$$

（3—10）

证明　根据矩阵乘法，有

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

因此

$$\frac{\partial \mathrm{Tr}AB}{\partial a_{ij}}=b_{ji}$$

根据定义6，便得

$$\frac{\partial \mathrm{Tr}AB}{\partial A}=B^T \quad （3—11）$$

由于

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

利用公式（3—11）就可推出公式（3—10）的其余等式。

$$（6）\frac{\partial \mathrm{Tr}(ACA^T)}{\partial A}=A(C+C^T) \quad （3—12）$$

证明　容易验证在此情况下的函数乘积的求导法则照样是适用的，就是说成立

$$\frac{\partial \mathrm{Tr}ACA^T}{\partial A}=\frac{\partial \mathrm{Tr}AB_1}{\partial A}+\frac{\partial \mathrm{Tr}B_2A^T}{\partial A}$$

而其中的$B_1=CA^T$，$B_2=AC$都可看作常数矩阵，因此

$$\frac{\partial \operatorname{Tr} ACA}{\partial A} = B_1^T + B_2 = A(C^T + C)$$

同理可证（设 C 为 $n \times n$ 矩阵）

$$\frac{\partial \operatorname{Tr} A^T C A}{\partial A} = (C + C^T) A \qquad (3-13)$$

例 9　求函数

$$f = (X - a - BZ)^T (X - a - BZ)$$

对矩阵 B 的导数

解　利用公式（3—8），得

$$\frac{\partial f}{\partial B} = \frac{\partial}{\partial B} (X - a - BZ)^T (X - a - BZ)$$

$$= \frac{\partial}{\partial B} (BZ + a - X)^T (BZ + a - X)$$

$$= 2(BZ + a - X) Z^T$$

$$= -2(X - a - BZ) Z^T$$

分析上述八个定义可以看出，定义 8 是最广义的，它全部概括了以前的七个定义，将它们作为自己的特款。就是说，不论函数 F 和自变量 A 是数量、向量、矩阵，F 对 A 的导数总是按照下述的两个步骤构成的。即

（1）将自变量 A 的元换成函数 F 对各元的导数，得到分块矩阵（3—4）；

（2）对矩阵（3—4）的每个分块，例如 $\frac{\partial F}{\partial a_{ij}}$，把其中的函数 F 的各元换成该元的对 a_{ij} 的导数，得到矩阵（3—5）。

将以上两步结合起来，就得到 F 对 A 的导数。

例如，$F = [f_{11}(A) \quad f_{12}(A)]$ 对 $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ 的导数就

可以这样得到：

（1）将A的各元换成F对该元的导数，得

$$\frac{dF}{dA}=\begin{bmatrix}\dfrac{\partial F}{\partial a_{11}} & \dfrac{\partial F}{\partial a_{12}}\\[2ex] \dfrac{\partial F}{\partial a_{21}} & \dfrac{\partial F}{\partial a_{22}}\end{bmatrix}$$

（2）将其中的F的各元换成该元的相应导数，便得到

$$\frac{dF}{dA}=\begin{bmatrix}\dfrac{\partial f_{11}}{\partial a_{11}} & \dfrac{\partial f_{12}}{\partial a_{11}} & \dfrac{\partial f_{11}}{\partial a_{12}} & \dfrac{\partial f_{12}}{\partial a_{12}}\\[2ex] \dfrac{\partial f_{11}}{\partial a_{21}} & \dfrac{\partial f_{12}}{\partial a_{21}} & \dfrac{\partial f_{11}}{\partial a_{22}} & \dfrac{\partial f_{12}}{\partial a_{22}}\end{bmatrix}$$

由此可知，导数$\dfrac{dF}{dA}$的行数等于F和A的行数的乘积；列数等于F和A的列数的乘积。按此定义，n维列向量对n维列向量的导数将是一个n^2维的列向量。

从以上的讨论可知，相对于向量的微分法把相对于数量的微分法作为自己的特殊情况。此一说法自然也适用于运算公式。由于相对于矩阵的微分法应用少些，只要根据上述说明，掌握相对于向量的导数含义及运算公式4，就能满足实用需要了。

顺便指出，以上的定义可以说是以分母为主的定义，这是目前常用的。当然也可以采用以分子为主的定义，就是说在定义F对A的导数时原则上可以把上述两步颠倒一下次序，例如把

$$F=[f_1\ \ f_2],\quad A=[a_1\ \ a_2]$$

的导数$\dfrac{dF}{dA}$定义为

$$\left[\frac{\partial f_1}{\partial a_1}\quad\frac{\partial f_1}{\partial a_2}\quad\frac{\partial f_2}{\partial a_1}\quad\frac{\partial f_2}{\partial a_2}\right]$$

而不是现在这样定义为

$$\left[\frac{\partial f_1}{\partial a_1}\quad\frac{\partial f_2}{\partial a_1}\quad\frac{\partial f_1}{\partial a_2}\quad\frac{\partial f_2}{\partial a_2}\right]$$

但当F和A一为列向量另一为行向量时，两种定义是相同的。

第四节　复合函数微分法

下面将介绍一些最常用的基本公式。我们用f代表数量函数，Z代表l维列向量函数，Y代表m维列向量或函数，X代表n维列向量或函数，t代表数量变量。

1．数量函数的公式

公式1　设$f=f(Y)$，$Y=Y(t)$，则

$$\frac{df}{dt}=\frac{df}{dY^T}\frac{dY}{dt}=\frac{dY^T}{dt}\frac{df}{dY}\tag{4—1}$$

公式2　设$f=f(Y)$，$Y=Y(X)$，则

$$\frac{df}{dX}=\frac{dY^T}{dX}\frac{df}{dY}\tag{4—2}$$

$$\frac{df}{dX^T}=\frac{df}{dY^T}\frac{dY}{dX^T}\tag{4—3}$$

公式1是容易直接验证的。公式2有两个形式，它们是互为转置的关系，现在给出其证明如下。

证明　由给定条件，我们有

$$df=\frac{df}{dY^T}dY\ \text{及}\quad dY=\frac{dY}{dX^T}dX$$

其中dY和dX分别代表如下的m维和n维列向量

$$dY = \begin{pmatrix} dy_1 \\ dy_2 \\ \vdots \\ dy_n \end{pmatrix}, \qquad dX = \begin{pmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_n \end{pmatrix}$$

上面的两个微分公式都不难利用多元函数微分法直接验证。把它们结合起来就得到

$$df = \frac{df}{dY^T}\frac{dY}{dX^T}dX$$

将右端的 dX 加以转置后移乘作除，见公式（2—9），便得公式 2 的第二种形式：

$$\frac{df}{dX^T} = \frac{df}{dY^T}\frac{dY}{dX^T}$$

将它转置，并利用根据定义 3 和定义 4 推出的关系式

$$\left(\frac{df}{dX^T}\right)^T = \frac{df}{dX}, \quad \left(\frac{df}{dY^T}\right)^T = \frac{df}{dY}, \quad \left(\frac{dY}{dX^T}\right)^T = \frac{dY^T}{dX}$$

就得到公式 2 的第一种形式：

$$\frac{df}{dX} = \frac{dY^T}{dX}\frac{df}{dY}$$

由公式 2 很容易得到下面的公式：

公式 3　设 $f = f(X, Y)$，$Y = Y(X)$，则

$$\frac{df}{dX} = \frac{\partial f}{\partial X} + \frac{dY^T}{dX}\frac{\partial f}{\partial Y} \qquad (4-4)$$

$$\frac{df}{dX^T} = \frac{\partial f}{\partial X^T} + \frac{\partial f}{\partial Y^T}\frac{dY}{dX^T} \qquad (4-5)$$

例10　求 $f = X^TAX$ 对 X 的导数。

解　令 $Y = AX$，由于

$$\frac{dY^T}{dX} = \frac{dX^T}{dX}A^T = A^T$$

再利用公式 3 就得到

$$\frac{df}{dX}=\frac{\partial}{\partial X}(X^TY)+\frac{dY^T\partial}{dX\partial Y}(X^TY)$$

$$=Y+A^TX$$

$$=(A+A^T)X$$

例11　求方程

$$AX=b$$

的最小二乘解，其中A为$m\times n$常数矩阵，其秩为$n<m$。

解　这实际就是求数量函数

$$f=(AX-b)^T(AX-b)$$

的极小值，令

$$Y=AX-b$$

利用公式2及例2，可得

$$\frac{df}{dX}=\frac{dY^T}{dX}\frac{df}{dY}$$

$$=A^T\cdot 2Y$$

$$=2A^T(AX-b)$$

令上式等于零，解出

$$X=(A^TA)^{-1}A^Tb$$

这就是第三章第一节定理4给出的结果。其中有些问题需要说明一下。由于矩阵A的秩为n，$n\times n$矩阵(A^TA)的秩也是n，所以其逆$(A^TA)^{-1}$存在。又从

$$A^T(AX-b)=0$$

不能推出

$$(AX-b)=0$$

因为两个非零矩阵的积可能是零，不能由此消去A^T，这是与数量乘积不同的。

例12　求函数

$$f=(X-a-BZ)^T(X-a-BZ)$$

对n维向量a的偏导数。

解 设

$$Y=(X-a-BZ)$$

则根据

$$\frac{d}{dY}Y^TY=2Y$$

和

$$\frac{\partial}{\partial a}Y^T=\frac{\partial}{\partial a}(X-a-BZ)^T=-I$$

再利用公式2，便得

$$\frac{\partial f}{\partial a}=\frac{\partial Y^T}{\partial a}\frac{\partial f}{\partial Y}$$

$$=-2(X-a-BZ)$$

2．向量函数的公式

公式4 设 $\underset{l\times1}{Z}=Z(Y)$，$\underset{m\times1}{Y}=Y(t)$，则

$$\frac{dZ}{dt}=\frac{dZ}{dY^T}\frac{dY}{dt} \tag{4—6}$$

公式5 设$Z=Z(Y)$，$Y=Y(X)$ 则

$$\underset{n\times l}{\frac{dZ^T}{dX}}=\underset{n\times m}{\frac{dY^T}{dX}}\underset{m\times l}{\frac{dZ^T}{dY}} \tag{4—7}$$

$$\underset{l\times n}{\frac{dZ}{dX^T}}=\underset{l\times m}{\frac{dZ}{dY^T}}\underset{m\times n}{\frac{dY}{dX^T}} \tag{4—8}$$

公式6 设 $Z=Z(X,Y)$，$Y=Y(X)$，则

$$\frac{dZ^T}{dX}=\frac{\partial Z^T}{\partial X}+\frac{dY^T}{dX}\frac{\partial Z^T}{\partial Y} \tag{4—9}$$

$$\frac{dZ}{dX^T}=\frac{\partial Z}{\partial X^T}+\frac{\partial Z}{\partial Y^T}\frac{dY}{dX^T} \tag{4—10}$$

下面将给出公式5的证明，其它两个公式的证明类似，建议读者自己完成。

证明　我们直接可以验证下列的两个微分公式：

$$\underset{l\times 1}{dZ} = \underset{l\times m}{\frac{dZ}{dY^T}}\underset{m\times 1}{dY},\quad \underset{m\times 1}{dY} = \underset{m\times n}{\frac{dY}{dX^T}}\underset{n\times 1}{dX}$$

依此有

$$dZ = \frac{dZ}{dY^T}\frac{dY}{dX^T}dX$$

将右端dX 移乘作除，则得

$$\frac{dZ}{dX^T} = \frac{dZ}{dY^T}\frac{dY}{dX^T}$$

这就是公式5的第二个形式。将它转置就得到第一个形式：

$$\frac{dZ^T}{dX} = \frac{dY^T}{dX}\frac{dZ^T}{dY}$$

公式5的组成是有规律的：它包含三项，每一项的分子和分母必一为列向量一为行向量，在形式上根据

$$\frac{dY^T}{dY} = \frac{dY}{dY^T} = 1$$

把dY^T 和dY 对消后，两边是相同的，右端第一项中的列向量（即不带转置符号的量）必与左端项的列向量是一样的。如在头个式子中的dX，后一式子中的 dZ。这种规律性其它的同类公式也都具备。

例13　求向量函数$\sin(C^TX)\cdot X^T$ 对X的导数，其中C^T 是常数行向量。

解　将$\sin(C^TX)$看成公式（2—7）中的λ，可得

$$\frac{d}{dX}[\sin(C^TX)\cdot X^T] = \frac{d\sin(C^TX)}{dX}X^T + \sin(C^TX)\frac{dX^T}{dX}$$

将上式中的C^TX 看成一个数量变量，可得

$$\frac{d}{dX}[\sin(C^TX)] = \frac{d(C^TX)}{dX}\frac{d\sin(C^TX)}{d(C^TX)}$$

$$= C\cos(C^TX)$$

由此便得

$$\frac{d}{dX}[\sin(C^TX)\cdot X^T] = C\cos(C^TX)\cdot X^T + \sin(C^TX)I$$

例14 求 $f=(AX-b)^TR(AX-b)$ 对X的导数，其中A是$m\times n$ 常数矩阵，R是$m\times m$ 常数矩阵，X和b各是n维和m维列向量，其中b 是定常的。

解 设

$$Y = AX-b$$

则由例7的结果，有

$$\frac{dY^T}{dX} = A^T$$

由例9的结果，有

$$\frac{df}{dY} = (R+R^T)Y$$

再利用公式2便得到

$$\frac{df}{dX} = \frac{dY^T}{dX}\frac{df}{dY}$$

$$= A^T(R+R^T)Y$$

$$= A^T(R+R^T)(AX-b)$$

如令上式等于零，则可解出

$$X = [A^T(R+R^T)A]^{-1}A^T(R+R^T)b$$

这就是使函数f取极小值的解。当R为对称阵时，上式化为

$$X = (A^TRA)^{-1}A^TRb$$