DERIVATIVES OF MATRIX-VALUED FUNCTIONS INVOLVING SEMI-TENSOR PRODUCTS IN A VECTOR VARIABLE

THANAPHON PHOONPHIPHAT

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE IN APPLIED MATHEMATICS

DEPARTMENT OF MATHEMATICS SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2024
KMITL-2024-SC-M-001-062

Thesis Title	Derivatives of Matrix-Valued Functions Involving Semi-Tensor Products in a Vector Variable
Student Name	Mr. Thanaphon Phoonphiphat
Student ID	65056043
Degree	Master of Science (Applied Mathematics)
Department	Mathematics
Year	2024
Thesis Advisor	Assoc. Prof. Dr. Pattrawut Chansangiam

Abstract

The propose of this paper to derive exact fomulas for the derivatives of certain matrix-valued functions with vector variables involving semi-tensor products. Moreover, we investigate the product rule of two matrix-valued functions with vector variables involving semi-tensor products, and its special cases. This paper enhances classical mathematical theory by incorporating semi-tensor products and eliminating restrictions on matrix dimensions.

Keywords : Matrix derivative, Kronecker product, semi-tensor product, vectorization, zero-one matrix.

Acknowledgements

I would like to express my sincere gratitude to Assoc. Prof. Dr. Pattrawut Chansangiam, my supervisor, for his guidance, advice, and support throughout this research. His expertise and encouragement have been instrumental to the success of this work. I am truly grateful for the opportunity to learn and grow under his mentorship.

I would like to thank my thesis committee members, Assoc. Prof. Dr. Wicharn Lewkeeratiyutkul and Asst. Prof. Dr. Kanchana Kumnungkit, for their valuable feedback and insightful comments, which have been essential in enhancing my research.

I would like to express my sincere gratitude for the scholarship provided for my master's degree from the King Mongkut's Institute of Technology Ladkrabang Research Fund, grant number KREF016517. This scholarship is categorized as a master's degree educational grant for the project titled "Gradient of Real Valued Functions in Vector or Matrix Variables Involving Semi-Tensor Products." This generous support has played a crucial role in facilitating my research endeavors and has allowed me to focus on my academic pursuits without the burden of financial constraints.

I am extremely thankful to my parents for all the support throughout academic years. I could not come this far without their love and motivation.

Thanaphon Phoonphiphat

Table of Contents

List of Tables

Chapter 1

Introduction

1.1 Research motivation

It is well-known that matrices are an important tool in many fields, especially in applied mathematics, data science, statistics, engineering, and economics; see e.g. [1, 2, 3, 4, 5]. Particularly, an interesting topic is matrix calculus which consists of studying matrix operations and matrix derivatives. Matrix calculus provides a convenient way to deal with multivariable calculus over spaces of matrices or vectors. Matrix derivatives are a crucial part of machine learning, particularly deep neural networks, which are trained by optimizing a loss function. Recall that the activation of a sole unit in an artificial neural network (ANN) is usually evaluated using the usual matrix multiplication between a weight vector w and an input vector x with an additional term called bias b which can be written formally as shown below

$$z(x) = \sum_{i}^{n} w_i x_i + b = w'x + b,$$

where z is called an activation function; see e.g. [6, 7]. In fact, ANN is made up from many of these units, assembled into collections of neurons. In order to fit this architecture, the training process is needed. Training ANN means choosing suitable weights w and bias b so that we get the desired output for all inputs x. To complete this step, we minimize a loss function which is the error between the activation function and the actual output. Obviously, all of those require the partial derivative of the loss function with respect to the parameters w and b.

As mentioned earlier, most of works applied matrix calculus based on the usual matrix multiplication which is the most familiar matrix operation. More precisely, many authors had derived several exact formulas for the derivatives of matrix/vector/scalar-valued functions with respect to matrix/vector/scalar as variables involving the usual matrix multiplication. Moreover, they also derived the product rule, chain rule and etc involving the usual matrix multiplication as well. In fact, the notation of usual matrix multiplication was generalized to the semi-tensor product. Recall that the multiplication of two matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{r\times s}$ is well-defined if $n = r$. The semi-tensor product (STP) of two matrices A and B of arbitrary dimensions is defined to be the usual product between the tensor product of each factor with the identity matrix of appropriate size. Indeed, for matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{r\times s}$ of arbitrary dimensions, D. Cheng [8] defined their semi-tensor product (STP) by

$$A \ltimes B = (A \otimes I_{\frac{\alpha}{n}})(B \otimes I_{\frac{\alpha}{p}}) \in \mathbb{R}^{\frac{\alpha m}{n} \times \frac{\alpha q}{p}}, \tag{1.1.1}$$

where α is the least common multiple of n and r. Here, the operation $\otimes$ is the Kronecker product. It turns out that the STP posseses rich algebraic properties like the usual matrix product, such as the associativity, the left/right distribution over the matrix addition, and certain identity-like properties. Furthermore, the STP is compatible with the scalar multiplication, the transposition, and the inversion; see e.g. [8, 9, 10, 11, 12]. The semi-tensor product provides a convenient way to convert higher-dimensional data or a multilinear function to a simple expression of matrices/vectors. For example, logical operations, fuzzy operations and Boolean operations can be represented in terms of the STP between certain representing matrix and a data vector. Thus, the STP is a useful tool in several areas of pure mathematics, such as classical and fuzzy logic; see e.g. [13, 14], multi-variable polynomials [8], lattice theory [8], finite-dimensional algebra [8], and connections in differential geometry [8].

From the above discussion, the STP is one of the powerful matrix operations. It is reasonable to generalize the formulas for the matrix derivatives. Instead of focusing on the usual matrix multiplication, it is worthy studying the matrix calculus involving the STP including matrix derivative, product rule and chain rule.

This thesis is arranged as follows. In Chapter 2, the symbolic notation and some useful definitions involving matrix algebra are given. In Chapter 3 and 4, the exact formulas for matrix derivatives, product rule and chain rule involving STP are given. Finally, Chapter 5 provides a brief conclusion.

1.2 Research Objectives

1) To derive exact formulas for the derivatives of matrix-valued functions with vector variables involving semi-tensor products.

2) To investigate the product rules of matrix-valued functions with vector variables involving semi-tensor products.

1.3 Scope

1) Let $A \in \mathbb{R}^{m\times n}$. We investigate the following derivatives:

$$\begin{aligned}
&\frac{\partial}{\partial \vec{x}}(A \ltimes \vec{x}),\\
&\frac{\partial}{\partial \vec{x}}(\vec{x}' \ltimes A),\\
&\frac{\partial}{\partial \vec{x}}(\vec{x}' \ltimes A \ltimes \vec{x}),\\
&\frac{\partial}{\partial \vec{x}}(\vec{y}' \ltimes A \ltimes \vec{x}),\\
&\frac{\partial}{\partial \vec{y}}(\vec{y}' \ltimes A \ltimes \vec{x}),
\end{aligned}$$

where $\vec{x}$ and $\vec{y}$ are vector variables.

2) Let $A \in \mathbb{R}^{m\times n}$. We investigate the following derivatives:

$$\frac{\partial}{\partial \vec{z}}(A \ltimes \vec{x}),$$
$$\frac{\partial}{\partial \vec{z}}(\vec{x}' \ltimes A),$$
$$\frac{\partial}{\partial \vec{z}}(\vec{x}' \ltimes A \ltimes \vec{x}),$$
$$\frac{\partial}{\partial \vec{z}}(\vec{y}' \ltimes A \ltimes \vec{x}),$$

where $\vec{x}$ and $\vec{y}$ are vector functions of a vector variable $\vec{z}$.

3) Let $A(x)$ and $B(x)$ be matrix-valued functions of a vector variable x. We investigate the following derivatives:

$$\frac{\partial}{\partial \vec{x}}(A(\vec{x}) \ltimes B(\vec{x})),$$

its special cases, and consequences.

1.4 Benefits

To obtain exact formulas for the derivatives of certain matrix-valued functions with vector variables involving semi-tensor products. Such derivative formulas can be applied in certain optimization problems in neuron networks.

1.5 Research Methodology

1) Study basic concepts in linear algebra and matrix theory.

2) Study matrix calculus, especially the derivatives of matrix-valued functions with vector variables involving the usual matrix multiplication.

3) Study semi-tensor products of matrices from research/survey papers and textbooks.

4) Prove exact formulas for the derivatives of matrix-valued functions with vector variables involving semi-tensor products.

5) Prove exact formulas for the product rules of matrix-valued functions with vector variables involving semi-tensor products.

6) Conclude the whole work. Write a research paper. Write the thesis.

Table 1.1: The research schedule

<table>
<tr><th rowspan="3">Activity</th><th colspan="9">Time frame</th></tr>
<tr><th colspan="2">2022</th><th colspan="4">2023</th><th colspan="3">2024</th></tr>
<tr><th>Aug.-Sep.</th><th>Oct.-Dec.</th><th>Jan.-Mar.</th><th>Aprl.-Jun.</th><th>Jun.-Sep.</th><th>Oct.-Dec.</th><th>Jan.-Mar.</th><th>Aprl.-May.</th><th>Jun.-Oct.</th></tr>
<tr><td>Step 1</td><td colspan="2">⟷</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Step 2</td><td></td><td colspan="3">⟷</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Step 3</td><td></td><td></td><td colspan="2">⟷</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Step 4</td><td></td><td></td><td></td><td></td><td colspan="2">⟷</td><td></td><td></td><td></td></tr>
<tr><td>Step 5</td><td></td><td></td><td></td><td></td><td></td><td colspan="3">⟷</td><td></td></tr>
<tr><td>Step 6</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td colspan="2">⟷</td></tr>
</table>

Chapter 2

Preliminaries

In this chapter, we introduce symbolic notation involving matrix algebra. The basic knowledge, elementary backgrounds and essential definitions are discussed here. These preparation will be applied throughout this thesis.

2.1 Vectorization

Througout, denote the transpose of a matrix A by A'. Each Each $A = [a_{ij}] \in \mathbb{R}^{m\times n}$ can be written in three ways as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \vec{a}_{*1} & \vec{a}_{*2} & \cdots & \vec{a}_{*n} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \text{——} & \vec{a}_{1*}^{'} & \text{——} \\ \text{——} & \vec{a}_{2*}^{'} & \text{——} \\ & \vdots & \\ \text{——} & \vec{a}_{m*}^{'} & \text{——} \end{bmatrix},$$

where $\vec{a}_{*j}$ is the j-th column of A and $\vec{a}_{i*}^{'}$ is the i-th row of A. The operator Vec : $\mathbb{R}^{m\times n} \to \mathbb{R}^{mn\times 1}$ maps each $A \in \mathbb{R}^{m\times n}$ to an $mn \times 1$ column vector given below

$$\text{Vec}(A) = \begin{bmatrix} \vec{a}_{*1} \\ \vec{a}_{*2} \\ \vdots \\ \vec{a}_{*n} \end{bmatrix}.$$

The Vec operator converts an $m \times n$ matrix A into an $mn \times 1$ column vector by stacking those columns of A one underneath another. In addition, the operator Devec : $\mathbb{R}^{m\times n} \to \mathbb{R}^{1\times mn}$, is defined as a $1 \times mn$ row vector given as follows

$$\text{Devec}(A) = \begin{bmatrix} \vec{a}_{1*}^{'} & \vec{a}_{2*}^{'} & \cdots & \vec{a}_{m*}^{'} \end{bmatrix}.$$

The Devec operator transforms $m \times n$ matrix A into $1 \times mn$ row vector by stacking those rows of A alongside each other. Obviously, these two operators are closely connected. We can see that

$$\text{Vec}(A)' = \text{Devec}(A')$$

and

$$\text{Vec}(A') = \text{Devec}(A)'.$$

2.2 Kronecker Product

Definition 2.1. Let A be an $m \times n$ matrix and let B be a $p \times q$ matrix. The Kronecker product or the tensor product, denoted by $\otimes$, is an operation on two arbitrary matrices resulting as an $mp \times nq$ matrix given by

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}.$$

Example 2.2. Let $A = \begin{bmatrix} 2 & 3 \\ 4 & 5 \\ 1 & 2 \end{bmatrix} \in \mathbb{R}^{3\times 2}$ and $B = \begin{bmatrix} -1 & 2 \end{bmatrix} \in \mathbb{R}^{1\times 2}$. Then

$$\begin{aligned} A \otimes B &= \begin{bmatrix} 2\begin{bmatrix} -1 & 2 \end{bmatrix} & 3\begin{bmatrix} -1 & 2 \end{bmatrix} \\ 4\begin{bmatrix} -1 & 2 \end{bmatrix} & 5\begin{bmatrix} -1 & 2 \end{bmatrix} \\ 1\begin{bmatrix} -1 & 2 \end{bmatrix} & 2\begin{bmatrix} -1 & 2 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} -2 & 4 & -3 & 6 \\ -4 & 8 & -5 & 10 \\ -1 & 2 & -2 & 4 \end{bmatrix} \in \mathbb{R}^{3\times 4}. \end{aligned}$$

The following properties involving Kronecker product are well-known as shown in the following.

Theorem 2.3. [15]. Let $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{p\times q}$ and $C \in \mathbb{R}^{r\times s}$. Then

$$A \otimes (B \otimes C) = (A \otimes B) \otimes C = A \otimes B \otimes C. \tag{2.2.1}$$

Theorem 2.4. [15]. Let $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{p\times q}$ and $C \in \mathbb{R}^{p\times q}$. Then

1. $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$,
2. $(B + C) \otimes A = (B \otimes A) + (C \otimes A)$.

Theorem 2.5. [15]. Let $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{p\times q}$, $C \in \mathbb{R}^{r\times s}$ and $D \in \mathbb{R}^{m\times n}$ such that $nq = rm$. Then

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD).$$

Theorem 2.6. [15]. Let $A \in \mathbb{R}^{m\times m}$ and $B \in \mathbb{R}^{n\times n}$. Then

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

Theorem 2.7. [15]. Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$. Then

$$(A \otimes B)' = A' \otimes B'.$$

Furthermore , the connection between the Vec, Devec (Section 2.1) and Kronecker product are demonstrated in the following lemma.

Lemma 2.8. [15]. Let $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times r}$ and $C \in \mathbb{R}^{r\times q}$. Then

1. $\text{Vec}(A) = (I_n \otimes A) \; = \; (A' \otimes I_m)\,\text{Vec}(I_m)$,
2. $\text{Devec}(I_n)(A' \otimes I_n) \; = \; \text{Devec}(I_m)(I_m \otimes A)$,
3. $\text{Vec}(ABC) \; = \; (C' \otimes A)\,\text{Vec}\,B$.

2.3 Tracy-Singh product

Definition 2.9. Let $A = [a_{ij}] \in \mathbb{R}^{m\times n}$ and $B = [b_{ij}] \in \mathbb{R}^{p\times q}$ be partitioned with block submatrices A_{ij} and B_{kl} of order $m_i \times n_j$ and $p_k \times q_l$, respectively. Then the Tracy-Singh product of A and B, denote by $A \boxtimes B$, is defined as follows

$$A \boxtimes B = \Big[[A_{ij} \otimes B_{kl}]_{kl}\Big]_{ij} \in \mathbb{R}^{mp\times nq}.$$

Example 2.10. Let $A = \left[\begin{array}{c|c} 1 & 2 \\ 3 & 4 \end{array}\right] \in \mathbb{R}^{2\times 2}$ and $B = \left[\begin{array}{cc|cc} 1 & 0 & 0 & 1 \end{array}\right] \in \mathbb{R}^{1\times 4}$, where $A_{11} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$, $A_{12} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$, $B_{11} = \begin{bmatrix} 1 & 0 \end{bmatrix}$ and $B_{12} = \begin{bmatrix} 0 & 1 \end{bmatrix}$. Then

$$\begin{aligned} A \boxtimes B &= \begin{bmatrix} A_{11} \boxtimes B & A_{12} \boxtimes B \end{bmatrix} \\ &= \begin{bmatrix} A_{11} \otimes B_{11} & A_{11} \otimes B_{12} & A_{12} \otimes B_{11} & A_{12} \otimes B_{12} \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \end{bmatrix} & \begin{bmatrix} 1 \\ 3 \end{bmatrix} \otimes \begin{bmatrix} 0 & 1 \end{bmatrix} & \begin{bmatrix} 2 \\ 4 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \end{bmatrix} & \begin{bmatrix} 2 \\ 4 \end{bmatrix} \otimes \begin{bmatrix} 0 & 1 \end{bmatrix} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 & 1 & 2 & 0 & 0 & 2 \\ 3 & 0 & 0 & 3 & 4 & 0 & 0 & 4 \end{bmatrix} \in \mathbb{R}^{2\times 8}. \end{aligned}$$

2.4 Zero-one Matrices

A matrix whose elements are all either one or zero is so-called a zero-one matrix. A simple zero-one matrix that is frequently used is the $n \times n$ identity matrix, I_n, which has the beautiful property such that for any matrix $A \in \mathbb{R}^{m\times n}$, $AI_n = A$. However, There are several types of zero-one matrices appearing in many researches. These matrices are crucial in our work, especially, the one which is called the commutation matrix. Let e_i^n be the i-th column of the $n \times n$ identity matrix I_n. The commutation matrix is compactly expressed as follows

$$K_{mn} = \begin{bmatrix} I_n \otimes e_1^{m'} \\ I_n \otimes e_2^{m'} \\ \vdots \\ I_n \otimes e_m^{m'} \end{bmatrix} = \begin{bmatrix} I_m \otimes e_1^n & I_m \otimes e_2^n & \dots & I_m \otimes e_n^n \end{bmatrix} \in \mathbb{R}^{nm\times mn}. \quad (2.4.1)$$

The commutation matrix K_{mn} has the property that

$$K_{mn}\,\mathrm{Vec}(A) = \mathrm{Vec}(A').$$

Example 2.11. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \in \mathbb{R}^{3\times 2}$. From(2.4.1), we can compute

$$K_{32} = \begin{bmatrix} I_2 \otimes e_1^{3'} \\ I_2 \otimes e_2^{3'} \\ I_2 \otimes e_3^{3'} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{6\times 6}.$$

Alternatively

$$K_{32} = \begin{bmatrix} I_3 \otimes e_1^2 & I_3 \otimes e_2^2 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{6\times 6}.$$

We can see that

$$K_{32}\,\mathrm{Vec}(A) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 5 & 2 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}$$

$$= \mathrm{Vec}(A').$$

Moreover, we can define

$$K_{mn}^{\tilde{\tau}_n} = \begin{bmatrix} I_n \otimes e_1^{m'} & I_n \otimes e_2^{m'} & \dots & I_n \otimes e_m^{m'} \end{bmatrix} \in \mathbb{R}^{n\times nm^2} \qquad (2.4.2)$$

and

$$K_{mn}^{\tau_n} = \begin{bmatrix} I_m \otimes e_1^n \\ I_m \otimes e_2^n \\ \vdots \\ I_m \otimes e_n^n \end{bmatrix} \in \mathbb{R}^{mn^2\times m}. \qquad (2.4.3)$$

The explicit expression of $K_{mn}^{\tilde{\tau}_n}$ and $K_{mn}^{\tau_n}$ are illustrated in the following example.

Example 2.12.

$$\begin{aligned} K_{32}^{\tilde{\tau}_2} &= \begin{bmatrix} I_2 \otimes e_1^{3'} & I_2 \otimes e_2^{3'} & I_2 \otimes e_m^{3'} \end{bmatrix} \\ &= \begin{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} & \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} \end{bmatrix}. \end{aligned}$$

Example 2.13.

$$K_{32}^{\tau_3} = \begin{bmatrix} I_3 \otimes e_1^2 \\ I_3 \otimes e_2^2 \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{bmatrix}.$$

Lemma 2.14. [16]. From the above notation,

1. if $n = 1$, then $K_{mn}^{\tilde{\tau}_n} = K_{m1}^{\tilde{\tau}_1} =$ Devec(I_m),
2. If $m = 1$, then $K_{1,n}^{\tilde{\tau}_n} = I_n$.

2.5 Semi-tensor Product

Definition 2.15. Let A be an $m \times n$ matrix, and let B be a $p \times q$ matrix. The **general left semi-tensor product** of A and B denoted by $A \ltimes B$, is defined as follows

$$A \ltimes B = (A \otimes I_{\frac{\alpha}{n}})(B \otimes I_{\frac{\alpha}{p}}),$$

which is of dimension $\frac{\alpha m}{n} \times \frac{\alpha q}{r}$, where $\alpha =$ lcm(n, p).

Remark 2.16.

1. If $n|p$, i.e., $nt = p$, $t \in \mathbb{N}$, then the general left semi-tensor product becomes the **left semi-tensor product,** that is,

$$A \ltimes B = (A \otimes I_t)B.$$

On the other hand, if $p|n$, i.e., $pt = n$, $t \in \mathbb{N}$, then

$$A \ltimes B = A(B \otimes I_t).$$

2. If $n = p$, then the general left semi-tensor product becomes the usual matrix multiplication, i.e.,

$$A \ltimes B = AB.$$

Example 2.17. Let $A = \begin{bmatrix} 2 & 3 \\ -1 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 2}$ and $B = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \in \mathbb{R}^{2\times 2}$.

Then $\alpha = \text{lcm}(2,3) = 6$. Thus,

$$\begin{aligned}
A \ltimes B &= (A \otimes I_{\frac{6}{2}})(B \otimes I_{\frac{6}{3}}) \\
&= (A \otimes I_3)(B \otimes I_2) \\
&= \left(\begin{bmatrix} 2 & 3 \\ -1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right) \left(\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \\
&= \begin{bmatrix} 2 & 0 & 0 & 3 & 0 & 0 \\ 0 & 2 & 0 & 0 & 3 & 0 \\ 0 & 0 & 2 & 0 & 0 & 3 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 3 & 0 \\ 0 & 3 \\ 2 & 0 \\ 0 & 2 \end{bmatrix} \\
&= \begin{bmatrix} 2 & 9 \\ 6 & 2 \\ 6 & 4 \\ -1 & 0 \\ 0 & -1 \\ -3 & 0 \end{bmatrix} \in \mathbb{R}^{6\times 2}.
\end{aligned}$$

The general **left semi-tensor product** has the following important properties as shown in the next theorem.

Theorem 2.18. [10]. Let A, B and C be matrices of such size that can form matrix product, and let $\alpha, \beta \in \mathbb{R}$. Then

1. $A \ltimes (B \ltimes C) = (A \ltimes B) \ltimes C$,
2. $(\alpha A + \beta B) \ltimes C = \alpha(A \ltimes B) + \beta(B \ltimes C)$,
3. $C \ltimes (\alpha A + \beta B) = \alpha(C \ltimes A) + \beta(C \ltimes B)$,
4. $(A \ltimes B)^{'} = B^{'} \ltimes A^{'}$,
5. $(A \ltimes B)^{-1} = B^{-1} \ltimes A^{-1}$.

2.6 Matrix Derivatives

Definition 2.19. Let y be a scalar function of elements of an $n\times1$ vector $\vec{x} = [x_1\ x_2\ \cdots\ x_n]'$. The derivative of y with respect to $\vec{x}$ is an $n \times 1$ vector given by

$$\frac{\partial y}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}.$$

Definition 2.20. Let $\vec{y} = [y_1\ y_2\ \cdots\ y_m]'$. be an $m \times 1$ vector whose elements are differentiable functions of a scalar x. The derivative of $\vec{y}$ with respect to x is a $1 \times m$ vector defined by

$$\frac{\partial \vec{y}}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \cdots & \frac{\partial y_m}{\partial x} \end{bmatrix}.$$

Definition 2.21. Let $\vec{y} = [y_1\ y_2\ \cdots\ y_m]'$. be an $m \times 1$ vector whose elements are differentiable functions of elements of an $n \times 1$ vector $\vec{x} = [x_1\ x_2\ \cdots\ x_n]'$. The derivative of $\vec{y}$ with respect to $\vec{x}$ is defined as an $n \times m$ matrix as follows

$$\frac{\partial \vec{y}}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \frac{\partial y_1}{\partial \vec{x}} & \frac{\partial y_2}{\partial \vec{x}} & \cdots & \frac{\partial y_m}{\partial \vec{x}} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \text{---} & \frac{\partial \vec{y}}{\partial x_1} & \text{---} \\ \text{---} & \frac{\partial \vec{y}}{\partial x_2} & \text{---} \\ & \vdots & \\ \text{---} & \frac{\partial \vec{y}}{\partial x_n} & \text{---} \end{bmatrix}.$$

Definition 2.22. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \vec{a}_{*1} & \vec{a}_{*2} & \cdots & \vec{a}_{*n} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \text{---} & \vec{a}_{1*}' & \text{---} \\ \text{---} & \vec{a}_{2*}' & \text{---} \\ & \vdots & \\ \text{---} & \vec{a}_{m*}' & \text{---} \end{bmatrix}$$

be an $m \times n$ matrix whose element are differentiable functions of element of an $p \times 1$ vector $\vec{x} = [x_1\ x_2\ \cdots\ x_n]'$, where $\vec{a}_{*j}$ is the j-th column of A and $\vec{a}_{i*}'$ is the i-th row of A. The derivative of A with respect to $\vec{x}$ is defined as a $p \times mn$ matrix as follows

$$\frac{\partial A}{\partial \vec{x}} = \frac{\partial\,\mathrm{Vec}(A)}{\partial \vec{x}} = \begin{bmatrix} \text{---} & \frac{\partial\,\mathrm{Vec}(A)}{\partial x_1} & \text{---} \\ \text{---} & \frac{\partial\,\mathrm{Vec}(A)}{\partial x_2} & \text{---} \\ & \vdots & \\ \text{---} & \frac{\partial\,\mathrm{Vec}(A)}{\partial x_p} & \text{---} \end{bmatrix}.$$

Lemma 2.23. [16, Ch. 4]. Let $\vec{x}$ be an $n \times 1$ matrix and let A be a matrix of constants. Then

1. $\dfrac{\partial}{\partial \vec{x}} A\vec{x} \;=\; A'$, for $A, m \times n$,

2. $\dfrac{\partial}{\partial \vec{x}} \vec{x}' A \;=\; A$, for $A, n \times p$,

3. $\dfrac{\partial}{\partial \vec{x}} \vec{x}' A\vec{x} \;=\; (A + A')\vec{x}$, for $A, n \times n$.

Theorem 2.24. [16, Ch. 4]. Let $\vec{x} = [x_1 \;\; x_2 \;\; \cdots \;\; x_p]'$ be a $p \times 1$ vector. Let $\vec{y} = [y_1(\vec{x}) \;\; y_2(\vec{x}) \;\; \cdots \;\; y_q(\vec{x})]'$ and $\vec{z} = [z_1(\vec{y}) \;\; z_2(\vec{y}) \;\; \cdots \;\; z_r(\vec{y})]'$ be $q \times 1$ and $r \times 1$ real-valued vector functions of variables $\vec{x}$ and $\vec{y}$, respectively. Then the chain rule is given by

$$\frac{\partial \vec{z}}{\partial \vec{x}} = \frac{\partial \vec{y}}{\partial \vec{x}} \cdot \frac{\partial \vec{z}}{\partial \vec{y}} \,.$$

Theorem 2.25. [16, Ch. 4]. Let A and B be $m \times n$ and $n \times r$ matrices, respectively. Assume that elements of both A and B are scalar functions of a vector $\vec{x} = [x_1 \; x_2 \; \cdots \; x_p]' \in \mathbb{R}^p$. The product rule of A and B is given by

$$\frac{\partial (A \cdot B)}{\partial \vec{x}} \;=\; \frac{\partial \text{Vec}\, A}{\partial \vec{x}} (B \otimes I_m) + \frac{\partial \text{Vec}\, B}{\partial \vec{x}} (I_r \otimes A').$$

Lemma 2.26. [16, Ch. 4]. Let $\vec{x} = [x_1 \; x_2 \; \cdots \; x_p]' \in \mathbb{R}^p$. Then

1. $\dfrac{\partial}{\partial \vec{x}} (\vec{x} \otimes I_n) \;=\; \dfrac{\partial}{\partial \vec{x}} \text{Vec}(\vec{x} \otimes I_n) \;=\; K_{np}^{\tau_p} \; \in \mathbb{R}^{n \times pn^2}$,

2. $\dfrac{\partial}{\partial \vec{x}} (I_n \otimes \vec{x}) \;=\; \text{Devec}(I_n) \otimes I_p \; \in \mathbb{R}^{p \times pn^2}$,

3. $\dfrac{\partial}{\partial \vec{x}} (\vec{x}' \otimes I_n) \;=\; I_p \otimes \text{Devec}(I_n) \in \mathbb{R}^{p \times pn^2}$.

Chapter 3

Derivatives of matrix-valued functions with vector variables involving semi-tensor products

3.1 Derivatives of matrix-valued functions with vector variables

We derive exact formulas of the derivatives of certain matrix-valued functions involving STPs in a vector variable.

Theorem 3.1. Let $A \in \mathbb{R}^{m\times n}$ be a constant matrix, and let $\vec{x}$ be a $p \times 1$ vector variable. Then

$$\frac{\partial}{\partial \vec{x}}(A \ltimes \vec{x}) \;=\; K^{\tilde{\tau}_p}_{\frac{\alpha}{p},p}\,[I_{\frac{\alpha}{p}} \otimes (A' \otimes I_{\frac{\alpha}{n}})], \tag{3.1.1}$$

$$\frac{\partial}{\partial \vec{x}}(\vec{x}' \ltimes A) \;=\; [I_p \otimes \mathrm{Devec}(I_{\frac{\beta}{p}})](A \otimes I_{\frac{\beta^2}{pm}}), \tag{3.1.2}$$

where $\alpha = \mathrm{lcm}(n,p)$ and $\beta = \mathrm{lcm}(p,m)$.

Proof. From Eq. (1.1.1), we have

$$A \ltimes \vec{x} = (A \otimes I_{\frac{\alpha}{n}})(\vec{x} \otimes I_{\frac{\alpha}{p}}).$$

It follows from the product rule in Theorem 2.25 that

$$\begin{aligned}
\frac{\partial}{\partial \vec{x}}(A \ltimes \vec{x}) &= \frac{\partial}{\partial \vec{x}}[(A \otimes I_{\frac{\alpha}{n}})(\vec{x} \otimes I_{\frac{\alpha}{p}})] \\
&= \frac{\partial (A \otimes I_{\alpha/n})}{\partial \vec{x}}\,[(\vec{x} \otimes I_{\frac{\alpha}{p}}) \otimes I_{\frac{m\alpha}{n}}] + \frac{\partial (\vec{x} \otimes I_{\alpha/p})}{\partial \vec{x}}\,[I_{\frac{\alpha}{p}} \otimes (A' \otimes I_{\frac{\alpha}{n}})] \\
&= \frac{\partial (\vec{x} \otimes I_{\alpha/p})}{\partial \vec{x}}\,[I_{\frac{\alpha}{p}} \otimes (A' \otimes I_{\frac{\alpha}{n}})].
\end{aligned}$$

Now, Lemma 2.26 implies that

$$\frac{\partial}{\partial \vec{x}}(A \ltimes \vec{x}) \;=\; K^{\tilde{\tau}_p}_{\frac{\alpha}{p},p}\,[I_{\frac{\alpha}{p}} \otimes (A' \otimes I_{\frac{\alpha}{n}})].$$

We compute the following derivative according to Theorem 2.25:

$$\begin{aligned}
\frac{\partial}{\partial \vec{x}}(\vec{x}' \ltimes A) &= \frac{\partial}{\partial \vec{x}}[(\vec{x}' \otimes I_{\frac{\beta}{p}})(A \otimes I_{\frac{\beta}{m}})] \\
&= \frac{\partial (\vec{x}' \otimes I_{\beta/p})}{\partial \vec{x}}\,[(A \otimes I_{\frac{\beta}{m}}) \otimes I_{\frac{\beta}{p}}] + \frac{\partial (A \otimes I_{\beta/m})}{\partial \vec{x}}\,[I_{\frac{n\beta}{m}} \otimes (\vec{x} \otimes I_{\frac{\beta}{p}})] \\
&= \frac{\partial (\vec{x}' \otimes I_{\beta/p})}{\partial \vec{x}}\,[(A \otimes I_{\frac{\beta}{m}}) \otimes I_{\frac{\beta}{p}}] \\
&= \frac{\partial (\vec{x}' \otimes I_{\beta/p})}{\partial \vec{x}}\,[A \otimes I_{\frac{\beta^2}{pm}}].
\end{aligned}$$

Now, Lemma 2.26 implies that

$$\frac{\partial}{\partial \vec{x}}(\vec{x}' \ltimes A) \;=\; [I_p \otimes \mathrm{Devec}(I_{\frac{\beta}{p}})](A \otimes I_{\frac{\beta^2}{pm}}).$$

□

Corollary 3.2.

1. If $p = nt$ for some $t \in \mathbb{N}$, then Eq. (3.1.1) becomes
$$\frac{\partial}{\partial \vec{x}}(A \ltimes \vec{x}) \; = \; A' \otimes I_t.$$
2. If $n = pt$ for some $t \in \mathbb{N}$, then Eq. (3.1.1) becomes
$$\frac{\partial}{\partial \vec{x}}(A \ltimes \vec{x}) \; = \; K_{t,p}^{\tilde{\tau}_p} \, (I_t \otimes A').$$
3. If $m = pt$ for some $t \in \mathbb{N}$, then Eq. (3.1.2) becomes
$$\frac{\partial}{\partial \vec{x}}(\vec{x}' \ltimes A) \; = \; [I_p \otimes \text{Devec}(I_t)](A \otimes I_t).$$
4. If $p = mt$ for some $t \in \mathbb{N}$, then Eq. (3.1.2) becomes
$$\frac{\partial}{\partial \vec{x}}(\vec{x}' \ltimes A) \; = \; A \otimes I_t.$$

Proof. For the case $p = nt$, we have $\alpha = \text{lcm}(n, p) = p$. By substituting $\alpha = p$ in (3.1.1) and applying Lemma 2.14, we obtain
$$\begin{aligned}\frac{\partial}{\partial \vec{x}}(A \ltimes \vec{x}) \; &= \; K_{1,p}^{\tilde{\tau}_p} \, [I_1 \otimes (A' \otimes I_t)] \\ &= \; I_p \, [A' \otimes I_t] \\ &= \; A' \otimes I_t.\end{aligned}$$
For the case $n = pt$, we have $\alpha = \text{lcm}(n, p) = n$ and thus
$$\frac{\partial}{\partial \vec{x}}(A \ltimes \vec{x}) \; = \; K_{t,p}^{\tilde{\tau}_p} \, (I_t \otimes A').$$
Similarly, the remaining results can be done in the same manner by substituting $\beta = \text{lcm}(p, m)$. $\square$

Theorem 3.3. Let $A \in \mathbb{R}^{m \times n}$ be a constant matrix, and let $\vec{x} = \vec{x}(\vec{z})$ be a $p \times 1$ vector function of an $r \times 1$ vector $\vec{z}$. Then
$$\frac{\partial}{\partial \vec{z}}(A \ltimes \vec{x}(\vec{z})) = \frac{\partial \vec{x}}{\partial \vec{z}} \, K_{\frac{\alpha}{p},p}^{\tilde{\tau}_p} \, [I_{\frac{\alpha}{p}} \otimes (A' \otimes I_{\frac{\alpha}{n}})], \tag{3.1.3}$$
$$\frac{\partial}{\partial \vec{z}}(\vec{x}'(\vec{z}) \ltimes A) = \frac{\partial \vec{x}}{\partial \vec{z}} \, [I_p \otimes \text{Devec}(I_{\frac{\alpha}{p}})] \, (A \otimes I_{\frac{\alpha^2}{pm}}), \tag{3.1.4}$$
where $\alpha = \text{lcm}(n, p)$ and $\beta = \text{lcm}(p, m)$.

Proof. The results follow directly by applying Theorem 2.24 to (3.1.1) and (3.1.2) in Theorem 3.1. $\square$

Corollary 3.4.

1. If $p = nt$ for some $t \in \mathbb{N}$, then Eq. (3.1.3) becomes
$$\frac{\partial}{\partial \vec{z}}(A \ltimes \vec{x}(\vec{z})) \; = \; \frac{\partial \vec{x}}{\partial \vec{z}} \, (A' \otimes I_t).$$

2. If $n = pt$ for some $t \in \mathbb{N}$, then Eq. (3.1.3) becomes

$$\frac{\partial}{\partial \vec{z}}(A \ltimes \vec{x}(\vec{z})) \;=\; \frac{\partial \vec{x}}{\partial \vec{z}}\, K_{t,p}^{\tilde{\tau}_p}\,(I_t \otimes A').$$

3. If $p = nt$ for some $t \in \mathbb{N}$, then Eq. (3.1.4) becomes

$$\frac{\partial}{\partial \vec{z}}(\vec{x}'(\vec{z}) \ltimes A) \;=\; \frac{\partial \vec{x}}{\partial \vec{z}}\,[I_p \otimes \mathsf{Devec}(I_t)]\,(A \otimes I_t).$$

4. If $n = pt$ for some $t \in \mathbb{N}$, then Eq. (3.1.4) becomes

$$\frac{\partial}{\partial \vec{z}}(\vec{x}'(\vec{z}) \ltimes A) \;=\; \frac{\partial \vec{x}}{\partial \vec{z}}\,(A \otimes I_t).$$

Proof. The results can be done in the same manner as those in Corollary 3.2. □

Theorem 3.5. Let $A \in \mathbb{R}^{m\times n}$ be a constant matrix. Let $\vec{x}$ and $\vec{y}$ be two independent vector variables of dimension $p \times 1$ and $q \times 1$, respectively. Then

$$\frac{\partial}{\partial \vec{x}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; K_{\frac{\beta}{p},p}^{\tilde{\tau}_p}\,[I_{\frac{\beta}{p}} \otimes ((A' \ltimes \vec{y}) \otimes I_{\frac{\beta m}{n\alpha}})], \tag{3.1.5}$$

where $\alpha = \mathrm{lcm}(q,m)$ and $\beta = \mathrm{lcm}(n\alpha/m, p)$.

Proof. Let $M = \vec{y}' \ltimes A = (\vec{y}' \otimes I_{\frac{\alpha}{q}})(A \otimes I_{\frac{\alpha}{m}})$. Then

$$\begin{aligned}(\vec{y}' \ltimes A) \ltimes \vec{x} \;&=\; [(\vec{y}' \otimes I_{\frac{\alpha}{q}})(A \otimes I_{\frac{\alpha}{m}})] \ltimes \vec{x}\\ &=\; (M \otimes I_{\frac{\beta m}{n\alpha}})(\vec{x} \otimes I_{\frac{\beta}{p}}).\end{aligned}$$

By taking derivative with respect to $\vec{x}$ and applying Theorem 2.25, it follows that

$$\begin{aligned}\frac{\partial}{\partial \vec{x}}(\vec{y}' \ltimes A \ltimes \vec{x}) &= \frac{\partial (M \otimes I_{\beta m/n\alpha})}{\partial \vec{x}}\,[(\vec{x} \otimes I_{\frac{\beta}{p}}) \otimes I_{\frac{\beta m}{q}}] + \frac{\partial (\vec{x} \otimes I_{\beta/p})}{\partial \vec{x}}\,[I_{\frac{\beta}{p}} \otimes (M' \otimes I_{\frac{\beta m}{n\alpha}})]\\ &= \frac{\partial (\vec{x} \otimes I_{\beta/p})}{\partial \vec{x}}\,[I_{\frac{\beta}{p}} \otimes (M' \otimes I_{\frac{\beta m}{n\alpha}})]\\ &= K_{\frac{\beta}{p},p}^{\tilde{\tau}_p}\,[I_{\frac{\beta}{p}} \otimes (M' \otimes I_{\frac{\beta m}{n\alpha}})].\end{aligned}$$

□

Corollary 3.6.

1. If a pair (m,q) is relatively prime, and $p = t(n\alpha/m)$ for some $t \in \mathbb{N}$, then Eq. (3.1.5) becomes

$$\frac{\partial}{\partial \vec{x}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; (A' \ltimes \vec{y}) \otimes I_t.$$

2. If a pair (m,q) is relatively prime, and $n\alpha/m = pt$ for some $t \in \mathbb{N}$, then Eq. (3.1.5) becomes

$$\frac{\partial}{\partial \vec{x}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; K_{t,p}^{\tilde{\tau}_p} \cdot [I_t \otimes (A' \ltimes \vec{y})]$$

3. If $m = qt$ for some $t \in \mathbb{N}$, and a pair $(n\alpha/m, p)$ is relatively prime, then Eq. (3.1.5) becomes

$$\frac{\partial}{\partial \vec{x}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; K^{\tilde{\tau}_p}_{n,p} \cdot [I_n \otimes [A'(\mathsf{y} \otimes \mathsf{I}_t)] \otimes I_p]$$

4. If $q = mt$ for some $t \in \mathbb{N}$, and a pair $(n\alpha/m, p)$ is relatively prime, then Eq. (3.1.5) becomes

$$\frac{\partial}{\partial \vec{x}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; K^{\tilde{\tau}_p}_{nt,p} \cdot [I_{nt} \otimes [(A' \otimes I_s)\vec{y}] \otimes I_p]$$

5. If $m = qt$ and $p = s(n\alpha/m)$ for some $t, s \in \mathbb{N}$, then Eq. (3.1.5) becomes

$$\frac{\partial}{\partial \vec{x}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; [A'(\mathsf{y} \otimes \mathsf{I}_t)] \otimes I_s.$$

6. If $m = qt$ and $n\alpha/m = ps$ for some $t, s \in \mathbb{N}$, then Eq. (3.1.5) becomes

$$\frac{\partial}{\partial \vec{x}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; K^{\tilde{\tau}_p}_{s,p} \cdot [I_s \otimes [A'(\mathsf{y} \otimes \mathsf{I}_t)]].$$

7. If $q = mt$ and $p = s(n\alpha/m)$ for some $t, s \in \mathbb{N}$, then Eq. (3.1.5) becomes

$$\frac{\partial}{\partial \vec{x}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; (A' \otimes I_t)\vec{y} \otimes I_s.$$

8. If $q = mt$ and $n\alpha/m = ps$ for some $t, s \in \mathbb{N}$, then Eq. (3.1.5) becomes

$$\frac{\partial}{\partial \vec{x}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; K^{\tilde{\tau}_p}_{s,p} \cdot [I_s \otimes [(A' \otimes I_s)\vec{y}]].$$

Proof. The condition that the pair (m, q) is relatively prime means that $\alpha = mq$. The condition that the pair $(n\alpha/m, p)$ is relatively prime means that $\beta = n\alpha p/m$. Now, the results follow from Eq. (3.1.5). □

Theorem 3.7. Let $A \in \mathbb{R}^{m\times n}$ be a constant matrix and let $\vec{x}$ and $\vec{y}$ be $p \times 1$ and $q \times 1$ vector functions, respectively. Then

$$\frac{\partial}{\partial \vec{y}}(\vec{y}' \ltimes A \ltimes \vec{x}) = (I_q \otimes \mathrm{Devec}(I_{\frac{\beta}{q}}))((A \ltimes \vec{x}) \otimes I_{\frac{\beta^2 n}{m\alpha q}}), \tag{3.1.6}$$

where $\alpha = \mathrm{lcm}(n, p)$ and $\beta = \mathrm{lcm}(q, m\alpha/n)$.

Proof. Let $N = A \ltimes \vec{x} = (A \otimes I_{\frac{\alpha}{n}})(\vec{x} \otimes I_{\frac{\alpha}{p}})$. Then

$$\begin{aligned}\vec{y}' \ltimes (A \ltimes \vec{x}) \;&=\; \vec{y}' \ltimes [(A \otimes I_{\frac{\alpha}{n}})(\vec{x} \otimes I_{\frac{\alpha}{p}})] \\ &=\; (\vec{y}' \otimes I_{\frac{\beta}{q}})(N \otimes I_{\frac{\beta n}{m\alpha}}).\end{aligned}$$

By taking derivative with respect to $\vec{y}$ and applying Theorem 2.25, we get

$$\begin{aligned}\frac{\partial}{\partial \vec{y}}(\vec{y}' \ltimes A \ltimes \vec{x}) &= \frac{\partial(\vec{y}' \otimes I_{\beta/q})}{\partial \vec{y}}\,[(N \otimes I_{\frac{\beta n}{m\alpha}}) \otimes I_{\frac{\beta}{q}}] + \frac{\partial(N \otimes I_{\beta n/m\alpha})}{\partial \vec{y}}\,[I_{\frac{\beta n}{pm}} \otimes (\vec{y} \otimes I_{\frac{\beta}{q}})] \\ &= \frac{\partial(\vec{y}' \otimes I_{\beta/q})}{\partial \vec{y}}\,[(N \otimes I_{\frac{\beta n}{m\alpha}}) \otimes I_{\frac{\beta}{q}}] \\ &= \frac{\partial(\vec{y}' \otimes I_{\beta/q})}{\partial \vec{y}}\,[N \otimes I_{\frac{\beta^2 n}{m\alpha q}}].\end{aligned}$$

Finally, Lemma 2.26 implies that

$$\frac{\partial}{\partial \vec{y}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; (I_q \otimes \text{Devec}(I_{\frac{\beta}{q}}))(N \otimes I_{\frac{\beta^2 n}{m\alpha q}}).$$

□

Corollary 3.8.

1. If a pair (n, p) is relatively prime, and $m\alpha/n = qt$ for some $t \in \mathbb{N}$, then Eq. (3.1.6) becomes
$$\frac{\partial}{\partial \vec{y}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; [I_q \otimes \text{Devec}(I_t)]\,((A \ltimes \vec{x}) \otimes I_t).$$
2. If a pair (n, p) is relatively prime, and $q = t(m\alpha/n)$ for some $t \in \mathbb{N}$, then Eq. (3.1.6) becomes
$$\frac{\partial}{\partial \vec{y}}(\vec{y}' \ltimes A \ltimes \vec{x}) = (A \ltimes \vec{x}) \otimes I_t.$$
3. If $p = nt$ for some $t \in \mathbb{N}$, and a pair $(q, m\alpha/n)$ is relatively prime, then Eq. (3.1.6) becomes
$$\frac{\partial}{\partial \vec{y}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; [I_q \otimes \text{Devec}(I_{mt})][(A \otimes I_t)\vec{x} \otimes I_{qmt}].$$
4. If $n = pt$ for some $t \in \mathbb{N}$, and a pair $(q, m\alpha/n)$ is relatively prime, then Eq. (3.1.6) becomes
$$\frac{\partial}{\partial \vec{y}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; [I_q \otimes \text{Devec}(I_m)]\,[A(\vec{x} \otimes I_t) \otimes I_{qm}].$$
5. If $p = nt$ and $(m\alpha/n) = qs$ for some $t, s \in \mathbb{N}$, then Eq. (3.1.6) becomes
$$\frac{\partial}{\partial \vec{y}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; [I_q \otimes \text{Devec}(I_s)]\,((A \otimes I_t)\vec{x} \otimes I_s).$$
6. If $p = nt$ and $q = s(m\alpha/n)$ for some $t, s \in \mathbb{N}$, then Eq. (3.1.6) becomes
$$\frac{\partial}{\partial \vec{y}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; ((A \otimes I_t)\vec{x} \otimes I_s).$$
7. If $n = pt$ and $m\alpha/n = qs$ for some $t, s \in \mathbb{N}$, then Eq.(3.1.6) becomes
$$\frac{\partial}{\partial \vec{y}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; [I_q \otimes \text{Devec}(I_s)]\,((A \otimes I_t)\vec{x} \otimes I_s).$$
8. If $n = pt$ and $q = s(m\alpha/n)$ for some $t, s \in \mathbb{N}$, then Eq. (3.1.6) becomes
$$\frac{\partial}{\partial \vec{y}}(\vec{y}' \ltimes A \ltimes \vec{x}) \;=\; A(\vec{x} \otimes I_t) \otimes I_s.$$

Proof. The condition that the pair (n, p) is relatively prime means $\alpha = np$. The condition that the pair $(q, m\alpha/n)$ is relatively prime means $\beta = qm\alpha/n$. Now, the results follow from Eq. (3.1.6). □

Theorem 3.9. Let $A \in \mathbb{R}^{m\times n}$ be a constant matrix, and let $\vec{x}$ be a $p \times 1$ vector variable. Then

$$\frac{\partial}{\partial \vec{x}}(\vec{x}' \ltimes A \ltimes \vec{x}) = [I_p \otimes \text{Devec}(I_{\frac{\alpha}{p}})]\,[(A \otimes I_{\frac{\alpha^2}{mp}})(\vec{x} \otimes I_{\frac{t\alpha}{p}})] + K_{t,p}^{\tilde{\tau}_p} \cdot [I_t \otimes (A' \ltimes \vec{x})], \qquad (3.1.7)$$

where $\alpha = \text{lcm}(p,m)$ and $n\alpha/m = pt$ for some $t \in \mathbb{N}$. Particularly, if $p = m = n$, then Eq. (3.1.7) becomes the item 3) in Lemma 2.23.

Proof. Since $A' \ltimes \vec{x} = (A' \otimes I_{\frac{\alpha}{m}})(\vec{x} \otimes I_{\frac{\alpha}{p}})$, it follows that

$$(\vec{x}' \ltimes A) \ltimes \vec{x} \;=\; [(\vec{x} \otimes I_{\frac{\alpha}{p}})(A \otimes I_{\frac{\alpha}{m}})]\,(\vec{x} \otimes I_t).$$

By taking derivative with respect to $\vec{x}$ and applying Theorem 2.25, we get

$$\begin{aligned}
\frac{\partial}{\partial \vec{x}}(\vec{x}' \ltimes A \ltimes \vec{x}) = {} & \frac{\partial[(\vec{x}' \otimes I_{\alpha/p})(A \otimes I_{\alpha/m})]}{\partial \vec{x}}\,[(\vec{x} \otimes I_t) \otimes I_{\frac{\alpha}{p}}] \\
& + \frac{\partial(\vec{x} \otimes I_t)}{\partial \vec{x}}\,[I_t \otimes (A' \otimes I_{\frac{\alpha}{m}})(\vec{x} \otimes I_{\frac{\alpha}{p}})] \\
= {} & \Bigg[\frac{\partial(\vec{x}' \otimes I_{\alpha/p})}{\partial \vec{x}}[(A \otimes I_{\frac{\alpha}{m}}) \otimes I_{\frac{\alpha}{p}}] \\
& + \frac{\partial(A \otimes I_{\alpha/m})}{\partial \vec{x}}[I_{\frac{\alpha n}{m}} \otimes (\vec{x} \otimes I_{\frac{\alpha}{p}})]\Bigg][(\vec{x} \otimes I_{\frac{t\alpha}{p}})] \\
& + \frac{\partial(\vec{x} \otimes I_t)}{\partial \vec{x}}\,[I_t \otimes (A' \ltimes \vec{x})].
\end{aligned}$$

Now, Lemma 2.26 implies that

$$\frac{\partial}{\partial \vec{x}}(\vec{x}' \ltimes A \ltimes \vec{x}) = [I_p \otimes \text{Devec}(I_{\frac{\alpha}{p}})]\,[(A \otimes I_{\alpha^2/mp})(\vec{x} \otimes I_{t\alpha/p})] + K_{t,p}^{\tilde{\tau}_p} \cdot [I_t \otimes (A' \ltimes \vec{x})].$$

For the special case $p = m = n$, we have $\alpha = p$ and $t = 1$. Thus, we obtain

$$\begin{aligned}
\frac{\partial}{\partial \vec{x}}(\vec{x}' A \vec{x}) &= [I_\alpha \otimes \text{Devec}(I_{\frac{\alpha}{\alpha}})]\,[(A \otimes I_{\frac{\alpha^2}{\alpha^2}})(\vec{x} \otimes I_1)] + K_{1,\alpha}^{\tilde{\tau}_\alpha} \cdot [I_1 \otimes (A' \otimes I_{\frac{\alpha}{\alpha}})(\vec{x} \otimes I_{\frac{\alpha}{\alpha}})]. \\
&= I_\alpha A \vec{x} + I_\alpha A' \vec{x} \\
&= (A + A')\vec{x}.
\end{aligned}$$

□

Theorem 3.10. Let $A \in \mathbb{R}^{m\times n}$ be a constant matrix, and let $\vec{x} = \vec{x}(\vec{z})$ be a $p \times 1$ vector function of an $r \times 1$ vector $\vec{z}$. Then

$$\begin{aligned}
\frac{\partial}{\partial \vec{z}}(\vec{x}'(\vec{z}) \ltimes A \ltimes \vec{x}(\vec{z})) = {} & \frac{\partial \vec{x}}{\partial \vec{z}}\,[I_p \otimes \text{Devec}(I_{\frac{\alpha}{p}})]\,[(A \otimes I_{\frac{\alpha^2}{mp}})(\vec{x} \otimes I_{\frac{t\alpha}{p}})] \\
& + \frac{\partial \vec{x}}{\partial \vec{z}}\,K_{t,p}^{\tilde{\tau}_p} \cdot [I_t \otimes (A' \ltimes \vec{x})]
\end{aligned}$$

where $\alpha = \text{lcm}(p,m)$ and $n\alpha/m = pt$ for some $t \in \mathbb{N}$.

Proof. By applying Theorem 2.24 on the formula (3.1.7), we can get the result. □

The results in this section generalize the classical results (e.g. [16, Ch. 4]) in the literature, particularly Lemma 2.23.

3.2 Neural networks applications

In a neural network, suppose we have an n-component input vector $\vec{x} \in \mathbb{R}^n$. In order to train a neuron, we choose a weight vector $\vec{w} \in \mathbb{R}^n$ with the same component number as that for the input vector. We also need a scalar bias $b \in \mathbb{R}$. Then, the activation of a single computation unit in a neuron is typically calculated as

$$\mathsf{F}(\vec{x}) \;=\; \sum_{i=1}^{n} w_i x_i + b \;=\; \vec{w}'\vec{x} + b.$$

The function F is known as the unit's affine function. To train this neuron, we choose weights $\vec{w}$ and the bias b that minimize an associated loss function. To minimize the loss function, we use matrix derivatives.

Now, suppose we have t collections of an n-component data. We can represent them with a single vector

$$\vec{x} \;=\; [x_1 \;\cdots\; x_n \;\cdots\; x_{n(t-1)+1} \;\cdots\; x_{nt}]' \in \mathbb{R}^{nt}.$$

Assume that

- we use the same weights for each data collection, namely,

$$\vec{w} \;=\; [w_1 \;\cdots\; w_n]' \in \mathbb{R}^n.$$

- we use different bias for different data sets, so we can form the bias vector to be

$$\vec{b} \;=\; [b_1 \;\cdots\; b_t]' \in \mathbb{R}^t.$$

Thus, the affine function is given by

$$\begin{aligned}\mathsf{F}(\vec{x}) \;&=\; (\vec{w}' \ltimes \vec{x}) + \vec{b} \\ &=\; \begin{bmatrix} \sum_{i=1}^{n} \vec{w}_i \vec{x}_{(i-1)t+1} + b_1 \\ \sum_{i=1}^{n} \vec{w}_i \vec{x}_{(i-1)t+2} + b_2 \\ \vdots \\ \sum_{i=1}^{n} \vec{w}_i \vec{x}_{(i-1)t+t} + b_t \end{bmatrix}.\end{aligned}$$

To minimize the associated loss function, we shall differentiate F with respect to $\vec{x}$, $\vec{w}$, and $\vec{b}$. Indeed, from Corollary 3.2, we obtain

$$\begin{aligned}\frac{\partial \mathsf{F}}{\partial \vec{x}} \;&=\; \frac{\partial}{\partial \vec{x}}(\vec{w}' \ltimes \vec{x}) + \frac{\partial}{\partial \vec{x}} b \\ &=\; (\vec{w}')' \otimes I_t \\ &=\; \vec{w} \otimes I_t\end{aligned}$$

The same corollary implies.

$$\begin{aligned}\frac{\partial \mathsf{F}}{\partial \vec{w}} \;&=\; \frac{\partial}{\partial \vec{w}}(\vec{w}' \ltimes \vec{x}) + \frac{\partial}{\partial \vec{w}} b \\ &=\; [I_n \otimes \mathsf{Devec}(I_t)](\vec{x} \otimes I_t).\end{aligned}$$

The gradient of F with respect to the bias is given by

$$\frac{\partial \mathsf{F}}{\partial \vec{b}} \;=\; \frac{\partial}{\partial \vec{b}}(\vec{w}' \ltimes \vec{x}) + \frac{\partial}{\partial \vec{b}}\vec{b} \;=\; I_t.$$

Chapter 4

Product rules

Previously, we only consider the formulation where A is a constant matrix. In general, A can be a matrix function of any vector variables and it is indeed more complicated to derive the matrix derivative formula. For convenience, let us first define the notation when working with an $m \times n$ matrix function $A(\vec{x})$, where $\vec{x} = [x_1\ x_2\ \cdots\ x_p]'$ is a $p \times 1$ vector variable. Suppose that

$$
\begin{aligned}
A(\vec{x}) &= \begin{bmatrix} a_{11}(\vec{x}) & a_{12}(\vec{x}) & \cdots & a_{1n}(\vec{x}) \\ a_{21}(\vec{x}) & a_{22}(\vec{x}) & \cdots & a_{2n}(\vec{x}) \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}(\vec{x}) & a_{m2}(\vec{x}) & \cdots & a_{mn}(\vec{x}) \end{bmatrix} \\
&= \begin{bmatrix} | & | & & | \\ \vec{a}_{*1}(\vec{x}) & \vec{a}_{*2}(\vec{x}) & \cdots & \vec{a}_{*n}(\vec{x}) \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \text{---} & \vec{a}_{1*}'(\vec{x}) & \text{---} \\ \text{---} & \vec{a}_{2*}'(\vec{x}) & \text{---} \\ & \vdots & \\ \text{---} & \vec{a}_{m*}'(\vec{x}) & \text{---} \end{bmatrix}
\end{aligned}
$$

4.1 A matrix derivative result involving Tracy-Singh products

Consider the following derivative

$$
\frac{\partial}{\partial \vec{x}} A(\vec{x}) = \begin{bmatrix} \text{---} & \dfrac{\partial\, \mathrm{Devec}(A(\vec{x}))}{\partial x_1} & \text{---} \\ \text{---} & \dfrac{\partial\, \mathrm{Devec}(A(\vec{x}))}{\partial x_2} & \text{---} \\ & \vdots & \\ \text{---} & \dfrac{\partial\, \mathrm{Devec}(A(\vec{x}))}{\partial x_p} & \text{---} \end{bmatrix}.
$$

The above expression can be viewed as n blocks shown below

$$
\begin{aligned}
\frac{\partial}{\partial \vec{x}} A(\vec{x}) &= \begin{bmatrix} \dfrac{\partial}{\partial x_1}\mathrm{a}_{*1}'(\vec{x}) & \dfrac{\partial}{\partial x_1}\mathrm{a}_{*2}'(\vec{x}) & \cdots & \dfrac{\partial}{\partial x_1}\mathrm{a}_{*n}'(\vec{x}) \\ \dfrac{\partial}{\partial x_2}\mathrm{a}_{*1}'(\vec{x}) & \dfrac{\partial}{\partial x_2}\mathrm{a}_{*2}'(\vec{x}) & \cdots & \dfrac{\partial}{\partial x_2}\mathrm{a}_{*n}'(\vec{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial}{\partial x_p}\mathrm{a}_{*1}'(\vec{x}) & \dfrac{\partial}{\partial x_p}\mathrm{a}_{*2}'(\vec{x}) & \cdots & \dfrac{\partial}{\partial x_p}\mathrm{a}_{*n}'(\vec{x}) \end{bmatrix} \\
&= \left[\left(\frac{\partial}{\partial \vec{x}} A(\vec{x})\right)_{*1} \left(\frac{\partial}{\partial \vec{x}} A(\vec{x})\right)_{*2} \cdots \left(\frac{\partial}{\partial \vec{x}} A(\vec{x})\right)_{*n} \right] \\
&= \left[\left(\frac{\partial}{\partial \vec{x}} A(\vec{x})\right)_{*j} \right]_{j=1}^{n}.
\end{aligned}
$$

From those we mentioned earlier, they allows us to formulate the following Lemma.

The Tracy-Singh product of these matrices, denoted by $\frac{\partial A(\vec{x})}{\partial \vec{x}} \boxtimes \text{Devec}(I_t)$, results in a block matrix where each block is given by the Kronecker product of the respective blocks from $\frac{\partial A(\vec{x})}{\partial \vec{x}} \boxtimes \text{Devec}(I_t)$.

Lemma 4.1. Let A($\vec{x}$) be an $m \times n$ matrix function of a vector $\vec{x} \in \mathbb{R}^p$. Then

1. $\frac{\partial}{\partial \vec{x}}(A(\vec{x}) \otimes I_t)) = \left[\left[\left(\frac{\partial}{\partial \vec{x}} A(\vec{x})\right)_{*j} \otimes e_r^{t'}\right]_{r=1}^{t}\right]_{j=1}^{n} = \frac{\partial A(\vec{x})}{\partial \vec{x}} \boxtimes \text{Devec}(I_t)$.

Proof. By using direct computation, we have that

$$\begin{aligned}
A(\vec{x}) \otimes I_t &= \begin{bmatrix} \mathsf{a}_{*1}(\vec{x}) \otimes I_t & \mathsf{a}_{*2}(\vec{x}) \otimes I_t & \cdots & \mathsf{a}_{*n}(\vec{x}) \otimes I_t \end{bmatrix} \\
&= \left[\left[\mathsf{a}_{*1}(\vec{x}) \otimes e_r^t\right]_{r=1}^{t} \; \left[\mathsf{a}_{*2}(\vec{x}) \otimes e_r^t\right]_{r=1}^{t} \cdots \left[\mathsf{a}_{*n}(\vec{x}) \otimes e_r^t\right]_{r=1}^{t}\right]_{j=1}^{n} \\
&= \left[\left[\mathsf{a}_{*j}(\vec{x}) \otimes e_r^t\right]_{r=1}^{t}\right]_{j=1}^{n}.
\end{aligned}$$

By taking derivative with respect to $\vec{x}$, the above equation becomes

$$\begin{aligned}
\frac{\partial}{\partial \vec{x}}(A(\vec{x}) \otimes I_t) &= \begin{bmatrix} \frac{\partial}{\partial x_1} \text{Devec} \left[\left[\mathsf{a}_{*j}(\vec{x}) \otimes e_r^t\right]_{r=1}^{t}\right]_{j=1}^{n} \\ \frac{\partial}{\partial x_2} \text{Devec} \left[\left[\mathsf{a}_{*j}(\vec{x}) \otimes e_r^t\right]_{r=1}^{t}\right]_{j=1}^{n} \\ \vdots \\ \frac{\partial}{\partial x_p} \text{Devec} \left[\left[\mathsf{a}_{*j}(\vec{x}) \otimes e_r^t\right]_{r=1}^{t}\right]_{j=1}^{n} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial}{\partial x_1} \left[\left[\mathsf{a}'_{*j}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t}\right]_{j=1}^{n} \\ \frac{\partial}{\partial x_2} \left[\left[\mathsf{a}'_{*j}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t}\right]_{j=1}^{n} \\ \vdots \\ \frac{\partial}{\partial x_p} \left[\left[\mathsf{a}'_{*j}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t}\right]_{j=1}^{n} \end{bmatrix}.
\end{aligned}$$

Hence,

$$\frac{\partial}{\partial \vec{x}}(A(\vec{x}) \otimes I_t) = \begin{bmatrix} \frac{\partial}{\partial x_1}\left[\mathsf{a}'_{*1}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} & \frac{\partial}{\partial x_1}\left[\mathsf{a}'_{*2}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} & \cdots & \frac{\partial}{\partial x_1}\left[\mathsf{a}'_{*n}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} \\ \frac{\partial}{\partial x_2}\left[\mathsf{a}'_{*1}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} & \frac{\partial}{\partial x_2}\left[\mathsf{a}'_{*2}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} & \cdots & \frac{\partial}{\partial x_2}\left[\mathsf{a}'_{*n}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} \\ \vdots & \vdots & \ddots & \\ \frac{\partial}{\partial x_p}\left[\mathsf{a}'_{*1}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} & \frac{\partial}{\partial x_p}\left[\mathsf{a}'_{*2}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} & \cdots & \frac{\partial}{\partial x_p}\left[\mathsf{a}'_{*n}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} \end{bmatrix}$$

$$= \begin{bmatrix} \left[\frac{\partial}{\partial x_1}\mathsf{a}'_{*1}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} & \left[\frac{\partial}{\partial x_1}\mathsf{a}'_{*2}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} & \cdots & \left[\frac{\partial}{\partial x_1}\mathsf{a}'_{*n}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} \\ \left[\frac{\partial}{\partial x_2}\mathsf{a}'_{*1}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} & \left[\frac{\partial}{\partial x_2}\mathsf{a}'_{*2}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} & \cdots & \left[\frac{\partial}{\partial x_2}\mathsf{a}'_{*n}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} \\ \vdots & \vdots & \ddots & \\ \left[\frac{\partial}{\partial x_p}\mathsf{a}'_{*1}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} & \left[\frac{\partial}{\partial x_p}\mathsf{a}'_{*2}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} & \cdots & \left[\frac{\partial}{\partial x_p}\mathsf{a}'_{*n}(\vec{x}) \otimes e_r^{t'}\right]_{r=1}^{t} \end{bmatrix}$$

$$= \left[\left[\left(\frac{\partial}{\partial \vec{x}}A(\vec{x})\right)_{*1} \otimes e_r^{t'}\right]_{r=1}^{t} \quad \left[\left(\frac{\partial}{\partial \vec{x}}A(\vec{x})\right)_{*2} \otimes e_r^{t'}\right]_{r=1}^{t} \quad \cdots \quad \left[\left(\frac{\partial}{\partial \vec{x}}A(\vec{x})\right)_{*n} \otimes e_r^{t'}\right]_{r=1}^{t}\right]$$

$$= \left[\left[\left(\frac{\partial}{\partial \vec{x}}A(\vec{x})\right)_{*j} \otimes e_r^{t'}\right]_{r=1}^{t}\right]_{j=1}^{n}$$

$$= \frac{\partial A(\vec{x})}{\partial \vec{x}} \boxtimes \mathsf{Devec}(I_t).$$

□

4.2 Product rules

In this section, we investigate the derivative of the semi-tensor product between two matrix functions with respect to a vector variable.

Theorem 4.2. Let $A \in \mathbb{R}^{m\times n}$ be a constant matrix, and let $\vec{x} = \vec{x}(\vec{z})$ and $\vec{y} = \vec{y}(\vec{z})$ be $p \times 1$ and $q \times 1$ vector functions of an $r \times 1$ vector $\vec{z}$. Then

$$\frac{\partial}{\partial \vec{z}}(\vec{y}'(\vec{z}) \ltimes A \ltimes \vec{x}(\vec{z})) = \frac{\partial \vec{y}}{\partial \vec{z}} \cdot \left[\frac{\partial(\vec{y}' \ltimes A)}{\partial \vec{y}} \boxtimes \mathsf{Devec}(I_{\frac{\beta}{nm}})\right][\vec{x}(\vec{z}) \otimes I_{\frac{\beta^2}{pnm}}] + \frac{\partial \vec{x}}{\partial \vec{z}} \cdot K^{\tilde{\tau}_p}_{\frac{\beta}{p},p}\,[I_{\frac{\beta}{p}} \otimes [(A' \ltimes \vec{y}(\vec{z})) \otimes I_{\frac{\beta}{nm}}]], \tag{4.2.1}$$

where $\alpha = \mathrm{lcm}(q, m)$ and $\beta = \mathrm{lcm}(nm, p)$.

Proof. For the formula (1.1.1), we have q, m, n and p, we obtain that

$$\begin{aligned} \vec{y}'(\vec{z}) \ltimes A \ltimes \vec{x}(\vec{z}) &= [(\vec{y}'(\vec{z}) \otimes I_{\frac{\alpha}{q}})(A \otimes I_{\frac{\alpha}{m}})] \ltimes \vec{x}(\vec{z}) \\ &= \left[[(\vec{y}'(\vec{z}) \otimes I_{\frac{\alpha}{q}})(A \otimes I_{\frac{\alpha}{m}})] \otimes I_{\frac{\beta}{nm}}\right](\vec{x}(\vec{z}) \otimes I_{\frac{\beta}{p}}) \\ &= [(\vec{y}'(\vec{z}) \ltimes A) \otimes I_{\frac{\beta}{nm}}]\,(\vec{x}(\vec{z}) \otimes I_{\frac{\beta}{p}}). \end{aligned}$$

Taking derivative with respect to $\vec{z}$ yields

$$\frac{\partial}{\partial \vec{z}}(\vec{y}'(\vec{z}) \ltimes A \ltimes \vec{x}(\vec{z})) = \frac{\partial[(\vec{y}'(\vec{z}) \ltimes A) \otimes I_{\beta/mn}]}{\partial \vec{z}}\,[\vec{x}(\vec{z}) \otimes I_{\frac{\beta}{p}} \otimes I_{\frac{\beta}{nm}}] + \frac{\partial(\vec{x}(\vec{z}) \otimes I_{\beta/p})}{\partial \vec{z}}\,[I_{\frac{\beta}{p}} \otimes [(A' \ltimes \vec{y}(\vec{z})) \otimes I_{\frac{\beta}{nm}}]].$$

Now, Theorem 2.24 implies

$$\frac{\partial}{\partial\vec{z}}(\vec{y}'(\vec{z}) \ltimes A \ltimes \vec{x}(\vec{z})) = \frac{\partial\vec{y}}{\partial\vec{z}} \cdot \frac{\partial[(\vec{y}'(\vec{z}) \ltimes A) \otimes I_{\beta/nm}]}{\partial\vec{y}} [\vec{x}(\vec{z}) \otimes I_{\frac{\beta^2}{pnm}}] + \frac{\partial\vec{x}}{\partial\vec{z}} \cdot \frac{\partial(\vec{x}(\vec{z}) \otimes I_{\beta/p})}{\partial\vec{x}} [I_{\frac{\beta}{p}} \otimes [(A' \ltimes \vec{y}(\vec{z})) \otimes I_{\frac{\beta}{nm}}]].$$

Finally, Lemmas 4.1 and 2.26 yield

$$\frac{\partial}{\partial\vec{z}}(\vec{y}'(\vec{z}) \ltimes A \ltimes \vec{x}(\vec{z})) = \frac{\partial\vec{y}}{\partial\vec{z}}\left[\frac{\partial(\vec{y}' \ltimes A)}{\partial\vec{y}} \boxtimes \text{Devec}(I_{\frac{\beta}{nm}})\right] [\vec{x}(\vec{z}) \otimes I_{\frac{\beta^2}{pnm}}] + \frac{\partial\vec{x}}{\partial\vec{z}} \cdot \frac{\partial}{\partial\vec{x}} K^{\tilde{\tau}_p}_{\frac{\beta}{p},p} [I_{\frac{\beta}{p}} \otimes [(A' \ltimes \vec{y}(\vec{z})) \otimes I_{\frac{\beta}{nm}}]].$$

Note that if we partition $\frac{\partial}{\partial\vec{y}}(\vec{y}' \ltimes A)$ as follows:

$$\left[\frac{\partial}{\partial\vec{y}}(\vec{y}' \ltimes A)_{*1}\ \frac{\partial}{\partial\vec{y}}(\vec{y}' \ltimes A)_{*2}\ \cdots \frac{\partial}{\partial\vec{y}}(\vec{y}' \ltimes A)_{*\frac{n\alpha}{m}}\right],$$

then the (r,l)-th submatrix of $\frac{\partial}{\partial\vec{y}}(\vec{y}' \ltimes A) \boxtimes \text{Devec}(I_{\frac{\beta}{nm}})$ is given by

$$\left(\frac{\partial}{\partial\vec{y}}(\vec{y}' \ltimes A)\right)_{*,r} \otimes e_l^{(\beta/nm)'}$$

for each $r = 1, 2, ..., n\alpha/m$ and $l = 1, 2, ..., \beta/(mn)$. □

Corollary 4.3. Under the assumptions of Theorem 4.2, suppose $n = p$ and $m = q$. Then

$$\frac{\partial}{\partial\vec{z}}\vec{y}'(\vec{z})A\vec{x}(\vec{z}) = \frac{\partial\vec{y}}{\partial\vec{z}} \cdot A\vec{x}(\vec{z}) + \frac{\partial\vec{x}}{\partial\vec{z}} \cdot A'\vec{y}(\vec{z}).$$

Proof. Since $n = p$ and $m = q$, we have $\alpha = m$ and $\beta = n$. Now, Eq. (4.2.1) reduces to

$$\frac{\partial}{\partial\vec{z}}\text{y}'(\text{z})A\text{x}(\text{z}) = \frac{\partial\vec{y}}{\partial\vec{z}} \cdot \left[\frac{\partial(\vec{y}'A)}{\partial\vec{y}} \boxtimes \text{Devec}(I_1)\right][\vec{x}(\vec{z}) \otimes I_1] + \frac{\partial\vec{x}}{\partial\vec{z}} \cdot K^{\tilde{\tau}_p}_{1,p} [I_1 \otimes [(A'\vec{y}(\vec{z})) \otimes I_1]].$$

It follows from Lemmas 2.14 and 2.23 respectively that

$$\begin{aligned}\frac{\partial}{\partial z}\vec{y}'(\vec{z})A\vec{x}(\vec{z}) &= \frac{\partial\vec{y}}{\partial\vec{z}} \cdot \left[\frac{\partial(\vec{y}'A)}{\partial\vec{y}}\right][\text{x}(\text{z})] + \frac{\partial\vec{x}}{\partial\vec{z}} \cdot K^{\tilde{\tau}_p}_{1,p} [(A'\vec{y}(\vec{z}))] \\ &= \frac{\partial\vec{y}}{\partial\vec{z}} \cdot [A][\vec{x}(\vec{z})] + \frac{\partial\vec{x}}{\partial\vec{z}} \cdot I_p [(A'\vec{y}(\vec{z}))] \\ &= \frac{\partial\vec{y}}{\partial\vec{z}} \cdot A\vec{x}(\vec{z}) + \frac{\partial\vec{x}}{\partial\vec{z}} \cdot A'\vec{y}(\vec{z}).\end{aligned}$$

□

Theorem 4.4. Let $A(\vec{x})$ and $B(\vec{x})$ be $m \times n$ and $c \times d$ matrix functions of a $p \times 1$ vector variable $\vec{x}$. Then

$$\frac{\partial}{\partial\vec{x}}(A(\vec{x}) \ltimes B(\vec{x})) = \left[\frac{\partial A(\vec{x})}{\partial\vec{x}} \boxtimes \text{Devec}(I_{\frac{\alpha}{n}})\right] [B(\vec{x}) \otimes I_{\frac{m\alpha^2}{cn}}] + \left[\frac{\partial B(\vec{x})}{\partial\vec{x}} \boxtimes \text{Devec}(I_{\frac{\alpha}{c}})\right] [I_{\frac{d\alpha}{c}} \otimes (A'(\vec{x}) \otimes I_{\frac{\alpha}{n}})]. \quad (4.2.2)$$

where $\alpha = \text{lcm}(n, c)$.

Proof. By denoting $\alpha = \text{lcm}(n, c)$, we get

$$A(\vec{x}) \ltimes B(\vec{x}) \;=\; (A(\vec{x}) \otimes I_{\frac{\alpha}{n}})(B(\vec{x}) \otimes I_{\frac{\alpha}{c}}).$$

Theorem 2.25 now implies that

$$\begin{aligned}\frac{\partial}{\partial \vec{x}}(A(\vec{x}) \ltimes B(\vec{x})) = \; & \frac{\partial [A(\vec{x}) \otimes I_{\alpha/n}]}{\partial \vec{x}} \, [(B(\vec{x}) \otimes I_{\frac{\alpha}{c}}) \otimes I_{\frac{m\alpha}{n}}] \\ & + \frac{\partial [B(\vec{x}) \otimes I_{\alpha/c}]}{\partial \vec{x}} \, [I_{\frac{d\alpha}{c}} \otimes (A'(\vec{x}) \otimes I_{\frac{\alpha}{n}})].\end{aligned}$$

By using Lemma 4.1, the above equation becomes

$$\begin{aligned}\frac{\partial}{\partial \vec{x}}(A(\vec{x}) \ltimes B(\vec{x})) = \; & \left[\frac{\partial A(\vec{x})}{\partial \vec{x}} \boxtimes \text{Devec}(I_{\frac{\alpha}{n}})\right] [B(\vec{x}) \otimes I_{\frac{m\alpha^2}{cn}}] \\ & + \left[\frac{\partial B(\vec{x})}{\partial \vec{x}} \boxtimes \text{Devec}(I_{\frac{\alpha}{c}})\right] [I_{\frac{d\alpha}{c}} \otimes (A'(\vec{x}) \otimes I_{\frac{\alpha}{n}})].\end{aligned}$$

Note that if we partition $\frac{\partial}{\partial \vec{x}}A(\vec{x})$ and $\frac{\partial}{\partial \vec{x}}B(\vec{x})$ as follows:

$$\left[\left(\frac{\partial}{\partial \vec{x}}A(\vec{x})\right)_{*1} \left(\frac{\partial}{\partial \vec{x}}A(\vec{x})\right)_{*2} \cdots \left(\frac{\partial}{\partial \vec{x}}A(\vec{x})\right)_{*n}\right]$$

and

$$\left[\left(\frac{\partial}{\partial \vec{x}}B(\vec{x})\right)_{*1} \left(\frac{\partial}{\partial \vec{x}}B(\vec{x})\right)_{*2} \cdots \left(\frac{\partial}{\partial \vec{x}}B(\vec{x})\right)_{*d}\right],$$

then the (j, r)-th and (l, s)-th submatrix of $\frac{\partial}{\partial \vec{x}}A(\vec{x}) \boxtimes \text{Devec}(I_{\frac{\alpha}{n}})$ and $\frac{\partial}{\partial \vec{x}}B(\vec{x}) \boxtimes \text{Devec}(I_{\frac{\alpha}{c}})$ are given by

$$\left(\frac{\partial}{\partial \vec{x}}A(\vec{x})\right)_{*j} \otimes e_r^{(\alpha/n)'}$$

for each $j = 1, 2, ..., n$ and $r = 1, 2, ..., \alpha/n$, and

$$\left(\frac{\partial}{\partial \vec{x}}B(\vec{x})\right)_{*l} \otimes e_s^{(\alpha/n)'}$$

for each $l = 1, 2, ..., d$ and $s = 1, 2, ..., \alpha/d$, respectively. □

From the product rule (Theorem 4.4), we can derive its special cases as follows.

Corollary 4.5.

1. If $n = c$, then Eq. (4.2.2) becomes the product rule involving TMP as discussed in Theorem 2.25.

2. If $n = 1$ (i.e., $\alpha = c$), then Eq. (4.2.2) becomes

$$\frac{\partial}{\partial \vec{x}}(A(\vec{x}) \ltimes B(\vec{x})) = \left[\frac{\partial A(\vec{x})}{\partial \vec{x}} \boxtimes \text{Devec}(I_c)\right] [B(\vec{x}) \otimes I_{mc}] + \left[\frac{\partial}{\partial \vec{x}}B(\vec{x})\right] [I_d \otimes (A'(\vec{x}) \otimes I_c)].$$

3. If $c = 1$ (i.e., $\alpha = n$), then Eq. (4.2.2) becomes

$$\frac{\partial}{\partial \vec{x}}(A(\vec{x}) \ltimes B(\vec{x})) = \left[\frac{\partial}{\partial \vec{x}}A(\vec{x})\right] [B(\vec{x}) \otimes I_{mn}] + \left[\frac{\partial B(\vec{x})}{\partial \vec{x}} \boxtimes \text{Devec}(I_n)\right] [I_{dn} \otimes A'(\vec{x})].$$

4. If $m = d = 1$, and a pair (n,c) is relatively prime, then Eq. (4.2.2) becomes

$$\frac{\partial}{\partial \vec{x}}(A(\vec{x}) \ltimes B(\vec{x})) = \left[\frac{\partial A(\vec{x})}{\partial \vec{x}} \boxtimes \text{Devec}(I_c)\right] [B(\vec{x}) \otimes I_{cn}] + \left[\frac{\partial B(\vec{x})}{\partial \vec{x}} \boxtimes \text{Devec}(I_n)\right] [I_n \otimes (A'(\vec{x}) \otimes I_c)].$$

Proof. All formulas follow from the assumption that $\alpha = \text{lcm}(n, c)$. □

The results in this section generalize the classical results (e.g. [16, Ch. 4]) in the literature, particularly Theorem 2.25.

Chapter 5

Conclusions and suggestions

5.1 Conclusions

This paper investigates matrix derivatives involving the semi-tensor product. The recipes of several product rule's forms are formulated here. Particulary, the notation of zero-one matrices, the versatility of usual product rule and chain rule, and Kronecker/Tracy-Singh products' properties allow us to derive concise and elegant expressions for those derivatives. Ours results generalize the classical ones in the literature, so that the matrix dimensions can the arbitrary. Looking ahead, further refinement of derivative formulas and techniques involving another matrix products would be found unlocking new insights.

5.2 Suggestions

One direction is to explore alternative definitions of the derivative that may be more suitable for specific applications. Another possible improvement involves modifying the product operation itself. By replacing the semi-tensor product with different forms of matrix products, we could potentially unlock new structural insights or simplify the computational framework.

References

[1] Liu, S. Leiva, V. Zhuang, D Ma, T. Figueroa-Zúñiga, J.I. 2023. "Matrix differential calculus with applications in the multivariate linear model and its diagnostics." *Science Talks*. 8

[2] Colombo, F. Gantner, J. Kimsey, D.P. Sabadini, I. 2022. "Universality property of the S-functional calculus, noncommuting matrix variables and Clifford operators." *Advances in Mathematics*. 410.

[3] Brinkhuis, J. 2015. "On the use of coordinate-free matrix calculus." *Journal of Multivariate Analysis*. 133: 377-381.

[4] Sauk, R. Markowski, A.S. Moskal, F. 2015. "Application of the graph theory and matrix calculus for optimal HAZOP nodes order determination." *Journal of Loss Prevention in the Process Industries*. 35: 377-386.

[5] Keshavarz, E. Ordokhani, Y. Razzaghi, M. 2019. "The Bernoulli wavelets operational matrix of integration and its applications for the solution of linear and nonlinear problems in calculus of variations." *Applied Mathematics and Computation*. 351: 83-98.

[6] Magnus, J.R. Neudecker, H. 2010. **Matrix Differential Calculus with Applications in Statistics and Econometrics.** third ed. Wiley Chichester: UK.

[7] Magnus, J.R. 2010. "On the concept of matrix derivative." *Journal of Multivariate Analysis"*. 101(9): 2200-2206.

[8] Cheng, D. 2001."Semi-tensor product of matrices and its application to Morgen's problem." *Sci. China Ser. F.* 44: 195–212.

[9] Cheng, D. Qi, H. Zhao, Y. 2012 **An introduction to Semi-Tensor Product of Matrices and Its Application.** World Scientific Publishing: Singapore.

[10] Cheng, D. Qi, H. Xue, A. 2007 "A survey on semi-tensor product of matrices." *Jrl. Syst. Sci. Complex.* 20: 304–322.

[11] Cheng, D. Qi, H. Li, Q. 2011. **Analysis and control of boolean networks: a semi-tensor product approach.** London: Springer-Verlag.

[12] Cheng, D. Dong, Y. 2003. "Semi-tensor product of matrices and its some applications to physics." *Methods Appl. Anal.* 10: 565–588.

[13] Cheng, D. Qi, H. 2005. "Matrix expression of logic and fuzzy control." *In Proceeding of the 44th IEEE Conference on Decision and Control.* 3273–3278.

[14] Fan, H. Feng, J. E. Meng, M. Wang, B. 2020. "General decomposition of fuzzy relations: Semi-tensor product approach." *Fuzzy Sets Syst.* 384: 75–90.

[15] Horn, J. R. and Johnson, C. R. 1991. **"Topics in Matrix Analysis."** Cambridge University Press.

[16] Turkington, D. A. 2002. **"Matrix Calculus and Zero-One Matrices: Statistical and Econometric Applications."** Cambridge: Cambridge University Press.

[17] Goodfellow, I Bengio, Y Conrville, A. (2016, April 16) "Deep Learning," (Online). Available: https://www.deeplearningbook.org

Appendix

Appendix A

The research paper

IAENG International Journal of Applied Mathematics

Derivatives of Matrix-Valued Functions Involving Semi-Tensor Products in Vector Variables

Thanaphon Phoonphiphat and Pattrawut Chansangiam

***Abstract*—The propose of this paper to derive exact formulae for the derivatives of certain matrix-valued functions with vector variables involving semi-tensor products. Moreover, we investigate the product rule of two matrix-valued functions with vector variables involving semi-tensor products, and its special cases. The paper results generalize the classical ones in the literature, so that the matrix dimensions can be arbitrary and the traditional matrix products are replaced by the semi-tensor products. Moreover, we apply our derivative formulas to compute the gradients of certain vector-valued function arising from neural networks. Furthermore, the derivative formulas can be applied to solve certain matrix equations that generalized classical linear systems. Indeed, a least-squares solution can be obtained as a minimizing vector of the least-squares error associated with the matrix equation.**

***Index Terms*—Matrix derivative, Kronecker product, semi-tensor product, vectorization, zero-one matrix, least-squares solution.**

I. Introduction

MATRIX differential calculus plays an important role in applied mathematics, statistics, data science, econometrics, and related areas. Matrix derivatives are fundamental topics for multivariate analysis, such as asymptotic distributions, linear regression models, and maximum likelihood estimation; see e.g. [1], [13]. The theory of matrix derivatives was developed with the utilization of matrix products, e.g., the traditional matrix product (TMP), and the Kronecker product $\otimes$. Moreover, matrix derivatives often involve vectorizations, e.g., Vec and Devec operators, and specific zero-one matrices, e.g., selection and permutation matrices. To derive derivative formulas, there are two approaches in the literature. The first one is by taking differentials as that in a pioneer work [13] and the paper [14]. Another one is by deriving a few general rules of differentiation such as the product rule and the chain rule; see e.g. [2]. The latter approach was beneficial in linear regression models, seemingly-unrelated regression models, and linear simultaneous equation models. Over the years, several authors had derived exact formulas for the derivatives of certain matrix/vector/scalar-valued functions with respect to matrix/vector/scalar variables. Moreover, they also derived product rules and chain rules involving TMPs as well.

A natural way to extend the study of matrix derivatives is to replace the TMP with the semi-tensor product (STP). Indeed, the STP of matrices, introduced by D. Cheng [5], is a generalization of the TMP so that the factor matrices can be of arbitrary dimensions. The STP of two real matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$ is defined as the TMP between each matrix factor expanding with certain identity matrices:

$$A \ltimes B \;=\; (A \otimes I_{\frac{\alpha}{n}})(B \otimes I_{\frac{\alpha}{p}}) \in \mathbb{R}^{\frac{\alpha m}{n}\times\frac{\alpha q}{p}}, \qquad (1)$$

where α is the least common multiple (lcm) of n and p. For the factor-dimension condition $n = tp$, the STP reduces to

$$A \ltimes B \;=\; A(B \otimes I_t). \qquad (2)$$

If $nt = p$ for some integer t, then $A \ltimes B = (A \otimes I_t)B$. For the matching-dimension condition $n = p$, the STP reduces to the TMP of A and B. Since the STP is based on the TMP, the STP posseses rich algebraic properties as those for TMP, such as associativity, bilinearity, and distributivity over the addition. Special features of STPs are the pseudo-commutativity concerning swap matrices and algebraic formulations of logical functions. See [3] for more information about theory of STPs. It turns out that STPs have a wide range of applications in mathematics and data science: classical and fuzzy logic [6], boolean networks ([6], [9], [10]), networked evolutionary games [7] and finite state machines [8]. Moreover, STPs have applications in physics [11] and engineering [12].

From the above discussion, the STP is one of powerful matrix operations. Instead of focusing on the TMP it is worthy to study matrix calculus in which the TMPs are generalized to the STPs. In this paper, we investigate the derivatives of certain matrix-valued functions involving STPs with respect to a vector variable. In particular, we observe the product rule for two matrix-valued functions. Our results extend the classical results for the case of matching-dimension condition (e.g. [2]) to the case of arbitrary dimensions. Our derivative formulas can be applied to solve matrix equations of the form $A \ltimes \mathrm{x} = B$, where A is a given matrix, B is a given vector/matrix, and x is an unknown column vector.

This work is arranged as following. In Section II, symbolic notations and useful results involving matrix algebra and derivatives are given. In Section III, we derive exact formulas of the derivative of certain matrix-valued functions involving STPs in a vector variable. Section IV deals with the product rule and its special cases. Applications of our theory to neuron networks are presented in Section V. Applications to matrix equations are discussed theoretically in Section VI, and computationally in Section VII . Finally, Section VIII provides a brief conclusion of the whole work.

II. Preliminaries on matrix calculus

This section provides useful tools and notations which will be used throughout this paper. Denote the set of natural

Manuscript received March 27, 2024; revised September 14, 2024. This work was financially supported by King Mongkut's Institute of Technology Ladkrabang Research Fund (KREF016517).

T. Phoonphiphat is a graduate student of King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand (e-mail: 65056043@kmitl.ac.th).

P. Chansangiam is a professor of Mathematics, School of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand (e-mail: pattrawut.ch@kmitl.ac.th).

numbers by $\mathbb{N}$. Let us denote the set of $m \times n$ real matrices by $\mathbb{R}^{m\times n}$. The transpose of a matrix A is denoted by A'.

A. Vectorizations and Matrix Products

Let $A \in \mathbb{R}^{m\times n}$ be a matrix denoted by

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \vec{a}_{*1} & \vec{a}_{*2} & \cdots & \vec{a}_{*n} \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \text{———} & \vec{a}'_{1*} & \text{———} \\ \text{———} & \vec{a}'_{2*} & \text{———} \\ & \vdots & \\ \text{———} & \vec{a}'_{m*} & \text{———} \end{bmatrix},$$

where $\vec{a}_{*j}$ is the j-th column of A and $\vec{a}'_{i*}$ is the i-th row of A. The operators Vec and Devec are defined as follows:

$$\mathrm{Vec}(A) = \begin{bmatrix} \vec{a}_{*1} \\ \vec{a}_{*2} \\ \vdots \\ \vec{a}_{*n} \end{bmatrix} \in \mathbb{R}^{mn\times 1},$$

$$\mathrm{Devec}(A) = \begin{bmatrix} \vec{a}'_{1*} & \vec{a}'_{2*} & \cdots & \vec{a}'_{m*} \end{bmatrix} \in \mathbb{R}^{1\times mn}.$$

We also recall the Kronecker product and Tracy-Singh product as well. The Kronecker product of $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$ is defined to be the following block matrix:

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix} \in \mathbb{R}^{mp\times nq}.$$

Lemma 1. *(e.g. [4]) The following properties hold for any matrices A,B,C,D:*

1) $A \otimes (B \otimes C) = (A \otimes B) \otimes C$,
2) $(A \otimes B)' = A' \otimes B'$,
3) $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$,

provided that all matrix products exist.

Let $A = [a_{ij}] \in \mathbb{R}^{m\times n}$ and $B = [b_{ij}] \in \mathbb{R}^{p\times}$ be partitioned with block submatrices A_{ij} and B_{kl} of dimension $m_i \times n_j$ and $p_k \times q_l$, respectively. Then the Tracy-Singh product of A and B is defined to be a block matrix as follows:

$$A \boxtimes B = \Big[[A_{ij} \otimes B_{kl}]_{kl} \Big]_{ij} \in \mathbb{R}^{mp\times nq}. \tag{3}$$

A zero-one matrix is a matrix whose elements are all either zero or one. It is an essential notation to deal with complexity when working with matrices. Let e_i^n be the i-th column of the $n \times n$ identity matrix I_n. The following commutation matrices will be used in later discussion:

$$K^{\tilde{\tau}_n}_{m,n} = \begin{bmatrix} I_n \otimes e_1^{m'} & I_n \otimes e_2^{m'} & \ldots & I_n \otimes e_m^{m'} \end{bmatrix} \in \mathbb{R}^{n\times nm^2},$$

$$K^{\tau_n}_{m,n} = \begin{bmatrix} I_m \otimes e_1^n \\ I_m \otimes e_2^n \\ \vdots \\ I_m \otimes e_n^n \end{bmatrix} \in \mathbb{R}^{mn^2\times m}.$$

Lemma 2. *(e.g. [2]). From the above notation, we have:*

1) *If* $n = 1$, *then* $K^{\tilde{\tau}_1}_{m,1} = \mathrm{Devec}(I_m)$.
2) *If* $m = 1$, *then* $K^{\tilde{\tau}_n}_{1,n} = I_n$.

B. Matrix Derivatives

We use the following layout conventions for matrix derivatives; see e.g. [2, Ch. 4].

Definition 3. *Let* $\mathrm{y} = [y_1\ y_2\ \cdots\ y_n]'$ *be an* $m \times 1$ *vector whose elements are differentiable functions of a scalar* x. *The derivative of* y *with respect to x is a* $1 \times m$ *vector defined by*

$$\frac{\partial \mathrm{y}}{\partial x} = \begin{bmatrix} \dfrac{\partial y_1}{\partial x} & \dfrac{\partial y_2}{\partial x} & \cdots & \dfrac{\partial y_m}{\partial x} \end{bmatrix}.$$

Definition 4. *Let* A *be an* $m \times n$ *matrix whose elements are differentiable functions of elements of a* $p \times 1$ *vector* $\mathrm{x} = [x_1\ x_2\ \cdots\ x_n]'$. *The derivative of* A *with respect to* x *is a* $p \times mn$ *matrix defined by*

$$\frac{\partial A}{\partial \mathrm{x}} = \begin{bmatrix} \text{———} & \frac{\partial\, \mathrm{Vec}(A)}{\partial x_1} & \text{———} \\ \text{———} & \frac{\partial\, \mathrm{Vec}(A)}{\partial x_2} & \text{———} \\ & \vdots & \\ \text{———} & \frac{\partial\, \mathrm{Vec}(A)}{\partial x_p} & \text{———} \end{bmatrix}.$$

Lemma 5. *(e.g. [2, Ch. 4]). Let* x *be an* $n \times 1$ *matrix and let A be a matrix of constants. Then*

1) $\dfrac{\partial}{\partial \mathrm{x}} A\mathrm{x} = A'$ *for A,* $m \times n$,
2) $\dfrac{\partial}{\partial \mathrm{x}} \mathrm{x}'A = A$ *for A,* $n \times p$,
3) $\dfrac{\partial}{\partial \mathrm{x}} \mathrm{x}'A\mathrm{x} = (A + A')\mathrm{x}$ *for A,* $n \times n$.

Theorem 6. *(e.g. [2, Ch. 4]). Let* $\mathrm{x} = [x_1\ x_2\ \cdots\ x_p]'$ *be a* $p \times 1$ *vector. Let* $\mathrm{y} = [y_1(\mathrm{x})\ y_2(\mathrm{x})\ \cdots\ y_q(\mathrm{x})]'$ *and* $\mathrm{z} = [z_1(\mathrm{y})\ z_2(\mathrm{y})\ \cdots\ z_r(\mathrm{y})]'$ *be* $q \times 1$ *and* $r \times 1$ *vector functions of* x *and* y, *respectively . Then the chain rule is given by*

$$\frac{\partial \mathrm{z}}{\partial \mathrm{x}} = \frac{\partial \mathrm{y}}{\partial \mathrm{x}} \cdot \frac{\partial \mathrm{z}}{\partial \mathrm{y}}.$$

Theorem 7. *(e.g. [2, Ch. 4]). Let A and B be* $m \times n$ *and* $n \times r$ *matrices, respectively. Assume that element of both A and B are scalar functions of a vector* $\mathrm{x} = [x_1\ x_2\ \cdots\ x_p]'$ *of size* $p \times 1$. *The product rule of A and B is given by*

$$\frac{\partial}{\partial \mathrm{x}}(A(\mathrm{x})B(\mathrm{x})) = \frac{\partial A(\mathrm{x})}{\partial \mathrm{x}}(B(\mathrm{x}) \otimes I_m) + \frac{\partial B(\mathrm{x})}{\partial \mathrm{x}}(I_r \otimes A'(\mathrm{x})).$$

Lemma 8. *(e.g. [2, Ch. 4]). Let* $\mathrm{x} \in \mathbb{R}^p$. *Then*

1) $\dfrac{\partial}{\partial \mathrm{x}}(\mathrm{x} \otimes I_n) = K^{\tilde{\tau}_p}_{n,p}$,
2) $\dfrac{\partial}{\partial \mathrm{x}}(I_n \otimes \mathrm{x}) = \mathrm{Devec}(I_n) \otimes I_p$,
3) $\dfrac{\partial}{\partial \mathrm{x}}(\mathrm{x}' \otimes I_n) = I_p \otimes \mathrm{Devec}(I_n)$.

III. Derivatives of Matrix-Valued Functions with Vector Variables Involving Semi-Tensor Products

We derive exact formulas of the derivatives of certain matrix-valued functions involving STPs in a vector variable.

Theorem 9. *Let* $A \in \mathbb{R}^{m\times n}$ *be a constant matrix, and let* x *be a* $p \times 1$ *vector variable. Then*

$$\frac{\partial}{\partial \mathrm{x}}(A \ltimes \mathrm{x}) = K^{\tilde{\tau}_p}_{\frac{\alpha}{p},p}\,[I_{\frac{\alpha}{p}} \otimes (A' \otimes I_{\frac{\alpha}{n}})], \tag{4}$$

$$\frac{\partial}{\partial \mathrm{x}}(\mathrm{x}' \ltimes A) = [I_p \otimes \mathrm{Devec}(I_{\frac{\beta}{p}})](A \otimes I_{\frac{\beta^2}{pm}}), \tag{5}$$

where $\alpha = \mathrm{lcm}(n,p)$ *and* $\beta = \mathrm{lcm}(p,m)$.

Proof: From Eq. (1), we have

$$A \ltimes \mathrm{x} = (A \otimes I_{\frac{\alpha}{n}})(\mathrm{x} \otimes I_{\frac{\alpha}{p}}).$$

It follows from the product rule in Theorem 7 that

$$\begin{aligned}\frac{\partial}{\partial \mathrm{x}}(A \ltimes \mathrm{x}) =& \frac{\partial}{\partial \mathrm{x}}[(A \otimes I_{\frac{\alpha}{n}})(\mathrm{x} \otimes I_{\frac{\alpha}{p}})] \\ =& \frac{\partial (A \otimes I_{\alpha/n})}{\partial \mathrm{x}} [(\mathrm{x} \otimes I_{\frac{\alpha}{p}}) \otimes I_{\frac{m\alpha}{n}}] \\ &+ \frac{\partial (\mathrm{x} \otimes I_{\alpha/p})}{\partial \mathrm{x}} [I_{\frac{\alpha}{p}} \otimes (A' \otimes I_{\frac{\alpha}{n}})] \\ =& \frac{\partial (\mathrm{x} \otimes I_{\alpha/p})}{\partial \mathrm{x}} [I_{\frac{\alpha}{p}} \otimes (A' \otimes I_{\frac{\alpha}{n}})].\end{aligned}$$

Now, Lemma 8 implies that

$$\frac{\partial}{\partial \mathrm{x}}(A \ltimes \mathrm{x}) \; = \; K^{\tilde{\tau}_p}_{\frac{\alpha}{p},p} \, [I_{\frac{\alpha}{p}} \otimes (A' \otimes I_{\frac{\alpha}{n}})].$$

We compute the following derivative according to Theorem 7:

$$\begin{aligned}\frac{\partial}{\partial \mathrm{x}}(\mathrm{x}' \ltimes A) =& \frac{\partial}{\partial \mathrm{x}}[(\mathrm{x}' \otimes I_{\frac{\beta}{p}})(A \otimes I_{\frac{\beta}{m}})] \\ =& \frac{\partial (\mathrm{x}' \otimes I_{\beta/p})}{\partial \mathrm{x}} [(A \otimes I_{\frac{\beta}{m}}) \otimes I_{\frac{\beta}{p}}] \\ &+ \frac{\partial (A \otimes I_{\beta/m})}{\partial \mathrm{x}} [I_{\frac{n\beta}{m}} \otimes (\mathrm{x} \otimes I_{\frac{\beta}{p}})] \\ =& \frac{\partial (\mathrm{x}' \otimes I_{\beta/p})}{\partial \mathrm{x}} [(A \otimes I_{\frac{\beta}{m}}) \otimes I_{\frac{\beta}{p}}] \\ =& \frac{\partial (\mathrm{x}' \otimes I_{\beta/p})}{\partial \mathrm{x}} [A \otimes I_{\frac{\beta^2}{pm}}].\end{aligned}$$

Now, Lemma 8 implies that

$$\frac{\partial}{\partial \mathrm{x}}(\mathrm{x}' \ltimes A) \; = \; [I_p \otimes \mathrm{Devec}(I_{\frac{\beta}{p}})](A \otimes I_{\frac{\beta^2}{pm}}).$$

■

Corollary 10.

1) *If* $p = nt$ *for some* $t \in \mathbb{N}$, *then Eq. (4) becomes*

$$\frac{\partial}{\partial \mathrm{x}}(A \ltimes \mathrm{x}) \; = \; A' \otimes I_t.$$

2) *If* $n = pt$ *for some* $t \in \mathbb{N}$, *then Eq. (4) becomes*

$$\frac{\partial}{\partial \mathrm{x}}(A \ltimes \mathrm{x}) \; = \; K^{\tilde{\tau}_p}_{t,p} \, (I_t \otimes A').$$

3) *If* $m = pt$ *for some* $t \in \mathbb{N}$, *then Eq. (5) becomes*

$$\frac{\partial}{\partial \mathrm{x}}(\mathrm{x}' \ltimes A) \; = \; [I_p \otimes \mathrm{Devec}(I_t)](A \otimes I_t).$$

4) *If* $p = mt$ *for some* $t \in \mathbb{N}$, *then Eq. (5) becomes*

$$\frac{\partial}{\partial \mathrm{x}}(\mathrm{x}' \ltimes A) \; = \; A \otimes I_t.$$

Proof: For the case $p = nt$, we have $\alpha = \mathrm{lcm}(n,p) = p$. By substituting $\alpha = p$ in (4) and applying Lemma 2, we obtain

$$\begin{aligned}\frac{\partial}{\partial \mathrm{x}}(A \ltimes \mathrm{x}) \; &= \; K^{\tilde{\tau}_p}_{1,p} \, [I_1 \otimes (A' \otimes I_t)] \\ &= \; I_p \, [A' \otimes I_t] \\ &= \; A' \otimes I_t.\end{aligned}$$

For the case $n = pt$, we have $\alpha = \mathrm{lcm}(n,p) = n$ and thus

$$\frac{\partial}{\partial \mathrm{x}}(A \ltimes \mathrm{x}) \; = \; K^{\tilde{\tau}_p}_{t,p} \, (I_t \otimes A').$$

Similarly, the remaining results can be done in the same manner by substituting $\beta = \mathrm{lcm}(p,m)$. ■

Theorem 11. *Let* $A \in \mathbb{R}^{m \times n}$ *be a constant matrix, and let* $\mathrm{x} = \mathrm{x}(\mathrm{z})$ *be a* $p \times 1$ *vector function of an* $r \times 1$ *vector* z. *Then*

$$\frac{\partial}{\partial \mathrm{z}}(A \ltimes \mathrm{x}(\mathrm{z})) = \frac{\partial \mathrm{x}}{\partial \mathrm{z}} \, K^{\tilde{\tau}_p}_{\frac{\alpha}{p},p} \, [I_{\frac{\alpha}{p}} \otimes (A' \otimes I_{\frac{\alpha}{n}})], \tag{6}$$

$$\frac{\partial}{\partial \mathrm{z}}(\mathrm{x}'(\mathrm{z}) \ltimes \mathrm{A}) = \frac{\partial \mathrm{x}}{\partial \mathrm{z}} \, [I_p \otimes \mathrm{Devec}(I_{\frac{\alpha}{p}})] \, (A \otimes I_{\frac{\alpha^2}{pm}}), \tag{7}$$

where $\alpha = \mathrm{lcm}(n,p)$ *and* $\beta = \mathrm{lcm}(p,m)$.

Proof: The results follow directly by applying Theorem 6 to (4) and (5) in Theorem 9. ■

We can observe certain special cases of Theorem 11 in a similar manner as Corollary 10.

Theorem 12. *Let* $A \in \mathbb{R}^{m \times n}$ *be a constant matrix. Let* x *and* y *be two independent vector variables of dimension* $p \times 1$ *and* $q \times 1$, *respectively. Then*

$$\frac{\partial}{\partial \mathrm{x}}(\mathrm{y}' \ltimes A \ltimes \mathrm{x}) \; = \; K^{\tilde{\tau}_p}_{\frac{\beta}{p},p} \, [I_{\frac{\beta}{p}} \otimes ((A' \ltimes \mathrm{y}) \otimes I_{\frac{\beta m}{n\alpha}})], \tag{8}$$

where $\alpha = \mathrm{lcm}(q,m)$ *and* $\beta = \mathrm{lcm}(n\alpha/m, p)$.

Proof: Let $M = \mathrm{y}' \ltimes A = (\mathrm{y}' \otimes I_{\frac{\alpha}{q}})(A \otimes I_{\frac{\alpha}{m}})$. *Then*

$$\begin{aligned}(\mathrm{y}' \ltimes A) \ltimes \mathrm{x} \; &= \; [(\mathrm{y}' \otimes I_{\frac{\alpha}{q}})(A \otimes I_{\frac{\alpha}{m}})] \ltimes \mathrm{x} \\ &= \; (M \otimes I_{\frac{\beta m}{n\alpha}})(\mathrm{x} \otimes I_{\frac{\beta}{p}}).\end{aligned}$$

By taking derivative with respect to x *and applying Theorem 7, it follows that*

$$\begin{aligned}\frac{\partial}{\partial \mathrm{x}}(\mathrm{y}' \ltimes A \ltimes \mathrm{x}) =& \frac{\partial (M \otimes I_{\beta m/n\alpha})}{\partial \mathrm{x}} [(\mathrm{x} \otimes I_{\frac{\beta}{p}}) \otimes I_{\frac{\beta m}{q}}] \\ &+ \frac{\partial (\mathrm{x} \otimes I_{\beta/p})}{\partial \mathrm{x}} [I_{\frac{\beta}{p}} \otimes (M' \otimes I_{\frac{\beta m}{n\alpha}})] \\ =& \frac{\partial (\mathrm{x} \otimes I_{\beta/p})}{\partial \mathrm{x}} [I_{\frac{\beta}{p}} \otimes (M' \otimes I_{\frac{\beta m}{n\alpha}})] \\ =& K^{\tilde{\tau}_p}_{\frac{\beta}{p},p} \, [I_{\frac{\beta}{p}} \otimes (M' \otimes I_{\frac{\beta m}{n\alpha}})].\end{aligned}$$

■

Corollary 13.

1) *If a pair* (m,q) *is relatively prime, and* $p = t(n\alpha/m)$ *for some* $t \in \mathbb{N}$, *then Eq. (8) becomes*

$$\frac{\partial}{\partial \mathrm{x}}(\mathrm{y}' \ltimes A \ltimes \mathrm{x}) \; = \; (A' \ltimes \mathrm{y}) \otimes I_t.$$

2) *If a pair* (m,q) *is relatively prime, and* $n\alpha/m = pt$ *for some* $t \in \mathbb{N}$, *then Eq. (8) becomes*

$$\frac{\partial}{\partial \mathrm{x}}(\mathrm{y}' \ltimes A \ltimes \mathrm{x}) \; = \; K^{\tilde{\tau}_p}_{t,p} \cdot [I_t \otimes (A' \ltimes \mathrm{y})].$$

3) *If* $m = qt$ *for some* $t \in \mathbb{N}$, *and a pair* $(n\alpha/m, p)$ *is relatively prime, then Eq. (8) becomes*

$$\frac{\partial}{\partial \mathrm{x}}(\mathrm{y}' \ltimes A \ltimes \mathrm{x}) \; = \; K^{\tilde{\tau}_p}_{n,p} \cdot [I_n \otimes [A'(\mathrm{y} \otimes \mathrm{I_t})] \otimes I_p].$$

4) *If* $q = mt$ *for some* $t \in \mathbb{N}$, *and a pair* $(n\alpha/m, p)$ *is relatively prime, then Eq. (8) becomes*

$$\frac{\partial}{\partial \mathrm{x}}(\mathrm{y}' \ltimes A \ltimes \mathrm{x}) \; = \; K^{\tilde{\tau}_p}_{nt,p} \cdot [I_{nt} \otimes [(A' \otimes I_s)\mathrm{y}] \otimes I_p].$$

5) *If $m = qt$ and $p = s(n\alpha/m)$ for some $t, s \in \mathbb{N}$, then Eq. (8) becomes*

$$\frac{\partial}{\partial \mathrm{x}}(\mathrm{y}' \ltimes A \ltimes \mathrm{x}) \;=\; [A'(\mathrm{y} \otimes \mathrm{I_t})] \otimes I_s.$$

6) *If $m = qt$ and $n\alpha/m = ps$ for some $t, s \in \mathbb{N}$, then Eq. (8) becomes*

$$\frac{\partial}{\partial \mathrm{x}}(\mathrm{y}' \ltimes A \ltimes \mathrm{x}) \;=\; K_{s,p}^{\tilde{\tau}_p} \cdot [I_s \otimes [A'(\mathrm{y} \otimes \mathrm{I_t})]].$$

7) *If $q = mt$ and $p = s(n\alpha/m)$ for some $t, s \in \mathbb{N}$, then Eq. (8) becomes*

$$\frac{\partial}{\partial \mathrm{x}}(\mathrm{y}' \ltimes A \ltimes \mathrm{x}) \;=\; (A' \otimes I_t)\mathrm{y} \otimes I_s.$$

8) *If $q = mt$ and $n\alpha/m = ps$ for some $t, s \in \mathbb{N}$, then Eq. (8) becomes*

$$\frac{\partial}{\partial \mathrm{x}}(\mathrm{y}' \ltimes A \ltimes \mathrm{x}) \;=\; K_{s,p}^{\tilde{\tau}_p} \cdot [I_s \otimes [(A' \otimes I_s)\mathrm{y}]].$$

Proof: The condition that the pair (m, q) is relatively prime means that $\alpha = mq$. The condition that the pair $(n\alpha/m, p)$ is relatively prime means that $\beta = n\alpha p/m$. Now, the results follow from Eq. (8). ■

Theorem 14. *Let $A \in \mathbb{R}^{m\times n}$ be a constant matrix, and let* x *and* y *be $p \times 1$ and $q \times 1$ independent vector variables, respectively. Then*

$$\begin{aligned}\frac{\partial}{\partial \mathrm{y}}&(\mathrm{y}' \ltimes A \ltimes \mathrm{x}) \\ &= (I_q \otimes \mathrm{Devec}(I_{\frac{\beta}{q}}))((A \ltimes \mathrm{x}) \otimes I_{\frac{\beta^2 n}{m\alpha q}}), \qquad (9)\end{aligned}$$

where $\alpha = \mathrm{lcm}(n, p)$ and $\beta = \mathrm{lcm}(q, m\alpha/n)$.

Proof: Let $N = A \ltimes \mathrm{x} = (A \otimes I_{\frac{\alpha}{n}})(\mathrm{x} \otimes I_{\frac{\alpha}{p}})$. Then

$$\begin{aligned}\mathrm{y}' \ltimes (A \ltimes \mathrm{x}) \;&=\; \mathrm{y}' \ltimes [(A \otimes I_{\frac{\alpha}{n}})(\mathrm{x} \otimes I_{\frac{\alpha}{p}})] \\ &=\; (\mathrm{y}' \otimes I_{\frac{\beta}{q}})(N \otimes I_{\frac{\beta n}{m\alpha}}).\end{aligned}$$

By taking derivative with respect to y *and applying Theorem 7, we get*

$$\begin{aligned}\frac{\partial}{\partial \mathrm{y}}(\mathrm{y}' \ltimes A \ltimes \mathrm{x}) = \;& \frac{\partial(\mathrm{y}' \otimes I_{\beta/q})}{\partial \mathrm{y}}\,[(N \otimes I_{\frac{\beta n}{m\alpha}}) \otimes I_{\frac{\beta}{q}}] \\ &+ \frac{\partial(N \otimes I_{\beta n/m\alpha})}{\partial \mathrm{y}}\,[I_{\frac{\beta n}{pm}} \otimes (\mathrm{y} \otimes I_{\frac{\beta}{q}})] \\ =\;& \frac{\partial(\mathrm{y}' \otimes I_{\beta/q})}{\partial \mathrm{y}}\,[(N \otimes I_{\frac{\beta n}{m\alpha}}) \otimes I_{\frac{\beta}{q}}] \\ =\;& \frac{\partial(\mathrm{y}' \otimes I_{\beta/q})}{\partial \mathrm{y}}\,[N \otimes I_{\frac{\beta^2 n}{m\alpha q}}].\end{aligned}$$

Finally, we arrive at Eq. (9) by using Lemma 8. ■

We can observe special cases of Theorem 14 in a similar manner as Corollary 13.

Lemma 15. *Let $A(\mathrm{x})$ be an $m\times n$ matrix function of a vector $\mathrm{x} \in \mathbb{R}^p$. Then*

1) $\dfrac{\partial}{\partial \mathrm{x}}A(\mathrm{x}) = \left[\left(\dfrac{\partial}{\partial \mathrm{x}}A(\mathrm{x})\right)_{*j}\right]_{j=1}^{n}$,
2) $\dfrac{\partial}{\partial \mathrm{x}}(A(\mathrm{x}) \otimes I_t)) \;=\; \dfrac{\partial A(\mathrm{x})}{\partial \mathrm{x}} \boxtimes \mathrm{Devec}(I_t)$.

Proof: A direct computation reveals that

$$\begin{aligned}\frac{\partial}{\partial \mathrm{x}}&A(\mathrm{x}) \\ &= \begin{bmatrix} \frac{\partial}{\partial x_1}\mathrm{Devec}(\mathrm{a}_{*1}(\mathrm{x})) & \cdots & \frac{\partial}{\partial x_1}\mathrm{Devec}(\mathrm{a}_{*n}(\mathrm{x})) \\ \frac{\partial}{\partial x_2}\mathrm{Devec}(\mathrm{a}_{*1}(\mathrm{x})) & \cdots & \frac{\partial}{\partial x_2}\mathrm{Devec}(\mathrm{a}_{*n}(\mathrm{x})) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_p}\mathrm{Devec}(\mathrm{a}_{*1}(\mathrm{x})) & \cdots & \frac{\partial}{\partial x_p}\mathrm{Devec}(\mathrm{a}_{*n}(\mathrm{x})) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial}{\partial x_1}\mathrm{a}'_{*1}(\mathrm{x}) & \frac{\partial}{\partial x_1}\mathrm{a}'_{*2}(\mathrm{x}) & \cdots & \frac{\partial}{\partial x_1}\mathrm{a}'_{*n}(\mathrm{x}) \\ \frac{\partial}{\partial x_2}\mathrm{a}'_{*1}(\mathrm{x}) & \frac{\partial}{\partial x_2}\mathrm{a}'_{*2}(\mathrm{x}) & \cdots & \frac{\partial}{\partial x_2}\mathrm{a}'_{*n}(\mathrm{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_p}\mathrm{a}'_{*1}(\mathrm{x}) & \frac{\partial}{\partial x_p}\mathrm{a}'_{*2}(\mathrm{x}) & \cdots & \frac{\partial}{\partial x_p}\mathrm{a}'_{*n}(\mathrm{x}) \end{bmatrix} \\ &= \left[\left(\frac{\partial}{\partial \mathrm{x}}A(\mathrm{x})\right)_{*j}\right]_{j=1}^{n}.\end{aligned}$$

We also have

$$\begin{aligned}A(\mathrm{x}) &\otimes I_t \\ &= [\mathrm{a}_{*1}(\mathrm{x}) \otimes I_t \quad \mathrm{a}_{*2}(\mathrm{x}) \otimes I_t \quad \cdots \quad \mathrm{a}_{*n}(\mathrm{x}) \otimes I_t] \\ &= \left[\left[\mathrm{a}_{*1}(\mathrm{x}) \otimes e_r^t\right]_{r=1}^{t} \cdots \left[\mathrm{a}_{*n}(\mathrm{x}) \otimes e_r^t\right]_{r=1}^{t}\right]_{j=1}^{n} \\ &= \left[\left[\mathrm{a}_{*j}(\mathrm{x}) \otimes e_r^t\right]_{r=1}^{t}\right]_{j=1}^{n}.\end{aligned}$$

By taking derivative with respect to x, the above equation becomes

$$\begin{aligned}\frac{\partial}{\partial \mathrm{x}}&(A(\mathrm{x}) \otimes I_t) \\ &= \begin{bmatrix} \frac{\partial}{\partial x_1}\mathrm{Devec}\left[\left[\mathrm{a}_{*j}(\mathrm{x}) \otimes e_r^t\right]_{r=1}^{t}\right]_{j=1}^{n} \\ \vdots \\ \frac{\partial}{\partial x_p}\mathrm{Devec}\left[\left[\mathrm{a}_{*j}(\mathrm{x}) \otimes e_r^t\right]_{r=1}^{t}\right]_{j=1}^{n} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial}{\partial x_1}\left[\left[\mathrm{a}'_{*j}(\mathrm{x}) \otimes e_r^{t'}\right]_{r=1}^{t}\right]_{j=1}^{n} \\ \vdots \\ \frac{\partial}{\partial x_p}\left[\left[\mathrm{a}'_{*j}(\mathrm{x}) \otimes e_r^{t'}\right]_{r=1}^{t}\right]_{j=1}^{n} \end{bmatrix}.\end{aligned}$$

Hence

$$\frac{\partial}{\partial \mathrm{x}}(A(\mathrm{x})\otimes I_t)$$
$$= \begin{bmatrix} \frac{\partial}{\partial x_1}\left[\mathrm{a}'_{*1}(\mathrm{x})\otimes e_r^{t'}\right]_{r=1}^{t} & \cdots & \frac{\partial}{\partial x_1}\left[\mathrm{a}'_{*n}(\mathrm{x})\otimes e_r^{t'}\right]_{r=1}^{t} \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_p}\left[\mathrm{a}'_{*1}(\mathrm{x})\otimes e_r^{t'}\right]_{r=1}^{t} & \cdots & \frac{\partial}{\partial x_p}\left[\mathrm{a}'_{*n}(\mathrm{x})\otimes e_r^{t'}\right]_{r=1}^{t} \end{bmatrix}$$
$$= \begin{bmatrix} \left[\frac{\partial}{\partial x_1}\mathrm{a}'_{*1}(\mathrm{x})\otimes e_r^{t'}\right]_{r=1}^{t} & \cdots & \left[\frac{\partial}{\partial x_1}\mathrm{a}'_{*n}(\mathrm{x})\otimes e_r^{t'}\right]_{r=1}^{t} \\ \vdots & \ddots & \vdots \\ \left[\frac{\partial}{\partial x_p}\mathrm{a}'_{*1}(\mathrm{x})\otimes e_r^{t'}\right]_{r=1}^{t} & \cdots & \left[\frac{\partial}{\partial x_p}\mathrm{a}'_{*n}(\mathrm{x})\otimes e_r^{t'}\right]_{r=1}^{t} \end{bmatrix}$$
$$= \left[\left[\left(\frac{\partial}{\partial \mathrm{x}}A(\mathrm{x})\right)_{*1}\otimes e_r^{t'}\right]_{r=1}^{t} \cdots \left[\left(\frac{\partial}{\partial \mathrm{x}}A(\mathrm{x})\right)_{*n}\otimes e_r^{t'}\right]_{r=1}^{t}\right]$$
$$= \left[\left[\left(\frac{\partial}{\partial \mathrm{x}}A(\mathrm{x})\right)_{*j}\otimes e_r^{t'}\right]_{r=1}^{t}\right]_{j=1}^{n}$$
$$= \frac{\partial A(\mathrm{x})}{\partial \mathrm{x}} \boxtimes \mathrm{Devec}(I_t).$$

■

Theorem 16. *Let $A \in \mathbb{R}^{m\times n}$ be a constant matrix, and let $\mathrm{x} = \mathrm{x}(\mathrm{z})$ and $\mathrm{y} = \mathrm{y}(\mathrm{z})$ be $p \times 1$ and $q \times 1$ vector functions of an $r \times 1$ vector* z*. Denote $\alpha = \mathrm{lcm}(q, m)$ and $\beta = \mathrm{lcm}(n\alpha/m, p)$. Then*

$$\begin{aligned} &\frac{\partial}{\partial \mathrm{z}}(\mathrm{y}'(\mathrm{z}) \ltimes A \ltimes \mathrm{x}(\mathrm{z})) \\ &\quad = \frac{\partial \mathrm{y}}{\partial \mathrm{z}} \cdot \left[\frac{\partial(\mathrm{y}' \ltimes A)}{\partial \mathrm{y}} \boxtimes \mathrm{Devec}(I_{\frac{\beta m}{n\alpha}})\right][\mathrm{x}(\mathrm{z}) \otimes I_{\frac{\beta^2 m}{pnq}}] \\ &\qquad + \frac{\partial \mathrm{x}}{\partial \mathrm{z}} \cdot K^{\tilde{\tau}_p}_{\frac{\beta}{p},p}\, [I_{\frac{\beta}{p}} \otimes [(A' \ltimes \mathrm{y}(\mathrm{z})) \otimes I_{\frac{\beta m}{n\alpha}}]]. \end{aligned} \tag{10}$$

In particular when $p = q$ and $\mathrm{x}(\mathrm{z}) = \mathrm{y}(\mathrm{z})$, by denoting $t = n\alpha/(mp)$, we get

$$\begin{aligned} &\frac{\partial}{\partial \mathrm{z}}(\mathrm{x}'(\mathrm{z}) \ltimes A \ltimes \mathrm{x}(\mathrm{z})) \\ &\quad = \frac{\partial \mathrm{x}}{\partial \mathrm{z}}\, [I_p \otimes \mathrm{Devec}(I_{\frac{\alpha}{p}})]\, [(A \otimes I_{\frac{\alpha^2}{mp}})(\mathrm{x} \otimes I_{\frac{t\alpha}{p}})] \\ &\qquad + \frac{\partial \mathrm{x}}{\partial \mathrm{z}}\, K^{\tilde{\tau}_p}_{t,p} \cdot [I_t \otimes (A' \ltimes \mathrm{x})]. \end{aligned} \tag{11}$$

Proof: From the formula (1), we have

$$\begin{aligned} &\mathrm{y}'(\mathrm{z}) \ltimes A \ltimes \mathrm{x}(\mathrm{z}) \\ &\quad = [(\mathrm{y}'(\mathrm{z}) \otimes I_{\frac{\alpha}{q}})(A \otimes I_{\frac{\alpha}{m}})] \ltimes \mathrm{x}(\mathrm{z}) \\ &\quad = \left[[(\mathrm{y}'(\mathrm{z}) \otimes I_{\frac{\alpha}{q}})(A \otimes I_{\frac{\alpha}{m}})] \otimes I_{\frac{\beta m}{n\alpha}}\right] (\mathrm{x}(\mathrm{z}) \otimes I_{\frac{\beta}{p}}) \\ &\quad = [(\mathrm{y}'(\mathrm{z}) \ltimes A) \otimes I_{\frac{\beta m}{n\alpha}}]\, (\mathrm{x}(\mathrm{z}) \otimes I_{\frac{\beta}{p}}). \end{aligned}$$

Taking derivative with respect to z yields

$$\begin{aligned} &\frac{\partial}{\partial \mathrm{z}}(\mathrm{y}'(\mathrm{z}) \ltimes A \ltimes \mathrm{x}(\mathrm{z})) \\ &\quad = \frac{\partial[(\mathrm{y}'(\mathrm{z}) \ltimes A) \otimes I_{\beta m/n\alpha}]}{\partial \mathrm{z}}\, [\mathrm{x}(\mathrm{z}) \otimes I_{\frac{\beta}{p}} \otimes I_{\frac{\beta m}{nq}}] \\ &\qquad + \frac{\partial(\mathrm{x}(\mathrm{z}) \otimes I_{\beta/p})}{\partial \mathrm{z}}\, [I_{\frac{\beta}{p}} \otimes [(A' \ltimes \mathrm{y}(\mathrm{z})) \otimes I_{\frac{\beta m}{n\alpha}}]]. \end{aligned}$$

Now, Theorem 6 implies

$$\begin{aligned} &\frac{\partial}{\partial \mathrm{z}}(\mathrm{y}'(\mathrm{z}) \ltimes A \ltimes \mathrm{x}(\mathrm{z})) \\ &\quad = \frac{\partial \mathrm{y}}{\partial \mathrm{z}} \cdot \frac{\partial[(\mathrm{y}'(\mathrm{z}) \ltimes A) \otimes I_{\beta m/n\alpha}]}{\partial \mathrm{y}}\, [\mathrm{x}(\mathrm{z}) \otimes I_{\frac{\beta^2 m}{pnq}}] \\ &\qquad + \frac{\partial \mathrm{x}}{\partial \mathrm{z}} \cdot \frac{\partial(\mathrm{x}(\mathrm{z}) \otimes I_{\beta/p})}{\partial \mathrm{x}}\, [I_{\frac{\beta}{p}} \otimes [(A' \ltimes \mathrm{y}(\mathrm{z})) \otimes I_{\frac{\beta m}{n\alpha}}]]. \end{aligned}$$

Finally, Lemmas 15 and 8 yield

$$\begin{aligned} &\frac{\partial}{\partial \mathrm{z}}(\mathrm{y}'(\mathrm{z}) \ltimes A \ltimes \mathrm{x}(\mathrm{z})) \\ &\quad = \frac{\partial \mathrm{y}}{\partial \mathrm{z}} \cdot \left[\frac{\partial(\mathrm{y}' \ltimes A)}{\partial \mathrm{y}} \boxtimes \mathrm{Devec}(I_{\frac{\beta m}{n\alpha}})\right][\mathrm{x}(\mathrm{z}) \otimes I_{\frac{\beta^2 m}{pnq}}] \\ &\qquad + \frac{\partial \mathrm{x}}{\partial \mathrm{z}} \cdot K^{\tilde{\tau}_p}_{\frac{\beta}{p},p}\, [I_{\frac{\beta}{p}} \otimes [(A' \ltimes \mathrm{y}(\mathrm{z})) \otimes I_{\frac{\beta m}{n\alpha}}]]. \end{aligned}$$

Note that if we partition $\frac{\partial}{\partial \mathrm{y}}(\mathrm{y}' \ltimes A)$ as

$$\left[\frac{\partial}{\partial \mathrm{y}}(\mathrm{y}' \ltimes A)_{*1}\ \frac{\partial}{\partial \mathrm{y}}(\mathrm{y}' \ltimes A)_{*2}\ \cdots \frac{\partial}{\partial \mathrm{y}}(\mathrm{y}' \ltimes A)_{*\frac{n\alpha}{m}}\right],$$

then the (r, l)-th submatrix of $\frac{\partial}{\partial \mathrm{y}}(\mathrm{y}' \ltimes A) \boxtimes \mathrm{Devec}(I_{\frac{\beta}{nm}})$ is given by

$$\left(\frac{\partial}{\partial \mathrm{y}}(\mathrm{y}' \ltimes A)\right)_{*,r} \otimes e_l^{(\beta/nm)'}$$

for each $r = 1, 2, \ldots, n\alpha/m$ and $l = 1, 2, \ldots, \beta/(mn)$. In particular, when $p = q$ and $\mathrm{x}(\mathrm{z}) = \mathrm{y}(\mathrm{z})$, Eq. (10) reduces to Eq. (11) due to an application of Theorem 9. ■

Corollary 17. *Under the assumptions of Theorem 16, suppose $n = p$ and $m = q$. Then*

$$\frac{\partial}{\partial \mathrm{z}}\mathrm{y}'(\mathrm{z})A\mathrm{x}(\mathrm{z}) = \frac{\partial \mathrm{y}}{\partial \mathrm{z}} \cdot A\mathrm{x}(\mathrm{z}) + \frac{\partial \mathrm{x}}{\partial \mathrm{z}} \cdot A'\mathrm{y}(\mathrm{z}).$$

Proof: Since $n = p$ and $m = q$, we have $\alpha = m$ and $\beta = n$. Now, Eq. (10) reduces to

$$\begin{aligned} &\frac{\partial}{\partial \mathrm{z}}\mathrm{y}'(\mathrm{z})A\mathrm{x}(\mathrm{z}) \\ &\quad = \frac{\partial \mathrm{y}}{\partial \mathrm{z}} \cdot \left[\frac{\partial(\mathrm{y}'A)}{\partial \mathrm{y}} \boxtimes \mathrm{Devec}(I_1)\right][\mathrm{x}(\mathrm{z}) \otimes I_1] \\ &\qquad + \frac{\partial \mathrm{x}}{\partial \mathrm{z}} \cdot K^{\tilde{\tau}_p}_{1,p}\, [I_1 \otimes [(A'\mathrm{y}(\mathrm{z})) \otimes I_1]]. \end{aligned}$$

It follows from Lemmas 2 and 5 respectively that

$$\begin{aligned} &\frac{\partial}{\partial \mathrm{z}}\mathrm{y}'(\mathrm{z})A\mathrm{x}(\mathrm{z}) \\ &\quad = \frac{\partial \mathrm{y}}{\partial \mathrm{z}} \cdot \left[\frac{\partial(\mathrm{y}'A)}{\partial \mathrm{y}}\right][\mathrm{x}(\mathrm{z})] + \frac{\partial \mathrm{x}}{\partial \mathrm{z}} \cdot K^{\tilde{\tau}_p}_{1,p}\, [(A'\mathrm{y}(\mathrm{z}))] \\ &\quad = \frac{\partial \mathrm{y}}{\partial \mathrm{z}} \cdot [A][\mathrm{x}(\mathrm{z})] + \frac{\partial \mathrm{x}}{\partial \mathrm{z}} \cdot I_p\, [(A'\mathrm{y}(\mathrm{z}))] \\ &\quad = \frac{\partial \mathrm{y}}{\partial \mathrm{z}} \cdot A\mathrm{x}(\mathrm{z}) + \frac{\partial \mathrm{x}}{\partial \mathrm{z}} \cdot\ A'\mathrm{y}(\mathrm{z}). \end{aligned}$$

■

Remark 18. *For the special case $p = m = n$, Eq. (11) becomes the item 3) in Lemma 5. To see this, note that $\alpha = p$ and $t = 1$. Thus, we obtain*

$$\begin{aligned} \frac{\partial}{\partial \mathrm{x}}(\mathrm{x}'A\mathrm{x}) &= [I_\alpha \otimes \mathrm{Devec}(I_{\frac{\alpha}{\alpha}})]\, [(A \otimes I_{\frac{\alpha^2}{\alpha^2}})(\mathrm{x} \otimes I_1)] \\ &\quad + K^{\tilde{\tau}_\alpha}_{1,\alpha} \cdot [I_1 \otimes (A' \otimes I_{\frac{\alpha}{\alpha}})(\mathrm{x} \otimes I_{\frac{\alpha}{\alpha}})]. \\ &= I_\alpha A\mathrm{x} + I_\alpha A'\mathrm{x} \\ &= (A + A')\mathrm{x}. \end{aligned}$$

The results in this section generalize the classical results (e.g. [2, Ch. 4]) in the literature, particularly Lemma 5.

IV. The product rule

In this section, we investigate the derivative of the semi-tensor product between two matrix functions with respect to a vector variable.

Theorem 19. *Let $A(\mathrm{x})$ and $B(\mathrm{x})$ be $m\times n$ and $c\times d$ matrix functions of a $p\times 1$ vector variable x. Then*

$$\begin{aligned}\frac{\partial}{\partial \mathrm{x}}&(A(\mathrm{x})\ltimes B(\mathrm{x}))\\ &=\left[\frac{\partial A(\mathrm{x})}{\partial \mathrm{x}}\boxtimes \mathrm{Devec}(I_{\frac{\alpha}{n}})\right][B(\mathrm{x})\otimes I_{\frac{m\alpha^2}{cn}}] \qquad (12)\\ &\quad+\left[\frac{\partial B(\mathrm{x})}{\partial \mathrm{x}}\boxtimes \mathrm{Devec}(I_{\frac{\alpha}{c}})\right][I_{\frac{d\alpha}{c}}\otimes(A'(\mathrm{x})\otimes I_{\frac{\alpha}{n}})].\end{aligned}$$

where $\alpha=\mathrm{lcm}(n,c)$.

Proof: By denoting $\alpha=\mathrm{lcm}(n,c)$, we get

$$A(\mathrm{x})\ltimes B(\mathrm{x}) = (A(\mathrm{x})\otimes I_{\frac{\alpha}{n}})(B(\mathrm{x})\otimes I_{\frac{\alpha}{c}}).$$

Theorem 7 now implies that

$$\begin{aligned}\frac{\partial}{\partial \mathrm{x}}&(A(\mathrm{x})\ltimes B(\mathrm{x}))\\ &=\frac{\partial[A(\mathrm{x})\otimes I_{\alpha/n}]}{\partial \mathrm{x}}[(B(\mathrm{x})\otimes I_{\frac{\alpha}{c}})\otimes I_{\frac{m\alpha}{n}}]\\ &\quad+\frac{\partial[B(\mathrm{x})\otimes I_{\alpha/c}]}{\partial \mathrm{x}}[I_{\frac{d\alpha}{c}}\otimes(A'(\mathrm{x})\otimes I_{\frac{\alpha}{n}})].\end{aligned}$$

By using Lemma 15, the above equation becomes

$$\begin{aligned}\frac{\partial}{\partial \mathrm{x}}&(A(\mathrm{x})\ltimes B(\mathrm{x}))\\ &=\left[\frac{\partial A(\mathrm{x})}{\partial \mathrm{x}}\boxtimes \mathrm{Devec}(I_{\frac{\alpha}{n}})\right][B(\mathrm{x})\otimes I_{\frac{m\alpha^2}{cn}}]\\ &\quad+\left[\frac{\partial B(\mathrm{x})}{\partial \mathrm{x}}\boxtimes \mathrm{Devec}(I_{\frac{\alpha}{c}})\right][I_{\frac{d\alpha}{c}}\otimes(A'(\mathrm{x})\otimes I_{\frac{\alpha}{n}})].\end{aligned}$$

Note that if we partition $\frac{\partial}{\partial \mathrm{x}}A(\mathrm{x})$ and $\frac{\partial}{\partial \mathrm{x}}B(\mathrm{x})$ as follows:

$$\left[\left(\frac{\partial}{\partial \mathrm{x}}A(\mathrm{x})\right)_{*1}\left(\frac{\partial}{\partial \mathrm{x}}A(\mathrm{x})\right)_{*2}\cdots\left(\frac{\partial}{\partial \mathrm{x}}A(\mathrm{x})\right)_{*n}\right]$$

and

$$\left[\left(\frac{\partial}{\partial \mathrm{x}}B(\mathrm{x})\right)_{*1}\left(\frac{\partial}{\partial \mathrm{x}}B(\mathrm{x})\right)_{*2}\cdots\left(\frac{\partial}{\partial \mathrm{x}}B(\mathrm{x})\right)_{*d}\right],$$

then each (j,r)-th submatrix of $\frac{\partial}{\partial \mathrm{x}}A(\mathrm{x})\boxtimes \mathrm{Devec}(I_{\frac{\alpha}{n}})$ and each (l,s)-th submatrix of $\frac{\partial}{\partial \mathrm{x}}B(\mathrm{x})\boxtimes \mathrm{Devec}(I_{\frac{\alpha}{c}})$ are given respectively by

$$\left(\frac{\partial}{\partial \mathrm{x}}A(\mathrm{x})\right)_{*j}\otimes e_r^{(\alpha/n)'},\quad \left(\frac{\partial}{\partial \mathrm{x}}B(\mathrm{x})\right)_{*l}\otimes e_s^{(\alpha/n)'}.$$

■

From the product rule (Theorem 19), we can derive its special cases as follows.

Corollary 20.

1) *If $n=c$, then Eq.* (12) *becomes the product rule involving TMP as discussed in Theorem 7.*
2) *If $n=1$ (i.e., $\alpha=c$), then Eq.* (12) *becomes*

$$\begin{aligned}\frac{\partial}{\partial \mathrm{x}}&(A(\mathrm{x})\ltimes B(\mathrm{x}))\\ &=\left[\frac{\partial A(\mathrm{x})}{\partial \mathrm{x}}\boxtimes \mathrm{Devec}(I_c)\right][B(\mathrm{x})\otimes I_{mc}]\\ &\quad+\left[\frac{\partial}{\partial \mathrm{x}}B(\mathrm{x})\right][I_d\otimes(A'(\mathrm{x})\otimes I_c)].\end{aligned}$$

3) *If $c=1$ (i.e., $\alpha=n$), then Eq.* (12) *becomes*

$$\begin{aligned}\frac{\partial}{\partial \mathrm{x}}&(A(\mathrm{x})\ltimes B(\mathrm{x}))\\ &=\left[\frac{\partial}{\partial \mathrm{x}}A(\mathrm{x})\right][B(\mathrm{x})\otimes I_{mn}]\\ &\quad+\left[\frac{\partial B(\mathrm{x})}{\partial \mathrm{x}}\boxtimes \mathrm{Devec}(I_n)\right][I_{dn}\otimes A'(\mathrm{x})].\end{aligned}$$

4) *If $m=d=1$, and a pair (n,c) is relatively prime, then Eq.* (12) *becomes*

$$\begin{aligned}\frac{\partial}{\partial \mathrm{x}}&(A(\mathrm{x})\ltimes B(\mathrm{x}))\\ &=\left[\frac{\partial A(\mathrm{x})}{\partial \mathrm{x}}\boxtimes \mathrm{Devec}(I_c)\right][B(\mathrm{x})\otimes I_{cn}]\\ &\quad+\left[\frac{\partial B(\mathrm{x})}{\partial \mathrm{x}}\boxtimes \mathrm{Devec}(I_n)\right][I_n\otimes(A'(\mathrm{x})\otimes I_c)].\end{aligned}$$

Proof: All formulas follow from the assumption that $\alpha=\mathrm{lcm}(n,c)$. ■

The results in this section generalize the classical results (e.g. [2, Ch. 4]) in the literature, particularly Theorem 7.

V. Applications to neural networks

In a neural network, suppose we have an n-component input vector $\mathrm{x}\in\mathbb{R}^n$. In order to train a neuron, we choose a weight vector $\mathrm{w}\in\mathbb{R}^n$ with the same component number as that for the input vector. We also need a scalar bias $b\in\mathbb{R}$. Then, the activation of a single computation unit in a neuron is typically calculated as

$$\mathrm{F}(\mathrm{x}) = \sum_{i=1}^{n} w_i x_i + b = \mathrm{w}'\mathrm{x}+b.$$

The function F is known as the unit's affine function. To train this neuron, we choose weights w and the bias b that minimize an associated loss function. To minimize the loss function, we use matrix derivatives.

Now, suppose we have t collections of an n-component data. We can represent them with a single vector

$$\mathrm{x} = [x_1\ \cdots\ x_n\ \cdots\ x_{n(t-1)+1}\ \cdots\ x_{nt}]'\in\mathbb{R}^{nt}.$$

Assume that

- we use the same weights for each data collection, namely,
$$\mathrm{w} = [w_1\ \cdots\ w_n]'\in\mathbb{R}^n.$$
- we use different bias for different data sets, so we can form the bias vector to be
$$\mathrm{b} = [b_1\ \cdots\ b_t]'\in\mathbb{R}^t.$$

IAENG International Journal of Applied Mathematics

Thus, the affine function is given by

$$\begin{aligned}\mathrm{F}(\mathrm{x}) &= (\mathrm{w}' \ltimes \mathrm{x}) + \mathrm{b} \\ &= \begin{bmatrix} \sum_{i=1}^{n} \mathrm{w}_i \mathrm{x}_{(i-1)t+1} + b_1 \\ \sum_{i=1}^{n} \mathrm{w}_i \mathrm{x}_{(i-1)t+2} + b_2 \\ \vdots \\ \sum_{i=1}^{n} \mathrm{w}_i \mathrm{x}_{(i-1)t+t} + b_t \end{bmatrix}.\end{aligned}$$

To minimize the associated loss function, we shall differentiate F with respect to x, w, and b. Indeed, from Corollary 10, we obtain

$$\begin{aligned}\frac{\partial \mathrm{F}}{\partial \mathrm{x}} &= \frac{\partial}{\partial \mathrm{x}}(\mathrm{w}' \ltimes \mathrm{x}) + \frac{\partial}{\partial \mathrm{x}} b \\ &= (\mathrm{w}')' \otimes I_t \\ &= \mathrm{w} \otimes I_t\end{aligned}$$

The same corollary implies.

$$\begin{aligned}\frac{\partial \mathrm{F}}{\partial \mathrm{w}} &= \frac{\partial}{\partial \mathrm{w}}(\mathrm{w}' \ltimes \mathrm{x}) + \frac{\partial}{\partial \mathrm{w}} b \\ &= [I_n \otimes \mathrm{Devec}(I_t)](\mathrm{x} \otimes I_t).\end{aligned}$$

The gradient of F with respect to the bias is given by

$$\frac{\partial \mathrm{F}}{\partial \mathrm{b}} = \frac{\partial}{\partial \mathrm{b}}(\mathrm{w}' \ltimes \mathrm{x}) + \frac{\partial}{\partial \mathrm{b}} \mathrm{b} = I_t.$$

VI. Least-squares solutions of the equation $A \ltimes \mathrm{x} = b$ via matrix derivatives

In this section, we shall apply certain derivative formulas to find least-squares solutions of a matrix equation related to linear systems. Recall that the Moore-Penrose inverse of a matrix A is denoted by $A^\dagger$; see, e.g., [13] for more details.

Recall that, in a classical linear system takes the form

$$A\mathrm{x} = b, \tag{13}$$

where $A \in \mathbb{R}^{m \times n}$ is a given constant matrix, $b \in \mathbb{R}^n$ is a given vector, and $\mathrm{x} \in \mathbb{R}^m$ is an unknown vector.

Lemma 21. *(e.g. [13]) Suppose that the linear system* (13) *is inconsistent. Then the least-squares solution to* (13) *is an exact solution to the normal equation*

$$A'A\mathrm{x} = A'b.$$

In fact, the general least-squares solutions of (13) *can be expressed as*

$$\mathrm{x} = (A'A)^\dagger A'b + \left[I_n - (A'A)^\dagger A'A\right] w, \tag{14}$$

where $w \in \mathbb{R}^n$ is arbitrary. The minimal-norm solution of Eq. (13) *is given by*

$$\mathrm{x} = (A'A)^\dagger A'b. \tag{15}$$

The system (13) *has a unique least-squares solution if and only if A is of full-column rank (i.e.* $\mathrm{rank}(A) = n$*). Moreover, such unique solution is given by* (15).

We can extend the classical case to that when $\mathrm{x} \in \mathbb{R}^p$, where p is a positive integer divided by n. Now, assume that $p = nt$ where t is a positive integer. We are given $b \in \mathbb{R}^{mt}$, and we would like to solve the following equation:

$$A \ltimes \mathrm{x} = b. \tag{16}$$

To find a least-squares solution of Eq. (16), we follow an idea of the works [15], [16], that is, we transform the matrix equation into a simple linear system. So, we look for a vector $\mathrm{x}^* \in \mathbb{R}^{nt}$ that minimizes the squared Euclidean norm

$$\|A \ltimes \mathrm{x} - b\|^2.$$

Indeed, the least-squares error can be computed as follows:

$$\begin{aligned}&\|A \ltimes \mathrm{x} - b\|^2 \\ &\quad = (A \ltimes \mathrm{x} - b)'(A \ltimes \mathrm{x} - b) \\ &\quad = (\mathrm{x}' \ltimes A' - b')(A \ltimes \mathrm{x} - b) \\ &\quad = \mathrm{x}' \ltimes A'A \ltimes \mathrm{x} - \mathrm{x}' \ltimes (A' \ltimes b) - (b' \ltimes A) \ltimes \mathrm{x} + b'b.\end{aligned} \tag{17}$$

The vector x^* is an exact solution of Eq. (16) if and only if the least-squares error (17) is zero. To minimize such error, we shall differentiate it with respect to the vector x. Indeed, we get

$$\begin{aligned}&\frac{\partial}{\partial \mathrm{x}}\|A \ltimes \mathrm{x} - b\|^2 \\ &= \frac{\partial}{\partial \mathrm{x}}(\mathrm{x}' \ltimes A'A \ltimes \mathrm{x}) - \frac{\partial}{\partial \mathrm{x}}(\mathrm{x}' \ltimes A' \ltimes b) - \frac{\partial}{\partial \mathrm{x}}(b' \ltimes A \ltimes \mathrm{x}).\end{aligned} \tag{18}$$

The first term in (18) can be computed using Theorem 16. The second term can be computed using (5) in Theorem 9, and Lemma 5. The last term can be computed using (4) in Theorem 9, and Lemma 2. Putting them together yields

$$\begin{aligned}&\frac{\partial}{\partial \mathrm{x}}\|A \ltimes \mathrm{x} - b\|^2 \\ &\quad = [(A'A \otimes I_t)(\mathrm{x} \otimes I_1)] + K_{1,p}^{\tilde{\tau}_p} \cdot [I_1 \otimes (A'A \ltimes \mathrm{x})] \\ &\qquad - (A' \ltimes b) - (b' \ltimes A)' \\ &\quad = (A'A \otimes I_t)\mathrm{x} + I_p \cdot (A'A \otimes I_t)\mathrm{x} \\ &\qquad - (A' \ltimes b) - (A' \ltimes b)' \\ &\quad = (A'A \otimes I_t)\mathrm{x} + (A'A \ltimes \mathrm{x}) - 2(A' \ltimes b) \\ &\quad = (A'A \otimes I_t)\mathrm{x} + (A'A \otimes I_t)\mathrm{x} - 2(A' \otimes I_t)b \\ &\quad = (A'A \otimes I_t)\mathrm{x} - (A' \otimes I_t)b.\end{aligned} \tag{19}$$

The least-squares solution can be obtained by setting the derivative (19) to be zero, and solve for x. Thus, the least-squares solutions of Eq. (16) can be obtained by solving the linear system

$$K\mathrm{x} = \mathrm{f}, \tag{20}$$

where $K = A'A \otimes I_t$ and $\mathrm{f} = (A' \otimes I_t)b$.

Now, we apply Lemma 21 to discuss theoretical details from the associated system (20). Note that

$$K^\dagger = (A'A \otimes I_t)^\dagger = (A'A)^\dagger \otimes I_t.$$

The general solutions of this system can be written as

$$\mathrm{x} = K^\dagger \mathrm{f} + (I_{nt} - K^\dagger K)\mathrm{w}, \tag{21}$$

where $\mathrm{w} \in \mathbb{R}^{nt}$ is arbitrary. From properties of the Kronecker product, the expression (21) becomes

$$\mathrm{x} = \left[(A'A)^\dagger A' \otimes I_t\right] b + \left[I_{nt} - \{(A'A)^\dagger A'A \otimes I_t\}\right] w. \tag{22}$$

Among such solutions, the minimal-norm solution is given by

$$\mathrm{x} = K^\dagger \mathrm{f} = \left[(A'A)^\dagger A' \otimes I_t\right] b. \tag{23}$$

In addition, Eq. (20) has a unique solution if and only if K is of full rank. Note that

$$\begin{aligned}\operatorname{rank} K &= \operatorname{rank}(A'A \otimes I_t) = (\operatorname{rank} A'A)\cdot(\operatorname{rank} I_t)\\ &= t \operatorname{rank} A'A.\end{aligned}$$

Thus, $\operatorname{rank} K = nt$ if and only if $\operatorname{rank} A'A = n$, or equivalently, A is of full-column rank. In this case, the unique solution is given by (23).

We summarize the above discussion as follows.

Theorem 22. *From the above notations, suppose that the matrix equation* (16) *is inconsistent. Then:*

(i) *The least-squares solutions of* (16) *is an exact solution of the linear system* (20) *where* $K = A'A \otimes I_t$ *and* $\mathrm{f} = (A' \otimes I_t)b$.
(ii) *The general least-squares solutions of* (16) *can be expressed as* (22), *where* $\mathrm{w} \in \mathbb{R}^{nt}$ *is arbitrary.*
(iii) *The minimal-norm least-squares solution of* (16) *is given by* (23).
(iv) *The equation* (16) *has a unique least-squares solution if and only if A is of full-column rank. Moreover, such unique least-squares solution is given by* (23).

Remark 23. *When $t = 1$, the matrix equation* (16) *reduces to the classical linear system* (13). *Hence, Theorem 22 is an extension of Lemma 21.*

In practice, to solve the linear system (20), we can use a modern iterative method such as a preconditioned AOR algorithm [17], and a gradient-descent algorithm [18].

VII. Gradient-descent algorithm for the matrix equation and numerical experiments

In this section, we propose an effective computational method to solve the matrix equation (16), and illustrate numerical experiments.

From Section VI, the least-squares solutions of Eq. (16) are equivalent to the solutions of the associated linear system (20). To solve the latter system, we adopt the gradient-descent optimization technique from the work [18]. The main idea is to minimize the residual error $\|K\mathrm{x} - \mathrm{f}\|$ at each iteration. We thus obtain the following gradient-descent iterative (GDI) algorithm:

Algorithm 1: GDI algorithm for solving Eq. (16)

```
A ∈ R^{m×n}, and b ∈ R^{mt} ;
Set i = 0. Choose x^(0) ∈ R^p. Compute
  K = A'A ⊗ I_t, f = (A' ⊗ I_t)b, M = (A'A)^2 ⊗ I_t.
for i = 0,1,2,3,... do
    r^(i) = f − K x^(i);
    if ||r^(i)|| ⩽ ε then
        x^(i) is a solution; break;
    else
        m_(i) = M r^(i);
        α_(i+1) = m_(i)^T r^(i) / (2 m_(i)^T m_(i)) ;
        x^(i+1) = x^(i) + α_(i+1) K r^(i) ;
    end
    update i;
end
```

We make experiments via Matlab R2024a on the same Mac Operating System (Intel i5 4C CPU/intel iris graphic 645GPU/8GB/128GB). The performance of the algorithm is evaluated by the iteration number, the residual error

$$R^{(i)} = \|r^{(i)}\| = \|f - K\mathrm{x}^{(i)}\|,$$

and the CPU time measured in seconds using the tic-toc function on MATLAB.

Example 24. *Consider the equation* $A \ltimes \mathrm{x} = \mathrm{b}$, *where*

$$A = \begin{bmatrix} 2 & -1 \\ 0 & 1 \\ -2 & 2 \end{bmatrix} \in \mathbb{R}^{3\times 2},$$
$$\mathrm{b} = \begin{bmatrix} 0 & 4 & 0 & -2 & 0 & -6 \end{bmatrix}' \in \mathbb{R}^6.$$

We would like to find a least-squares solution

$$\mathrm{x} = [x_1\ x_2\ \cdots x_6]' \in \mathbb{R}^6$$

satisfying the above equation. Due to Theorem 22, this task is equivalent to finding a solution of the associated linear system $K\mathrm{x} = \mathrm{f}$, *where*

$$K = A'A \otimes I_3 = \begin{bmatrix} 8 & 0 & -6 & 0 \\ 0 & 8 & 0 & -6 \\ -6 & 0 & 6 & 0 \\ 0 & -6 & 0 & 6 \end{bmatrix} \in \mathbb{R}^{4\times 4},$$
$$\mathrm{f} = (A' \otimes I_3)\mathrm{b} = \begin{bmatrix} 0 & 20 & 0 & -18 \end{bmatrix}' \in \mathbb{R}^4.$$

We apply Algorithm 1 with an initial guess

$$\mathrm{x}^{(0)} = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}'$$

and a tolerance error $\epsilon = 0.05$. *The experimental results are illustrated numerically in Table I and graphically in Figure 1.*

TABLE I
Numerical solutions for each iteration

i	x_1	x_2	x_3	x_4	$R^{(i)}$
1	1	1	1	1	25.5342
2	0.9533	1.7362	1.0351	0.3690	12.9427
3	0.9288	2.1025	1.0513	0.0513	6.8096
4	0.9154	2.2839	1.0581	-0.1097	4.0079
5	0.9076	2.3730	1.0602	-0.1925	2.9093
6	0.9026	2.4159	1.0599	-0.2363	2.5578
7	0.8988	2.4360	1.0584	-0.2609	2.4575
8	0.8897	2.4448	1.0559	-0.2771	2.4264
9	0.8897	2.4464	1.0504	-0.2952	2.4083
10	0.7958	2.2619	0.9360	-0.4540	2.2017
⋮	⋮	⋮	⋮	⋮	⋮
49	0.0099	1.0158	0.0117	-1.9808	0.0272

In this problem, Algorithm 1 takes 49 *iterations and consumes only* 0.036119 *seconds to reach an approximate solution*

$$\mathrm{x}^{(49)} = \begin{bmatrix} 0.0099 & 1.0158 & 0.0117 & -1.9808 \end{bmatrix}'.$$

We can check the least-squares error

$$\|A \ltimes x^{(49)} - b\|^2 = 0.0280.$$

Thus, $x^{(49)}$ *is a desire least-squares solution. Hence, Algorithm 1 is capable and effective.*

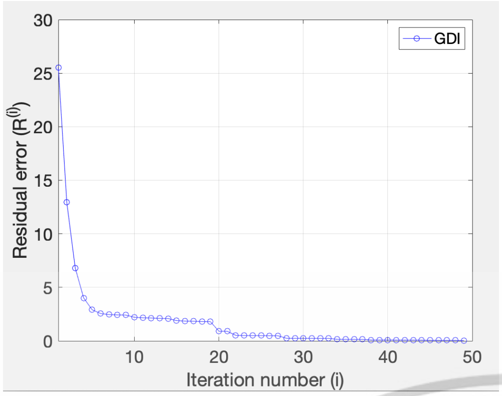

Fig. 1. The residual error at each iteration for Ex. 24

VIII. Conclusions

This paper investigates matrix derivatives involving the semi-tensor products. The recipes of several product rule's forms are formulated in Sections III and IV. Particularly, the notation of zero-one matrices, the versatility of usual product rule and chain rule, and Kronecker/Tracy-Singh products' properties allow us to derive concise and elegant expressions for those derivatives. Our results generalize the classical ones in the literature, so that the matrix dimensions can be arbitrary. As applications in neural networks, we apply derivative formulas to compute the gradient of a vector-valued function with respect to certain vector variables. The derivative formulas can be applied to solve a matrix equation of the form $A \ltimes x = b$. A least-squares solution can be obtained as a minimizing vector of the associated least-squares error. We can seek for a least-squares solution of this matrix equation by solving the associated linear system. Thus, we get formulas of general/minimal-norm/unique least-squares solutions as in Theorem 22. Moreover, we propose a gradient-descent iterative procedure to solve the matrix equation for a least-squares solution. Looking ahead, further refinement of derivative formulas and techniques involving another matrix products would be found unlocking new insights.

References

[1] S. Liu, V. Leiva, D. Zhuang, T. Ma, J.I. Figueroa-Zúñiga, "Matrix differential calculus with applications in the multivariate linear model and its diagnostics," *Science Talks*, vol. 8, 100274, 2023.

[2] DA. Turkington, "Matrix Calculus and Zero-One Matrices: Statistical and Econometric Applications," *Cambridge University*, 2001.

[3] D. Cheng, H. Qi, A. Xue, "A survey on semi-tensor product of matrices," *Jrl. Syst. Sci. Complex.*, vol. 20, pp. 304–322, 2007.

[4] R.A. Horn, CR. Johnson, "Topics in Matrix Analysis.", *Cambridge University*, 1991.

[5] D. Cheng, "Semi-tensor product of matrices and its application to Morgen's problem", *Sci. China Ser. F* vol. 44, pp. 195–212, 2001.

[6] D. Cheng, H. Qi, Q. Li, "Analysis and control of boolean networks: a semi-tensor product approach," *London: Springer-Verlag*, 2011.

[7] G. Zhao, H. Li, P. Duan, F. E. Alsaadi, "Survey on applications of semi-tensor product method in networked evolutionary games," *J. Appl. Anal. Comput.*, vol. 10, pp. 32–54, 2020.

[8] Y. Yan, D. Cheng, J. Feng, H. Li, J. Yue, "Survey on applications of algebraic state space theory of logical systems to finite state machines," *Sci. China Inf. Sci.,* vol. 66, 111201, 2023.

[9] Y. Yan, J. Yue, Z. Chen, "Algebraic method of simplifying Boolean networks using semi-tensor product of Matrices," *Asian J. Control*, vol. 21, pp. 2569–2577, 2019.

[10] H. Ji, Y. Li, X. Ding, J. Lu, "Stability analysis of Boolean networks with Markov jump disturbances and their application in apoptosis networks," *Electronic Res. Arch.,* vol. 30, pp. 3422–3434, 2022.

[11] D. Cheng, Y. Dong, "Semi-tensor product of matrices and its some applications to physics," *Methods Appl. Anal.,* vol. 10, pp. 565–588, 2003.

[12] H. Li, G. Zhao, M. Meng, J. Feng, "A survey on applications of semi-tensor product method in engineering," *Sci. China Inf. Sci.,* vol. 61, 010202, 2018.

[13] J.R. Magnus, H. Neudecker, "Matrix Differential Calculus with Applications in Statistics and Econometrics," *third ed., Wiley, Chichester, UK,* 2019.

[14] J.R. Magnus, "On the concept of matrix derivative," *Journal of Multivariate Analysis*, vol. 101, 9, pp. 2200-2206, 2010.

[15] J. Jaipraseft, P. Chansangiam, "Solving the Sylvester-transpose matrix equation under the semi-tensor product," *Symmetry,* vol. 14, 6, 1094, 2022.

[16] J. Jaiprasert, P. Chansangiam, "Exact and least-squares solutions of a generalized Sylvester-transpose matrix equation over generalized quaternions," *Electronic Res. Arch.,* vol. 32, 4, pp. 2789–2804, 2024.

[17] A. Li, "A new preconditioned AOR iterative method and comparison theorems for linear systems," *IAENG Int. J. Appl. Math.,* vol. 42, 3, pp. 161-163. 2012.

[18] K. Tansri, P. Chansangiam, "Gradient-descent iterative algorithm for solving exact and weighted least-squares solutions of rectangular linear systems," *AIMS Mathematics*, vol. 8, 5, pp. 11781–11798, 2023.

[19] I. Goodfellow, Y. Bengio, A. Conrville. (2016, April 16) "Deep Learning," (Online). Available: https://www.deeplearningbook.org

Author Biography

Name	Mister Thanaphon Phoonphiphat
Date of Birth	18 August 1999
Address	20 moo 2, Nongnamsai, Phachi, Phra Nakhon Si Ayutthaya, 13140
Education	(2022) Bachelor of Science in Applied Mathematics
	King Mongkut's Institute of Technology Ladkrabang
Scholarship	KMITL Research and Innovation Services (KRIS)