Lecture 3-6

We conclude the new material this term with a proof that the mixed partials of a function at a point are equal whenever both are continuous, following §15.6 of the text; next week will be spent entirely on review.

Let $f : \mathbb{R}^2 \to \mathbb{R}$ have second-order partials $f_{xy} = \partial/\partial y(\partial f/\partial x), f_{yx} = \partial/\partial x(\partial f/\partial y)$ existing in a neighborhood of $\vec{a} \in \mathbb{R}^2$ and suppose that f_{xy}, f_{yx} are continuous at $\vec{a}$. For nonzero h, k, consider the second-order difference $D(h,k) = f(a+h, b+k) - f(a, b+k) - f(a+h, b) + f(a,b)$. On the one hand, we have $D(h,k) = G(a+h) - G(a)$, where $G(x) = f(x, b+k) - f(x,b)$; applying the mean value theorem first to $D(h,k)$ and then to $G(x)$, we get $D(h,k) = hkf_{yx}(a', b')$ for some a', b' respectively between $a, a+h$ and $b, b+k$. On the other hand, we also have $D(h,k) = H(b+k) - H(b)$, where $H(y) = f(a+h, y) - f(a,y)$. Applying the mean value theorem first to $D(h,k)$ and then to $H(y)$, we get $D(h,k) = hkf_{xy}(a'', b'')$ for some a'', b'' respectively between $a, a+h$ and $b, b+k$. Dividing by hk, taking the limit as $h, k \to 0$, and using the continuity of f_{xy} and f_{yx} at $\vec{a}$ at the last step, we get $f_{xy}(\vec{a}) = f_{yx}(\vec{a})$, as desired. The result generalizes immediately to functions of more than two variables: *in taking a higher-order partial derivative of a function at a point, it makes no difference in what order the variables occur provided that all partial derivatives in question are continuous.* This holds simply because in interchanging two variables and taking partial derivatives with respect to them, all other variables are treated as constants anyway, so it makes no difference whether they are present or not.

This being a course where we are not afraid of skeletons in the closet, we give the standard example of a function whose mixed partials are not equal at a point. Take $f(x,y) = xy\frac{x^2-y^2}{x^2+y^2}$ for $(x,y) \neq (0,0), f(0,0) = 0$. I claim that $f_{xy}(0,0) \neq f_{yx}(0,0)$. To prove this it is essential that we *not* use the quotient rule, which leads to a big mess and doesn't apply at $(0,0)$ anyway, since the formula for $f(x,y)$ at other points makes no sense at $(0,0)$. Instead we compute $f_x(0,k)$ for all k directly from the definition, getting $f_x(0,k) = \lim_{h\to 0}(f(h,k) - f(0,k))/h = \lim_{h\to 0} \frac{hk\frac{h^2-k^2}{h^2+k^2}}{h} = -k, f_{xy}(0,k) = -1$, for all k. Similarly $f_y(h,0) = \lim_{k\to 0}(f(h,k) - f(h,0))/k = \lim_{k\to 0} \frac{\frac{hk(h^2-k^2)}{h^2+k^2}}{k} = h$, so $f_{yx}(h,0) = 1$ for all h. In particular, $f_{xy}(0,0) = -1 \neq 1 = f_{yx}(0,0)$, as claimed.

What is really going on in this last example is that the function f is differentiable but not twice differentiable at the point $(0,0)$. If you were to guess the definition of twice differentiability at a point, you would probably get some of it (but I very much doubt all of it) right. A function $g : \mathbb{R}^2 \to \mathbb{R}$ is said to be twice differentiable at $(a,b) \in \mathbb{R}^2$ if the first- and second-order partials $g_x, g_y, g_{xx}, g_{xy}, g_{yx}$, and g_{yy} all exist at that point, if $g_{xy}(a,b) = g_{yx}(a,b)$, and finally $\lim_{(h,k)\to(0,0)} \frac{g(a+b,h+k)-g(a,b)-\nabla g(a,b)\cdot(h,k)-((1/2)g_{xx},g_{xy},(1/2)g_{yy})(a,b)\cdot(h^2,hk,k^2)}{h^2+k^2} = 0$. Why on earth do we see the factors of $1/2$ here before the values of g_{xx} and g_{yy} at (a,b)? A partial answer is that even for a homogeneous quadratic function $h(x,y) = ax^2 + bxy + cy^2$, the constant b equals the value of h_{xy} at $(0,0)$, but the constants a and c are *half* the values of h_{xx}, h_{yy} at $(0,0)$. A more complete answer is that if we fix h,k as well as a,b and look at the Taylor series expansion of a general $g(a+th, b+tk)$, regarded as a function of the single variable t, then the second derivative term appears with coefficient $1/2$ in this expansion; since by the chain rule there are two terms contributing to the hk term in this expansion, it winds up having coefficient $g_{xy}(0,0)$, while the coefficients of h^2, k^2 are $(1/2)g_{xx}(a,b), (1/2)g_{yy}(a,b)$, respectively. We call the sum $\nabla g(a,b)\cdot(h,k) + ((1/2)g_{xx}(, g_{xy}, g_{yy})(a,b)\cdot(h^2, hk, k^2)$ the *quadratic approximation* to the difference $g(a+h, b+k) - g(a,b)$; it involves three more terms than the linear approximation of this difference, but achieves a greater accuracy, being small compared to h^2+k^2 rather than just to $\sqrt{h^2+k^2}$. A function f with continuous first- and second-order partials at a point is always twice differentiable there and has equal mixed partials at that point; the preceding argument shows that in fact for the mixed partials to be equal, one needs only that both be continuous at that point; one does not need the unmixed partials f_{xx}, f_{yy} even to exist at that point. In fact it turns out that even if only one of the mixed partials f_{xy} or f_{yx} is continuous at (a,b), but both mixed partials exist there, then again they must be equal.

As you might expect from the preceding paragraph, linear and quadratic approximations are just the tip of the iceberg: any function with continuous partials at $\vec{a}\in\mathbb{R}^m$ of order up to n (where we compute the order of a partial derivative by counting how many times we differentiate with respect to each variable and then add the results) is n times differentiable at that point, where this last condition means that $f(\vec{a}+\vec{h}) - f(\vec{a})$ is approximated by a polynomial of degree at most n in the coordinates $h_1,\ldots,h_n$ of $\vec{h}$, with an error which when divided by $||\vec{h}||^n$ goes to 0 as $\vec{h}\to\vec{0}$. A typical coefficient of a monomial in the h_i in this approximation is the corresponding partial of f evaluated at $\vec{a}$ divided by a product of factorials, one for each power of every variable in the monomial. The only thing that matters about each partial is the number of times differentiation with respect to each variable in turn takes place; the order in which the differentiations are performed is irrelevant. The sum of $f(\vec{a})$ and the approximation is called the *Taylor polynomial of order n for f at $\vec{a}$.*