Matrix Calculus

Tor Aksel N. Heirung

January 23, 2017

Abstract

This note explains differentiation of some functions expressed in terms of vectors and matrices. The choice of formulas is based on what is most relevant for the course TTK4135 Optimization and Control, but the explanations are general and should enable derivation of other differentiation formulas.

Contents

1 The Definition of the Gradient

For a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ (see footnote[1] if you are not familiar with this notation), the *row vector* $\partial f/\partial x$ is defined by

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} \tag{1}$$

The *gradient vector* $\nabla f(x)$ is

$$\nabla f(x) = \left[\frac{\partial f}{\partial x}\right]^\top = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \tag{2}$$

It should be noted that the gradient vector is defined as a row vector in some fields (and textbooks). This is a matter of definition, and has no fundamental meaning. All the derivatives we derive in this note use the definition (2).

For a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$, the *Jacobian matrix* $[\partial f/\partial x]$ is an $m \times n$ matrix whose element in the ith row and the jth column is $\partial f_i/\partial x_j$, that is,

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{3}$$

Note that (1) is a special case of (3).

2 Linear Functions

2.1 Linear Scalar Function

We first consider linear scalar functions $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(x) = c^\top x \tag{4}$$

It is fairly simple to derive the gradient of this function using the definition (2). First, we write out the function f in order to clarify the subsequent steps:

$$f(x) = c^\top x = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = c_1x_1 + c_2x_2 + \cdots + c_nx_n \tag{5}$$

Now, the gradient can be calculated as

$$\nabla f(x) = \left[\frac{\partial f}{\partial x}\right]^\top = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = c \tag{6}$$

[1] $f : \mathbb{R}^n \to \mathbb{R}^m$ means that f is a vector of the m functions $f_1, \ldots, f_m$, and that all of them have $x \in \mathbb{R}^n$ as argument. Hence, $f : \mathbb{R}^n \to \mathbb{R}$ means that f is a scalar function of n variables.

Since the function f is scalar, it follows that

$$f(x) = c^\top x = (c^\top x)^\top = x^\top c \tag{7}$$

Hence, we have the two differentiation formulas

$$\nabla(c^\top x) = c \tag{8a}$$

$$\nabla(x^\top c) = c \tag{8b}$$

2.2 Linear Vector Functions

By vector function, we mean "a vector of functions". An example is

$$f(x) = Ax = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix} = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} \tag{9}$$

where f is a vector containing the two functions f_1 and f_2. This also means that $f : \mathbb{R}^2 \to \mathbb{R}^2$. This type of function should be familiar from linear dynamic systems, where systems of differential equations are commonly written $\dot{x} = Ax$.

We can find the Jacobian of f in (9) using the definition (3);

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = A \tag{10}$$

This of course holds for any A, so that

$$\frac{\partial}{\partial x}(Ax) = A \tag{11}$$

in the general case.

3 Quadratic Forms

The most common quadratic form encountered in this course is

$$f(x) = \frac{1}{2}x^\top Gx \tag{12}$$

However, before differentiating this function we will look at two other derivatives. First, consider

$$f(x, y) = \frac{1}{2}x^\top Gy, \quad x,\, y \in \mathbb{R}^n \tag{13}$$

Keep in mind that Gy is a (column) vector, and that f is a scalar function. The gradient of $f(x, y)$ with respect to x is then

$$\nabla_x f(x, y) = \frac{1}{2}Gy \tag{14}$$

This follows immediately from (8b). Since $x^\top G$ is a (row) vector, we can in a similar manner use (8a) to establish that the gradient of $f(x, y)$ with respect to y is

$$\nabla_y f(x, y) = \frac{1}{2}(x^\top G)^\top = \frac{1}{2}G^\top x \tag{15}$$

We are now ready to find the gradient of

$$f(x) = \frac{1}{2} x^\top G x \tag{12}$$

Keeping (14) and (15) in mind and using a variant of the product rule, we differentiate first with respect to the "first x" and then with respect to the "second x". The gradient becomes

$$\nabla f(x) = \frac{1}{2} G x + \frac{1}{2} G^\top x \tag{16}$$

Since (16) is not exactly intuitive on its own, consider the 2×2 case

$$\begin{aligned} f(x) &= \frac{1}{2} x^\top G x, \quad G = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \frac{1}{2} \big(g_{11} x_1^2 + (g_{12} + g_{21}) x_1 x_2 + g_{22} x_2^2 \big) \end{aligned} \tag{17}$$

Going directly from the definition (2), we have that the gradient is

$$\begin{aligned} \nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} &= \frac{1}{2} \begin{bmatrix} 2 g_{11} x_1 + g_{12} x_2 + g_{21} x_2 \\ g_{12} x_1 + g_{21} x_1 + 2 g_{22} x_2 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} g_{11} x_1 + g_{12} x_2 \\ g_{21} x_1 + g_{22} x_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} g_{11} x_1 + g_{21} x_2 \\ g_{12} x_1 + g_{22} x_2 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} g_{11} & g_{21} \\ g_{12} & g_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \frac{1}{2} G x + \frac{1}{2} G^\top x \end{aligned} \tag{18}$$

A special case of (12), which is both common and important, is that of $G = G^\top$ (symmetric G). The gradient then simplifies to

$$\nabla f(x) = G x, \quad G = G^\top \tag{19}$$

4 Summary

The following formulas are the most important ones from the sections above.

If $f : \mathbb{R}^n \to \mathbb{R}$, then

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} \tag{1}$$

and

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x} \end{bmatrix}^\top = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \tag{2}$$

For the linear function $f(x) = c^\top x = x^\top c$, we have

$$\nabla(c^\top x) = c \tag{8a}$$

$$\nabla(x^\top c) = c \tag{8b}$$

The Jacobian of the vector function $f(x) = Ax$ is

$$\frac{\partial}{\partial x}(Ax) = A \tag{11}$$

The gradient of the quadratic function $f(x) = \frac{1}{2}x^\top Gx$ is

$$\nabla f(x) = \frac{1}{2}Gx + \frac{1}{2}G^\top x \tag{16}$$

or, when G is a symmetric matrix:

$$\nabla f(x) = Gx, \quad G = G^\top \tag{19}$$

5 Request for Comments

Please email andreas.flaten@itk.ntnu.no if you feel that:

- this note is confusing,
- something should be explained better,
- something is missing,
- something is wrong, or
- you have a suggestion for improvement.

All feedback is appreciated.