



Munich Personal RePEc Archive

A note on matrix differentiation

Kowal, Pawel

December 2006

Online at <https://mpra.ub.uni-muenchen.de/3917/>

MPRA Paper No. 3917, posted 09 Jul 2007 UTC

A note on matrix differentiation

Paweł Kowal

July 9, 2007

Abstract

This paper presents a set of rules for matrix differentiation with respect to a vector of parameters, using the flattened representation of derivatives, i.e. in form of a matrix. We also introduce a new set of Kronecker tensor products of matrices. Finally we consider a problem of differentiating matrix determinant, trace and inverse.

JEL classification: C00

Keywords: matrix differentiation, generalized Kronecker products

1 Introduction

Derivatives of matrices with respect to a vector of parameters can be expressed as a concatenation of derivatives with respect to a scalar parameters. However such a representation of derivatives is very inconvenient in some applications, e.g. if higher order derivatives are considered, and or even are not applicable if matrix functions (like determinant or inverse) are present. For example finding an explicit derivative of $\det(\partial X/\partial \theta)$ would be a quite complicated task. Such a problem arise naturally in many applications, e.g. in maximum likelihood approach for estimating model parameters.

The same problems emerges in case of a tensor representation of derivatives. Additionally, in this case additional effort is required to find the flattened representation of resulting tensors, which is required, since running numerical computations efficiently is possible only in case of two dimensional data structures.

In this paper we derive formulas for differentiating matrices with respect to a vector of parameters, when one requires the flattened form of resulting derivatives, i.e. representation of derivatives in form of matrices. To do this

we introduce a new set of the Kronecker matrix products as well as the generalized matrix transposition. Then, first order and higher order derivatives of functions being compositions of primitive function using elementary matrix operations like summation, multiplication, transposition and the Kronecker product, can be expressed in a closed form based on primitive matrix functions and their derivatives, using these elementary operations, the generalized Kronecker products and the generalized transpositions.

We consider also more general matrix functions containing matrix functions (inverse, trace and determinant). Defining the generalized trace function we are able to express derivatives of such functions in closed form.

2 Matrix differentiation rules

Let us consider smooth functions $\Omega \ni \theta \mapsto X(\theta) \in \mathbb{R}^{m \times n}$, $\Omega \ni \theta \mapsto Y(\theta) \in \mathbb{R}^{p \times q}$, where $\Omega \subset \mathbb{R}^k$ is an open set. Functions X, Y associate a $m \times n$ and $p \times q$ matrix for a given vector of parameters, $\theta = \text{col}(\theta_1, \theta_2, \dots, \theta_k)$. Let the differential of the function X with respect to θ is defined as

$$\frac{\partial X}{\partial \theta} = \left[\begin{array}{cccc} \frac{\partial X}{\partial \theta_1} & \frac{\partial X}{\partial \theta_2} & \cdots & \frac{\partial X}{\partial \theta_k} \end{array} \right]$$

for $\partial X / \partial \theta_i \in \mathbb{R}^{m \times n}$, $i = 1, 2, \dots, k$.

Proposition 2.1. *The following equations hold*

1. $\frac{\partial}{\partial \theta}(\alpha X) = \alpha \frac{\partial X}{\partial \theta}$
2. $\frac{\partial}{\partial \theta}(X + Y) = \frac{\partial X}{\partial \theta} + \frac{\partial Y}{\partial \theta}$
3. $\frac{\partial}{\partial \theta}(X \times Y) = \frac{\partial X}{\partial \theta} \times (I_k \otimes Y) + X \times \frac{\partial Y}{\partial \theta}$

where $\alpha \in \mathbb{R}$ and I_k is a $k \times k$ dimensional identity matrix, assuming that differentials exist and matrix dimensions coincide.

Proof. The first two cases are obvious. We have

$$\begin{aligned} \frac{\partial}{\partial \theta}(X \times Y) &= \left[\begin{array}{cccc} \frac{\partial X}{\partial \theta_1} \times Y + X \times \frac{\partial Y}{\partial \theta_1} & \cdots & \frac{\partial X}{\partial \theta_k} \times Y + X \times \frac{\partial Y}{\partial \theta_k} \end{array} \right] \\ &= \left[\begin{array}{ccc} \frac{\partial X}{\partial \theta_1} & \cdots & \frac{\partial X}{\partial \theta_k} \end{array} \right] \times \left[\begin{array}{ccc} Y & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Y \end{array} \right] + X \times \left[\begin{array}{ccc} \frac{\partial Y}{\partial \theta_1} & \cdots & \frac{\partial Y}{\partial \theta_k} \end{array} \right] \\ &= \frac{\partial X}{\partial \theta} \times (I_k \otimes Y) + X \times \frac{\partial Y}{\partial \theta} \end{aligned}$$

□

Differentiating matrix transposition is a little bit more complicated. Let us define a generalized matrix transposition

Definition 2.2. Let $X = [X_1, X_2, \dots, X_n]$, where $X_i \in \mathbb{R}^{p \times q}$, $i = 1, 2, \dots, n$ is a $p \times q$ matrix is a partition of $p \times nq$ dimensional matrix X . Then

$$\mathbb{T}_n(X) \doteq [X'_1, X'_2, \dots, X'_n]$$

Proposition 2.3. The following equations hold

1. $\frac{\partial}{\partial \theta}(X') = \mathbb{T}_k(\frac{\partial X}{\partial \theta})$
2. $\frac{\partial}{\partial \theta}(\mathbb{T}_n(X)) = \mathbb{T}_{k \times n}(\frac{\partial X}{\partial \theta})$

Proof. The first condition is a special case of the second condition for $n = 1$. We have

$$\begin{aligned} \frac{\partial}{\partial \theta}(\mathbb{T}_{(n)}(X)) &= [\mathbb{T}_{(n)}(\frac{\partial X}{\partial \theta_1}) \quad \dots \quad \mathbb{T}_{(n)}(\frac{\partial X}{\partial \theta_k})] \\ &= [\frac{\partial X'_1}{\partial \theta_1}, \dots, \frac{\partial X'_n}{\partial \theta_1} \quad \dots \quad \frac{\partial X'_1}{\partial \theta_k}, \dots, \frac{\partial X'_n}{\partial \theta_k}] = \mathbb{T}_{(k \times n)}\left(\frac{\partial X}{\partial \theta}\right) \end{aligned}$$

since

$$\frac{\partial X}{\partial \theta} = [\frac{\partial X_1}{\partial \theta_1}, \dots, \frac{\partial X_n}{\partial \theta_1} \quad \dots \quad \frac{\partial X_1}{\partial \theta_k}, \dots, \frac{\partial X_n}{\partial \theta_k}]$$

□

Let us now turn to differentiating tensor products of matrices. Let for any matrices X, Y , where $X \in \mathbb{R}^{p \times q}$ is a matrix with elements $x_{ij} \in \mathbb{R}$ for $i = 1, 2, \dots, p$, $j = 1, 2, \dots, q$. The Kronecker product, $X \otimes Y$ is defined as

$$X \otimes Y \doteq \begin{bmatrix} x_{11}Y & \dots & x_{1q}Y \\ \vdots & \ddots & \vdots \\ x_{p1}Y & \dots & x_{pq}Y \end{bmatrix}$$

Similarly as in case of differentiating matrix transposition we need to introduce the generalized Kronecker product

Definition 2.4. Let $X = [X_1, X_2, \dots, X_m]$, where $X_i \in \mathbb{R}^{p \times q}$, $i = 1, 2, \dots, m$ is a $p \times q$ matrix is a partition of $p \times mq$ dimensional matrix X . Let $Y = [Y_1, Y_2, \dots, Y_n]$, where $Y_i \in \mathbb{R}^{r \times s}$, $i = 1, 2, \dots, n$ is a $r \times s$ matrix is a partition of $r \times ns$ dimensional matrix Y . Then

$$\begin{aligned} X \otimes_n^1 Y &\doteq [X \otimes Y_1, \dots, X \otimes Y_n] \\ X \otimes_n^m Y &\doteq [X_1 \otimes_n^1 Y, \dots, X_m \otimes_n^1 Y] \\ X \otimes_{n_1, n_2, \dots, n_s}^{1, m_2, \dots, m_s} Y &\doteq [X \otimes_{n_2, \dots, n_s}^{m_2, \dots, m_s} Y_1, \dots, X \otimes_{n_2, \dots, n_s}^{m_2, \dots, m_s} Y_{n_1}] \\ X \otimes_{n_1, n_2, \dots, n_s}^{m_1, m_2, \dots, m_s} Y &\doteq [X_1 \otimes_{n_1, n_2, \dots, n_s}^{1, m_2, \dots, m_s} Y, \dots, X_{m_1} \otimes_{n_1, n_2, \dots, n_s}^{1, m_2, \dots, m_s} Y] \end{aligned}$$

assuming that appropriate matrix partitions exist.

Proposition 2.5. *The following equations hold*

1. $\frac{\partial}{\partial \theta}(X \otimes Y) = \frac{\partial X}{\partial \theta} \otimes Y + X \otimes_k^1 \frac{\partial Y}{\partial \theta}$
2. $\frac{\partial}{\partial \theta}(X \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} Y) = \frac{\partial X}{\partial \theta} \otimes_{1, n_1, \dots, n_s}^{k, m_1, \dots, m_s} Y + X \otimes_{k, n_1, \dots, n_s}^{1, m_1, \dots, m_s} \frac{\partial Y}{\partial \theta}$

Proof. We have

$$\begin{aligned} \frac{\partial}{\partial \theta}(X \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} Y) &= \left[\frac{\partial}{\partial \theta_1}(X \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} Y) \quad \dots \quad \frac{\partial}{\partial \theta_k}(X \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} Y) \right] \\ &= \left[\frac{\partial X}{\partial \theta_1} \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} Y \quad \dots \quad \frac{\partial X}{\partial \theta_k} \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} Y \right] \\ &+ \left[X \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} \frac{\partial Y}{\partial \theta_1} \quad \dots \quad X \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} \frac{\partial Y}{\partial \theta_k} \right] \\ &= \frac{\partial X}{\partial \theta} \otimes_{1, n_1, \dots, n_s}^{k, m_1, \dots, m_s} Y + X \otimes_{k, n_1, \dots, n_s}^{1, m_1, \dots, m_s} \frac{\partial Y}{\partial \theta} \end{aligned}$$

Since $X \otimes Y = X \otimes_1^1 Y$, in case of the standard Kronecker product we obtain

$$\frac{\partial}{\partial \theta}(X \otimes Y) = \frac{\partial X}{\partial \theta} \otimes_1^k Y + X \otimes_k^1 \frac{\partial Y}{\partial \theta} = \frac{\partial X}{\partial \theta} \otimes Y + X \otimes_k^1 \frac{\partial Y}{\partial \theta}$$

□

In proposition 2.1 we have omitted the case of multiplication of a matrix by a scalar function, using proposition 2.5 we obtain

Proposition 2.6. *Let α is a scalar function of θ and X is a matrix valued function of θ , $X(\theta) \in \mathbb{R}^{p \times q}$. Then*

$$\frac{\partial}{\partial \theta}(\alpha X) = \alpha \times \frac{\partial X}{\partial \theta} + \frac{\partial \alpha}{\partial \theta} \otimes X$$

Proof. Expression αX can be represented as $\alpha X = (\alpha \otimes I_p) \times X$, where I_p is a $p \times p$ dimensional identity matrix. Hence

$$\begin{aligned} \frac{\partial}{\partial \theta}(\alpha X) &= \frac{\partial}{\partial \theta}((\alpha \otimes I_p) \times X) = \frac{\partial(\alpha \otimes I_p)}{\partial \theta} \times (I_k \otimes X) + (\alpha \otimes I_p) \times \frac{\partial X}{\partial \theta} \\ &= \left(\frac{\partial \alpha}{\partial \theta} \otimes I_p\right) \times (I_k \otimes X) + \alpha \times \frac{\partial X}{\partial \theta} = \frac{\partial \alpha}{\partial \theta} \otimes X + \alpha \times \frac{\partial X}{\partial \theta} \end{aligned}$$

□

Let \mathcal{S} is a set of smooth matrix valued functions $\Omega \ni \theta \mapsto X(\theta) \in \mathbb{R}^{p \times q}$, where $\Omega \subset \mathbb{R}^k$ is an open set, for any integers $p, q \geq 1$ not necessary the same for all functions in \mathcal{S} . Let $\text{dif } \mathcal{S} \doteq \{\partial X / \partial \theta : X \in \mathcal{S}\}$. The set \mathcal{S} may contain scalars and matrices, which are interpreted as constant functions.

Let $\text{ext}(\mathcal{S})$ is a set of functions obtained by applying elementary matrix operations on the set \mathcal{S} , i.e. $\text{ext}(\mathcal{S})$ is a smallest set such that if $X, Y \in \text{ext}(\mathcal{S})$, then matrix valued functions $X + Y$, $X \times Y$, $T_n(X)$, $X \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} Y$ if exist belong to $\text{ext}(\mathcal{S})$, where $n, n_1, \dots, n_s, m_1, \dots, m_s$ are any positive integers.

Theorem 2.7. $\text{dif}(\text{ext}(\mathcal{S})) = \text{ext}(\mathcal{S} \cup \text{dif}(\mathcal{S}))$.

Proof. By induction using propositions 2.1, 2.3, 2.5, 2.6. □

The theorem 2.7 states, that derivatives of matrix valued functions obtained by applying elementary operations like summation, matrix multiplication, generalized transposition and generalized Kronecker tensor product can be expressed as a combination of these functions and their derivatives using these elementary operations. Applying the theorem 2.7 to a set $\mathcal{T} = \text{dif}(\text{ext}(\mathcal{S}))$ we can see that also higher order derivatives can be expressed, using these elementary operations, as combinations of elementary functions \mathcal{S} and their higher order derivatives.

3 Derivatives of matrix determinant, trace and inverse

Let us consider derivatives of matrix inverse, determinant and trace. We need to introduce the generalized trace defined analogously as the generalized transposition.

Definition 3.1. Let $X = [X_1, X_2, \dots, X_n]$, where $X_i \in \mathbb{R}^{p \times p}$, $i = 1, 2, \dots, n$ is a $p \times p$ matrix, is a partition of $p \times np$ dimensional matrix X . Then

$$\text{tr}_n(X) \doteq [\text{tr } X_1, \text{tr } X_2, \dots, \text{tr } X_n]$$

Proposition 3.2. The following equations hold

1. $\frac{\partial \det(X)}{\partial \theta} = \det(X) \times \text{tr}_k(X^{-1} \times \frac{\partial X}{\partial \theta})$
2. $\frac{\partial \text{tr}_n(X)}{\partial \theta} = \text{tr}_{k \times n}(\frac{\partial X}{\partial \theta})$
3. $\frac{\partial X^{-1}}{\partial \theta} = -X^{-1} \times \frac{\partial X}{\partial \theta} \times (I_k \otimes X^{-1})$

Proof. We have

$$\begin{aligned}\frac{\partial \det(X)}{\partial \theta} &= \left[\frac{\partial \det(X)}{\partial \theta_1} \quad \dots \quad \frac{\partial \det(X)}{\partial \theta_k} \right] \\ &= \left[\det(X) \operatorname{tr}(X^{-1} \times \frac{\partial X}{\partial \theta_1}) \quad \dots \quad \det(X) \times \operatorname{tr}(X^{-1} \times \frac{\partial X}{\partial \theta_k}) \right] \\ &= \det(X) \times \operatorname{tr}_k(X^{-1} \times \frac{\partial X}{\partial \theta}) \\ \frac{\partial \operatorname{tr}_n(X)}{\partial \theta} &= \left[\frac{\partial \operatorname{tr}_n(X)}{\partial \theta_1} \quad \dots \quad \frac{\partial \operatorname{tr}_n(X)}{\partial \theta_k} \right] = \left[\operatorname{tr}_n(\frac{\partial X}{\partial \theta_1}) \quad \dots \quad \operatorname{tr}_n(\frac{\partial X}{\partial \theta_k}) \right] = \operatorname{tr}_{k \times n}(\frac{\partial X}{\partial \theta})\end{aligned}$$

Similarly

$$\begin{aligned}\frac{\partial X^{-1}}{\partial \theta} &= \left[\frac{\partial X^{-1}}{\partial \theta_1} \quad \dots \quad \frac{\partial X^{-1}}{\partial \theta_k} \right] = - \left[X^{-1} \frac{\partial X}{\partial \theta_1} X^{-1} \quad \dots \quad X^{-1} \frac{\partial X}{\partial \theta_k} X^{-1} \right] \\ &= -X^{-1} \times \left[\frac{\partial X}{\partial \theta_1} \quad \dots \quad \frac{\partial X}{\partial \theta_k} \right] \times (I_k \otimes X^{-1}) = -X^{-1} \frac{\partial X}{\partial \theta} (I_k \otimes X^{-1})\end{aligned}$$

since in case of scalar parameter $\theta \in \mathbb{R}$, $\partial \det(X)/\partial \theta = \det(X) \operatorname{tr}(X^{-1} \partial X/\partial \theta)$, $\partial \operatorname{tr}(X)/\partial \theta = \operatorname{tr}(\partial X/\partial \theta)$, and $\partial X^{-1}/\partial \theta = -X^{-1}(\partial X/\partial \theta)X^{-1}$ (see for example Petersen, Petersen, (2006)). \square

Let a set \mathcal{S} and operation dif are defined as in the previous section. Let $\operatorname{ext}_2(\mathcal{S})$ is a set of functions obtained by applying elementary matrix operations and matrix determinant, trace and inverse on the set \mathcal{S} , i.e. $\operatorname{ext}(\mathcal{S})$ is a smallest set such that if $X, Y \in \operatorname{ext}_2(\mathcal{S})$, then matrix valued functions $X + Y$, $X \times Y$, $\operatorname{T}_n(X)$, $X \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} Y$, $\det(X)$, $\operatorname{tr}_n(X)$, X^{-1} if exist belong to $\operatorname{ext}_2(\mathcal{S})$, where $n, n_1, \dots, n_s, m_1, \dots, m_s$ are any positive integers.

Theorem 3.3. $\operatorname{dif}(\operatorname{ext}_2(\mathcal{S})) = \operatorname{ext}_2(\mathcal{S} \cup \operatorname{dif}(\mathcal{S}))$.

Proof. By induction using propositions 2.1, 2.3, 2.5, 2.6, 3.2. \square

4 Derivatives of function composition

Let f is a matrix valued function given by

$$R^p \ni x \mapsto f(x) \in R^{m \times n}$$

and g is a vector valued function $\Omega \ni \theta \mapsto g(\theta) \in R^p$. We can define a function composition $\Omega \ni \theta \mapsto f(g(\theta)) \in R^{m \times n}$.

Proposition 4.1. *The following condition holds*

$$\frac{\partial}{\partial \theta} f(g(\theta)) = \frac{\partial f(g(\theta))}{\partial x} \times \left(\frac{\partial g(\theta)}{\partial \theta} \otimes I_n \right)$$

Proof. Let

$$f(x) = \begin{bmatrix} f_{11}(x) & \dots & f_{1n}(x) \\ \vdots & \ddots & \vdots \\ f_{m1}(x) & \dots & f_{mn}(x) \end{bmatrix}$$

where $f_{ij}(x)$ are scalar valued functions. Then for $s = 1, \dots, k$

$$\frac{\partial f_{ij}(x)}{\partial \theta_s} = \sum_{k=1}^p \frac{\partial f_{ij}(x)}{\partial x_k} \times \frac{\partial x_k}{\partial \theta_s} = \frac{\partial f_{ij}(x)}{\partial x} \times \frac{\partial x}{\partial \theta_s}$$

since $\partial x / \partial \theta_s$ is a column vector. Further

$$\begin{aligned} \frac{f(x)}{\partial \theta_s} &= \sum_{k=1}^p \begin{bmatrix} \frac{\partial f_{11}(x)}{\partial x_k} & \dots & \frac{\partial f_{1n}(x)}{\partial x_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{m1}(x)}{\partial x_k} & \dots & \frac{\partial f_{mn}(x)}{\partial x_k} \end{bmatrix} \times \frac{\partial x_k}{\partial \theta_s} = \sum_{k=1}^p \frac{\partial f(x)}{\partial x_k} \times \frac{\partial x_k}{\partial \theta_s} \\ &= \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} & \dots & \frac{\partial f(x)}{\partial x_p} \end{bmatrix} \times \begin{bmatrix} I_n \times \frac{\partial x_1}{\partial \theta_s} \\ \vdots \\ I_n \times \frac{\partial x_p}{\partial \theta_s} \end{bmatrix} = \frac{\partial f(x)}{\partial x} \times \left(\frac{\partial x}{\partial \theta_s} \otimes I_n \right) \end{aligned}$$

Finally

$$\frac{f(x)}{\partial \theta_s} = \frac{\partial f(x)}{\partial x} \times \left[\frac{\partial x}{\partial \theta_1} \otimes I_n \quad \dots \quad \frac{\partial x}{\partial \theta_k} \otimes I_n \right] = \frac{\partial f(x)}{\partial x} \times \left(\frac{\partial x}{\partial \theta} \otimes I_n \right)$$

□

5 Properties of the generalized Kronecker product

Proposition 5.1. *For any matrices A, B*

1. $A \otimes_1^k B = A \otimes B$.
2. $A \otimes_{\dots, n_k, 1, \dots}^{\dots, 1, 1, \dots} B = A \otimes_{\dots, n_k, \dots}^{\dots, m_k, \dots} B$.
3. $A \otimes_{\dots, n_k, n_{k+1}, \dots}^{\dots, 1, 1, \dots} B = A \otimes_{\dots, n_k \times n_{k+1}, \dots}^{\dots, 1, \dots} B$.
4. $A \otimes_{\dots, 1, n_{k+1}, \dots}^{\dots, m_k, m_{k+1}, \dots} B = A \otimes_{\dots, n_{k+1}, \dots}^{\dots, m_k \times m_{k+1}, \dots} B$.

assuming that the Kronecker products exist.

Proposition 5.2. For any matrices A, B, C

1. $A \otimes_{n_1, \dots, n_k}^{m_1, \dots, m_k} (B + C) = A \otimes_{n_1, \dots, n_k}^{m_1, \dots, m_k} B + A \otimes_{n_1, \dots, n_k}^{m_1, \dots, m_k} C.$
2. $(A + B) \otimes_{n_1, \dots, n_k}^{m_1, \dots, m_k} C = A \otimes_{n_1, \dots, n_k}^{m_1, \dots, m_k} C + B \otimes_{n_1, \dots, n_k}^{m_1, \dots, m_k} C.$

assuming that the Kronecker products exist and matrix dimensions coincide.

Proposition 5.3. For any matrices A, B, C, D

$$(AB) \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} (CD) = (A \otimes C) \times (B \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} D)$$

assuming that products AB and CD , as well as Kronecker products exist.

Proof. Observe that $X \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} Y = X \otimes_{n_1, \dots, n_s, 1}^{m_1, \dots, m_s, 1} Y$, and $(AB) \otimes_1^1 (CD) = (A \otimes C) \times (B \otimes_1^1 D)$, since $(AB) \otimes (CD) = (A \otimes C) \times (B \otimes D)$. Let $(AB) \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} (CD) = (A \otimes C) \times (B \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} D)$ for $k \geq 0$. Then

$$\begin{aligned} & (AB) \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} (CD) \\ &= \left[(AB) \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} (CD_1) \quad \dots, (AB) \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} (CD_{n_{k+1}}) \right] \\ &= \left[(A \otimes C)(B \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} D_1) \quad \dots, (A \otimes C)(B \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} D_{n_{k+1}}) \right] \\ &= (A \otimes C) \times (B \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} D) \end{aligned}$$

Similarly

$$\begin{aligned} & (AB) \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{m_{k+1}, m_k, \dots, m_1, 1} (CD) \\ &= \left[(AB_1) \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} (CD) \quad \dots, (AB_{m_{k+1}}) \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} (CD) \right] \\ &= \left[(A \otimes C)(B_1 \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} D) \quad \dots, (A \otimes C)(B_{m_{k+1}} \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} D) \right] \\ &= (A \otimes C) \times (B \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{m_{k+1}, m_k, \dots, m_1, 1} D) \end{aligned}$$

□

Proposition 5.4. For any matrices A, B of size $p_1 \times q_1$ and $p_2 \times q_2$

$$A \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} B = (A \otimes B) \times (I_{q_1} \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} I_{q_2})$$

assuming that Kronecker product exists.

Proposition 5.5. For any matrices A, B, C

$$A \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} (B \otimes C) = (A \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} B) \otimes C$$

assuming that Kronecker products exist.

Proof. Observe that $X \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} Y = X \otimes_{n_1, \dots, n_s, 1}^{m_1, \dots, m_s, 1} Y$, and $A \otimes_1^1 (B \otimes C) = (A \otimes_1^1 B) \otimes C$, since $A \otimes (B \otimes C) = (A \otimes B) \otimes C$. Let $A \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} (B \otimes C) = (A \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} B) \otimes C$ for $k \geq 0$. Then

$$\begin{aligned} & A \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} (B \otimes C) \\ &= \left[A \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} (B_1 \otimes C) \quad \dots, A \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} (B_{n_k} \otimes C) \right] \\ &= \left[(A \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} B_1) \otimes C \quad \dots, (A \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} B_{n_k}) \otimes C \right] \\ &= (A \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} B) \otimes C \end{aligned}$$

Similarly

$$\begin{aligned} & A \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{m_{k+1}, m_k, \dots, m_1, 1} (B \otimes C) \\ &= \left[A_1 \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} (B \otimes C) \quad \dots, A_{m_{k+1}} \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} (B \otimes C) \right] \\ &= \left[(A_1 \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} B) \otimes C \quad \dots, (A_{m_{k+1}} \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} B) \otimes C \right] \\ &= (A \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{m_{k+1}, m_k, \dots, m_1, 1} B) \otimes C \end{aligned}$$

□

Proposition 5.6. For any matrices A, B, C

$$A \otimes (B \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} C) = (A \otimes B) \otimes_{1, n_1, \dots, n_s}^{q, m_1, \dots, m_s} C$$

where q is a number of columns of the matrix A , assuming that Kronecker products exist.

Proof. Observe that $X \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} Y = X \otimes_{n_1, \dots, n_s, 1}^{m_1, \dots, m_s, 1} Y$, and $A \otimes (B \otimes_1^1 C) = (A \otimes B) \otimes_1^1 C = (A \otimes B) \otimes_{1, 1}^{q, 1} C$, since $A \otimes (B \otimes C) = (A \otimes B) \otimes C$ and $A \otimes_1^q C = A \otimes C$ if the Kronecker product exists. Let $A \otimes (B \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} C) = (A \otimes B) \otimes_{1, n_k, \dots, n_1, 1}^{q, m_k, \dots, m_1, 1} C$ for $k \geq 0$. Let $A = [A_1, \dots, A_q]$. Then

$$\begin{aligned} & A_i \otimes (B \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} C) \\ &= \left[A_i \otimes (B \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} C_1) \quad \dots, A_i \otimes (B \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} C_{n_k}) \right] \\ &= \left[(A_i \otimes B) \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} C_1 \quad \dots, (A_i \otimes B) \otimes_{n_k, \dots, n_1, 1}^{m_k, \dots, m_1, 1} C_{n_k} \right] \\ &= (A_i \otimes B) \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} C \end{aligned}$$

Similarly

$$\begin{aligned} & A_i \otimes (B \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{m_{k+1}, m_k, \dots, m_1, 1} C) \\ &= \left[A_i \otimes (B_1 \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} C) \quad \dots, A_i \otimes (B_{m_{k+1}} \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} C) \right] \\ &= \left[(A_i \otimes B_1) \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} C \quad \dots, (A_i \otimes B_{m_{k+1}}) \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{1, m_k, \dots, m_1, 1} C \right] \\ &= (A_i \otimes B) \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{m_{k+1}, m_k, \dots, m_1, 1} C \end{aligned}$$

Finally,

$$\begin{aligned}
& A \otimes (B \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{m_{k+1}, m_k, \dots, m_1, 1} C) \\
&= \left[A_1 \otimes (B \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{m_{k+1}, m_k, \dots, m_1, 1} C) \quad \dots \quad A_q \otimes (B \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{m_{k+1}, m_k, \dots, m_1, 1} C) \right] \\
&= \left[(A_1 \otimes B) \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{m_{k+1}, m_k, \dots, m_1, 1} C \quad \dots \quad (A_q \otimes B) \otimes_{n_{k+1}, n_k, \dots, n_1, 1}^{m_{k+1}, m_k, \dots, m_1, 1} C \right] \\
&= (A \otimes B) \otimes_{1, n_{k+1}, n_k, \dots, n_1, 1}^{q, m_{k+1}, m_k, \dots, m_1, 1} C
\end{aligned}$$

□

Proposition 5.7. *Let A is $m \times n$ matrix. Let B is $p \times q$ matrix. Then*

$$A \otimes_q^1 B = (I_m \otimes_p^1 I_p) \times (B \otimes A)$$

Proof. Let A^i is i -th column of A and B^j is j -th column of B . Let I_p^k denotes k -th column of $p \times p$ identity matrix and let B_j^i denotes element of B at j -th row and i -th column. Then

$$\begin{aligned}
(I_m \otimes_p^1 I_p) \times (B^j \otimes A^i) &= \left[I_m \otimes I_p^1 \quad \dots \quad I_m \otimes I_p^p \right] \times \begin{bmatrix} B_1^j A^i \\ \dots \\ B_p^j A^i \end{bmatrix} \\
&= \sum_{r=1}^p (I_m \otimes I_p^r) \times (A^i \otimes B_r^j) = \sum_{r=1}^p A^i \otimes (I_p^r \times B_r^j) = A^i \otimes B^j
\end{aligned}$$

Further

$$\begin{aligned}
(I_m \otimes_p^1 I_p) \times (B^j \otimes A) &= (I_m \otimes_p^1 I_p) \times \left[B^j \otimes A^1 \quad \dots \quad B^j \otimes A^n \right] \\
&= \left[A^1 \otimes B^j \quad \dots \quad A^n \otimes B^j \right] = A \otimes B^j \\
(I_m \otimes_p^1 I_p) \times (B \otimes A) &= (I_m \otimes_p^1 I_p) \times \left[B^1 \otimes A \quad \dots \quad B^q \otimes A \right] \\
&= \left[A \otimes B^1 \quad \dots \quad A \otimes B^q \right] = A \otimes_q^1 B
\end{aligned}$$

□

Proposition 5.8. *Let A is $m \times n$ matrix. Let B is $p \times q$ matrix. Then*

$$A \otimes_{m_1, 1, \dots, m_s, 1, q/\bar{m}}^{1, n_1, \dots, 1, n_s, 1} B = (I_m \otimes_p^1 I_p) \times (B \otimes_{n_1, \dots, n_s}^{m_1, \dots, m_s} A)$$

where $\bar{m} = m_1 \times \dots \times m_s$, assuming that the Kronecker products exist.

Proof. Proposition holds for $s = 0$. Let for any $s \geq 0$ $A \otimes_{m_s, 1, \dots, m_1, 1, q/\bar{m}_s}^{1, n_s, \dots, 1, n_1} B = (I_m \otimes_p^1 I_p) \times (B \otimes_{n_s, \dots, n_1}^{m_s, \dots, m_1} A)$, where $\bar{m}_s = m_1 \times \dots \times m_s$. Then

$$\begin{aligned} & (I_m \otimes_p^1 I_p) \times (B \otimes_{n_{s+1}, n_s, \dots, n_1}^{1, m_s, \dots, m_1} A) \\ &= (I_m \otimes_p^1 I_p) \times \left[B \otimes_{n_s, \dots, n_1}^{m_s, \dots, m_1} A_1 \quad \dots \quad B \otimes_{n_s, \dots, n_1}^{m_s, \dots, m_1} A_{n_{s+1}} \right] \\ &= \left[A_1 \otimes_{m_s, 1, \dots, m_1, 1, q/\bar{m}_s}^{1, n_s, \dots, 1, n_1, 1} B \quad \dots \quad A_{n_{s+1}} \otimes_{m_s, \dots, m_1, q/\bar{m}_s}^{n_s, \dots, n_1, 1} B \right] \\ &= A \otimes_{1, 1, m_s, 1, \dots, m_1, 1, q/\bar{m}_s}^{1, n_{s+1}, 1, n_s, \dots, 1, n_1, 1} B \end{aligned}$$

and

$$\begin{aligned} & (I_m \otimes_p^1 I_p) \times (B \otimes_{n_{s+1}, n_s, \dots, n_1}^{m_{s+1}, m_s, \dots, m_1} A) \\ &= (I_m \otimes_p^1 I_p) \times \left[B_1 \otimes_{n_{s+1}, n_s, \dots, n_1}^{1, m_s, \dots, m_1} A \quad \dots \quad B_{m_{s+1}} \otimes_{n_{s+1}, n_s, \dots, n_1}^{1, m_s, \dots, m_1} A \right] \\ &= \left[A \otimes_{1, 1, m_s, 1, \dots, m_1, 1, q/\bar{m}_{s+1}}^{1, n_{s+1}, 1, n_s, \dots, 1, n_1, 1} B_1 \quad \dots \quad A \otimes_{1, 1, m_s, 1, \dots, m_1, 1, q/\bar{m}_{s+1}}^{1, n_{s+1}, 1, n_s, \dots, 1, n_1, 1} B_{m_{s+1}} \right] \\ &= A \otimes_{m_{s+1}, 1, m_s, 1, \dots, m_1, 1, q/\bar{m}_{s+1}}^{1, n_{s+1}, 1, n_s, \dots, 1, n_1, 1} B \end{aligned}$$

□

Proposition 5.9.

$$(I_m \otimes_q^1 I_q)^{-1} = I_q \otimes_m^1 I_m$$

Proof. Observe that $(I_m \otimes_q^1 I_q)$ is an orthogonal matrix since this matrix can be obtained permuting columns of the matrix I_{mq} . Hence $(I_m \otimes_q^1 I_q)^{-1} = (I_m \otimes_q^1 I_q)^T$. Further

$$(I_m \otimes_q^1 I_q)^T = \begin{bmatrix} I_m \otimes (I_q^1)^T \\ \dots \\ I_m \otimes (I_q^q)^T \end{bmatrix} = \begin{bmatrix} (I_q^1)^T \otimes_m^1 I_m \\ \dots \\ (I_q^q)^T \otimes_m^1 I_m \end{bmatrix} = I_q \otimes_m^1 I_m$$

The second equality can be shown using for example proposition 5.7. □

Proposition 5.10.

$$(I_m \otimes_n^k I_q)^{-1} = (I_{nm} \otimes_{q/n}^{nk} I_{q/n}) \times (I_{kq} \otimes_{m/k}^k I_{m/k})$$

assuming that the Kronecker product exists.

Proof. Observe that $(I_m \otimes_n^1 I_q)$ is an orthogonal matrix since this matrix can be obtained permuting columns of the matrix I_{mq} . Hence $(I_m \otimes_n^1 I_q)^{-1} =$

$(I_m \otimes_n^1 I_q)^T$. Further

$$\begin{aligned} I_n \otimes (I_{q/n} \otimes_m^1 I_m) \times (I_m \otimes_n^1 I_q)^T &= \begin{bmatrix} (I_{q/n} \otimes_m^1 I_m) \times I_m \otimes (I_q^1)^T \\ \dots \\ (I_{q/n} \otimes_m^1 I_m) \times I_m \otimes (I_q^n)^T \end{bmatrix} \\ &= \begin{bmatrix} (I_q^1)^T \otimes_m^1 I_m \\ \dots \\ (I_q^n)^T \otimes_m^1 I_m \end{bmatrix} = I_q \otimes_m^1 I_m \end{aligned}$$

The second equality can be shown using for example proposition 5.7. Hence

$$\begin{aligned} (I_m \otimes_n^1 I_q)^{-1} &= \left(I_n \otimes (I_{q/n} \otimes_m^1 I_m)^{-1} \right) \times (I_q \otimes_m^1 I_m) \\ &= \left(I_n \otimes (I_m \otimes_{q/n}^1 I_{q/n}) \right) \times (I_q \otimes_m^1 I_m) \\ &= (I_{nm} \otimes_{q/n}^n I_{q/n}) \times (I_q \otimes_m^1 I_m) \end{aligned}$$

Further

$$\begin{aligned} (I_m \otimes_n^k I_q)^{-1} &= \left((I_k \otimes I_{m/k}) \otimes_{1,n}^{k,1} I_q \right)^{-1} = \left(I_k \otimes (I_{m/k} \otimes_n^1 I_q) \right)^{-1} \\ &= I_k \otimes \left((I_{nm/k} \otimes_{q/n}^n I_{q/n}) \times (I_q \otimes_{m/k}^1 I_{m/k}) \right) \\ &= I_k \otimes (I_{nm/k} \otimes_{q/n}^n I_{q/n}) \times I_k \otimes (I_q \otimes_{m/k}^1 I_{m/k}) \\ &= (I_{nm} \otimes_{q/n}^{nk} I_{q/n}) \times (I_{kq} \otimes_{m/k}^k I_{m/k}) \end{aligned}$$

□

Proposition 5.11. *Let A is $m \times n$ matrix. Let B is $p \times q$ matrix. Then*

$$A \otimes B = (I_m \otimes_p^1 I_p) \times (B \otimes A) \times (I_q \otimes_n^1 I_n)$$

Proof.

$$\begin{aligned} (I_m \otimes_p^1 I_p) \times (B \otimes A) &= A \otimes_q^1 B = (A \otimes B) \times (I_n \otimes_q^1 I_q) \\ &= (A \otimes B) \times (I_q \otimes_n^1 I_n)^{-1} \end{aligned}$$

□

6 Concluding remarks

Derived formulas requires matrix tensor products, which are absent, when representing derivatives as the concatenation of derivatives with respect to a scalar parameters. Hence, this approach may decrease numerical efficiency. This problem however can be resolved using appropriate data structures.

References

- [1] T.P. Minka. Old and new matrix algebra useful for statistics. notes, December 2000.
- [2] K.P. Petersen and M. S. Petersen. The matrix cookbook. notes, 2006.