

Understanding the Directional Derivative and the Gradient

Irene Markelic

November 17, 2024

1 Introduction

Understanding how functions change in different directions is crucial in many fields. For example in the context of neural networks where gradients are used to update weights during the training process. In this article, we will explore the concepts of the directional derivative and the gradient, and we will demonstrate their relationship through the equation:

$$g'(0) = \nabla_u f(\mathbf{x}) = \nabla f(\mathbf{x}) \mathbf{u} \quad (1)$$

This article is best understood if you have some basic knowledge of dot products, derivatives, Leibniz and Lagrange notation, partial derivatives, and the chain rule.

TL;DR: jump to 6.

2 What are the Gradient and the Directional Derivative?

To begin with, let's look at the derivative of a univariate function. It's defined as:

$$\frac{df}{dx} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \quad (2)$$

(Here we have used Leibniz notation for denoting the derivative, $\frac{df}{dx}$. This is simply a different notation for the Lagrange notation $f'(x)$, which many are familiar with from school.) The derivative of a univariate function $f(x)$ is its slope, defined as the rise (vertical change) over the run (horizontal change), with the latter being infinitesimally small. It indicates how the function changes at a given point when we increase its parameter by an infinitesimal amount. This "change in the function" means we learn how steep the function is at that point and whether it is increasing or decreasing.

2.1 The Gradient

The **gradient** is a generalization of the derivative for the case of a scalar-valued function, $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ (many values in, one value out). It's defined as:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}. \quad (3)$$

Thus, the gradient is a vector of all partial derivatives. The symbol ∇ (the Greek letter nabla, read "del") can be interpreted as an operator that is applied to the function. Each element of the gradient vector gives the slope of the function with respect to the corresponding variable: $\frac{\partial f}{\partial x_1}$ gives the slope of the function with respect to x_1 treating all other variables (x_2, \dots, x_n) as constants. In fact, **the gradient points to the direction of steepest ascent**. We will look at a proof for this shortly in section 4.

2.2 The Directional Derivative

The notation for the **directional derivative** is $\nabla_{\mathbf{u}} f(\mathbf{x})$, or $D_{\mathbf{u}} f(\mathbf{x})$, but here we will use the former. Its definition is given by:

$$\nabla_{\mathbf{u}} f(\mathbf{x}) = \lim_{s \rightarrow 0} \frac{f(\mathbf{x} + \mathbf{u}s) - f(\mathbf{x})}{s}, \quad (4)$$

Here, \mathbf{u} is a unit vector, meaning its magnitude (or length) is 1. (Recall that the magnitude of a vector is the square root of the sum of the squares of its individual components, and it is denoted by two vertical bars around the vector, although in some references just one pair of vertical bars is used.

$$\|\mathbf{a}\| = \sqrt{a_1^2 + \dots + a_n^2} \quad (5)$$

with $\mathbf{a} \in \mathbb{R}^n$.) You can see that the definition of the directional derivative is very similar to the univariate case. (I deliberately chose to use the variable s instead of h to make the distinction from the univariate case clearer.) However, in the case of the directional derivative, the input to f is $\mathbf{x} + s\mathbf{u}$, which represents a line equation. You start from the vector \mathbf{x} and you move s units in the direction of \mathbf{u} . Thus, the above expression examines how the function changes as we move an infinitesimal amount into the direction of \mathbf{u} . In other words: the directional derivative represents the rate of change of the function at a given point when the point is moved a tiny bit (an infinitesimal amount) in the direction of \mathbf{u} . As mentioned, \mathbf{u} serves as a directional vector, and this is reflected in the notation $\nabla_{\mathbf{u}} f(\mathbf{x})$, where \mathbf{u} appears as a subscript.

- a specific point
- function value of specific point
- direction vector \mathbf{u}
- line connecting function value of specific point to function value of \mathbf{u}

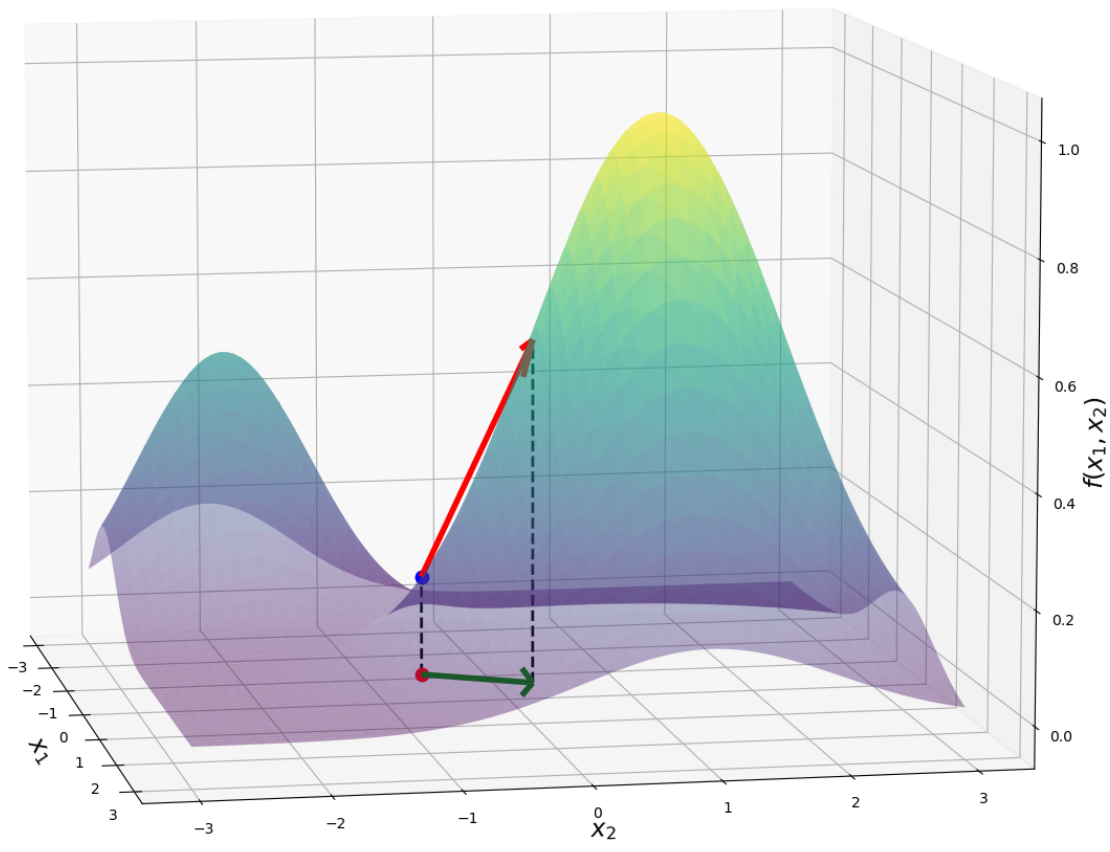


Figure 1: A specific 2d point (red dot) is the input to a function. How does the function value change when we travel along the directional vector ($s\mathbf{u}$ indicated by the green line?). The directional derivative is the rate of change of f when we travel an infinitesimal bit ($s \rightarrow 0$) in the direction of \mathbf{u} .

For a better understanding of the above let's have a look at figure 1. We see a two-dimensional function and the red dot in the plane is a random but specific point. Its function value is marked by the blue dot. Now, in the case of a univariate function, we have only one single variable, e.g. x that is input to the function and we can only increase or decrease it. When computing the derivative of a univariate function we imagine computing the rise over the run when the change in x is infinitesimally small. But in the case of functions that take more than one input variable, e.g. x_1 and x_2 , we can change each of these. Maybe we want to go one step in direction x_1 and two steps in direction x_2 . So, if we want to make a statement about the change of the value of the function, we must first specify in which direction we want to move. This is shown in the figure as the green vector. It corresponds to the directional vector \mathbf{u} , only that it is not normalized, so actually the green vector can be interpreted as $s\mathbf{u}$. The scalar s tells how far we want to move in the given direction, e.g. 2 times the length and direction of \mathbf{u} , hence you can also interpret s as a scaling factor. Now, that we have set a direction the entire idea is analogous to the concept of the derivative of the univariate case. The directional derivative tells us how the function value changes when we move a tiny bit in the given direction. Imagine the red vector is very, very small, (because s is approaching 0), then it corresponds to the tangent line of the function at the specific point. The slope of this tangent line represents the directional derivative.

To make this more concrete, imagine you are standing blindfolded on a mountain (the function) and your goal is to avoid falling off a cliff. You could use one leg or a stick and carefully feel around you, testing the terrain to determine in which direction (the directional vector \mathbf{u}) it is safe to move.

2.3 The Relationship between Gradient and Directional Derivative

Let's quickly look at the differences between the gradient and the directional derivative. While the gradient is a vector that indicates the direction in which the rate of change is maximal, the directional derivative is a scalar that tells us the rate of change of the function when moving in a specific direction. If that direction happens to be the direction of steepest ascent, the directional derivative coincides with the magnitude of the gradient, as we will demonstrate shortly in section 4. Also, the directional derivative equals the dot product of the gradient and the directional vector, which we will also show in section 4. The differences and relationship between the gradient and the directional derivative are summarized in Table 1.

Table 1: Comparison between gradient and directional derivative

Gradient	Directional Derivative
$\nabla f(\mathbf{x})$	$\nabla_{\mathbf{u}} f(\mathbf{x})$
is a vector, e.g. $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$	is a scalar, e.g. 4
indicates direction of steepest ascent	indicates rate of change in direction specified by \mathbf{u}
The directional derivative is the dot product of the gradient and the directional vector $\nabla_{\mathbf{u}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \mathbf{u}$	
if \mathbf{u} is direction of steepest ascent, then the directional derivative equals the magnitude of the gradient, e.g. $\nabla f(\mathbf{x}) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ then $\nabla_{\mathbf{u}} f(\mathbf{x}) = \sqrt{1^2+2^2}$	

3 How to compute the directional derivative?

In the following we will show - and this is the main point of this article - that:

$$g'(0) = \nabla_{\mathbf{u}} f(\mathbf{x}) = \nabla f(\mathbf{x}) \mathbf{u} \tag{6}$$

But let's start slowly.

3.1 Limit Definition and Derivative

We have now stated many times (repetitions were intended) that the directional derivative is the rate of change of the function at a point \mathbf{x} when moving an infinitesimal amount into the direction of \mathbf{u} . We have shown how the directional derivative was defined as a limit in equation 4. The limit interpretation means that we take two points on the surface of f (as indicated in Figure 1). One point is moved closer and closer to the point of interest and based on that we learn the rate of change of that point. Of course, for a multivariable function this only makes sense if we move the point along the line given by $s\mathbf{u}$.

But we can also interpret this "rate of change of the function at a point when moving an infinitesimal amount into the direction of \mathbf{u} " differently: Since \mathbf{x} and \mathbf{u} are fixed, only s is a free variable. Therefore, we can interpret $\mathbf{x} + s\mathbf{u}$ as a line equation and f as a function that outputs a function value for each value of s . Have a look at Figure 2. Here we see a zoomed-in section of the previous plot. In addition, I marked a few points along the line $s\mathbf{u}$ and I marked the corresponding function values as dots, too. So you can clearly see that we have a function of a line, parametrized by s . The derivative of this function is - per definition - the rate of change of this function at each point. Take note: the rate of change of the point where $s = 0$ is the "rate of change of the function at a point when moving

an infinitesimal amount in the direction of \mathbf{u} ". Therefore, it is the same as the directional derivative.

This was a very wordy explanation. Next, we do the same reasoning but with math.

3.2 Part 1: Showing that $g'(0) = \nabla_{\mathbf{u}}f(\mathbf{x})$

I have previously stated "we can interpret $\mathbf{x} + s\mathbf{u}$ as a line equation and f as a function that outputs a function value for each value of s ." To make this easier to understand, we give this alternative interpretation a new name, so that we can distinguish it from the function $f(\mathbf{x})$. Therefore, we define:

$$g(s) = f(\mathbf{x} + s\mathbf{u}) \quad (7)$$

I've also previously stated that the derivative of this function at $s = 0$ is the same as the directional derivative. Thus, what we need to show is:

$$g'(0) = \nabla_{\mathbf{u}}f(\mathbf{x}) \quad (8)$$

To show that this statement is true, we take the derivative of our new function g with respect to s according to the definition of the derivative of a univariate function.

$$g'(s) = \lim_{h \rightarrow 0} \frac{g(s+h) - g(s)}{h} \quad (9)$$

Next, we plug in 0 for s , because we are looking for $g'(0)$.

$$g'(0) = \lim_{h \rightarrow 0} \frac{g(0+h) - g(0)}{h} = \lim_{h \rightarrow 0} \frac{g(h) - g(0)}{h} \quad (10)$$

Now we substitute the definition of g (i.e. $g(s) = f(\mathbf{x} + s\mathbf{u})$). We must be careful not to confuse the symbols here; g is a function that takes the parameter s . When we write $g(h)$, it means that the parameter s takes the specific value of h , thus $g(s = h)$. Therefore, we can substitute h for s in the definition: $g(s = h) = f(\mathbf{x} + h\mathbf{u})$. Similarly, for $g(0) = f(\mathbf{x} + 0\mathbf{u}) = f(\mathbf{x})$. Thus, we have:

$$g'(0) = \lim_{h \rightarrow 0} \frac{g(0+h) - g(0)}{h} = \lim_{h \rightarrow 0} \frac{g(h) - g(0)}{h} = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{u}) - f(\mathbf{x})}{h} \quad (11)$$

This expression is just the definition of the directional derivative, with the difference that the variable that we take the limit of is called h instead of s . However, the variable name can be anything. We can rename it to s , leading us to:

$$g'(0) = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{u}) - f(\mathbf{x})}{h} = \lim_{s \rightarrow 0} \frac{f(\mathbf{x} + s\mathbf{u}) - f(\mathbf{x})}{s} = \nabla_{\mathbf{u}}f(\mathbf{x}) \quad (12)$$

Which is what we wanted to show:

$$g'(0) = \nabla_{\mathbf{u}}f(\mathbf{x}) \quad (13)$$

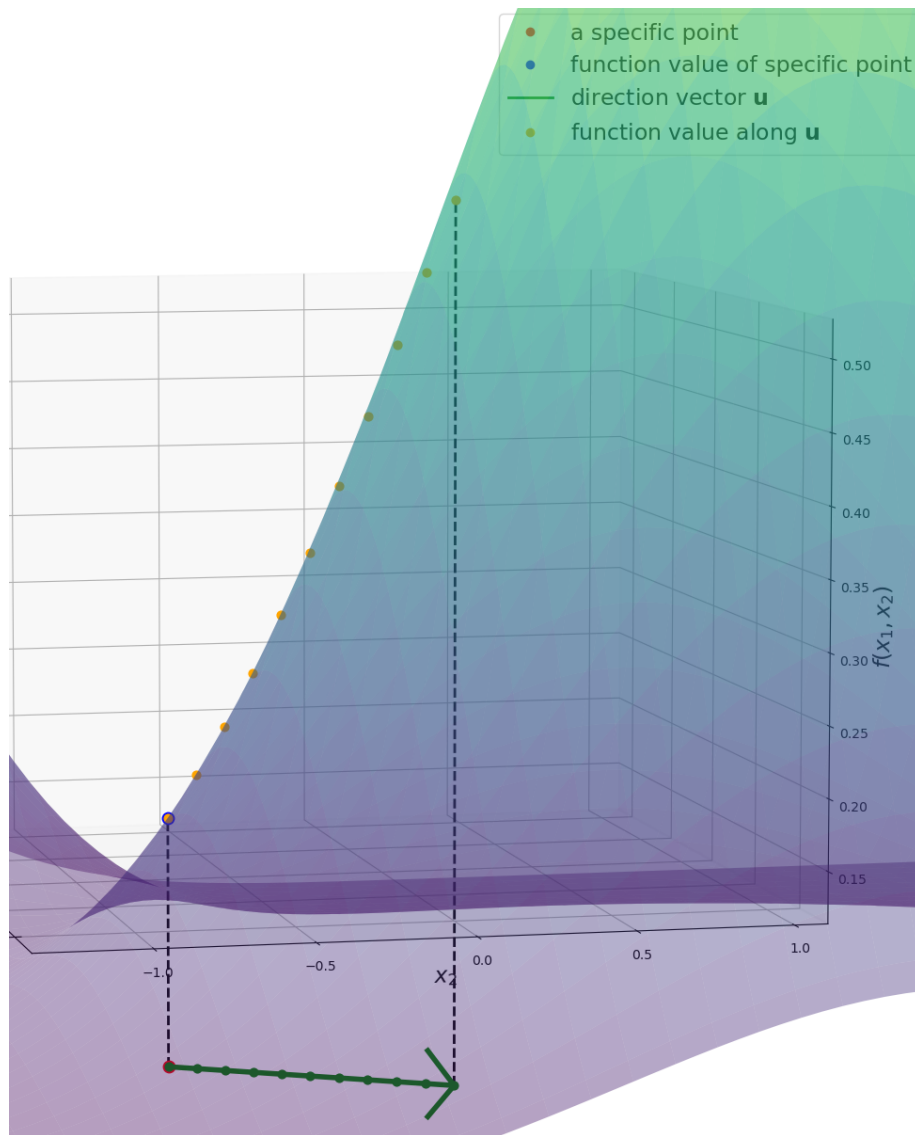


Figure 2: Showing how the points along the line given by $\mathbf{x} + s\mathbf{u}$ (green points) are mapped to $f(\mathbf{x} + s\mathbf{u})$ (orange points). This can be interpreted as a function of s with fixed \mathbf{x} and \mathbf{u} . We define $g(s) = f(\mathbf{x} + s\mathbf{u})$. The rate of change (the derivative) of $g(s)$ at $s = 0$, the orange dot with the black outline, is the directional derivative of the function f at point \mathbf{x} and the directional vector \mathbf{u} .

3.3 Part 2: Showing that $\nabla_u f(\mathbf{x}) = \nabla f(\mathbf{x})\mathbf{u}$

Now, we will show that the second part of the statement is true, i.e. $\nabla_u f(\mathbf{x}) = \nabla f(\mathbf{x})\mathbf{u}$. In other words, the directional derivative is the dot product of the gradient and the directional vector. We will reuse our new function $g(s)$. What we want to show is:

$$\nabla_u f(\mathbf{x}) = \nabla f(\mathbf{x})\mathbf{u} \quad (14)$$

Let's do it:

$$g(s) = f(\mathbf{x} + s\mathbf{u}) \quad (15)$$

$$g'(s) = \frac{d}{d(\mathbf{x} + s\mathbf{u})} f(\mathbf{x} + s\mathbf{u}) \frac{d}{ds} (\mathbf{x} + s\mathbf{u}) \quad (16)$$

$$g'(s) = \frac{d}{d(\mathbf{x} + s\mathbf{u})} f(\mathbf{x} + s\mathbf{u})\mathbf{u} \quad (17)$$

$$g'(0) = \frac{d}{d(\mathbf{x} + 0\mathbf{u})} f(\mathbf{x} + 0\mathbf{u})\mathbf{u} \quad (18)$$

$$g'(0) = \frac{d}{d(\mathbf{x})} f(\mathbf{x})\mathbf{u} \quad (19)$$

$$g'(0) = \nabla f(\mathbf{x})\mathbf{u} \quad (20)$$

$$g'(0) = \nabla_u f(\mathbf{x}) = \nabla f(\mathbf{x})\mathbf{u} \quad (21)$$

In line 16, we apply the chain rule. In line 17, we realize that the derivative of $\mathbf{x} + s\mathbf{u}$ w.r.t. (with respect to) s is simply \mathbf{u} . However, we need to leave the outer derivative in its symbolic form, since we do not know the specifics of f . Since we want to know the rate of change at $s = 0$, we plug this into our equations in line 18. In line 19, we just evaluate the results of the equations when $s = 0$. However, there is an error in line 18! Take note of the following: I write $\frac{d}{d(\mathbf{x})} f(\mathbf{x})$. The input to f is a vector, and the derivative of a vector is its gradient, and this has a different notation; I correct this in line 20. In the final line 21, we exploit what we have shown in subsection 3.2, namely that $g'(0) = \nabla_u f(\mathbf{x})$. And that's it! That's what we wanted to show.

4 Why is the gradient the direction of steepest ascent?

Previously, we stated that (a) the gradient points in the direction of steepest ascent and (b) the directional derivative equals the magnitude of the gradient when the direction it points to is the direction of steepest ascent. Now, we will explore why this is the case. Recall that the directional derivative represents the rate of change at a given point when the point is moved in the direction of \mathbf{u} . The rate of change will be highest when the direction we move in is the steepest. Therefore, the \mathbf{u} that yields the maximal value for

the directional derivative is the direction of steepest ascent. We will demonstrate that this occurs when \mathbf{u} points in the same direction as the gradient.

We have just shown that $\nabla_u f(\mathbf{x}) = \nabla f(\mathbf{x})\mathbf{u}$. Thus, the directional derivative is dot product of the gradient and the directional vector. The dot product is defined as the product of the lengths of the vectors times the cosine of the angle between them:

$$\mathbf{a}\mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\|\cos(\theta) \quad (22)$$

Where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, and θ denotes the angle between the two vectors \mathbf{a} and \mathbf{b} , and the vertical bars denote the vector lengths.

Since \mathbf{u} is a unit vector, as we stated in the beginning of this text, its length is 1. Thus, we have:

$$\nabla_u f(\mathbf{x}) = \nabla f(\mathbf{x})\mathbf{u} = |\nabla f(\mathbf{x})|1\cos(\theta) \quad (23)$$

So when is this expression maximal? The cosine function always returns a value between -1 and 1, meaning the above expression is maximal when cosine equals 1. In this case, we have

$$\nabla_u f(\mathbf{x}) = \nabla f(\mathbf{x})\mathbf{u} = |\nabla f(\mathbf{x})| \quad (24)$$

Again, this statement holds true only when the cosine of the angle between the two vectors $\nabla f(\mathbf{x})$ and \mathbf{u} is 1. This occurs when the angle is 0, indicating that both vectors point in the same direction. Since this is the direction that maximizes the expression, it is the direction of steepest ascent.

Thus, the directional derivative equals the magnitude of the gradient when it points in the direction of steepest ascent. This is simply a mathematical consequence of the fact that, in this case, the cosine of θ is maximal, leading to a maximal value for the expression.

This is also why the gradient points in the direction of steepest ascent; simply because this is exactly the case when the cosine of theta is 1 which does not change the value of the product.

5 Examples

Let's conclude this with two simple examples.

5.1 Example 1, $\nabla_u f(\mathbf{x}) = \nabla f(\mathbf{x})\mathbf{u}$

In which direction should you move if you are at position (4,5) on the curve $f(\mathbf{x}) = x_1^2 + x_2^2$ to ascend as quickly as possible? Additionally, what is the rate of change at the given point in the direction of steepest ascent?

Solution:

The direction of steepest ascent is given by the gradient, which is the collection of the

partial derivatives of the function. Thus, all we have to do here is to compute these partial derivatives:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{df}{dx_1} \\ \frac{df}{dx_2} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} \quad (25)$$

This gradient represents the direction of steepest descent for any point on the curve. We are at the position $x_1 = 4$ and $x_2 = 5$, so that we can substitute these values into the gradient:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2(4) \\ 2(5) \end{bmatrix} = \begin{bmatrix} 8 \\ 10 \end{bmatrix} \quad (26)$$

Thus, the direction of steepest ascent at the point (4,5) is given by the vector $\begin{bmatrix} 8 \\ 10 \end{bmatrix}$ and the rate of change in that direction is the magnitude of the gradient, $\sqrt{8^2 + 10^2}$.

5.2 Example 2, $g'(0) = \nabla_{\mathbf{u}} f(\mathbf{x})$

This is Example 2.3 from the book [1] on p.24: "We wish to compute the directional derivative of

$$f(\mathbf{x} = x_1 x_2) \quad (27)$$

at $\mathbf{x} = [1, 0]$ in the direction $\mathbf{u} = [-1, -1]^T$.

Solution:

We could compute the gradient and then multiply it by the directional vector. However, we can also solve it using the new function g as described above. For that we define the function

$$g(s) = f(\mathbf{x} + s\mathbf{u}) = f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + s \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}\right) \quad (28)$$

Since $f(\mathbf{x} = x_1 x_2)$ we have:

$$g(s) = (x_1 + s u_1)(x_2 + s u_2) \quad (29)$$

Plugging in the values for \mathbf{x} and \mathbf{u} gives:

$$g(s) = (1 + s(-1))(0 + s(-1)) = (1 - s)(-s) = s^2 - s \quad (30)$$

Now, we want the derivative of g at position $s=0$:

$$g'(s) = 2s - 1 \quad (31)$$

$$g'(0) = 2(0) - 1 = -1 \quad (32)$$

Thus, the directional derivative of $f(\mathbf{x}) = x_1 x_2$ at $\mathbf{x} = [1, 0]$ in the direction $\mathbf{u} = [-1, -1]$ is -1.

¹In the book the directional vector is called s , but here I called it u to be consistent with the previous notation.

6 Things to remember

- The directional derivative is a scalar (a number). It is the rate of change of a function at a given point that moves in a certain direction. It is denoted as $\nabla_u f(\mathbf{x})$.
- The gradient of a scalar function, $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$, is a vector that consists of the partial derivatives of the function. It indicates the direction of steepest ascent, and its magnitude represents the directional derivative in that direction.
- We can compute the directional derivative in two ways: a) by introducing the function $g(s) = f(\mathbf{x} + s\mathbf{u})$, where the directional derivative equals $g'(0)$ and b) (more important probably) by using the fact that **the directional derivative equals the dot product of the gradient and the directional vector:** $\nabla_u f(\mathbf{x}) = \nabla f(\mathbf{x})\mathbf{u}$
- Knowing all this might save your life when you find yourself blindfolded on a mountain.

References

- [1] Mykel Kochenderfer and Tim Wheeler. *Algorithms for Optimization*. The MIT Press, 2019. ISBN: 0262039427.
- [2] James Stewart. *Calculus: Early Transcendentals*. 8th. Cengage Learning, 2015. ISBN: 978-1-305-25380-2.