

## Derivatives

Suppose we have a function of several variables  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  — a function with possibly multiple inputs and multiple outputs. What would it mean to be the **derivative** of such a function?

We've seen that the partial derivatives of a function  $z = f(x, y)$  give the rates of change of  $f$  in the  $x$ - and  $y$ -directions. This is a partial answer to our question. However, what do we do if we have a function like

$$(u, v) = f(x, y, z) = (2xy - 3z, yz - 3xz)?$$

A naive approach might be to write the function in component form:

$$u = 2xy - 3z, \quad v = yz - 3xz.$$

Now we can take partial derivatives:

$$\begin{aligned} \frac{\partial u}{\partial x} &= 2y, & \frac{\partial u}{\partial y} &= 2x, & \frac{\partial u}{\partial z} &= -3, \\ \frac{\partial v}{\partial x} &= -3z, & \frac{\partial v}{\partial y} &= z, & \frac{\partial v}{\partial z} &= y - 3x. \end{aligned}$$

Well, we have 6 partial derivatives — do we want to call all of these (collectively) the “derivative” of  $f$ ? This seems somewhat *ad hoc* and unsatisfying.

You might think of arranging these partial derivatives in a matrix:

$$\begin{bmatrix} \frac{\partial u}{\partial x} = 2y & \frac{\partial u}{\partial y} = 2x & \frac{\partial u}{\partial z} = -3 \\ \frac{\partial v}{\partial x} = -3z & \frac{\partial v}{\partial y} = z & \frac{\partial v}{\partial z} = y - 3x \end{bmatrix}.$$

Now we have a single entity, and it is reasonable to say that it contains information about the rate(s) of change of  $f$ . However, this still seems a bit arbitrary — for example, why put the  $u$  derivatives in the first row (as opposed to, say, the first column)?

Actually, the matrix we've written down above *will* turn out to be *the* derivative of  $f$ . In order to give a rationale for this, we have to think about derivatives in a conceptual way.

You probably encountered the derivative of a function  $y = f(x)$  as the slope of the tangent line to a curve. What *is* the tangent line, in a conceptual sense? It's the best “flat” approximation to the curve near the point of tangency. That is, near the point of tangency, the tangent and the curve are very close.

We could generalize this to a function  $z = f(x, y)$  by considering the tangent *plane* to the graph, which is a surface in three dimensions. And though it's difficult to picture, we can *imagine* continuing this procedure to higher dimensions. A line in the plane looks like this:

$$ax + by = c.$$

A plane in space looks like this:

$$ax + by + cz = d.$$

With more input variables for  $f$ , we just keep adding more variables to the equation.

These equations are **linear**, which means roughly that the variables occur only to the first power, multiplied by constants. Equations of this kind describe things that are “flat”.

With multiple inputs and multiple outputs, an equation for a “flat” thing involves a *matrix*. For example, here's a **linear function** with 3 inputs and 2 outputs:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & 3 & -8 \\ 4 & 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Written out, this would be

$$u = x + 3y - 8z, \quad v = 4x + 3z.$$

Here's the general definition.

**Definition.** A function  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is a **linear transformation** if:

- (a)  $T(x + y) = T(x) + T(y)$  for all  $x, y \in \mathbb{R}^n$ .
- (b)  $T(c \cdot x) = c \cdot T(x)$  for all  $c \in \mathbb{R}$  and  $x \in \mathbb{R}^n$ .

You've seen things which have these properties — derivatives and antiderivatives, for example.

A linear transformation is a function that behaves like the linear functions you know:  $y = mx$  or  $z = ax + by$ , for instance. The derivative of a function will be the linear transformation that best approximates the function. Here's the definition.

**Definition.** Let  $f : U \rightarrow \mathbb{R}^m$  be a function, where  $U$  is an open set in  $\mathbb{R}^n$ , and let  $x \in U$ . The **derivative** of  $f$  at  $x$  is a linear transformation  $Df : \mathbb{R}^n \rightarrow \mathbb{R}^m$  which satisfies

$$\lim_{\|h\| \rightarrow 0} \frac{\|f(x+h) - f(x) - Df(h)\|}{\|h\|} = 0.$$

People sometimes write “ $Df(x)(h)$ ” or “ $Df_x(h)$ ” to indicate the dependence of  $Df$  on  $x$ . It is also common to see “ $df$ ” instead of “ $Df$ ”.

Notice that this resembles our old difference quotient definition for derivatives:

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

Why do we have  $\|\cdot\|$ 's in the equation? The idea is that  $h$ ,  $f(x+h)$ , and  $f(x)$  are elements of  $\mathbb{R}^m$  — that is,  $m$ -dimensional *vectors*.  $h$  is really  $(h_1, \dots, h_n)$ . And  $f$  is a function which produces  $m$  numbers as its output, so  $f(x+h)$ , and  $f(x)$  look like  $(y_1, \dots, y_m)$ . We can't take the quotient of two vectors, but we *can* take the quotient of their lengths.

As with the difference quotient definition for the derivative  $f'(x)$  of a function  $y = f(x)$ , the definition we've given is not so great for actually computing derivatives. The following result makes it a lot easier.

**Theorem.** Suppose  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is differentiable at  $c \in \mathbb{R}^n$ . Write  $x = (x_1, x_2, \dots, x_n)$  and  $f = (f_1, f_2, \dots, f_m)$ , and let  $h \in \mathbb{R}^n$ . Then

$$Df(h) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \cdot h.$$

In other words,  $Df$  is represented by the matrix of partial derivatives of the components of  $f$  with respect to the input variables.  $\square$

The converse isn't true. If the partial derivatives of a function exist, the function might fail to be differentiable (in the sense that the limit above might be undefined). On the other hand, we have the following result.

**Theorem.** Suppose  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Write  $x = (x_1, x_2, \dots, x_n)$  and  $f = (f_1, f_2, \dots, f_m)$ . If the partial derivatives  $\frac{\partial f_i}{\partial x_j}$  at  $c \in \mathbb{R}^n$  are continuous for  $i = 1, \dots, m$  and  $j = 1, \dots, n$ , then  $f$  is differentiable at  $c$ .  $\square$

In this case, by the preceding theorem,  $Df$  is represented by the matrix of partial derivatives.

**Remark.** A function with continuous partial derivatives is called **continuously differentiable**.

---

**Example.** Compute  $Df$  for

$$(u, v) = f(x, y, z) = (7xy + z^2, x + 3y - z^4).$$

Since  $f$  has 3 inputs and 2 outputs, the matrix for  $Df$  will be  $2 \times 3$ :

$$Df = \begin{bmatrix} 7y & 7x & 2z \\ 1 & 3 & -4z^3 \end{bmatrix}. \quad \square$$

---

**Example.** Compute  $Df$  for

$$(x, y, z) = f(s, t) = (s^2 - t^2, 3st, s^3t^3)$$

We have 2 inputs and 3 outputs, so  $Df$  will be a  $3 \times 2$  matrix:

$$Df = \begin{bmatrix} 2s & 2t \\ 3t & 3s \\ 3s^2t^3 & 3s^3t^2 \end{bmatrix}. \quad \square$$

---

**Example.** Compute  $Df$  for

$$y = \tan x.$$

$$Df = [(\sec x)^2].$$

You can see this is the ordinary derivative of a function of 1 variable. We drop the matrix brackets and just write  $Df = (\sec x)^2$ , or  $y' = (\sec x)^2$  like usual.  $\square$

---

**Definition.** Let  $f : U \rightarrow \mathbb{R}$ , where  $U \subset \mathbb{R}^n$ . The derivative  $Df$  is called the **gradient** of  $f$ , and is denoted  $\nabla f$ .

**Example.** Find the gradient (the derivative) of

$$z = f(x, y) = x^2 - xy - y^2.$$

$$\nabla f = [2x - y \quad -x - 2y].$$

Notice that the gradient is a **row vector** — that is, a  $1 \times n$  matrix.  $\square$

---

**Example.** Find the gradient (the derivative) of

$$w = f(x, y, z) = \frac{x + 3y}{z}.$$

$$\nabla f = \left[ \frac{1}{z} \quad \frac{3}{z} \quad -\frac{x+3y}{z^2} \right]. \quad \square$$

Since you'd expect that the best linear approximation to a linear function is the linear function, and since the derivative of  $f$  as the best linear approximation to  $f$ , you'd expect the derivative of a linear function to be the linear function.

**Proposition.** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear function. Then  $f$  is differentiable, and  $Df = f$ .

**Proof.** Since  $f$  is linear,  $f(x+h) = f(x) + f(h)$ . Therefore,

$$\lim_{\|h\| \rightarrow 0} \frac{\|f(x+h) - f(x) - f(h)\|}{\|h\|} = \lim_{\|h\| \rightarrow 0} \frac{\|f(x) + f(h) - f(x) - f(h)\|}{\|h\|} = \lim_{\|h\| \rightarrow 0} \frac{0}{\|h\|} = 0.$$

This shows that  $Df = f$ .  $\square$

**Corollary.** (a) The function  $s : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  given by  $s(x, y) = x + y$  is differentiable.

Note:  $\mathbb{R}^n \times \mathbb{R}^n$  consists of pairs  $(x, y)$  where  $x, y \in \mathbb{R}^n$ .

(b) For  $c \in \mathbb{R}$ , the function  $p : \mathbb{R}^n \rightarrow \mathbb{R}^n$  given by  $p(x) = c \cdot x$  is differentiable.

**Proof.** Both  $s$  and  $p$  are linear functions, so (a) and (b) follow from the proposition.  $\square$

**Example.** The following function  $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  is linear:

$$f(x, y, z) = (2x + 3y, 5y - y, 7x + 10y).$$

Verify that it's equal to its derivative.

Computing the partial derivatives directly,

$$Df = \begin{bmatrix} 2 & 3 \\ 5 & -1 \\ 7 & 10 \end{bmatrix}.$$

Writing  $f$  as a matrix multiplication, I have

$$f \left( \begin{bmatrix} x \\ y \\ z \end{bmatrix} \right) = \begin{bmatrix} 2 & 3 \\ 5 & -1 \\ 7 & 10 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

You can see that the matrices are the same.  $\square$

Here are some other standard and unsurprising results about derivatives. You can prove them using the preceding corollary and the **Chain Rule**, which I'll discuss later.

**Theorem.** Let  $U \subset \mathbb{R}^n$ , let  $f, g : U \rightarrow \mathbb{R}^m$ , and let  $c \in \mathbb{R}$ . Suppose  $f$  and  $g$  are differentiable at  $c$ . Then:

(a)  $f + g$  is differentiable at  $c$ , and

$$D(f + g) = Df + Dg.$$

(b)  $c \cdot f$  is differentiable at  $c$ , and

$$D(c \cdot f) = c \cdot Df. \quad \square$$