Lecture 3-2

Temporarily skipping Chapter 15 of the text, we proceed directly to Chapter 16, in which the derivative of a real-valued function of n variables is (finally) defined. Consider the simplest case of a function f from $\mathbb{R}^2$ to $\mathbb{R}$. What should the derivative of f at a point (a,b) be? The most straightforward approach would be to define it as the limit $\lim_{(x,y)\to(a,b)}(f(x,y)-f(a,b))/(x-a,y-b)$, but this makes no sense as we cannot divide by a vector. We recall at this point that the function $g(x,y)$ defined to be $xy/(x^2+y^2)$ if $(x,y)\neq(0,0)$ and 0 if $(x,y)=(0,0)$ must not be differentiable at 0 (even though both partials of this function exist at $(0,0)$), since this function is not even continuous at $(0,0)$.

For a general $f:\mathbb{R}^2\to\mathbb{R}$ and a point $(a,b)\in\mathbb{R}^2$ such that f is defined at all points sufficiently close to (a,b), we have $f(a+h,b+k)-f(a,b)=f(a+h,b+k)-f(a,b+k)+f(a,b+k)-f(a,b)$, which is approximately equal to $hf_x(a,b+k)+kf_y(a,b)$. We therefore agree to call f *differentiable* at (a,b) in this situation if its partials exist at (a,b) *and* for sufficiently small h,k the difference $d(h,k)=f(a+h,b+k)-f(a,b)-hf_x(a,b)-kf_y(a,b)$ is "small enough" in the sense that $\lim_{(h,k)\to(0,0)} d(h,k)/||(h,k)||=0$. Thus for example the function $g(x,y)$ defined above is indeed not differentiable at $(0,0)$, since here $d(h,k)=hk/(h^2+k^2)$ (for h,k not both 0) and $\frac{hk}{(h^2+k^2)\sqrt{h^2+k^2}}\not\to 0$ as $(h,k)\to(0,0)$. On the other hand, the function $f(x,y)=(x^2+y^2)\sin(1/\sqrt{x^2+y^2})$ for $(x,y)\neq(0,0), f(0,0)=0$, is differentiable at $(0,0)$: its partial derivatives are both easily seen to be 0 there and $\lim_{(h,k)\to(0,0}\frac{(h^2+k^2)\sin(1/\sqrt{h^2+k^2})}{\sqrt{h^2+k^2}}=0$, as desired. This last function is the two-variable analogue of the function $h(x)=x^2\sin(1/x)$ for $x\neq 0, h(0)=0$, that we saw in the fall, whose derivative exists everywhere but is discontinuous at 0; our function $f(x,y)$ is differentiable at $(0,0)$ but its differential (defined below) is not. In words, we can say that the partial derivatives of a function f at a point $\vec{a}$ see only how that function changes in the coordinate directions, so their existence is not sufficient even to guarantee continuity at that point. On the other hand, if the partials are also continuous at $\vec{a}$, then we will see below that that condition gives us enough control over the behavior of $f(\vec{x})$ as $\vec{x}$ approaches $\vec{a}$ from any direction to say that f is differentiable there. We call a function f with continuous partials at a point $\vec{a}$ *continuously differentiable* there. It is easy to check that any function differentiable at a point $\vec{a}$ is also continuous there.

We also see that that there is a fundamental difference between derivatives of real-valued functions of one variable, regarded as rates of change, and their counterparts for $n>1$ variables: *the derivative of a function on $\mathbb{R}^n$ for $n>1$, even at a single point, is too complicated to be describable by a single real number.* If $f:\mathbb{R}^n\to\mathbb{R}$ is differentiable at $\vec{a}=(a_1,\ldots,a_n)$, and if $\nabla f(\vec{a})=(\partial f/\partial x_1,\ldots,\partial f/\partial x_n)(\vec{a})$, then the derivative (actually usually called the *differential*) of f at $\vec{a}$ is the *function* taking $\vec{h}=(h_1,\ldots,h_n)\in\mathbb{R}^n$ to the dot product $\nabla f(\vec{a})\cdot\vec{h}=\sum\partial f/\partial x_i(\vec{a})h_i$. Speaking loosely, we may (and often do) identify this function with the vector $\nabla f(\vec{a})$ that determines it. We can describe the differential of f at $\vec{a}$, regarded as a function of the difference vector $\vec{h}$, as the closest linear approximation to $f(\vec{a}+\vec{h})-f(\vec{a})$.

For a vector-valued function $\vec{f}=(f_1,\ldots,f_m):\mathbb{R}^n\to\mathbb{R}^m$ we have that $\vec{f}$ is differentiable at $\vec{a}\in\mathbb{R}^n$ (by definition) if and only if each coordinate function f_i is differentiable there. The differential of $\vec{f}$ at $\vec{a}$ then identifies with a "stack" of m vectors in $\mathbb{R}^n$. Such a stack (i.e. a rectangular array of numbers, said to be of size $m \times n$ in this case because it has m rows and n columns) is called an $m \times n$ *matrix*. We will later see that any $m \times n$ matrix may be multiplied by any column vector $\vec{v} \in \mathbb{R}^n$ by taking the successive dot products of its rows with $\vec{v}$ and arranging the results as a column vector in $\mathbb{R}^m$. This is how we compute the differential of $\vec{f}$ at $\vec{a}$, regarding elements of the domain and range of this differential as column vectors in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively.

Now we can show that continuous differentiability implies differentiability: *if ∇f exists and is continuous in a neighborhood of $\vec{a}$, then f is differentiable there.* Indeed, using the mean value theorem n times, we have that a typical difference $f(\vec{a}+\vec{h}) - f(\vec{a})$ can be written as $h_n(\partial f/\partial x_n(a_1+h_1,\ldots,a_{n-1}+h_{n-1},a'_n)) + h_{n-1}(\partial f/\partial x_{n-1}(a_1+h_1,\ldots,a'_{n-1}, a_n+h_n)) + \ldots + h_1 \partial f/\partial x_1(a'_1, a_2+h_2,\ldots,a_n+h_n)$, for some a'_n in (a_n, a_n+h_n), a'_{n-1} in $(a_{n_1}+h_{n-1})$, and so on, where $\vec{h} = (h_1,\ldots,h_n)$. Subtracting $\nabla f(\vec{a})\cdot\vec{h}$ and dividing by $||\vec{h}||$, we get a sum of n products, all of them involving one quantity of absolute value less than 1 times another quantity going off to 0 as $\vec{h} \to \vec{0}$, so the limit of $D(\vec{h})/||\vec{h}||$ is 0 as $\vec{h} \to \vec{0}$, as desired. Given the function $f(x,y) = (x^2+y^4)/(x^2+y^2)$ for $(x,y) \neq (0,0)$, $f(0,0) = 0$, we can use either this criterion to show that f is differentiable at $(0,0)$ or check this directly from the definition.

There are useful formulas for the gradients of particular functions often arising in physics. Recall the standard notation r introduced earlier for the length of the vector $\vec{r} = (x,y,z)$. Then one can check immediately that $\nabla r = (1/r)\vec{r}$. In fact, we have the more general formula $\nabla f(r) = (f'(r)/r)\vec{r}$; we will prove this later, using the chain rule for multivariable functions.