MA1C, WEEK 2: TOTAL DERIVATIVES, TANGENT PLANES AND THE CHAIN RULE.

TA: PADRAIC BARTLETT

These notes, like all future notes, can be found on my website.

1. Last Week's HW

Average: was around 90%. There were no major problems to speak of; just calculation difficulties in places.

2. Random Question

The Four-Color Theorem is a famous theorem in mathematics that is notable for being the first major theorem whose proof required the use of a computer; it says that any map in a plane can be colored with four colors so that no two countries who share a border are of the same color. This kind of question, however, extends to other shapes – for example, we can ask how many colors would be required to color a map on a torus, i.e. a division of a torus into countries.

It turns out that you need up to 7 colors to color any map on the torus. Can you find a map which requires you to use all seven colors?

3. Total Derivatives and Tangent Planes - Definitions

Definition 3.1. For a function $f : \mathbb{R}^n \to \mathbb{R}$, recall that we defined the **partial derivative** $\frac{\partial(f)}{\partial x_i}$ of f as

$$\frac{\partial(f)}{\partial x_i}(a_1 \ldots a_n) := \lim_{h \to 0} \frac{f(a_1, \ldots a_i + h, \ldots a_n) - f(a_1 \ldots a_n)}{h}$$

The idea here was that this gives us the rate of change of f in the x_i-"direction" at a point a.

However, when we talk about a derivative what we really want is a way to talk about some nice linear approximation to a function – this is what we always want to use derivatives for, as they give us an idea of where a function is "going" at a certain point in time. So: the partial derivatives in of themselves aren't necessarily enough information to really talk about what the function is doing at a point, as it's a priori possible that the function's derivatives might look like one kind of thing if you approach it along the paths $t \mapsto x_i$, but does something completely crazy on the paths that the partial derivatives aren't telling you about! What we really want is for the derivative to give us not a collection of n paths, but a **tangent plane** of dimension n which approximates our function along any path! This motivates the definition below of the total derivative:

Definition 3.2. For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, we define the total derivative as the $m \times n$ matrix

$$Df(a) = \left[\frac{\partial f_i}{\partial x_j}(a)\right],$$

and say that f is differentiable at a if and only if the limit

$$\lim_{x\ to 0} \frac{||f(x) - f(a) - Df(a) \cdot (x - a)||}{||x - a||} = 0.$$

If this holds and $m = 1$, we can say that the **tangent plane** to f at a exists, and define it as the plane

$$x_{n+1} = f(a) + \left(\frac{\partial f}{\partial x_1}(a)\right) \cdot (x_1 - a_1) + \ldots + \left(\frac{\partial f}{\partial x_n}(a)\right) \cdot (x_n - a_n). \tag{3.3}$$

Remark 3.4. The above definition is often quite ponderous for showing that something is differentiable: it is often easier to simply show the stronger statement that a function f is C^1 – i.e. that all of its partial derivatives exist and are continuous – because we proved in class that C^1 functions are all differentiable. it bears noting, however, that there are differentiable functions that are not C^1! Can you think of any?

4. Total Derivatives and Tangent Planes - Examples

So: to clarify the definitions above, we work several examples.

Example 4.1. Let

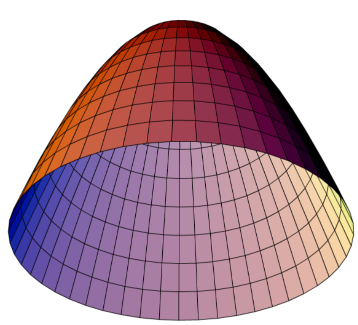
$$f(x, y) = -x^2 - y^2.$$

Decide whether f is differentiable, and find its tangent plane at 0, if it exists.

Proof. So: the partials of f are simply $\frac{\partial f}{\partial x} = -2x, \frac{\partial f}{\partial y} = -2y$; so, as both of these are continuous, we have that f is C^1 and thus differentiable. Its tangent plane at an arbitrary point a, b in space is given by the equations

$$z = -a^2 - b^2 - 2a(x - a) - 2b(y - b) = a^2 + b^2 - 2ax - 2by;$$

in specific, at $(a, b) = (0, 0)$, this is just the tangent plane $z = 0$; i.e. the xy-plane. Looking at the picture of the graph below, this is in accord with our intuition of what the tangent plane to 0 of this function should look like.



□

Example 4.2. Let

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

Decide whether f is differentiable, and find its tangent plane at 0, if it exists.

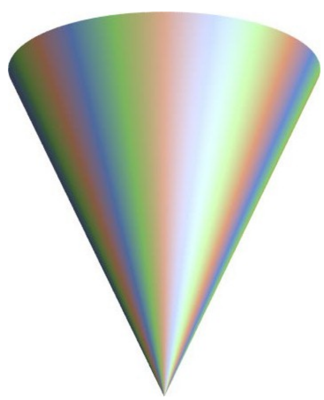
Proof. So: the partials of f are $\frac{\partial f}{\partial x} = \frac{x}{\sqrt{x^2+y^2}}, \frac{\partial f}{\partial y} = \frac{x}{\sqrt{x^2+y^2}}$. However, neither of these functions are even defined at 0, nor can they be, as

$$\frac{\partial f}{\partial x}(\epsilon, 0) = \frac{\epsilon}{|\epsilon|} = \pm 1,$$

depending on whether ϵ is greater than 0 or less than 0, and thus the limit

$$\frac{\partial(f)}{\partial x_i}(0,0) := \lim_{h\to 0} \frac{f(\epsilon,0) - f(0,0))}{h}$$

doesn't exist. So the total derivative cannot be said to exist at (0,0), as the partials do not even **exist** there. This, again, agrees with our intution: because this shape is simply a **cone**, and as such there is intuitively no way to approximate the "point" of the cone with a plane of some sort.



□

Example 4.3. Let

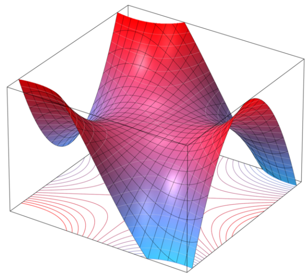
$$f(x,y) = x^3 - 3xy^2$$

Decide whether f is differentiable, and find its tangent plane at 0, if it exists.

Proof. So: the partials of f are $\frac{\partial f}{\partial x} = 3x^2 - 3y^2, \frac{\partial f}{\partial y} = -6xy$; so, as both of these are continuous, we have that f is C^1 and thus differentiable. Its tangent plane at an arbitrary point a, b in space is given by the equations

$$z = a^3 - 3ab^2 + (3a^2 - 3b^2)(x-a) - 6ab(y-b)$$

and thus specifically at $(a,b) = (0,0)$, this is just the tangent plane $z = 0$; again, the xy-plane. Looking at the picture of the graph below, this yet again makes sense.



This surface is called a "monkey saddle" (because it looks like a normal "saddle point," but with a third depression; the idea being that a monkey would need a saddle with three depressions to ride properly, as it would need a place for its tail as well as its legs.) ☐

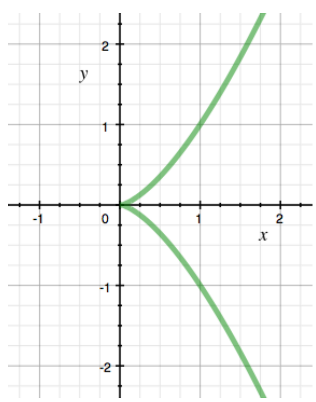
It's also worth defining a concept similar to that of the tangent plane here, that of a **tangent vector**: i.e. for a function $f : \mathbb{R} \to \mathbb{R}^n$ (i.e. a path in $\mathbb{R}^n$), we define the tangent vector at any point $x \in \mathbb{R}$ to be the vector $(\frac{\partial f_1}{\partial t}(x) \ldots \frac{\partial f_n}{\partial t}(x))$. This can be thought of as the velocity vector of the path at any given point x: this exists whenever the partials all exist and aren't all identically 0.

Example 4.4. Let

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

Find f's tangent vectors, wherever they exist.

Proof. So: the partials of f are $\frac{\partial f_1}{\partial t} = 2t, \frac{\partial f_2}{\partial t} = 3t^2$; so they're defined everywhere and give us that f has a well-defined tangent vector everywhere **except for** at $t = 0$; this is because at $t = 0$ both of the partials above vanish, and we have no well-defined "vector." This, again, makes sense when compared with its graph, as



visually has no well-defined tangent vector at 0. ☐

5. The Chain Rule - Definition

So: first, recall the single-variable formulation of the chain rule:

Theorem 5.1. *For $f, g : \mathbb{R} \to \mathbb{R}$, g differentiable at x_0, f differentiable at $g(x_0)$, we have that*

$$(f \circ g)'(x_0) = f'(g(x_0)) \cdot g'(x_0).$$

The multivariable definition is almost precisely the same:

Theorem 5.2. *For $f : \mathbb{R}^m \to \mathbb{R}^p, g : \mathbb{R}^n \to \mathbb{R}^m$, g differentiable at x_0, f differentiable at $g(x_0)$, we have that*

$$D(f \circ g)(x_0) = (Df)(g(x_0)) \cdot (Dg)(x_0).$$

The only thing to stress here is that this definition of the chain rule is about the total derivative – trying to apply the chain rule to the partials of a function will frequently cause horrible things to happen to you.

6. The Chain Rule - Examples

Example 6.1. Check that the chain rule works when applied to $g(x, y) = x^2 + y^2, f(x) = \sqrt{(x)}$.

Proof. So, we calculate:

$$D(f \circ g)(x_0, y_0) = D(\sqrt{x^2 + y^2})(x_0, y_0) = \left(\frac{x_0}{\sqrt{x_0^2 + y_0^2}}, \frac{y_0}{\sqrt{x_0^2 + y_0^2}} \right)$$

from our calculations earlier in the HW. As well,

$$(Df)(g(x_0, y_0) \cdot (Dg)(x_0, y_0) = \frac{1}{2\sqrt{g(x_0, y_0)}} \cdot (2x_0, 2y_0) = \left(\frac{x_0}{\sqrt{x_0^2 + y_0^2}}, \frac{y_0}{\sqrt{x_0^2 + y_0^2}} \right).$$

So these two quantities are equal; thus, the chain rule worked here. □

Example 6.2. Check that the chain rule works when applied to $g(t) = (t^2, t^3, t), f(x, y, z) = x + y + z$.

Proof. So, we calculate:

$$D(f \circ g)(t_0) = D(t + t^2 + t^3)(t_0) = 1 + 2t_0 + 3t_0^2,$$
$$(Df)(g(t_0) \cdot (Dg)(t_0) = (1, 1, 1) \cdot (2t_0, 3t_0^2, 1)^T = 1 + 2t_0 + 3t_0^2.$$

So these two quantities are equal; thus, the chain rule again worked here. □