The Guide for Matrix Calculus

Rubem Vasconcelos Pacelli
`rubem.engenharia@gmail.com`

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

`Version:`June 4, 2023

Contents

4 Solution of Matrix Differentiations   **13**

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 [1], but they use an ~~weird~~ unfamiliar notation.

After delving a lot, I finally found a good reference from Professor Randal's class note [2]. However, to my surprise, when I tried to apply those matrix differentiation propositions, I got "wrong" answers! The truth is that *Matrix Calculus notation is severely fragmented and there is no consensus among the researchers over which notation to follow.* Fortunately, there are two major ways to represent a derivative of a vector [3]. If you do not select the author's representation, you will end up with the transposed result. The first representation is called Jacobian formulation or numerator layout, while the second one is called Hessian formulation or denominator layout. Nevertheless, even for the same layout, some conventions need to be stated.

Due to the lack of references and the need to have one, I decided to make this quick guide. The goal here is twofold: make an introduction to both representations and derive the partial derivatives for the most common Matrix Calculus expressions you came across. Most of the problems will be defined in real space, but some will be extended to the complex plane. In the end, I hope to have a consistent reference guide to understand an author's book that uses a different notation. The way I define in Section 2.2 will be the way I adopt in my papers and in this guide, but I will let you know when some point is not consensus among the authors. Furthermore, I will adopt only apply the Hessian (or demodulator) formulation since it matches with the notation commonly used by authors of the field of Signal Processing, Machine Learning, and Optimization Theory.

1.1 Curated reference list

The following list shows some references you can rely on besides this guide when it comes to Matrix Calculus (in decreasing order of importance):

- Searle, Shayle R., and Andre I. Khuri. Matrix algebra useful for statistics [4]: A classical book with the first edition in 1987. It treats Matrix Calculus in chapter 9 of the second edition.

- Hjørungnes, Are, and David Gesbert. Complex-valued matrix differentiation: Techniques and key results [5]. An advanced book that uses Wirtinger calculus to differentiate complex-valued matrices. You might prefer his article [6] which summarizes the key results.

- Matrix Calculus You Need For Deep Learning [7]: Preprint article focused on its application on Machine Learning. It offers the best tradeoff between comprehension and brevity as it is just 33 pages long.

- You can find good references in these book appendices: Dattoro [8], appendix D; Bishop [9], appendix C; Simon Haykin [10], appendix B (Wirtinger calculus).

- Old and New Matrix Algebra Useful for Statistics, Thomas Minka [11]: Another good guide for matrix algebra hosted on the personal site of the author. It has very well-referenced and focuses more on the concepts of Matrix Calculus than solving the derivatives.

- The Matrix Cookbook [12]: Good reference to see Matrix Calculus results rather than understanding how they were solved.

- Lecture notes in Introduction to Machine Learning, from Carnegie Mellon University [3]: A short lecture but with a clear explanation of the difference between numerator and denominator layout.

- Wikipedia [13]: Maybe it is not the most reliable source, but it is very comprehensive.

- Professor Randal's class note [2]: It has the same purpose as this guide, that is, to solve common matrix differentiation, but it is done using the numerator layout. Such a layout is rarely used for references herein cited.

Differentiation solutions that were collected from other sources will be referenced, while solution that I derived by myself will not have any reference. Obviously, this guide may have errors (I hope not). If you find it, feel free to reach out through email or simply make a pull request on my Github.

2 Notation and nomeclatures

Let

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \in \mathbb{C}^{m \times n} \tag{1}$$

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

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

which may also be denoted as an n-tuple, $(x_1, x_2, \dots, x_n)$, when more convenient.

For scalars, italic Roman $(a, b, c, \dots)$ represents constants (known values), while italic Greek $(\alpha, \beta, \gamma, \dots)$ represents variables (unknown values). For vectors, Roman bold lowercase letters will be adopted, where the initial letters $(\mathbf{a}, \mathbf{b}, \mathbf{c}, \dots)$ represent constants (known values) and the final letters $(\mathbf{x}, \mathbf{y}, \mathbf{w}, \dots)$ represent variables (unknown values). The same thing goes for matrices, but it is denoted as uppercase letters instead of lowercase. The letter $\mathbf{z}$ (or $\mathbf{Z}$, for matrices) will only be used when the vector (or matrix) is complex-valued. In any other case, the vector or matrix will be real-valued. Finally, the operators $\cdot^\top$, $\cdot^{\mathsf{H}}$, $\cdot^*$ $\mathrm{tr}(\cdot)$, $\mathrm{adj}(\cdot)$, and $|\cdot|$ denote, respectively, the transpose, the hermitian, the conjugate, the trace, the adjoint, and the determinant (or absolute value when the operand is a scalar).

2.1 Jacobian formulation (numerator layout)

In the Jacobian formulation (also called numerator layout), the derivative matrix is written laying out the numerator in its shape, while the denominator has its shape transposed (you will understand it better as soon as you see the definitions).

2.1.1 Vector-vector, scalar-vector, and vector-scalar derivatives

Consider two vectors $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. The partial derivative of each element in $\mathbf{y}$ by each element in $\mathbf{x}$ is represented as

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

We can infer what is the shape of $\dfrac{\partial y}{\partial \mathbf{x}}$ and $\dfrac{\partial \mathbf{y}}{\partial x}$ by changing the respective vector sizes in (3).

2.1.2 Matrix-scalar derivative (tangent matrix)

The partial derivative $\dfrac{\partial \mathbf{Y}}{\partial x}$ (usually called tangent matrix) is defined for the numerator layout as

$$\left[\frac{\partial \mathbf{Y}}{\partial x}\right]_{\mathrm{Num}} = \begin{bmatrix} \dfrac{\partial y_{11}}{\partial x} & \dfrac{\partial y_{12}}{\partial x} & \cdots & \dfrac{\partial y_{1n}}{\partial x} \\ \dfrac{\partial y_{21}}{\partial x} & \dfrac{\partial y_{22}}{\partial x} & \cdots & \dfrac{\partial y_{2n}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial y_{m1}}{\partial x} & \dfrac{\partial y_{m2}}{\partial x} & \cdots & \dfrac{\partial y_{mn}}{\partial x} \end{bmatrix} \in \mathbb{R}^{m \times n}, \tag{4}$$

where $\mathbf{Y} \in \mathbb{R}^{m \times n}$.

2.1.3 Scalar-matrix derivative (gradient matrix)

The partial derivative of $\dfrac{\partial y}{\partial \mathbf{X}}$ (usually called gradient matrix) is given by

$$\left[\frac{\partial y}{\partial \mathbf{X}}\right]_{\mathrm{Num}} = \begin{bmatrix} \dfrac{\partial y}{\partial x_{11}} & \dfrac{\partial y}{\partial x_{21}} & \cdots & \dfrac{\partial y}{\partial x_{m1}} \\ \dfrac{\partial y}{\partial x_{12}} & \dfrac{\partial y}{\partial x_{22}} & \cdots & \dfrac{\partial y}{\partial x_{m2}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial y}{\partial x_{1n}} & \dfrac{\partial y}{\partial x_{2n}} & \cdots & \dfrac{\partial y}{\partial x_{mn}} \end{bmatrix} \in \mathbb{R}^{n \times m}, \tag{5}$$

where $\mathbf{X} \in \mathbb{R}^{m \times n}$.

2.1.4 Row vector-scalar and scalar-row vector derivatives

From these definitions, we can infer two nonobvious equalities that are rather useful when handling matrix differentiations. If we consider a special case of the gradient matrix (5), when $m = 1 \therefore \mathbf{X} = \mathbf{x}^\top \in \mathbb{R}^{1 \times n}$, where $\mathbf{x} \in \mathbb{R}^n$, we have that

$$\left[\frac{\partial \alpha}{\partial \mathbf{x}^\top}\right]_{\mathrm{Num}} = \begin{bmatrix} \dfrac{\partial \alpha}{\partial x_1} \\ \dfrac{\partial \alpha}{\partial x_2} \\ \vdots \\ \dfrac{\partial \alpha}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n. \tag{6}$$

However, by comparing with (3), one can state that

$$\left[\frac{\partial \alpha}{\partial \mathbf{x}}\right]_{\mathrm{Num}} = \begin{bmatrix} \dfrac{\partial \alpha}{\partial x_1} & \dfrac{\partial \alpha}{\partial x_2} & \cdots & \dfrac{\partial \alpha}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{1 \times n}, \tag{7}$$

Therefore,

$$\left[\frac{\partial \alpha}{\partial \mathbf{x}^\top}\right]_{\mathrm{Num}} = \left[\frac{\partial \alpha}{\partial \mathbf{x}}\right]_{\mathrm{Num}}^\top \tag{8}$$

Similarly, from (4), when $m = 1 \therefore \mathbf{Y} = \mathbf{y}^\top \in \mathbb{R}^{1 \times n}$, where $\mathbf{y} \in \mathbb{R}^n$, we have that

$$\left[\frac{\partial \mathbf{y}^\top}{\partial \alpha}\right]_{\text{Num}} = \begin{bmatrix} \dfrac{\partial y_1}{\partial \alpha} & \dfrac{\partial y_2}{\partial \alpha} & \cdots & \dfrac{\partial y_n}{\partial \alpha} \end{bmatrix} \in \mathbb{R}^{1 \times n}. \tag{9}$$

However, from (3), we also have that

$$\left[\frac{\partial \mathbf{y}}{\partial \alpha}\right]_{\text{Num}} = \begin{bmatrix} \dfrac{\partial y_1}{\partial \alpha} \\[2mm] \dfrac{\partial y_2}{\partial \alpha} \\[2mm] \cdots \\[2mm] \dfrac{\partial y_n}{\partial \alpha} \end{bmatrix} \in \mathbb{R}^{n}. \tag{10}$$

Therefore,

$$\left[\frac{\partial \mathbf{y}^\top}{\partial \alpha}\right]_{\text{Num}} = \left[\frac{\partial \mathbf{y}}{\partial \alpha}\right]_{\text{Num}}^\top. \tag{11}$$

2.1.5 Jacobian matrix for the numerator layout

In the numerator layout, the notation for the Jacobian matrix is given by

$$\mathbf{J} = \left[\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}\right]_{\text{Num}}. \tag{12}$$

where $f : \mathbb{R}^n \to \mathbb{R}^m$ is a vector function. By calling (3), it is clear that the Jacobian takes the form

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial f_1}{\partial \mathbf{x}} \\[2mm] \dfrac{\partial f_2}{\partial \mathbf{x}} \\[2mm] \vdots \\[2mm] \dfrac{\partial f_m}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\[2mm] \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\[2mm] \vdots & \vdots & \ddots & \vdots \\[2mm] \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m \times n}, \tag{13}$$

where $f = (f_1, f_2, \ldots, f_m)$, being $f_i : \mathbb{R}^n \to \mathbb{R}$, for $1 \le i \le m$. The numerator layout is also called the Jacobian formulation due to the fact that it is represented without the need for the transpose operator.

2.1.6 Hessian matrix for the numerator layout

The Matrix Calculus notation for the Hessian matrix in the numerator layout is given by

$$\mathbf{H} = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^\top} = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\[2mm] \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\[2mm] \vdots & \vdots & \ddots & \vdots \\[2mm] \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \dfrac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \in \mathbb{R}^{n \times n}, \tag{14}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$. We can rewrite it by recalling (8), that is,

$$\mathbf{H} = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^\top} = \frac{\partial}{\partial \mathbf{x}} \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^\top} \right) = \frac{\partial}{\partial \mathbf{x}} \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right)^\top . \tag{15}$$

Note that $\left(\dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right)^\top = \left(\dfrac{\partial f(\mathbf{x})}{\partial x_1}, \dfrac{\partial f(\mathbf{x})}{\partial x_2}, \ldots, \dfrac{\partial f(\mathbf{x})}{\partial x_n} \right)$ is a vector. Therefore, we have a vector-vector derivate and, by applying the differentiation as shown in (3), we get $\mathbf{H}$.

2.2   Hessian formulation (denominator layout)

In the Hessian formulation (also called denominator layout), the derivative matrix is written laying out the denominator in its shape, while the numerator has its shape transposed.

2.2.1   Vector-vector, scalar-vector, and vector-scalar derivatives

Consider two vectors $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. The partial derivative of each element in $\mathbf{y}$ by each element in $\mathbf{x}$ is represented as

$$\left[\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right]_{\text{Den}} = \begin{bmatrix} \dfrac{\partial y_1}{\partial \mathbf{x}} & \dfrac{\partial y_2}{\partial \mathbf{x}} & \cdots & \dfrac{\partial y_3}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial y_1}{\partial x_1} & \dfrac{\partial y_2}{\partial x_1} & \cdots & \dfrac{\partial y_m}{\partial x_1} \\ \dfrac{\partial y_1}{\partial x_2} & \dfrac{\partial y_2}{\partial x_2} & \cdots & \dfrac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial y_1}{\partial x_n} & \dfrac{\partial y_2}{\partial x_n} & \cdots & \dfrac{\partial y_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{n \times m}. \tag{16}$$

We can infer what is the shape of $\dfrac{\partial y}{\partial \mathbf{x}}$ and $\dfrac{\partial \mathbf{y}}{\partial x}$ by changing the respective vector sizes in (16).

2.2.2   Matrix-scalar derivative (tangent matrix)

The partial derivative $\dfrac{\partial \mathbf{Y}}{\partial x}$ (usually called tangent matrix) is defined for the denominator layout as

$$\left[\frac{\partial \mathbf{Y}}{\partial x} \right]_{\text{Den}} = \begin{bmatrix} \dfrac{\partial y_{11}}{\partial x} & \dfrac{\partial y_{21}}{\partial x} & \cdots & \dfrac{\partial y_{m1}}{\partial x} \\ \dfrac{\partial y_{12}}{\partial x} & \dfrac{\partial y_{22}}{\partial x} & \cdots & \dfrac{\partial y_{m2}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial y_{1n}}{\partial x} & \dfrac{\partial y_{2n}}{\partial x} & \cdots & \dfrac{\partial y_{mn}}{\partial x} \end{bmatrix} \in \mathbb{R}^{n \times m}, \tag{17}$$

where $\mathbf{Y} \in \mathbb{R}^{m \times n}$.

2.2.3   Scalar-matrix derivative (gradient matrix)

The partial derivative of $\dfrac{\partial y}{\partial \mathbf{X}}$ (usually called gradient matrix) is given by

$$\left[\frac{\partial y}{\partial \mathbf{X}} \right]_{\text{Den}} = \begin{bmatrix} \dfrac{\partial y}{\partial x_{11}} & \dfrac{\partial y}{\partial x_{12}} & \cdots & \dfrac{\partial y}{\partial x_{1n}} \\ \dfrac{\partial y}{\partial x_{21}} & \dfrac{\partial y}{\partial x_{22}} & \cdots & \dfrac{\partial y}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial y}{\partial x_{m1}} & \dfrac{\partial y}{\partial x_{m2}} & \cdots & \dfrac{\partial y}{\partial x_{mn}} \end{bmatrix} \in \mathbb{R}^{m \times n}, \tag{18}$$

where $\mathbf{X} \in \mathbb{R}^{m \times n}$.

2.2.4 Row vector-scalar and scalar-row vector derivatives

From these definitions, we can infer two nonobvious equalities that are rather useful when handling matrix differentiations. If we consider a special case of the gradient matrix, (18), when $m = 1 \therefore \mathbf{X} = \mathbf{x}^\top \in \mathbb{R}^{1 \times n}$, where $\mathbf{x} \in \mathbb{R}^n$, we have that

$$\left[\frac{\partial \alpha}{\partial \mathbf{x}^\top} \right]_{\text{Den}} = \begin{bmatrix} \dfrac{\partial \alpha}{\partial x_1} & \dfrac{\partial \alpha}{\partial x_2} & \cdots & \dfrac{\partial \alpha}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{1 \times n}, \tag{19}$$

however, by using we definition from (16), it is also true to state that

$$\left[\frac{\partial \alpha}{\partial \mathbf{x}} \right]_{\text{Den}} = \begin{bmatrix} \dfrac{\partial \alpha}{\partial x_1} \\ \dfrac{\partial \alpha}{\partial x_2} \\ \vdots \\ \dfrac{\partial \alpha}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n. \tag{20}$$

Therefore,

$$\left[\frac{\partial \alpha}{\partial \mathbf{x}^\top} \right]_{\text{Den}} = \left[\frac{\partial \alpha}{\partial \mathbf{x}} \right]_{\text{Den}}^\top \tag{21}$$

Similarly, from (17), when $m = 1 \therefore \mathbf{Y} = \mathbf{y}^\top \in \mathbb{R}^{1 \times n}$, where $\mathbf{y} \in \mathbb{R}^n$, we have that

$$\left[\frac{\partial \mathbf{y}^\top}{\partial \alpha} \right]_{\text{Den}} = \begin{bmatrix} \dfrac{\partial y_1}{\partial \alpha} \\ \dfrac{\partial y_2}{\partial \alpha} \\ \cdots \\ \dfrac{\partial y_n}{\partial \alpha} \end{bmatrix} \in \mathbb{R}^n. \tag{22}$$

However, from (16), we also have that

$$\left[\frac{\partial \mathbf{y}}{\partial \alpha} \right]_{\text{Den}} = \begin{bmatrix} \dfrac{\partial y_1}{\partial \alpha} & \dfrac{\partial y_2}{\partial \alpha} & \cdots & \dfrac{\partial y_n}{\partial \alpha} \end{bmatrix} \in \mathbb{R}^{1 \times n}. \tag{23}$$

Therefore,

$$\left[\frac{\partial \mathbf{y}^\top}{\partial \alpha} \right]_{\text{Den}} = \left[\frac{\partial \mathbf{y}}{\partial \alpha} \right]_{\text{Den}}^\top. \tag{24}$$

2.2.5 The Jacobian matrix for the denominator layout

The Jacobian matrix in denominator notation is given by

$$\mathbf{J} = \left[\frac{\partial f(\mathbf{x})^\top}{\partial \mathbf{x}}\right]_{\text{Den}} \tag{25}$$

$$= \begin{bmatrix} \dfrac{\partial f_1(\mathbf{x})^\top}{\partial \mathbf{x}} \\[2mm] \dfrac{\partial f_2(\mathbf{x})^\top}{\partial \mathbf{x}} \\[2mm] \vdots \\[2mm] \dfrac{\partial f_m(\mathbf{x})^\top}{\partial \mathbf{x}} \end{bmatrix} \tag{26}$$

$$= \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\[2mm] \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\[2mm] \vdots & \vdots & \ddots & \vdots \\[2mm] \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m \times n}, \tag{27}$$

where $f : \mathbb{R}^n \to \mathbb{R}^m$ and $f = (f_1, f_2, \ldots, f_m)$, being $f_i : \mathbb{R} \to \mathbb{R}$ for $1 \leq i \leq m$.

2.2.6 The Hessian matrix for the denominator layout

The Hessian matrix in the denominator layout is given by

$$\mathbf{H} = \left[\frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}^2}\right]_{\text{Den}} = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\[2mm] \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\[2mm] \vdots & \vdots & \ddots & \vdots \\[2mm] \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \dfrac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \in \mathbb{R}^{n \times n}, \tag{28}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$. The element $\partial \mathbf{x}^2$ seems to be merely a convention which means that $\partial \mathbf{x}$ is repeated twice, in the very same way it is done with scalar differentiations. It can be scaled to n consecutive derivatives without making the notation cumbersome (points to the denominator layout team).

2.3 Comparative between Jacobian and Hessian formulations

As you could have noticed,

$$\left[\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right]_{\text{Num}} = \left[\frac{\partial \mathbf{y}}{\partial \mathbf{x}}\right]_{\text{Den}}^{\top}, \tag{29}$$

$$\left[\frac{\partial \mathbf{Y}}{\partial x}\right]_{\text{Num}} = \left[\frac{\partial \mathbf{Y}}{\partial x}\right]_{\text{Den}}^{\top}, \tag{30}$$

$$\left[\frac{\partial y}{\partial \mathbf{X}}\right]_{\text{Num}} = \left[\frac{\partial y}{\partial \mathbf{X}}\right]_{\text{Den}}^{\top}, \tag{31}$$

$$\left[\frac{\partial \alpha}{\partial \mathbf{x}\partial \mathbf{x}^{\top}}\right]_{\text{Num}} = \left[\frac{\partial \alpha}{\partial \mathbf{x}^2}\right]_{\text{Den}}, \tag{32}$$

where that last equation can be inferred from the discussion about the Hessian matrix.

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 Signal Processing and Machine Learning. That is why we will focus on the denominator layout hereafter (the notation $[\cdot]_{\text{Den}}$ will be dropped out since we do not need it anymore).

As a rule of thumb, keep in mind that:

- $\dfrac{\partial \mathbf{y}}{\partial \mathbf{x}}$ will yield a matrix.

- $\dfrac{\partial \mathbf{Y}}{\partial x}$ will yield a matrix.

- $\dfrac{\partial x}{\partial \mathbf{X}}$ will yield a matrix.

- $\dfrac{\partial y}{\partial \mathbf{x}}$ will yield a vector.

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

- $\mathbf{J} = \dfrac{\partial f(\mathbf{x})}{\partial \mathbf{x}}^{\top}$.

- $\mathbf{H} = \dfrac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}^2}$.

- $\dfrac{\partial \alpha}{\partial \mathbf{x}^{\top}} = \dfrac{\partial \alpha}{\partial \mathbf{x}}^{\top}$ for $\alpha \in \mathbb{R}$.

- $\dfrac{\partial \mathbf{y}^{\top}}{\partial \alpha} = \dfrac{\partial \mathbf{y}}{\partial \alpha}^{\top}$ for $\alpha \in \mathbb{R}$.

2.4 Notations not widely agreed upon

Expressions such as $\dfrac{\partial \mathbf{y}}{\partial \mathbf{X}}, \dfrac{\partial \mathbf{Y}}{\partial \mathbf{X}}$, or $\dfrac{\partial \mathbf{Y}}{\partial \mathbf{x}}$ have no agreement for both Jacobian and Hessian notations. It is possible, however, to define the matrix-matrix derivative for both representations. The problem is that some authors define it in the most intuitive manner: For $\dfrac{\partial \mathbf{Y}}{\partial \mathbf{X}}$, element in the position (i, j) is $\partial y_{ij}/\partial x_{ij}$. However, as we saw, it is inconsistent for both formulations as the matrix for the Jacobian (Hessian) formulation must lay out its denominator (numerator) in its transposed shape. Therefore,

for a consistent numerator layout, we would have

$$\left[\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}\right]_{\text{Num}} = \begin{bmatrix} \dfrac{\partial y_{1,1}}{\partial x_{1,1}} & \dfrac{\partial y_{1,2}}{\partial x_{2,1}} & \cdots & \dfrac{\partial y_{1,m}}{\partial x_{m,1}} \\ \dfrac{\partial y_{2,1}}{\partial x_{1,2}} & \dfrac{\partial y_{2,2}}{\partial x_{2,2}} & \cdots & \dfrac{\partial y_{2,n}}{\partial x_{m,2}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial y_{n,1}}{\partial x_{1,n}} & \dfrac{\partial y_{n,2}}{\partial x_{2,n}} & \cdots & \dfrac{\partial y_{n,m}}{\partial x_{m,n}} \end{bmatrix} \in \mathbb{R}^{n \times m}, \tag{33}$$

and for a consistent denominator layout, we would have

$$\left[\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}\right]_{\text{Den}} = \begin{bmatrix} \dfrac{\partial y_{1,1}}{\partial x_{1,1}} & \dfrac{\partial y_{2,1}}{\partial x_{1,2}} & \cdots & \dfrac{\partial y_{n,1}}{\partial x_{1,n}} \\ \dfrac{\partial y_{1,2}}{\partial x_{2,1}} & \dfrac{\partial y_{2,2}}{\partial x_{2,2}} & \cdots & \dfrac{\partial y_{n,2}}{\partial x_{2,n}} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial y_{1,m}}{\partial x_{m,1}} & \dfrac{\partial y_{2,m}}{\partial x_{m,2}} & \cdots & \dfrac{\partial y_{n,m}}{\partial x_{m,n}} \end{bmatrix} \in \mathbb{R}^{m \times n}, \tag{34}$$

where $\mathbf{X} \in \mathbb{R}^{m \times n}$ and $\mathbf{Y} \in \mathbb{R}^{n \times m}$. For expressions $\dfrac{\partial \mathbf{y}}{\partial \mathbf{X}}$ and $\dfrac{\partial \mathbf{Y}}{\partial \mathbf{x}}$ can the derived by taking them as special cases of the matrix-matrix formulation. Notwithstanding, keep in mind that both equations are not standard and their usage must be acknowledged by the reader.

2.5   On the ambiguity of the notation ∇^2

It is common for some authors, such as Simon Haykin [10], to denote the gradient vector and the Hessian matrix as

$$\mathbf{g} = \nabla f = \frac{\partial f}{\partial \mathbf{x}} \tag{35}$$

and

$$\mathbf{H} = \nabla^2 f = \frac{\partial^2 f}{\partial \mathbf{x}^2}, \tag{36}$$

respectively. It is important to point out, however, that the notation ∇^2 might be ambiguous since it is also used in Vector Calculus to denote the Laplacian operator.

Vector Calculus is another branch of Mathematics that deals with differentiations and integrals of vector fields in three-dimensional space. It is used in different areas of Physics and Engineering, such as electromagnetic fields, fluid flow, etc. In Vector Calculus, the gradient vector is commonly defined as

$$\nabla = \begin{bmatrix} \dfrac{\partial}{\partial x_1} \\ \dfrac{\partial}{\partial x_2} \\ \vdots \\ \dfrac{\partial}{\partial x_n} \end{bmatrix} \tag{37}$$

And the gradient vector is denoted in the very same way we denote in Matrix Calculus, i.e., ∇f, where f is a potential (or scalar-valued) function. On the other hand, the symbol $\nabla^2 f$ is often used to denote the Laplacian operator, which is defined as

$$\nabla^2 f = \nabla \cdot \nabla f = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2} + \dots \frac{\partial f}{\partial x_n}, \tag{38}$$

where $\cdot$ is the dot or inner product. Since Vector Calculus is far enough from Matrix Calculus in many applications, the notation $\nabla^2 f$ is used in both fields without problems. However, when it is not the case, the notation $\nabla^2 f$ must be avoided to denote the Hessian matrix.

3 Identities

We need to be cautious when applying the matrix differentiation identities since the element orders matter. For instance, for scalar elements, the product rule may be written as either $(fg)' = f'g + g'f$ or $(fg)' = gf' + fg'$. In Matrix Calculus, we do not have such a privilege.

3.1 Chain rule

3.1.1 Univariate functions

For scalar elements, the chain rule is given by

$$\frac{\partial w}{\partial v} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial y}\frac{\partial y}{\partial v}. \tag{39}$$

Similarly, in matrix notation, we have

$$\frac{\partial \mathbf{w}}{\partial \mathbf{v}} = \frac{\partial \mathbf{y}}{\partial \mathbf{v}}\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\frac{\partial \mathbf{w}}{\partial \mathbf{x}}, \tag{40}$$

where $\mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^m, \mathbf{v} \in \mathbb{R}^p$, and $\mathbf{w} \in \mathbb{R}^q$. In this expression, $\mathbf{w}$ depends on $\mathbf{x}$, $\mathbf{x}$ depends on $\mathbf{y}$, and $\mathbf{y}$ depends on $\mathbf{z}$. The number of elements in the chain rule can be increased indiscriminately. The main point here is that *the chain rule in Matrix Calculus notation must be placed backward when compared with the standard chain rule of scalar elements.*

3.1.2 Multivariate functions

In the previous section, we had a case where $\mathbf{w}$ depends on $\mathbf{x}$, which depends on $\mathbf{y}$, which depends on $\mathbf{v}$. If $\mathbf{w} = f(\mathbf{x}), \mathbf{x} = g(\mathbf{y})$, and $\mathbf{y} = h(\mathbf{v})$, then f, g, and h are functions of one variable, also called univariate functions. However, we might find a situation where $\mathbf{w} = f(\mathbf{x}, \mathbf{y})$ is a function of two (or more) variables.

For scalar elements, we can find partial derivatives of multivariate functions by considering that $w = f(x, y)$ is differentiable on x and y. The chain rule becomes

$$\frac{\partial w}{\partial z} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial z} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial z}. \tag{41}$$

Similarly, for Matrix Calculus notation, we have

$$\frac{\partial \mathbf{w}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}}\frac{\partial \mathbf{w}}{\partial \mathbf{x}} + \frac{\partial \mathbf{y}}{\partial \mathbf{z}}\frac{\partial \mathbf{w}}{\partial \mathbf{y}} \tag{42}$$

Note the backward placement of each summation term. This expression can be used for an unrestricted number of variables.

3.2 Sum (or minus) rule

3.2.1 Vector-vector derivative

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ and $a, b \in \mathbb{R}$, where $\mathbf{x}$ and $\mathbf{y}$ depend on $\mathbf{w} \in \mathbb{R}^n$, but a and b do not. Therefore,

$$\frac{\partial(a\mathbf{x} \pm b\mathbf{y})}{\partial \mathbf{w}} = a\frac{\partial \mathbf{x}}{\partial \mathbf{w}} \pm b\frac{\partial \mathbf{y}}{\partial \mathbf{w}} \tag{43}$$

3.2.2 Matrix-scalar derivative

Another is when you have

$$\frac{\partial(a\mathbf{X} \pm b\mathbf{Y})}{\partial \alpha}, \tag{44}$$

where $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m \times n}$ depend on $\alpha \in \mathbb{R}$. The solution is

$$\frac{\partial(a\mathbf{X} \pm b\mathbf{Y})}{\partial \alpha} = a\frac{\partial \mathbf{X}}{\partial \alpha} \pm b\frac{\partial \mathbf{Y}}{\partial \alpha}. \tag{45}$$

3.2.3 Scalar-matrix derivative

The scalar-matrix derivative has a similar result, i.e.,

$$\frac{\partial \left(ax \pm by\right)}{\partial \mathbf{W}} = a\frac{\partial x}{\partial \mathbf{W}} \pm b\frac{\partial y}{\partial \mathbf{W}}, \tag{46}$$

where $x, y \in \mathbb{R}$ depend on $\mathbf{W} \in \mathbb{R}^{m \times n}$, but $a, b \in \mathbb{R}$ do not.

3.3 Product rule

3.3.1 Vector-vector derivative

Let $w \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^m$, where both depend on $\mathbf{x} \in \mathbb{R}^n$. Then,

$$\frac{\partial w\mathbf{v}}{\partial \mathbf{x}} = w\frac{\partial \mathbf{v}}{\partial \mathbf{x}} + \frac{\partial w}{\partial \mathbf{x}}\mathbf{v}^\top. \tag{47}$$

Note that is not possible to apply the product rule when you have $\mathbf{Wv}$, where $\mathbf{W} \in \mathbb{R}^{n \times m}$ also depends on $\mathbf{x}$. If you tried, you would get $\partial \mathbf{W}/\partial \mathbf{x}$, which there is no consensus about this layout (vide Section 2.4).

3.3.2 Scalar-vector derivative

Another possibility of applying the product rule is when you have $\mathbf{w}^\top \mathbf{v}$, where $\mathbf{w} \in \mathbb{R}^m$ also depends on $\mathbf{x} \in \mathbb{R}^n$. In this case, the dot product is given by

$$\frac{\partial \mathbf{w}^\top \mathbf{v}}{\partial \mathbf{x}} = \frac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{w} + \frac{\partial \mathbf{w}}{\partial \mathbf{x}}\mathbf{v}. \tag{48}$$

3.3.3 Scalar-matrix derivative

It is still possible to apply the product rule to

$$\frac{\partial wv}{\partial \mathbf{X}}, \tag{49}$$

where $w, v \in \mathbb{R}$ depend on $\mathbf{X} \in \mathbb{R}^{m \times n}$. In this case, we have

$$\frac{\partial wv}{\partial \mathbf{X}} = w\frac{\partial v}{\partial \mathbf{X}} + v\frac{\partial w}{\partial \mathbf{X}}. \tag{50}$$

3.3.4 Matrix-scalar derivative

The last case is when you have

$$\frac{\partial \mathbf{WV}}{\partial \alpha}, \tag{51}$$

where both $\mathbf{W} \in \mathbb{R}^{m \times p}$ and $\mathbf{V} \in \mathbb{R}^{p \times n}$ depend on $\alpha \in \mathbb{R}$. In this case, we have

$$\frac{\partial \mathbf{WV}}{\partial \alpha} = \frac{\partial \mathbf{V}}{\partial \alpha}\mathbf{W}^\top + \mathbf{V}^T\frac{\mathbf{W}}{\partial \alpha} \tag{52}$$

4 Solution of Matrix Differentiations

We usually have two ways to solve matrix differentiation:

1. Performing element-by-element operations in matrices and vectors;

2. Preserving the Matrix Calculus notation, performing operations on the whole matrix/vector and, eventually, using some identities.

The latter is usually more straightforward and less toilsome than the former and is therefore preferable.

The solutions in this Section will usually show the element-by-element solution and the solution by preserving the Matrix Calculus notation. For the element-by-element solutions, you only need to know that a scalar-vector derivate results in a vector for the denominator layout. All other shapes will naturally arise. For solutions with Matrix Calculus notation, you need to be acquainted with some of the identities shown in Section 3.

4.1 $\quad \dfrac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top$

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

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \tag{53}$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ \sum_{j=1}^{n} a_{2j}x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj}x_j \end{bmatrix}^\top \right) \tag{54}$$

$$= \begin{bmatrix} \dfrac{\partial}{\partial \mathbf{x}} \left(\sum_{j=1}^{n} a_{1j}x_j \right) & \dfrac{\partial}{\partial \mathbf{x}} \left(\sum_{j=1}^{n} a_{2j}x_j \right) & \dots & \dfrac{\partial}{\partial \mathbf{x}} \left(\sum_{j=1}^{n} a_{mj}x_j \right) \end{bmatrix} \tag{55}$$

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

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \dfrac{\partial}{\partial x_1} \left(\sum_{j=1}^{n} a_{1j}x_j \right) & \dfrac{\partial}{\partial x_1} \left(\sum_{j=1}^{n} a_{21j}x_j \right) & \dots & \dfrac{\partial}{\partial x_1} \left(\sum_{j=1}^{n} a_{mj}x_j \right) \\ \dfrac{\partial}{\partial x_2} \left(\sum_{j=1}^{n} a_{1j}x_j \right) & \dfrac{\partial}{\partial x_2} \left(\sum_{j=1}^{n} a_{21j}x_j \right) & \dots & \dfrac{\partial}{\partial x_2} \left(\sum_{j=1}^{n} a_{mj}x_j \right) \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial}{\partial x_n} \left(\sum_{j=1}^{n} a_{1j}x_j \right) & \dfrac{\partial}{\partial x_n} \left(\sum_{j=1}^{n} a_{21j}x_j \right) & \dots & \dfrac{\partial}{\partial x_n} \left(\sum_{j=1}^{n} a_{mj}x_j \right) \end{bmatrix} \tag{56}$$

$$= \begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{bmatrix} \tag{57}$$

$$\boxed{\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top \in \mathbb{R}^{n \times m}} \tag{58}$$

4.2 $\quad \dfrac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{v}} = \dfrac{\partial \mathbf{x}}{\partial \mathbf{v}} \mathbf{A}^\top$

Let $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{v} \in \mathbb{R}^p$ and $\mathbf{A} \in \mathbb{R}^{m\times n}$, where $\mathbf{x}$ depends on $\mathbf{v}$, but $\mathbf{A}$ does not. Then

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{v}} = \begin{bmatrix} \dfrac{\partial}{\partial \mathbf{v}} \sum_{i=1}^{n} a_{1i}x_i & \dfrac{\partial}{\partial \mathbf{v}} \sum_{i=1}^{n} a_{2i}x_i & \dots & \dfrac{\partial}{\partial \mathbf{v}} \sum_{i=1}^{n} a_{mi}x_i \end{bmatrix} \tag{59}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} a_{1i}\dfrac{\partial x_i}{\partial \mathbf{v}} & \sum_{i=1}^{n} a_{2i}\dfrac{\partial x_i}{\partial \mathbf{v}} & \dots & \sum_{i=1}^{n} a_{mi}\dfrac{\partial x_i}{\partial \mathbf{v}} \end{bmatrix} \tag{60}$$

$$= \underbrace{\begin{bmatrix} \dfrac{\partial x_1}{\partial \mathbf{v}} & \dfrac{\partial x_2}{\partial \mathbf{v}} & \dots & \dfrac{\partial x_n}{\partial \mathbf{v}} \end{bmatrix}}_{p \times n} \underbrace{\begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix}}_{n \times m} \tag{61}$$

$$\boxed{\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{v}} = \frac{\partial \mathbf{x}}{\partial \mathbf{v}} \mathbf{A}^\top \in \mathbb{R}^{p \times m}} \tag{62}$$

Observe that this result is equivalent to applying the chain rule (c.f. (40)), that is,

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{v}} = \frac{\partial \mathbf{x}}{\partial \mathbf{v}} \frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}}{\partial \mathbf{v}} \mathbf{A}^\top. \tag{63}$$

4.3 $\dfrac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$

Let $\mathbf{a}, \mathbf{x} \in \mathbb{R}^n$, in which $\mathbf{a}$ does not depend on $\mathbf{x}$. You can derive the derivative for the inner product by considering that $\mathbf{a}^\top$ 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,

$$\frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}^{\top\top} = \mathbf{a}. \tag{64}$$

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

$$\frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \tag{65}$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^{n} a_i x_i \right) \tag{66}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1} \left(\sum_{i=1}^{n} a_i x_i \right) \\ \frac{\partial}{\partial x_2} \left(\sum_{i=1}^{n} a_i x_i \right) \\ \vdots \\ \frac{\partial}{\partial x_n} \left(\sum_{i=1}^{n} a_i x_i \right) \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \tag{67}$$

$$\boxed{\frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \in \mathbb{R}^n} \tag{68}$$

4.4 $\dfrac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}$

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

$$\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \tag{69}$$

Nevertheless, here is the step-by-step:

$$\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \right) \tag{70}$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^{n} x_i a_i \right) \tag{71}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1} \left(\sum_{i=1}^{n} x_i a_i \right) \\ \frac{\partial}{\partial x_2} \left(\sum_{i=1}^{n} x_i a_i \right) \\ \vdots \\ \frac{\partial}{\partial x_n} \left(\sum_{i=1}^{n} x_i a_i \right) \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \tag{72}$$

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a} \in \mathbb{R}^n} \tag{73}$$

4.5 $\dfrac{\partial \mathbf{a}^{\mathsf{H}}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}^{*}$

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

$$\frac{\partial \mathbf{a}^{\mathsf{H}}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}^{\mathsf{H}\top} = \mathbf{a}^{*} \tag{74}$$

Nevertheless, here is the step-by-step:

$$\frac{\partial \mathbf{a}^{\mathsf{H}}\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} a_1^* & a_2^* & \cdots & a_n^* \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) = \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^{n} a_i^* x_i \right) \tag{75}$$

$$\tag{76}$$

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

$$\frac{\partial \mathbf{a}^{\mathsf{H}}\mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \dfrac{\partial}{\partial x_1} \left(\sum_{i=1}^{n} a_i^* x_i \right) \\ \dfrac{\partial}{\partial x_2} \left(\sum_{i=1}^{n} a_i^* x_i \right) \\ \vdots \\ \dfrac{\partial}{\partial x_n} \left(\sum_{i=1}^{n} a_i^* x_i \right) \end{bmatrix} = \begin{bmatrix} a_1^* \\ a_2^* \\ \vdots \\ a_n^* \end{bmatrix} \tag{77}$$

$$\boxed{\frac{\partial \mathbf{a}^{\mathsf{H}}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}^{*} \in \mathbb{C}^n} \tag{78}$$

4.6 $\quad \dfrac{\partial \mathbf{z}^{\mathsf{H}}\mathbf{a}}{\partial \mathbf{z}} = \mathbf{0}$

Let us define $\mathbf{z}, \mathbf{a} \in \mathbb{C}^n$, where $\mathbf{a}$ does not depend on $\mathbf{z}$. Since $\mathbf{z}^{\mathsf{H}}\mathbf{a} \neq \mathbf{a}^{\mathsf{H}}\mathbf{z}$, we have no choice but derive it:

$$\frac{\partial \mathbf{z}^{\mathsf{H}}\mathbf{a}}{\partial \mathbf{z}} = \frac{\partial}{\partial \mathbf{z}} \left(\begin{bmatrix} z_1^* & z_2^* & \cdots & z_n^* \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \right) \tag{79}$$

$$= \frac{\partial}{\partial \mathbf{z}} \left(\sum_{i=1}^{n} z_i^* a_i \right) \tag{80}$$

$$= \begin{bmatrix} \frac{\partial}{\partial z_1} \left(\sum_{i=1}^{n} z_i^* a_i \right) \\ \frac{\partial}{\partial z_2} \left(\sum_{i=1}^{n} z_i^* a_i \right) \\ \vdots \\ \frac{\partial}{\partial z_n} \left(\sum_{i=1}^{n} z_i^* a_i \right) \end{bmatrix} \tag{81}$$

$$= \begin{bmatrix} \frac{\partial z_1^*}{\partial z_1} \\ \frac{\partial z_2^*}{\partial z_2} \\ \vdots \\ \frac{\partial z_n^*}{\partial z_n} \end{bmatrix}. \tag{82}$$

By recalling that $\dfrac{\partial z^*}{\partial z} = 0$ [5], we have that

$$\frac{\partial \mathbf{z}^{\mathsf{H}}\mathbf{a}}{\partial \mathbf{z}} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{83}$$

$$\boxed{\frac{\partial \mathbf{z}^{\mathsf{H}}\mathbf{a}}{\partial \mathbf{z}} = \mathbf{0} \in \mathbb{C}^n} \tag{84}$$

where $\mathbf{0}$ is the zero vector.

4.7 $\quad \dfrac{\partial \mathbf{z}^{\mathsf{H}}\mathbf{a}}{\partial \mathbf{z}^*} = \mathbf{a}$

(TODO) [5]

17

4.8 $\dfrac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{v}} = \dfrac{\partial \mathbf{x}}{\partial \mathbf{v}}\mathbf{y} + \dfrac{\partial \mathbf{y}}{\partial \mathbf{v}}\mathbf{x}$

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^m$. Where $\mathbf{x}$ and $\mathbf{y}$ depend on $\mathbf{v}$. Thus,

$$\frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{v}} = \frac{\partial}{\partial \mathbf{v}} \sum_{i=1}^{n} y_i x_i \tag{85}$$

$$= \sum_{i=1}^{n} \frac{\partial y_i x_i}{\partial \mathbf{v}}. \tag{86}$$

Recalling that $(fg)' = f'g + g'f$, we have

$$\frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{v}} = \sum_{i=1}^{n} x_i \frac{\partial y_i}{\partial \mathbf{v}} + \sum_{i=1}^{n} y_i \frac{\partial x_i}{\partial \mathbf{v}} \tag{87}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} x_i \dfrac{\partial y_i}{\partial v_1} \\ \sum_{i=1}^{n} x_i \dfrac{\partial y_i}{\partial v_2} \\ \vdots \\ \sum_{i=1}^{n} x_i \dfrac{\partial y_i}{\partial v_m} \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^{n} y_i \dfrac{\partial x_i}{\partial v_1} \\ \sum_{i=1}^{n} y_i \dfrac{\partial x_i}{\partial v_2} \\ \vdots \\ \sum_{i=1}^{n} y_i \dfrac{\partial x_i}{\partial v_m} \end{bmatrix} \tag{88}$$

$$= \begin{bmatrix} \dfrac{\partial y_1}{\partial v_1} & \dfrac{\partial y_2}{\partial v_1} & \cdots & \dfrac{\partial y_n}{\partial v_1} \\ \dfrac{\partial y_1}{\partial v_2} & \dfrac{\partial y_2}{\partial v_2} & \cdots & \dfrac{\partial y_n}{\partial v_2} \\ \vdots & \ddots & \vdots & \vdots \\ \dfrac{\partial y_1}{\partial v_m} & \dfrac{\partial y_2}{\partial v_m} & \cdots & \dfrac{\partial y_n}{\partial v_m} \end{bmatrix} \mathbf{x} \tag{89}$$

$$+ \begin{bmatrix} \dfrac{\partial x_1}{\partial v_1} & \dfrac{\partial x_2}{\partial v_1} & \cdots & \dfrac{\partial x_n}{\partial v_1} \\ \dfrac{\partial x_1}{\partial v_2} & \dfrac{\partial x_2}{\partial v_2} & \cdots & \dfrac{\partial x_n}{\partial v_2} \\ \vdots & \ddots & \vdots & \vdots \\ \dfrac{\partial x_1}{\partial v_m} & \dfrac{\partial x_2}{\partial v_m} & \cdots & \dfrac{\partial x_n}{\partial v_m} \end{bmatrix} \mathbf{y} \tag{90}$$

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{v}} = \frac{\partial \mathbf{x}}{\partial \mathbf{v}}\mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{v}}\mathbf{x} \in \mathbb{R}^m} \tag{91}$$

Note that, if either $\mathbf{x}$ or $\mathbf{y}$ does not depend on $\mathbf{v}$, just disregard $\dfrac{\partial \mathbf{x}}{\partial \mathbf{v}}\mathbf{y}$ or $\dfrac{\partial \mathbf{y}}{\partial \mathbf{v}}\mathbf{x}$, respectively. When none depends on $\mathbf{v}$, the obvious result is the zero vector, $\mathbf{0}$. A simpler way to solve it is to apply the scalar-vector product rule (see (48)), that is,

$$\frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{v}} = \frac{\partial \mathbf{x}}{\partial \mathbf{v}} \frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{x}} + \frac{\partial \mathbf{y}}{\partial \mathbf{v}} \frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{y}} = \frac{\partial \mathbf{x}}{\partial \mathbf{v}}\mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{v}}\mathbf{x} \tag{92}$$

4.9 $\dfrac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}$

Let $\mathbf{x} \in \mathbb{R}^n$. Thus,

$$\frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \sum_{i=1}^{n} x_i^2 \tag{93}$$

$$= \sum_{i=1}^{n} \frac{\partial x_i^2}{\partial \mathbf{x}} \tag{94}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} \dfrac{\partial x_i^2}{\partial x_1} \\ \sum_{i=1}^{n} \dfrac{\partial x_i^2}{\partial x_2} \\ \vdots \\ \sum_{i=1}^{n} \dfrac{\partial x_i^2}{\partial x_m} \end{bmatrix} \tag{95}$$

$$= \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{bmatrix} \tag{96}$$

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x} \in \mathbb{R}^n} \tag{97}$$

Note that this perfectly matches with the derivate of a quadratic scalar value, i.e., $\frac{\mathrm{d}x^2}{\mathrm{d}x} = 2x$.

4.10 $\quad \dfrac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{v}} = 2\dfrac{\partial \mathbf{x}}{\partial \mathbf{v}}\mathbf{x}$

Let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^m$, where $\mathbf{x}$ depends on $\mathbf{v}$. Thus,

$$\frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{v}} = \frac{\partial}{\partial \mathbf{v}} \sum_{i=1}^{n} x_i^2 \tag{98}$$

$$= \sum_{i=1}^{n} \frac{\partial x_i^2}{\partial \mathbf{v}} \tag{99}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} \dfrac{\partial x_i^2}{\partial v_1} \\[2ex] \sum_{i=1}^{n} \dfrac{\partial x_i^2}{\partial v_2} \\[2ex] \vdots \\[2ex] \sum_{i=1}^{n} \dfrac{\partial x_i^2}{\partial v_m} \end{bmatrix} \tag{100}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} 2x_i \dfrac{\partial x_i}{\partial v_1} \\[2ex] \sum_{i=1}^{n} 2x_i \dfrac{\partial x_i}{\partial v_2} \\[2ex] \vdots \\[2ex] \sum_{i=1}^{n} 2x_i \dfrac{\partial x_i}{\partial v_m} \end{bmatrix} \tag{101}$$

$$= 2 \begin{bmatrix} \dfrac{\partial x_1}{\partial \mathbf{v}} & \dfrac{\partial x_2}{\partial \mathbf{v}} & \dots & \dfrac{\partial x_n}{\partial \mathbf{v}} \end{bmatrix} \mathbf{x} \tag{102}$$

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{v}} = 2\frac{\partial \mathbf{x}}{\partial \mathbf{v}}\mathbf{x} \in \mathbb{R}^m} \tag{103}$$

Note that this solution could also be solved by the chain rule (c.f. (40)) as follows

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{v}} = \frac{\partial \mathbf{x}}{\partial \mathbf{v}} \frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{x}} = 2\frac{\partial \mathbf{x}}{\partial \mathbf{v}}\mathbf{x} \in \mathbb{R}^m} \tag{104}$$

4.11 $\quad \dfrac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \left(\mathbf{A}^\top + \mathbf{A}\right)\mathbf{x}$

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

$$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \tag{105}$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} \sum_{i=1}^{n} x_i a_{i1} & \sum_{i=1}^{n} x_i a_{i2} & \dots & \sum_{i=1}^{n} x_i a_{in} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \tag{106}$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j \right). \tag{107}$$

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

$$
\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =
\begin{bmatrix}
\dfrac{\partial}{\partial x_1}\left(\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j\right) \\[2ex]
\dfrac{\partial}{\partial x_2}\left(\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j\right) \\[2ex]
\vdots \\[2ex]
\dfrac{\partial}{\partial x_n}\left(\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} x_i a_{ij} x_j\right)
\end{bmatrix}
\tag{108}
$$

$$
=
\begin{bmatrix}
2x_1 a_{11} + \displaystyle\sum_{\substack{j=1 \\ j\neq 1}}^{n} a_{1j} x_j + \sum_{\substack{i=1 \\ i\neq 1}}^{n} a_{i1} x_i \\[2ex]
2x_2 a_{22} + \displaystyle\sum_{\substack{j=1 \\ j\neq 2}}^{n} a_{2j} x_j + \sum_{\substack{i=1 \\ i\neq 2}}^{n} a_{i2} x_i \\[2ex]
\vdots \\[2ex]
2x_n a_{nn} + \displaystyle\sum_{\substack{j=1 \\ j\neq n}}^{n} a_{nj} x_j + \sum_{\substack{i=1 \\ i\neq n}}^{n} a_{in} x_i
\end{bmatrix}
\tag{109}
$$

$$
=
\begin{bmatrix}
\displaystyle\sum_{j=1}^{n} a_{1j} x_j \\[2ex]
\displaystyle\sum_{j=1}^{n} a_{2j} x_j \\[2ex]
\vdots \\[2ex]
\displaystyle\sum_{j=1}^{n} a_{nj} x_j
\end{bmatrix}
+
\begin{bmatrix}
\displaystyle\sum_{i=1}^{n} a_{i1} x_i \\[2ex]
\displaystyle\sum_{i=1}^{n} a_{i2} x_i \\[2ex]
\vdots \\[2ex]
\displaystyle\sum_{i=1}^{n} a_{in} x_i
\end{bmatrix}
\tag{110}
$$

$$
= \mathbf{A}^\top \mathbf{x} + \mathbf{A}\mathbf{x}
\tag{111}
$$

$$
\boxed{\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \left(\mathbf{A}^\top + \mathbf{A}\right)\mathbf{x} \in \mathbb{R}^n}
\tag{112}
$$

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

$$
\boxed{\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} \in \mathbb{R}^n}
\tag{113}
$$

4.12 $\dfrac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{v}} = \dfrac{\partial \mathbf{x}}{\partial \mathbf{v}}\left(\mathbf{A} + \mathbf{A}^\top\right)\mathbf{x}$

Let $\mathbf{v} \in \mathbb{R}^m, \mathbf{x} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{n \times n}$, where $\mathbf{x}$ depends on $\mathbf{v}$, but $\mathbf{A}$ does not. For the quadratic form, it follows that

$$\frac{\partial \mathbf{x}^\top \mathbf{A}\mathbf{x}}{\partial \mathbf{v}} = \frac{\partial}{\partial \mathbf{v}}\left(\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) \tag{114}$$

$$= \frac{\partial}{\partial \mathbf{v}}\left(\begin{bmatrix} \sum\limits_{i=1}^{n} x_i a_{i1} & \sum\limits_{i=1}^{n} x_i a_{i2} & \dots & \sum\limits_{i=1}^{n} x_i a_{in} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) \tag{115}$$

$$= \frac{\partial}{\partial \mathbf{v}}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j\right) \tag{116}$$

$$= \begin{bmatrix} \dfrac{\partial}{\partial v_1}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} x_i a_{ij} x_j\right) \\ \dfrac{\partial}{\partial v_2}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} x_i a_{ij} x_j\right) \\ \vdots \\ \dfrac{\partial}{\partial v_n}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} x_i a_{ij} x_j\right) \end{bmatrix} \tag{117}$$

$$= \begin{bmatrix} \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_{ij}\dfrac{\partial x_i x_j}{\partial v_1} \\ \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_{ij}\dfrac{\partial x_i x_j}{\partial v_2} \\ \vdots \\ \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_{ij}\dfrac{\partial x_i x_j}{\partial v_n} \end{bmatrix} \tag{118}$$

Recalling that $(fg)' = f'g + g'f$, we have that

$$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{v}} = \begin{bmatrix} \sum_{i=1}^{n} \sum_{j=1}^{n} x_j a_{ij} \frac{\partial x_i}{\partial v_1} + \sum_{i=1}^{n} \sum_{j=1}^{n} x_i a_{ij} \frac{\partial x_j}{\partial v_1} \\ \sum_{i=1}^{n} \sum_{j=1}^{n} x_j a_{ij} \frac{\partial x_i}{\partial v_2} + \sum_{i=1}^{n} \sum_{j=1}^{n} x_i a_{ij} \frac{\partial x_j}{\partial v_2} \\ \vdots \\ \sum_{i=1}^{n} \sum_{j=1}^{n} x_j a_{ij} \frac{\partial x_i}{\partial v_n} + \sum_{i=1}^{n} \sum_{j=1}^{n} x_i a_{ij} \frac{\partial x_j}{\partial v_n} \end{bmatrix} \tag{119}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} \sum_{j=1}^{n} x_j a_{ij} \frac{\partial x_i}{\partial v_1} \\ \sum_{i=1}^{n} \sum_{j=1}^{n} x_j a_{ij} \frac{\partial x_i}{\partial v_2} \\ \vdots \\ \sum_{i=1}^{n} \sum_{j=1}^{n} x_j a_{ij} \frac{\partial x_i}{\partial v_n} \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i a_{ij} \frac{\partial x_j}{\partial v_1} \\ \sum_{i=1}^{n} \sum_{j=1}^{n} x_i a_{ij} \frac{\partial x_j}{\partial v_2} \\ \vdots \\ \sum_{i=1}^{n} \sum_{j=1}^{n} x_i a_{ij} \frac{\partial x_j}{\partial v_n} \end{bmatrix} \tag{120}$$

$$= \begin{bmatrix} \frac{\partial x_1}{\partial \mathbf{v}} & \frac{\partial x_2}{\partial \mathbf{v}} & \cdots & \frac{\partial x_n}{\partial \mathbf{v}} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} +$$

$$\begin{bmatrix} \frac{\partial x_1}{\partial \mathbf{v}} & \frac{\partial x_2}{\partial \mathbf{v}} & \cdots & \frac{\partial x_n}{\partial \mathbf{v}} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{121}$$

$$= \frac{\partial \mathbf{x}}{\partial \mathbf{v}} \mathbf{A} \mathbf{x} + \frac{\partial \mathbf{x}}{\partial \mathbf{v}} \mathbf{A}^\top \mathbf{x} \tag{122}$$

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{v}} = \frac{\partial \mathbf{x}}{\partial \mathbf{v}} \left(\mathbf{A} + \mathbf{A}^\top \right) \mathbf{x} \in \mathbb{C}^m} \tag{123}$$

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

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{v}} = 2 \frac{\partial \mathbf{x}}{\partial \mathbf{v}} \mathbf{A} \mathbf{x} \in \mathbb{C}^m} \tag{124}$$

Note that the solution is much easier if we maintain the Matrix Calculus notation and apply the chain rule, that is,

$$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{v}} = \frac{\partial \mathbf{x}}{\partial \mathbf{v}} \frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}}{\partial \mathbf{v}} \left(\mathbf{A}^\top + \mathbf{A} \right) \mathbf{x}, \tag{125}$$

where the last equality comes from (112).

4.13 $\quad \dfrac{\partial \mathbf{b}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top \mathbf{b}$

Let $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$ and $\mathbf{A} \in \mathbb{R}^{m \times n}$, where neither $\mathbf{b}$ nor $\mathbf{A}$ depend on $\mathbf{x}$. It follows that

$$\frac{\partial \mathbf{b}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} b_1 & b_2 & \ldots & b_m \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \tag{126}$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} \sum\limits_{i=1}^{m} a_{i1} b_i & \sum\limits_{i=1}^{m} a_{i2} b_i & \ldots & \sum\limits_{i=1}^{m} a_{in} b_i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \tag{127}$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_i x_j \right) \tag{128}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_i \frac{\partial x_j}{\partial \mathbf{x}} \tag{129}$$

$$= \begin{bmatrix} \sum_{i=1}^{m} a_{i1} b_i \\ \sum_{i=1}^{m} a_{i2} b_i \\ \vdots \\ \sum_{i=1}^{m} a_{in} b_i \end{bmatrix} \tag{130}$$

$$= \begin{bmatrix} a_{11} & a_{21} & \ldots & a_{m1} \\ a_{12} & a_{22} & \ldots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \tag{131}$$

$$\boxed{\frac{\partial \mathbf{b}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top \mathbf{b} \in \mathbb{R}^n} \tag{132}$$

Note that this solution could solve by simply observing that $\mathbf{b}^\top \mathbf{A}$ is actually a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}$. Thus,

$$\frac{\partial \mathbf{b}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = \left(\mathbf{b}^\top \mathbf{A} \right)^\top = \mathbf{A}^\top \mathbf{b}, \tag{133}$$

where the first equality comes from the (58).

4.14 $\quad \dfrac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{b}}{\partial \mathbf{x}} = \mathbf{A} \mathbf{b}$

Let $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{b} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{m \times n}$, where neither $\mathbf{b}$ nor $\mathbf{A}$ depend on $\mathbf{x}$. The quickest way to solve it is to note that $\mathbf{x}^\top \mathbf{A} \mathbf{b} = \mathbf{b}^\top \mathbf{A}^\top \mathbf{x}$, which is the problem solved by the Section 4.13. Thus,

$$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{b}}{\partial \mathbf{x}} = \frac{\partial \mathbf{b}^\top \mathbf{A}^\top \mathbf{x}}{\partial \mathbf{x}} = \left(\mathbf{b}^\top \mathbf{A}^\top \right)^\top = \mathbf{A} \mathbf{b}, \tag{134}$$

where the second equality comes from (58). Nevertheless, here is the step-by-step

$$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{b}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} x_1 & x_2 & \dots & x_m \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \right) \tag{135}$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} \sum_{i=1}^m a_{i1} x_i & \sum_{i=1}^m a_{i2} x_i & \dots & \sum_{i=1}^m a_{in} x_i \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \right) \tag{136}$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^m \sum_{j=1}^n a_{ij} b_j x_i \right) \tag{137}$$

$$= \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_j \frac{\partial x_i}{\partial \mathbf{x}} \tag{138}$$

$$= \begin{bmatrix} \sum_{j=1}^n a_{1j} b_j \\ \sum_{j=1}^n a_{2j} b_j \\ \vdots \\ \sum_{j=1}^n a_{nj} b_j \end{bmatrix} \tag{139}$$

$$= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \tag{140}$$

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{b}}{\partial \mathbf{x}} = \mathbf{A} \mathbf{b} \in \mathbb{R}^m} \tag{141}$$

4.15 $\quad \dfrac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{v}} = \dfrac{\partial \mathbf{x}}{\partial \mathbf{v}} \mathbf{A}^\top \mathbf{y} + \dfrac{\partial \mathbf{y}}{\partial \mathbf{v}} \mathbf{A} \mathbf{x}$

Let $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$, $\mathbf{v} \in \mathbb{R}^p$, and $\mathbf{A} \in \mathbb{R}^{m \times n}$, where $\mathbf{x}$ and $\mathbf{y}$ depend on $\mathbf{v}$, but $\mathbf{A}$ does not. Therefore,

$$\frac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{v}} = \frac{\partial}{\partial \mathbf{v}} \left(\begin{bmatrix} y_1 & y_2 & \dots & y_m \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \tag{142}$$

$$= \frac{\partial}{\partial \mathbf{v}} \left(\begin{bmatrix} \sum_{i=1}^m a_{i1} y_i & \sum_{i=1}^m a_{i2} y_i & \dots & \sum_{i=1}^m a_{in} y_i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \tag{143}$$

$$= \frac{\partial}{\partial \mathbf{v}} \left(\sum_{i=1}^m \sum_{j=1}^n a_{ij} x_j y_i \right) \tag{144}$$

$$= \sum_{i=1}^m \sum_{j=1}^n a_{ij} \frac{\partial y_i x_j}{\partial \mathbf{v}} \tag{145}$$

Recalling that $(fg)' = f'g + g'f$, we have that

$$\frac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{v}} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} y_i \frac{\partial x_j}{\partial \mathbf{v}} + \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_j \frac{\partial y_i}{\partial \mathbf{v}} \tag{146}$$

$$= \begin{bmatrix} \dfrac{\partial x_1}{\partial \mathbf{v}} & \dfrac{\partial x_2}{\partial \mathbf{v}} & \cdots & \dfrac{\partial x_n}{\partial \mathbf{v}} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} + \tag{147}$$

$$\begin{bmatrix} \dfrac{\partial y_1}{\partial \mathbf{v}} & \dfrac{\partial y_2}{\partial \mathbf{v}} & \cdots & \dfrac{\partial y_m}{\partial \mathbf{v}} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{148}$$

$$\boxed{\frac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{v}} = \frac{\partial \mathbf{x}}{\partial \mathbf{v}} \mathbf{A}^\top \mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{v}} \mathbf{A} \mathbf{x}} \tag{149}$$

Even though this problem is trickier, we can find the same solution in a clever way by preserving the Matrix Calculus notation and applying the chain rule. Note that $\mathbf{y}^\top \mathbf{A} \mathbf{x}$ depends on both $\mathbf{x}$ and $\mathbf{y}$ which, in turn, depend on $\mathbf{v}$. Therefore (c.f. (42)),

$$\frac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{v}} = \frac{\partial \mathbf{x}}{\partial \mathbf{v}} \frac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} + \frac{\partial \mathbf{y}}{\partial \mathbf{v}} \frac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{y}} \tag{150}$$

$$= \frac{\partial \mathbf{x}}{\partial \mathbf{v}} \mathbf{A}^\top \mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{v}} \mathbf{A} \mathbf{x}. \tag{151}$$

4.16 $\quad \dfrac{\partial \mathrm{tr}\,(\mathbf{A} \mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^\top$

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

$$\frac{\partial \mathrm{tr}\,(\mathbf{A} \mathbf{X})}{\partial \mathbf{X}} = \frac{\partial}{\partial \mathbf{X}} \left(\mathrm{tr} \left(\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix} \right) \right)$$

$$= \frac{\partial}{\partial \mathbf{X}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right)$$

$$= \begin{bmatrix} \dfrac{\partial}{\partial x_{11}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) & \dfrac{\partial}{\partial x_{12}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) & \cdots & \dfrac{\partial}{\partial x_{1n}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) \\[2em] \dfrac{\partial}{\partial x_{21}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) & \dfrac{\partial}{\partial x_{22}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) & \cdots & \dfrac{\partial}{\partial x_{2n}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) \\[2em] \vdots & \vdots & \ddots & \vdots \\[1em] \dfrac{\partial}{\partial x_{n1}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) & \dfrac{\partial}{\partial x_{n2}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) & \cdots & \dfrac{\partial}{\partial x_{nn}} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_{ji} \right) \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix}$$

$$\boxed{\frac{\partial \mathrm{tr}\left(\mathbf{AX}\right)}{\partial \mathbf{X}} = \mathbf{A}^{\top}} \tag{152}$$

I have no idea how to make this solution simpler.

4.17 $\quad \dfrac{\partial \left|\mathbf{X}\right|}{\partial \mathbf{X}} = \mathrm{adj}\left(\mathbf{X}\right)$

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

$$\left|\mathbf{X}\right| = \sum_{i=1}^{n} x_{ki}\left|\mathbf{C}_{ki}\right| = \sum_{i=1}^{n} x_{ik}\left|\mathbf{C}_{ik}\right| \quad \forall\, k \in \{1, 2, ..., n\}, \tag{153}$$

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

$$\frac{\partial \left|\mathbf{X}\right|}{\partial \mathbf{X}} = \frac{\partial}{\partial \mathbf{X}}\left(\sum_{i=1}^{n} x_{ki}\left|\mathbf{C}_{ki}\right|\right) \quad \forall\, k \in \{1, 2, ..., n\} \tag{154}$$

$$= \frac{\partial}{\partial \mathbf{X}}\left(\begin{bmatrix} \sum_{i=1}^{n} x_{1i}\left|\mathbf{C}_{1i}\right| & \sum_{i=1}^{n} x_{1i}\left|\mathbf{C}_{1i}\right| & \cdots & \sum_{i=1}^{n} x_{1i}\left|\mathbf{C}_{1i}\right| \\ \sum_{i=1}^{n} x_{2i}\left|\mathbf{C}_{2i}\right| & \sum_{i=1}^{n} x_{2i}\left|\mathbf{C}_{2i}\right| & \cdots & \sum_{i=1}^{n} x_{2i}\left|\mathbf{C}_{2i}\right| \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ni}\left|\mathbf{C}_{ni}\right| & \sum_{i=1}^{n} x_{ni}\left|\mathbf{C}_{ni}\right| & \cdots & \sum_{i=1}^{n} x_{ni}\left|\mathbf{C}_{ni}\right| \end{bmatrix}\right) \tag{155}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_{11}}\left(\sum_{i=1}^{n} x_{1i}\left|\mathbf{C}_{1i}\right|\right) & \frac{\partial}{\partial x_{12}}\left(\sum_{i=1}^{n} x_{1i}\left|\mathbf{C}_{1i}\right|\right) & \cdots & \frac{\partial}{\partial x_{13}}\left(\sum_{i=1}^{n} x_{1i}\left|\mathbf{C}_{1i}\right|\right) \\ \frac{\partial}{\partial x_{21}}\left(\sum_{i=1}^{n} x_{2i}\left|\mathbf{C}_{2i}\right|\right) & \frac{\partial}{\partial x_{22}}\left(\sum_{i=1}^{n} x_{2i}\left|\mathbf{C}_{2i}\right|\right) & \cdots & \frac{\partial}{\partial x_{33}}\left(\sum_{i=1}^{n} x_{2i}\left|\mathbf{C}_{2i}\right|\right) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_{n1}}\left(\sum_{i=1}^{n} x_{ni}\left|\mathbf{C}_{ni}\right|\right) & \frac{\partial}{\partial x_{n2}}\left(\sum_{i=1}^{n} x_{ni}\left|\mathbf{C}_{ni}\right|\right) & \cdots & \frac{\partial}{\partial x_{n3}}\left(\sum_{i=1}^{n} x_{ni}\left|\mathbf{C}_{ni}\right|\right) \end{bmatrix} \tag{156}$$

$$= \begin{bmatrix} \left|\mathbf{C}_{11}\right| & \left|\mathbf{C}_{12}\right| & \cdots & \left|\mathbf{C}_{1n}\right| \\ \left|\mathbf{C}_{21}\right| & \left|\mathbf{C}_{22}\right| & \cdots & \left|\mathbf{C}_{2n}\right| \\ \vdots & \vdots & \ddots & \vdots \\ \left|\mathbf{C}_{n1}\right| & \left|\mathbf{C}_{n2}\right| & \cdots & \left|\mathbf{C}_{nn}\right| \end{bmatrix} \tag{157}$$

$$\boxed{\frac{\partial \left|\mathbf{X}\right|}{\partial \mathbf{X}} = \mathrm{adj}\left(\mathbf{X}\right)} \tag{158}$$

I have no idea how to make this solution simpler.

4.18 $\quad \dfrac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -\left(\mathbf{A}^{-1}\right)^{\top}\dfrac{\partial \mathbf{A}}{\partial \alpha}\left(\mathbf{A}^{-1}\right)^{\top}$

Let $\mathbf{A} \in \mathbb{C}^{m \times n}$ and $\alpha \in \mathbb{C}$. Remember that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. Differentiating both sides of this equation with respect to α, we get

$$\frac{\partial \mathbf{A}^{-1}\mathbf{A}}{\partial \alpha} = \frac{\partial \mathbf{I}}{\partial \alpha} = \mathbf{0}_{m \times n}, \tag{159}$$

where $\mathbf{0}_{m \times n}$ is a zero matrix with dimension $m \times n$. By applying the product rule of a matrix-matrix derivate, we get (c.f. (52))

$$\frac{\partial \mathbf{A}^{-1} \mathbf{A}}{\partial \alpha} = \frac{\partial \mathbf{A}}{\partial \alpha} \left(\mathbf{A}^{-1}\right)^\top + \mathbf{A}^\top \frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = \mathbf{0}_{m \times n} \tag{160}$$

Using the property $\left(\mathbf{A}^\top\right)^{-1} = \left(\mathbf{A}^{-1}\right)^\top$ and rearranging this expression, we get

$$\boxed{\frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -\left(\mathbf{A}^{-1}\right)^\top \frac{\partial \mathbf{A}}{\partial \alpha} \left(\mathbf{A}^{-1}\right)^\top} \tag{161}$$

References

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

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

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

[4] Shayle R Searle and Andre I Khuri. *Matrix algebra useful for statistics*. John Wiley & Sons, 2017.

[5] Are Hjørungnes. *Complex-valued matrix derivatives: with applications in signal processing and communications*. Cambridge University Press, 2011.

[6] Are Hjorungnes and David Gesbert. "Complex-valued matrix differentiation: Techniques and key results". In: *IEEE Transactions on Signal Processing* 55.6 (2007), pp. 2740–2746.

[7] Terence Parr and Jeremy Howard. *The Matrix Calculus You Need For Deep Learning*. July 2, 2018. arXiv: 1802.01528 [cs, stat]. URL: http://arxiv.org/abs/1802.01528. preprint.

[8] Jon Dattorro. *Convex Optimization & Euclidean Distance Geometry*. Lulu. com, 2010. ISBN: 0-615-19368-4.

[9] Christopher M Bishop and Nasser M Nasrabadi. *Pattern Recognition and Machine Learning*. Vol. 4. 4. Springer, 2006.

[10] Simon S Haykin et al. *Neural networks and learning machines*. 2009.

[11] Old and New Matrix Algebra Useful for Statistics. *Lecture notes of Macroeconomics*. 2000. URL: https://tminka.github.io/papers/matrix/minka-matrix.pdf.

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

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

[14] Cesar Augusto Taconeli. *Lecture notes in Análise de Regressão Linear*. 2018. URL: https://docs.ufpr.br/~taconeli/CE07118/Algebra.pdf.

[15] Paul Klein. *Lecture notes of Macroeconomics*. 1999. URL: http://paulklein.ca/newsite/teaching/calcvec.pdf.

[16] Simon S Haykin. *Adaptive Filter Theory*. Pearson Education India, 2002. ISBN: 81-317-0869-1.