Matrix Derivatives with Chain Rule and Rules for Simple, Hadamard, and Kronecker Products(2)

Matrix Derivatives with Kronecker Products

Wenchao Zhang, Rongpeng Li and Mubing Zhou

South University of Science and Technology of China

May 6, 2014

Notations and Definitions

I_n or I^n is the $n\times n$ identity matrix.

X' is the transpose of matrix X.

$\bar{X}$ is the $1\times nm$ row vectorization of a $n\times m$ matrix.

$D_{\bar{X}}$ is the $nm\times nm$ diagonal matrix whose diagonal elements $d_{ii}=x_i$, where x_i is entries of $\bar{X}$.

$X*Y$ is the $n\times m$ Hadamard product of two $n\times m$ matrices $X$and Y, i.e. $X*Y=[x_{ij}y_{ij}]$.

$X\otimes Y$ is the right Kronecker product of matrices X and Y, i.e. $X\otimes Y=[x_{ij}Y]$.

E^{mn} is an $nm\times nm$ matrix such that $e_{gh}=1$, if $1\leqslant g=n(j-1)+k, h=m(k-1)+j\leqslant mn$, where $0<j\leqslant m; 0<k\leqslant n$, and $e_{gh}=0$ otherwise.

Example. To make a better understanding, we give an example for E^{32} as follows:

$$E^{32} = \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right) \tag{1}$$

It's easy to see that we can view it as a block matrix, in each 2×3 block $E_{ij}(1 \leqslant i \leqslant 3,\ 1 \leqslant j \leqslant 2)$, where E_{ij} is the 2×3 matrices which only entry $e_{\mathrm{ij}} = 1$, and others are 0. Obviously, this conclusion is also true for any E^{nm}(Since first entry (says (1,1)) for E_{ij} is $((i-1)n+1, (j-1)m+1)$.).

Definition. (Partial Derivative of Matrices) *Let Z be a $p \times q$ matrix and Y is an $m \times n$ matrix whose elements are differentiable functions of all elements in Z. Define $\frac{\partial Y}{\partial Z}$ to be a $pq \times mn$ matrix whose i-th row is the mn-vector $\frac{\partial \bar{Y}}{\partial z_i}$, where z_i is the i-th element in $\bar{Z}$.*

Obviously from the definition, $\frac{\partial Y}{\partial Z} = \frac{\partial \bar{Y}}{\partial Z} = \frac{\partial \bar{Y}}{\partial \bar{Z}}$, i.e. there is an equivalence of matrix and vector derivatives.

Remark. (Vector Form) A row vector Y can be expressed as $Y = [Y_1, Y_2]$ where the elements of Y_1 are mathematical variables which are independent of each other, and Y_2 are functions of elements in Y_1.

Typically, if $Y = \bar{X}$,where X are symmetric matrix, Y_1 contains the nonredundant, lower(resp. upper) triangular elements of X.

Lemma 1. *Let* $X = [X_1', X_2']'$ *where* X_1 *is an* $m_1 \times n$ *matrix and* X_2 *is an* $m_2 \times n$ *matrix. Suppose each element of* X *is a differentiable function of all elements of a matrix* Z*. Then*

$$\frac{\partial X}{\partial Z} = \left[\frac{\partial X_1}{\partial Z}, \frac{\partial X_2}{\partial Z} \right]$$

Proof. *From the definition,*

$$\frac{\partial X}{\partial Z} = \frac{\partial \bar{X}}{\partial \bar{Z}} = \frac{\partial [\bar{X}_1, \bar{X}_2]}{\partial \bar{Z}} = \left[\frac{\partial \bar{X}_1}{\partial \bar{Z}}, \frac{\partial \bar{X}_2}{\partial \bar{Z}} \right] = \left[\frac{\partial X_1}{\partial Z}, \frac{\partial X_2}{\partial Z} \right]$$

$\square$

Remark. Actually, $X = [X_1', X_2']' = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$, therefore $\bar{X} = [\bar{X}_1, \bar{X}_2]$.

Lemma 2. *Let* $A = [A_{11}, \ldots, A_{1r}, \ldots, A_{n1}, \ldots, A_{nr}]$ *where* A_{ij} *are* $p \times s$ *matrices. Then*

$$A[E^{nr} \otimes I^s] = [A_{11}, \ldots, A_{n1}, \ldots, A_{1r}, \ldots, A_{nr}]$$

Proof. *Sustitute* I_s *into the places which is 1 in* E^{nr}*, it's obvious by the matrix multiplication.* $\square$

Example. *Consider* $n = 3, r = 2$*, from equation (1) we have*

$$E^{32} \otimes I^s = \left(\begin{array}{ccc|ccc} I_s & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_s & 0 & 0 \\ \hline 0 & I_s & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_s & 0 \\ \hline 0 & 0 & I_s & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_s \end{array}\right)$$

Hence,

$$A[E^{nr} \otimes I^s] = [A_{11}, A_{21}, A_{31}, A_{12}, A_{22}, A_{32}]$$

Lemma 3. *Let Z be a $p \times q$ matrix. Let X be a $1 \times n$ vector each of whose elements is a differentiable function of all elements of Z; and let Y be an $r \times s$ matrix each of whose elements is constant with respect to all elements of Z. Then*

$$\frac{\partial(X \otimes Y)}{\partial Z} = \left(\frac{\partial X}{\partial Z} \otimes \bar{Y}\right)(E^{nr} \otimes I^s)$$

Proof. *We have*

$$\left[\frac{\partial X}{\partial Z} \otimes \bar{Y}\right] = \left[\frac{\partial x_1}{\partial Z} Y_1, \ldots, \frac{\partial x_1}{\partial Z} Y_r, \ldots, \frac{\partial x_n}{\partial Z} Y_1, \ldots, \frac{\partial x_n}{\partial Z} Y_r\right]$$

where Y_i are rows of Y. By lemma 2,

$$\left[\frac{\partial X}{\partial Z} \otimes \bar{Y}\right][E^{nr} \otimes I^s] = \left[\frac{\partial x_1}{\partial Z} Y_1, \ldots, \frac{\partial x_n}{\partial Z} Y_1, \ldots, \frac{\partial x_1}{\partial Z} Y_r, \ldots, \frac{\partial x_n}{\partial Z} Y_r\right]$$

On the left hand side,

$$\frac{\partial(X \otimes Y)}{\partial Z} = \frac{\partial(\overline{X \otimes Y})}{\partial Z} = \left[\frac{\partial x_1}{\partial Z} Y_1, \ldots, \frac{\partial x_n}{\partial Z} Y_1, \ldots, \frac{\partial x_1}{\partial Z} Y_r, \ldots, \frac{\partial x_n}{\partial Z} Y_r\right]$$

$\square$

Lemma 4. *Let Z be a $p \times q$ matrix. Let X be an $m \times n$ matrix each of whose elements is a differentiable function of all elements of Z; and let Y an $r \times s$ matrix each of whose elements is constant with respect to all elements of Z. Then*

$$\frac{\partial(X \otimes Y)}{\partial Z} = \left(\frac{\partial X}{\partial Z} \otimes \bar{Y} \right)(I^m \otimes E^{nr} \otimes I^s)$$

Proof. *We prove by induction on m, Lemma 3 implies this lemma is true when $m = 1$. Now for m, partion X into$[X_1', X_m']'$ where X_1 consists of the first $m-1$ rows of X, and X_m is the m-th row of X. Then by lemmas 1 and 3*

$$\begin{aligned}
\frac{\partial(X \otimes Y)}{\partial Z} &= \frac{\partial(\overline{X \otimes Y})}{\partial Z} = \frac{\partial[(\overline{X_1 \otimes Y}), (\overline{X_m \otimes Y})]}{\partial Z} \\
&= \left[\left(\frac{\partial X_1}{\partial Z} \otimes \bar{Y} \right)(I^{m-1} \otimes E^{nr} \otimes I^s), \left(\frac{\partial X_m}{\partial Z} \otimes \bar{Y} \right)(E^{nr} \otimes I^s) \right] \\
&= \left(\frac{\partial X}{\partial Z} \otimes \bar{Y} \right)(I^m \otimes E^{nr} \otimes I^s)
\end{aligned}$$

□

Remark. Note that since $\bar{Y}$ is a row vector, we have

$$\left(\frac{\partial X}{\partial Z} \otimes \bar{Y} \right)(I^m \otimes E^{nr} \otimes I^s) = \frac{\partial X}{\partial Z}(I^{mn} \otimes \bar{Y})(I^m \otimes E^{nr} \otimes I^s)$$

Lemma 5. *Let Z be a $p\times q$ matrix. Let X be a $1\times n$ vector each of whose elements is constant with respect to all elements of Z; and let Y be an $r\times s$ matrix each of whose elements is a differentiable function of all elements of Z. Then*

$$\frac{\partial(X\otimes Y)}{\partial Z}=\left(\bar{X}\otimes\frac{\partial Y}{\partial Z}\right)(E^{nr}\otimes I^{s})$$

Proof. *Similar to the proof of lemma 3.* □

Lemma 6. *Let Z be a $p\times q$ matrix. Let X be an $m\times n$ matrix each of whose elements is a constant with respect to all elements of Z; and let Y an $r\times s$ matrix each of whose elements is differentiable function of all elements of Z. Then*

$$\frac{\partial(X\otimes Y)}{\partial Z}=\left(\bar{X}\otimes\frac{\partial Y}{\partial Z}\right)(I^{m}\otimes E^{nr}\otimes I^{s})$$

Proof. *Follows the same induction process as the proof of lemma 4.* □

Remark. Note that since $\bar{X}$ is a row vector, we have

$$\left(\bar{X}\otimes\frac{\partial Y}{\partial Z}\right)(I^{m}\otimes E^{nr}\otimes I^{s})=\frac{\partial Y}{\partial Z}(\bar{X}\otimes I^{rs})(I^{m}\otimes E^{nr}\otimes I^{s})$$

Theorem 7. (Derivative of Kronecker Products) *Let Z be a $p \times q$ matrix. Let X and Y be $m \times n$ and $r \times s$ matrices respectively, each of whose elements is differentiable function of all elements of Z. Then*

$$\frac{\partial (X \otimes Y)}{\partial Z} = \left[\frac{\partial X}{\partial Z}(I^{mn} \otimes \bar{Y}) + \frac{\partial Y}{\partial Z}(\bar{X} \otimes I^{rs}) \right](I^{m} \otimes E^{nr} \otimes I^{s})$$

Proof. *By lemmas 4 and 6.* □

Thank You!