Quick-but-comprehensive guide to Matrix Differentiation

Rubem Vasconcelos Pacelli
rubem.engenharia@gmail.com

Department of Teleinformatics Engineering, Federal University of Ceará.
Fortaleza, Ceará, Brazil.

September 23, 2022

Contents

1 Introduction

Since my Master's degree, I've been struggling with matrix differentiation as I could not find good references that cover it nicely. The bibliographies I found at that time were books from Economics (see references), and they use a ~~weird~~ distinct notation.

After delving a lot, I finally found a good reference of the Professor Randal (honorable mention for the Matrix Cookbook too). However, to my surprise, when I tried to apply these matrix differentiation propositions, I got "wrong" answers! Only in my Doctorate, I discovered what was going on: *there are two ways to represent a derivative of a vector* [1]. If you do not select the author's representation, you will end up with the same result, but in a row vector[1] instead of a column vector and vice-versa. For the cases where the resulting derivative is a matrix, you will get its transpose. The first representation is called Jacobian formulation or numerator layout, while the second one is called Hessian formulation or denominator layout.

Due to the lack of references and the need to get my own guide, I decided to make this quick guide. I will use the notation that most Engineers might be used to. Moreover, I will only cover the Hessian formulation since this is the one that matches the derivative results I find in my books. If you are looking for the Jacobian formulation, I highly recommend Professor Randal's note, which uses this representation. The unique drawback is that he does not use complex numbers.

Some of the differentiation solutions here were collected from class notes, while others I derived by myself. Obviously, this guide may have errors (I hope not). If you find it, feel free to reach me out through email or simply make a pull request on my Github.

2 Notation and nomeclature

Let

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{12} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{m2} & \dots & a_{mn} \end{bmatrix} \in \mathbb{C}^{m \times n} \tag{1}$$

be a complex matrix with dimension equal to $m \times n$, where $a_{ij} \in \mathbb{C}$ is its element in the position (i, j). Similarly, a complex vector is defined by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{C}^n. \tag{2}$$

Nonbold Romain and Greek alphabets represent scalars, while bold uppercases and bold lowercases represent matrices and vectors, respectively. In Section 3, I will try to use the initial letters of the Romain alphabet $(a, b, c, \dots)$ to represent knows variables, and the final letters of the Romain alphabet $(x, y, z, w, \dots)$ to represent unknown variables. The operators $\cdot^{\mathsf{T}}$, $\cdot^{\mathsf{H}}$, $\cdot^{*}$ $\mathrm{tr}(\cdot)$, $\mathrm{adj}(\cdot)$, and $|\cdot|$ represent, respectively, the transpose, the hermitian, the conjulgate, the trace, the adjoint, and the determinant (or absolute value when the operand is a scalar).

[1] Although the expression "row vector" is quite common, I really advocate to avoid it since, once defined a vector as a column, $\mathbf{y}^{\mathsf{T}} \in \mathbb{C}^{1 \times n}$ is actually a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}$. That is, it has nothing to do with a vector, which is an entity in a n-dimensional space. Therefore, throughout this note, I will refer to is as $1 \times n$ matrix.

2.1 Jacobian formulation (numerator layout)

Consider two vectors $\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{y} \in \mathbb{C}^m$. In the Jacobian formulation (also called numerator layout), the partial derivative of each element in $\mathbf{y}$ by each element in $\mathbf{x}$ is represented as

$$\left[\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right]_{\text{Num}} = \begin{bmatrix} \frac{\partial y_1^{\mathsf{T}}}{\partial \mathbf{x}} \\ \frac{\partial y_2^{\mathsf{T}}}{\partial \mathbf{x}} \\ \vdots \\ \frac{\partial y_m^{\mathsf{T}}}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \in \mathbb{C}^{m \times n}. \tag{3}$$

Note that it perfectly matches the Jacobian matrix definition,

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1^{\mathsf{T}}}{\partial \mathbf{x}} \\ \frac{\partial f_2^{\mathsf{T}}}{\partial \mathbf{x}} \\ \vdots \\ \frac{\partial f_m^{\mathsf{T}}}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}, \tag{4}$$

where $f : \mathbb{R}^n \to \mathbb{R}$. Perhaps that is why it is called the "Jacobian formulation".

2.2 Hessian formulation (denominator layout)

The Hessian formulation (or denominator layout) has the following notation

$$\left[\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right]_{\text{Den}} = \begin{bmatrix} \frac{\partial y_1}{\partial \mathbf{x}} & \frac{\partial y_2}{\partial \mathbf{x}} & \cdots & \frac{\partial y_3}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \in \mathbb{C}^{n \times m}. \tag{5}$$

I've tried to find some analogy with the Hessian matrix but, unfortunately, I haven't discovered it yet.

2.3 Comparative between Jacobian and Hessian formulations

As you could have noticed,

$$\left[\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right]_{\text{Num}} = \left[\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right]_{\text{Den}}^{\mathsf{T}}. \tag{6}$$

That is the difference when you try to differentiate without paying attention to which representation the author adopted. The good news is that, as long as you differentiate it correctly, you can switch between the Jacobian and Hessian formulations by simply transposing the final result. Fortunately, the denominator layout is the most adopted by authors from areas related to Electrical Engineering. That is why we will focus on the denominator layout hereafter.

As a rule of thumb, keep in mind that:

- $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ will yield a matrix.
- $\frac{\partial \mathbf{Y}}{\partial x}$ will yield a matrix.
- $\frac{\partial x}{\partial \mathbf{X}}$ will yield a matrix.
- $\frac{\partial y}{\partial \mathbf{x}}$ will yield a vector.
- $\frac{\partial \mathbf{y}}{\partial x}$ will yield a $1 \times n$ matrix ("row vector").

3 Matrix Differentiation

3.1 $\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}}$

Let $\mathbf{A} \in \mathbb{C}^{m \times n}$ and $\mathbf{x} \in \mathbb{C}^n$, in which $\mathbf{A}$ does not depend on $\mathbf{x}$, we have that:

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \tag{7}$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} \sum_{j=1}^n a_{1j} x_j \\ \sum_{j=1}^n a_{2j} x_j \\ \vdots \\ \sum_{j=1}^n a_{mj} x_j \end{bmatrix}^\mathsf{T} \right) \tag{8}$$

$$= \begin{bmatrix} \frac{\partial}{\partial \mathbf{x}} \left(\sum_{j=1}^n a_{1j} x_j \right) & \frac{\partial}{\partial \mathbf{x}} \left(\sum_{j=1}^n a_{2j} x_j \right) & \dots & \frac{\partial}{\partial \mathbf{x}} \left(\sum_{j=1}^n a_{mj} x_j \right) \end{bmatrix} \tag{9}$$

Since a scalar-vector derivative is represented by a vector, we have that

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1} \left(\sum_{j=1}^n a_{1j} x_j \right) & \frac{\partial}{\partial x_1} \left(\sum_{j=1}^n a_{21j} x_j \right) & \dots & \frac{\partial}{\partial x_1} \left(\sum_{j=1}^n a_{mj} x_j \right) \\ \frac{\partial}{\partial x_2} \left(\sum_{j=1}^n a_{1j} x_j \right) & \frac{\partial}{\partial x_2} \left(\sum_{j=1}^n a_{21j} x_j \right) & \dots & \frac{\partial}{\partial x_2} \left(\sum_{j=1}^n a_{mj} x_j \right) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_n} \left(\sum_{j=1}^n a_{1j} x_j \right) & \frac{\partial}{\partial x_n} \left(\sum_{j=1}^n a_{21j} x_j \right) & \dots & \frac{\partial}{\partial x_n} \left(\sum_{j=1}^n a_{mj} x_j \right) \end{bmatrix} \tag{10}$$

$$= \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{bmatrix} \tag{11}$$

$$\boxed{\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\mathsf{T}} \tag{12}$$

3.2 $\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{z}}$

3.3 $\frac{\partial \mathbf{a}^\mathsf{T}\mathbf{x}}{\partial \mathbf{x}}$

Let $\mathbf{a}, \mathbf{x} \in \mathbb{C}^n$, in which $\mathbf{a}$ does not depend on $\mathbf{x}$. You can derive the derivative for the inner product by considering that $\mathbf{a}^\mathsf{T}$ is actually a $1 \times n$ matrix that transforms $\mathbb{R}^n$ into $\mathbb{R}$, and we already know what is the derivate of a $\mathbf{Ax}$. Thus,

$$\frac{\partial \mathbf{a}^\mathsf{T}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}^{\mathsf{T}^\mathsf{T}} = \mathbf{a}. \tag{13}$$

Even though, if you want the step-by-step, here it is:

$$\frac{\partial \mathbf{a}^\mathsf{T}\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \tag{14}$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^{n} a_i x_i \right) \tag{15}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1} \left(\sum_{i=1}^{n} a_i x_i \right) \\ \frac{\partial}{\partial x_2} \left(\sum_{i=1}^{n} a_i x_i \right) \\ \vdots \\ \frac{\partial}{\partial x_n} \left(\sum_{i=1}^{n} a_i x_i \right) \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \tag{16}$$

$$\boxed{\frac{\partial \mathbf{a}^\mathsf{T}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}} \tag{17}$$

3.4 $\frac{\partial \mathbf{x}^\mathsf{T}\mathbf{a}}{\partial \mathbf{x}}$

This one can be solved quickly by noticing that $\mathbf{x}^\mathsf{T}\mathbf{a} = \mathbf{a}^\mathsf{T}\mathbf{x}$. Hence,

$$\frac{\partial \mathbf{x}^\mathsf{T}\mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^\mathsf{T}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \tag{18}$$

Nevertheless, here is the step-by-step:

$$\frac{\partial \mathbf{x}^\mathsf{T}\mathbf{a}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \right) \tag{19}$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^{n} x_i a_i \right) \tag{20}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1} \left(\sum_{i=1}^{n} x_i a_i \right) \\ \frac{\partial}{\partial x_2} \left(\sum_{i=1}^{n} x_i a_i \right) \\ \vdots \\ \frac{\partial}{\partial x_n} \left(\sum_{i=1}^{n} x_i a_i \right) \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \tag{21}$$

$$\boxed{\frac{\partial \mathbf{x}^\mathsf{T}\mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}} \tag{22}$$

3.5 $\dfrac{\partial \mathbf{a}^{\mathsf{H}}\mathbf{x}}{\partial \mathbf{x}}$

Let $\mathbf{a}, \mathbf{x} \in \mathbb{C}^n$, in which $\mathbf{a}$ does not depend on $\mathbf{x}$. Once again, we could say that

$$\frac{\partial \mathbf{a}^{\mathsf{H}}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}^{\mathsf{H}^{\mathsf{T}}} = \mathbf{a}^* \tag{23}$$

Nevertheless, here is the step-by-step:

$$\frac{\partial \mathbf{a}^{\mathsf{H}}\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} a_1^* & a_2^* & \dots & a_n^* \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) = \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^{n} a_i^* x_i \right) \tag{24}$$

$$\tag{25}$$

Since a scalar-vector derivative is represented by a vector, we have that

$$\frac{\partial \mathbf{a}^{\mathsf{H}}\mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1} \left(\sum_{i=1}^{n} a_i^* x_i \right) \\ \frac{\partial}{\partial x_2} \left(\sum_{i=1}^{n} a_i^* x_i \right) \\ \vdots \\ \frac{\partial}{\partial x_n} \left(\sum_{i=1}^{n} a_i^* x_i \right) \end{bmatrix} = \begin{bmatrix} a_1^* \\ a_2^* \\ \vdots \\ a_n^* \end{bmatrix} \tag{26}$$

$$\boxed{\frac{\partial \mathbf{a}^{\mathsf{H}}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}^*} \tag{27}$$

3.6 $\dfrac{\partial \mathbf{x}^{\mathsf{H}}\mathbf{a}}{\partial \mathbf{x}}$

Notice that $\mathbf{x}^{\mathsf{H}}\mathbf{a} \neq \mathbf{a}^{\mathsf{H}}\mathbf{x}$. Therefore, we have no choice but derive it. Let $\mathbf{a}, \mathbf{x} \in \mathbb{C}^n$, in which $\mathbf{a}$ does not depend on $\mathbf{x}$, we have that

$$\frac{\partial \mathbf{x}^{\mathsf{H}}\mathbf{a}}{\mathbf{x}} = \frac{\partial}{\mathbf{x}} \left(\begin{bmatrix} x_1^* & x_2^* & \dots & x_n^* \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \right) \tag{28}$$

$$= \frac{\partial}{\mathbf{x}} \left(\sum_{i=1}^{n} x_i^* a_i \right) \tag{29}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1} \left(\sum_{i=1}^{n} x_i^* a_i \right) \\ \frac{\partial}{\partial x_2} \left(\sum_{i=1}^{n} x_i^* a_i \right) \\ \vdots \\ \frac{\partial}{\partial x_n} \left(\sum_{i=1}^{n} x_i^* a_i \right) \end{bmatrix}. \tag{30}$$

By recalling that $\dfrac{\partial x^*}{x} = 0$ (reference required), we have that

$$\frac{\partial \mathbf{x}^{\mathsf{H}}\mathbf{a}}{\mathbf{x}} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{31}$$

$$\boxed{\frac{\partial \mathbf{x}^{\mathsf{H}} \mathbf{a}}{\mathbf{x}} = \mathbf{0}} \tag{32}$$

where $\mathbf{0} \in \mathbb{C}^n$ is the zero vector.

3.7 $\dfrac{\partial \mathbf{x}^{\mathsf{T}} \mathbf{y}}{\partial \mathbf{z}}$

3.8 $\dfrac{\partial \mathbf{x}^{\mathsf{T}} \mathbf{x}}{\partial \mathbf{z}}$

3.9 $\dfrac{\partial \mathbf{x}^{\mathsf{T}} \mathbf{A} \mathbf{x}}{\partial \mathbf{x}}$

Let $\mathbf{A} \in \mathbb{C}^{n \times n}$ and $\mathbf{x} \in \mathbb{C}^n$, in which $\mathbf{A}$ does not depend on $\mathbf{x}$. For the quadratic form, it follows that

$$\begin{aligned}
\frac{\partial \mathbf{x}^{\mathsf{T}} \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} &= \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) && (33) \\
&= \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} \sum_{i=1}^{n} x_i a_{i1} & \sum_{i=1}^{n} x_i a_{i2} & \dots & \sum_{i=1}^{n} x_i a_{in} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) && (34) \\
&= \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \right). && (35)
\end{aligned}$$

Note that the element inside the parentheses is a scalar and that a scalar-vector derivative results in a vector, that is,

$$\frac{\partial \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j\right) \\ \frac{\partial}{\partial x_2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j\right) \\ \vdots \\ \frac{\partial}{\partial x_n}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j\right) \end{bmatrix} \tag{36}$$

$$= \begin{bmatrix} 2x_1 a_{11} + \sum_{\substack{j=1\\ j\neq 1}}^{n} a_{1j}x_j + \sum_{\substack{i=1\\ i\neq 1}}^{n} a_{i1}x_i \\ 2x_2 a_{22} + \sum_{\substack{j=1\\ j\neq 2}}^{n} a_{2j}x_j + \sum_{\substack{i=1\\ i\neq 2}}^{n} a_{i2}x_i \\ \vdots \\ 2x_n a_{nn} + \sum_{\substack{j=1\\ j\neq n}}^{n} a_{nj}x_j + \sum_{\substack{i=1\\ i\neq n}}^{n} a_{in}x_i \end{bmatrix} \tag{37}$$

$$= \begin{bmatrix} \sum_{j=1}^{n} a_{1j}x_j + \sum_{i=1}^{n} a_{i1}x_i \\ \sum_{j=1}^{n} a_{2j}x_j + \sum_{i=1}^{n} a_{i2}x_i \\ \vdots \\ \sum_{j=1}^{n} a_{nj}x_j + \sum_{i=1}^{n} a_{in}x_i \end{bmatrix} \tag{38}$$

$$\boxed{\frac{\partial \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\mathsf{T}\mathbf{x} + \mathbf{A}\mathbf{x}} \tag{39}$$

For the special case where $\mathbf{A}$ is symmetric, it we obtain

$$\boxed{\frac{\partial \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}} \tag{40}$$

3.10 $\dfrac{\partial \mathbf{x}^\intercal \mathbf{A} \mathbf{x}}{\partial \mathbf{z}}$

3.11 $\dfrac{\partial \mathbf{y}^\intercal \mathbf{A} \mathbf{x}}{\partial \mathbf{x}}$

3.12 $\dfrac{\partial \mathbf{y}^\intercal \mathbf{A} \mathbf{x}}{\partial \mathbf{y}}$

3.13 $\dfrac{\partial \mathbf{y}^\intercal \mathbf{A} \mathbf{x}}{\partial \mathbf{z}}$

3.14 $\dfrac{\partial \mathrm{tr}(\mathbf{A}\mathbf{X})}{\partial \mathbf{X}}$

Let $\mathbf{A}, \mathbf{X} \in \mathbb{R}^{n \times n}$, where $\mathbf{A}$ does not depend on the elements in $\mathbf{X}$.

$$
\begin{aligned}
\frac{\partial \mathrm{tr}\left(\mathbf{A}\mathbf{X}\right)}{\partial \mathbf{X}} &= \frac{\partial}{\partial \mathbf{X}} \left(\mathrm{tr} \left(\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix} \right) \right) \\
&= \frac{\partial}{\partial \mathbf{X}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) \\
&= \begin{bmatrix} \frac{\partial}{\partial x_{11}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) & \frac{\partial}{\partial x_{12}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) & \dots & \frac{\partial}{\partial x_{1n}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) \\ \frac{\partial}{\partial x_{21}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) & \frac{\partial}{\partial x_{22}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) & \dots & \frac{\partial}{\partial x_{2n}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_{n1}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) & \frac{\partial}{\partial x_{n2}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) & \dots & \frac{\partial}{\partial x_{nn}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) \end{bmatrix} \\
&= \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{bmatrix}
\end{aligned}
$$

$$\boxed{\frac{\partial \mathrm{tr}\left(\mathbf{A}\mathbf{X}\right)}{\partial \mathbf{X}} = \mathbf{A}^\intercal} \tag{41}$$

3.15 $\dfrac{\partial \left|\mathbf{X}\right|}{\partial \mathbf{X}}$

Let $\mathbf{X} \in \mathbb{R}^{n \times n}$. Through Laplace expansion (cofactor expansion), we can rewrite the determinant of $\mathbf{X}$ as the sum of the cofactors of any row or column, multiplied by its generating element, that is

$$\left|\mathbf{X}\right| = \sum_{i=1}^{n} x_{ki} \left|\mathbf{C}_{ki}\right| = \sum_{i=1}^{n} x_{ik} \left|\mathbf{C}_{ik}\right| \quad \forall\, k \in \{1, 2, ..., n\}, \tag{42}$$

where $\mathbf{C}_{ij}$ denotes the cofactor matrix of $\mathbf{X}$ generated from element x_{ij}. It is worth noting that the cofactor of $\mathbf{C}_{ij}$ is independent of the value of any element (i,j) in $\mathbf{X}$. Therefore, it follows that

$$\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = \frac{\partial}{\partial \mathbf{X}} \left(\sum_{i=1}^{n} x_{ki} |\mathbf{C}_{ki}| \right) \quad \forall\ k \in \{1, 2, ..., n\} \tag{43}$$

$$= \frac{\partial}{\partial \mathbf{X}} \left(\begin{bmatrix} \sum_{i=1}^{n} x_{1i} |\mathbf{C}_{1i}| & \sum_{i=1}^{n} x_{1i} |\mathbf{C}_{1i}| & \dots & \sum_{i=1}^{n} x_{1i} |\mathbf{C}_{1i}| \\ \sum_{i=1}^{n} x_{2i} |\mathbf{C}_{2i}| & \sum_{i=1}^{n} x_{2i} |\mathbf{C}_{2i}| & \dots & \sum_{i=1}^{n} x_{2i} |\mathbf{C}_{2i}| \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ni} |\mathbf{C}_{ni}| & \sum_{i=1}^{n} x_{ni} |\mathbf{C}_{ni}| & \dots & \sum_{i=1}^{n} x_{ni} |\mathbf{C}_{ni}| \end{bmatrix} \right) \tag{44}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_{11}} \left(\sum_{i=1}^{n} x_{1i} |\mathbf{C}_{1i}| \right) & \frac{\partial}{\partial x_{12}} \left(\sum_{i=1}^{n} x_{1i} |\mathbf{C}_{1i}| \right) & \dots & \frac{\partial}{\partial x_{13}} \left(\sum_{i=1}^{n} x_{1i} |\mathbf{C}_{1i}| \right) \\ \frac{\partial}{\partial x_{21}} \left(\sum_{i=1}^{n} x_{2i} |\mathbf{C}_{2i}| \right) & \frac{\partial}{\partial x_{22}} \left(\sum_{i=1}^{n} x_{2i} |\mathbf{C}_{2i}| \right) & \dots & \frac{\partial}{\partial x_{33}} \left(\sum_{i=1}^{n} x_{2i} |\mathbf{C}_{2i}| \right) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_{n1}} \left(\sum_{i=1}^{n} x_{ni} |\mathbf{C}_{ni}| \right) & \frac{\partial}{\partial x_{n2}} \left(\sum_{i=1}^{n} x_{ni} |\mathbf{C}_{ni}| \right) & \dots & \frac{\partial}{\partial x_{n3}} \left(\sum_{i=1}^{n} x_{ni} |\mathbf{C}_{ni}| \right) \end{bmatrix} \tag{45}$$

$$= \begin{bmatrix} |\mathbf{C}_{11}| & |\mathbf{C}_{12}| & \dots & |\mathbf{C}_{1n}| \\ |\mathbf{C}_{21}| & |\mathbf{C}_{22}| & \dots & |\mathbf{C}_{2n}| \\ \vdots & \vdots & \ddots & \vdots \\ |\mathbf{C}_{n1}| & |\mathbf{C}_{n2}| & \dots & |\mathbf{C}_{nn}| \end{bmatrix} \tag{46}$$

$$\boxed{\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = \operatorname{adj}(\mathbf{X})} \tag{47}$$

3.16 $\frac{\partial \mathbf{A}^{-1}}{\partial \alpha}$

References

[1] Aarti Singh. *Lecture notes in Introduction to Machine Learning.* 2013-2016. URL: https://www.cs.cmu.edu/~aarti/Class/10315_Spring22/315S22_Rec4.pdf.

[2] Randal J Barnes. "Matrix differentiation". In: *Springs Journal* (2006), pp. 1–9.

[3] Kaare Brandt Petersen, Michael Syskind Pedersen, et al. "The matrix cookbook". In: *Technical University of Denmark* 7.15 (2008), p. 510.

[4] Phoebus J Dhrymes and Phoebus J Dhrymes. *Mathematics for econometrics.* Vol. 984. Springer, 1978.

[5] *Matrix calculus - Wikipedia.* (Accessed on 09/22/2022). URL: https://en.wikipedia.org/wiki/Matrix_calculus#Numerator-layout_notation.