Math 1c | TA: Padraic Bartlett

Recitation 2: The Derivative

Week 2 | *Caltech 2012*

1 Random Question

A **Barker sequence** of length n is a sequence

$$a_1, a_2, \dots a_n$$

of numbers, all equal to either to $+1$ or -1, such that the **autocorrelation coefficients**

$$c_k = \left| \sum_{i=1}^{n-k} a_i \cdot a_{i+k} \right|$$

are all ≤ 1. Roughly speaking, this means that if you have two streams of the same Barker sequence coming in from two different sources, and you combine these two streams by multiplying them together, you will always get a very weak output (≤ 1) whenever these two streams are out of synch at all, but a really strong output ($= n$, the length of the sequence!) when they are completely synched up.

Step 1: Find Barker sequences with lengths 2,3,4,5,7,11, and 13.

Step 2: Show that there are no Barker sequences with length longer than 13. (This is an open question! As such, let me / the world of mathematics and engineering know if you solve or disprove it.)

2 The Derivative

This recitation is centered around the idea of a **derivative** in $\mathbb{R}^n$, and how we can develop a "good" notion of the derivative in the multivariable case using our intuition from the single-variable case. There are a number of possible ways to do this! Perhaps the easiest is the concept of a **partial derivative**, which we define below:

Definition. The **partial derivative** $\frac{\partial f}{\partial x_i}$ of a function $f : \mathbb{R}^n \to \mathbb{R}$ along its i-th coördinate at some point $\mathbf{a}$, formally speaking, is the limit

$$\lim_{h \to 0} \frac{f(\mathbf{a} + h \cdot \mathbf{e}_i) - f(\mathbf{a})}{h}.$$

(Here, e_i is the i-th basis vector, which has its i-th coördinate equal to 1 and the rest equal to 0.)

However, this is not necessarily the best way to think about the partial derivative, and certainly not the easiest way to calculate it! Typically, we think of the i-th partial derivative of f as the derivative of f when we "hold all of f's other variables constant" – i.e. if we think of f as a single-variable function with variable x_i, and treat all of the other x_j's as constants. This method is markedly easier to work with, and is how we actually *calculate* a partial derivative.

For an example, consider the following function, whose graph in $\mathbb{R}^3$ is a **hyperboloid of one sheet**:

Example. Suppose that we let $f(x, y) = \sqrt{x^2 + y^2 - 1}$. What are the partial derivatives of f at $(1, -1)$? Geometrically speaking, how can we interpret these partial derivatives?

Answer. If we hold y constant above and think of f as a single-variable function with variable x, we can use our Ma1a skills to take the derivative of this as a function solely in f:

$$\frac{\partial f}{\partial x} = \frac{2x}{2\sqrt{x^2 + y^2 - 1}} = \frac{x}{\sqrt{x^2 + y^2 - 1}}.$$

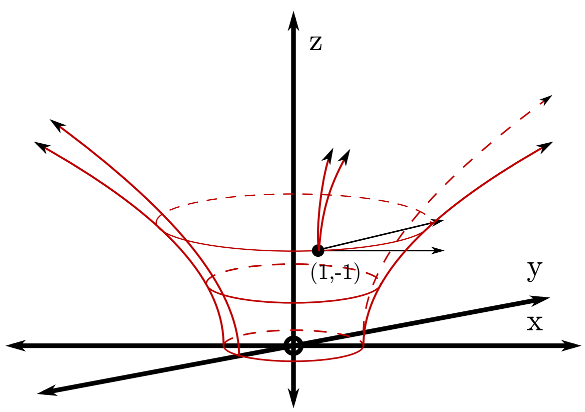
Similarly, if we hold y constant and think of f as a single-variable function with variable y, we can calculate

$$\frac{\partial f}{\partial y} = \frac{2y}{2\sqrt{x^2 + y^2 - 1}} = \frac{y}{\sqrt{x^2 + y^2 - 1}}.$$

Plugging in $(1, 1)$ gives us

$$\frac{\partial f}{\partial x} = \frac{1}{\sqrt{1^2 + (-1)^2 - 1}} = 1,$$
$$\frac{\partial f}{\partial y} = \frac{-1}{\sqrt{1^2 + (-1^{)}2 - 1}} = -1.$$

Visually, if we look at the graph of $f(x, y)$ near the point (1,1), we can see that these partial derivatives are telling us how the z-co”ordinate changes as we vary our inputs x or y near (1,1). Specifically, these partial derivatives are telling us that near (1,1), relatively small increases in x will produce increases of roughly the same magnitude in z, while relatively small increases in y will produce *decreases* of about the same magnitude in z. I.e. near (1,1), our function has roughly “slope 1” in the xz-plane, and “slope -1” in the yz-plane:



Partial derivatives are great in the sense that they're really easy to calculate: all we have to do is use our single-variable skills, and we can easily calculate pretty much any partial derivative! However, in of itself, a partial derivative is not as immediately **useful** as the derivative was back in single-variable calculus.

Specifically, in Math 1a, we thought of the derivative of a function as giving us a way to create a "linear approximation" of that function! I.e. for $\mathbb{R}^1$, the derivative $f'(a)$ was a constant such that the function $f(a) + xf'(a)$ was "very close" to $f(x)$ near a: i.e. it was a constant chosen such that the limit

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - f'(a)\cdot h}{h} = 0.$$

(i.e. in the above limit, subtracting $f(a) + f'(a)\cdot h$ took away the "linear" part of f, leaving it with only (if f had a Taylor series $\sum c_i x^i$) terms that are quadratic or higher-order.)

Analogously, for $f : \mathbb{R}^n \to \mathbb{R}$, we can ask that the derivative be something similar! Specifically, consider the following definition:

Definition. The function $f : \mathbb{R}^n \to \mathbb{R}$ has a **total derivative** $T_{\mathbf{a}}$ at some point $\mathbf{a}$ if $f(\mathbf{a}) + T_{\mathbf{a}} \cdot (\mathbf{x})$ is a "linear approximation" of f at $\mathbf{a}$: i.e. if the limit

$$\lim_{||\mathbf{h}||\to 0} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - T_{\mathbf{a}}\cdot \mathbf{h}}{||\mathbf{h}||} = 0.$$

While this definition of the derivative has the advantage that it captures this idea of a "linear approximation" in a way that the directional derivative doesn't obviously do, it has the downside that it seems impossible to calculate! How can we find such a thing?

Consider the following definition and theorem:

Definition. The **differential** of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is the following matrix of partial derivatives:

$$D(f)\big|_{\mathbf{a}} = \left(\frac{\partial f_i}{\partial x_j}(\mathbf{a})\right).$$

Theorem 1 *If $f : \mathbb{R}^n \to \mathbb{R}^m$ has a total derivative at the point $\mathbf{a}$, then this total derivative is simply the differential of f: i.e.*

$$T_{\mathbf{a}} = D(f)\big|_{\mathbf{a}} \left(\frac{\partial f_i}{\partial x_j}(\mathbf{a})\right).$$

A quick consequence of the above theorem is that if f has a total derivative at some point $\mathbf{a}$, it has all of its partial derivatives at that point a. A question we could then ask is the following: does the converse hold? In other words, if a function f has all of its partial derivatives at some point, does it have a total derivative at that point?

As it turns out: no! Consider the following example:

Example. The function

$$f(x,y) = \begin{cases} \frac{y^3}{x^2+y^2}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0) \end{cases}$$

has partial derivatives $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ defined on all of $\mathbb{R}^2$, and yet has no total derivative at $(0,0)$.

Solution. To find f's partial derivatives, we simply calculate and break things apart into cases.

Specifically, for $\frac{\partial f}{\partial x}$, there are two possible situations we can find ourself in: either $y \neq 0$, or $y = 0$. In the first case, if we hold y constant and differentiate with respect to x, we have (by differentiating)

$$\frac{\partial f}{\partial x} = \frac{-2xy^3}{(x^2+y^2)^2}.$$

In the second, because $y = 0 \Rightarrow f(x,y) = 0$, we're looking at the derivative with respect to x of a function that's identically 0! This is obviously 0: therefore, in this situation we have

$$\frac{\partial f}{\partial x} = 0.$$

(We have to calculate things in cases above because f is piecewise-defined: therefore, it's possible that the derivative changes when we run into the piecewise-defined part.)

Similarly, for $\frac{\partial f}{\partial y}$, we have that whenever $x \neq 0$, we have (by holding x constant and differentiating with respect to y)

$$\frac{\partial f}{\partial y} = \frac{3y^2}{x^2+y^2} - \frac{2y^4}{(x^2+y^2)^2}.$$

Whenever $x = 0$, we have $f(0,y) = \frac{y^3}{y^2} = y$ for $y \neq 0$, and $f(0,0) = 0 = y$ for $y = 0$. In other words, we have $f(0,y) = y$; therefore, we can see that in this situation we have

$$\frac{\partial f}{\partial y} = 1.$$

So: we know that if our function *did* have a total derivative at $(0,0)$, it would be given by the partials – i.e. that $T_{(0,0)}$, if it exists, must be $\left(\frac{\partial f}{\partial x}\Big|_{(0,0)}, \frac{\partial f}{\partial y}\Big|_{(0,0)}\right) = (0,1)$.

However, when we examine the limit

$$\begin{aligned}
&\lim_{||(h_1,h_2)||\to 0} \frac{\left|f((0,0)+(h_1,h_2)) - f(0,0) - T_{(0,0)}\cdot(h_1,h_2)\right|}{||(h_1,h_2)||} \\
&= \lim_{||(h_1,h_2)||\to 0} \frac{\left|\frac{h_2^3}{h_1^2+h_2^2} - 0 - (0,1)\cdot(h_1,h_2)\right|}{||(h_1,h_2)||} \\
&= \lim_{||(h_1,h_2)||\to 0} \frac{\left|\frac{h_2^3}{||(h_1,h_2)||^2} - h_2\right|}{||(h_1,h_2)||} \\
&= \lim_{||(h_1,h_2)||\to 0} \frac{\left|h_2(h_2^2 - ||(h_1,h_2)||^2)\right|}{||(h_1,h_2)||^3},
\end{aligned}$$

we can see that along the line $h_1 = h_2$, we have

$$\begin{aligned}
\lim_{||(h_1,h_2)||\to 0} \frac{h_2(h_2^2 - ||(h_1,h_2)||^2)}{||(h_1,h_2)||^3} &= \lim_{h_2\to 0} \frac{h_2(h_2^2 - (\sqrt{h_2^2+h_2^2})^2)}{(\sqrt{h_2^2+h_2^2})^3} \\
&= \lim_{h_2\to 0} \frac{h_2(h_2^2 - 2h_2^2)}{(\sqrt{2h_2^2})^3} \\
&= \lim_{h_2\to 0} \frac{-h_2^3}{(\sqrt{2})^3 \cdot |h_2|^3} \\
&= \lim_{h_2\to 0} -\frac{h_2/|h_2|}{2\sqrt{2}},
\end{aligned}$$

which goes to $-\frac{1}{2\sqrt{2}}$ when h_2 is positive and goes to 0, and $\frac{1}{2\sqrt{2}}$ when h_2 is negative and goes to 0. In either case, this limit is certainly not 0; therefore, our function is not totally differentiable at $(0,0)$.

What went wrong above? Well, while our function *did* have all of its partial derivatives, they weren't exactly the nicest partial derivatives you could hope for. For example, if you looked at $\frac{\partial f}{\partial x}$ along the line $x = y$, whenever $x \neq 0$ we had

$$\frac{\partial f}{\partial x} = \frac{-2x \cdot x^3}{(x^2+x^2)^2} = \frac{-2x^4}{4x^4} = -\frac{1}{2}.$$

However when $x = y = 0$, we have $\frac{\partial f}{\partial x}(0,0) = 0$! So, this partial is not a **continuous** function on all of $\mathbb{R}^2$. As a result, its behavior at $(0,0)$ – which we use to define $T_{(0,0)}$ – doesn't accurately represent its behavior *near* (0,0), which is why $T_{(0,0)}$ failed to be a good linear approximation to f near $(0,0)$.

As it turns out, this is the *only* way in which we can have partials and yet not have a total derivative! Specifically, we have the following theorem:

Theorem 2 *For a function $f : \mathbb{R}^n \to \mathbb{R}$, if all of f's partials exist in a neighborhood of* $\mathbf{a}$ *and are continuous at the point* $\mathbf{a}$*, then f has a total derivative at* $\mathbf{a}$.

Throughout this class, we'll use this theorem pretty much whenever we want to prove something is differentiable, because (as you saw above!) actually using the definition of differentiablilty is typically fairly awful/calculation-intensive.

So: we have a derivative. What can we do with it?

3 Applications of the Derivative

One of the most natural/useful applications of the derivative is the concept of the **tangent plane**, which we define here:

Definition. Take a function $f : \mathbb{R}^n \to \mathbb{R}$ that's differentiable at some point $\mathbf{a}$. We can define the **tangent plane** to f at $\mathbf{a}$ as the set of all points $(x_1, \ldots x_{n+1})$ that satisfy the following linear equation:

$$x_{n+1} - f(\mathbf{a}) = \left(\frac{\partial f}{x_1}(\mathbf{a}), \frac{\partial f}{x_2}(\mathbf{a}), \dots \frac{\partial f}{x_n}(\mathbf{a}) \right) \cdot (x - a_1, x - a_2, \dots x - a_n)\,.$$

In the specific case where $n = 2$ and we're looking at a function of the form $f(x, y)$, note that this forms an actual plane in $\mathbb{R}^3$, with slopes in the xz and yz- planes that match the slopes $\frac{\partial f}{x}, \frac{\partial f}{y}$.

In other words, the tangent plane to a function is the plane with slopes given by its partial derivatives.

This concept is perhaps best illustrated with an example:

Example. The function

$$f(x, y) = \sqrt{1 - \frac{x^2}{4} - \frac{y^2}{4}}$$

has the tangent plane $z - 1 = 0$ at $(0, 0, 1)$.

Proof. Using our understanding of how partial derivatives work, we can calculate that

$$\frac{\partial f}{\partial x} = \frac{-2x/4}{2\sqrt{1 - \frac{x^2}{4} - \frac{y^2}{4}}} = -\frac{x}{4\sqrt{1 - \frac{x^2}{4} - \frac{y^2}{4}}},$$

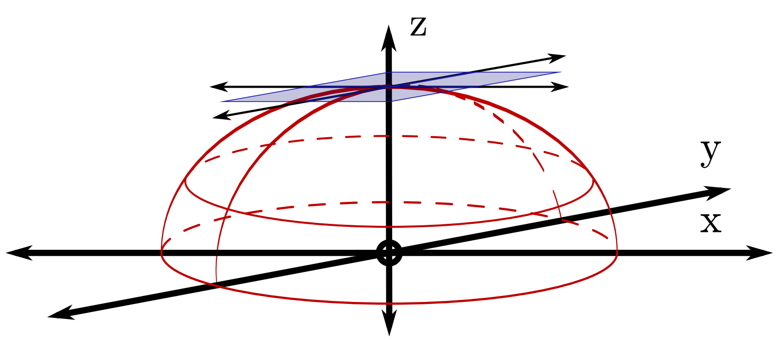
and

$$\frac{\partial f}{\partial y} = \frac{-2y/4}{2\sqrt{1 - \frac{x^2}{4} - \frac{y^2}{4}}} = -\frac{y}{4\sqrt{1 - \frac{x^2}{4} - \frac{y^2}{4}}}.$$

In particular, at (0,0) we can calculate that both of these quantities are 0, and therefore that the equation for our tangent plane there is

$$\begin{aligned} z - f(0, 0) &= \left(\frac{\partial f}{\partial x}(0, 0), \frac{\partial f}{\partial y}(0, 0) \right) \cdot (x - 0, y - 0) \\ \Rightarrow z - \sqrt{1 - 0^2 - 0^2} &= (0, 0) \cdot (x, y) \\ \Rightarrow z - 1 &= 0 \end{aligned}$$

We graph this situation below:



Another application of the derivative is to the concept of a **tangent vector**, which we define here:

Definition. For a function $f(t) : \mathbb{R} \to \mathbb{R}^n$, we can define the **tangent vector** to f at some value a as the vector

$$\left(\frac{\partial f_1}{\partial t}(a), \frac{\partial f_2}{\partial t}(a), \dots \frac{\partial f_n}{\partial t}(a) \right),$$

whenever this quantity is nonzero. (When it is 0, it's not really quite a "tangent" vector as it doesn't have a direction; so that's why we ask that this vector is not identically 0.)

Visually, if you're thinking of f as the function corresponding to the position of a particle in $\mathbb{R}^n$ at some time t, the **tangent vector** can be interpreted as the **velocity vector** for the particle at time t.

Again, we illustrate this with an example:

Example. The function

$$f(t) = (t^2, t^3)$$

has the tangent vector $(2, 3)$ at time $t = 1$.

Proof. This is a pretty trivial calculation: by the definition above, we have that

$$\begin{aligned} D(f)(t) &= \left(\frac{\partial f_1}{\partial t}(a), \frac{\partial f_2}{\partial t}(a) \right) \\ &= (2t, 3t^2) \\ \Rightarrow D(f)(1) &= (2, 3). \end{aligned}$$

4 Tools for Taking Derivatives

Switching gears somewhat, we now turn from the theory of derivatives to a more practical approach – how do we calculate these things? For functions $\mathbb{R}^1 \to \mathbb{R}^1$, in particular, we had things like the product and chain rule; are there analogues for functions $\mathbb{R}^n \to \mathbb{R}^m$?

Well: yes! Specifically, for the product rule, we have the following theorem:

Theorem 3 *Suppose that f, g are a pair of functions $\mathbb{R}^n \to \mathbb{R}^m$, and we're looking at the inner product*[1] *$f \cdot g$ of these two functions. Then, we have that*

$$D(f \cdot g)\Big|_{\mathbf{a}} = f(\mathbf{a}) \cdot (D(g))\Big|_{\mathbf{a}} + g(\mathbf{a}) \cdot (D(f))\Big|_{\mathbf{a}}.$$

We have a similar extension of the chain rule to the multivariable case, as well:

[1] Recall that the **inner product** of two vectors $\mathbf{u}, \mathbf{v}$ is just the real number $\sum_{i=1}^m u_i v_i$.

Theorem 4 *Take any function $g : \mathbb{R}^m \to \mathbb{R}^l$, and any function $f : \mathbb{R}^n \to \mathbb{R}^m$. Then, we have*

$$D(g \circ f)\Big|_{\mathbf{a}} = D(g)\Big|_{f(\mathbf{a})} \cdot D(f)\Big|_{\mathbf{a}}.$$

One interesting/cautionary tale to notice from the above calculations is that the partial derivative of $g \circ f$ with respect to one variable x_i can depend on **many** of the variables and coördinates in the functions f and g!

I.e. something many first-year calculus students are tempted to do on their sets is to write

$$\frac{\partial (g \circ f)_i}{\partial x_j}\Big|_{\mathbf{a}} = \frac{\partial g_i}{\partial x_j}\Big|_{f(\mathbf{a})} \cdot \frac{\partial f_i}{\partial x_j}\Big|_{\mathbf{a}}.$$

DO NOT DO THIS. Do not do this. Do not do this. Ever. Because it is wrong. Indeed, if you expand how we've stated the chain rule above, you can see that $\frac{\partial (g \circ f)_i}{\partial x_j}\Big|_{\mathbf{a}}$ – the (i, j)-th entry in the matrix $D(g \circ f)$ – is actually equal to the i-th row of $D(g)\Big|_{f(\mathbf{a})}$ multipled by the j-th column of $D(f)\Big|_{\mathbf{a}}$ – i.e. that

$$\frac{\partial (g \circ f)_i}{\partial x_j}\Big|_{\mathbf{a}} = \begin{bmatrix} \frac{\partial g_i}{\partial x_1}\Big|_{f(\mathbf{a})} & \dots \frac{\partial g_i}{\partial x_m}\Big|_{f(\mathbf{a})} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial f_1}{\partial x_j}\Big|_{\mathbf{a}} \\ \vdots \\ \frac{\partial f_m}{\partial x_j}\Big|_{\mathbf{a}} \end{bmatrix}.$$

Notice how this is much more complex! In particular, it means that the partials of $g \circ f$ depend on all sorts of things going on with g and f, and aren't restricted to worrying about just the one coördinate you're finding partials with respect to.

The moral here is basically if you're applying the chain rule without doing a *lot* of derivative calculations, you've almost surely messed something up. So, when in doubt, just find the matrices $D(f)$ and $D(g)$!

We work one example, to illustrate how to do these kinds of calculations:

Example. If $f(x) = (x, x^2, x^3)$ and $g(x, y, z) = \sin(xyz)$, use the chain rule to find $D(g \circ f)\Big|_{\mathbf{a}}$, for any $a \in \mathbb{R}$.

Solution. If we straightforwardly apply the chain rule, we have that

$$\begin{aligned} D(g \circ f)\Big|_{\mathbf{a}} &= D(g)\Big|_{f(\mathbf{a})} \cdot D(f)\Big|_{\mathbf{a}} \\ &= \begin{bmatrix} \frac{\partial g}{\partial x}\Big|_{f(\mathbf{a})} & \frac{\partial g}{\partial y}\Big|_{f(\mathbf{a})} & \frac{\partial g}{\partial z}\Big|_{f(\mathbf{a})} \end{bmatrix} \cdot \begin{bmatrix} \frac{\partial f_1}{\partial x}\Big|_{\mathbf{a}} \\ \frac{\partial f_2}{\partial x}\Big|_{\mathbf{a}} \\ \frac{\partial f_3}{\partial x}\Big|_{\mathbf{a}} \end{bmatrix} \end{aligned}$$

$$= \left[\begin{array}{ccc} yz \cdot \cos(xyz)\Big|_{(a,a^2,a^3)} & xz \cdot \cos(xyz)\Big|_{(a,a^2,a^3)} & xy \cdot \cos(xyz)\Big|_{(a,a^2,a^3)} \end{array}\right] \cdot \left[\begin{array}{c} 1\Big|_{\mathbf{a}} \\ 2x\Big|_{\mathbf{a}} \\ 3x^2\Big|_{\mathbf{a}} \end{array}\right]$$

$$= \left[\begin{array}{ccc} a^5 \cdot \cos(a^6) & a^4 \cdot \cos(a^6) & a^3 \cdot \cos(a^6) \end{array}\right] \cdot \left[\begin{array}{c} 1 \\ 2a \\ 3a^2 \end{array}\right]$$

$$= a^5 \cdot \cos(a^6) + 2a^5 \cdot \cos(a^6) + 3a^5 \cdot \cos(a^6)$$

$$= 6a^5 \cdot \cos(a^6).$$

As a quick sanity check, we can verify that this makes sense by just looking at the function $g \circ f$ directly: $g \circ f(x) = \sin(x \cdot x^2 \cdot x^3) = \sin(x^6)$, and therefore $(g \circ f)'(a) = 6a^5 \cdot \cos(a^6)$ by applying the one-dimensional version of the chain rule.