MA 1C RECITATION 04/10/13

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 10, 2013.

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

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

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 2.2. 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. The Chain Rule

Since we can take partial derivatives and partial derivatives behave much like derivatives from single-variable calculus, we must have a chain rule for partial derivatives as well. Namely, we have the following result.

Theorem 3.1. *If $\boldsymbol{r} : \boldsymbol{R} \to \boldsymbol{R}^n$ and $f : \boldsymbol{R}^n \to \boldsymbol{R}$ are functions such that $g = f \circ \boldsymbol{r}$, then*

$$g'(t) = \nabla f(\boldsymbol{a}) \cdot \boldsymbol{r}'(t)$$

where $\boldsymbol{a} = \boldsymbol{r}(t)$.

This formula is useful. Personally, I find the following version of the chain rule more intuitive. I'll state it in a special case, but you can see how you can generalize this.

Proposition 3.2. *Suppose that $z = f(x, y)$ is a differentiable function of x and y, where $x = g(s, t)$ and $t = h(s, t)$ are differentiable functions of s and t. Then*

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}$$
$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}$$

The way I think about this is that since f depends on x and y, and x and y depend on s and t, then if you want to the take the derivative of f with respect to (say) s, you need to sum over the ways that f changes with respect to s.

Remark 3.3. Why is this summing enough? Can't there be some other kind of contribution to the way f changes with respect to s? Actually, no. Recall that the partial derivatives are just the directional derivatives take with respect to the standard bases e_i, and that these form an orthonormal basis. Since they are orthogonal and span the entire space, we can write the way a function changes just by seeing how it with respect to a certain basis.

Remark 3.4. This kind of calculation might get hairy once you get more variables, or run into more complicated compositions, so it might help to draw a dependency graph.

Example 2. If $z = e^x \sin y$, where $x = st^2$ and $y = s^2t$, find $\partial z/\partial s$.

Solution. We have

$$\begin{aligned}\frac{\partial z}{\partial s} &= \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \\ &= (e^x \sin y)(t^2) + (e^x \cos y)(2st) \\ &= t^2 e^{st^2} \sin(s^2 t) + 2ste^{st^2} \cos(s^2 t).\end{aligned}$$

For extra practice, show that

$$\frac{\partial z}{\partial t} = 2ste^{st^2} \sin(s^2 t) + s^2 e^{st^2} \cos(s^2 t).$$

□

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

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

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.

5. General Chain Rule

The general chain rule is elegant. Suppose that $\mathbf{h} = \mathbf{f} \circ \mathbf{g}$, and $\mathbf{g}$ is differentiable at $\mathbf{a}$ and $\mathbf{f}$ is differentiable at $\mathbf{g}(\mathbf{a})$. Then

$$D\mathbf{h}(\mathbf{a}) = D(\mathbf{f}(\mathbf{g}(\mathbf{a}))D\mathbf{g}(\mathbf{a})$$

where multiplication is given by matrix multiplication. Note that the usual chain rule is just a special case of this formula. An easy way to remember this is to see that *the derivative of the composition is the composition of the derivatives.*

Let's first see this abstractly, as a generalization of our example above.

Example 3. Suppose that we have a function $f : \mathbf{R}^2 \to \mathbf{R}$ and we set variables $x = X(s,t)$ and $y = Y(s,t)$, so that $f(x,y)$ is also a function of s and t. What are $\partial f/\partial s$ and $\partial f/\partial t$?

Solution. We first calculate

$$Df = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right].$$

Now, let's write f as a composition of two things to take advantage of the chain rule. Let $g : \mathbf{R}^2 \to \mathbf{R}^2$ be given by $g(s,t) = (X(s,t), Y(s,t))$. Therefore, we have

$$Dg = \begin{bmatrix} \frac{\partial X}{\partial s} & \frac{\partial X}{\partial t} \\ \frac{\partial Y}{\partial s} & \frac{\partial Y}{\partial t} \end{bmatrix}.$$

Hence, by the general chain rule, we obtain

$$Df = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]\Bigg|_{(X(s,t),Y(s,t))} \begin{bmatrix} \frac{\partial X}{\partial s} & \frac{\partial X}{\partial t} \\ \frac{\partial Y}{\partial s} & \frac{\partial Y}{\partial t} \end{bmatrix}.$$

It is *critical* that you evaluate Df at $g(s,t)$. This is a very common mistake, so make sure that you keep track of it!

Expanding Df, we have

$$Df = \left[\left.\frac{\partial f}{\partial x}\right|_{(X(s,t),Y(s,t))} \frac{\partial X}{\partial s} + \left.\frac{\partial f}{\partial y}\right|_{(X(s,t),Y(s,t))} \frac{\partial Y}{\partial s}, \left.\frac{\partial f}{\partial x}\right|_{(X(s,t),Y(s,t))} \frac{\partial X}{\partial t} + \left.\frac{\partial f}{\partial y}\right|_{(X(s,t),Y(s,t))} \frac{\partial Y}{\partial t} \right],$$

which gives us the same partials that we would have found if we used the shortcut.

□

However, note that in this form, it is harder to remember to evaluate the terms in the right way. This is why I recommend that you compute the Jacobian for every function you need to differentiate for the chain rule and multiply them, even in the single variable case. It doesn't take too long and keeps you from making silly mistakes and from getting lost when working with many indices.

Let's see another example, which you may recognize from physics.

Example 4. Let $f(x, y, z) = x^2 - y^2 + z$ and let $x(r, \theta, \rho) = r\cos\theta$ and $y(r, \theta, \rho) = r\sin\theta$, and $z(r, \theta, \rho) = \rho$. What is $\partial f/\partial r$?

Solution. I'll do this via the shortcut and a dependence tree. Define $g(r, \theta, \rho) = (r\cos\theta, r\sin\theta, \rho)$. Then we have

$$\begin{aligned}
\frac{\partial f}{\partial r} &= \left.\frac{\partial f}{\partial x}\right|_{g(r,\theta,\rho)} \frac{\partial x}{\partial r} + \left.\frac{\partial f}{\partial y}\right|_{g(r,\theta,\rho)} \frac{\partial y}{\partial r} + \left.\frac{\partial f}{\partial z}\right|_{g(r,\theta,\rho)} \frac{\partial z}{\partial r} \\
&= 2r\cos^2\theta - 2r\sin^2\theta + 0 \\
&= 2r\cos(2\theta).
\end{aligned}$$

□

Show that this matches up with the answer you would have gotten by multiplying Jacobians and seeing what you get in the first entry of the Jacobian of the composition.