

# Quick-but-comprehensive guide to Matrix Differentiation

Rubem Vasconcelos Pacelli  
rubem.engenharia@gmail.com

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

September 22, 2022

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Notation and nomenclature</b>	<b>2</b>
2.1	Jacobian formulation (numerator layout) . . . . .	2
2.2	Hessian formulation (denominator layout) . . . . .	3
2.3	Comparative between Jacobian and Hessian formulations . . . . .	3
<b>3</b>	<b>Matrix Differentiation</b>	<b>3</b>

## 1 Introduction

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

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

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

Some of the proposals 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](#).

---

<sup>1</sup>Although the expression “row vector” is quite common, I really advocate to avoid it since, once defined a vector as a column,  $\mathbf{y}^T \in \mathbb{C}^{1 \times n}$  is actually a linear transformation from  $\mathbb{R}^n$  to  $\mathbb{R}$ . That is, it does not have anything to do with a vector, which is an entity in a  $n$ -dimensional space. Therefore, throughout this note, I will refer to it as  $1 \times n$  matrix.

## 2 Notation and nomenclature

Let

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

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

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

Nonbold Roman and Greek alphabets represent scalars, while bold uppercases and bold lowercases represent matrices and vectors, respectively. The operators  $\cdot^T$ ,  $\cdot^H$ ,  $\text{tr}(\cdot)$ ,  $\text{adj}(\cdot)$ , and  $|\cdot|$  represent, respectively, the transpose, the hermitian, the trace, the adjoint, and the determinant (or absolute value when the operand is a scalar).

### 2.1 Jacobian formulation (numerator layout)

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

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

Note that it perfectly matches with the Jacobian matrix definition,

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1^T}{\partial \mathbf{x}} \\ \frac{\partial f_2^T}{\partial \mathbf{x}} \\ \vdots \\ \frac{\partial f_m^T}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}, \quad (4)$$

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

## 2.2 Hessian formulation (denominator layout)

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

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

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

## 2.3 Comparative between Jacobian and Hessian formulations

As you could have noticed,

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

# 3 Matrix Differentiation

## References

### Cited references

- [1] Aarti Singh. *Lecture notes in Introduction to Machine Learning*. 2013-2016. URL: [https://www.cs.cmu.edu/~aarti/Class/10315\\_Spring22/315S22\\_Rec4.pdf](https://www.cs.cmu.edu/~aarti/Class/10315_Spring22/315S22_Rec4.pdf).

### Supplementary papers

- [2] Randal J Barnes. “Matrix differentiation”. In: *Springs Journal* (2006), pp. 1–9.
- [3] Kaare Brandt Petersen, Michael Syskind Pedersen, et al. “The matrix cookbook”. In: *Technical University of Denmark 7.15* (2008), p. 510.
- [4] Phoebus J Dhrymes and Phoebus J Dhrymes. *Mathematics for econometrics*. Vol. 984. Springer, 1978.
- [5] *Matrix calculus - Wikipedia*. (Accessed on 09/22/2022). URL: [https://en.wikipedia.org/wiki/Matrix\\_calculus#Numerator-layout\\_notation](https://en.wikipedia.org/wiki/Matrix_calculus#Numerator-layout_notation).