

Differentiation done correctly: 2. Higher derivatives

Posted on February 22, 2013 by wj32 / 0 Comment

Navigation: 1. The derivative | 2. Higher derivatives | 3. Partial derivatives | 4. Inverse and implicit functions | 5. Maxima and minima

Last time, we covered the definition of the derivative and its basic properties, which all turn out to be quite similar to their single variable counterparts. Now we are going to explore higher derivatives. In traditional multivariable calculus, true higher derivatives do not exist (except in a specific situation which will be discussed in part 5). Of course, we have so-called “mixed/higher partial derivatives”, which are coordinate-dependent and notationally tricky to work with. As a consequence, the usual statement of Taylor’s theorem in \mathbb{R}^n ends up being ugly and hard to remember. In reality, Taylor’s theorem for Banach spaces looks almost exactly the same as the single variable Taylor’s theorem!

Recall that for any Banach space F , the space of continuous linear maps $L(E, F)$ is also a Banach space. If $A \subseteq E$ is an open set and $f: A \rightarrow F$ is differentiable, then $Df = f' : A \rightarrow L(E, F)$ is a map between Banach spaces. Therefore we may consider the second derivative

$$D^2f = f'' : A \rightarrow L(E, L(E, F))$$

obtained by differentiating f' . Continuing the process, we have higher order derivatives

$$D^pf = f^{(p)} : A \rightarrow L^p(E, F),$$

where $L^p(E, F) = L(E, L(E, \dots, L(E, F)\dots))$. It is clear that $D^p(f + g) = D^pf + D^pg$ and $D^p(cf) = cD^pf$ for all scalars c . We say that f is **of class C^p** or is **p times continuously differentiable** if $D^pf(x)$ exists for all $x \in A$ and D^pf is continuous on A . Note that if f is of class C^p , then D^kf is automatically continuous for all $0 \leq k < p$ as well. We identify $L^p(E, F)$ with the space of multilinear maps $L(E, \dots, E; F)$, and write $D^pf(x_1, \dots, x_p)$ for $D^pf(x_1)\cdots(x_p)$. Our definition of p times continuous differentiability is not the usual one that is stated in terms of mixed partial derivatives, but part 3 will make it clear that these definitions are equivalent. We have the following extension of [Theorem 8](#), which follows easily by induction.

Theorem 18. *Let $A \subseteq E$ be an open set, let F_1, \dots, F_n be Banach spaces, let $f: A \rightarrow F_1 \times \cdots \times F_n$ and let $f_i = \pi_i \circ f$ be the component functions of f , where $\pi_i: F_1 \times \cdots \times F_n \rightarrow F_i$ is the canonical projection. Then f is of class C^p if and only if every f_i is of class C^p . In that case we have $D^pf_i(x) = \pi_i \circ D^pf(x)$, i.e.*

$$D^pf(x) = \begin{bmatrix} D^pf_1(x) \\ \vdots \\ D^pf_n(x) \end{bmatrix}.$$

Theorem 19. *Let $A \subseteq E$ and $B \subseteq F$ be open sets. Let $f: A \rightarrow F$ and $g: B \rightarrow G$ be class C^p maps with $f(A) \subseteq B$. Then $g \circ f$ is of class C^p .*

Proof. We use induction on p , with the chain rule proving the case $p = 1$ (the case $p = 0$ also holds because a composition of continuous maps is also continuous). Assume that the result holds for $p - 1$ and suppose f and g are of class C^p . By the chain rule we have

$$D(g \circ f)(x) = Dg(f(x)) \circ Df(x).$$

As a function of x , the right hand side is a composition of C^{p-1} maps, so the induction hypothesis shows that $D(g \circ f)$ is of class C^{p-1} and therefore $g \circ f$ is of class C^p . ■

Symmetry

An important fact is that $D^pf(x)$ is always symmetric (as a multilinear map) if f is of class C^p . In part 3, we will show that the well-known [equality of mixed partial derivatives](#) is a special case of this. To prove this result, we start with the case $p = 2$.

Lemma 20. *Let $\varphi: E \times E \rightarrow F$ be a bilinear map. If there is a map ψ into F defined for sufficiently small $(v, w) \in E \times E$ such that*

$$\lim_{(v, w) \rightarrow (0, 0)} \psi(v, w) = 0$$

and

$$|\varphi(v, w)| \leq |\psi(v, w)| \, |v| \, |w|,$$

then $\varphi = 0$.

Proof. Let $v, w \in E$. For sufficiently small $s > 0$ we have

$$|\varphi(sv, sw)| \leq |\psi(sv, sw)| |sv| |sw|,$$

so

$$s^2 |\varphi(v, w)| \leq s^2 |\psi(sv, sw)| |v| |w|.$$

Dividing by s^2 and taking $s \rightarrow 0$ proves the result. ■

Theorem 21. Let $A \subseteq E$ be an open set and let $f: A \rightarrow F$ be a class C^2 map. Then for every $x \in A$, the bilinear map $D^2f(x)$ is symmetric. That is,

$$D^2f(x)(v, w) = D^2f(x)(w, v)$$

for all $v, w \in E$.

Proof. Let $x \in A$ and choose $r > 0$ so that the open ball of radius r around x is contained in A . Let $v, w \in E$ with $|v|, |w| < r/2$. Define $g(x) = f(x + v) - f(x)$. The [mean value theorem](#) then gives

$$\begin{aligned} g(x + w) - g(x) &= \int_0^1 g'(x + tw)w \, dt \\ &= \int_0^1 [Df(x + v + tw) - Df(x + tw)]w \, dt \\ &= \int_0^1 \left(\int_0^1 D^2f(x + sv + tw)v \, ds \right) w \, dt \\ &= \int_0^1 \int_0^1 D^2f(x)(v, w) \, ds \, dt + \int_0^1 \int_0^1 \psi(sv, tw)(v, w) \, ds \, dt \\ &= D^2f(x)(v, w) + \varphi(v, w) \end{aligned}$$

where $\psi(\alpha, \beta) = D^2f(x + \alpha + \beta) - D^2f(x)$ and

$$\varphi = \int_0^1 \int_0^1 \psi(sv, tw) \, ds \, dt.$$

If we repeat the above process starting with g_1 in place of g , where $g_1(x) = f(x + w) - f(x)$, we obtain that

$$g_1(x + v) - g_1(x) = D^2f(x)(w, v) + \varphi(w, v).$$

Since $g(x + w) - g(x) = g_1(x + v) - g_1(x)$, we have

$$D^2f(x)(w, v) - D^2f(x)(v, w) = \varphi(v, w) - \varphi(w, v),$$

and from the definitions of φ and ψ we see that

$$|D^2f(x)(w, v) - D^2f(x)(v, w)| \leq 2 \sup_{0 \leq s, t \leq 1} |\psi(sv, tw)| |v| |w|.$$

Since D^2f is continuous, we can apply Lemma 20 to the bilinear map

$$(v, w) \mapsto D^2f(x)(w, v) - D^2f(x)(v, w)$$

to obtain the result. ■

Theorem 22. Let $A \subseteq E$ be an open set and let $f: A \rightarrow F$ be a class C^p map. Then for every $x \in A$, the multilinear map $D^pf(x)$ is symmetric.

Proof. We use induction on p , with Theorem 21 proving the case $p = 2$. Suppose the result holds for $2, \dots, p - 1$. If $v_1, \dots, v_p \in E$, then

$$\begin{aligned} D^pf(x)(v_1, \dots, v_p) &= D^2D^{p-2}f(x)(v_1, v_2)(v_3, \dots, v_p) \\ &= D^2D^{p-2}f(x)(v_2, v_1)(v_3, \dots, v_p) \\ &= D^2D^{p-2}f(x)(v_2, v_1, v_3, \dots, v_p) \end{aligned}$$

by applying Theorem 21 to the C^2 map $D^{p-2}f$. Also, the induction hypothesis shows that

$$D^{p-1}f(x)(v_{\sigma(2)}, \dots, v_{\sigma(p)}) = D^{p-1}f(x)(v_2, \dots, v_p)$$

for any permutation σ of $\{2, \dots, p\}$. If $\varphi_\sigma: L^{p-1}(E, F) \rightarrow F$ is the linear map given by $\lambda \mapsto \lambda(v_{\sigma(2)}, \dots, v_{\sigma(p)})$ then

$$\begin{aligned} D^pf(x)(v_1, v_{\sigma(2)}, \dots, v_{\sigma(p)}) &= \varphi_\sigma(D^pf(x)(v_1)) \\ &= D(\varphi_\sigma \circ D^{p-1}f)(x)(v_1) \\ &= D(\varphi_e \circ D^{p-1}f)(x)(v_1) \\ &= \varphi_e(D^pf(x)(v_1)) \\ &= D^pf(x)(v_1, \dots, v_p) \end{aligned}$$

where e is the identity permutation. Since any permutation of $\{1, \dots, p\}$ can be expressed as a composition of the permutations considered in (*) and (**), $D^pf(x)$ is symmetric. ■

Taylor's theorem

Taylor’s theorem is an important tool for approximating a function based on its derivatives. It will be important in part 5, where we look at necessary and sufficient conditions for a point to be a local minimum or local maximum.

Theorem 23 (Taylor’s theorem). Let $A \subseteq E$ be an open set and let $f: A \rightarrow F$ be a class C^p map. Let $x \in A$ and let $v \in E$. Assume that the line segment $x + tv$ with $0 \leq t \leq 1$ is contained in A . Write $v^{(k)}$ for the k -tuple (v, \dots, v) . Then

$$f(x + v) = \sum_{k=0}^{p-1} \frac{D^k f(x) v^{(k)}}{k!} + R_p,$$

where

$$R_p = \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!} D^p f(x + tv) v^{(p)} dt.$$

Proof. We use induction on p , with [the mean value theorem](#) proving the case $p = 1$. Assume that the result holds for $p - 1$. Let

$$g(t) = \frac{(1-t)^{p-1}}{(p-1)!} \quad \text{and} \quad h(t) = D^{p-1} f(x + tv) v^{(p-1)}$$

so that

$$g'(t) = -\frac{(1-t)^{p-2}}{(p-2)!} \quad \text{and} \quad h'(t) = D^p f(x + tv) v^{(p)}.$$

(Note that for convenience, we are again identifying $h'(t)$ with an element of F .) Applying [integration by parts](#) with the vector space product $\mathbb{R} \times F \rightarrow F$ gives

$$\begin{aligned} & \int_0^1 \frac{(1-t)^{p-2}}{(p-2)!} D^{p-1} f(x + tv) v^{(p-1)} dt + \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!} D^p f(x + tv) v^{(p)} dt \\ &= -\frac{1}{(p-1)!} D^{p-1} f(x) v^{(p-1)}, \end{aligned}$$

and the result follows. ■

Corollary 24 (Taylor’s theorem with estimate). In Theorem 23, we also have

$$f(x + v) = \sum_{k=0}^p \frac{D^k f(x) v^{(k)}}{k!} + \theta(v)$$

where

$$|\theta(v)| \leq \sup_{0 \leq t \leq 1} \frac{|D^p f(x + tv) - D^p f(x)|}{p!} |v|^p$$

and

$$\lim_{v \rightarrow 0} \frac{\theta(v)}{|v|^p} = 0.$$

Proof. Let $\psi(\alpha) = D^p f(x + \alpha) - D^p f(x)$. We can write R_p as

$$\int_0^1 \frac{(1-t)^{p-1}}{(p-1)!} D^p f(x) v^{(p)} dt + \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!} D^p \psi(tv) v^{(p)} dt.$$

The first integral gives the p th term, and the second integral is bounded by

$$\sup_{0 \leq t \leq 1} |\psi(tv)| |v|^p \int_0^1 \frac{(1-t)^{(p-1)}}{(p-1)!} dt = \frac{1}{p!} \sup_{0 \leq t \leq 1} |\psi(tv)| |v|^p.$$

The result follows from the continuity of $D^p f$ at x . ■

Next time, we will examine partial derivatives and how they relate to higher derivatives.

Navigation: [1. The derivative](#) | [2. Higher derivatives](#) | [3. Partial derivatives](#) | [4. Inverse and implicit functions](#) | [5. Maxima and minima](#)

