Unconstrained Optimization

Definition: Let $f : X \to \mathbb{R}$ be a real-valued function and let $\overline{x} \in X$. [Note that X can be any set, not necessarily a subset of $\mathbb{R}^n$.] We say that

$\overline{x}$ is a (global) **maximum** or **maximum point** of f if $\forall x \in X : f(\overline{x}) \geqq f(x)$; and

$\overline{x}$ is a (global) **minimum** or **minimum point** of f if $\forall x \in X : f(\overline{x}) \leqq f(x)$.

We say the $\overline{x}$ is an **extremum** if it's either a maximum or a minimum.

Definition: Let $f : X \to \mathbb{R}$ be a real-valued function on a set X in $\mathbb{R}^n$ and let $\overline{x} \in X$. We say that

$\overline{x}$ is a **local maximum** of f if there is neighborhood $\mathfrak{N}$ of $\overline{x}$ such that $\forall x \in \mathfrak{N} : f(\overline{x}) \geqq f(x)$;

$\overline{x}$ is a **local minimum** of f if there is neighborhood $\mathfrak{N}$ of $\overline{x}$ such that $\forall x \in \mathfrak{N} : f(\overline{x}) \leqq f(x)$.

We say that $\overline{x}$ is a **locally unique** or **strict** maximum or minimum if there is a neighborhood $\mathfrak{N}$ in which the inequality is strict for all $x \in \mathfrak{N}$ except $\overline{x}$ itself.

Obviously, a global maximum/minimum is a local maximum/minimum, and a local maximum/minimum is not necessarily a global maximum/minimum.

Definition: A point $\overline{\mathbf{x}} \in \mathbb{R}^n$ is a **critical point** of a function $f : \mathbb{R}^n \to \mathbb{R}$ if $\nabla f(\overline{x}) = \mathbf{0}$ — *i.e.*, if all partial derivatives of f at $\overline{x}$ are zero.

Our first theorem establishes a *necessary* condition for a point to be a maximum or a minimum of a function: the point must be a critical point. We first state and prove the theorem for real functions, and only for a local maximum. The proof is easily altered to cover the alternative case of a local minimum. Then we state and prove the theorem for functions on $\mathbb{R}^n$.

Theorem: Let $f : \mathbb{R} \to \mathbb{R}$. If $\overline{x}$ is a local maximum of f and f is differentiable at $\overline{x}$, then $\overline{x}$ is a critical point of f.

Proof: Suppose that $f'(\overline{x}) > 0$. Then there is a $\delta > 0$ such that

$$0 < |\Delta x| < \delta \Rightarrow \frac{f(\overline{x} + \Delta x) - f(\overline{x})}{\Delta x} > 0.$$

Therefore for $0 < \Delta x < \delta$ we have $f(\overline{x} + \Delta x) - f(\overline{x}) > 0$ — *i.e.*, $f(\overline{x} + \Delta x) > f(\overline{x})$. In every neighborhood of $\overline{x}$ there are points $\overline{x} + \Delta x$ at which $f(\overline{x} + \Delta x) > f(\overline{x})$, so $\overline{x}$ is not a local maximum of f, contrary to the theorem's assumption. Therefore $f'(\overline{x}) > 0$ cannot be true. A parallel argument shows that $f'(\overline{x}) < 0$ also cannot be true. Since $f'(\overline{x})$ does exist, we have $f'(\overline{x}) = 0$. ■

While a local maximum or minimum point must be a critical point — a necessary condition — the reverse is not true: "$\overline{x}$ is a critical point" is not *sufficient* to ensure that $\overline{x}$ is either a maximum or minimum point. The classical counterexample is $f(x) = x^3$: $\overline{x} = 0$ is a critical point of f, but is obviously neither a local maximum nor a local minimum.

We often refer to properties of a function's derivative as "first-order" properties, and properties of a function's second derivative as "second-order" properties of the function. So here we say that a *first-order necessary condition* for $\overline{x}$ to be a local maximum or minimum of a function is that it be a critical point of the function.

Now we obtain the generalization to multivariate functions, the first-order necessary condition for a local maximum of a function $f : \mathbb{R}^n \to \mathbb{R}$. As with $n = 1$ case above, the proof is easily altered to cover the case of a local minimum.

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$. If $\overline{\mathbf{x}} \in \mathbb{R}^n$ is a local maximum of f and f is differentiable at $\overline{\mathbf{x}}$, then $\overline{\mathbf{x}}$ is a critical point of f.

Proof: We apply the same argument as in the preceding proof to each of the partial derivatives of f. Suppose $\dfrac{\partial f}{\partial x_i}(\overline{\mathbf{x}}) > 0$. Then there is a $\delta > 0$ such that

$$0 < |h| < \delta \Rightarrow \frac{f(\overline{\mathbf{x}} + h\mathbf{e}_i) - f(\overline{\mathbf{x}})}{h} > 0,$$

where $\mathbf{e}_i$ is the i-th unit vector in $\mathbb{R}^n$. Then for $0 < h < \delta$ we have $f(\overline{\mathbf{x}} + h\mathbf{e}_i) - f(\overline{\mathbf{x}}) > 0$ — *i.e.*, $f(\overline{\mathbf{x}} + h\mathbf{e}_i) > f(\overline{\mathbf{x}})$. Therefore there is no neighborhood of $\overline{\mathbf{x}}$ on which $f(\mathbf{x}) \leqq f(\overline{\mathbf{x}})$, so $\overline{\mathbf{x}}$ is not a local maximum of f. Since this contradicts the assumption of the theorem, we cannot have $\dfrac{\partial f}{\partial x_i}(\overline{\mathbf{x}}) > 0$. A parallel argument shows that we cannot have $\dfrac{\partial f}{\partial x_i}(\overline{\mathbf{x}}) < 0$. Therefore $\dfrac{\partial f}{\partial x_i}(\overline{\mathbf{x}}) = 0$. ■

Now we develop a *sufficient* condition for a local extremum, which includes a second-order condition as well as the first-order critical-point condition — a condition on the function's second derivative. We again begin with the case $n = 1$ before dealing with the case of an arbitrary n. And this time we prove the theorem for the case of a local minimum; the proof for a local maximum is a straightforward alteration of this proof.

Theorem: Let $f : \mathbb{R} \to \mathbb{R}$ and let $\overline{x}$ be a point at which f is twice differentiable. If $f'(\overline{x}) = 0$ and $f''(\overline{x}) > 0$, then $\overline{x}$ is a local minimum point of f.

Proof: As we've done before, we define the function F by $F(\Delta x) = f(\overline{x} + \Delta x) - f(\overline{x})$. We'll show that for some neighborhood $\mathfrak{N}$ around zero

$$\left[\Delta x \in \mathfrak{N} \;\&\; \Delta x \neq 0\right] \Rightarrow F(\Delta x) > 0,$$

i.e., $f(\overline{x} + \Delta x) > f(\overline{x})$ for all nonzero $\Delta x \in \mathfrak{N}$, so this will show that $\overline{x}$ is indeed a locally unique minimum point of f.

Using the 2^{nd}-degree Taylor polynomial, we have

$$F(\Delta x) = f'(\overline{x})\Delta x + \frac{1}{2}f''(\overline{x})(\Delta x)^2 + R_2(\Delta x), \tag{1}$$

where $R_2(\Delta x)$ is the remainder term, which satisfies $\lim\limits_{\Delta x \to 0} \dfrac{1}{(\Delta x)^2} R_2(\Delta x) = 0$. Because the theorem assumes that $f'(\overline{x}) = 0$, equation (1) becomes

$$F(\Delta x) = \frac{1}{2}f''(\overline{x})(\Delta x)^2 + R_2(\Delta x); \tag{2}$$

and because we're excluding $\Delta x = 0$, we can divide both sides of equation (2) by $(\Delta x)^2$, so we have

$$\frac{1}{(\Delta x)^2}F(\Delta x) = \frac{1}{2}f''(\overline{x}) + \frac{1}{(\Delta x)^2}R_2(\Delta x). \tag{3}$$

Equation (3) is the key element of the proof: the two terms on the right-hand side are $\frac{1}{2}f''(\overline{x})$, which is a given positive number, and $\dfrac{1}{(\Delta x)^2}R_2(\Delta x)$, which is very small when Δx is small. So for small values of Δx, the first term dominates the second term — *i.e.*, since the first term is a positive number, when we add the second term to it (even if the second term is negative), the right-hand side will still be positive. (Note, in fact, that when $R_2(\Delta x)$ is positive, the right-hand side is clearly positive, so it's only those Δx for which $R_2(\Delta x)$ is negative that we have to worry about.) This will guarantee that $F(\Delta x)$ on the left-hand side is positive, which is the conclusion we're trying to establish.

Now let's write down formally what we described informally in the preceding paragraph. Since $\lim\limits_{\Delta x\to 0} \dfrac{1}{(\Delta x)^2}R_2(\Delta x) = 0$, there is a $\delta > 0$ such that

$$0 < |\Delta x| < \delta \Rightarrow \left|\frac{1}{(\Delta x)^2}R_2(\Delta x)\right| < \frac{1}{2}f''(\overline{x}).$$

Therefore, when $0 < |\Delta x| < \delta$ we have

$$\frac{1}{2}f''(\overline{x}) + \frac{1}{(\Delta x)^2}R_2(\Delta x) > 0,$$

and therefore $F(\Delta x) > 0$, *i.e.*, $f(\overline{x} + \Delta x) > f(\overline{x})$, for all Δx such that $0 < |\Delta x| < \delta$, completing the proof. ■

Now we want to prove the theorem for functions on any Euclidean space $\mathbb{R}^n$. Conceptually, we can copy the $n = 1$ proof we just completed. The key idea of the proof will still be the idea expressed in equation (3): we'll have an equation in which the quantity on the left is the one we want to prove is positive when Δx is small; and the right-hand side will have one term that's known to be positive, plus a Taylor remainder term that's smaller than the first term when Δx is small.

But because the displacements Δx will be vectors rather than just numbers, three features of the proof will have to be adjusted: (i) we can't divide by $(\Delta x)^2$ when Δx is a vector; (ii) when we work with the limit of a function on $\mathbb{R}^n$, we're not just working with *numbers* that are small, but with *vectors* that are small, *i.e.*, whose lengths are small; and (iii) the second-derivative term in equation (3) — the first term on the right-hand side — won't be a number, but the Hessian matrix of f at $\overline{x}$, and we'll have to turn that into a number for the equation to make sense.

We'll use a tool — the Weierstrass Theorem — that we haven't seen yet in the course, one that we'll take on faith for now, and we'll develop it in depth later in the course. The theorem deals with two properties of sets that we'll also develop later: sets that are closed, and sets that are bounded.

The Weierstrass Theorem: Let X be a closed and bounded subset of $\mathbb{R}^n$ and let $f : X \to \mathbb{R}$. If f is continuous, then f attains both a maximum and a minimum on X:

$$\exists \overline{x} \in X : x \in X \Rightarrow f(\overline{x}) \geqq f(x) \quad \text{and} \quad \exists \widehat{x} \in X : x \in X \Rightarrow f(\widehat{x}) \leqq f(x).$$

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $\overline{\mathbf{x}}$ be a point at which f is twice differentiable. If $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$ and the Hessian matrix $H_f(\overline{\mathbf{x}})$ is positive definite, then $\overline{\mathbf{x}}$ is a locally unique minimum point of f.

Proof: We again define the function F by $F(\Delta\mathbf{x}) = f(\overline{\mathbf{x}} + \Delta\mathbf{x}) - f(\overline{\mathbf{x}})$. Using the 2^{nd}-degree Taylor polynomial, we have

$$F(\Delta\mathbf{x}) = \nabla f(\overline{\mathbf{x}})\Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}\, H_f(\overline{\mathbf{x}})\, \Delta\mathbf{x} + R_2(\Delta\mathbf{x}), \tag{4}$$

where $R_2(\Delta\mathbf{x})$ is the remainder term, which satisfies $\lim\limits_{\Delta\mathbf{x}\to\mathbf{0}} \frac{1}{\|\Delta\mathbf{x}\|^2} R_2(\Delta\mathbf{x}) = 0$.

Because the theorem assumes that $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$, equation (4) becomes

$$F(\Delta\mathbf{x}) = \frac{1}{2}\Delta\mathbf{x}\, H_f(\overline{\mathbf{x}})\, \Delta\mathbf{x} + R_2(\Delta\mathbf{x}); \tag{5}$$

and because we're excluding $\Delta\mathbf{x} = \mathbf{0}$, we can divide both sides of equation (5) by $\|\Delta\mathbf{x}\|^2$:

$$\frac{1}{\|\Delta\mathbf{x}\|^2} F(\Delta\mathbf{x}) = (\frac{1}{2}) \frac{1}{\|\Delta\mathbf{x}\|^2} \Delta\mathbf{x}\, H_f(\overline{\mathbf{x}})\, \Delta\mathbf{x} + \frac{1}{\|\Delta\mathbf{x}\|^2} R_2(\Delta\mathbf{x}). \tag{6}$$

Now note that unlike equation (3) in the $n = 1$ proof, the first right-hand-side term in (6) is not the product of two numbers, like $f''(\overline{x})(\Delta\mathbf{x})^2$ as in the $n = 1$ proof, but is instead a quadratic form. So now what we have to establish is that there is some positive number, say $\beta > 0$, such that the quadratic form in (6) always satisfies $Q(\Delta\mathbf{x}) > \beta$. Then β can substitute for the number $f''(\overline{x})(\Delta\mathbf{x})^2$ in the $n = 1$ proof.

This is where the Weierstrass Theorem comes in. First note that for any nonzero vector $\mathbf{z} \neq \mathbf{0} \in \mathbb{R}^n$, if we multiply $\mathbf{z}$ by the scalar $\frac{1}{\|\mathbf{z}\|}$, then the length of the resulting vector $\frac{1}{\|\mathbf{z}\|}\mathbf{z}$ is exactly 1:

$$\left\| \frac{1}{\|\mathbf{z}\|}\mathbf{z} \right\| = \left| \frac{1}{\|\mathbf{z}\|} \right| \|\mathbf{z}\| = \frac{1}{\|\mathbf{z}\|} \|\mathbf{z}\| = 1.$$

Next note that (i) the set $\{\mathbf{z} \in \mathbb{R}^n \mid \|\mathbf{z}\| = 1\}$, the unit sphere in $\mathbb{R}^n$, which we'll denote by S, is a closed and bounded set, and (ii) all quadratic forms are continuous functions (why is that?). Therefore the Weierstrass Theorem tells us that on the sphere S a quadratic form $Q(\mathbf{z})$ will attain a minimum value, say β. And if the quadratic form is positive definite, then we must have $\beta > 0$: $Q(\mathbf{z})$ *attains* its minimum value β, so if we have $\beta = 0$ then there will be a vector $\mathbf{z} \in S$ at which $Q(\mathbf{z}) = 0$, contradicting the fact that $Q(\cdot)$ is positive definite. Finally, note that we can rewrite the quadratic term in (6) as follows, so that the quadratic form is only evaluated on the unit sphere:

$$(\frac{1}{2}) \frac{1}{\|\Delta\mathbf{x}\|^2} \Delta\mathbf{x}\, H_f(\overline{\mathbf{x}})\, \Delta\mathbf{x} = \frac{1}{2} \left(\frac{1}{\|\Delta\mathbf{x}\|} \Delta\mathbf{x} \right) H_f(\overline{\mathbf{x}}) \left(\frac{1}{\|\Delta\mathbf{x}\|} \Delta\mathbf{x} \right). \tag{7}$$

Now we can return to copying the $n = 1$ proof. Since the Hessian matrix $H_f(\overline{\mathbf{x}})$ is positive definite, there is a $\beta > 0$ such that $\mathbf{z}H\mathbf{z} \geqq \beta$ for all $\mathbf{z}$ in the unit sphere S. Let $\delta > 0$ be such that

$$0 < \|\Delta x\| < \delta \;\Rightarrow\; \left| \frac{1}{\|\Delta\mathbf{x}\|^2} R_2(\Delta\mathbf{x}) \right| < \frac{1}{2}\beta.$$

Then for all $\Delta\mathbf{x}$ that satisfy $0 < \|\Delta\mathbf{x}\| < \delta$, the right-hand side of (6) is positive, and therefore $F(\Delta\mathbf{x}) > 0$, *i.e.*, $f(\overline{x} + \Delta x) > f(\overline{x})$, for all $\Delta\mathbf{x}$ such that $0 < \|\Delta\mathbf{x}\| < \delta$, completing the proof. ■

An obvious corollary is the parallel theorem that provides a sufficient condition for a point $\overline{x}$ to be a local maximum:

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $\overline{\mathbf{x}}$ be a point at which f is twice differentiable. If $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$ and the Hessian matrix $H_f(\overline{\mathbf{x}})$ is negative definite, then $\overline{\mathbf{x}}$ is a locally unique maximum point of f.

This follows immediately from the local-minimum theorem: $\overline{x}$ is a local maximum of f if and only if it's a local minimum of $-f$; and the Hessian of $-f$ is the negative of the Hessian of f, so the associated quadratic forms satisfy $\Delta\mathbf{x}\, H_{-f}(\overline{\mathbf{x}})\, \Delta\mathbf{x} = -\Delta\mathbf{x}\, H_f(\overline{\mathbf{x}})\, \Delta\mathbf{x}$.

We've now developed both a *sufficient* condition to ensure that a point $\overline{x}$ is a local maximum or minimum (consisting of a first-order condition together with a second-order condition), and a *necessary* first-order condition that a point has to satisfy if it's a local maximum or minimum. But it seems pretty clear that there is also a second-order condition that a local maximum $\overline{x}$ has to satisfy. The condition can't be as strong as the second-order sufficient condition — *viz.*, that $f''(\overline{x}) < 0$, in the case $n = 1$, and that the Hessian of f at $\overline{x}$ is negative definite more generally. That condition is violated by any constant function f, for example, where every point $\overline{x}$ in the function's domain is a global (and therefore local) maximizer, but $f''(x) = 0$ at every point x, in the case $n = 1$, and the Hessian of f is the zero matrix (all second partial derivatives are zero) for general n.

The second-order necessary condition for a point to be a local maximum or minimum is the slightly weaker condition that the Hessian matrix must be negative or positive *semi*definite instead of definite:

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $\overline{\mathbf{x}}$ be a point at which f is twice differentiable. If $\overline{x}$ is a local maximum of f, then $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$ and the Hessian matrix $H_f(\overline{\mathbf{x}})$ is negative semidefinite. If $\overline{x}$ is a local minimum of f, then $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$ and the Hessian matrix $H_f(\overline{\mathbf{x}})$ is positive semidefinite.

Proof for $n = 1$: The proof for the case $n = 1$ is elementary. Assume that $\overline{x}$ is a local maximum of f. Therefore, according to our earlier theorem, $f'(\overline{x}) = 0$. If $f''(\overline{x}) > 0$, then according to our sufficient condition, $\overline{x}$ is a *strict* local minimum, and therefore can't also be a local maximum, a contradiction. Therefore $f''(\overline{x}) \leqq 0$.

Proof for general n: Assume that $\overline{x}$ is a local maximum of f. Therefore, according to our earlier theorem, $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$. Suppose that $\Delta\mathbf{x}\, H_f(\overline{\mathbf{x}})\, \Delta\mathbf{x} > 0$ for some $\Delta\mathbf{x} \neq \mathbf{0} \in \mathbb{R}^n$. We will show that for any real number λ that's sufficiently small, $f(\overline{x} + \lambda\Delta\mathbf{x}) > f(\overline{x})$, contradicting the fact that $\overline{x}$ is a local maximum. For every λ we have

$$f(\overline{x} + \lambda\Delta\mathbf{x}) - f(\overline{x}) = \nabla f(\overline{x})(\lambda\Delta\mathbf{x}) + \frac{1}{2}(\lambda\Delta\mathbf{x})H_f(\overline{x})(\lambda\Delta\mathbf{x}) + R_2(\lambda\Delta\mathbf{x}),$$

where

$$\lim_{\lambda\to 0} \frac{1}{\lambda^2} R_2(\lambda\Delta\mathbf{x}) = 0.$$

Therefore, since $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$, we have

$$f(\overline{x} + \lambda\Delta\mathbf{x}) - f(\overline{x}) = \frac{1}{2}\lambda^2 \Delta\mathbf{x}\, H_f(\overline{x})\, \Delta\mathbf{x} + R_2(\lambda\Delta\mathbf{x}),$$

and for $\lambda \neq 0$ we have

$$\frac{1}{\lambda^2}\big[f(\overline{x} + \lambda\Delta\mathbf{x}) - f(\overline{x})\big] = \frac{1}{2}\Delta\mathbf{x}\, H_f(\overline{x})\, \Delta\mathbf{x} + \frac{1}{\lambda^2} R_2(\lambda\Delta\mathbf{x}).$$

Since $\lim_{\lambda\to 0} \frac{1}{\lambda^2} R_2(\lambda\Delta\mathbf{x}) = 0$ and $\Delta\mathbf{x}\, H_f(\overline{x})\, \Delta\mathbf{x} > 0$, there is a $\delta > 0$ such that

$$0 < |\lambda| < \delta \Rightarrow \frac{1}{\lambda^2} R_2(\lambda\Delta\mathbf{x}) < \frac{1}{2}\Delta\mathbf{x}\, H_f(\overline{x})\, \Delta\mathbf{x},$$

and therefore

$$0 < |\lambda| < \delta \Rightarrow \frac{1}{\lambda^2}\big[f(\overline{x} + \lambda\Delta\mathbf{x}) - f(\overline{x})\big] > 0;$$

i.e.,

$$0 < |\lambda| < \delta \Rightarrow f(\overline{x} + \lambda\Delta\mathbf{x}) > f(\overline{x}).$$

which contradicts our assumption that $\overline{x}$ is a local maximum of f. ■

As before, the result for a local minimum of f follows immediately by applying the theorem to $-f$.