wj32

information when you need it

Differentiation done correctly: 5. Maxima and minima

Posted on February 25, 2013 by wj32 / 2 Comments

In this final post, we are going to look at some applications of differentiation to locating maxima and minima of real valued functions. In order to do this, we will be using Taylor's theorem (covered in part 2) to prove the higher derivative test for functions on Banach spaces, and the implicit function theorem (covered in part 4) to prove a special case of the method of Lagrange multipliers.

Consequences of Taylor's theorem

Definition 35. Let $f: X \to \mathbb{R}$ be a map defined on a topological space X. If there is a neighborhood U of $x \in X$ such that $f(t) \leq f(x)$ for all $t \in U$, then we say that f has a **local maximum** at x. Similarly, if $f(t) \geq f(x)$ for all $t \in U$ then we say that f has a **local minimum** at x. If f has a local maximum or local minimum at x, then we say that f has an **extreme value** at x. If strict inequality holds, then we say that f has a **strict** local maximum or minimum.

In single variable calculus, a differentiable function $f: \mathbb{R} \to \mathbb{R}$ has a local maximum or minimum at a point $x \in \mathbb{R}$ only if $f'(x) = 0$. It is easy to extend this result to maps defined on Banach spaces.

Theorem 36. Let $A \subseteq E$ be an open set and let $f: A \to \mathbb{R}$. If f is differentiable at $x \in A$ and has an extreme value at x, then $f'(x) = 0$.

Proof. Let $v \in E$ and let $g(t) = x + tv$. Then $f \circ g$ has an extreme value at 0, so $0 = (f \circ g)'(0) = f'(g(0))g'(0) = f'(x)v$. Therefore $f'(x) = 0$. ∎

Also recall that if $f: \mathbb{R} \to \mathbb{R}$ is of class C^1 and there is a point $x \in \mathbb{R}$ such that $f'(x) = 0$, then $f(x)$ is a local minimum if $f''(x) > 0$ and $f(x)$ is a local maximum if $f''(x) < 0$. There is a similar test for higher derivatives that follows from Taylor's theorem. Again, we can prove analogous statements for maps defined on Banach spaces. If $q \in L(E, \ldots, E; \mathbb{R})$ is a multilinear map from E^p to $\mathbb{R}$, then we say that q is a **multilinear form**.

Definition 37. Write $h^{(p)}$ for the p-tuple $(h, \ldots, h)$. We say that a form q is **positive semidefinite** if $qh^{(p)} \geq 0$ for all h and **positive definite** if $qh^{(p)} > 0$ for all $h \neq 0$. The terms **negative semidefinite** and **negative definite** are defined similarly. If $qh^{(p)}$ takes on both positive and negative values, then we say that q is **indefinite**.

Theorem 38 (Higher derivative test). Let $A \subseteq E$ be an open set and let $f: A \to \mathbb{R}$. Assume that f is $(p-1)$ times continuously differentiable and that $D^p f(x)$ exists for some $p \geq 2$ and $x \in A$. Also assume that $f'(x), \ldots, f^{(p-1)}(x) = 0$ and $f^{(p)}(x) \neq 0$. Write $h^{(p)}$ for the p-tuple $(h, \ldots, h)$.

1. If f has an extreme value at x, then p is even and the form $f^{(p)}(x)h^{(p)}$ is semidefinite.
2. If there is a constant c such that $f^{(p)}(x)h^{(p)} \geq c > 0$ for all $|h| = 1$, then f has a strict local minimum at x and (1) applies.
3. If there is a constant c such that $f^{(p)}(x)h^{(p)} \leq c < 0$ for all $|h| = 1$, then f has a strict local maximum at x and (1) applies.

Proof. By Corollary 24 and the given assumptions, we can write

$$f(x + h) - f(x) = \frac{1}{p!}f^{(p)}(x)h^{(p)} + \theta(h)|h|^p$$

where $\theta(h) \to 0$ as $h \to 0$. First assume that f has an extreme value at x. Choose a vector $h_0 \neq 0$ such that $f^{(p)}(x)h_0^{(p)} \neq 0$. Then for sufficiently small $t \in \mathbb{R}$ we have both

$$f(x + th_0) - f(x) = \left(\frac{1}{p!}f^{(p)}(x)h_0^{(p)} \pm \theta(th_0)\left|h_0\right|^p\right)t^p$$

and

$$\left|\theta(th_0)\right|\left|h_0\right|^p < \frac{1}{p!}f^{(p)}(x)h_0^{(p)}.$$

For these t, the sign of (*) is the same as the sign of $f^{(p)}(x)h_0^{(p)}$. Since x is an extreme value, the sign of (*) must remain constant for small t, which cannot happen unless p is even. Similarly, if $f^{(p)}(x)h^{(p)}$ is not semidefinite then there is some vector $h_1 \neq 0$ such that $f^{(p)}(x)h_1^{(p)}$ and $f^{(p)}(x)h_0^{(p)}$ have opposite

signs, which contradicts the fact that the sign of (*) is constant for small t. Now suppose that the condition in (2) holds. Then

$$f(x + h) - f(x) = \frac{1}{p!}f^{(p)}(x)h^{(p)} + \theta(h)|h|^p$$

$$= \left[\frac{1}{p!}f^{(p)}(x)\left(\frac{h}{|h|}\right)^{(p)} + \theta(h)\right]|h|^p$$

$$\geq \left[\frac{c}{p!} + \theta(h)\right]|h|^p.$$

Since $\theta(h) \to 0$ as $h \to 0$, the last term is positive for sufficiently small $h \neq 0$. For these h we have $f(x + h) > f(x)$, so f has a strict local minimum at x. The proof for (3) is similar. ▫

Corollary 39 (Higher derivative test, finite-dimensional case). *In Theorem 38, further assume that E is finite-dimensional. Then $h \mapsto f^{(p)}(x)h^{(p)}$ has both a minimum and maximum value on the set $\{h \in E : |h| = 1\}$, and:*

1. *If the form $f^{(p)}(x)h^{(p)}$ is indefinite, then f does not have an extreme value at x.*
2. *If the form $f^{(p)}(x)h^{(p)}$ is positive definite, then f has a strict local minimum at x.*
3. *If the form $f^{(p)}(x)h^{(p)}$ is negative definite, then f has a strict local maximum at x.*

Proof. Since E is finite-dimensional, the set $S = \{h \in E : |h| = 1\}$ is compact. Therefore the continuous map $h \mapsto f^{(p)}(x)h^{(p)}$ attains a minimum c and a maximum C on S. Part (1) follows directly from part (1) of Theorem 38. If $f^{(p)}(x)h^{(p)}$ is positive definite then $c > 0$, so part (2) of Theorem 38 applies. If $f^{(p)}(x)h^{(p)}$ is negative definite then $C < 0$, so part (3) of Theorem 38 applies. ▫ The simplest form of Corollary 39 occurs when $p = 2$. Let E be an n-dimensional real Banach space, let $A \subseteq E$ be an open set, and let $f : A \to \mathbb{R}$ be a class C^1 map. Suppose that $f''(x)$ exists at $x \in A$. Let $\{e_1, ..., e_n\}$ be a basis for E so that $E = E_1 \times \cdots \times E_n$, where E_i is the subspace generated by e_i. **Definition 40.** The **Hessian** matrix of f at x is the real matrix

$$\begin{bmatrix} D_1D_1f(x) & \cdots & D_1D_nf(x) \\ \vdots & \ddots & \vdots \\ D_nD_1f(x) & \cdots & D_nD_nf(x) \end{bmatrix},$$

where each element

$$D_iD_jf(x) \in L(E_i, L(E_j, \mathbb{R}))$$

is identified with $D_iD_jf(x)(e_i, e_j) \in \mathbb{R}$.

Theorem 29 shows that this matrix is symmetric. We can restate Corollary 39 in terms of the Hessian matrix.

Corollary 41. *Suppose that $f'(x) = 0$ and $f''(x)$ exists. Let H be the Hessian matrix of f at x.*

1. *If H has both positive and negative eigenvalues, then f does not have an extreme value at x.*
2. *If H is positive definite, then f has a strict local minimum at x.*
3. *If H is negative definite, then f has a strict local maximum at x.*

Proof. It is clear that $f''(x)(h, h) = \tilde{h}^T H \tilde{h}$, where $\tilde{h}$ is the column vector representing h. ▫

Lagrange multipliers

The method of Lagrange multipliers provides a necessary condition for a function $f : A \to \mathbb{R}$ to be maximized or minimized subject to a constraint expressed as a function $g : A \to \mathbb{R}$. We first need an elementary result from linear algebra.

Lemma 42. *Let $f, g : E \to \mathbb{R}$ be nonzero linear functionals. If $\ker f \subseteq \ker g$, then $f = \lambda g$ for some $\lambda \in \mathbb{R}$.*

Proof. $\ker f$ cannot be a strict subset of $\ker g$ since $\dim(E/\ker f) = 1$, so $\ker f = \ker g$. Let $v \notin \ker f$ and take $\lambda = f(v)/g(v)$. Clearly $f = \lambda g$ on $\ker f = \ker g$. If $x \notin \ker f$ then $x = rv$ for some $r \in \mathbb{R}$, so

$$f(x) = rf(v) = \lambda rg(v) = \lambda g(x).$$

Therefore $f = \lambda g$ on E. ▫

Theorem 43 (Method of Lagrange multipliers, single constraint). *Let $A \subseteq E$ be an open set. Let $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$ be of class C^1, and let $S = g^{-1}(\{0\})$. If $f|_S$ has an extreme value at $x \in S$ and $g'(x) \neq 0$, then there is a number $\lambda \in \mathbb{R}$ such that $f'(x) = \lambda g'(x)$.*

Proof. If we can prove that $\ker g'(x) \subseteq \ker f'(x)$ then the result follows from Lemma 42. Choose some $w \notin \ker g'(x)$, let $F = \ker g'(x)$ and let $G = \langle w \rangle$; then $E = F \oplus G$. Let $B = A \cap F$ and let $C = A \cap G$. Write $x = (x_1, x_2)$ where $x_1 \in B$ and $x_2 \in C$. Since $g'(x) \neq 0$, $D_2g(x)$ is invertible; we also have $g(x_1, x_2) = 0$. By the implicit function theorem, there exists a neighborhood $U \subseteq B$ of x_1 and a C^1 map $h : U \to C$ such that $h(x_1) = x_2$ and $g(x_1, h(x_1)) = 0$. Let $\tilde{h} : U \to A$ be given by $t \mapsto (t, h(t))$ so that $\tilde{h}(U) \subseteq S$ and $h'(x_1)|_F$ is the identity map on F. Since $f|_S$ has an extreme value at x we have $(f \circ \tilde{h})'(x_1) = 0$ by Theorem 36, so $f'(x) \circ \tilde{h}'(x_1) = 0$ by the chain rule. In particular, if $v \in \ker g'(x) = F$ then

$$0 = [f'(x) \circ \tilde{h}'(x_1)](v) = f'(x)v,$$

so $v \in \ker f'(x)$. ▪

There is also a more general version for a constraint function that maps into an infinite-dimensional space. We omit the proof because it requires a few theorems from functional analysis.

Theorem 44 (Method of Lagrange multipliers, multiple constraints). *Let $A \subseteq E$ be an open set. Let $f : A \to R$ and $g : A \to F$ be of class C^1, and let $S = g^{-1}(\{0\})$. If $f|_S$ has an extreme value at $x \in S$ and $g'(x)$ is surjective, then there is a continuous linear map $\lambda : F \to R$ such that $f'(x) = \lambda \circ g'(x)$.*

Conclusion

We have seen how the theorems of multivariable calculus in R^n generalize easily to more general Banach spaces. Because we can work coordinate-free, the proofs are often easier to understand than their R^n counterparts. By constructing the derivative on Banach spaces, we gain a powerful tool that allows us to both do computations and prove things much more easily than before.

wj32

[View all posts by wj32 →](#)

2 Comments

ab cd

March 8, 2018 at 8:55 pm

There is a minor error in proof of lemma 42: if x is not in ker(f), that doesn't imply that x = rv. It implies that x = rv + w for some w \in ker(f) = ker(g).

But since f(w) = g(w) = 0 your next step still holds, viz.:

f(x) = f(rv+w) = rf(v) + f(w) = rf(v) = r \lambda g(v) = \lambda [r g(v) + g(w)] = \lambda g(x)

Pingback: [What is the higher-order derivative test in multivariable calculus? – Math Solution](#)