ISSN 2186-6570

Matrix differentiation with diagrammatic notation

Kenji Nakahira

Quantum Information Science Research Center,
Quantum ICT Research Institute, Tamagawa University
6-1-1 Tamagawa-gakuen, Machida, Tokyo 194-8610 Japan

Tamagawa University Quantum ICT Research Institute Bulletin, Vol.12, No.1, 1-6, 2022

Tamagawa University Quantum ICT Research Institute Bulletin
Vol.12 No.1 : 1–6 (2022)

Matrix differentiation with diagrammatic notation

Kenji Nakahira
Quantum Information Science Research Center,
Quantum ICT Research Institute, Tamagawa University
6-1-1 Tamagawa-gakuen, Machida, Tokyo 194-8610 Japan
E-mail: nakahira@lab.tamagawa.ac.jp

***Abstract*—We propose a diagrammatic notation for matrix differentiation. Our new notation enables us to derive formulas for matrix differentiation more easily than the usual matrix (or index) notation. We demonstrate the effectiveness of our notation through several examples.**

I. Introduction

Matrix differentiation (or matrix calculus) is widely accepted as an essential tool in various fields including estimation theory, signal processing, and machine learning. This is also utilized in many fields of quantum information theory (e.g., quantum tomography [1], [2], the optimal control of quantum systems [3], and the perturbative analysis of entanglement negativity [4]). Matrix differentiation provides a convenient way to collect the derivative of each component of the dependent variable with respect to each component of the independent variable, where the dependent and independent variables can be a scalar, a vector, or a matrix. However, the usual matrix (or index) notation often suffers from cumbersome calculations and difficulty in the intuitive interpretation of the final results.

It is known that diagrammatic representations using string diagrams can be successfully applied in linear algebra (see [5] and references therein). In this paper, we provide a simple diagrammatic approach to derive useful formulas for matrix differentiation. Note that positive semidefinite matrices and completely positive maps, which can respectively represent quantum states and quantum processes, are regarded as vectors and matrices in the real Hilbert space of Hermitian matrices.

Here we mention some related work. In Ref. [6], the way of graphically representing the del operator (i.e., ∇) is presented, in which calculations are limited to the case of three-dimensional Euclidean space. Reference [7] presents a diagrammatic notation for manipulating tensor derivatives with respect to one parameter. We adopt a similar notation to those given in these references.

II. Definition of matrix differentiation

Let $\mathbb{R}$ be the set of all real numbers and $\mathbb{R}^{m\times n}$ be the set of all $m\times n$ real matrices. Also, let $\{|i\rangle\}_{i=1}^m$ denote the standard basis of $\mathbb{R}^m$. We are concerned only with finite-dimensional real Hilbert spaces. Given a map f from $\mathbb{R}^{m\times n}$ to $\mathbb{R}$ and a matrix $X \in \mathbb{R}^{m\times n}$ of independent variables, we denote the $m\times n$ real matrix with (i,j)-th component $\frac{\partial}{\partial X_{i,j}} f(X)$ by $\frac{\partial}{\partial X} f(X)$, where $X_{i,j} := \langle i|X|j\rangle$ is the (i,j)-th component of X. We have

$$\frac{\partial}{\partial X} f(X) = \sum_{i=1}^m \sum_{j=1}^n \frac{\partial}{\partial X_{i,j}} |i\rangle\langle j| f(X) = \begin{bmatrix} \frac{\partial}{\partial X_{1,1}} f(X) & \frac{\partial}{\partial X_{1,2}} f(X) & \cdots & \frac{\partial}{\partial X_{1,n}} f(X) \\ \frac{\partial}{\partial X_{2,1}} f(X) & \frac{\partial}{\partial X_{2,2}} f(X) & \cdots & \frac{\partial}{\partial X_{2,n}} f(X) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial X_{m,1}} f(X) & \frac{\partial}{\partial X_{m,2}} f(X) & \cdots & \frac{\partial}{\partial X_{m,n}} f(X) \end{bmatrix}. \quad (1)$$

In the special case of $n = 1$, X is a column vector, which is denoted by $|x\rangle$. In this case, we have

$$\frac{\partial}{\partial |x\rangle} f(|x\rangle) = \sum_{i=1}^m \frac{\partial}{\partial x_i} |i\rangle f(|x\rangle) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(|x\rangle) \\ \frac{\partial}{\partial x_2} f(|x\rangle) \\ \vdots \\ \frac{\partial}{\partial x_m} f(|x\rangle) \end{bmatrix},$$

where $x_i := \langle i|x\rangle$.

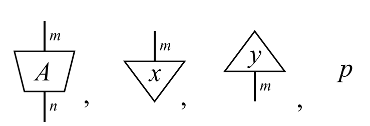
A similar notation is used when f is a map from $\mathbb{R}^{m\times n}$ to $\mathbb{R}^{m'\times n'}$. For such f, $\frac{\partial}{\partial X} f(X)$ is an $m\times n\times m'\times n'$ fourth-order tensor with components $\{\frac{\partial}{\partial X_{i,j}} \langle i'|f(X)|j'\rangle\}_{i,j,i',j'}$. This can be written as the following $mm' \times nn'$ matrix:

$$\frac{\partial}{\partial X} f(X) = \begin{bmatrix} \frac{\partial}{\partial X_{1,1}} f(X) & \frac{\partial}{\partial X_{1,2}} f(X) & \cdots & \frac{\partial}{\partial X_{1,n}} f(X) \\ \frac{\partial}{\partial X_{2,1}} f(X) & \frac{\partial}{\partial X_{2,2}} f(X) & \cdots & \frac{\partial}{\partial X_{2,n}} f(X) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial X_{m,1}} f(X) & \frac{\partial}{\partial X_{m,2}} f(X) & \cdots & \frac{\partial}{\partial X_{m,n}} f(X) \end{bmatrix},$$

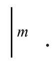
where, for each i and j, $\frac{\partial}{\partial X_{i,j}} f(X)$ is the $m' \times n'$ matrix whose (i',j')-th component is $\frac{\partial}{\partial X_{i,j}} \langle i'|f(X)|j'\rangle$.

III. Diagrammatic notation

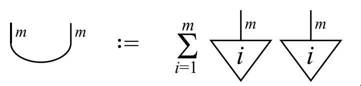
In diagrammatic terms, a matrix is represented as a box with an input wire at the bottom and an output wire at the top. Column vectors, row vectors, and scalars are regarded as special cases of matrices. For example, $A \in \mathbb{R}^{m\times n}$, $|x\rangle \in \mathbb{R}^m := \mathbb{R}^{m\times 1}$, $|y\rangle \in \mathbb{R}^{m*} := \mathbb{R}^{1\times m}$, and $p \in \mathbb{R}$ are diagrammatically depicted as



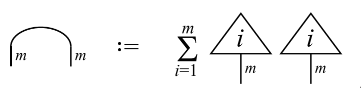
The Hilbert space $\mathbb{R}^m$ is represented by the wire with label m, while the Hilbert space $\mathbb{R}$ is represented by 'no wire'. For a scalar, the box will be omitted. Matrix multiplication and tensor products are represented as the sequential and parallel compositions, respectively. The identity matrix $\mathbb{1} \in \mathbb{R}^{m\times m}$ is depicted as



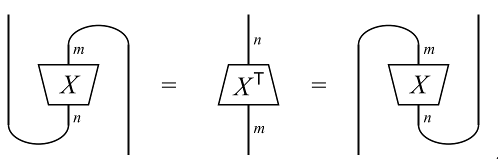
We often use a special column vector $|\cup_n\rangle \in \mathbb{R}^n \otimes \mathbb{R}^n$, called a cup, and a special row vector $\langle\cap_n| \in \mathbb{R}^{n*} \otimes \mathbb{R}^{n*}$, called a cap. The cup $|\cup_n\rangle$ is depicted as



The cap $\langle\cap_n|$ is the transpose of $|\cup_n\rangle$, which is depicted as

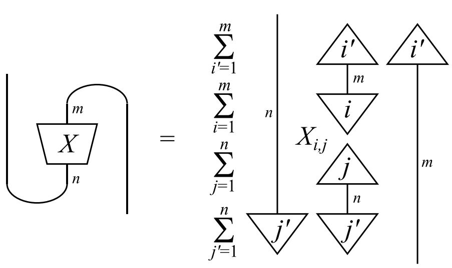


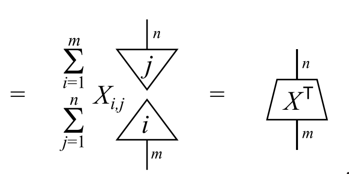
We have that, for any $X \in \mathbb{R}^{m\times n}$,



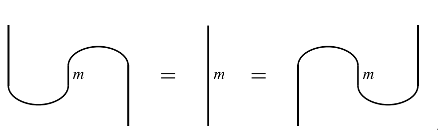
(2)

Indeed, the left equality is obtained from



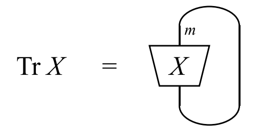


and the same argument works for the right equality. Equation (2) implies that the transpose acts diagrammatically by rotating boxes 180°. Substituting $X = \mathbb{1}$ with Eq. (2) yields

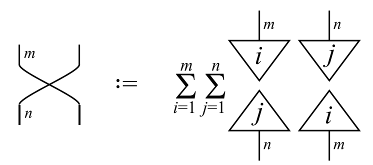


(3)

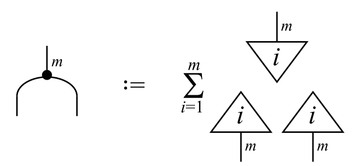
The trace of $X \in \mathbb{R}^{m\times m}$ satisfies $\operatorname{Tr} X = \langle\cap|X \otimes \mathbb{1}|\cup\rangle$, i.e.,



We also use the swap matrix $\times_{n,m}$, depicted by

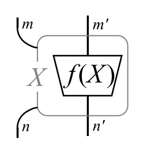


and the matrix called "spider", depicted by

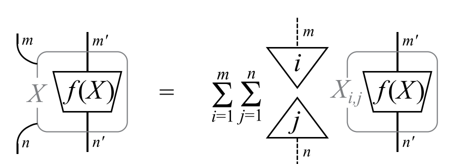


For details regarding the properties of these matrices, see, e.g., Ref. [5].

We write $\frac{\partial}{\partial X} f(X)$ with a map $f : \mathbb{R}^{m\times n} \to \mathbb{R}^{m'\times n'}$ as



From Eq. (1), we have

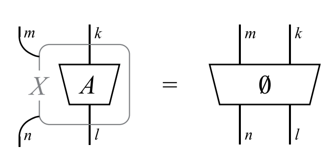


IV. Basic formulas

We review some basic formulas that we shall frequently use later.

A. Derivatives of A and X

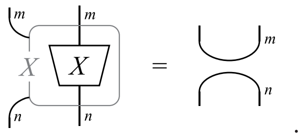
For any matrix A that is independent of X, $\frac{\partial}{\partial X} A = \mathbb{0}$, i.e.,



(4)

holds, where $\mathbb{0}$ is the zero matrix of size $mk \times nl$. In what follows, we assume that matrices $A, B, \ldots$ are independent of X, unless otherwise mentioned. Also, from

$\frac{\partial}{\partial X_{i,j}} X_{k,l} = \delta_{i,k}\delta_{j,l}$ (where $\delta_{i,k}$ is the Kronecker delta), we have $\frac{\partial}{\partial X} X = |\cup_m\rangle\langle\cap_n|$, i.e.,



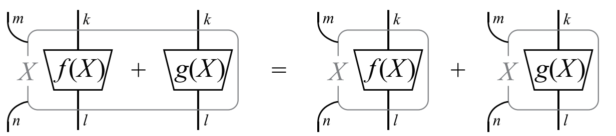
(5)

B. Rules for sums and products

The following sum rule holds:

$$\frac{\partial}{\partial X}[f(X) + g(X)] = \frac{\partial}{\partial X} f(X) + \frac{\partial}{\partial X} g(X),$$

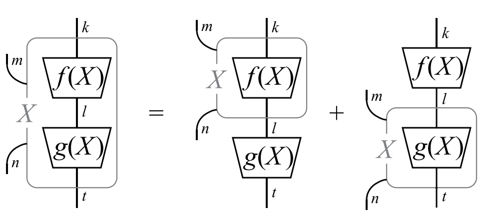
which is diagrammatically represented as



As for matrix multiplication and tensor products, we have

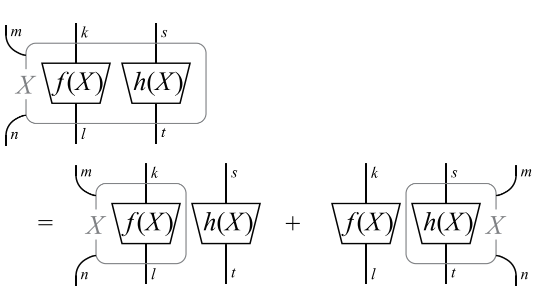
$$\frac{\partial}{\partial X} f(X)g(X) = \left[\frac{\partial}{\partial X} f(X)\right] g(X) + f(X)\left[\frac{\partial}{\partial X} g(X)\right],$$
$$\frac{\partial}{\partial X} f(X) \otimes h(X) = \left[\frac{\partial}{\partial X} f(X)\right] \otimes h(X) + f(X) \otimes \left[\frac{\partial}{\partial X} h(X)\right],$$

which are depicted as



(6)

and



(7)

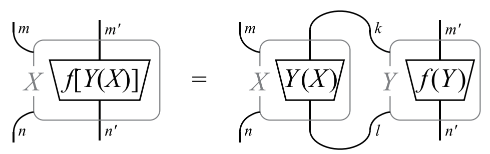
Note that we assume that the order of wires does not matter in a diagram.

C. Chain rules

Given a matrix $X \in \mathbb{R}^{m\times n}$, a map $Y : \mathbb{R}^{m\times n} \to \mathbb{R}^{m'\times n'}$, and a map $f : \mathbb{R}^{m'\times n'} \to \mathbb{R}^{k\times l}$, the derivative of $f[Y(X)]$ with respect to $X_{i,j}$ satisfies

$$\frac{\partial}{\partial X_{i,j}} f[Y(X)] = \sum_{i'=1}^{k}\sum_{j'=1}^{l} \frac{\partial f[Y(X)]}{\partial Y_{i',j'}} \frac{\partial Y_{i',j'}}{\partial X_{i,j}},$$

where $Y_{i',j'} := \langle i'|Y(X)|j'\rangle$. Thus, $\frac{\partial}{\partial X} f[Y(X)]$ can be diagrammatically represented by



(8)

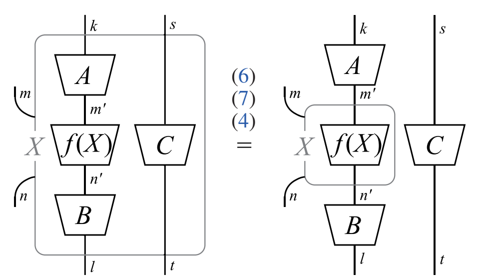
All the formulas presented in this paper can be obtained using the above-mentioned equations. It is noteworthy that this paper is focused on the matrix differentiation, but our notation can be easily extended to the case of high-order tensors.

V. Other basic formulas

We derive several basic formulas.

A. Derivatives of matrix multiplication and tensor products

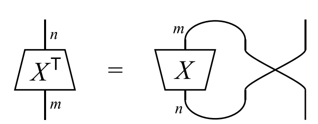
We immediately obtain



(9)

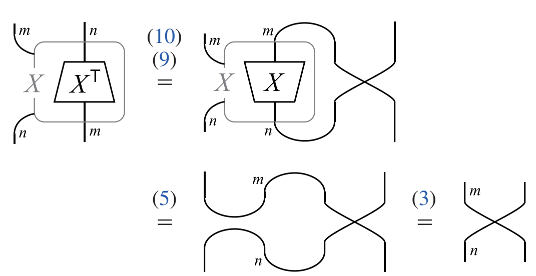
B. Derivative of X^T

Since X^T is represented by



(10)

we have



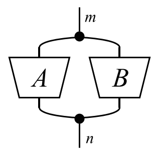
(11)

C. Derivatives of Hadamard products

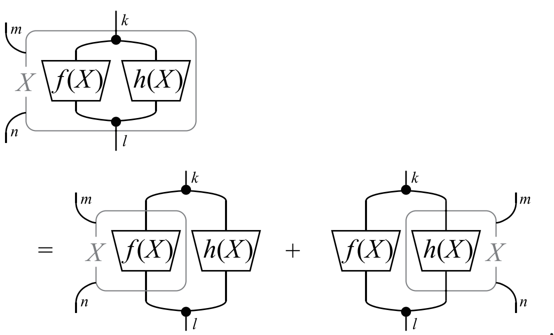
The Hadamard product of $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{m\times n}$, denoted by $A \circ B$, is the component-wise product, i.e.,

$$A \circ B := \sum_{i=1}^{m} \sum_{j=1}^{n} \langle i|A|j\rangle \langle i|B|j\rangle |i\rangle \langle j| ,$$

which is diagrammatically depicted as



From Eq. (11), we can readily verify



(12)

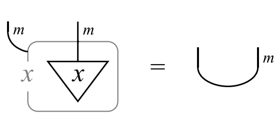
VI. Examples

We will give some concrete examples that are directly derived from the above basic formulas.

A. Derivatives with respect to column vectors

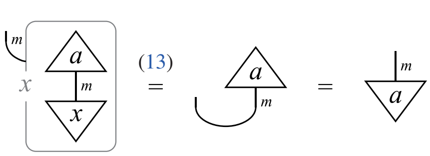
1) $\frac{\partial}{\partial |x\rangle} \langle a|x\rangle = |a\rangle$:

Substituting $n = 1$ into Eq. (5) gives



(13)

Thus, we have

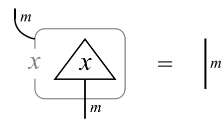


(14)

Note that $\langle a|^{\mathsf{T}} = |a\rangle$ holds since $|a\rangle$ is a real column vector.

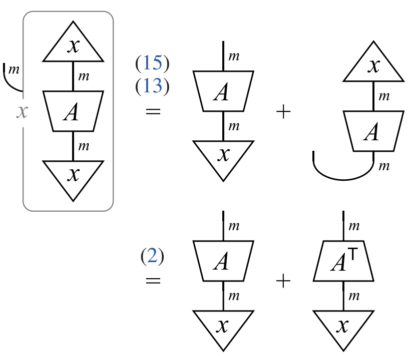
2) $\frac{\partial}{\partial |x\rangle} \langle x|A|x\rangle = (A + A^{\mathsf{T}}) |x\rangle$:

Substituting $n = 1$ into Eq. (11) gives



(15)

and thus



(16)

holds.

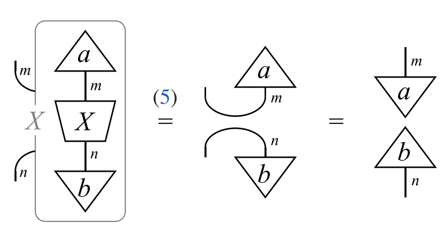
3) Other important examples:

We can easily obtain the following formulas (the proofs are left to the readers) [1]:

$$\frac{\partial}{\partial |x\rangle} \| A|x\rangle - |b\rangle \|_2^2 = 2A^{\mathsf{T}}(A|x\rangle - |b\rangle),$$
$$\frac{\partial}{\partial |x\rangle} \| |x\rangle - |b\rangle \|_2 = \frac{|x\rangle - |b\rangle}{\| |x\rangle - |b\rangle \|_2}.$$

B. Derivatives with respect to matrices

1) $\frac{\partial}{\partial X} \langle a|X|b\rangle = |a\rangle \langle b|$:

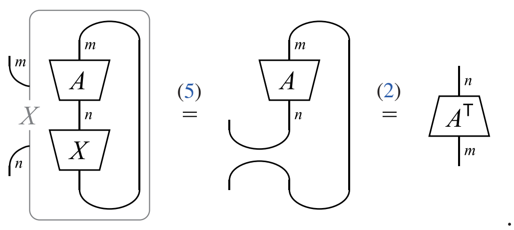


[1] We remind that $\| |v\rangle \|^2 = \langle v|v\rangle$ holds. The second line follows from substituting $u := \| |x\rangle - |b\rangle \|_2^2$ into

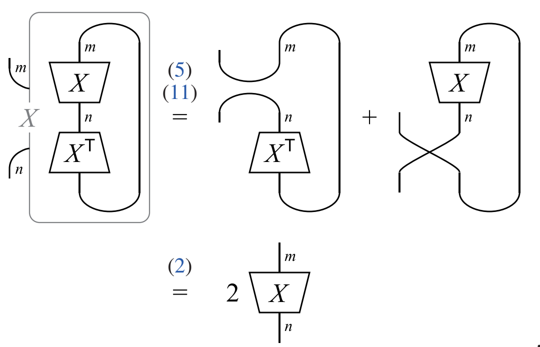
$$\frac{\partial}{\partial |x\rangle} \sqrt{u} = \frac{\partial u}{\partial |x\rangle} \frac{\partial \sqrt{u}}{\partial u} = \frac{\partial u}{\partial |x\rangle} \cdot \frac{1}{2\sqrt{u}},$$

which is immediately obtained by the chain rule, and using $\frac{\partial u}{\partial |x\rangle} = 2(|x\rangle - |b\rangle)$, which is obtained from substituting $A = \mathbb{1}$ into the first line.

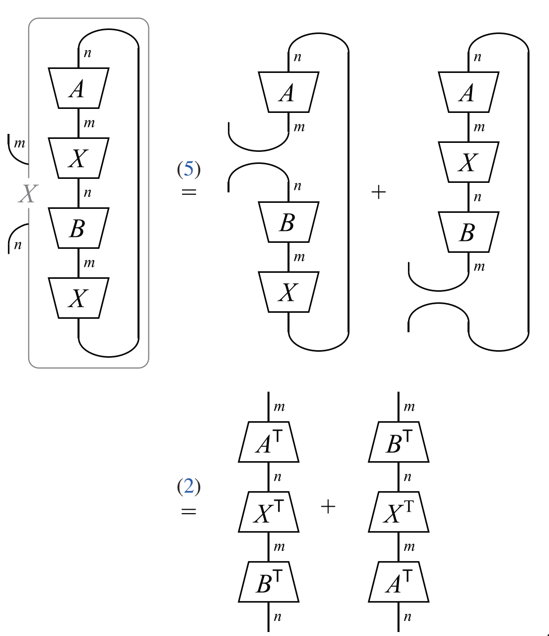
2) $\frac{\partial}{\partial X}\operatorname{Tr}(AX) = A^{\mathsf{T}}$*:*



3) $\frac{\partial}{\partial X}\operatorname{Tr}(XX^{\mathsf{T}}) = 2X$*:*

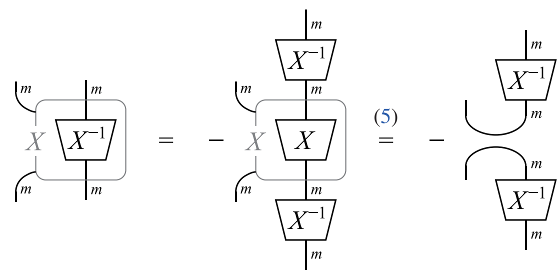


4) $\frac{\partial}{\partial X}\operatorname{Tr}(AXBX) = A^{\mathsf{T}}X^{\mathsf{T}}B^{\mathsf{T}} + B^{\mathsf{T}}X^{\mathsf{T}}A^{\mathsf{T}}$*:*



5) $\frac{\partial}{\partial X}X^{-1} = -(\mathbb{1}\otimes X^{-1})\,|{\cup}\rangle\,\langle{\cap}|\,(\mathbb{1}\otimes X^{-1})$*:*

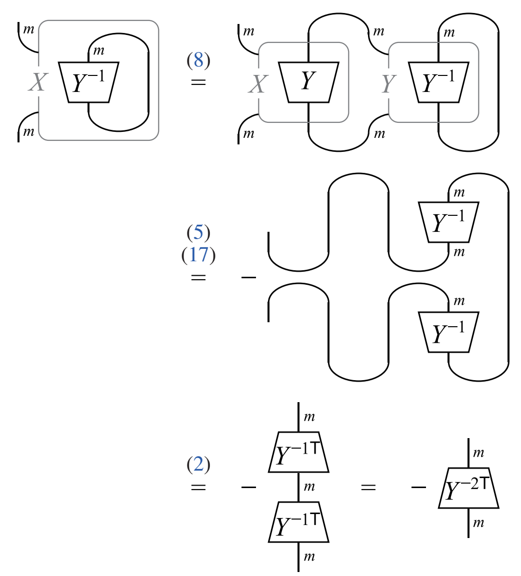
Letting $Z \coloneqq \frac{\partial}{\partial X}X^{-1}$ and differentiating $X^{-1} = X^{-1}XX^{-1}$ with respect to X gives $Z = Z + X^{-1}\frac{\partial X}{\partial X}X^{-1} + Z$. Thus, we have



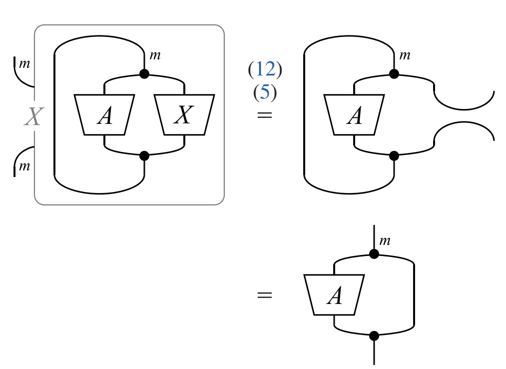
(17)

6) $\frac{\partial}{\partial X}\operatorname{Tr}[(X+A)^{-1}] = -[(X+A)^{-2}]^{\mathsf{T}}$*:*

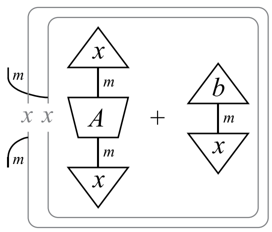
Letting $Y \coloneqq X + A$ and using the chain rule, we obtain



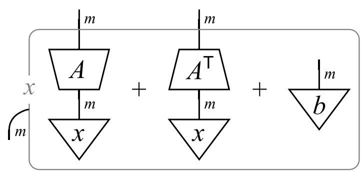
7) $\frac{\partial}{\partial X}\operatorname{Tr}(A \circ X) = A \circ \mathbb{1}$*:*



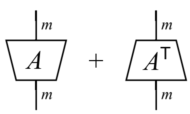
8) $\dfrac{\partial^2}{\partial\,|x\rangle\,\partial\,\langle x|}(\langle x|A|x\rangle + \langle b|x\rangle) = A + A^{\mathsf{T}}$*:*



(16)
(14)
=



(13)
(3)
=



.

This formula shows that the Hessian matrix of the quadratic function $\langle x|A|x\rangle + \langle b|x\rangle + c$ with $A \in \mathbb{R}^{m\times m}$, $|b\rangle \in \mathbb{R}^m$, and $c \in \mathbb{R}$ is $A + A^{\mathsf{T}}$.

9) Other important examples:

We can easily obtain the following formulas (the proofs are left to the readers):

$$\frac{\partial}{\partial X}\operatorname{Tr}(AXB) = A^{\mathsf{T}}B^{\mathsf{T}},$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(X\otimes X) = (2\operatorname{Tr}X)\mathbb{1},$$
$$\frac{\partial}{\partial X}\langle a|X^{\mathsf{T}}CX|b\rangle = CX\,|b\rangle\,\langle a| + C^{\mathsf{T}}X\,|a\rangle\,\langle b|\,,$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(X^k) = k(X^{k-1})^{\mathsf{T}},$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(AX^k) = \sum_{s=0}^{k-1}(X^sAX^{k-1-s})^{\mathsf{T}},$$
$$\frac{\partial}{\partial X}\operatorname{Tr}(AX^{-1}B) = -(X^{-1}BAX^{-1})^{\mathsf{T}},$$

where k is a natural number.

VII. Conclusion

We introduced a diagrammatic notation for matrix differentiation. We demonstrated through some interesting examples that our notation makes it possible to easily and intuitively calculate matrix differentiation.

Acknowledgment

I am grateful to O. Hirota for support.

References

[1] M. S. Kaznady and D. F. James, "Numerical strategies for quantum tomography: Alternatives to full optimization," *Phys. Rev. A*, vol. 79, no. 2, 022109, 2009.

[2] G. C. Knee, E. Bolduc, J. Leach, and E. M. Gauger, "Quantum process tomography via completely positive and trace-preserving projection," *Physical Review A*, vol. 98, no. 6, 062336, 2018.

[3] N. Leung, M. Abdelhafez, J. Koch, and D. Schuster, "Speedup for quantum optimal control from automatic differentiation based on graphics processing units," *Phys. Rev. A*, vol. 95, no. 4, 042318, 2017.

[4] J. C. Cresswell, I. Tzitrin, and A. Z. Goldberg, "Perturbative expansion of entanglement negativity using patterned matrix calculus," *Phys. Rev. A*, vol. 99, no. 1, 012322, 2019.

[5] B. Coecke, "Quantum picturalism," *Contemporary physics*, vol. 51, no. 1, pp. 59–83, 2010.

[6] J.-H. Kim, M. S. H. Oh, and K.-Y. Kim, "Boosting vector calculus with the graphical notation," *arXiv preprint arXiv:1911.00892*, 2019.

[7] A. Toumi, R. Yeung, and G. de Felice, "Diagrammatic differentiation for quantum machine learning," *arXiv preprint arXiv:2103.07960*, 2021.