A note on matrix differentiation

Pawel Kowal

December 2006

Online at http://mpra.ub.uni-muenchen.de/3917/
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 flattered 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 $\frac{\partial X}{\partial \theta}$ 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 flatted 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 flatted 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 gen-
eralized 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 func-
tions and their derivatives, using these elementary operations, the generalized
Kronecker products and the generalized transpositions.

We consider also more general matrix functions containing matrix func-
tions (inverse, trace and determinant). Defining the generalized trace func-
tion 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, \( \theta = \text{col}(\theta_1, \theta_2, \ldots, \theta_k) \). Let

the differential of the function \( X \) with respect to \( \theta \) is defined as

\[
\frac{\partial X}{\partial \theta} = \left[ \frac{\partial X}{\partial \theta_1} \frac{\partial X}{\partial \theta_2} \ldots \frac{\partial X}{\partial \theta_k} \right]
\]

for \( \frac{\partial X}{\partial \theta_i} \in \mathbb{R}^{m \times n}, i = 1, 2, \ldots, k \).

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

\[
\frac{\partial}{\partial \theta} (X \times Y) = \left[ \frac{\partial X}{\partial \theta_1} \times Y + X \times \frac{\partial Y}{\partial \theta_1} \ldots \frac{\partial X}{\partial \theta_k} \times Y + X \times \frac{\partial Y}{\partial \theta_k} \right]
\]

\[
= \left[ \frac{\partial X}{\partial \theta_1} \ldots \frac{\partial X}{\partial \theta_k} \right] \times \begin{bmatrix} Y & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & Y \end{bmatrix} + X \times \begin{bmatrix} \frac{\partial Y}{\partial \theta_1} \ldots \frac{\partial Y}{\partial \theta_k} \end{bmatrix}
\]

\[
= \frac{\partial X}{\partial \theta} \times (I_k \otimes Y) + X \times \frac{\partial Y}{\partial \theta}
\]

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

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

$$T_n(X) = [X'_1, X'_2, \ldots, X'_n]$$

**Proposition 2.3.** The following equations hold

1. $\frac{\partial}{\partial \theta} (X') = T_k(\frac{\partial X}{\partial \theta})$
2. $\frac{\partial}{\partial \theta} (T_n(X)) = T_{k \times n}(\frac{\partial X}{\partial \theta})$

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

$$\frac{\partial}{\partial \theta} (T_n(X)) = [T_n(\frac{\partial X}{\partial \theta_1}) \ldots T_n(\frac{\partial X}{\partial \theta_k})]$$

$$= [\frac{\partial X'_1}{\partial \theta_1}, \ldots, \frac{\partial X'_k}{\partial \theta_1}, \ldots, \frac{\partial X'_1}{\partial \theta_k}, \ldots, \frac{\partial X'_n}{\partial \theta_k}] = T_{k \times n}(\frac{\partial X}{\partial \theta})$$

since

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

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 $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, q$. The Kronecker product, $X \otimes Y$ is defined as

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

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

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

$$X \otimes_{n}^{1} Y = [X \otimes Y_1, \ldots, X \otimes Y_n]$$

$$X \otimes_{m}^{n} Y = [X_1 \otimes Y, \ldots, X_m \otimes Y]$$

$$X \otimes_{n_1,n_2,\ldots,n_s}^{1,m_2,\ldots,m_s} Y = [X \otimes_{n_1,n_2,\ldots,n_s}^{m_2,\ldots,m_s} Y_1, \ldots, X \otimes_{n_1}^{m_2,\ldots,m_s} Y_{n_1}]$$

$$X \otimes_{n_1,n_2,\ldots,n_s}^{m_1,m_2,\ldots,m_s} Y = [X_1 \otimes_{n_1,n_2,\ldots,n_s}^{1,m_2,\ldots,m_s} Y, \ldots, X_m \otimes_{n_1,n_2,\ldots,n_s}^{1,m_2,\ldots,m_s} Y]$$

assuming that appropriate matrix partitions exist.
Proposition 2.5. The following equations hold

1. \( \frac{\partial}{\partial \theta} (X \otimes Y) = \frac{\partial X}{\partial \theta} \otimes Y + X \otimes \frac{\partial Y}{\partial \theta} \)

2. \( \frac{\partial}{\partial \theta} (X \otimes_{n_1, \ldots, n_s} Y) = \frac{\partial X}{\partial \theta} \otimes_{k, n_1, \ldots, n_s} Y + X \otimes_{1, n_1, \ldots, n_s} \frac{\partial Y}{\partial \theta} \)

Proof. \( \frac{\partial}{\partial \theta} (X \otimes_{n_1, \ldots, n_s} Y) = \frac{\partial X}{\partial \theta} \otimes_{k, n_1, \ldots, n_s} Y + X \otimes_{1, n_1, \ldots, n_s} \frac{\partial Y}{\partial \theta} \)

Since \( X \otimes Y = X \otimes 1 Y \), in case of the standard Kronecker product we obtain

\[
\frac{\partial}{\partial \theta} (X \otimes Y) = \frac{\partial X}{\partial \theta} \otimes Y + X \otimes \frac{\partial Y}{\partial \theta}
\]

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

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

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

\[
\frac{\partial}{\partial \theta} (\alpha X) = \frac{\partial}{\partial \theta} ((\alpha \otimes I_p) \times X) = \frac{\partial (\alpha \otimes I_p)}{\partial \theta} \times (I_k \otimes X) + (\alpha \otimes I_p) \times \frac{\partial X}{\partial \theta}
\]

\[
= (\frac{\partial \alpha}{\partial \theta} \otimes I_p) \times (I_k \otimes X) + \alpha \times \frac{\partial X}{\partial \theta} = \frac{\partial \alpha}{\partial \theta} \otimes X + \alpha \times \frac{\partial X}{\partial \theta}
\]

Let \( S \) 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 \( S \). Let \( \text{dif} S \equiv \{ \partial X/\partial \theta : X \in S \} \). The set \( S \) 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 $X + Y$, $X \times Y$, $T_n(X)$, $X \otimes_{m_1, \ldots, m_s} Y$ if exist belong to $\text{ext}(\mathcal{S})$, where $n, n_1, \ldots, n_s, m_1, \ldots, m_s$ 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 expresses, 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 $X = [X_1, X_2, \ldots, X_n]$, where $X_i \in \mathbb{R}^{p \times p}$, $i = 1, 2, \ldots, n$ is a $p \times p$ matrix, is a partition of $p \times np$ dimensional matrix $X$. Then

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

**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. $\frac{\partial \text{tr}_n(X)}{\partial \theta} = \text{tr}_{k \times n} \left( \frac{\partial X}{\partial \theta} \right)$
3. $\frac{\partial X^{-1}}{\partial \theta} = -X^{-1} \times \frac{\partial X}{\partial \theta} \times (I_k \otimes X^{-1})$
Proof. We have

\[
\frac{\partial \det(X)}{\partial \theta} = \begin{bmatrix} \frac{\partial \det(X)}{\partial \theta_1} & \ldots & \frac{\partial \det(X)}{\partial \theta_k} \end{bmatrix} = \left[ \det(X) \operatorname{tr}(X^{-1} \times \frac{\partial X}{\partial \theta_1}) \ldots \det(X) \times \operatorname{tr}(X^{-1} \times \frac{\partial X}{\partial \theta_k}) \right] = \det(X) \times \operatorname{tr}_k(X^{-1} \times \frac{\partial X}{\partial \theta})
\]

\[
\frac{\partial \operatorname{tr}_n(X)}{\partial \theta} = \begin{bmatrix} \frac{\partial \operatorname{tr}_n(X)}{\partial \theta_1} & \ldots & \frac{\partial \operatorname{tr}_n(X)}{\partial \theta_k} \end{bmatrix} = \left[ \operatorname{tr}_n(\frac{\partial X}{\partial \theta_1}) \ldots \operatorname{tr}(\frac{\partial X}{\partial \theta_k}) \right] = \operatorname{tr}_k n(\frac{\partial X}{\partial \theta})
\]

Similarly

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

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

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

**Theorem 3.3.** \( \text{dif}(\text{ext}_2(S)) = \text{ext}_2(S \cup \text{dif}(S)) \).

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

---

### 4 Derivatives of function composition

Let \( f \) is a matrix valued function given by

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

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

**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
\[
f(x) = \begin{bmatrix}
  f_{11}(x) & \cdots & f_{1n}(x) \\
  \vdots & \ddots & \vdots \\
  f_{m1}(x) & \cdots & f_{mn}(x)
\end{bmatrix}
\]
where \(f_{ij}(x)\) are scalar valued functions. Then for \(s = 1, \ldots, k\)
\[
\frac{\partial f_{ij}(x)}{\partial \theta_s} = \sum_{k=1}^{p} \frac{\partial f_{ij}(x)}{\partial x_k} \times \frac{\partial x_k}{\partial \theta_s} = \frac{\partial f_{ij}(x)}{\partial x} \times \frac{\partial x}{\partial \theta_s}
\]
since \(\frac{\partial x}{\partial \theta_s}\) is a column vector. Further
\[
f(x) = \sum_{k=1}^{p} \begin{bmatrix}
  \frac{\partial f_{11}(x)}{\partial x_k} & \cdots & \frac{\partial f_{1n}(x)}{\partial x_k} \\
  \vdots & \ddots & \vdots \\
  \frac{\partial f_{m1}(x)}{\partial x_k} & \cdots & \frac{\partial f_{mn}(x)}{\partial x_k}
\end{bmatrix} \times \frac{\partial x_k}{\partial \theta_s} = \sum_{k=1}^{p} \frac{\partial f(x)}{\partial x} \times \frac{\partial x}{\partial \theta_s} = \sum_{k=1}^{p} \frac{\partial f(x)}{\partial x} \times \frac{\partial x}{\partial \theta_s} \otimes I_n
\]
Finally
\[
\frac{f(x)}{\partial \theta_s} = \frac{\partial f(x)}{\partial x} \times \left[ \frac{\partial x}{\partial \theta_j} \otimes I_n \ldots \frac{\partial x}{\partial \theta_k} \otimes I_n \right] = \frac{\partial f(x)}{\partial x} \times \left( \frac{\partial x}{\partial \theta_j} \otimes I_n \right)
\]

5 Properties of the generalized Kronecker product

Proposition 5.1. For any matrices \(A, B\)

1. \(A \otimes^k_1 B = A \otimes B\).

2. \(A \otimes_{\cdots, 1, \ldots}^{\cdots, 1, \ldots} B = A \otimes_{\cdots, n_k, \ldots}^{\cdots, m_k, \ldots} B\).

3. \(A \otimes_{\cdots, n_k, \ldots, n_k+1, \ldots}^{\cdots, 1, \ldots} B = A \otimes_{\cdots, n_k, \ldots}^{\cdots, m_k, \ldots, n_k+1, \ldots} B\).

4. \(A \otimes_{\cdots, 1, \ldots, n_k+1, \ldots}^{\cdots, m_k, \ldots} B = A \otimes_{\cdots, n_k+1, \ldots}^{\cdots, m_k \times m_k+1, \ldots} B\).

assuming that the Kronecker products exist.
Proposition 5.2. For any matrices $A$, $B$, $C$

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

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

assuming that the Kronecker products exist and matrix dimensions coincide.

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

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

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

Proof. Observe that $X \otimes_{n_1,\ldots,n_s}^{m_1,\ldots,m_s} Y = X \otimes_{n_1,\ldots,n,s+1}^{m_1,\ldots,m_s+1} Y$, 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 $(AB) \otimes_{n_k,\ldots,n_{k+1}}^{m_k,\ldots,m_{k+1}} (CD) = (A \otimes C) \times (B \otimes_{n_k,\ldots,n_{k+1}}^{m_k,\ldots,m_{k+1}} D)$ for $k \geq 0$. Then

$$(AB) \otimes_{n_k+1,\ldots,n_{k+1}}^{1,m_k,\ldots,m_{k+1}} (CD)$$

$$= [ (AB) \otimes_{n_k,\ldots,n_{k+1}}^{m_k,\ldots,m_{k+1}} (CD) ]$$

$$= [ (A \otimes C)(B \otimes_{n_k,\ldots,n_{k+1}}^{m_k,\ldots,m_{k+1}} D) ]$$

$$= (A \otimes C) \times (B \otimes_{n_k+1,\ldots,n_{k+1}}^{1,m_k,\ldots,m_{k+1}} D)$$

Similarly

$$(AB) \otimes_{n_k+1,\ldots,n_{k+1}}^{m_{k+1},m_k,\ldots,m_{k+1}} (CD)$$

$$= [ (AB) \otimes_{n_k+1,\ldots,n_{k+1}}^{m_k,\ldots,m_{k+1}} (CD) ]$$

$$= [ (A \otimes C)(B \otimes_{n_k+1,\ldots,n_{k+1}}^{m_k,\ldots,m_{k+1}} D) ]$$

$$= (A \otimes C) \times (B \otimes_{n_k+1,\ldots,n_{k+1}}^{m_{k+1},m_k,\ldots,m_{k+1}} D)$$

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

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

assuming that Kronecker product exists.

Proposition 5.5. For any matrices $A$, $B$, $C$

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

assuming that Kronecker products exist.
Proof. Observe that \( X \otimes_{n_1 \ldots n_s}^m Y = X \otimes_{n_1 \ldots n_s}^{m_1 \ldots m_s} Y \), 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 \( A \otimes_{n_k \ldots n_1}^{m_k \ldots m_1} (B \otimes C) = (A \otimes_{n_k \ldots n_1}^{m_k \ldots m_1} B) \otimes C \) for \( k \geq 0 \). Then

\[
A \otimes_{n_k+1}^{1, m_k \ldots m_1, 1} (B \otimes C)
= \left[ A \otimes_{n_k+1}^{m_k \ldots m_1, 1} (B_1 \otimes C) \ldots, A \otimes_{n_k+1}^{m_k \ldots m_1, 1} (B_{n_k} \otimes C) \right]
= \left[ (A \otimes_{n_k+1}^{m_k \ldots m_1, 1} B_1) \otimes C \ldots, (A \otimes_{n_k+1}^{m_k \ldots m_1, 1} B_{n_k}) \otimes C \right]
= (A \otimes_{n_k+1}^{1, m_k \ldots m_1, 1} B) \otimes C
\]

Similarly

\[
A \otimes_{n_k+1}^{m_k+1, m_k \ldots m_1, 1} (B \otimes C)
= \left[ A_1 \otimes_{n_k+1}^{1, m_k \ldots m_1, 1} (B \otimes C) \ldots, A_{m_k+1} \otimes_{n_k+1}^{1, m_k \ldots m_1, 1} (B \otimes C) \right]
= \left[ (A_1 \otimes_{n_k+1}^{m_k \ldots m_1, 1} B) \otimes C \ldots, (A_{m_k+1} \otimes_{n_k+1}^{m_k \ldots m_1, 1} B) \otimes C \right]
= (A \otimes_{n_k+1}^{1, m_k \ldots m_1, 1} B) \otimes C
\]

\[ \square \]

**Proposition 5.6.** For any matrices \( A, B, C \)

\[ A \otimes (B \otimes_{n_1 \ldots n_s}^{m_1 \ldots m_s} C) = (A \otimes B) \otimes_{n_1 \ldots n_s}^{q, m_1 \ldots m_s} C \]

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

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

\[
A_i \otimes (B \otimes_{n_k+1}^{1, m_k \ldots m_1, 1} C)
= \left[ A_i \otimes (B \otimes_{n_k+1}^{m_k \ldots m_1, 1} C_1) \ldots, A_i \otimes (B \otimes_{n_k+1}^{m_k \ldots m_1, 1} C_{n_k}) \right]
= \left[ (A_i \otimes B) \otimes_{n_k \ldots n_1}^{m_k \ldots m_1, 1} C_1 \ldots, (A_i \otimes B) \otimes_{n_k \ldots n_1}^{m_k \ldots m_1, 1} C_{n_k} \right]
= (A_i \otimes B) \otimes_{n_k+1}^{1, m_k \ldots m_1, 1} C
\]

Similarly

\[
A_i \otimes (B \otimes_{n_k+1}^{m_k+1, m_k \ldots m_1, 1} C)
= \left[ A_i \otimes (B_1 \otimes_{n_k+1}^{m_k \ldots m_1, 1} C) \ldots, A_i \otimes (B_{m_k+1} \otimes_{n_k+1}^{m_k \ldots m_1, 1} C) \right]
= \left[ (A_i \otimes B_1) \otimes_{n_k+1}^{1, m_k \ldots m_1, 1} C \ldots, (A_i \otimes B_{m_k+1}) \otimes_{n_k+1}^{1, m_k \ldots m_1, 1} C \right]
= (A_i \otimes B) \otimes_{n_k+1}^{m_k+1, m_k \ldots m_1, 1} C
\]
Finally,
\[
A \otimes (B \otimes_{n_{k+1},n_k,\ldots,n_1,1}^{m_{k+1},m_k,\ldots,m_1,1} C) = 
\begin{bmatrix}
A_1 \otimes (B \otimes_{n_{k+1},n_k,\ldots,n_1,1}^{m_{k+1},m_k,\ldots,m_1,1} C) \\
\vdots \\
A_q \otimes (B \otimes_{n_{k+1},n_k,\ldots,n_1,1}^{m_{k+1},m_k,\ldots,m_1,1} C)
\end{bmatrix} 
= (A \otimes B) \otimes_{1,n_{k+1},n_k,\ldots,n_1,1}^{q,m_{k+1},m_k,\ldots,m_1,1} C
\]

\[\square\]

**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 \( B^i_j \) denotes element of \( B \) at j-th row and i-th column. Then
\[
(I_m \otimes_p^1 I_p) \times (B^j \otimes A^i) = 
\begin{bmatrix}
I_m \otimes I_p^1 \\
\vdots \\
I_m \otimes I_p^p
\end{bmatrix} \times 
\begin{bmatrix}
B^j_1 A^i \\
\vdots \\
B^j_p A^i
\end{bmatrix} 
= \sum_{r=1}^p (I_m \otimes I_p^r) \times (A^i \otimes B^j_r) = \sum_{r=1}^p A^i \otimes (I_p^r \times B^j_r) = A^i \otimes B^j
\]

Further
\[
(I_m \otimes_p^1 I_p) \times (B^j \otimes A) = (I_m \otimes_p^1 I_p) \times 
\begin{bmatrix}
B^j_1 A^1 \\
\vdots \\
B^j_q A^n
\end{bmatrix} 
= [A^1 \otimes B^j \\
\vdots \\
A^n \otimes B^j] = A \otimes B^j
\]

\( (I_m \otimes_p^1 I_p) \times (B \otimes A) = (I_m \otimes_p^1 I_p) \times 
\begin{bmatrix}
B^1_1 A \\
\vdots \\
B^q_1 A
\end{bmatrix} 
= [A \otimes B^1 \\
\vdots \\
A \otimes B^q] = A \otimes_q^1 B
\]

\[\square\]

**Proposition 5.8.** Let \( A \) is \( m \times n \) matrix. Let \( B \) is \( p \times q \) matrix. Then
\[
A \otimes_{m_1,1,\ldots,m_s,1}^{1,n_1,\ldots,n_s,1} B = (I_m \otimes_p^1 I_p) \times (B \otimes_{m_1,\ldots,m_s}^{m_1,\ldots,m_s} A)
\]
where \( \tilde{m} = m_1 \times \cdots \times m_s \), assuming that the Kronecker products exist.
Proof. Proposition holds for $s = 0$. Let for any $s \geq 0$ $A \otimes_{m_s}^{1, n_{s+1}, \ldots, m_1, 1, q/m_s} B = (I_m \otimes_p^{1_p} I_p) \times (B \otimes_{m_s}^{m_{s+1}, \ldots, m_1, 1, q/m_s} A)$, where $m_s = m_1 \times \cdots \times m_s$. Then

\[
(I_m \otimes_p^{1_p} I_p) \times (B \otimes_{m_s}^{1, m_{s+1}, \ldots, m_1, 1, q/m_s} A)
= (I_m \otimes_p^{1_p} I_p) \times \left[ B \otimes_{m_s}^{m_{s+1}, \ldots, m_1, 1, q/m_s} A_1 \quad \cdots \quad B \otimes_{m_s}^{m_{s+1}, \ldots, m_1, 1, q/m_s} A_{n_{s+1}} \right]
= \left[ A_1 \otimes_{m_s}^{1, m_{s+1}, \ldots, m_1, 1, q/m_s} B \quad \cdots \quad A_{n_{s+1}} \otimes_{m_s}^{m_{s+1}, \ldots, m_1, 1, q/m_s} B \right]
= A \otimes_{1, 1, m_s, 1, m_{s+1}, 1, 1, \ldots, m_1, 1, q/m_s} B
\]

and

\[
(I_m \otimes_p^{1_p} I_p) \times (B \otimes_{m_s}^{m_{s+1}, \ldots, m_1, 1, q/m_s} A)
= (I_m \otimes_p^{1_p} I_p) \times \left[ B_1 \otimes_{m_s}^{m_{s+1}, \ldots, m_1, 1, q/m_s} A \quad \cdots \quad B_{m_{s+1}} \otimes_{m_s}^{m_{s+1}, \ldots, m_1, 1, q/m_s} A \right]
= \left[ A \otimes_{1, 1, m_s, 1, m_{s+1}, 1, 1, \ldots, m_1, 1, q/m_s} B_1 \quad \cdots \quad A \otimes_{1, 1, m_s, 1, m_{s+1}, 1, 1, \ldots, m_1, 1, q/m_s} B_{m_{s+1}} \right]
= A \otimes_{m_{s+1}, 1, m_{s+1}, 1, m_1, 1, q/m_s} B
\]

$\square$

Proposition 5.9.

\[
(I_m \otimes_q^{1_q} I_q)^{-1} = I_q \otimes_m^{1_m} I_m
\]

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

\[
(I_m \otimes_q^{1_q} I_q)^T = \begin{bmatrix} I_m \otimes (I_q^T)^T \\ \cdots \\ I_m \otimes (I_q^T)^T \end{bmatrix} = \begin{bmatrix} (I_q^T)^T \otimes_m^{1_m} I_m \\ \cdots \\ (I_q^T)^T \otimes_m^{1_m} I_m \end{bmatrix} = I_q \otimes_m^{1_m} I_m
\]

The second equality can be shown using for example proposition 5.7. $\square$

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

\[ I_n \otimes (I_{q/n} \otimes_m^1 I_m) \times (I_m \otimes_n^1 I_q)^T = \left[ \begin{array}{c} (I_{q/n} \otimes_m^1 I_m) \times (I_m \otimes_n^1 I_q)^T \\ \vdots \\ (I_{q/n} \otimes_m^1 I_m) \times (I_m \otimes_n^1 I_q)^T \end{array} \right] = \left[ \begin{array}{c} (I_q^m)^T \otimes_m^1 I_m \\ \vdots \\ (I_q^n)^T \otimes_m^1 I_m \end{array} \right] = I_q \otimes_m^1 I_m \]

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

\[ (I_m \otimes_n^1 I_q)^{-1} = \left( I_n \otimes (I_{q/n} \otimes_m^1 I_m)^{-1} \right) \times (I_q \otimes_m^1 I_m) = \left( I_n \otimes (I_m \otimes_{q/n}^1 I_q)/n \right) \times (I_q \otimes_m^1 I_m) = (I_{nm} \otimes_{q/n}^n I_{q/n}) \times (I_q \otimes_m^1 I_m) \]

Further

\[ (I_m \otimes_n^k I_q)^{-1} = \left( I_k \otimes I_{m/k} \otimes_{1,n}^k I_q \right)^{-1} = \left( I_k \otimes (I_{m/k} \otimes_{1,n}^1 I_q) \right)^{-1} = I_k \otimes \left( (I_{nm/k} \otimes_{q/n}^n I_{q/n}) \times (I_q \otimes_{m/k}^n I_{m/k}) \right) = I_k \otimes (I_{nm/k} \otimes_{q/n}^n I_{q/n}) \times I_k \otimes (I_q \otimes_{m/k}^1 I_{m/k}) = (I_{nm} \otimes_{q/n}^n I_{q/n}) \times (I_{kq} \otimes_{m/k}^k I_{m/k}) \]

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.

\[ (I_m \otimes_p^1 I_p) \times (B \otimes A) = A \otimes_q^1 B = (A \otimes B) \times (I_q \otimes_n^1 I_q) = (A \otimes B) \times (I_q \otimes_n^1 I_n)^{-1} \]

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
