Note on Matrix Derivatives with Chain Rule and Rules for Simple, Hadamard, and Kronecker Products

Rongpeng Li, Wenchao Zhang and Mubing Zhou

South University of Science and Techonology of China

Table of contents

1. Introduction 2

2. Notations and Definitions 2

3. Matrix Derivatives with Chain Rule and Simple Products 3

4. Matrix Derivatives with Rule For Hadamard Products 4

4.1. Hadamard Product 4
4.2. Properties of Hadamard Product 5
4.3. Schur Product Theorem 5
4.4. Product Rule for Hadamard Product 5
4.5. Physical Meaning of Hadamard Product 5
4.6. Hadamard Matrix 6
4.6.1. Introduction 6
4.6.2. Sylvester's construction 6

5. Matrix Derivatives with Rule For Kronecker Products 6

5.1. Main Result 6
5.2. Some Lemmas and Proofs 6
5.3. An Example (by Jöreskog 1970) 8

Bibliography 9

1. Introduction

Multivariable calculus has been a mature mathematical tool since 1970s, but people didn't have a very clear concept regarding to matrix derivative at that time. Though the chain rule and product rule for derivatives of composite scalar functions are very elementary result even for high school students, when it comes to matrix derivatives, people find it difficult. Simply transforming the results in scalar calculus to matrix calculus may yield wrong answers. For instance, the chain rule for a composite function is $\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial x}$. But if we replace x, y, z by matrices, the chain rule becomes invalid. The same situation is true for product rule. Moreover, not only simple product but other various products are defined, like Hadamard and Kronecker product. Finding the product rules for them is a even more hard work. So deriving rules for matrix calculus "would be virtually beyond the power of human thought to obtain in practice without the compact notation and the logical principles of the algebra of matrices". This paper, mainly based on the work of P.M. Bentler and S.Y. Lee [1], is devoted to seeking for the right expression for chain rule and the rules for simple, Hadamard and Kronecker product in matrix calculus.

2. Notations and Definitions

We are studying the general matrix calculus with all entries in the matrices either real numbers or real-valued functions.

We use the following 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.

Remark. 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. (A 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.

By now, it is necessary to give some more explict examples to illustrate the above concepts. Suppose $Z = (z_1, z_2), Y = \begin{pmatrix} y_1 & y_2 \\ y_3 & y_4 \end{pmatrix}$ and $X = \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix}$. Then

$$\begin{aligned}
\bar{Y} &= (y_1, y_2, y_3, y_4) \\
D_{\bar{Y}} &= diag\{y_1, y_2, y_3, y_4\} \\
X * Y &= \begin{pmatrix} x_1 y_1 & x_2 y_2 \\ x_3 y_3 & x_4 y_4 \end{pmatrix} \\
Z \otimes Y &= (z_1 Y, z_2 Y) \\
&= \begin{pmatrix} z_1 y_1 & z_1 y_2 & z_2 y_1 & z_2 y_2 \\ z_1 y_3 & z_1 y_4 & z_3 y_3 & z_2 y_4 \end{pmatrix}
\end{aligned}$$

where $*$ and $\otimes$ denote Hadamard product and Krnoecker product respectively. Then the derivative of Y with respect to Z can be shown to be

$$\frac{\partial Y}{\partial Z} = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \frac{\partial y_2}{\partial z_1} & \frac{\partial y_3}{\partial z_1} & \frac{\partial y_4}{\partial z_1} \\ \\ \frac{\partial y_1}{\partial z_2} & \frac{\partial y_2}{\partial z_2} & \frac{\partial y_3}{\partial z_2} & \frac{\partial y_4}{\partial z_2} \end{pmatrix}$$

3. MATRIX DERIVATIVES WITH CHAIN RULE AND SIMPLE PRODUCTS

The chain rule is given in the following theorem.

Theorem 1. *Suppose each element of a* ***vector*** $Y = [Y_1, Y_2]$ *is a differential function of all elements of a* ***vector*** Z*, where* Y_1 *contains all the mathematical variables which are independent to each other in* Y *and* Y_2 *are functions of elements in* Y_1*. Furthermore, suppose that each element of a* ***vector*** X *is a differentiable function of all elements of* Y_1*. Then*

$$\frac{\partial X}{\partial Z} = \frac{\partial Y_1}{\partial Z}\frac{\partial X}{\partial Y_1} \tag{2}$$

Proof. Suppose X is $1 \times r$, Y_1 is $1 \times s$ and Z is $1 \times t$. Then we have

$$\frac{\partial X}{\partial Z} = \begin{pmatrix} \frac{\partial x_1}{\partial z_1} & \cdots & \frac{\partial x_r}{\partial z_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial z_t} & \cdots & \frac{\partial x_r}{\partial z_t} \end{pmatrix}$$

and

$$\frac{\partial Y_1}{\partial Z}\frac{\partial X}{\partial Y_1} = \begin{pmatrix} \frac{\partial y_1}{\partial z_1} & \cdots & \frac{\partial y_s}{\partial z_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial z_t} & \cdots & \frac{\partial y_s}{\partial z_t} \end{pmatrix}\begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_r}{\partial y_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial y_s} & \cdots & \frac{\partial x_r}{\partial y_s} \end{pmatrix}$$

The (i,j)-th element in $\frac{\partial X}{\partial Z}$ is $\frac{\partial x_i}{\partial z_j}$ while the corresponding element in $\frac{\partial Y_1}{\partial Z}\frac{\partial X}{\partial Y_1}$ is $\sum_{l=1}^{s}\frac{\partial y_l}{\partial z_i}\frac{\partial x_j}{\partial y_l}$. By the chain rule in scalar calculus, we verify Eq.2. □

Note that in Theorem 1, we split Y into two parts. If we don't do that, the result may be wrong. Generally speaking, the following equation is NOT true:

$$\frac{\partial X}{\partial Z}=\frac{\partial Y}{\partial Z}\frac{\partial X}{\partial Y}$$

A simple counterexample is to take $X=[v_1,v_1,v_2,v_2]$, $Y=[v_1,v_2,v_1,v_2]$ and $Z=[v_1,v_2,v_2,v_1]$. Now $\frac{\partial X}{\partial Z}=\begin{pmatrix}1&1&0&0\\0&0&1&1\\0&0&1&1\\1&1&0&0\end{pmatrix}$ But $\frac{\partial Y}{\partial Z}\frac{\partial X}{\partial Y}=\begin{pmatrix}2&2&0&0\\0&0&2&2\\0&0&2&2\\2&2&0&0\end{pmatrix}$. Note that here $Y_1=[v_1,v_2]$.

The product rule is obtain in Theorem.2

Theorem 2. *Suppose each element of a $p\times r$ matrix Y and an $r\times q$ matrix X is a differentiable function of all elements of a $m\times n$ matrix Z. Then*

$$\frac{\partial (YX)}{\partial Z}=\frac{\partial Y}{\partial Z}(I^p\otimes X)+\frac{\partial X}{\partial Z}(Y'\otimes I^q) \tag{3}$$

Proof. $\frac{\partial (YX)}{\partial Z}$ is a $mn\times pq$ matrix. Its (i,j)th element corresponds to the i-th element in $\bar{Z}$, say z_β, and j-th element in $\overline{XY}$, say $(XY)_{(s,t)}$. $(XY)_{(s,t)}$ is the product of the s-th row of Y and t-th column of X, or $(XY)_{(s,t)}=\sum_{l=1}^{r} y_{sl}\,x_{lt}$. So

$$\frac{\partial (YX)}{\partial Z}=[\frac{\partial \sum_{l=1}^{r} y_{sl}\,x_{lt}}{\partial z_\beta}]=\sum_{l=1}^{r}\frac{\partial y_{sl}}{\partial z_\beta}x_{lt}+\frac{\partial x_{lt}}{\partial z_\beta}y_{sl}$$

Now consider $\frac{\partial Y}{\partial Z}(I^p\otimes X)$. $\frac{\partial Y}{\partial Z}$ is a $mn\times pr$ matrix while $I^p\otimes X$ is a $pr\times pq$ matrix. The (i,j)th element of $\frac{\partial Y}{\partial Z}(I^p\otimes X)$ is the product of i-th row of $\frac{\partial Y}{\partial Z}$ and j-th column of $I^p\otimes X$. The i-th row of corresponds to the i-th element of $\bar{Z}$, which is z_β.

Recall that $I^p\otimes X$ is in the form of $\operatorname{diag}\{X,X,...,X\}$ So the j-th column of $I^p\otimes X$ corresponds to the t-column of X. And the t-th column of X should be multiplied by the s-th row of Y. So the (i,j)th element of $\frac{\partial Y}{\partial Z}(I^p\otimes X)$ is $\sum_{l=1}^{k}\frac{\partial y_{sl}}{\partial z_\beta}x_{lt}$. Similarly, we can prove that the (i,j)th element of $\frac{\partial X}{\partial Z}(Y'\otimes I^q)$ is $\sum_{l=1}^{k}\frac{\partial x_{lt}}{\partial z_\beta}y_{sl}$. Add these two parts up is just the (i,j)-th element of $\frac{\partial (YX)}{\partial Z}$. □

4. MATRIX DERIVATIVES WITH RULE FOR HADAMARD PRODUCTS

4.1. Hadamard Product.

The Hadamard product (also known as the Schur product or the entrywise product is a binary operation that takes two matrices of the same dimensions, and produces another matrix where each element ij is the product of elements ij of the original two matrices.It is attributed to, and named after, either French mathematician Jacques Hadamard, or German mathematician Issai Schur. The Hadamard product is commutative, associative and distributive over addition. That is

$$A*B = B*A \tag{4}$$
$$A*(B*C) = (A*B)*C \tag{5}$$
$$A*(B+C) = A*B+A*C \tag{6}$$

The identity matrix under Hadamard multiplication of two m-by-n matrices is m-by-n matrix where all elements are equal to 1, which is different from simple identity matrix I.

A matrix has an inverse under Hadamard multiplication if and only if none of the elements are equal to zero.

In programming language like MATLAB, Hadamard product is done by using $.*$ where $*$ stands for the normal multiplication.

4.2. Properties of Hadamard Product.

For square A and B, the row-sums of their Hadamard product are the diagonal elements of $A\,B^T$

$$\sum_j (A*B)_{i,j} = (A\,B^T)_{i,j} \tag{7}$$

The Hadamard product is a principal submatrix of the Kronecker product.

A principal submatrix is a square submatrix where the distinguished rows and columns are the same.

4.3. Schur Product Theorem.

Theorem 3. *The Hadamard product of two positive-semidefinite matrices is positive-semidefinite. This is known as the Schur product theorem.*

Proof. Proof is done using eigendecomposition.
Let $M = \sum u_i\, m_i\, m_i^T$ and $N = \sum v_i\, n_i\, n_i^T$. Then $M * N = \sum_{i,j} u_i\, v_j\, (m_i\, m_i^T) * (n_j\, n_j^T) = \sum_{i,j} u_i\, v_j\, (m_i * n_j)\,(m_i * n_j)^T$. Each $(m_i * n_j)\,(m_i * n_j)^T$ is positive and $u_i\, v_j > 0$, thus the sum giving $M * N$ is also positive. □

4.4. Product Rule for Hadamard Product.

Theorem 4. *Suppose Y and X are $m-n$ matrices and each of whose elements is a differentiable function of all elements of a $p-n$ matrix Z. Then*

$$\frac{\partial (Y*X)}{\partial Z} = \frac{\partial Y}{\partial Z}(D_{\hat{x}}) + \frac{\partial X}{\partial Z}(D_{\hat{y}}). \tag{8}$$

Proof.

$$\frac{\partial Y}{\partial Z}(D_{\hat{x}}) + \frac{\partial X}{\partial Z}(D_{\hat{y}}) = [\sum_{ij} (\frac{\partial y_{ij}}{\partial z_{st}}\, x_{ij}) + \sum_{jk} (y_{jk}\, \frac{\partial x_{jk}}{\partial z_{st}})] \tag{9}$$

By the definition of D_x,

$$RHS = \frac{\partial (Y*X)}{\partial Z}. \tag{10}$$

□

4.5. Physical Meaning of Hadamard Product.

A very intuitive way to understand Hadamard product is to view it as a amplification of image processing. Here is a demonstration of Hadamard product that reduces the brightness of a picture. The matrix of original picture Hadamard-multiplies another matrix of exponential decay from the center.

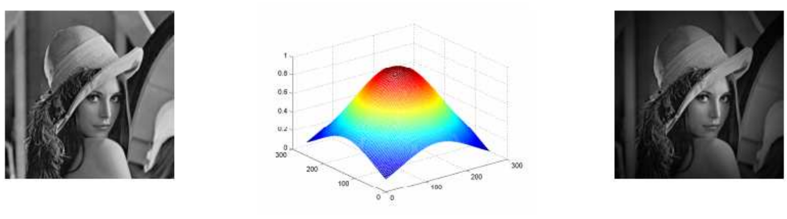

Figure 1.

4.6. Hadamard Matrix.

4.6.1. Introduction.

The Hadamard matrix H is a square with elements 1 or -1. It has several interesting properties.It is first studied by the same French mathematician.

$$H H^T = n I_n \tag{11}$$

A Hadamard matrix has maximal determinant among matrices with entries of absolute value less than or equal to 1. If M is a complex matrix of order n, whose entries are bounded by $M_{ij} \leq 1$, then $|det(M)| \leq n^{n/2}$. The Hadamard matrix can be constructed by Sylvester's construction.

4.6.2. Sylvester's construction.

If H is a Hadamard matrix, then the matrix below is also a Hadamard matrix. $\begin{pmatrix} H & H \\ H & -H \end{pmatrix}$ For example, $H_1 = [1]$, $H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. $H_{2^k} = \begin{pmatrix} H_{2^{k-1}} & H_{2^{k-1}} \\ H_{2^{k-1}} & H_{2^{k-1}} \end{pmatrix} = H_2 \otimes H_{2^{k-1}}$ where $\otimes$ denotes the Kronecker product. There is a famous conjecture about Hadamard matrix that there must exist the Hadamard matrix with order of $4k$ where k is a positive number. Up to now, people find very large Hadamard matrix but some small ones remain un discovered like the matrix of order of 668.

5. MATRIX DERIVATIVES WITH RULE FOR KRONECKER PRODUCTS

5.1. Main Result.

Theorem 5. (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*

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

5.2. Some Lemmas and Proofs.

Lemma 6. *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,*

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

□

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

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

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

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 8. *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*

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

where Y_i are rows of Y. By lemma 7,

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

On the left hand side,

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

□

Lemma 9. *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 8 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 6 and 8*

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

□

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 10. *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 8.* □

Lemma 11. *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 9.* □

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

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

Hence the proof of Theorem 5 is followed by Lemma 9 and 11.

5.3. An Example (by Jöreskog 1970).

Jöreskog shows how six classes of stochastic processes can be represented as covariance structure models whose parameters and structure can be estimated and tested.

We might consider short-term memory in a given context to be representable by one of Jöreskog 's model. Let this process have a covariance matrix Σ. Suppose now that long-term memory in the same context can be represented by a different stochastic structure whose covariance matrix is Σ_1. We propose that these two processes operate simultaneously, yielding a combined generalized stochastic structure $(\Sigma_1\otimes\Sigma)$. Such a model can be estimated and tested.

Here, we suppose that $\Sigma_1 = A$ is a known and a specific short-term process is postulated for Σ. Let $X(t)$ be the stochastic process with arbitatrary points $t_1 < t_2 < \ldots < t_p$. Following Jöreskog(1970), the Markov simplex can be defined by $E(X(t_j)) = \mu_j$, $\mathrm{var}(X(t_j)) = \sigma_j^2$ and $\mathrm{Cov}(X(t_j), X(t_k)) = \rho^{|t_j-t_k|}$, where $0<\rho<1$. Let $t_1^*=t_1, t_2^*=t_2-t_1, \ldots, t_p^*=t_p-t_{p-1}$. Then the covariance matrix of $x'=(X(t_1), \ldots, X(t_p))$ can be represented as $\Sigma=D_\sigma P D_\sigma$, where D_σ is the diagonal matrix of $\sigma_1,\ldots,\sigma_p$, and

$$P=\begin{pmatrix} 1 & \rho^{t_2^*} & \rho^{t_2^*+t_3^*} & \ldots & \rho^{t_2^*+t_3^*+\ldots+t_p^*}\\ \rho^{t_2^*} & 1 & \rho^{t_3^*} & \ldots & \rho^{t_3^*+\ldots+t_p^*}\\ \rho^{t_2^*+t_3^*} & \rho^{t_3^*} & 1 & \ldots & \ldots\\ \ldots & \ldots & \ldots & \ldots & \ldots\\ \rho^{t_2^*+t_3^*+\ldots+t_p^*} & \rho^{t_3^*+\ldots+t_p^*} & \ldots & \ldots & 1 \end{pmatrix}$$

There are $2p-1$ unknown parameters in Σ, namely $\rho^{t_2^*},\ldots,\rho^{t_p^*},\sigma_1,\ldots,\sigma_p$.

Now consider $\Sigma^*=(A\otimes\Sigma)$, where A is the diagonal matrix with elements $a_1, a_2, \ldots, a_n$, which represent the long-term decay structure that underlies repeated short-term process.

One problem is to estimate the parameter vector $\Theta = (\Theta_1, \Theta_2)$, where $\Theta_1 = \left(\rho^{t_2^*}, \ldots, \rho^{t_p^*}\right)$ and $\Theta_2 = (\sigma_1, \sigma_2, \ldots, \sigma_p)$. Let an abitrary fit function be $f(\Sigma^*(\Theta))$. Typically, one wants to estimate Θ by minimizing (or maximizing) f (i.e. maximal likehood method), and we need to calculate $\partial\Sigma^*/\partial\Theta$.

We denote $t_{i,j}^* = \sum_{k=i}^{j} t_k^*$ and $t_{i,j,r}^* = t_{i,j}^* - t_r^*$. Note that $\partial P/\partial\Theta_1$, is a matrix with rows $\left(\partial P/\partial\rho^{t_2^*}, \ldots, \partial P/\partial\rho^{t_p^*}\right)$. We see that $\partial P/\partial\rho^{t_i^*}$ is the row vector $[B_{i1}, \ldots, B_{ip}]$ for $2 \leqslant i \leqslant p$, where B_{ij} are $1 \times p$ row vector defined by

$$B_{ij} = \begin{cases} \left(0, \ldots, 0, \rho^{t_{j+1,i-1}^*}, \rho^{t_{j+1,i+1,i}^*}, \ldots, \rho^{t_{\mathrm{j}c+1,p,i}^*}\right) & \text{if } j < i-1 \\ \left(0, \ldots, 0, 1, \rho^{t_{i+1}^*}, \ldots, \rho^{t_{i+1,p}^*}\right) & \text{if } j = i-1 \\ \left(\rho^{t_{2,i-1}^*}, \rho^{t_{3,i-1}^*}, \rho^{t_{i-1}^*}, 1, 0, \ldots, 0\right) & \text{if } j = i \\ \left(\rho^{t_{2,j,i}^*}, \rho^{t_{3,j,i}^*}, \ldots, \rho^{t_{i+1,p}^*}, 0, \ldots, 0\right) & \text{if } j > i \end{cases}$$

From lemma 11,

$$\frac{\partial\Sigma^*}{\partial\Theta_1} = \frac{\partial(A \otimes E)}{\partial\Theta_1} = \left(\bar{A} \otimes \frac{\partial\Sigma}{\partial\Theta_1}\right)(I^n \otimes E^{np} \otimes I^p) = \left(\bar{A} \otimes \frac{\partial P}{\partial\Theta_1}(D_\sigma \otimes D_\sigma)\right)(I^n \otimes E^{np} \otimes I^p)$$

where $\partial P/\partial\Theta_1$ is given previously.

Obviously, $\partial\Sigma/\partial\Theta_2$ yields $K((I \otimes PD_\sigma) + (PD_\sigma \otimes I))$, where K is a matrix operator of zeros and ones which selects the appropriate rows of the subsequent matrix.

From lemma 6 we obtain,

$$\begin{aligned} \frac{\partial\Sigma^*}{\partial\Theta_2} &= \left(\bar{A} \otimes \frac{\partial\Sigma}{\partial\Theta_2}\right)(I^n \otimes E^{np} \otimes I^p) \\ &= (\bar{A} \otimes K((I \otimes PD_\sigma) + (PD_\sigma \otimes I)))(I^n \otimes E^{np} \otimes I^p) \end{aligned}$$

This completes $\partial\Sigma^*/\partial\Theta$.

BIBLIOGRAPHY

[1] BENTLER, P. and LEE, S.-Y. Matrix derivatives with chain rule and rules for simple, hadamard, and kronecker products. *Journal of Mathematical Psychology 17*, 3 (1978), p. 255–262.