Analysis II: Higher Derivatives and Taylor's Theorem

Jesse Ratzkin

October 14, 2009

In this section of notes we discuss second and higher derivatives of a function of several variables. First, recall that if $f : \mathbb{R}^n \to \mathbb{R}^m$ and $x_0 \in \mathbb{R}^n$ then f is differentiable at x_0 if there is a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{x \to x_0} \frac{|f(x) - f(x_0) - A(x - x_0)|}{|x - x_0|} = 0.$$

We call this linear transformation the derivative $Df|_{x_0}$. Now suppose $U \subset \mathbb{R}^n$ is open and $f : U \to \mathbb{R}^m$ is C^1, *i.e.* has a derivative which varies continuously in U. This derivative is now a function

$$U \ni x \mapsto Df|_x \in L(\mathbb{R}^n, \mathbb{R}^m) = \mathbb{R}^{mn},$$

and so it makes sense to see if Df itself has a derivative. If it exists, this derivative will now be a linear map $D^2 f : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^m) = \mathbb{R}^{mn}$. We can clarify some of the notation by using the bilinear maps we introduced in the last set of notes.

Let V, W, Z be vector spaces, and denote the space of bilinear maps $\beta : V \times W \to Z$ as $L(V, W; Z)$.

Lemma 1. *The spaces $L(V, W; Z)$, $L(W, V; Z)$, $L(V, L(W, Z))$, and $L(W, L(V, Z))$ are all naturally isomorphic (that is, the isomorphism doesn't depend on any choices). The common dimension of these vector spaces is* $\dim(V)\dim(W)\dim(Z)$.

Notice that(again) the order of operations is important. If we restrict our attention to symmetric bilinear transformations we don't have to pay attention to the order, though.

Proof. To $\beta \in L(V, W; Z)$ we have already associated the linear transformation $T_\beta : V \to L(W, Z)$, where $(T_beta(v))(w) = \beta(v, w)$. We define the inverse map by associating a bilinear map β_T to a linear transformation $T : V \to L(W, Z)$. In fact, $\beta_T(v, w) = (T(v))(w)$. By construction, $\beta_{T_beta} = \beta$ for all $\beta \in L(V, W; Z)$ and $T_{\beta_T} = T$ for all $T \in L(V, L(W, Z))$, and so we've just shown $L(V, W; Z)$ is naturally isomorphic to $L(V, L(W, Z))$. Permuting the arguments of the bilinear and linear maps, the same line of reasoning shows $L(W, V; Z)$ is naturally isomorphic to $L(W, L(V, Z))$. Finally, if $\beta \in L(V, W; Z)$ we define $\hat{\beta} \in L(W, V; Z)$ by $\hat{\beta}(v, w) = \beta(w, v)$, which is an involution. (An involution is an isomorphism which is its own inverse.)

All that remains is to count the dimension of all four of these vector spaces. Choose bases $\{E_i\}$, $\{F_j\}$, and $\{G_k\}$ (respectively) for V, W, and Z. Write $v \in V$ as $v = \sum v_i E_i$ and $w \in W$ as $w = \sum w_j F_j$. Now define the bilinear maps $\beta_{ijk} \in L(V, W; Z)$ by $\beta(v, w) = v_i w_j G_k \in Z$; these form a basis of $L(V, W; Z)$. $\square$

Recall we defined the norm of the bilinear form β by

$$\|\beta\| = \sup\{|\beta(v, w)| : |v| = 1 = |w|\} = \|T_\beta\|,$$

and so these isomorphisms in fact preserve the norms of all the spaces.

Corollary 2. *The dimension of the space of bilinear maps $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ is $n^2 m$.*

In fact, we can define k-linear maps, where $k \in \mathbb{N}$, as those maps which take in k vectors and are linear in each factor. By induction on the lemma above and its corollary, the space of k-linear maps $\mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}^m$ is $n^k m$. (Here there are k factors of $\mathbb{R}^n$ in the domain of the k-linear map.)

Another fact about symmetric bilinear transformations (and symmetric k-linear transformations in general) will be useful. A bilinear transformation β is symmetric if $\beta(v, w) = \beta(w, v)$ for all possible pairs v and w. Such a bilinear transformation is uniquely determined by its action on the diagonal. In other words, a symmetric bilinear transformation is determined by the values $\beta(v, v)$ for all possible v. Indeed, if β is symmetric, then

$$\beta(v - w, v - w) = \beta(v, v) - 2\beta(v, w) + \beta(w, w),$$

which implies

$$\beta(v, w) = \frac{1}{2}(\beta(v, v) + \beta(w, w) - \beta(v - w, v - w)).$$

By the same reasoning (or induction) a symmetric k-linear transformation $\beta : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}^m$ is determined by its action on the diagonal, that is $\beta(v, v, \ldots, v)$ for all $v \in \mathbb{R}^n$.

Now let's return to the question of second and higher order derivatives. Let's say $f : \mathbb{R}^n \to \mathbb{R}^m$ is a C^1 function; then $Df : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^m)$ is continuous, and we can ask if it has a linear approximation at a basepoint $x_0 \in \mathbb{R}^n$. Such a linear approximation would be a linear map

$$A : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^m)$$

which satisfies

$$\lim_{x \to x_0} \frac{Df|_x(w) - Df|_{x_0}(w) - (A(x - x_0))(w)}{|x - x_0|} = 0 \tag{1}$$

for **all** $w \in \mathbb{R}^n$. Using the identification $L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)) = L(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R}^m)$ above (and abusing notation just a little) we can write $D^2 f|_x$ as the bilinear transformation which best approximates the function Df, base at x. The limit estimate above says $D^2 f|_x(v, w)$ is smaller than $|v||w|$, so, in particular, we obtain the second order Taylor estimate

$$\lim_{v \to 0} \frac{|f(x+v) - f(x) - Df|_x(v) - \frac{1}{2} D^2 f|_x(v, v)|}{|v|^2} = 0. \tag{2}$$

The factor of $1/2$ in front of the second derivative term in the equation (2) occurs because $Df|_x$ appears twice in equation (1).

By induction on the number of derivatives, we prove the following Taylor series expansion for function of several variables.

Theorem 3. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ have k derivatives at the point x. Then the kth order Taylor remainder $r_k(v)$ defined by*

$$\begin{aligned} f(x+v) &= f(x) + Df|_x(v) + \frac{1}{2} D^2 f|_x(v, v) + \cdots + \\ &\quad + \frac{1}{(k-1)!} D^{k-1} f|_x(v, v, \cdots, v) + \frac{1}{k!} D^k f|_x(v, v, \cdots, v) + r_k(v) \end{aligned} \tag{3}$$

satisfies

$$\lim_{v \to 0} \frac{|r_k(v)|}{|v|^k} = 0.$$

In this theorem, the derivative $D^l f$ of order l, for $l = 1, 2, \ldots, k$, is an l-linear map from $\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}^m$.

At this point, we write out the second derivative in terms of partial derivatives. As was the case for the first derivative, the existence of the second derivative implies the existence of the second partials, and in this case

$$\left. \frac{\partial^2 f_i}{\partial x_j \partial x_k} \right|_{x_0} = D^2 f_i|_{x_0}(e_j, e_k),$$

where f_i is the ith component of the function f. Indeed, we can borrow the proof in the previous set of notes for the first derivative and apply it to all higher derivatives. However, with second and higher order derivatives we have a choice of the order in which we take derivatives. Does the answer we get depend on this choice? In other words, is the second derivative a symmetric bilinear form?

Theorem 4. *If f has a second derivative it is a symmetric bilinear form, so that*

$$\frac{\partial^2 f_i}{\partial x_j \partial x_k} = \frac{\partial^2 f_i}{\partial x_k \partial x_j}$$

for all i, j, k. In other words,

$$D^2 f|_{x_0}(v, w) = D^2 f|_{x_0}(w, v)$$

for all $v, w \in \mathbb{R}^n$.

The same proof shows the kth order derivative is in fact symmetric in all its k entries.

Remark 1. *The usual proof one sees that the second derivative is symmetric actually requires that the second derivative varies continuously with the basepoint. The proof below does not assume continuity of the second derivative. We outline the (slightly) weaker proof in exercises below.*

Proof. Without loss of generality we can assume $f : \mathbb{R}^n \to \mathbb{R}$; the symmetry assertion we want to prove concerns only interchanging v and w, not on the values f can assume.

Fix a basepoint x in the domain of f and vectors $v, w \in \mathbb{R}^n$ and consider the function

$$F(t, v, w) = f(x + tv + tw) - f(x + tv) - f(x + tw) + f(x).$$

Notice that F is symmetric in v and w: $F(t, v, w) = F(t, w, v)$.

We will complete the proof by showing $\lim_{t\to 0} t^{-2}F(t, v, w) = D^2 f|_x (v, w)$. Then, because the thing inside the limit is symmetric in v and w, we conclude $D^2 f|_x (v, w) = D^2 f|_x (w, v)$ as we claimed.

For a given t, we will need an auxiliary function

$$g(s) = f(x + sv + tw) - f(x + sv).$$

Then $F(t, v, w) = g(t) - g(0)$. Also, because f is differentiable, g is also a differentiable function of the one variable s. So, by the mean value theorem for one variable, there is $\theta \in (0, t)$ such that $F(t, v, w) = tg'(\theta)$. However, we can also compute $g'(\theta)$ in terms of Df using the chain rule, obtaining

$$F(t, v, w) = tg'(\theta) = t[Df|_{x+\theta v+tw} (v) - Df|_{x+\theta v} (v).$$

Next we expand both the Df terms in Taylor series.

$$\begin{aligned}
\frac{F(t, v, w)}{t^2} &= \frac{1}{t}\left[Df|_x (v) + D^2 f\big|_x (v, \theta v + tw) + r(v, \theta v + tw)\right] \\
&\quad - \left[Df|_x (v) + D^2 f\big|_x (v, \theta v) + r(v, \theta v)\right] \\
&= D^2 f\big|_x (v, w) + \frac{r(v, \theta v + w)}{t} + \frac{r(v, \theta v)}{t}
\end{aligned}$$

The two remainder terms $r(v, \theta v+w)$ and $r(v, \theta v)$ are sublinear, so the corresponding ratios go to zero as $t \to 0$. This completes the proof. $\square$

We can apply all these ideas to compute higher order derivatives. For instance, the third derivative $D^3 f|_x$, should it exist, is a trilinear map from $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^m$. It is also fully symmetric, that is

$$D^3 f\big|_x (u, v, w) = D^3 f\big|_x (v, u, w) = D^3 f\big|_x (w, v, u),$$

and so on. In particular, if f has derivatives up to order k, then for $l \leq k$ the lth order mixed partial derivatives are equal.

We say a function f is of class C^k, for $k = 0, 1, 2, 3, \ldots$, if f has derivatives up to and including order k, and the kth derivative is continuous. Borrowing the proof we had in the last set of notes for first derivatives, we see that $f \in C^k$ if and only if all partial derivatives of f of order $l \leq k$ exist and are continuous. If f has derivatives to all orders, then we write $f \in C^\infty$.

Observe that the Taylor series expansion in (3) approximates a function f as a polynomial, just as in the one variable case. When does this series converge to f? Actually, the answer is the same as in the one variable case. Write out

$$\begin{aligned} f(x+v) &= f(x) + Df(v) + \frac{1}{2} Df(v,v) + \cdots + \frac{1}{k!} Df(v, v, \ldots, v) + \cdots \\ &= c_0 + c_1(v) + c_2(v,v) + \ldots, \end{aligned}$$

where in each term c_k is a symmetric k-linear form. Using the same estimates one does in the one variable case (with operator norms replacing absolute values), we see that the radius of convergence of this series is R where

$$R = \limsup_{k \to \infty} \|c_k\|^{1/k}.$$

So long as $|v| < R$, the series converges uniformly and absolutely, and we can differentiate it term by term. A function f which is given by a convergent power series is analytic, which we write as $f \in C^\omega$. The condition $f \in C^\omega$ is much stronger than the condition $f \in C^\infty$, which we can even see for functions of one variable. Let

$$f(x) = \begin{cases} e^{-1/x} & x > 0 \\ 0 & x \leq 0. \end{cases}$$

One can verify that all derivatives of f exist, even at $x = 0$. Indeed, the kth derivative at zero is $f^{(k)}(0) = 0$, so the Taylor polynomial (centered at 0) to any order is the zero polynomial. If f did have a convergent power series centered at 0, it would have to be the zero series, which would then mean f is identically equal to 0 in a neighborhood of 0. Which it isn't.

Here are some exercises to think about, written in no particular order.

1. Write down a function $f : \mathbb{R} \to \mathbb{R}$, with $f \in C^\infty$, such that $f(x) \geq 0$ for all x and
$$f(x) = \begin{cases} 1 & |x| < \frac{1}{2} \\ 0 & |x| > 1. \end{cases}$$
Such a function is called a bump function, and is useful in many applications.