MA 1C RECITATION 04/19/12

1. 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 1.1 (Apostol Theorem 8.8)**.** *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 1.2. *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*

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

Note this is the same formula given in Theorem 8.8. 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.

Warning 1.3. If you use this method, I recommend that you argue somewhere that it's equivalent to the one given in the Theorem 8.8 and then cite it whenever you want to use it.

Remark 1.4. 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 1.5. 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 1. If $z = e^x \sin y$, where $x = st^2$ and $y = s^2 t$, find $\partial z/\partial s$.

Date: April 19, 2012.

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

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

□

2. 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 2.1. 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 just remember it's just the total derivative from last time. You give $D\mathbf{f}$ a direction, and it tells you how $\mathbf{f}$ changes in that direction.

3. 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 2. 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
$$Df = \left[\frac{\partial f}{\partial x}\bigg|_{(X(s,t),Y(s,t))}\frac{\partial X}{\partial s} + \frac{\partial f}{\partial y}\bigg|_{(X(s,t),Y(s,t))}\frac{\partial Y}{\partial s}, \frac{\partial f}{\partial x}\bigg|_{(X(s,t),Y(s,t))}\frac{\partial X}{\partial t} + \frac{\partial f}{\partial y}\bigg|_{(X(s,t),Y(s,t))}\frac{\partial Y}{\partial t}\right],$$
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 3. 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
$$\begin{aligned}\frac{\partial f}{\partial r} &= \frac{\partial f}{\partial x}\bigg|_{g(r,\theta,\rho)}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\bigg|_{g(r,\theta,\rho)}\frac{\partial y}{\partial r} + \frac{\partial f}{\partial z}\bigg|_{g(r,\theta,\rho)}\frac{\partial z}{\partial r} \\ &= 2r\cos^2\theta - 2r\sin^2\theta + 0 \\ &= 2r\cos(2\theta).\end{aligned}$$
□

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.

4. Critical Points

We now return to the scalar field case. Last time we saw that if a point is a local extremum of a scalar function f, then the gradient of f is zero at that point. However, the converse statement is not true. Consider $f(x,y) = x^2 - y^2$ at the point $(0,0)$. Since we are always interested in finding extrema, even if we have to search for them among a given set of points, we give these points a name. A point where the gradient of f is zero is called a **stationary point** (or **critical point**) of f.

Why are these called "stationary points"? I think the definition makes the most sense if you think about the problem physically. Say f measures the temperature in a certain $\mathbf{R}^n$ space. Recalling our reasoning on the last problem from last week, we saw that the gradient points in the *direction of greatest change*, in other words, it will always point to the locally "hottest point." Suppose you take the gradient of a scalar field at point p, then move a tiny bit in the direction the gradient is pointing at p, then evaluate the gradient and repeat the process. If the temperature function is bounded, then by following this process, you will eventually will arrive at the "hottest point," and since you are at the hottest point, no direction will point to a hotter point, and so the gradient will be zero, and you will remain stationary.

This thinking is slightly misleading in general, since it applies to "coldest points" as well, but it gives you an idea behind the terminology. Actually, for functions $f(x, y)$ in two variables, the stationary points correspond to peaks, pits, and saddle points. In Math 2a, you will study how points converges to these stationary points, which is a fascinating subject in itself.

To study stationary points, we need to look at higher derivatives, much like how we studied maxima in the one-variable case by looking at the second-derivatives at that point. To do this, we introduce the following important object.

Definition 4.1. The **Hessian** matrix of a scalar field $f : \mathbf{R}^n \to \mathbf{R}$ is defined to be

$$H(\mathbf{x}) = [D_{ij}f(\mathbf{x})]_{i,j=1}^n$$

Remark 4.2. Note that

$$H(f)(x) = J(\nabla f)(x).$$

In particular, the Hessian matrix is symmetric, so there exists a basis of $\mathbf{R}^n$ consisting of the eigenvectors of H.

Observe that if $\mathbf{x}$ is an eigenvector of H, then the sign of the derivative in the direction of $\mathbf{x}$ is $\mathbf{x}H(\mathbf{a})\mathbf{x}^T$. Therefore, the sign of the derivative of f in the various directions at $\mathbf{a}$ corresponds with the signs of the eigenvalues of H. In particular, we have three cases.

- (H is *negative definite*) If all the eigenvalues of H are negative, then f has a maximum.
- (H is *positive definite*) If all the eigenvalues of H are positive, then f has a minimum.
- If there are eigenvalues of both signs, then f has a saddle point.

Example 4. Find and classify the critical points of $f(x, y) = x^2 + y^2$.

Solution. We have $\nabla f = (2x, 2y)$. Therefore, the only critical point of f is $(0, 0)$. Computing the Hessian, we find that

$$H_f = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

Therefore, the eigenvalues of H_f are 2 and 2, so H_f is positive definite, implying that $(0, 0)$ is a local minimum. $\square$

Critical points are subtle objects. They may not always behave like you'd imagine.

For instance, $f(x, y, z) = \sin(x^2 + y^2 + z^2)$ is an example that shows that critical points don't need to be isolated, since every point in the 3-sphere is a critical point of f. (Work it out!)

A function may also have no critical points whatsoever. Consider the (slightly modified) Gaussian function $f(x, y) = \int_x^y e^{-t^2}\, dt$. Then $\nabla f(x, y) = (-e^{-x^2}, e^{-y^2})$, which is never zero.

Warning 4.3. To find extrema for a function restricted to a region, you must check the critical points *and the boundary points*, just like in the single-variable case. This is because if a function is defined on a region, you may have global extrema that are not critical points.

4.1. **How to find extrema of multivariable functions.** This is for your last homework problem this week, where you have to find and classify all extrema and saddle points.

(a) Find and classify all critical points in the defined region and classify them as local/relative extrema or saddle points by looking at the Hessian at that point.
(b) Find the extrema on the boundary by evaluating the function at the boundary points, and see if any of these are global extrema (larger or smaller than your local extrema found above). Since the gradient may not be zero here, you don't have to worry about finding relative extrema or saddle points.
(c) Neatly and clearly describe all the local extrema, the saddles, and the global extrema on the interior of the region defined, as well as the global extrema that may lie on the boundary.