MA1C ANALYTIC RECITATION 4/19/12

ALDEN WALKER

1. Recall: Derivatives in General

The derivative of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ has exactly the same definition as for scalar valued functions, as does the definition of differentiability, e.g.

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

etc. It is convenient to split up into component functions, so $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \cdots, f_m(\mathbf{x}))$. Now instead of the gradient, the total derivative expressed in terms of the standard basis is the Jacobian:

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

This may seem complicated, but think about this: $\mathbf{f}$ eats things in $\mathbb{R}^n$ and spits out things in $\mathbb{R}^m$. If you give $D\mathbf{f}(\mathbf{a})$ a direction in $\mathbb{R}^n$, it will tell you how the image of $\mathbf{a}$ under $\mathbf{f}$ is changing in that direction.

2. The Chain Rule in General

The chain rule in general is about as nice as it could possibly be: 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 the multiplication is just matrix multiplication. Looking again at the chain rule from before, we see that that was a special case of this.

2.1. **Note.** Unless you are way better than I am at keeping track of indices (which is not unlikely, actually), I would recommend computing the entire Jacobian of the functions you need to differentiate for the chain rule and multiplying them, even if you are only asked for a single derivative. It's usually not too hard, and it keeps you from having to worry about lots of mess.

2.2. **Thinking About It.** That formula is totally fine. However, I think that it might be helpful to imagine the chain rule in the following way. Say f depends on x and y, and x and y in turn depend on t (through $\mathbf{r}$). Therefore, to take the derivative of f with respect to t, we sum up over all the ways in which f is affected by t, using the chain rule for each. That is,

$$\frac{\partial f}{\partial t} = \left.\frac{\partial f}{\partial x}\right|_{x(t)} \frac{\partial x}{\partial t} + \left.\frac{\partial f}{\partial y}\right|_{y(t)} \frac{\partial y}{\partial t}$$

Notice that this is the same formula as before. It may help to draw a dependency tree.

2.3. **Example.** If some of the spaces involved are one-dimensional ($\mathbb{R}$), then things become easier, since the gradient appears. If $\mathbf{r} : \mathbb{R} \to \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$, and $g = f \circ \mathbf{r}$, then $g'(t) = \nabla f(\mathbf{a}) \cdot \mathbf{r}'(t)$, where $\mathbf{a} = \mathbf{r}(t)$.

Let $\mathbf{r}(t) = (a\cos(t), a\sin(t))$, let $f = x^2 - y^2$, and $g = f \circ \mathbf{r}$. Then $\nabla f(\mathbf{r}(t)) = (2x, -2y)|_{\mathbf{r}(t)} = (2a\cos(t), -2a\sin(t))$. We also have $\mathbf{r}'(t) = (-a\sin(t), a\cos(t))$. Therefore,

$$\begin{aligned} g'(t) &= \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t) \\ &= (2a\cos(t), -2a\sin(t)) \cdot (-a\sin(t), a\cos(t)) \\ &= -2a^2\cos(t)\sin(t) - 2a^2\cos(t)\sin(t) \\ &= -2a^2\sin(2t) \end{aligned}$$

If you picture in your head the saddle and then the path that $g(t)$ takes, you will see that this answer is intuitively correct.

2.4. **Example.** Say we have $f(x,y) : \mathbb{R}^2 \to \mathbb{R}$, and we set $x = X(s,t)$ and $y = Y(s,t)$. Then f is a function of s and t–let's symbolically find $\partial f/\partial s$ and $\partial f/\partial t$. We have

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

and we define $g : \mathbb{R}^2 \to \mathbb{R}^2$ by $g(s,t) = (X(s,t), Y(s,t))$, so

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

and thus:

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

Keeping track of the fact that Df is evaluated at $g(s,t)$ is the tricky part I think. We expand out to get

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

Which gives us the partials desired. Note that if we draw out a dependence tree, we would get the same thing.

2.5. **Example.** Define $f(x,y,z) = x^2 - y^2 + z$, and let $x(r,\theta,p) = r\cos\theta$ and $y(r,\theta,p) = r\sin\theta$ and $z(r,\theta,p) = p$. Let's find $\partial f/\partial r$.

There are two ways, as above. First, we can make a dependence tree, so letting $g(r,\theta,p) = (r\cos\theta, r\sin\theta, p)$, we have

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

Or, we could see

$$Df = \begin{bmatrix} 2x & -2y & 1 \end{bmatrix} \quad \text{and} \quad Dg = \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\sin\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

So

$$\begin{aligned} D(f \circ g)(r,\theta,p) &= Df(g(r,\theta,p))Dg(r,\theta,p) \\ &= \begin{bmatrix} 2r\cos\theta & -2r\sin\theta & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 2r\cos(2\theta) & -4r^2\cos\theta\sin\theta & 1 \end{bmatrix} \end{aligned}$$

Notice that $\frac{\partial f}{\partial r}$ is the left hand entry in the matrix, which agrees with what we got before.

3. Tangent Spaces

If f is differentiable at a point, then we can define what we mean by the tangent space. However, note that we define the tangent space to a level set, not to the function itself. The notation used for the tangent space isn't standard, I don't think, but I'm not sure that the book defines any notation.

Suppose that f is differentiable at $\mathbf{a} \in L_c(f)$. If $\nabla f(\mathbf{a}) \neq 0$, then the tangent space $\theta_{\mathbf{a}}(L_c(f))$ to $L_c(f)$ at $\mathbf{a}$ is

$$\theta_{\mathbf{a}}(L_c(f)) = \{\mathbf{x} \in \mathbb{R}^n \mid \nabla f(\mathbf{a}) \cdot \mathbf{x} = 0\}$$

Notice that the tangent space lives in $\mathbb{R}^n$, and not on the graph of the function! To see that that is the correct formula, remember that the gradient points in the direction of greatest increase. Therefore, if you are perpendicular to the gradient, you are a direction of 0 increase. Therefore, going in any direction in the plane $\nabla f(\mathbf{a}) \cdot \mathbf{x} = 0$ keeps you inside the level set because there is no change in the function.

3.1. **Example.** Let $f(x, y) = x^2 + y^2$. Then the level set $f(x, y) = 1$ is the circle of radius 1 in $\mathbb{R}^2$. At $(1, 0)$, we have $\nabla f(1, 0) = (2, 0)$, so $\theta_{(1,0)}(L_1(f)) = \{\mathbf{x} \mid (2, 0) \cdot \mathbf{x} = 0\}$, otherwise known as a vertical line.

3.2. **Example (What you might have thought the tangent space is).** Suppose you want the tangent plane to the *graph* of $f(x, y) = x^2 + y^2$. Then you have to do this trick. Set $g(x, y, z) = x^2 + y^2 - z$, and consider the level set $L_0(g)$. Note that this level set is exactly where $x^2 + y^2 = z$, i.e. it is the graph of f. Then the gradient of g at, say, $(1, 1, 2)$ is $(2, 2, -1)$. Therefore $\theta_{(1,1,2)}(L_0(g)) = \{\mathbf{x} \in \mathbb{R}^3 \mid (2, 2, -1) \cdot \mathbf{x} = 0\}$. A picture is helpful. Notice that the vector $(2, 2, -1)$ points straight out from the graph of f, as desired.

3.3. **Note.** It will be helpful on your homework to use the preceding trick, for example when finding a direction vector normal to a surface.

4. Critical Points

A place where the gradient is zero is called a stationary point of the function. To explore the nature of the stationary point (it might not be a local extreme point), we must look at higher order derivatives (just like the 2nd derivative test in one dimension).

The Hessian matrix of f is defined to be $H(\mathbf{x}) = [D_{ij} f(\mathbf{x})]_{i,j=1}^n$, and it is important because if f has continuous second partial derivatives in a ball about $\mathbf{a}$, then for $\mathbf{y}$ such that $\mathbf{a} + \mathbf{y} \in B(\mathbf{a})$ we have

$$f(\mathbf{a} + \mathbf{y}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{y} + \frac{1}{2}\mathbf{y}H(\mathbf{a})\mathbf{y}^T + \|\mathbf{y}\|^2 E_2(\mathbf{a}, \mathbf{y})$$

The last term is just an error term—don't worry about it. The point is that if you rearrange that you see that the sign of the derivative in the direction $\mathbf{y}$ is $\frac{1}{2}\mathbf{y}H(\mathbf{a})\mathbf{y}^T$. The reason this is useful is that the Hessian matrix, under the above assumptions, is symmetric (I would say it is self-adjoint, because the following theorem applies to that more general case), which means that there is a basis of $\mathbb{R}^n$ consisting of eigenvectors of H. If $\mathbf{y}$ is an eigenvector, then the sign of $\frac{1}{2}\mathbf{y}H(\mathbf{a})\mathbf{y}^T$ is the sign of the eigenvalue of $\mathbf{y}$. This means that 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,

- If all the eigenvalues of H are negative, then f has a maximum.
- If all the eigenvalues of H are positive, then f has a minimum.
- If there are eigenvalues of both signs, f has a saddle.

4.1. **Example.** Let's look at the critical points of $f(x, y) = x^2+y^2$. First, we find the gradient $\nabla f = (2x, 2y)$. Clearly, the only place where this is 0 is at $(0, 0)$. To see what kind of stationary point this is, we compute the Hessian:

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

The eigenvalues of which are clearly $(2, 2)$. Therefore this matrix is positive definite, and $(0, 0)$ is therefore a local minimum.

The same calculation with the function $f(x, y) = x^2 - y^2$ yields the Hessian $H_f = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$, showing that the stationary point $(0, 0)$ is a saddle for f.

4.2. **Example.** Let $f(x, y, z) = \sin(x^2 + y^2 + z^2)$. Then $\nabla f = (2x\cos(x^2 + y^2 + z^2), 2y\cos(x^2 + y^2 + z^2), 2z\cos(x^2+y^2+z^2))$. For this to be $(0, 0, 0)$, we must have either $(x, y, z) = (0, 0, 0)$ or $\cos(x^2+y^2+z^2) = 0$; that is, we must have $x^2 + y^2 + z^2 = (2n + 1)\frac{\pi}{2}$. Notice that every point in the spheres is a critical point. This example shows that critical points need not be isolated.

4.3. **Example.** Let $f(x, y) = \int_x^y e^{-t^2} dt$. Then $\nabla f(x, y) = (-e^{-x^2}, e^{-y^2})$. Not hard, but I just want to remind you that you know how to differentiate many things. Notice that this function has no critical points.

4.4. **Note On Homework.** Finding extreme points for a function restricted to a region is *not* the same as finding extreme points and saddles for a function which happen to lie in a region! In the former case, you must do the boundary separately since you might find maxima and minima which aren't stationary points. Here is what you need to know in order to do your homework problem about this:

(1) Find all stationary points in the interior and classify them as relative extrema/saddle points
(2) Find all extreme points on the boundary by hand (restrict the function to the boundary and do a one-dimensional search)
(3) Saddle points and relative extrema cannot, by definition, occur on the boundary. Record only those extrema on the boundary which are global extreme points.
(4) You solution should contain (in addition to your work!) all relative/global extrema and saddle points on the interior, and all global extrema on the boundary.

By "global extrema", I mean global extrema for the function restricted to the region.