

The Guide for Matrix Calculus

Rubem Vasconcelos Pacelli
rubem.engenharia@gmail.com

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

Version: October 24, 2022

Contents

1	Introduction	3
2	Notation and nomenclatures	3
2.1	Jacobian formulation (numerator layout)	4
2.1.1	Vector-vector, scalar-vector, and vector-scalar derivatives	4
2.1.2	Matrix-scalar derivative (tangent matrix)	4
2.1.3	Scalar-matrix derivative (gradient matrix)	4
2.1.4	The hidden equality for the numerator layout	5
2.1.5	Jacobian matrix for the numerator layout	5
2.1.6	Hessian matrix for the numerator layout	6
2.2	Hessian formulation (denominator layout)	6
2.2.1	Vector-vector, scalar-vector, and vector-scalar derivatives	6
2.2.2	Matrix-scalar derivative (tangent matrix)	6
2.2.3	Scalar-matrix derivative (gradient matrix)	7
2.2.4	The hidden equality for the denominator layout	7
2.2.5	The Jacobian matrix for the denominator layout	7
2.2.6	The Hessian matrix for the denominator layout	8
2.3	Comparative between Jacobian and Hessian formulations	9
2.4	Notations not widely agreed upon	10
3	Identities	10
3.1	Chain rule	10
3.1.1	Univariate functions	10
3.1.2	Multivariate functions	11
3.2	Sum (or minus) rule	11
3.2.1	Vector-vector derivative	11
3.2.2	Matrix-scalar derivative	11
3.2.3	Scalar-matrix derivative	11
3.3	Product rule	11
3.3.1	Vector-vector derivative	11
3.3.2	Scalar-vector derivative	12
3.3.3	Scalar-matrix derivative	12
3.3.4	Matrix-scalar derivative	12
4	Solution of Matrix Differentiations	12
4.1	$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \mathbf{A}^\top$	12
4.2	$\frac{\partial \mathbf{Ax}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{A}^\top$	13
4.3	$\frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$	13

4.4	$\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{a}^H \mathbf{x}} = \mathbf{a}$	14
4.5	$\frac{\partial \mathbf{x}}{\partial \mathbf{a}^H \mathbf{x}} = \mathbf{a}^*$	14
4.6	$\frac{\partial \mathbf{x}}{\partial \mathbf{x}^H \mathbf{a}} = \mathbf{0}$	15
4.7	$\frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{z}} \mathbf{x}$	16
4.8	$\frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{z}} = 2\mathbf{x}$	17
4.9	$\frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{z}} = 2 \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{x}$	18
4.10	$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{z}} = (\mathbf{A}^\top + \mathbf{A}) \mathbf{x}$	18
4.11	$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} (\mathbf{A} + \mathbf{A}^\top) \mathbf{x}$	20
4.12	$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{b}^\top \mathbf{A} \mathbf{x}} = \mathbf{A}^\top \mathbf{b}$	21
4.13	$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{b}}{\partial \mathbf{b}^\top \mathbf{A} \mathbf{x}} = \mathbf{A} \mathbf{b}$	22
4.14	$\frac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{A}^\top \mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{z}} \mathbf{A} \mathbf{x}$	23
4.15	$\frac{\partial \text{tr}(\mathbf{A} \mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^\top$	24
4.16	$\frac{\partial \mathbf{X} }{\partial \mathbf{X}} = \text{adj}(\mathbf{X})$	25
4.17	$\frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -(\mathbf{A}^{-1})^\top \frac{\partial \mathbf{A}}{\partial \alpha} (\mathbf{A}^{-1})^\top$	25

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] (honorable mention for the Matrix Cookbook too [3]). 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 [4]. If you do not select the author’s representation, you will end up with the same result, but in a row vector¹ instead of a column vector or 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. Nevertheless, even for the same layout, some conventions need to be stated as there are some disagreements.

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 a quick-but-comprehensive explanation of both representations, and derive the partial derivatives for the most common matrix calculus expressions you came across. In the end, I hope to have a consistent reference to guide me when writing my articles, or to help me when I need to understand an author’s book that uses notations different from mine. The way I define in the Section 2.2 will be the way I going to adopt throughout my papers, but I will let you know when some point is not consensus among the authors. If you are only looking for a table of results, the most thorough I’ve seen so far is on Wikipedia [5], but maybe it is not the most reliable source.

I will adopt the notation that most Engineers might be used to and only apply the Hessian formulation on the derivations since this is the notation I will adopt on my papers (it matches the derivative results I find in my books, too). If you are looking for the Jacobian formulation for the same derivations I do here, I highly recommend Professor Randal’s class note, which uses this representation. The unique drawback is that he does not use complex numbers and has fewer derivations than mine.

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 nomenclatures

Let

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

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

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

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

Nonbold Romain and Greek alphabets represent scalars, while bold uppercase and bold lowercase represent matrices and vectors, respectively. I make no distinction between scalar- and vector-valued functions, both are denoted as nonbold letters. In Section 4, I will try to use the initial letters of

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

the Romain alphabet (a, b, c, \dots) to represent constants (known values), and the final letters of the Romain alphabet (x, y, z, w, \dots) to represent variables (unknown values). Greek letters will be preferred to represent independent and unknown scalars that are not within a vector or matrix. Finally, the operators \cdot^\top , \cdot^H , \cdot^* , $\text{tr}(\cdot)$, $\text{adj}(\cdot)$, 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 $\mathbf{x} \in \mathbb{C}^n$ and $\mathbf{y} \in \mathbb{C}^m$. The partial derivative of each element in \mathbf{y} by each element in \mathbf{x} is represented as

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

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

2.1.2 Matrix-scalar derivative (tangent matrix)

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

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

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

2.1.3 Scalar-matrix derivative (gradient matrix)

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

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

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

2.1.4 The hidden equality for the numerator layout

From these definitions, we can infer a nonobvious equality that is rather useful when handling matrix differentiations. If we consider a special case of the gradient matrix (Eq.(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

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

However, by using we definition from the Eq.(3), it is also true to state that

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

Therefore,

$$\left[\frac{\partial \alpha}{\partial \mathbf{x}^\top} \right]_{\text{Num}} = \left[\frac{\partial \alpha}{\partial \mathbf{x}} \right]_{\text{Num}}^\top \quad (8)$$

2.1.5 Jacobian matrix for the numerator layout

The matrix calculus notation for the Jacobian matrix is given by

$$\mathbf{J} = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}, \quad (9)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a vector function. As shown below, the matrix differentiation notation used to represent the Jacobian matrix is the numerator layout (c.f. Eq.(3))

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1}{\partial \mathbf{x}} \\ \frac{\partial f_2}{\partial \mathbf{x}} \\ \vdots \\ \frac{\partial f_m}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m \times n}, \quad (10)$$

where $f = (f_1, f_2, \dots, f_m)$, being $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, for $1 \leq i \leq m$. Therefore, it is true to state that

$$\mathbf{J} = \left[\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right]_{\text{Num}}. \quad (11)$$

That is why it is also called “Jacobian formulation”.

2.1.6 Hessian matrix for the numerator layout

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

$$\mathbf{H} = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^\top} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad (12)$$

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

$$\mathbf{H} = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^\top} = \frac{\partial}{\partial \mathbf{x}} \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^\top} \right) = \frac{\partial}{\partial \mathbf{x}} \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right)^\top. \quad (13)$$

Note that $\left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right)^\top = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right)$ is a vector. Therefore, we have a vector-vector derivate and, by applying the differentiation as shown in Eq.(3), we get \mathbf{H} . As we will see further on, the Hessian in matrix calculus notation for the denominator layout needs to be written differently.

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

$$\left[\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right]_{\text{Den}} = \begin{bmatrix} \frac{\partial y_1}{\partial \mathbf{x}} & \frac{\partial y_2}{\partial \mathbf{x}} & \cdots & \frac{\partial y_m}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \in \mathbb{C}^{n \times m}. \quad (14)$$

We can infer what is the shape of $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ and $\frac{\partial \mathbf{y}}{\partial x}$ by changing the respective vector sizes in Equation (14).

2.2.2 Matrix-scalar derivative (tangent matrix)

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

$$\left[\frac{\partial \mathbf{Y}}{\partial x} \right]_{\text{Den}} = \begin{bmatrix} \frac{\partial y_{11}}{\partial x} & \frac{\partial y_{21}}{\partial x} & \cdots & \frac{\partial y_{m1}}{\partial x} \\ \frac{\partial y_{12}}{\partial x} & \frac{\partial y_{22}}{\partial x} & \cdots & \frac{\partial y_{m2}}{\partial x} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{1n}}{\partial x} & \frac{\partial y_{2n}}{\partial x} & \cdots & \frac{\partial y_{mn}}{\partial x} \end{bmatrix} \in \mathbb{C}^{n \times m}, \quad (15)$$

where $\mathbf{Y} \in \mathbb{C}^{m \times n}$. However, be aware that some authors do not follow this convention for the tangent matrix for the denominator layout [5]. For sake of consistency (laying out the denominator and the transpose of the numerator), I will follow the convention as denoted in the Eq.(15).

2.2.3 Scalar-matrix derivative (gradient matrix)

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

$$\left[\frac{\partial y}{\partial \mathbf{X}} \right]_{\text{Den}} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{m1}} & \frac{\partial y}{\partial x_{m2}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{bmatrix} \in \mathbb{C}^{m \times n}, \quad (16)$$

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

2.2.4 The hidden equality for the denominator layout

From these definitions, we can infer a nonobvious equality that is rather useful when handling matrix differentiations. If we consider a special case of the gradient matrix (Eq.(16)) 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

$$\left[\frac{\partial \alpha}{\partial \mathbf{x}^\top} \right]_{\text{Den}} = \begin{bmatrix} \frac{\partial \alpha}{\partial x_1} & \frac{\partial \alpha}{\partial x_2} & \cdots & \frac{\partial \alpha}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{1 \times n}, \quad (17)$$

however, by using we definition from the Eq.(14), it is also true to state that

$$\left[\frac{\partial \alpha}{\partial \mathbf{x}} \right]_{\text{Den}} = \begin{bmatrix} \frac{\partial \alpha}{\partial x_1} \\ \frac{\partial \alpha}{\partial x_2} \\ \vdots \\ \frac{\partial \alpha}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n. \quad (18)$$

Therefore,

$$\left[\frac{\partial \alpha}{\partial \mathbf{x}^\top} \right]_{\text{Den}} = \left[\frac{\partial \alpha}{\partial \mathbf{x}} \right]_{\text{Den}}^\top \quad (19)$$

2.2.5 The Jacobian matrix for the denominator layout

The matrix calculus notation for the Jacobian matrix is given by

$$\mathbf{J} = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}, \quad (20)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a vector function. We have seen that this equation follows the numerator layout, not the denominator. As we will see further on, to transform from one layout to another, we

need to apply the transpose operator, that is (c.f. Eq.(10)),

$$[\mathbf{J}]_{\text{Den}} = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}^\top \quad (21)$$

$$= \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}^\top \quad (22)$$

$$= \left[\frac{\partial f_1(\mathbf{x})}{\partial \mathbf{x}} \quad \frac{\partial f_2(\mathbf{x})}{\partial \mathbf{x}} \quad \cdots \quad \frac{\partial f_m(\mathbf{x})}{\partial \mathbf{x}} \right]^\top \quad (23)$$

$$= \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial \mathbf{x}}^\top \\ \frac{\partial f_2(\mathbf{x})}{\partial \mathbf{x}}^\top \\ \vdots \\ \frac{\partial f_m(\mathbf{x})}{\partial \mathbf{x}}^\top \end{bmatrix} \quad (24)$$

$$= \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m \times n}. \quad (25)$$

2.2.6 The Hessian matrix for the denominator layout

The matrix calculus notation for the Hessian matrix in the denominator layout is given by

$$\mathbf{H} = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}^2} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad (26)$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^n$. Some authors, such as Simon Haykin [6], adopt this notation. The element $\partial \mathbf{x}^2$ seems to be merely a convention which means that f is being differentiated twice by \mathbf{x} . It perfectly matches scalar differentiation notation and can be scaled in n consecutive derivatives without making the notation cumbersome (points to the Hessian team).

We can rewrite the equation (26) as

$$\mathbf{H} = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}^2} = \frac{\partial}{\partial \mathbf{x}} \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right), \quad (27)$$

Since $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right)$ is a vector, we have a vector-vector differentiation which,

for the denominator layout, yields

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}. \quad (28)$$

Therefore, it is true to state that

$$\mathbf{H} = \left[\frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}^2} \right]_{\text{Den}}. \quad (29)$$

2.3 Comparative between Jacobian and Hessian formulations

As you could have noticed,

$$\left[\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right]_{\text{Num}} = \left[\frac{\partial \mathbf{y}}{\partial \mathbf{x}} \right]_{\text{Den}}^{\top}, \quad (30)$$

$$\left[\frac{\partial \mathbf{Y}}{\partial x} \right]_{\text{Num}} = \left[\frac{\partial \mathbf{Y}}{\partial x} \right]_{\text{Den}}^{\top}, \quad (31)$$

$$\left[\frac{\partial y}{\partial \mathbf{X}} \right]_{\text{Num}} = \left[\frac{\partial y}{\partial \mathbf{X}} \right]_{\text{Den}}^{\top}. \quad (32)$$

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 result². 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 (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:

- $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ 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”).
- $\mathbf{J} = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}}^{\top}$.
- $\mathbf{H} = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}^2}$.
- $\frac{\partial \alpha}{\partial \mathbf{x}^{\top}} = \frac{\partial \alpha}{\partial \mathbf{x}}^{\top}$ for $\alpha \in \mathbb{R}$.

²First, you should apply $[\cdot]_{\text{Den}} = [\cdot]_{\text{Num}}^{\top}$ on partial derivatives that you get in the solution. Then, you apply the transpose to the whole solution.

2.4 Notations not widely agreed upon

Expressions such as $\frac{\partial \mathbf{y}}{\partial \mathbf{X}}$, $\frac{\partial \mathbf{Y}}{\partial \mathbf{X}}$, or $\frac{\partial \mathbf{Y}}{\partial \mathbf{x}}$ have no agreement for 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: a matrix whose 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

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

and, for a consistent denominator layout, we would have

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

where $\mathbf{X} \in \mathbb{C}^{m \times n}$ and $\mathbf{Y} \in \mathbb{C}^{n \times m}$. Notwithstanding, keep in mind that both equations are not standard and their usage must be acknowledged by the reader.

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

$$\frac{\partial w}{\partial z} = \frac{\partial w}{\partial x} \frac{\partial x}{\partial y} \frac{\partial y}{\partial z}. \quad (35)$$

Similarly, in matrix notation, we have

$$\frac{\partial \mathbf{w}}{\partial \mathbf{z}} = \frac{\partial \mathbf{y}}{\partial \mathbf{z}} \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \frac{\partial \mathbf{w}}{\partial \mathbf{x}}, \quad (36)$$

where $\mathbf{x} \in \mathbb{C}^n$, $\mathbf{y} \in \mathbb{C}^m$, $\mathbf{z} \in \mathbb{C}^p$, 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{z} . If $\mathbf{w} = f(\mathbf{x})$, $\mathbf{x} = g(\mathbf{y})$, and $\mathbf{y} = h(\mathbf{z})$, 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

$$\frac{\partial w}{\partial z} = \frac{\partial w}{\partial x} \frac{\partial x}{\partial z} + \frac{\partial w}{\partial y} \frac{\partial y}{\partial z}. \quad (37)$$

Similarly, for matrix calculus notation, we have

$$\frac{\partial \mathbf{w}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \frac{\partial \mathbf{w}}{\partial \mathbf{x}} + \frac{\partial \mathbf{y}}{\partial \mathbf{z}} \frac{\partial \mathbf{w}}{\partial \mathbf{y}} \quad (38)$$

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

$$\frac{\partial(a\mathbf{x} \pm b\mathbf{y})}{\partial \mathbf{w}} = a \frac{\partial \mathbf{x}}{\partial \mathbf{w}} \pm b \frac{\partial \mathbf{y}}{\partial \mathbf{w}} \quad (39)$$

3.2.2 Matrix-scalar derivative

Another is when you have

$$\frac{\partial(a\mathbf{X} \pm b\mathbf{Y})}{\partial \alpha}, \quad (40)$$

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

$$\frac{\partial(a\mathbf{X} \pm b\mathbf{Y})}{\partial \alpha} = a \frac{\partial \mathbf{X}}{\partial \alpha} \pm b \frac{\partial \mathbf{Y}}{\partial \alpha}. \quad (41)$$

3.2.3 Scalar-matrix derivative

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

$$\frac{\partial(ax \pm by)}{\partial \mathbf{W}} = a \frac{\partial x}{\partial \mathbf{W}} \pm b \frac{\partial y}{\partial \mathbf{W}}, \quad (42)$$

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

3.3 Product rule

3.3.1 Vector-vector derivative

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

$$\frac{\partial w \mathbf{z}}{\partial \mathbf{x}} = w \frac{\partial \mathbf{z}}{\partial \mathbf{x}} + \frac{\partial w}{\partial \mathbf{x}} \mathbf{z}^\top. \quad (43)$$

Note that is not possible to apply the product rule when you have $\mathbf{W}\mathbf{z}$, where $\mathbf{W} \in \mathbb{C}^{n \times m}$ also depends on \mathbf{x} . If you tried, you would get $\partial \mathbf{W} / \partial \mathbf{x}$, which does not exist.

3.3.2 Scalar-vector derivative

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

$$\frac{\partial \mathbf{w}^\top \mathbf{z}}{\partial \mathbf{x}} = \frac{\partial \mathbf{z}}{\partial \mathbf{x}} \mathbf{w} + \frac{\partial \mathbf{w}}{\partial \mathbf{x}} \mathbf{z}. \quad (44)$$

3.3.3 Scalar-matrix derivative

It is still possible to apply the product rule to

$$\frac{\partial wz}{\partial \mathbf{X}}, \quad (45)$$

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

$$\frac{\partial wz}{\partial \mathbf{X}} = w \frac{\partial z}{\partial \mathbf{X}} + z \frac{\partial w}{\partial \mathbf{X}}. \quad (46)$$

3.3.4 Matrix-scalar derivative

The last case is when you have

$$\frac{\partial \mathbf{WZ}}{\partial \alpha}, \quad (47)$$

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

$$\frac{\partial \mathbf{WZ}}{\partial \alpha} = \frac{\partial \mathbf{Z}}{\partial \alpha} \mathbf{W}^\top + \mathbf{Z}^\top \frac{\partial \mathbf{W}}{\partial \alpha} \quad (48)$$

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 derivative results in a vector for the Hessian formulation. All other shapes will naturally arise. For solutions with matrix calculus notation, you need to be acquainted with some of its identities.

4.1 $\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \mathbf{A}^\top$

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:

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \quad (49)$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} \sum_{j=1}^n a_{1j} x_j \\ \sum_{j=1}^n a_{2j} x_j \\ \vdots \\ \sum_{j=1}^n a_{mj} x_j \end{bmatrix} \right)^\top \quad (50)$$

$$= \left[\frac{\partial}{\partial \mathbf{x}} \left(\sum_{j=1}^n a_{1j} x_j \right) \quad \frac{\partial}{\partial \mathbf{x}} \left(\sum_{j=1}^n a_{2j} x_j \right) \quad \dots \quad \frac{\partial}{\partial \mathbf{x}} \left(\sum_{j=1}^n a_{mj} x_j \right) \right] \quad (51)$$

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

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1} \left(\sum_{j=1}^n a_{1j} x_j \right) & \frac{\partial}{\partial x_1} \left(\sum_{j=1}^n a_{21j} x_j \right) & \cdots & \frac{\partial}{\partial x_1} \left(\sum_{j=1}^n a_{mj} x_j \right) \\ \frac{\partial}{\partial x_2} \left(\sum_{j=1}^n a_{1j} x_j \right) & \frac{\partial}{\partial x_2} \left(\sum_{j=1}^n a_{21j} x_j \right) & \cdots & \frac{\partial}{\partial x_2} \left(\sum_{j=1}^n a_{mj} x_j \right) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_n} \left(\sum_{j=1}^n a_{1j} x_j \right) & \frac{\partial}{\partial x_n} \left(\sum_{j=1}^n a_{21j} x_j \right) & \cdots & \frac{\partial}{\partial x_n} \left(\sum_{j=1}^n a_{mj} x_j \right) \end{bmatrix} \quad (52)$$

$$= \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{bmatrix} \quad (53)$$

$$\boxed{\frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \mathbf{A}^\top \in \mathbb{C}^{n \times m}} \quad (54)$$

4.2 $\frac{\partial \mathbf{Ax}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{A}^\top$

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

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{z}} = \begin{bmatrix} \frac{\partial}{\partial \mathbf{z}} \sum_{i=1}^n a_{1i} x_i & \frac{\partial}{\partial \mathbf{z}} \sum_{i=1}^n a_{2i} x_i & \cdots & \frac{\partial}{\partial \mathbf{z}} \sum_{i=1}^n a_{mi} x_i \end{bmatrix} \quad (55)$$

$$= \begin{bmatrix} \sum_{i=1}^n a_{1i} \frac{\partial x_i}{\partial \mathbf{z}} & \sum_{i=1}^n a_{2i} \frac{\partial x_i}{\partial \mathbf{z}} & \cdots & \sum_{i=1}^n a_{mi} \frac{\partial x_i}{\partial \mathbf{z}} \end{bmatrix} \quad (56)$$

$$= \underbrace{\begin{bmatrix} \frac{\partial x_1}{\partial \mathbf{z}} & \frac{\partial x_2}{\partial \mathbf{z}} & \cdots & \frac{\partial x_n}{\partial \mathbf{z}} \end{bmatrix}}_{p \times n} \underbrace{\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}}_{n \times m} \quad (57)$$

$$\boxed{\frac{\partial \mathbf{Ax}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{A}^\top \in \mathbb{C}^{p \times m}} \quad (58)$$

Observe that this result is equivalent to applying the chain rule (c.f. Equation (36)), that is,

$$\frac{\partial \mathbf{Ax}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \frac{\partial \mathbf{Ax}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{A}^\top. \quad (59)$$

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

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}^\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,

$$\frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}^\top = \mathbf{a}. \quad (60)$$

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

$$\frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \quad (61)$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^n a_i x_i \right) \quad (62)$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1} (\sum_{i=1}^n a_i x_i) \\ \frac{\partial}{\partial x_2} (\sum_{i=1}^n a_i x_i) \\ \vdots \\ \frac{\partial}{\partial x_n} (\sum_{i=1}^n a_i x_i) \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \quad (63)$$

$$\boxed{\frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \in \mathbb{C}^n} \quad (64)$$

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

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

$$\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^\top \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \quad (65)$$

Nevertheless, here is the step-by-step:

$$\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \right) \quad (66)$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^n x_i a_i \right) \quad (67)$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1} (\sum_{i=1}^n x_i a_i) \\ \frac{\partial}{\partial x_2} (\sum_{i=1}^n x_i a_i) \\ \vdots \\ \frac{\partial}{\partial x_n} (\sum_{i=1}^n x_i a_i) \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \quad (68)$$

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a} \in \mathbb{C}^n} \quad (69)$$

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

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

$$\frac{\partial \mathbf{a}^H \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}^H{}^\top = \mathbf{a}^* \quad (70)$$

Nevertheless, here is the step-by-step:

$$\frac{\partial \mathbf{a}^H \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} a_1^* & a_2^* & \dots & a_n^* \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) = \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^n a_i^* x_i \right) \quad (71)$$

$$(72)$$

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

$$\frac{\partial \mathbf{a}^H \mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1} (\sum_{i=1}^n a_i^* x_i) \\ \frac{\partial}{\partial x_2} (\sum_{i=1}^n a_i^* x_i) \\ \vdots \\ \frac{\partial}{\partial x_n} (\sum_{i=1}^n a_i^* x_i) \end{bmatrix} = \begin{bmatrix} a_1^* \\ a_2^* \\ \vdots \\ a_n^* \end{bmatrix} \quad (73)$$

$$\boxed{\frac{\partial \mathbf{a}^H \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}^* \in \mathbb{C}^n} \quad (74)$$

4.6 $\frac{\partial \mathbf{x}^H \mathbf{a}}{\partial \mathbf{x}} = \mathbf{0}$

Notice that $\mathbf{x}^H \mathbf{a} \neq \mathbf{a}^H \mathbf{x}$. 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

$$\frac{\partial \mathbf{x}^H \mathbf{a}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} x_1^* & x_2^* & \dots & x_n^* \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \right) \quad (75)$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^n x_i^* a_i \right) \quad (76)$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1} (\sum_{i=1}^n x_i^* a_i) \\ \frac{\partial}{\partial x_2} (\sum_{i=1}^n x_i^* a_i) \\ \vdots \\ \frac{\partial}{\partial x_n} (\sum_{i=1}^n x_i^* a_i) \end{bmatrix}. \quad (77)$$

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

$$\frac{\partial \mathbf{x}^H \mathbf{a}}{\partial \mathbf{x}} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (78)$$

$$\boxed{\frac{\partial \mathbf{x}^H \mathbf{a}}{\partial \mathbf{x}} = \mathbf{0} \in \mathbb{C}^n} \quad (79)$$

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

$$4.7 \quad \frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{z}} \mathbf{x}$$

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

$$\frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{z}} = \frac{\partial}{\partial \mathbf{z}} \sum_{i=1}^n y_i x_i \quad (80)$$

$$= \sum_{i=1}^n \frac{\partial y_i x_i}{\partial \mathbf{z}}. \quad (81)$$

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

$$\frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{z}} = \sum_{i=1}^n x_i \frac{\partial y_i}{\partial \mathbf{z}} + \sum_{i=1}^n y_i \frac{\partial x_i}{\partial \mathbf{z}} \quad (82)$$

$$= \begin{bmatrix} \sum_{i=1}^n x_i \frac{\partial y_i}{\partial z_1} \\ \sum_{i=1}^n x_i \frac{\partial y_i}{\partial z_2} \\ \vdots \\ \sum_{i=1}^n x_i \frac{\partial y_i}{\partial z_m} \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^n y_i \frac{\partial x_i}{\partial z_1} \\ \sum_{i=1}^n y_i \frac{\partial x_i}{\partial z_2} \\ \vdots \\ \sum_{i=1}^n y_i \frac{\partial x_i}{\partial z_m} \end{bmatrix} \quad (83)$$

$$= \begin{bmatrix} \frac{\partial y_1}{\partial z_1} & \frac{\partial y_2}{\partial z_1} & \cdots & \frac{\partial y_n}{\partial z_1} \\ \frac{\partial y_1}{\partial z_2} & \frac{\partial y_2}{\partial z_2} & \cdots & \frac{\partial y_n}{\partial z_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial z_m} & \frac{\partial y_2}{\partial z_m} & \cdots & \frac{\partial y_n}{\partial z_m} \end{bmatrix} \mathbf{x} \quad (84)$$

$$+ \begin{bmatrix} \frac{\partial x_1}{\partial z_1} & \frac{\partial x_2}{\partial z_1} & \cdots & \frac{\partial x_n}{\partial z_1} \\ \frac{\partial x_1}{\partial z_2} & \frac{\partial x_2}{\partial z_2} & \cdots & \frac{\partial x_n}{\partial z_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial z_m} & \frac{\partial x_2}{\partial z_m} & \cdots & \frac{\partial x_n}{\partial z_m} \end{bmatrix} \mathbf{y} \quad (85)$$

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{z}} \mathbf{x} \in \mathbb{C}^m} \quad (86)$$

Note that, if either \mathbf{x} or \mathbf{y} does not depend on \mathbf{z} , just disregard $\frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{y}$ or $\frac{\partial \mathbf{y}}{\partial \mathbf{z}} \mathbf{x}$, respectively. When neither depends on \mathbf{z} , the obvious result is the zero vector, $\mathbf{0}$. A simpler way to solve it is to apply the scalar-vector product rule (see Equation (44)), that is,

$$\frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{x}} + \frac{\partial \mathbf{y}}{\partial \mathbf{z}} \frac{\partial \mathbf{x}^\top \mathbf{y}}{\partial \mathbf{y}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{z}} \mathbf{x} \quad (87)$$

$$4.8 \quad \frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}$$

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

$$\frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \sum_{i=1}^n x_i^2 \quad (88)$$

$$= \sum_{i=1}^n \frac{\partial x_i^2}{\partial \mathbf{x}} \quad (89)$$

$$= \begin{bmatrix} \sum_{i=1}^n \frac{\partial x_i^2}{\partial x_1} \\ \sum_{i=1}^n \frac{\partial x_i^2}{\partial x_2} \\ \vdots \\ \sum_{i=1}^n \frac{\partial x_i^2}{\partial x_m} \end{bmatrix} \quad (90)$$

$$= \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{bmatrix} \quad (91)$$

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x} \in \mathbb{C}^n} \quad (92)$$

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

$$4.9 \quad \frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{z}} = 2 \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{x}$$

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

$$\frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{z}} = \frac{\partial}{\partial \mathbf{z}} \sum_{i=1}^n x_i^2 \quad (93)$$

$$= \sum_{i=1}^n \frac{\partial x_i^2}{\partial \mathbf{z}} \quad (94)$$

$$= \begin{bmatrix} \sum_{i=1}^n \frac{\partial x_i^2}{\partial z_1} \\ \sum_{i=1}^n \frac{\partial x_i^2}{\partial z_2} \\ \vdots \\ \sum_{i=1}^n \frac{\partial x_i^2}{\partial z_m} \end{bmatrix} \quad (95)$$

$$= \begin{bmatrix} \sum_{i=1}^n 2x_i \frac{\partial x_i}{\partial z_1} \\ \sum_{i=1}^n 2x_i \frac{\partial x_i}{\partial z_2} \\ \vdots \\ \sum_{i=1}^n 2x_i \frac{\partial x_i}{\partial z_m} \end{bmatrix} \quad (96)$$

$$= 2 \begin{bmatrix} \frac{\partial x_1}{\partial \mathbf{z}} & \frac{\partial x_2}{\partial \mathbf{z}} & \dots & \frac{\partial x_n}{\partial \mathbf{z}} \end{bmatrix} \mathbf{x} \quad (97)$$

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{z}} = 2 \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{x} \in \mathbb{C}^m} \quad (98)$$

Note that this solution could also be solved by the chain rule (c.f. Equation (36)) as follows

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \frac{\partial \mathbf{x}^\top \mathbf{x}}{\partial \mathbf{x}} = 2 \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{x} \in \mathbb{C}^m} \quad (99)$$

$$4.10 \quad \frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A}^\top + \mathbf{A}) \mathbf{x}$$

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

$$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \quad (100)$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} \sum_{i=1}^n x_i a_{i1} & \sum_{i=1}^n x_i a_{i2} & \dots & \sum_{i=1}^n x_i a_{in} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \quad (101)$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \right). \quad (102)$$

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

$$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1} \left(\sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} x_j \right) \\ \frac{\partial}{\partial x_2} \left(\sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} x_j \right) \\ \vdots \\ \frac{\partial}{\partial x_n} \left(\sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} x_j \right) \end{bmatrix} \quad (103)$$

$$= \begin{bmatrix} 2x_1 a_{11} + \sum_{\substack{j=1 \\ j \neq 1}}^n a_{1j} x_j + \sum_{\substack{i=1 \\ i \neq 1}}^n a_{i1} x_i \\ 2x_2 a_{22} + \sum_{\substack{j=1 \\ j \neq 2}}^n a_{2j} x_j + \sum_{\substack{i=1 \\ i \neq 2}}^n a_{i2} x_i \\ \vdots \\ 2x_n a_{nn} + \sum_{\substack{j=1 \\ j \neq n}}^n a_{nj} x_j + \sum_{\substack{i=1 \\ i \neq n}}^n a_{in} x_i \end{bmatrix} \quad (104)$$

$$= \begin{bmatrix} \sum_{j=1}^n a_{1j} x_j \\ \sum_{j=1}^n a_{2j} x_j \\ \vdots \\ \sum_{j=1}^n a_{nj} x_j \end{bmatrix} + \begin{bmatrix} \sum_{\substack{i=1 \\ i \neq 1}}^n a_{i1} x_i \\ \sum_{i=1}^n a_{i2} x_i \\ \vdots \\ \sum_{i=1}^n a_{in} x_i \end{bmatrix} \quad (105)$$

$$= \mathbf{A}^\top \mathbf{x} + \mathbf{A} \mathbf{x} \quad (106)$$

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A}^\top + \mathbf{A}) \mathbf{x} \in \mathbb{C}^n} \quad (107)$$

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

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A} \mathbf{x} \in \mathbb{C}^n} \quad (108)$$

$$4.11 \quad \frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} (\mathbf{A} + \mathbf{A}^\top) \mathbf{x}$$

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

$$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{z}} = \frac{\partial}{\partial \mathbf{z}} \left(\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \quad (109)$$

$$= \frac{\partial}{\partial \mathbf{z}} \left(\begin{bmatrix} \sum_{i=1}^n x_i a_{i1} & \sum_{i=1}^n x_i a_{i2} & \dots & \sum_{i=1}^n x_i a_{in} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \quad (110)$$

$$= \frac{\partial}{\partial \mathbf{z}} \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \right) \quad (111)$$

$$= \begin{bmatrix} \frac{\partial}{\partial z_1} \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \right) \\ \frac{\partial}{\partial z_2} \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \right) \\ \vdots \\ \frac{\partial}{\partial z_n} \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \right) \end{bmatrix} \quad (112)$$

$$= \begin{bmatrix} \sum_{i=1}^n \sum_{j=1}^n a_{ij} \frac{\partial x_i x_j}{\partial z_1} \\ \sum_{i=1}^n \sum_{j=1}^n a_{ij} \frac{\partial x_i x_j}{\partial z_2} \\ \vdots \\ \sum_{i=1}^n \sum_{j=1}^n a_{ij} \frac{\partial x_i x_j}{\partial z_n} \end{bmatrix} \quad (113)$$

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

$$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{z}} = \begin{bmatrix} \sum_{i=1}^n \sum_{j=1}^n x_j a_{ij} \frac{\partial x_i}{\partial z_1} + \sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} \frac{\partial x_j}{\partial z_1} \\ \sum_{i=1}^n \sum_{j=1}^n x_j a_{ij} \frac{\partial x_i}{\partial z_2} + \sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} \frac{\partial x_j}{\partial z_2} \\ \vdots \\ \sum_{i=1}^n \sum_{j=1}^n x_j a_{ij} \frac{\partial x_i}{\partial z_n} + \sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} \frac{\partial x_j}{\partial z_n} \end{bmatrix} \quad (114)$$

$$= \begin{bmatrix} \sum_{i=1}^n \sum_{j=1}^n x_j a_{ij} \frac{\partial x_i}{\partial z_1} \\ \sum_{i=1}^n \sum_{j=1}^n x_j a_{ij} \frac{\partial x_i}{\partial z_2} \\ \vdots \\ \sum_{i=1}^n \sum_{j=1}^n x_j a_{ij} \frac{\partial x_i}{\partial z_n} \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} \frac{\partial x_j}{\partial z_1} \\ \sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} \frac{\partial x_j}{\partial z_2} \\ \vdots \\ \sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} \frac{\partial x_j}{\partial z_n} \end{bmatrix} \quad (115)$$

$$= \begin{bmatrix} \frac{\partial x_1}{\partial \mathbf{z}} & \frac{\partial x_2}{\partial \mathbf{z}} & \cdots & \frac{\partial x_n}{\partial \mathbf{z}} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \quad (116)$$

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

$$= \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{A} \mathbf{x} + \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{A}^\top \mathbf{x} \quad (117)$$

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} (\mathbf{A} + \mathbf{A}^\top) \mathbf{x} \in \mathbb{C}^m} \quad (118)$$

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

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{z}} = 2 \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{A} \mathbf{x} \in \mathbb{C}^m} \quad (119)$$

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

$$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} (\mathbf{A}^\top + \mathbf{A}) \mathbf{x}, \quad (120)$$

where the last equality comes from Equation (107).

$$\mathbf{4.12} \quad \frac{\partial \mathbf{b}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top \mathbf{b}$$

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

$$\frac{\partial \mathbf{b}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left([b_1 \ b_2 \ \dots \ b_m] \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \quad (121)$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} \sum_{i=1}^m a_{i1} b_i & \sum_{i=1}^m a_{i2} b_i & \dots & \sum_{i=1}^m a_{in} b_i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \quad (122)$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^m \sum_{j=1}^n a_{ij} b_i x_j \right) \quad (123)$$

$$= \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_i \frac{\partial x_j}{\partial \mathbf{x}} \quad (124)$$

$$= \begin{bmatrix} \sum_{i=1}^m a_{i1} b_i \\ \sum_{i=1}^m a_{i2} b_i \\ \vdots \\ \sum_{i=1}^m a_{in} b_i \end{bmatrix} \quad (125)$$

$$= \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \quad (126)$$

$$\boxed{\frac{\partial \mathbf{b}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^\top \mathbf{b} \in \mathbb{C}^n} \quad (127)$$

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,

$$\frac{\partial \mathbf{b}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{b}^\top \mathbf{A})^\top = \mathbf{A}^\top \mathbf{b}, \quad (128)$$

where the first equality comes from the Equation (54).

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

Let $\mathbf{x} \in \mathbb{C}^m$, $\mathbf{b} \in \mathbb{C}^n$ and $\mathbf{A} \in \mathbb{C}^{m \times n}$, 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 4.12. Thus,

$$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{b}}{\partial \mathbf{x}} = \frac{\partial \mathbf{b}^\top \mathbf{A}^\top \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{b}^\top \mathbf{A}^\top)^\top = \mathbf{A} \mathbf{b}, \quad (129)$$

where the second equality comes from Equation (54). Nevertheless, here is the step-by-step

$$\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{b}}{\partial \mathbf{x}} = \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} x_1 & x_2 & \dots & x_m \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \right) \quad (130)$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\begin{bmatrix} \sum_{i=1}^m a_{i1} x_i & \sum_{i=1}^m a_{i2} x_i & \dots & \sum_{i=1}^m a_{in} x_i \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \right) \quad (131)$$

$$= \frac{\partial}{\partial \mathbf{x}} \left(\sum_{i=1}^m \sum_{j=1}^n a_{ij} b_j x_i \right) \quad (132)$$

$$= \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_j \frac{\partial x_i}{\partial \mathbf{x}} \quad (133)$$

$$= \begin{bmatrix} \sum_{j=1}^n a_{1j} b_j \\ \sum_{j=1}^n a_{2j} b_j \\ \vdots \\ \sum_{j=1}^n a_{mj} b_j \end{bmatrix} \quad (134)$$

$$= \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \quad (135)$$

$$\boxed{\frac{\partial \mathbf{x}^\top \mathbf{A} \mathbf{b}}{\partial \mathbf{x}} = \mathbf{A} \mathbf{b} \in \mathbb{C}^m} \quad (136)$$

$$4.14 \quad \frac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{A}^\top \mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{z}} \mathbf{A} \mathbf{x}$$

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

$$\frac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{z}} = \frac{\partial}{\partial \mathbf{z}} \left(\begin{bmatrix} y_1 & y_2 & \dots & y_m \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \quad (137)$$

$$= \frac{\partial}{\partial \mathbf{z}} \left(\begin{bmatrix} \sum_{i=1}^m a_{i1} y_i & \sum_{i=1}^m a_{i2} y_i & \dots & \sum_{i=1}^m a_{in} y_i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right) \quad (138)$$

$$= \frac{\partial}{\partial \mathbf{z}} \left(\sum_{i=1}^m \sum_{j=1}^n a_{ij} x_j y_i \right) \quad (139)$$

$$= \sum_{i=1}^m \sum_{j=1}^n a_{ij} \frac{\partial y_i x_j}{\partial \mathbf{z}} \quad (140)$$

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

$$\frac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{z}} = \sum_{i=1}^m \sum_{j=1}^n a_{ij} y_i \frac{\partial x_j}{\partial \mathbf{z}} + \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_j \frac{\partial y_i}{\partial \mathbf{z}} \quad (141)$$

$$= \begin{bmatrix} \frac{\partial x_1}{\partial \mathbf{z}} & \frac{\partial x_2}{\partial \mathbf{z}} & \cdots & \frac{\partial x_n}{\partial \mathbf{z}} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} + \quad (142)$$

$$\begin{bmatrix} \frac{\partial y_1}{\partial \mathbf{z}} & \frac{\partial y_2}{\partial \mathbf{z}} & \cdots & \frac{\partial y_m}{\partial \mathbf{z}} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad (143)$$

$$\boxed{\frac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{A}^\top \mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{z}} \mathbf{A} \mathbf{x}} \quad (144)$$

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{z} . Therefore (c.f. Equation (38)),

$$\frac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{z}} = \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \frac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} + \frac{\partial \mathbf{y}}{\partial \mathbf{z}} \frac{\partial \mathbf{y}^\top \mathbf{A} \mathbf{x}}{\partial \mathbf{y}} \quad (145)$$

$$= \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \mathbf{A}^\top \mathbf{y} + \frac{\partial \mathbf{y}}{\partial \mathbf{z}} \mathbf{A} \mathbf{x}, \quad (146)$$

where the second equality comes from previous solutions.

4.15 $\frac{\partial \text{tr}(\mathbf{A} \mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^\top$

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

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

$$\boxed{\frac{\partial \text{tr}(\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = \mathbf{A}^\top} \quad (147)$$

I have no idea how to make this solution simpler.

$$4.16 \quad \frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = \text{adj}(\mathbf{X})$$

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

$$|\mathbf{X}| = \sum_{i=1}^n x_{ki} |\mathbf{C}_{ki}| = \sum_{i=1}^n x_{ik} |\mathbf{C}_{ik}| \quad \forall k \in \{1, 2, \dots, n\}, \quad (148)$$

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

$$\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = \frac{\partial}{\partial \mathbf{X}} \left(\sum_{i=1}^n x_{ki} |\mathbf{C}_{ki}| \right) \quad \forall k \in \{1, 2, \dots, n\} \quad (149)$$

$$= \frac{\partial}{\partial \mathbf{X}} \begin{pmatrix} \sum_{i=1}^n x_{1i} |\mathbf{C}_{1i}| & \sum_{i=1}^n x_{2i} |\mathbf{C}_{2i}| & \dots & \sum_{i=1}^n x_{ni} |\mathbf{C}_{ni}| \\ \sum_{i=1}^n x_{2i} |\mathbf{C}_{2i}| & \sum_{i=1}^n x_{2i} |\mathbf{C}_{2i}| & \dots & \sum_{i=1}^n x_{2i} |\mathbf{C}_{2i}| \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ni} |\mathbf{C}_{ni}| & \sum_{i=1}^n x_{ni} |\mathbf{C}_{ni}| & \dots & \sum_{i=1}^n x_{ni} |\mathbf{C}_{ni}| \end{pmatrix} \quad (150)$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_{11}} \left(\sum_{i=1}^n x_{1i} |\mathbf{C}_{1i}| \right) & \frac{\partial}{\partial x_{12}} \left(\sum_{i=1}^n x_{1i} |\mathbf{C}_{1i}| \right) & \dots & \frac{\partial}{\partial x_{1n}} \left(\sum_{i=1}^n x_{1i} |\mathbf{C}_{1i}| \right) \\ \frac{\partial}{\partial x_{21}} \left(\sum_{i=1}^n x_{2i} |\mathbf{C}_{2i}| \right) & \frac{\partial}{\partial x_{22}} \left(\sum_{i=1}^n x_{2i} |\mathbf{C}_{2i}| \right) & \dots & \frac{\partial}{\partial x_{2n}} \left(\sum_{i=1}^n x_{2i} |\mathbf{C}_{2i}| \right) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_{n1}} \left(\sum_{i=1}^n x_{ni} |\mathbf{C}_{ni}| \right) & \frac{\partial}{\partial x_{n2}} \left(\sum_{i=1}^n x_{ni} |\mathbf{C}_{ni}| \right) & \dots & \frac{\partial}{\partial x_{nn}} \left(\sum_{i=1}^n x_{ni} |\mathbf{C}_{ni}| \right) \end{bmatrix} \quad (151)$$

$$= \begin{bmatrix} |\mathbf{C}_{11}| & |\mathbf{C}_{12}| & \dots & |\mathbf{C}_{1n}| \\ |\mathbf{C}_{21}| & |\mathbf{C}_{22}| & \dots & |\mathbf{C}_{2n}| \\ \vdots & \vdots & \ddots & \vdots \\ |\mathbf{C}_{n1}| & |\mathbf{C}_{n2}| & \dots & |\mathbf{C}_{nn}| \end{bmatrix} \quad (152)$$

$$\boxed{\frac{\partial |\mathbf{X}|}{\partial \mathbf{X}} = \text{adj}(\mathbf{X})} \quad (153)$$

I have no idea how to make this solution simpler.

$$4.17 \quad \frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -(\mathbf{A}^{-1})^\top \frac{\partial \mathbf{A}}{\partial \alpha} (\mathbf{A}^{-1})^\top$$

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

$$\frac{\partial \mathbf{A}^{-1} \mathbf{A}}{\partial \alpha} = \frac{\partial \mathbf{I}}{\partial \alpha} = \mathbf{0}_{m \times n}, \quad (154)$$

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. Equation (48))

$$\frac{\partial \mathbf{A}^{-1} \mathbf{A}}{\partial \alpha} = \frac{\partial \mathbf{A}}{\partial \alpha} (\mathbf{A}^{-1})^\top + \mathbf{A}^\top \frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = \mathbf{0}_{m \times n} \quad (155)$$

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

$$\boxed{\frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -(\mathbf{A}^{-1})^\top \frac{\partial \mathbf{A}}{\partial \alpha} (\mathbf{A}^{-1})^\top} \quad (156)$$

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] Kaare Brandt Petersen, Michael Syskind Pedersen, et al. “The matrix cookbook”. In: *Technical University of Denmark* 7.15 (2008), p. 510.
- [4] Aarti Singh. *Lecture notes in Introduction to Machine Learning*. 2013-2016. URL: https://www.cs.cmu.edu/~aarti/Class/10315_Spring22/315S22_Rec4.pdf.
- [5] *Matrix calculus - Wikipedia*. (Accessed on 09/22/2022). URL: https://en.wikipedia.org/wiki/Matrix_calculus#Numerator-layout_notation.
- [6] Simon S Haykin et al. *Neural networks and learning machines*. 2009.
- [7] Cesar Augusto Taconeli. *Lecture notes in Análise de Regressão Linear*. 2018. URL: <https://docs.ufpr.br/~taconeli/CE07118/Algebra.pdf>.
- [8] Paul Klein. *Lecture notes of Macroeconomics*. 1999. URL: <http://paulklein.ca/newsite/teaching/calvec.pdf>.