MA 1C RECITATION 04/09/15

1. Directional and Partial Derivatives

To begin, we will specialize our study from a general multivariable function ($f : \mathbf{R}^n \to \mathbf{R}^m$) to the case of a real-valued multivariable function $f : \mathbf{R}^n \to \mathbf{R}$. Such a function is sometimes called a *scalar field*. This is the simplest interesting case of a multivariable function, but it allows us to illustrate many of the features that are also present in a more general setting.

Since there is "calculus" in the title of this class, the first natural question is to ask is "how do you differentiate such a function"? Intuitively, differentation allows us to find the instantaneous rate of change as you move in the domain. This is easy enough in $\mathbf{R}$, since we can only move back and forth on the real line, but as you saw with limits, in $\mathbf{R}^2$ and higher dimensions, there are infinitely many ways to move! This is why we cannot just take a derivative in multivariable calculus. We must also choose a *direction*. This leads us to the following definition.

Definition 1.1. The **directional derivative** of a function f at a point $\mathbf{a}$ along a *unit vector* $\mathbf{u}$ is defined as

$$f'(\mathbf{a}; \mathbf{u}) := \lim_{h \to 0} \frac{f(\mathbf{a} + h\mathbf{u}) - f(\mathbf{a})}{h}.$$

Remark 1.2. You may also see the directional derivative denoted as $f_u(a)$.

Warning 1.3. Just like in the one-dimensional case, the limit may not exist. For instance, suppose that a function is not continuous at the point.

Warning 1.4. It is important that the vector $\mathbf{u}$ is normalized, or else you will get the wrong number.

Looking at the definition of directional derivative, we see that this is just a natural generalization of the one-dimensional derivative you know and love. Often the $\mathbf{u}$ that you choose will be the directions of your basis, such as $e_1 = (1, 0)$ and $e_2 = (0, 1)$ in $\mathbf{R}^2$. We use these so much that we give these a special name.

Definition 1.5. The **partial derivative** of $f(x_1, \ldots, x_n)$ at $\mathbf{a}$ with respect to x_k is given by

$$\frac{\partial f}{\partial x_k}(\mathbf{a}) := f'(\mathbf{a}, e_k).$$

Note that $\frac{\partial f}{\partial x_k}$ is again a multivariate function, so we can differentiate this with respect to another variable x_ℓ to get a function $\frac{\partial}{\partial x_\ell} \frac{\partial f}{\partial x_k} = \frac{\partial^2 f}{\partial x_\ell \partial x_k}$, which is called a *mixed partial derivative*.

Warning 1.6. Note that the order matters! It is *not* always true that the equality

$$\frac{\partial^2 f}{\partial x_\ell \partial x_k} = \frac{\partial^2 f}{\partial x_k \partial x_\ell}$$

Date: April 9, 2015.

holds. However, in many "nice" contexts, this equality does hold. For instance, it is a basic theorem that if all partial derivatives exist and *are continuous* (important!) on a domain D, then (mixed) partial derivatives commute on that domain.

In this class, we will *usually* be in these "nice" situations where two partial derivatives will commute, but just keep in mind that this fails in general.

1.1. **How to calculate partial derivatives.** In practice, computing partial derivatives is very easy. For instance, to compute $\partial f/\partial x_k$, just think of everything except for the variable x_k as constant and differentiate just like you would for one variable.

Let's start with a simple example.

Example 1. Let's compute the partial derivatives for the function $f(x,y) = x^2 - 2xy + y^2$. Let's first do it the long way—no shortcuts—along x, just from the definitions.

We have

$$\begin{aligned}\frac{\partial f}{\partial x} &= f'((x,y);(1,0)) \\ &= \lim_{h\to 0} \frac{((x+h)^2 - 2(x+h)y + y^2) - (x^2 - 2x + y^2)}{h} \\ &= \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} + \lim_{h\to 0} \frac{2(x+h)y - 2xy}{h} \\ &= 2x - 2y,\end{aligned}$$

where the last step is from the usual power rule for a one-dimensional derivative. Note in particular $\frac{\partial f}{\partial x}$ is a function in two variables, even though we have not explicitly written it as such.

Note that f is symmetric with respect to x and y, so by our shortcut technique, we easily compute

$$\frac{\partial f}{\partial y} = 2y - 2x.$$

Remark 1.7. We should probably write $\frac{\partial f}{\partial x}(x,y)$, but this notation is cumbersome, especially once you work in more than two variables. Just make sure that you understand what the shorthand $\frac{\partial f}{\partial x}$ stands for.

2. Total Derivative

We have infinitely many directional derivatives at a given point, each corresponding to a path through that point, but to do calculus, we want a notion of "the" derivative of a function on $\mathbf{R}^n$. This is given by the notion of total derivative.

Definition 2.1. The **total derivative** of f at $\mathbf{a}$ is a *linear transformation* $T_{\mathbf{a}}$ such that

$$f(\mathbf{a}+\mathbf{v}) = f(\mathbf{a}) + T_{\mathbf{a}}(\mathbf{v}) + ||\mathbf{v}||E(\mathbf{a},\mathbf{v})$$

for all $\mathbf{v}$ in some ball around $\mathbf{a}$ and where $E(\mathbf{a},\mathbf{v}) \to 0$ as $||\mathbf{v}|| \to 0$.

This is the rigorous of working with the following intuitive idea: if you take the limit of $(f(\mathbf{a}+\mathbf{v}) - f(\mathbf{a}))/||\mathbf{v}||$ as $||v|| \to 0$, then it (a) has a limit and (b) that limit is given by $T_{\mathbf{a}}$.

How should you think about $T_{\mathbf{a}}$? If you feed $T_{\mathbf{a}}$ a vector, it spits out a number that tells you how f is changing in that direction. From this, we see how it encodes *all* the information about directional derivatives into one object.

2.1. **Differentiability implies that Directional Derivatives Exist.** A big theorem that we saw in class if that a function is differentiable at a point (i.e. has a total derivative at that point), then it has all directional derivatives. Indeed, we have

$$f'(\mathbf{a};\mathbf{u}) := T_{\mathbf{a}}(\mathbf{u})$$

for a *unit* vector $\mathbf{u}$.

There are many examples of functions with only some directional derivatives and not others. Then these functions cannot be differentiable, however, there are a number of other subtleties that you should keep in mind.

- A function can have all its partial derivatives and still not be differentiable! (So there's no "basis" for directional derivatives.) Consider the function

  $$f(x,y) = \begin{cases} 0, & xy = 0 \\ \frac{1}{xy} & xy \neq 0. \end{cases}$$

  This doesn't have most directional derivatives (so is not differentiable), but its partial derivatives do exist.

  For a more subtle example, consider something like

  $$g(x,y) = \begin{cases} \frac{xy^2}{x^2+y^4}, & x \neq 0 \\ 0, & x = 0. \end{cases}$$

- A function can have all its directional derivatives and still be continuous! (See Apostol, p. 257.) However, if we know that a function has a total derivative, then it is continuous. (So "differentiability implies continuity" still works in the multivariable setting.)
- Just because a function is differentiable doesn't mean its partial derivatives have to be continuous! This is a warning to not assume anything is true because it "seems like it should be true."

3. The Gradient

Here comes the first of the three grand objects in multivariable calculus. (The others are the divergence and the curl, which you will see later. It's not too far-fetched of a statement to say that multivariable calculus is the study of div, grad, and curl.)

Definition 3.1. The **gradient** of a function $f : \mathbf{R}^n \to \mathbf{R}$ at a point $\mathbf{a}$ is defined as

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(\mathbf{a}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{a}) \right).$$

In particular, note that if f is *differentiable*, then the gradient of f is precisely the total derivative of f, written with respect to the standard basis.

Remark 3.2. Think of the gradient as a function from the set of n-variable functions to the set of n-vectors. You feed it a function, it spits out a vector. You can feed this vector a point to get information about partial derivatives in every direction at that point.

3.1. **Computing directional derivatives easily.** If a function is differentiable, we noted above that $f'(\mathbf{a};\mathbf{u}) = T_{\mathbf{a}}(\mathbf{u})$. We also observed that $T_{\mathbf{a}}$, expressed in terms of the standard basis, is just $\nabla f(\mathbf{a})$. Therefore, if we have differentiability, the easiest way to compute directional derivatives to note that

$$f'(\mathbf{a};\mathbf{u}) = \nabla f(\mathbf{a}) \cdot \mathbf{u}.$$

Here is a problem that exploits a property of the gradient that is a simple consequence of the above method of computing directional derivatives: for differentiable functions, *gradients point in the direction of maximal increase.*

Example 2. Find a differentiable function whose maximal directional derivative at $(2,3)$ is equal to 1 in the direction $(1,0)$.

Since the directional derivative in the $(1,0)$ direction is $\nabla f(2,3) \cdot (1,0)$, to have this equal to 1, we need to find a f such that

$$\frac{\partial f}{\partial x}(2,3) = 1.$$

There are many ways to find such functions.

Now, for the other condition. In other to be the direction of maximal increase, $(1,0)$ needs to maximize $\nabla f(2,3) \cdot \mathbf{v}$. Recall that a property of the dot product $\mathbf{R}^n$ is that $|\mathbf{a} \cdot \mathbf{b}| = \|\mathbf{a}\|\|\mathbf{b}\| \cos\theta$ where θ is the "angle" between the vectors. Thus, we maximize by setting $\mathbf{v}$ equal to the (normalized) gradient, that is,

$$\nabla f(2,3) = (1,0).$$

We now just need to find a function that satisfies these two conditions. One example is

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

4. General Derivatives

We're now going to move beyond the scalar field case and consider functions $f : \mathbf{R}^n \to \mathbf{R}^m$. Many of the definitions carry over verbatim to this more general case. For instance,

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

We usually split up such functions f into scalar-valued functions in each coordinate, that is, we write

$$\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})).$$

However, now instead of the gradient, the total derivative expressed in terms of the standard basis is the Jacobian:

$$D\mathbf{f}(\mathbf{a}) = \left[\frac{\partial f_i}{\partial x_j}\right] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Remark 4.1. Note that the coordinate functions form the *rows* of the Jacobian and the variable of which you are taking the partial derivative form the *columns*. I like to think of this as "stacking rows of gradients" to distinguish the Jacobian from its transpose.

This looks complicated, but but conceptually nothing has changed: you give $D\mathbf{f}$ a direction, and it tells you how $\mathbf{f}$ changes in that direction.

We'll see more of this next week, so I'll just stop with introducing this definition.

5. How can you tell if a function is differentiable?

As we saw above, it is very convenient to know that a function is differentiable, since we have all of these shortcuts that we can use. However, we have not mentioned any way for us to *determine* when a function is differentiable.

The most important such condition is the following result, which is how we usually show that a function is differentiable at a point.

5.1. An important existence proof.

The following theorem is essential; it says that "usually" (that is, in "nice" circumstances) a total derivative exists.

Theorem 5.1. *("If the partial derivatives are continuous, then the total derivative exists.") Let $f = (f_1, \ldots, f_m) : D \subseteq \boldsymbol{R}^n \to \boldsymbol{R}^m$ for an $a \in Int(D)$ and let $\epsilon > 0$ be such that $B_a(\epsilon) \subseteq D$ and all partial derivatives $\frac{\partial f_i}{\partial x_j}$ exist in $B_a(\epsilon)$ and are continuous at a. Then $f'(a)$ exists, i.e. f is differentiable at a.*

(And of course, $f'(a)$ is given by the Jacobian matrix $[\frac{\partial f_i}{\partial x_j}(x)] = Df(a)$.)

Proof. By taking coordinates, we can see that $f'(a)$ exists if and only if $f_1'(a), \ldots, f_m'(a)$ all exist (and the matrix of $f'(a)$ has vectors $f_1'(a), \ldots, f_m'(a)$). So we can assume $f : D \subseteq \mathbf{R}^n \to \mathbf{R}$.

Define the linear functions $L : \mathbf{R}^n \to \mathbf{R}$ by

$$L(h) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a) h_i$$

where $h = (h_1, \ldots, h_n) = \nabla f(a) \cdot h$. (We know that if $f'(a)$ exists, it has to be L). Fix small h and for $a = (a_1, \ldots, a_n)$, let

$$\phi_j(t) = f(a_1 + h_1, \ldots, a_{j-1} + h_{j-1}, t, a_{j+1}, \ldots, a_n).$$

Then

$$f(a + h) - f(a) = \sum_{j=1}^{n} \phi_j(a_j + h_j) - \phi_j(a_j).$$

Note that $\phi_j(t)$ is continuous in $[a_j, a_j + h_j]$ and differentiable there since

$$\phi_j'(x) = \frac{\partial f}{\partial j}(a_1 + h_1, \ldots, a_{j-1} + h_{j-1}, x, a_{j+1}, \ldots, a_n).$$

By the Mean Value Theorem,

$$\frac{\phi_j(a_j + h_j) - \phi_j(a_j)}{h_j} = \phi_j'(t_j(h_j)) = \frac{\partial f}{\partial x_j}(s_j(h_j))$$

where $t_j(h_j) \in [a_j, a_j + h_j]$ and

$$s_j(h_j) = (a_1 + h_1, \ldots, a_{j-1} + h_{j-1}, t_j(h_j), a_{j+1}, \ldots, a_n).$$

We want to show that

$$\lim_{h \to 0} \frac{f(a + h) - f(a) - \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a) h_j}{||h||} = 0.$$

But the above expression (inside the limit) is

$$\frac{1}{||h||}\sum_{j=1}^{n} h_j\left(\frac{\partial f}{\partial x_j}(s_j(h_j)) - \frac{\partial f}{\partial x_j}(a)\right) \in \mathbf{R}.$$

Since $||h_j|| \leq ||h||$, the absolute value of this is bounded by

$$\sum_{j=1}^{n}\left|\frac{\partial f}{\partial x_j}(s_j(h_j)) - \frac{\partial f}{\partial x_j}(a)\right|.$$

But as $h \to 0$, $t_j(h) \to a_j$, so $s_j(h_j) \to a$ and then by the continuity of the partial derivatives, at a the last sum converges to 0. □

5.2. **Examples.**

Example 3. Let's show that $f(x, y) = -\cos(xy)$ is differentiable at $(0, 0)$.

By Theorem 8.3, we can find the partials by taking the other variable to be constant. We have

$$\frac{\partial f}{\partial x} = y\sin(xy)$$

which is continuous at $(0, 0)$, e.g. by the procedure we discussed in recitation last week. Similarly,

$$\frac{\partial f}{\partial y} = x\sin(xy)$$

is continuous. Thus, by the theorem above, $f(x, y)$ is differentiable at $(0, 0)$.

Let's try something a little more complicated.

Example 4. Let's show that $f(x, y) = x^2+y$ is differentiable at $(0, 0)$, directly from the definition. In other words, we want to find a linear transformation $T = T_{(0,0)}$ such that

$$f(\mathbf{v}) = 0 + T(\mathbf{v}) + ||\mathbf{v}||E((0, 0), \mathbf{v})$$

for all $\mathbf{v}$ such that $||\mathbf{v}|| < \delta$ where we can pick $\delta > 0$ and such that $E(\mathbf{v}) = E((0, 0), \mathbf{v}) \to 0$ as $\mathbf{v} \to (0, 0)$.

This is often difficult, and why we want to use the theorem above when we can; then all we need to do is produce T. If the partial derivatives exists, we have a good idea of what the T should be. If you can picture the function like this (it's a function from $\mathbf{R}^2 \to \mathbf{R}$, so we can visualize it in 3-space, then the tangent place should be what we should get if we ignore the E-term.

Anyway, by one of the methods above, we guess $T = (0, 1)$. By substituting it in the above equality, we can find the E:

$$\begin{aligned} f(\mathbf{v}) &= T(\mathbf{v}) + ||\mathbf{v}||E(\mathbf{v}) \\ v_1^2 + v_2 &= (0, 1)\cdot(v_1, v_2) + \sqrt{v_1^2 + v_2^2}E(\mathbf{v}) \\ \frac{v_1^2}{\sqrt{v_1^2+v_2^2}} &= E(\mathbf{v}). \end{aligned}$$

In the second line, technically, $\mathbf{v}$ should be a column vector, but I replaced it with the dot product, which gives the same result in this case.

Finally, let's confirm that the E works. First, the equality part of definition certainly holds in some ball around $(0, 0)$. We then need to verify that $E \to 0$ as $\mathbf{v} \to (0, 0)$. We do so with a short $\epsilon - \delta$ proof.

Proof. Let $\epsilon > 0$. Set $\delta = \epsilon$. We note that if $||\mathbf{v}|| < \delta$, then $|v_1| < \delta = \epsilon$. We have

$$E(\mathbf{v}) = \frac{v_1^2}{\sqrt{v_1^2 + v_2^2}} \le |v_1| < \epsilon$$

and so we have shown that $E \to 0$ as $||\mathbf{v}|| \to 0$. □