THE HESSIAN AND CONVEXITY

Let $f \in C^2(U), U \subset R^n$ open, $x_0 \in U$ a critical point.

Nondegenerate critical points are isolated. A critical point $x_0 \in U$ is *non degenerate* if the quadratic form $d^2f(x_0)$ is non-degenerate: $(\forall v \neq 0)(\exists w \neq 0)(d^2f(x_0)(v,w) \neq 0)$. Equivalently, the symmetric linear operator $H(x_0) \in \mathcal{L}(R^n)$ associated with $d^2f(x_0)$ by the standard inner product in R^n does not have zero as an eigenvalue. This is equivalent to $H(x_0)$ being invertible, or $\det H(x_0) \neq 0$.

Consider the map $F : U \to R^n$, $F(x) = df(x) \in \mathcal{L}(R^n, R) \sim R^n$ (where we use the inner product to identify $\mathcal{L}(R^n, R)$ with R^n in the first equality below). Then, for the standard basis (e_i) of R^n:

$$\langle \partial_{e_i} F(x), e_j \rangle = \partial_{e_i}(df)(x)[e_j] = d^2f(x)(e_i, e_j) = \langle H(x)e_i, e_j \rangle$$

or $dF(x)[e_i] = \partial_{e_i} F(x) = H(x)e_i$ for all i, and we conclude $dF(x) = H(x) \in \mathcal{L}(R^n)$. Thus $H(x_0)$ invertible is equivalent to $dF(x_0)$ invertible, and we know this implies the existence of a ball $B_r(x_0)$ and $c > 0$ so that

$$|F(x) - F(x_0)| > c|x - x_0| \text{ for all } x \in B_r(x_0);$$

in particular (since $F(x_0) = df(x_0) = 0$) we have $df(x) = F(x) \neq 0$ if $x \in B_r(x_0), x \neq x_0$. We conclude *nondegenerate critical points are isolated.*

Hessian test for local max/min.

A. Sufficient conditions.

(i) If $d^2f(x_0) > 0$, x_0 is a strict local min. If $d^2f(x_0) < 0$, x_0 is a strict local max.

Proof. Since x_0 is a critical point, we have for $|v|$ small enough:

$$f(x_0 + v) - f(x_0) = [\frac{1}{2}d^2f(x_0)(\hat{v}, \hat{v}) + \frac{r(v)}{|v|^2}]|v|^2, \quad \hat{v} = \frac{v}{|v|}, \quad \lim_{v \to 0} \frac{r(v)}{|v|^2} = 0.$$

Since $d^2f(x_0) > 0$, by compactness of the unit sphere in R^n we have $\frac{1}{2}d^2f(x_0)(\hat{v}, \hat{v}) > \mu > 0$, for some $\mu > 0$ depending only on x_0. Then choose $\delta > 0$ so that $\frac{|r(v)|}{|v|^2} < \mu/2$ if $|v| < \delta, v \neq 0$. We have:

$$f(x_0 + v) - f(x_0) > |v|^2[\mu + \frac{r(v)}{|v|^2}] > \frac{\mu}{2}|v|^2 > 0 \text{ if } |v| < \delta \textit{ and } v \neq 0.$$

Thus x_0 is a strict local min.

Remark: If we only know $d^2f(x_0) \geq 0$, it doesn't follow that x_0 is a local min. (Example: $x \mapsto x^3, x_0 = 0$.)

(ii) If $d^2f(x_0)$ is indefinite, x_0 is neither a local max, nor a local min. ('x_0 is a 'saddle point'.)

Proof. Let $0 \neq v \in R^n$ be such that $d^2f(x_0)(v,v) > 0$. Then (since $\lim_{t\to 0} r(tv)/t^2|v|^2 = 0$ in the second order Taylor approximation above) we may find $\delta > 0$ (depending on x_0 *and on* v) so that

$$|t| \leq \delta, t \neq 0 \Rightarrow \frac{|r(tv)|}{t^2|v|^2} < \frac{1}{4}d^2f(x_0)(\hat{v},\hat{v}),$$

where $\hat{v} = \frac{v}{|v|}$. For such t:

$$f(x_0+tv)-f(x_0) = t^2|v|^2[\frac{1}{2}d^2f(x_0)(\hat{v},\hat{v})+\frac{r(tv)}{t^2|v|^2}] > t^2|v|^2\frac{1}{4}d^2f(x_0)(\hat{v},\hat{v}) > 0,$$

so x_0 cannot be a local max.

The same argument shows that if $d^2f(x_0)(w,w) < 0$ and $|t|$ is small enough (with $t \neq 0$), then $f(x_0+tw) - f(x_0) < 0$, so x_0 cannot be a local min, either.

B. Necessary conditions. If x_0 is a local min, $d^2f(x_0) \geq 0$. If x_0 is a local max, $d^2f(x_0) \leq 0$.

Remark. Even if x_0 is a strict local min, we can't guarantee $d^2f(x_0) > 0$. (Example: $x \mapsto x^4, x_0 = 0$.)

Proof. Assume $d^2f(x_0)(v,v) < 0$, for some $v \in R^n$. By continuity at x_0 of $d^2f(\cdot)(v,v) : U \to R$, we may find $\delta > 0$ (depending on x_0 and on v) so that $d^2f(x_0+sv)(v,v) < 0$ whenever $|s| \leq \delta$. *Fix such an* s. By the second-order Mean Value Theorem, we may find $\theta \in (0,1)$ (depending on x_0, v and s) so that:

$$f(x_0+sv_0) - f(x_0) = \frac{1}{2}d^2f(x_0+\theta sv)(sv,sv) = \frac{s^2}{2}d^2f(x_0+\theta sv)(v,v) < 0,$$

since $|\theta s| < |s| \leq \delta$. Thus x_0 can't be a local min.

Convexity.

Definition. Let $F : K \to R$ ($K \subset R^n$ open, convex) is *convex* if:

$$(\forall x_0, x_1 \in K) f(x_t) \leq tf(x_1) + (1-t)f(x_0), \quad x_t := tx_1 + (1-t)x_0, t \in [0,1].$$

(*Strictly convex* if we have strict inequality for $t \in (0,1)$.)

Proposition 1. Let f be differentiable in K. Then f is strictly convex iff the graph of f lies above its tangent plane at each point:

$$(\forall x_0, x \in K) \quad f(x) > f(x_0) + df(x_0)[x - x_0].$$

(f is convex iff weak inequality obtains.)

Proof. Assuming f strictly convex, we write down the defining inequality, subtract $f(x_0) + df(x_0)[x_t - x_0]$ from both sides and divide by $t \in (0,1)$, noting $x_t - x_0 = t(x - x_0)$:

$$\frac{1}{t}(f(x_t) - f(x_0) - df(x_0)[x_t - x_0]) < f(x) - f(x_0) - df(x_0)[x - x_0], \quad t \in (0,1).$$

The left-hand side is $r(x_t)/t$ (first-order Taylor remainder), so letting $t \to 0_+$ we conclude the right-hand side is positive.

Conversely, assume strict inequality holds for all $x, x_0 \in K$. Let $x, y \in K$, and take x_0 interior to the segment from x to y:

$$x_0 = (1-t)x + ty, \quad y - x_0 = -\frac{1-t}{t}(x - x_0), t \in (0,1)$$

$$f(x) > f(x_0) + df(x_0)[x - x_0] \quad f(y) > f(x_0) + df(x_0)[y - x_0].$$

Multiplying the first inequality by $1 - t$, the second by t, and adding the results, we eliminate the derivative terms:

$$f(x_0) < tf(y) + (1-t)f(x), t \in (0,1),$$

the inequality defining strict convexity.

Proposition 2. Let $f \in C^2(K)$. Then (i) If $d^2 f(x) > 0$ for all $x \in K$, then f is strictly convex. (And $d^2 f(x)$ semi-positive definite implies convex.) (ii) If f is convex in K, $d^2 f \geq 0$ for all $x \in K$.

Proof (i) Let $x, x_0 \in K$ and write the second-order mean value theorem for this pair: for some $\theta \in (0,1)$:

$$f(x) - f(x_0) - df(x_0)[x - x_0] = \frac{1}{2} d^2 f(x_0 + \theta(x - x_0))(x - x_0, x - x_0).$$

Since the right-hand side is assumed to be positive, Proposition 1 implies f is strictly convex.

Proof(ii). Assume $d^2 f(x_0)(v,v) < 0$, for some $x_0 \in K, v \in R^n, |v| = 1$. By continuity of $d^2 f(\cdot)(v,v)$, we may find $r > 0$ (depending on x_0 and

v) so that $x \in B_r(x_0) \Rightarrow d^2 f(x)(v,v) < 0$; in particular this is true if $x = x_0 + sv, 0 \le s < r$. Fix such an $s > 0$. By the second-order mean value theorem, there exists $\theta \in (0,1)$ (depending on x_0, v and s) so that:

$$f(x_0+sv)-f(x_0)-df(x_0)[sv] = \frac{1}{2}d^2 f(x_0+\theta sv)(sv,sv) = \frac{s^2}{2}d^2 f(x_0+\theta sv)(v,v) < 0,$$

since $0 < \theta s < r$. This contradicts the convexity criterion in Proposition 1.

Remark: Of course, f strictly convex *does not imply* $d^2 f > 0$ at all points (for instance $x \mapsto x^4$.)

Convexity and differentiability in one dimension.

We consider $f : I \to R$ convex, where $I \subset R$ is an *open* interval.

Three slopes lemma. Let $a < b < c$ be points in I. Then:

$$\frac{f(b)-f(a)}{b-a} \le \frac{f(c)-f(a)}{c-a} \le \frac{f(b)-f(c)}{b-c}.$$

It suffices to write:

$$b = tc + (1-t)a, \quad t \in (0,1), \quad b-a = t(c-a), \quad b-c = (1-t)(c-a).$$

Then the first inequality is seen to be equivalent to:

$$f(b) - f(a) \le t(f(c)-f(a)), \text{ or } f(b) \le tf(c) + (1-t)f(a),$$

the definition of convexity. The second inequality is proved similarly.

Continuity. Let $x_0 \in I, \delta > 0$ be such that $[x_0 - \delta, x_0 + \delta] \subset I$; set $M = \max\{f(x_0-\delta), f(x_0+\delta)\}$. (Note $f(x_0) \le M$, by convexity). Let $x \in (x_0, x_0+\delta)$. Applying the lemma to $x_0 < x < x_0 + \delta$, we find:

$$\frac{f(x)-f(x_0)}{x-x_0} \le \frac{f(x_0+\delta)-f(x_0)}{\delta}, \text{ or } f(x)-f(x_0) \le \frac{|x-x_0|}{\delta}(M - f(x_0)),$$

and considering the points $x_0 - \delta < x_0 < x$:

$$\frac{f(x_0)-f(x_0-\delta)}{\delta} \le \frac{f(x)-f(x_0)}{x-x_0}, \text{ or} f(x)-f(x_0) \ge \frac{|x-x_0|}{\delta}(f(x_0)-M).$$

We conclude (similarly for $x \in (x_0 - \delta, x_0)$):

$$|f(x)-f(x_0)| \le \frac{|x-x_0|}{\delta}|M - f(x_0)|.$$

Thus:

$$f \text{ convex on } I \subset R \text{ open} \Rightarrow\ f \text{ continuous in } I.$$

The proof of continuity for $f : K \to R$ convex (where $K \subset R^n$ is convex and open) is similar (see [Fleming].) The fact that I is an open interval is important here. The function $f : [-1, 1] \to R$:

$$f(x) = x^2 \text{ for } |x| < 1, \quad f(\pm 1) = 2$$

is convex and not continuous in $[-1, 1]$.

Differentiability. For $h > 0$, denote by $m_f(x, x+h) = \frac{f(x+h)-f(x)}{h}$ the slope of the (oriented) secant to the graph of f from x to $x + h$. The three-slopes lemma implies this is monotone increasing in h; thus its limit as $h \to 0_+$ exists, the *right-derivative* of f at x:

$$f'_+(x) = \lim_{h \to 0_+} m_f(x, x+h) = \inf_{h>0} m_f(x, x+h).$$

Similarly, the left-derivative exists for each $x \in I$:

$$f'_-(x) = \lim_{k \to 0_+} m_f(x-k, x) = \sup_{k>0} m_f(x-k, x).$$

And since $m_f(x-k, x) \le m_f(x, x+h)$ for each $h, k > 0$, it follows that $f'_-(x) \le f'_+(x)$, for each $x \in I$. Also, if $x < y$ are points in I and $h > 0, k > 0$ are chosen so that $x + h = y - k$, we have:

$$f'_-(x) \le f'_+(x) \le m_f(x, x+h) \le m_f(y-k, y) \le f'_-(y) \le f'_+(y).$$

We see that f'_-, f'_+ are both monotone increasing (nondecreasing) in I, with $f'_+(x) \le f'_-(y)$ if $x < y$.

And now it is easy to see that, if $a \in I$ is a point of continuity of f'_-, then $f'_-(a) = f'_+(a)$:

$$f'_-(a) \le f'_+(a) \le \lim_{y \to a_+} f'_-(y) = f'_-(a).$$

Thus f is differentiable at a. Being monotone, f'_- is continuous at all but a countable set of points in I. We conclude:

$$f \text{ convex in } I \subset R \Rightarrow f \text{ differentiable in the complement of a countable set } D \subset I.$$

The corresponding results in higher dimensions will be seen later.