IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 4, 2018, accepted August 16, 2018, date of publication August 31, 2018, date of current version September 21, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2867234

Towards a Unified Framework of Matrix Derivatives

JIANYU XU[1], GUOQI LI[1], CHANGYUN WEN[2], (Fellow, IEEE), KUN WU[3], AND LEI DENG[4]
[1]Center for Brain Inspired Computing Research, Department of Precision Instrument, Tsinghua University, Beijing 100084, China
[2]School of Electrical and Electronic Engineering , Nanyang Technological University, Singapore 639798
[3]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[4]Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106, USA

Corresponding author: Guoqi Li (liguoqi@tsinghua.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under under Grant 61603209 and Grant 61876215 and in part by the Independent Research Plan of Tsinghua University under Grant 20151080467.

ABSTRACT The need of processing and analyzing massive statistics simultaneously requires the derivatives of matrix-to-scalar functions (scalar-valued functions of matrices) or matrix-to-matrix functions (matrix-valued functions of matrices). Although derivatives of a matrix-to-scalar function have already been defined, the way to express it in algebraic expression, however, is not as clear as that of scalar-to-scalar functions (scalar-valued functions of scalars). Due to the fact that there does not exist a uniform way of applying ''chain rule'' on matrix derivation, we classify approaches utilized in existing schemes into two ways: the first relies on the index notation of several matrices, and they would be eliminated while being multiplied; the second relies on the vectorizing of matrices and thus they can be dealt with in the way we treat vector-to-vector functions (vector-valued functions of vectors), which has already been settled. On one hand, we find that the first approach holds a much lower time complexity than that of the second approach in general. On the other hand, until now though we know most typical functions that can be derived in the first approach, theoretically the second approach is more generally fit for any routine of "chain rule." The result of the second approach, nevertheless, can be also simplified to the same order of time complexity with the first approach under certain conditions. Therefore, it is important to establish these conditions. In this paper, we establish a sufficient condition under which not only the first approach can be applied but also the time complexity of results obtained from the second approach can be reduced. This condition is described in two equivalent individual conditions, each of which is a counterpart of an approach sequentially. In addition, we generalize the methods and use these two approaches to do the derivatives under the two conditions individually. This paper enables us to unify the framework of matrix derivatives, which would result in various applications in science and engineering.

INDEX TERMS Matrix derivatives, index notation, Kronecker product, chain rule, matrix calculus, time complexity.

I. INTRODUCTION

With the deepening of researches on the function of matrices, it is gradually important to find a method to calculate the matrix derivatives, i.e., the derivative of matrix-to-scalar functions (scalar-valued functions of matrices) or matrix-to-matrix functions (matrix-valued functions of matrices). The applications of matrix derivatives have been extensively involved in in many real life optimization problems such as signal processing [1]–[7], machine learning [8]–[10], image processing [11], [12], complex networks [13]–[15], social science [16], [17], and various optimization problems [18]–[20].

However, the principle of matrix derivatives is yet to be clearly defined. Firstly, although matrix-to-scalar function derivatives has been mentioned in a lot in existing works, see for examples in [21]–[23], the chain rule that plays a key role in doing the derivatives, nevertheless, is not based on a clear definition [24]. Actually the chain rule in this situation is very difficult, if not impossible, to be put into use directly, as it falls short of approach generality and theoretical analysis. Secondly, for the matrix-to-matrix function derivatives, researchers pointed out that some existing notations may be unsuitable because they would have ''no interpretation'' in some cases, and thus an useful chain rule does not exist [21], [25]. Considering these situations, we focus on the relatively better studied matrix-to-scalar function derivatives in this work. However, our results shall still have prevalent meaning of guidance in matrix derivatives. That is not only because matrix-to-matrix function derivatives are generally implicitly involved as numerous intermediate matrix variables, but also because there exist operations or mutually dependent relations among those variables. To make a conclusion on existing works that have applied chain rule on matrix derivatives, we classify their methods into two approaches. In the following parts, we suppose that any ordinary $m \times n$ matrix (which means it is not a Kronecker product or commutation matrix or similar-sized matrix) satisfies that $\frac{m}{n} = O(1)$.

On one hand, some articles, including [13] and [26], express the derivatives by calculating scalar-to-scalar functions derivatives for every element of the variable matrix, and re-arrange these derivatives according to the index notation of the elements of the variable matrix. This is because the matrix-to-scalar function can be considered as a scalar-to-scalar function with its variable being an element of the variable matrix. In this work, we mainly consider a scalar-valued function $F(U)$ where matrix U is a function of X and want to find the derivative of $F(.)$ with respect to X. In this case, the element $\frac{\partial U_{p,q}}{\partial X_{i,j}}$ of a fourth order tensor is always involved [27], [28]. While applying the chain rule, it would be multiplied by many elements of other fourth-order tensors which serve as "links" in the chain of derivatives. After that, those elements are summed up according to their index notations. We name this method as the "***Index Approach***". Note that the time complexity of determining matrix differentiation can be very high. Although it can be reduced to $O\left(n^6\right)$ by noting those tensors as matrices and doing matrix product, and to $O\left(n^4\right)$ by calculating the product of the vector in the first position with each of the following matrices sequentially, $O\left(n^4\right)$ is still much more than tolerance of computation. However, through properly arranging the index notation in the Index Approach, the process of the derivatives can be realized through matrix products and summation. It is well known that the time complexity of products[1] of the matrices involved in the approach is $O\left(n^3\right)$, therefore we can reduce the complexity of matrix-to-scalar scalar function derivatives to $O\left(n^3\right)$. Note that such a $O\left(n^3\right)$ complexity can be achieved in various cases, as almost all the derivatives of matrix-to-scalar functions in science and engineering applications can be exactly decomposed into the product of two or more matrices with their index notations, which represent the position of each element of the derivatives, corresponding to those of the functions and variables. It is necessary to mention we could indeed construct a function that the Index Approach fails to handle, as illustrated in the Case study section. This suggests that the procedures of Index Approach may be limited to certain condition.

On the other hand, we know that a vector is a special form of a matrix in which all elements are organized in a line, and a matrix can always be stacked to a vector form. Since derivatives of a vector-to-vector function have been well defined and discussed, derivatives of a matrix-to-matrix function can be operable. In [29] and [30] and textbooks [21] and [31] the authors introduced another approach, which is to vectorize the function matrix and the variable matrix by its columns separately and then derive the vectorized function matrix to the vectorized variable matrix according to the principle of vector derivatives. Therefore, we can call it "***Vec Approach***". It is worthwhile to notice that, although vectorizing matrix to vectors provides a general way for matrix derivatives, the Vec Approach may be inapplicable in some situations. For example, considering the problem proposed in [32] and [33], each column of an input matrix represents an external source for controlling. In this case, stacking the input matrix into a vector will lose the physical meaning within each column, and what's worse, makes the network becomes uncontrollable.

Besides, if we process the derivation in the Vec Approach, from an $m \times n$ sized matrix to a $p \times q$ one, we would get a derivative with its size $mn \times pq$. Since Vec Approach is the same as derivatives of vector-to-vector function, conditions fit for this approach are general. However, supposing that $O\left(\frac{m}{n}\right) = O\left(\frac{p}{q}\right) = O\left(\frac{n}{p}\right) = O(1)$ as done earlier, and that the derivatives are going to be multiplied by one another sequentially, the time complexity of derivatives achieved in this approach can be as high as $O\left(n^4\right)$ (only when products are calculated under proper sequence, similar to that of the Index Approach mentioned earlier), which is unwieldy for application. Also, the outcomes in [21]–[23] have all involved the Kronecker product $\otimes$ and the commutation matrix $T_{p,q}$. Since there are many elements occurring repeatedly in a Kronecker-product matrix, a waste of space complexity forms and would lead to an unnecessarily high time complexity while the Kronecker product is involved in multiplication.

This kind of difference leads to a huge variation on the time complexity of the outcomes, even though the outcomes belong to the same definition and algorithm which are just different seemingly. There did exist some attempts to reduce the complexity of the results that come from the Vec Approach. Al-Zhour and Aziz [34] introduced an access that has reduced the time complexity of the original result, which consists of Kronecker product, from $O\left(n^4\right)$ to $O\left(n^3\right)$, a result similar to that of the Index Approach. However, the access in [34] is not a general approach. Therefore, a question should be put forward that under which condition would the ***Index Approach*** be equivalent to[2] the ***Vec Approach***.

In this paper, we propose two sufficient conditions to get $O\left(n^3\right)$-sized derivatives via the Index Approach, or the Vec Approach, respectively, when determining the matrix-to-scalar function derivatives. Also, for those satisfying the conditions, we give general ways to get $O\left(n^3\right)$ results.

[1] Here we only consider the naïve matrix multiplication.

[2] 2 Here "equivalent" means that any matrix function which can be derived in one approach if and only if it can be derived in another.

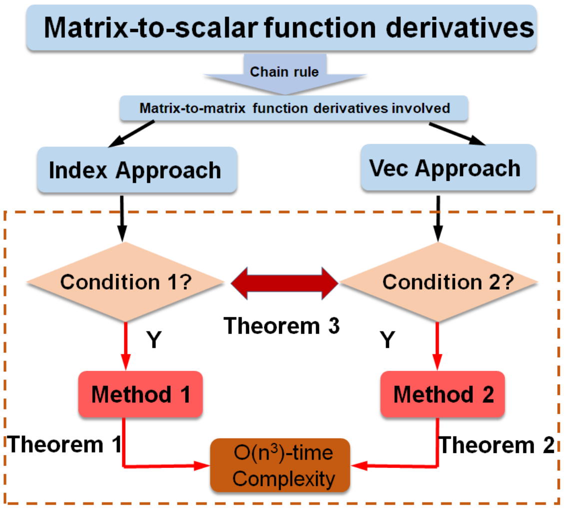

FIGURE 1. **The structure of matrix derivatives while applying chain rule for matrix-to-scalar function derivatives, in which matrix-to-matrix function derivatives are generally involved. Parts in the dashed rectangular frame are our contribution.**

With respect to this, we build up a general framework in Figure 1 while applying chain rule for matrix-to-scalar function derivatives, in which matrix-to-matrix function derivatives are generally involved. The main contributions of this work are illustrated in the parts in the dashed rectangular frame in this Figure. For the two approaches under certain (but common) conditions, we give constructive proofs for the equivalence of the two conditions and their validity respectively, which demonstrate the pathway for the transformation of the two definitions.

II. INDEX APPROACH AND VEC APPROACH FOR MATRIX DERIVATIVES

A. PRELIMINARY KNOWLEDGE AND DEFINITION

Definition 1 (Kronecker Product, $\otimes$ [32]): It is an operation that transforms matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$ into matrix $C \in \mathbb{R}^{mp\times nq}$.

$$C = A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix},$$

where a_{ij} is the i^{th} row and j^{th} column element of matrix A.

Definition 2 (Commutation Matrix [21]): For $A \in \mathbb{R}^{m\times n}$, there exists a matrix $T_{m,n}$ such that:

$$T_{m,n}vec\left(A\right) = vec\left(A^T\right)$$

where $vec(.)$ vectorizes a matrix by stacking its columns.

Definition 3: Four functions XH, XL, KH and KL are defined as follows:

(1) For $(A \otimes B), A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{p\times q}$, define:

$$XH\left(A \otimes B\right) = m$$
$$XL\left(A \otimes B\right) = n$$
$$KH\left(A \otimes B\right) = p$$
$$KL\left(A \otimes B\right) = q$$

(2) For $T_{p,q}$, define:

$$XH\left(T_{p,q}\right) = KL\left(T_{p,q}\right) = p$$
$$XL\left(T_{p,q}\right) = KH\left(T_{p,q}\right) = q$$

*Definition 4 (**Rational**):* Consider the following matrix product:

$$\prod_{l=1}^{t} R_l$$

where $t \in \mathbb{Z}^+$. For every R_l, $l = 1, 2, \dots, t$, there either exist matrices $A_l \in \mathbb{R}^{m_l\times n_l}$ and $B_l \in \mathbb{R}^{p_l\times q_l}$ such that $R_l = A_l \otimes B_l$, or there exists $p_l, q_l \in \mathbb{Z}^+$, such that $R_l = T_{p_l,q_l}$. If $KL\left(R_l\right) = KH\left(R_{l+1}\right)$ is true for every $l \in \{1, 2, 3, \dots, t\}$, then we call the matrix product **Rational**.

Based on the above definitions, it is easy to know that if $\prod_{l=1}^{t} R_l$ is Rational, then for any $1 \le t_1 \le t_2 \le t, t_1, t_2 \in \mathbb{Z}^+$, $\prod_{l=t_1}^{t_2} R_l$ is Rational. Also, if for any $1 \le t_1 \le t_2 \le t$, $t_1, t_2 \in \mathbb{Z}^+$, $\prod_{l=t_1}^{t_2} R_l$ is Rational, then $\prod_{l=1}^{t} R_l$ is Rational. Also, if the product R_lR_{l+1} is well defined, the column number of R_l is equal to the row number R_{l+1}, which means $XL\left(R_l\right) KL\left(R_l\right) = XH\left(R_{l+1}\right) KH\left(R_{l+1}\right)$. Therefore,

$$KL\left(R_l\right) = KH\left(R_{l+1}\right) \Leftrightarrow XL\left(R_l\right) = XH\left(R_{l+1}\right)$$

Now we quote three formulae that have been proved before. For $(A \otimes B), A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{p\times q}$, according to [34], we have

$$\left(A \otimes B\right) T_{n,q} = T_{m,p}\left(B \otimes A\right) \tag{1}$$

It is apparent that $\left(A \otimes B\right) T_{n,q}$ and $T_{m,p}\left(B \otimes A\right)$ are Rational.

And based on [1] and [23], we have

$$\left(A \otimes B\right)\left(C \otimes D\right) = AC \otimes BD \tag{2}$$

This apparently implies that the product of A and C as well as that of B and D is well defined, which is equivalent to that $\left(A \otimes B\right)\left(C \otimes D\right)$ is Rational.

Also, according to [31], it is easy to obtain that

$$T_{p,m}T_{n,q} = T_{p,m}T_{m,p} = I_{mp} \tag{3}$$

when $m = n$ and $p = q$. That is to say, when $T_{p,m}T_{n,q}$ is Rational, formula (3) is true.

Now we have the following lemma.

Lemma 1: Consider the following matrix product $\prod_{l=1}^{t} R_l$ and only apply (1), (2) and (3) on the product $\prod_{l=1}^{t} R_l$ in order to deform it into the following expression

$$\prod_{l'=1}^{t'} S_{l'}$$

where t' is the number of items in the deformed product. If $\prod_{l=1}^{t} R_l$ is Rational, then $\prod_{l'=1}^{t'} R'_l$ is Rational, and $KH\left(R_1\right) = KH\left(R'_1\right), KL\left(R_t\right) = KL\left(R'_{t'}\right)$.

Proof: Note that each of the formulae (1), (2) and (3) only involves two matrices, and according to the definition of Rational, whether the product is Rational or not only depends on the relationship between each matrix and its neighborhood.

Therefore, for $R_l R_{l+1} = S_{l'}$, we should prove that $KH(R_l) = KH(S_{l'})$, and $KL(R_{l+1}) = KL(S_{l'})$

For $R_l R_{l+1} = S_{l'} S_{l'+1}$, we should prove that $KH(R_l) = KH(S_{l'})$, and $KL(R_{l+1}) = KL(S_{l'+1})$

That is to say, in order to prove Lemma 1, we just need to prove the following four statements:

1) When $(A \otimes B) T_{n,q} = T_{m,p}(B \otimes A)$, it is true that $KH(A \otimes B) = KH(T_{m,p})$ and $KL(T_{n,q}) = KL(B \otimes A)$
2) When $(A \otimes B)(C \otimes D) = AC \otimes BD$, $KH(A \otimes B) = KH(AC \otimes BD)$ and $KL(C \otimes D) = KL(AC \otimes BD)$
3) When $T_{p,m} T_{m,p} R$ is Rational, $KH(T_{p,m}) = KH(R)$
4) When $S T_{p,m} T_{m,p}$ is Rational, $KL(T_{m,p}) = KL(S)$

Actually, all above statements can be obtained by directly using Definition 3. ■

By applying the formulae (1), (2) and (3), the KH (as well as XH) value of the first matrix of a Rational product remains unchanged, so does the KL (as well as XL) value of the last matrix of a Rational product. Therefore, we can show the KH, XH, KL and XL value of a Rational product.

Definition 5 (Enlargement of XH, XL, KH, KL): For a Rational product

$$\prod_{l=1}^{t} R_l$$

define:

$$XH\left(\prod_{l=1}^{t} R_l\right) = XH(R_1)$$
$$XL\left(\prod_{l=1}^{t} R_l\right) = XL(R_t)$$
$$KH\left(\prod_{l=1}^{t} R_l\right) = KH(R_1)$$
$$KL\left(\prod_{l=1}^{t} R_l\right) = KL(R_t)$$

According to *Definition 5*, the value of the four function remain unchanged while applying formulae (1), (2) and (3).

Definition 6 (Expression of Matrix-to-Scalar Derivatives): For a matrix-to-scalar function $f(X)$, we define: 1)

1) $\frac{\partial f}{\partial X}$ is a matrix with the same size of X;
2) $f'(X) = \frac{\partial f}{\partial X}$;
3) $\left(\frac{\partial f}{\partial X}\right)_{i,j} = \frac{\partial f}{\partial X_{i,j}}$;
4) $\frac{\partial f}{\partial vec(X)} = \left(vec\left(\frac{\partial f}{\partial X}\right)\right)^T$.

B. INDEX APPROACH AND VEC APPROACH

Now we suppose that there is a matrix-to-scalar composite function $f(X)$, and that there is a sequence of matrix-to-matrix functions $\{F_k\}$, $k = 1, 2, \ldots\ldots, r-1$. Assume that F_0 is a matrix-to-scalar function, such that

$$f(X) = F_0(F_1(F_2(\ldots\ldots(F_{r-1}(X)))))$$

In the following part, for $k = 0, 1, 2, 3, \ldots\ldots, r$, we denote $F_k(F_{k+1}(\ldots\ldots(F_{r-1}(X))))$ as F_k, especially $F_r = X$ such that $F_k \in \mathbb{R}^{m_k \times n_k}$, where $m_k = O(n)$, $n_k = O(n)$.

We firstly propose the definition of Index Approach and Vec Approach as follows:

Definition 7 (Index Approach): For a matrix-to-scalar function $f(X) = F_0(F_1(F_2(\ldots\ldots(F_{r-1}(X)))))$, we define the following way, which is based on chain-rule, to calculate its derivatives with respect to every element of X an *Index Approach*:

$$\frac{\partial f(X)}{\partial X_{i,j}} = \sum_{i_1}\sum_{j_1}\sum_{i_2}\sum_{j_2}\cdots\cdots\sum_{i_{r-1}}\sum_{j_{r-1}} \frac{\partial F_0}{\partial F_{1_{(i_1 j_1)}}} \cdot \prod_{k=1}^{r-1} \frac{\partial F_{k_{(i_k j_k)}}}{\partial F_{k+1_{(i_{k+1} j_{k+1})}}}$$

where $i = i_r$ and$j = j_r$.

Definition 8 (Vec Approach): For a matrix-to-scalar function $f(X) = F_0(F_1(F_2(\ldots\ldots(F_{r-1}(X)))))$, we define the following way, which is also based on chain-rule, to calculate its derivatives with respect to the vectorized matrix $Vec(X)$ an *Vec Approach*:

$$\frac{\partial f}{\partial vec(X)} = \prod_{k=0}^{r-1} \frac{\partial vec(F_k)}{\partial vec(F_{k+1})} = \frac{\partial F_0(F_1)}{\partial vec(F_1)} \prod_{k=1}^{r-1} \frac{\partial vec(F_k)}{\partial vec(F_{k+1})}$$

In the following part, we will respectively discuss how to find the $\boldsymbol{O}(\boldsymbol{n}^3)$ derivative of $f(X)$ with respect to X based on Index Approach and Vec Approach.

1) $O(n^3)$ COMPLEXITY CONDITION FOR INDEX APPROACH

Based on the Index Approach, the derivative $f'(X)$ is derived as follows:

$$\left(f'(X)\right)_{i,j} = \frac{\partial f(X)}{\partial X_{i,j}} = \sum_{i_1}\sum_{j_1}\sum_{i_2}\sum_{j_2}\cdots\cdots\sum_{i_{r-1}}\sum_{j_{r-1}} \frac{\partial F_0}{\partial F_{1_{(i_1 j_1)}}} \cdot \prod_{k=1}^{r-1} \frac{\partial F_{k_{(i_k j_k)}}}{\partial F_{k+1_{(i_{k+1} j_{k+1})}}}$$

where $i = i_r$ and$j = j_r$.

Originally the time complexity of this expression can be as high as $O(n^{2r})$, but it is demonstrated in [13] that it can be reduced to $O(n^3)$ if the following condition is true.

Condition 1: For $\forall k = 2, \ldots, r$, there exist:

Either matrices $A_k \in \mathbb{R}^{m_{k-1} \times m_k}$, $C_k \in \mathbb{R}^{n_{k-1} \times n_k}$ such that

$$\frac{\partial F_{k-1_{(i_{k-1} j_{k-1})}}}{\partial F_{k_{(i_k j_k)}}} = A_{k_{(i_{k-1}, i_k)}} \cdot C_{k_{(j_{k-1}, j_k)}}$$

Or matrices $A_k \in \mathbb{R}^{m_{k-1}\times n_k}$, $C_k \in \mathbb{R}^{n_{k-1}\times m_k}$ such that

$$\frac{\partial F_{k-1_{(i_{k-1},j_{k-1})}}}{\partial F_{k_{(i_k,j_k)}}} = A_{k_{(i_{k-1},j_k)}} \cdot C_{k_{(j_{k-1},i_k)}}$$

When the condition above is satisfied, we term this approach as ***Method 1***

2) $O\left(n^3\right)$ COMPLEXITY CONDITION FOR VEC APPROACH

Based on the Vec Approach, the derivative $f'(X)$ is derived as follows

$$\begin{aligned} f'(X) &= \frac{\partial f}{\partial vec(X)} \\ &= \prod_{k=0}^{r-1} \frac{\partial vec(F_k)}{\partial vec(F_{k+1})} \\ &= \frac{\partial F_0(F_1)}{\partial vec(F_1)} \prod_{k=1}^{r-1} \frac{\partial vec(F_k)}{\partial vec(F_{k+1})} \end{aligned}$$

As is shown above, the derivatives are expressed as a $1 \times O\left(n^2\right)$-sized vector multiplied by a series of $O\left(n^2\right) \times O\left(n^2\right)$-sized matrices. Although the time complexity of a product of $O\left(n^2\right) \times O\left(n^2\right)$-sized matrices is $O\left(n^6\right)$, it can be reduced to $O\left(n^4\right)$ while calculating the product of the vector with the following matrices sequentially. However, in [21], [29], and [31], we notice that the derivatives achieved via Vec Approach are all Rational, and Al-Zhour and Aziz [34] deform an Rational derivatives from an $O\left(n^4\right)$ expression to an $O\left(n^3\right)$ one with the utilization of the following formula

$$vec(AXB) = \left(B^T \otimes A\right) vec(X) \tag{4}$$

Therefore, we propose the following condition.

Condition 2: For $\forall k = 2, \ldots, r$, $\frac{\partial vec(F_{k-1})}{\partial vec(F_k)}$ can be expressed as a Rational product of matrices, and its KH value equals m_{k-1} and its KL value equals m_k.

When the condition above is satisfied, we term this approach as ***Method 2***.

In the next section, we shall prove that the above Condition 1 and Condition 2 guarantee the $O(n^3)$ complexity of Method 1 and Method 2, respectively. In addition, they are equivalent when taking the derivatives.

III. THEORETICAL ANALYSIS

Theorem 1: Under Condition 1, the derivative of $f(X)$ can be expressed as a product based on the Index Approach with its time complexity $O\left(n^3\right)$

Proof: We prove this theorem by mathematical induction.

(i) When $r = 2$, there are two cases:

Case 1:

$$\frac{\partial F_{1_{(i_1,j_1)}}}{\partial F_{2_{(i_2,j_2)}}} = A_{2_{(i_1,i_2)}} \cdot C_{2_{(j_1,j_2)}}$$

Case 2:

$$\frac{\partial F_{1_{(i_1,j_1)}}}{\partial F_{2_{(i_2,j_2)}}} = A_{2_{(i_1,j_2)}} \cdot C_{2_{(j_1,i_2)}}$$

For Case 1, the original problem can be written as

$$\begin{aligned} \left(f'(X)\right)_{i,j} &= \frac{\partial f(X)}{\partial X_{i,j}} \\ &= \sum_{i_1}\sum_{j_1} \frac{\partial F_0}{\partial F_{1_{(i_1,j_1)}}} \cdot \frac{\partial F_{1_{(i_1,j_1)}}}{\partial F_{2_{(i_2,j_2)}}} \\ &= \sum_{i_1}\sum_{j_1} \frac{\partial F_0}{\partial F_{1_{(i_1,j_1)}}} \cdot A_{2_{(i_1,i_2)}} \cdot C_{2_{(j_1,j_2)}} \\ &= \sum_{i_1}\sum_{j_1} A^T_{2_{(i_2,i_1)}} \cdot \left(\frac{\partial F_0}{\partial F_1}\right)_{(i_1,j_1)} \cdot C_{2_{(j_1,j_2)}} \\ &= \left(A_2^T \left(\frac{\partial F_0}{\partial F_1}\right) C_2\right)_{(i_2,j_2)} \end{aligned}$$

That is to say,

$$f'(X) = A_2^T \left(\frac{\partial F_0}{\partial F_1}\right) C_2$$

For Case 2, similar to the substantiation above, we have

$$\begin{aligned} \left(f'(X)\right)_{i,j} &= \sum_{i_1}\sum_{j_1} A^T_{2_{(j_2,i_1)}} \cdot \left(\frac{\partial F_0}{\partial F_1}\right)_{(i_1,j_1)} \cdot C_{2_{(j_1,i_2)}} \\ &= \left(A_2^T \left(\frac{\partial F_0}{\partial F_1}\right) C_2\right)_{(j_2,i_2)} \\ &= \left(C_2^T \left(\frac{\partial F_0}{\partial F_1}\right)^T A_2\right)_{(i_2,j_2)} \end{aligned}$$

That is to say,

$$f'(X) = C_2^T \left(\frac{\partial F_0}{\partial F_1}\right)^T A_2$$

In conclusion, when $r = 2$, $f'(X)$ can be expressed as a product with its time complexity $O\left(n^3\right)$.

(ii) Assume that when $r = t \left(t \geq 2, t \in \mathbb{Z}^+\right)$, $f'(X)$ can be expressed as a product with its time complexity $O\left(n^3\right)$.

Then considering the case $r = t + 1$, we have

$$\begin{aligned} \left(f'(X)\right)_{i,j} &= \frac{\partial f(X)}{\partial X_{i,j}} \\ &= \sum_{i_1}\sum_{j_1}\sum_{i_2}\sum_{j_2} \cdots\cdots \sum_{i_t}\sum_{j_t} \frac{\partial F_0}{\partial F_{1_{(i_1,j_1)}}} \\ &\quad \cdot \prod_{k=1}^{t} \frac{\partial F_{k_{(i_k,j_k)}}}{\partial F_{k+1_{(i_{k+1},j_{k+1})}}} \\ &= \sum_{i_1}\sum_{j_1} \frac{\partial F_0}{\partial F_{1_{(i_1,j_1)}}} \cdot \frac{\partial F_{1_{(i_1,j_1)}}}{\partial F_{2_{(i_2,j_2)}}} \cdot \sum_{i_2}\sum_{j_2} \\ &\quad \ldots \sum_{i_t}\sum_{j_t}\prod_{k=2}^{t} \frac{\partial F_{k_{(i_k,j_k)}}}{\partial F_{k+1_{(i_{k+1},j_{k+1})}}} \end{aligned}$$

According to our proof in (i), if

$$\left(\frac{\partial F_0}{\partial F_2}\right)_{(i_2,j_2)} = \sum_{i_2}\sum_{j_2} \frac{\partial F_0}{\partial F_{1_{(i_1,j_1)}}} \cdot \frac{\partial F_{1_{(i_1,j_1)}}}{\partial F_{2_{(i_2,j_2)}}}$$

and $\frac{\partial F_{1(i_1,j_1)}}{\partial F_{2(i_2,j_2)}}$ satisfies *Condition 1*, then

$$\left(\frac{\partial F_0}{\partial F_2}\right)_{(i_2,j_2)} = \left(A_2^T\left(\frac{\partial F_0}{\partial F_1}\right)C_2\right)_{(i_2,j_2)}$$

or

$$\left(\frac{\partial F_0}{\partial F_2}\right)_{(i_2,j_2)} = \left(C_2^T\left(\frac{\partial F_0}{\partial F_1}\right)^T A_2\right)_{(i_2,j_2)}$$

Now we rearrange the functions as

$$G_0 = F_0$$

For $k = 1, 2, 3, \ldots, t$,

$$G_k = F_{k+1}$$

Therefore,

$$\left(f'(X)\right)_{i,j} = \frac{\partial f(X)}{\partial X_{i,j}} = \sum_{i_2}\sum_{j_2}\sum_{i_3}\sum_{j_3}\cdots\cdots\sum_{i_t}\sum_{j_t}\frac{\partial F_0}{\partial F_{2(i_2,j_2)}} \cdot \prod_{k=2}^{t}\frac{\partial G_{k-1(i_k,j_k)}}{\partial G_{k(i_{k+1},j_{k+1})}}$$

It is apparent that $\{G_k\}$ satisfies *Condition 1*.

Note that $\frac{\partial F_0}{\partial F_2}$ is a product of $O(n) \times O(n)$-sized matrices. That is to say, calculating $\frac{\partial F_0}{\partial F_2}$ whose time complexity is $O(n^3)$ does not increase the time complexity of calculating $f'(X)$. Therefore, in the case that $r = t + 1$, the result of Theorem 1 is also true. Thus the theorem is proved. ■

We are going to consider Condition 2. Before doing that, we present the following lemma first.

Lemma 2: For any Rational product

$$\prod_{l=1}^{t} R_l$$

there exist either matrix G and H such that

$$\prod_{l=1}^{t} R_l = (G \otimes H)$$

or matrix G and H such that

$$\prod_{l=1}^{t} R_l = T_{p,m}(G \otimes H)$$

where $G \in \mathbb{R}^{m\times n}$ and $H \in \mathbb{R}^{p\times q}$

Proof: From the definition of Rational product, we know that for every $l = 1, 2, \ldots\ldots, t$, there exists either matrix A_l, B_l such that $R_l = (A_l \otimes B_l)$ or exist $p_l, m_l \in \mathbb{Z}^+$ such that $R_l = T_{p_l,m_l}$.

Suppose that there exists a series of n integers $\{a_n\}$ such that $1 \le a_1 < a_2 < \ldots < a_n \le t$, and that: 1)

1) For any integer $k \in \{a_n\}$, $R_k = T_{p_k,m_k}$;
2) For any integer $l \in \{1, 2, \ldots, t\}$ but $l \notin \{a_n\}$, $R_l = (A_l \otimes B_l)$.

We then can define the sum

$$S = \sum_{i=1}^{n} a_i$$

Now we make a series of operation on $\prod_{l=1}^{t} R_l$ according to the following rules:

Operation: For $l \in \{1, 2, 3, \ldots, t-2, t-1\}$, if $R_l \cdot R_{l+1} = (A_l \otimes B_l)\, T_{p_{l+1},m_{l+1}}$, then according to the formula (1), there exist A_{l+1}, B_{l+1}, and $p_l, m_l \in \mathbb{Z}^+$, such that

$$R_l \cdot R_{l+1} = T_{p_l,m_l}(A_{l+1} \otimes B_{l+1})$$

Therefore, we can update the $\prod_{l=1}^{t} R_l$ by supposing that

$$R_l = T_{p_l,m_l}$$

and

$$R_{l+1} = (A_{l+1} \otimes B_{l+1})$$

According to Lemma 1, the updated $\prod_{l=1}^{t} R_l$ is also a Rational product. Thus the value of

$$S = \sum_{i=1}^{n} a_i$$

would reduce by 1, with the n remaining unchanged.

Since we know that $1 \le a_1 < a_2 < \ldots < a_n \le t$, if we can show that

$$S = \sum_{i=1}^{n} a_i \ge \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

then the number of the times by which we can operate on the $\prod_{l=1}^{t} R_l$ is limited. That is to say, the number of operations we would carry out is finite.

Now we prove that the minimum value of S is $\frac{n(n+1)}{2}$.

If the Rational product $\prod_{l=1}^{t} R_l$ cannot be operated furthermore, and $S > \frac{n(n+1)}{2}$ at present, there must exist $k \in \{1, 2, \ldots\ldots, n\}$ such that $k \notin \{a_n\}$. Therefore, $k < a_n$. Suppose that a_r is the minimal value among $\{a_n\}$ that satisfies $a_r > k$. That is to say, $a_{r-1} < k < a_r$. So,$(a_r - 1) \notin \{a_n\}$, and $R_{a_r-1} \cdot R_{a_r} = \left(A_{a_r-1} \otimes B_{a_r-1}\right) T_{p_{a_r},m_{a_r}}$, which is capable of being operated. This gives a contradiction. Therefore, the minimum of S equals $\frac{n(n+1)}{2}$.

In this situation, we know that in the Rational product $\prod_{l=1}^{t} R_l$,

$$R_l = \begin{cases} T_{p_l,m_l} & 1 \le l \le n \\ A_l \otimes B_l & l > n \end{cases}$$

If $n \le t - 1$,

$$\prod_{l=1}^{t} R_l = \prod_{l=1}^{n} T_{p_l,m_l} \prod_{l'=n+1}^{t} A_{l'} \otimes B_{l'}$$

According to the formula (2),

$$\prod_{l'=n+1}^{t} A_{l'} \otimes B_{l'} = \left(\prod_{i=n+1}^{t} A_i\right) \otimes \left(\prod_{j=n+1}^{t} B_j\right)$$

According to the formula (3),

$$\prod_{l=1}^{n} T_{p_l,m_l} = \begin{cases} T_{p_1,m_1} \text{when } n \text{ is an odd number} \\ I_{p_1 m_1} \text{when } n \text{ is an even number} \end{cases}$$

Thus

$$\prod_{l=1}^{t} R_l = \begin{cases} T_{p_1,m_1}\left(\prod_{i=n+1}^{t} A_i\right)\otimes\left(\prod_{j=n+1}^{t} B_j\right) \\ \text{when } n \text{ is an odd number} \\ I_{p_1 m_1}\left(\prod_{i=n+1}^{t} A_i\right)\otimes\left(\prod_{j=n+1}^{t} B_j\right) \\ \quad =\left(\prod_{i=n+1}^{t} A_i\right)\otimes\left(\prod_{j=n+1}^{t} B_j\right) \\ \text{when } n \text{ is an even number} \end{cases}$$

which is the statement of Lemma 2. If $n = t$, then

$$\begin{aligned} \prod_{l=1}^{t} R_l &= \prod_{l=1}^{t} T_{p_l,m_l} \\ &= \begin{cases} T_{p_1,m_1} \text{when } t \text{ is an odd number} \\ I_{p_1 m_1} \text{when } t \text{ is an even number} \end{cases} \\ &= \begin{cases} T_{p_1,m_1}\left(I_{m_1}\otimes I_{p_1}\right) \text{when } t \text{ is an odd number} \\ \left(I_{m_1}\otimes I_{p_1}\right) \text{when } t \text{ is an even number} \end{cases} \end{aligned}$$

which also satisfies Lemma 2. ■

Theorem 2: Under Condition 2, the derivative of $f(X)$ can be obtained based on Vec Approach with its time complexity $O\left(n^3\right)$.

Proof: Note that the original derivatives is given as

$$\begin{aligned} f'(X) &= \frac{\partial f(F_1)}{\partial vec(F_1)} \prod_{i=1}^{r-1} \frac{\partial vec(F_i)}{\partial vec(F_{i+1})} \\ &= \left(vec\left(\frac{\partial f(F_1)}{\partial F_1}\right)\right)^T \prod_{i=1}^{r-1} \frac{\partial vec(F_i)}{\partial vec(F_{i+1})} \end{aligned}$$

If

$$\prod_{i=1}^{r-1} \frac{\partial vec(F_i)}{\partial vec(F_{i+1})}$$

is Rational, according to Lemma 2, there are two cases:

Case 1:

$$\prod_{i=1}^{r-1} \frac{\partial vec(F_i)}{\partial vec(F_{i+1})} = (G\otimes H)$$

Then

$$\begin{aligned} \frac{\partial f}{\partial vec(X)} &= \left(vec\left(\frac{\partial f(F_1)}{\partial F_1}\right)\right)^T (G\otimes H) \\ \Leftrightarrow \left(vec\left(\frac{\partial f}{\partial X}\right)\right)^T &= \left(vec\left(\frac{\partial f(F_1)}{\partial F_1}\right)\right)^T (G\otimes H) \\ &= \left((G\otimes H)^T vec\left(\frac{\partial f(F_1)}{\partial F_1}\right)\right)^T \\ &= \left(\left(G^T\otimes H^T\right) vec\left(\frac{\partial f(F_1)}{\partial F_1}\right)\right)^T \end{aligned}$$

According to (4)

$$\begin{aligned} &= \left(vec\left(H^T \frac{\partial f(F_1)}{\partial F_1} G\right)\right)^T \\ \Leftrightarrow \frac{\partial f}{\partial X} &= H^T \frac{\partial f(F_1)}{\partial F_1} G \end{aligned}$$

Case 2:

$$\prod_{i=1}^{r-1} \frac{\partial vec(F_i)}{\partial vec(F_{i+1})} = T_{p,m}(G\otimes H)$$

Then

$$\begin{aligned} \frac{\partial f}{\partial vec(X)} &= \left(vec\left(\frac{\partial f(F_1)}{\partial F_1}\right)\right)^T T_{p,m}(G\otimes H) \\ \Leftrightarrow \left(vec\left(\frac{\partial f}{\partial X}\right)\right)^T &= \left(vec\left(\frac{\partial f(F_1)}{\partial F_1}\right)\right)^T (H\otimes G)T_{q,n} \\ &= \left(T_{n,q} vec\left(G^T \frac{\partial f(F_1)}{\partial F_1} H\right)\right)^T \\ &= \left(vec\left(\left(G^T \frac{\partial f(F_1)}{\partial F_1} H\right)^T\right)\right)^T \\ &= \left(vec\left(H^T \left(\frac{\partial f(F_1)}{\partial F_1}\right)^T G\right)\right)^T \\ \Leftrightarrow \frac{\partial f}{\partial X} &= H^T \left(\frac{\partial f(F_1)}{\partial F_1}\right)^T G \end{aligned}$$

This Theorem holds. ■

Before proving the equivalence of Condition 1 and Condition 2 in Theorem 3, we propose the following lemma.

Lemma 3: For any Rational product $\prod_{l=1}^{t} R_l$ there exists Matrix A&B s.t., for any integer $i, j, u, v > 0$,

$$\left(\prod_{l=1}^{t} R_l\right)_{((v-1)q+u,(j-1)p+i)} = A_{(i,u)}\cdot B_{(j,v)}$$

Or

$$\left(\prod_{l=1}^{t} R_l\right)_{((v-1)q+u,(j-1)p+i)} = A_{(i,v)}\cdot B_{(j,u)}$$

where $q = KH(R_1)$ and $p = KL(R_t)$.

Proof:

$$\begin{aligned} &\left(\prod_{l=1}^{t} R_l\right)_{((v-1)q+u,(j-1)p+i)} \\ &= \sum_{x_1}\sum_{x_2}\sum_{x_3}\cdots\sum_{x_{t-1}} R_{1_{((v-1)q+u,x_1)}} \\ &\quad \cdot R_{2_{(x_1,x_2)}}\cdot\ldots\cdot R_{t-1_{(x_{t-2},x_{t-1})}} R_{t_{(x_{t-1},(j-1)p+i)}} \end{aligned}$$

Suppose

$$\begin{aligned} x_0 &= (v-1)q+u \\ x_t &= (j-1)p+i \end{aligned}$$

Because $\prod_{l=1}^{t} R_l$ is Rational product, then

$$KL(R_l) = KH(R_{l+1})$$

For $l \in \{1, 2, \ldots, t\}$, we can suppose

$$x_l = (j_l - 1)\, p_l + i_l$$

where

$$p_l = KL(R_l) = KH(R_{l+1})$$

and

$$p_0 = q,\ v = j_0,\ u = i_0$$

such that

$$x_0 = (j_0 - 1)\, p_0 + i_0$$

Therefore,

$$\begin{aligned}&\left(\prod_{l=1}^{t} R_l\right)_{((v-1)q+u,(j-1)p+i)}\\ &= \sum_{i_1}\sum_{i_2}\sum_{i_3}\cdots\sum_{i_{t-2}}\sum_{i_{t-1}}\sum_{j_1}\sum_{j_2}\sum_{j_3}\cdots\sum_{j_{t-2}}\sum_{j_{t-1}}\\ &\quad\times R_{1((v-1)q+u,(j_1-1)p_1+i_1)}\cdot R_{2((j_1-1)p_1+i_1,(j_2-1)p_2+i_2)}\\ &\quad\cdot\ldots\cdot R_{t-1((j_{t-2}-1)p_{t-2}+i_{t-2},(j_{t-1}-1)p_{t-1}+i_{t-1})}\\ &\quad\cdot R_{t((j_{t-1}-1)p_{t-1}+i_{t-1},(j_t-1)p_t+i_t)}\\ &= \sum_{i_1}\sum_{i_2}\sum_{i_3}\cdots\sum_{i_{t-2}}\sum_{i_{t-1}}\sum_{j_1}\sum_{j_2}\sum_{j_3}\cdots\sum_{j_{t-2}}\sum_{j_{t-1}}\prod_{l=1}^{t}\\ &\quad\times (R_l)_{((j_{l-1}-1)p_{l-1}+i_{l-1},(j_l-1)p_l+i_l)}\end{aligned}$$

According to the definition of Rational product,

(i) If $R_l = A \otimes B$, then

from the definition of Kronecker product, we have

$$(A \otimes B)_{((v-1)q+u,(j-1)p+i)} = A_{(v,j)} \cdot B_{(u,i)}$$

where $q = KH(A \otimes B) = KH(R_l) = p_{l-1}$ and $p = KL(A \otimes B) = KL(R_l) = p_l$.

(ii) If $R_l = T_{p,q}$, then

$q = KH(R_l) = KH\left(T_{p,q}\right)$ and $p = KL(R_l) = KL\left(T_{p,q}\right)$. Thus,

$$\left(T_{p,q}\right)_{((v-1)q+u,(j-1)p+i)} = \delta_{v,i} \cdot \delta_{j,u}$$

where

$$\delta_{a,b} = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases}$$

In summary, for $\forall l \in \{1, 2, 3, \ldots, t\}$, their exist Matrix S_l and T_l, such that

$$(R_l)_{((j_{l-1}-1)p_{l-1}+i_{l-1},(j_l-1)p_l+i_l)} = (S_l)_{(i_{l-1},j_l)} \cdot (T_l)_{(j_{l-1},i_l)}$$

or

$$(R_l)_{((j_{l-1}-1)p_{l-1}+i_{l-1},(j_l-1)p_l+i_l)} = (S_l)_{(i_{l-1},i_l)} \cdot (T_l)_{(j_{l-1},j_l)}$$

Therefore,

$$\begin{aligned}&\left(\prod_{l=1}^{t} R_l\right)_{((v-1)q+u,(j-1)p+i)}\\ &= \sum_{i_1}\sum_{i_2}\sum_{i_3}\cdots\sum_{i_{t-2}}\sum_{i_{t-1}}\sum_{j_1}\sum_{j_2}\\ &\quad\times\sum_{j_3}\cdots\sum_{j_{t-2}}\sum_{j_{t-1}}\prod_{l=1}^{t}(S_l)_{(i_{l-1},*)}\cdot(T_l)_{(j_{l-1},**)}\end{aligned} \quad (5)$$

where "*" and "**" represents "j_l and i_l" or "$i_l$and j_l", which depend on l.

Considering the two series $\{S_l\}$ and $\{T_l\}$ on the right-hand side of the equation (5), we find that each i_l and j_l in the subscript would occur twice, where $l = 1, 2, \ldots, t-1$. For $j_0, i_0, i_t and j_t$, each of them occurs only once. It also shows that their occurrence is sequential.

Now we give two series of matrixes $\{G_l\}$ and $\{H_l\}$ as follows: 1)

1) $G_1 = S_1, H_1 = T_1$;
2) for $l = 2, 3, \ldots, t$, if the subscript of S_{l-1} in equation (5) is (i_{l-1}, j_l), choose $G_l = T_l$, and $H_l = S_l$; otherwise the subscript of S_{l-1} in equation (5) is (i_{l-1}, i_l), then we choose $G_l = S_l$, and $H_l = T_l$;

In this way, we can find that the column subscript of G_{l-1} is equal to the row subscript of G_l, and so is that of H_{l-1} to H_l. Therefore, we can denote G_l and H_l respectively as $G_{l_{(g_{l-1},g_l)}}$ and $H_{l_{(h_{l-1},h_l)}}$, where $l = 1, 2, 3, \ldots, t$ and $(g_l, h_l) =$ either (i_l, j_l) or (j_l, i_l) and $g_0 = i_0 = u$, $h_0 = j_0 = v$

Now transform equation (5) into the following equation:

$$\begin{aligned}&\left(\prod_{l=1}^{t} R_l\right)_{((v-1)q+u,(j-1)p+i)}\\ &= \sum_{g_1}\sum_{g_2}\sum_{g_3}\cdots\sum_{g_{t-2}}\sum_{g_{t-1}}\sum_{h_1}\sum_{h_2}\\ &\quad\times\sum_{h_3}\cdots\sum_{h_{t-2}}\sum_{h_{t-1}}\prod_{l=1}^{t}(G_l)_{(g_{l-1},g_l)}\cdot(H_l)_{(h_{l-1},h_l)}\\ &= \sum_{g_1}\sum_{g_2}\sum_{g_3}\cdots\sum_{g_{t-2}}\sum_{g_{t-1}}\prod_{l=1}^{t}(G_l)_{(g_{l-1},g_l)}\sum_{h_1}\sum_{h_2}\\ &\quad\times\sum_{h_3}\cdots\sum_{h_{t-2}}\sum_{h_{t-1}}\prod_{l=1}^{t}(H_l)_{(h_{l-1},h_l)}\\ &= \left(\prod_{l=1}^{t} G_l\right)_{(g_0,g_t)}\cdot\left(\prod_{l=1}^{t} H_l\right)_{(h_0,h_t)}\end{aligned}$$

Because

$$g_0 = i_0 = u$$
$$h_0 = j_0 = v$$

and $(g_t, h_t) =$ either $(i_t, j_t) = (i, j)$ or $(j_t, i_t) = (j, i)$, so

$$\left(\prod_{l=1}^{t} R_l\right)_{((v-1)q+u,(j-1)p+i)} = \left(\prod_{l=1}^{t} G_l\right)_{(u,i)}\cdot\left(\prod_{l=1}^{t} H_l\right)_{(v,j)}$$

or

$$\left(\prod_{l=1}^{t} R_l\right)_{((v-1)q+u,(j-1)p+i)} = \left(\prod_{l=1}^{t} G_l\right)_{(u,j)} \cdot \left(\prod_{l=1}^{t} H_l\right)_{(v,i)}$$

Thus, the lemma 3 is proved. ■

Theorem 3: Condition 1 and Condition 2 are equivalent.

Proof: The idea is to first prove that Condition 1 implies Condition 2 (*Condition* 1 ⇒ *Condition* 2), and then vice versa(*Condition* 2 ⇒ *Condition* 1).

Condition 1 ⇒ *Condition* 2: In the case that

$$\frac{\partial F_{k-1_{(i_{k-1},j_{k-1})}}}{\partial F_{k_{(i_k,j_k)}}} = A_{k_{(i_{k-1},i_k)}} \cdot C_{k_{(j_{k-1},j_k)}},$$

On one hand,

$$A_{k_{(i_{k-1},i_k)}} \cdot C_{k_{(j_{k-1},j_k)}} = (C_k \otimes A_k)_{((j_{k-1}-1)m_{k-1}+i_{k-1},(j_k-1)m_k+i_k)}$$

On the other hand,

$$\frac{\partial F_{k-1_{(i_{k-1},j_{k-1})}}}{\partial F_{k_{(i_k,j_k)}}} = \left(\frac{\partial vec\,(F_{k-1})}{\partial vec\,(F_k)}\right)_{((j_{k-1}-1)m_{k-1}+i_{k-1},(j_k-1)m_k+i_k)}$$

Therefore,

$$\frac{\partial vec\,(F_{k-1})}{\partial vec\,(F_k)} = C_k \otimes A_k$$

From Definition 4, we know that $C_k \otimes A_k$ is Rational. Also

$$KH\,(C_k \otimes A_k) = m_{k-1}$$

and

$$KL\,(C_k \otimes A_k) = m_k$$

In the case that

$$\frac{\partial F_{k-1_{(i_{k-1},j_{k-1})}}}{\partial F_{k_{(i_k,j_k)}}} = A_{k_{(i_{k-1},j_k)}} \cdot C_{k_{(j_{k-1},i_k)}},$$

it is equivalent to

$$\frac{\partial F_{k-1_{(i_{k-1},j_{k-1})}}}{\partial F^T_{k_{(j_k,i_k)}}} = A_{k_{(i_{k-1},j_k)}} \cdot C_{k_{(j_{k-1},i_k)}}$$

If we define $D_{k-1} = F_{k-1}$, and $D_k = F_k^T$, then the case turns to be:

$$\frac{\partial D_{k-1_{(i_{k-1},j_{k-1})}}}{\partial D_{k_{(j_k,i_k)}}} = A_{k_{(i_{k-1},j_k)}} \cdot C_{k_{(j_{k-1},i_k)}}$$

which is identical to the former case. Therefore,

$$\frac{\partial vec\,(D_{k-1})}{\partial vec\,(D_k)} = C_k \otimes A_k$$

That is to say

$$\frac{\partial vec\,(F_{k-1})}{\partial vec\,(F_k^T)} = C_k \otimes A_k$$

According to the chain rule

$$\begin{aligned}\frac{\partial vec\,(F_{k-1})}{\partial vec\,(F_k)} &= \frac{\partial vec\,(F_{k-1})}{\partial vec\,(F_k^T)} \cdot \frac{\partial vec\,(F_k^T)}{\partial vec\,(F_k)}\\ &= \frac{\partial vec\,(F_{k-1})}{\partial vec\,(F_k^T)} \cdot \frac{\partial T_{m_k,n_k} vec\,(F_k)}{\partial vec\,(F_k)}\\ &= \frac{\partial vec\,(F_{k-1})}{\partial vec\,(F_k^T)} \cdot T_{m_k,n_k}\\ &= (C_k \otimes A_k)\,T_{m_k,n_k}\end{aligned}$$

Also

$$XL\,(C_k \otimes A_k) = m_k$$

Thus, $(C_k \otimes A_k)\,T_{m_k,n_k}$ is Rational, and

$$KH\,(C_k \otimes A_k) = m_{k-1}$$
$$KL\,(T_{m_k,n_k}) = m_k$$

Condition 2 ⇒ *Condition* 1:
Suppose

$$\frac{\partial vec\,(F_{k-1})}{\partial vec\,(F_k)} = \prod_{l=1}^{t} R_l$$

Because

$$\frac{\partial F_{k-1_{(i_{k-1},j_{k-1})}}}{\partial F_{k_{(i_k,j_k)}}} = \left(\frac{\partial vec\,(F_{k-1})}{\partial vec\,(F_k)}\right)_{((j_{k-1}-1)m_{k-1}+i_{k-1},(j_k-1)m_k+i_k)}$$

so we have

$$\frac{\partial F_{k-1_{(i_{k-1},j_{k-1})}}}{\partial F_{k_{(i_k,j_k)}}} = \left(\prod_{l=1}^{t} R_l\right)_{((j_{k-1}-1)m_{k-1}+i_{k-1},(j_k-1)m_k+i_k)}$$

Because

$$m_{k-1} = KH\,(R_1)$$

and

$$m_k = KL\,(R_t)$$

according to the Lemma 3, there exists Matrix A & B such that

$$\left(\prod_{l=1}^{t} R_l\right)_{((j_{k-1}-1)m_{k-1}+i_{k-1},(j_k-1)m_k+i_k)} = A_{(i_k,i_{k-1})} \cdot B_{(j_k,j_{k-1})}$$

or

$$\left(\prod_{l=1}^{t} R_l\right)_{((j_{k-1}-1)m_{k-1}+i_{k-1},(j_k-1)m_k+i_k)} = A_{(i_k,j_{k-1})} \cdot B_{(j_k,i_{k-1})}$$

Let

$$A_k = A$$
$$C_k = B$$

Therefore

$$\frac{\partial F_{k-1_{(i_{k-1}, j_{k-1})}}}{\partial F_{k_{(i_k, j_k)}}} = A_{k_{(i_k, i_{k-1})}} \cdot C_{k_{(j_k, j_{k-1})}}$$

or

$$\frac{\partial F_{k-1_{(i_{k-1}, j_{k-1})}}}{\partial F_{k_{(i_k, j_k)}}} = A_{k_{(i_k, j_{k-1})}} \cdot C_{k_{(j_k, i_{k-1})}}$$

This Theorem holds. ■

IV. CASE STUDIES

A. MATRIX DERIVATIVES BASED ON INDEX APPROACH OR VEC APPROACH

As mentioned, the applications of matrix-to-scalar function derivatives have been extensively involved in various optimization problems. As claimed in Figure 1, the main contributions of this work are illustrated in the parts in the dashed rectangular frame in this Figure 1. Here we present an example to illustrate this.

In [13], the problem of minimum-cost control of complex networks is addressed, which has received lots of attention recently, see for examples [14], [35], [36]. The problem is modelled as driving the network's state to the origin in the time interval $[0, t_f]$ so as to minimize the following cost

$$\varepsilon(t_f) = \min_{u(t), B} \quad E\left[\int_0^{t_f} ||u(t)||^2 dt\right]$$

where $u(t)$ is the input to be designed and $B \in \mathbb{R}^{n\times m}$ is an input matrix variable that usually subject to certain constraints. The physical meaning of n and m are seen in [14]. Existing results in [13] reveal that such a problem can be converted into minimizing the following cost function

$$\min tr\left[W_B^{-1}X_f\right] = tr\left[\left(\int_0^{t_f} e^{At}BB^Te^{A^Tt}dt\right)^{-1} X_f\right] \tag{6}$$

where X_f is a constant matrix given by $X_f \triangleq \varepsilon\left[x_f x_f^T\right] = e^{At_f}X_0e^{A^Tt_f}$, $E(B) \triangleq tr\left[W_B^{-1}X_f\right]$ and $W_B \triangleq \int_0^{t_f} e^{At}BB^Te^{A^Tt}dt$ and $A \in \mathbb{R}^{n\times n}$. Obviously, the above problem is an optimization problem on matrix manifold.

How to obtain the derivative of the cost function with respect to the matrix variable in the above problem is a key technique when solving the problem by exploring the gradient information. Without losing generality, suppose that $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$. Then, we have $X_f \in \mathbb{R}^{n\times n}$ and $W_B \in \mathbb{R}^{n\times n}$ and obtain that

$$\frac{\partial E(B)}{\partial B} = \frac{\partial tr\left(W_B^{-1}X_f\right)}{\partial vec\left(W_B^{-1}X_f\right)} \cdot \frac{\partial vec\left(W_B^{-1}X_f\right)}{\partial vec\left(W_B^{-1}\right)} \cdot \frac{\partial vec\left(W_B^{-1}\right)}{\partial vec(W_B)} \cdot \frac{\partial vec\left(\int_0^{t_f} e^{At}BB^Te^{A^Tt}dt\right)}{\partial vec(B)} \tag{7}$$

In [22], the following results are shown:

1) $\frac{\partial tr(X)}{\partial X} = I$
2) $\frac{\partial (XA)_{i,j}}{\partial X_{m,n}} = \delta_{i,m}A_{n,j}$
3) $\frac{\partial (X^{-1})_{k,l}}{\partial X_{i,j}} = -\left(X^{-1}\right)_{k,i}\left(X^{-1}\right)_{j,l}$
4) $\frac{\partial (X^TX)_{k,l}}{\partial X_{i,j}} = \delta_{l,j}X_{k,i}^T + \delta_{k,j}X_{i,l}$

According to these four conclusions, we contend that the right hand side of (7) is qualified for Condition 1. Therefore, we can get the following result via Index Approach, which has been shown in [13].

$$\frac{\partial E(B)}{\partial B} = -2\int_0^{t_f} e^{A^Tt}W_B^{-1}X_fW_B^{-1}e^{At}dtB \tag{8}$$

Note that the time complexity of calculating (8) is $O\left(n^3\right)$ since only $n\times n$ dimensional matrix inverse and multiplication are involved.

Consequently, because Condition 1 is proven to be equivalent to Condition 2, there should exist a Vec Approach derivative of $E(B)$, which is derived as follows.

$$\begin{aligned}
&\frac{\partial E(B)}{\partial vec(B)} \\
&= \frac{\partial tr\left(W_B^{-1}X_f\right)}{\partial vec\left(W_B^{-1}\right)} \cdot \frac{\partial vec\left(W_B^{-1}\right)}{\partial vec(W_B)} \cdot \frac{\partial vec(W_B)}{\partial vec(B)} \\
&= vec\left(X_f^T\right)^T \cdot \left(-W_B^{-T} \otimes W_B^{-1}\right) \cdot \int_0^{t_f} \frac{\partial vec\left(e^{At}BB^Te^{A^Tt}\right)}{\partial vec(B)}dt \\
&= -vec\left(X_f^T\right)^T \cdot \left(W_B^{-T} \otimes W_B^{-1}\right) \cdot \int_0^{t_f} \frac{\partial vec\left(e^{At}BB^Te^{A^Tt}\right)}{\partial vec(B)}dt \\
&= -vec\left(X_f^T\right)^T \cdot \left(W_B^{-T} \otimes W_B^{-1}\right) \cdot \int_0^{t_f} \left(\left(e^{A^Tt}\right)^T \otimes e^{At}\right) \\
&\quad \cdot \left(I_{m^2} + T_{m,m}\right) \cdot (B \otimes I_n)\, dt \\
&= -vec\left(X_f^T\right)^T \cdot \left(W_B^{-T} \otimes W_B^{-1}\right) \cdot \int_0^{t_f} \left(e^{At} \otimes e^{At}\right) \\
&\quad \cdot \left(I_{n^2} + T_{n,n}\right) \cdot (B \otimes I_n)\, dt \\
&= -\int_0^{t_f} vec\left(X_f^T\right)^T \cdot \left(W_B^{-T} \otimes W_B^{-1}\right) \cdot \left(e^{At} \otimes e^{At}\right) \\
&\quad \cdot \left(I_{n^2} + T_{n,n}\right) \cdot (B \otimes I_n)\, dt
\end{aligned} \tag{9}$$

Obviously the time complexity of calculating (9) is $O\left(n^6\right)$ as multiplication of $n^2 \times n^2$ dimensional matrices is involved. Although the complexity can be reduced to $O\left(n^4\right)$ by calculating the product of vector with the following matrices sequentially, it is much higher than calculating (8).

Note that (9) is a Rational product, which means (9) satisfies Condition 2. As a result, by omitting '·' and using the three equations (1), (2) and (3), the above formula can be transformed into

$$\begin{aligned}
&-\int_0^{t_f} vec\left(X_f^T\right)^T \left(W_B^{-T} \otimes W_B^{-1}\right)\left(e^{At} \otimes e^{At}\right) \\
&\quad \times \left(I_{n^2} + T_{n,n}\right)(B \otimes I_n)\, dt
\end{aligned}$$

$$= -\int_0^{t_f} \left(\left(W_B^{-1} \otimes W_B^{-T}\right) vec\left(X_f^T\right)\right)^T \left(e^{At} \otimes e^{At}\right)$$
$$\times \left(I_{n^2} + T_{n,n}\right)(B \otimes I_n)\, dt$$
$$= -\int_0^{t_f} \left(vec\left(W_B^{-T} X_f^T W_B^{-T}\right)\right)^T \left(e^{At} \otimes e^{At}\right)$$
$$\times \left(I_{n^2} + T_{n,n}\right)(B \otimes I_n)\, dt$$
$$= -\int_0^{t_f} \left(\left(e^{A^T t} \otimes e^{A^T t}\right) vec\left(W_B^{-T} X_f^T W_B^{-T}\right)\right)^T$$
$$\times \left(I_{n^2} + T_{n,n}\right)(B \otimes I_n)\, dt$$
$$= -\int_0^{t_f} \left(vec\left(e^{A^T t} W_B^{-T} X_f^T W_B^{-T} e^{At}\right)\right)^T$$
$$\times \left(I_{n^2} + T_{n,n}\right)(B \otimes I_n)\, dt$$
$$= -\int_0^{t_f} \left(vec\left(e^{A^T t} W_B^{-T} X_f^T W_B^{-T} e^{At}\right)\right)^T$$
$$\times \left((B \otimes I_n) + T_{n,n}(B \otimes I_n)\right) dt$$
$$= -\int_0^{t_f} \left(\left(B^T \otimes I_n\right) vec\left(e^{A^T t} W_B^{-T} X_f^T W_B^{-T} e^{At}\right)\right)^T$$
$$+ \left(\left(B^T \otimes I_n\right) T_{n,n} vec\left(e^{A^T t} W_B^{-T} X_f^T W_B^{-T} e^{At}\right)\right)^T dt$$
$$= -\int_0^{t_f} \left(vec\left(e^{A^T t} W_B^{-T} X_f^T W_B^{-T} e^{At} B\right)\right)^T$$
$$\times \left(vec\left(e^{A^T t} W_B^{-1} X_f W_B^{-1} e^{At} B\right)\right)^T dt$$
$$= -\int_0^{t_f} \left(vec\left(e^{A^T t}\left(W_B^{-T} X_f^T W_B^{-T}\right.\right.\right.$$
$$\left.\left.\left.+ W_B^{-1} X_f W_B^{-1}\right) e^{At} B\right)\right)^T dt$$

Because $X_f = \varepsilon\left[x_f x_f^T\right] = e^{At_f} X_0 e^{A^T t_f}$ is a symmetrical matrix, we have that $X_f^T = X_f$. Also, $W_B = \int_0^{t_f} e^{At} B B^T e^{A^T t} dt$ is symmetrical. Therefore, W_B^{-1} is a symmetrical matrix. That is to say, $W_B^{-1} = W_B^{-T}$. Then,

$$-\int_0^{t_f} \left(vec\left(e^{A^T t}\left(W_B^{-T} X_f^T W_B^{-T} + W_B^{-1} X_f W_B^{-1}\right) e^{At} B\right)\right)^T dt$$
$$= -2\int_0^{t_f} \left(\text{vec}\left(e^{A^T t} W_B^{-T} X_f W_B^{-T} e^{At} B\right)\right)^T dt$$
$$= vec\left(-2\int_0^{t_f} e^{A^T t} W_B^{-T} X_f W_B^{-T} e^{At} dt B\right)^T$$

Apparently, this is the same result obtained in [13]. Thus, as the Condition 1 is satisfied, we can obtain $O\left(n^3\right)$ complexity for both Index Approach and Vector Approach.

B. A CASE WHEN CONDITION 1 OR 2 DOES NOT HOLD

As mentioned in the Introduction, the original time complexity of determining matrix differentiation can be very high. We can reduce the complexity of determining matrix differentiation to $O\left(n^3\right)$ by employing the Index Approach or Vector Approach if either Condition 1 or 2 holds. It is also pointed out that such a $O\left(n^3\right)$ complexity can be achieved in various cases, as almost all matrix derivatives in science and engineering applications can be exactly decomposed into the product of two or more matrices with their index notations, which represent the position of each element of the derivatives, corresponding to those of the functions and variables. However, it is necessary to mention that we could indeed construct a function that the Index Approach fails to handle. In this case, the time complexity of matrix differentiation based on vector approach is much higher than $O\left(n^3\right)$.

Suppose $A = \{a_{i,j}\}$, $A \in \mathbb{R}^{n\times n}$, define a function $F(A) \in \mathbb{R}^{n\times n}$ such that

$$F(A)_{p,q} = \begin{cases} \prod_{k=1}^{p+q-1} a_{(k,p+q-k)}^{p\cdot q\cdot k} & \text{when } p+q \le n \\ \prod_{k=p+q-n}^{n} a_{(k,p+q-k)}^{p\cdot q\cdot k} & \text{when } p+q \ge n+1 \end{cases}$$

Then, we have

$$\frac{\partial F(A)_{p,q}}{\partial A_{u,v}} = \delta_{(p+q,u+v)} \cdot \begin{cases} \frac{p\cdot q\cdot u}{a_{u,v}} \prod_{k=1}^{p+q-1} a_{(k,p+q-k)}^{p\cdot q\cdot k} \\ \quad \text{when } p+q \le n \\ \frac{p\cdot q\cdot u}{a_{u,v}} \prod_{k=p+q-n}^{n} a_{(k,p+q-k)}^{p\cdot q\cdot k} \\ \quad \text{when } p+q \ge n+1 \end{cases}$$
$$= \delta_{(p+q,u+v)} \cdot p \cdot q \cdot F(A)_{p,q} \cdot \frac{u}{a_{u,v}}$$

It is easy to see that the above equation does not satisfy the Condition 1. Therefore, we define a scalar-to-matrix function, specifically a determinant function $E(A) = det(F(A))$ and consider the derivatives

$$\frac{\partial det(F(A))}{\partial A}$$

We get each element of the derivatives as follows.

$$\left(\frac{\partial det(F(A))}{\partial A}\right)_{u,v}$$
$$= \frac{\partial det(F(A))}{\partial A_{u,v}}$$
$$= \sum_{p=1}^{n}\sum_{q=1}^{n} \frac{\partial det(F(A))}{\partial F(A)_{p,q}} \cdot \frac{\partial F(A)_{p,q}}{\partial A_{u,v}}$$
$$= \sum_{p=1}^{n}\sum_{q=1}^{n} det(F(A)) \cdot (F(A))_{p,q}^{-T} \cdot \delta_{(p+q,u+v)}$$
$$\cdot p \cdot q \cdot F(A)_{p,q} \cdot \frac{u}{a_{u,v}}$$
$$= det(F(A)) \cdot \sum_{p=1}^{n}\sum_{q=1}^{n} \delta_{(p+q,u+v)} \cdot p \cdot q$$
$$\cdot \left((F(A))^{-T} \circ F(A)\right)_{p,q} \cdot \frac{u}{a_{u,v}}$$

where the notation $\circ$ means "Hadamard Product", according to [22] . Thus, it is seen that the above derivative cannot be expressed as a product of a series of ordinary matrices.

Now we express the derivatives in Vec Approach. Because elements in the variable matrix remain unchanged compared to that of Index Approach, the derivatives only differ in expression. As

$$\left(\frac{\partial vec\left(F\left(A\right)\right)}{\partial vec\left(A\right)}\right)_{((q-1)n+p,(v-1)n+u)} = \frac{\partial vec\left(F\left(A\right)\right)_{((q-1)n+p,1)}}{\partial vec\left(A\right)_{((v-1)n+u,1)}} = \frac{\partial F\left(A\right)_{p,q}}{\partial A_{u,v}} = \delta_{(p+q,u+v)} \cdot p \cdot q \cdot F\left(A\right)_{p,q} \cdot \frac{u}{a_{u,v}}$$

we obtain that

$$\frac{\partial det\left(F\left(A\right)\right)}{\partial vec\left(A\right)} = \frac{\partial det\left(F\left(A\right)\right)}{\partial vec\left(F\left(A\right)\right)} \frac{\partial vec\left(F\left(A\right)\right)}{\partial vec\left(A\right)} = det\left(F\left(A\right)\right) \cdot \left(vec\left(F\left(A\right)^{-T}\right)\right)^{T} \frac{\partial vec\left(F\left(A\right)\right)}{\partial vec\left(A\right)}$$

which is a $1 \times n^2$-sized vector multiplied by an $n^2 \times n^2$-sized matrix, and its time complexity on calculating is $O\left(n^4\right)$.

V. DISCUSSION AND CONCLUSION

In the Section 3 we have proposed two conditions for Index Approach and Vec Approach sequentially, and under each condition we have presented the methods to get $O\left(n^3\right)$-time-complexity for the matrix-to-scalar function derivatives. Also, equivalence of the two conditions has been proved.

As a result, any derivative that satisfies Condition 1 or Condition 2 can be transformed to an expression whose time complexity is $O\left(n^3\right)$ in Index Approach and Vec Approach sequentially. Also, since the two conditions are equivalent to each other, any of the two methods is as well suitable for the other one. To summarize, the originally-$O\left(n^4\right)$ derivatives can be therefore written in an $O\left(n^3\right)$ form in those two ways.

There are still some problems for further investigation. For instance, what is a sufficient and necessary condition under which Index Approach and Vec Approach are equivalent to each other? Is there any other approach that can be applied in matrix-to-scalar function derivatives? The application of chain-rule for calculating derivatives will greatly rely on the universality and complexity of those approaches, so these two points are also critical. What's more, extending the results in this work to more general matrix-valued functions of matrices or tensor-valued functions of tensors light up the way of future research.

REFERENCES

[1] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jan. 2017.

[2] C. Xing, S. Ma, and Y. Zhou, "Matrix-monotonic optimization for MIMO systems," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 334–348, Jan. 2015.

[3] M. Jianpeng, Z. Shun, L. Hongyan, Z. Nan, and N. Arumugam, "Iterative LMMSE individual channel estimation over relay networks with multiple antennas," *IEEE Trans. Veh. Technol.*, vol. 67, no. 1, pp. 423–435, Jan. 2018.

[4] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, Jun. 2007.

[5] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[6] Z.-X. Yang, G.-S. Zhao, G. Li, and H.-J. Rong, "Matrix differentiation for capacity region of Gaussian multiple access channels under weighted total power constraint," *Ann. Telecommun.*, vol. 72, nos. 11–12, pp. 703–715, 2017.

[7] D. Xu and D. P. Mandic, "The theory of quaternion matrix derivatives," *IEEE Trans. Signal Process.*, vol. 63, no. 6, pp. 1543–1556, Mar. 2015.

[8] S. Ji and J. Ye, "Linear dimensionality reduction for multi-label classification," in *Proc. IJCAI*, vol. 9. 2009, pp. 1077–1082.

[9] Y. Jia, F. Nie, and C. Zhang, "Trace ratio problem revisited," *IEEE Trans. Neural Netw.*, vol. 20, no. 4, pp. 729–735, Apr. 2009.

[10] C. Shen, J. Kim, F. Liu, L. Wang, and A. van den Hengel, "Efficient dual approach to distance metric learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 394–406, Feb. 2014.

[11] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*. Boston, MA, USA: Cengage Learning, 2014.

[12] R. Wang, F. Nie, X. Yang, F. Gao, and M. Yao, "Robust 2DPCA with non-greedy ℓ_1-norm maximization for image analysis," *IEEE Trans. Cybern.*, vol. 45, no. 5, pp. 1108–1112, May 2015.

[13] G. Li *et al.*, "Minimum-cost control of complex networks," *New J. Phys.*, vol. 18, p. 013012, Jan. 2015.

[14] G. Li, P. Tang, C. Wen, and Z. Meng, "Boundary constraints for minimum cost control of directed networks," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4196–4207, Dec. 2017.

[15] G. Li, F. Zeng, H. Li, and A. K. Qin, "Matrix function optimization problems under orthonormal constraint," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2017.2724068.

[16] M. E. J. Newman, "Modularity and community structure in networks," *Proc. Nat. Acad. Sci. USA*, vol. 103, no. 23, pp. 8577–8582, 2006.

[17] S. Zhang and H. Zhao, "Normalized modularity optimization method for community identification with degree adjustment," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 88, p. 052802, Nov. 2013.

[18] M. Rajih, P. Comon, and R. A. Harshman, "Enhanced line search: A novel method to accelerate PARAFAC," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1128–1147, 2008.

[19] A. Uschmajew, "Local convergence of the alternating least squares algorithm for canonical tensor approximation," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 2, pp. 639–652, 2012.

[20] W. Espig, M. Hackbusch, and A. Khachatryan. (2015). "On the convergence of alternating least squares optimisation in tensor format representations." [Online]. Available: https://arxiv.org/abs/1506.00062

[21] J. Magnus and H. Neudecker, *Matrix Differential Calculus With Applications in Statistics and Econometrics*, 3rd ed. New York, NY, USA: Wiley, 2007.

[22] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Tech. Univ. Denmark, Kongens Lyngby, Denmark, Tech. Rep., 2012. [Online]. Available: http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=3274

[23] T. P. Minka. (1997). *Old and New Matrix Algebra Useful for Statistics*. [Online]. Available: www.stat.cmu.edu/minka/papers/matrix.html

[24] K. M. Abadir and J. R. Magnus, "Matrix algebra," in *Econometric Theory*, vol. 22. Cambridge, U.K.: Cambridge Univ. Press, 2006, pp. 968–972.

[25] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus With Applications in Statistics and Econometrics*. New York, NY, USA: Wiley, 1999, pp. 171–173.

[26] P. S. Dwyer, "Some applications of matrix derivatives in multivariate analysis," *J. Amer. Stat. Assoc.*, vol. 62, no. 318, pp. 607–625, 1967.

[27] E. W. Grafarend and R.-J. You, "Fourth order Taylor–Kármán structured covariance tensor for gravity gradient predictions by means of the Hankel transformation," *GEM-Int. J. Geomath.*, vol. 6, no. 2, pp. 319–342, 2015.

[28] S. Ge, M. Han, and X. Hong, "A fully automatic ocular artifact removal from EEG based on fourth-order tensor method," *Biomed. Eng. Lett.*, vol. 4, no. 1, pp. 55–63, 2014.

[29] W. Vetter, "Derivative operations on matrices," *IEEE Trans. Autom. Control*, vol. AC-15, no. 2, pp. 241–244, Apr. 1970.

[30] E. C. MacRae, "Matrix derivatives with an application to an adaptive linear decision problem," *Ann. Statist.*, vol. 2, no. 2, pp. 337–346, 1974.

[31] P. L. Fackler, *Notes on Matrix Calculus* (Lecture Notes). Raleigh, NC, USA: North Carolina State Univ., 2005. [Online]. Available: http://www4.ncsu.edu/unity/users/p/pfackler/www/MatCalc.pdf

[32] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.

[33] J. Zhong, J. Lu, T. Huang, and D. W. C. Ho, "Controllability and synchronization analysis of identical-hierarchy mixed-valued logical control networks," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3482–3493, Nov. 2017.

[34] Z. A. A. Al-Zhour, "The general (vector) solutions of such linear (coupled) matrix fractional differential equations by using Kronecker structures," *Appl. Math. Comput.*, vol. 232, pp. 498–510, Apr. 2014.

[35] G. Yan, J. Ren, Y.-C. Lai, C.-H. Lai, and B. Li, "Controlling complex networks: How much energy is needed?" *Phys. Rev. Lett.*, vol. 108, p. 218703, May 2012.

[36] A. Li, S. P. Cornelius, Y.-Y. Liu, L. Wang, and A.-L. Barabasi, "The fundamental advantages of temporal networks," *Science*, vol. 358, no. 6366, pp. 1042–1046, 2017.

JIANYU XU is currently pursuing the bachelor's degree with the Department of Precision Instrument, Tsinghua University, China. He is also a member of the 11th group of the Spark Research Training Program at Tsinghua University. His research interests include tensor computation, complex network, and control and optimization. He was a winner of the First Prize in the National Senior High School Mathematical Contest and ranked 1st in Heilongjiang Province, and the Second Prize twice in Chinese Mathematical Olympiad.

GUOQI LI received the B.Eng. degree from the Xi'an University of Technology, China, in 2004, the M.Eng. degree from Xi'an Jiaotong University, China, in 2007, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2011. From 2011 to 2014, he was a Scientist with the Data Storage Institute and the Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore. Since 2014, he has been with the Department of Precision Instrument, Tsinghua University, China, where he is currently an Associate Professor. His current research interests include brain inspired computing, complex systems, machine learning, and neuromorphic computing. Dr. Li has authored over 70 journal and conference papers. He is an Editorial-Board Member and a Guest Associate Editor of the *Frontiers in Neuroscience*. He has also actively involved in professional services such as serving as an International Technical Program Committee Member and as a Track Chair for international conferences. He serves as reviewer for a number of international journals.

CHANGYUN WEN (F'10) received the B.Eng. degree from Xi'an Jiaotong University, China, in 1983, and the Ph.D. degree from The University of Newcastle, Australia, in 1990. From 1989 to 1991, he was a Post-Doctoral Fellow with The University of Adelaide. Since 1991, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, where he is currently a Full Professor. His main research activities are control systems and applications, intelligent power management system, smart grids, model-based online learning, and system identification. He received the Institution of Engineers (IES) Prestigious Engineering Achievement Award 2005 from IES, Singapore, in 2005. He is an Associate Editor of a number of journals including *Automatica*, the IEEE Transactions on Industrial Electronics, and the *IEEE Control Systems Magazine*. He is the Executive Editor-in-Chief of the *Journal of Control and Decision*. He served as an Associate Editor for the IEEE Transactions on Automatic Control from 2000 to 2002. He has been actively involved in organizing international conferences playing the roles of the General Chair, the General Co-Chair, the Technical Program Committee Chair, a Program Committee Member, a General Advisor, and the Publicity Chair. He is a member of the IEEE Fellow Committee from 2011 to 2013 and a Distinguished Lecturer of the IEEE Control Systems Society from 2010 to 2013.

KUN WU is currently pursuing the B.Eng. degree with the Department of Electronic Engineering, Tsinghua University, China. He is also an Undergraduate Research Assistant with the Department of Electronic Engineering, Tsinghua University. His current research interests include large-scale graph processing systems, and distributed systems and algorithms.

LEI DENG received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2012, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2017. He currently holds a post-doctoral position with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA, USA. His research interests include brain-inspired computing, machine learning algorithm and architecture, neuromorphic system, computational neuroscience, and complex networks.