Caltech Division of the Humanities and Social Sciences

Supergradients

KC Border

Fall 2001

v. 2018.01.25::15.45

1 The supergradient of a concave function

There is a useful way to characterize the concavity of differentiable functions.

Theorem 1 (Concave functions lie below tangent lines) *Suppose f is concave on a convex neighborhood $C \subset \boldsymbol{R}^n$ of x, and differentiable at x. Then for every y in C,*

$$f(x) + f'(x) \cdot (y - x) \geqslant f(y). \tag{1}$$

Proof: Let $y \in C$. Rewrite the definition of concavity as

$$f\big(x + \lambda(y - x)\big) \geqslant f(x) + \lambda\big(f(y) - f(x)\big).$$

Rearranging and dividing by $\lambda > 0$,

$$\frac{f\big(x + \lambda(y - x)\big) - f(x)}{\lambda} \geqslant f(y) - f(x).$$

Letting $\lambda \downarrow 0$, the left hand side converges to $f'(x) \cdot (y - x)$. ∎

The converse is true as the following clever argument shows.

Theorem 2 *Let f be differentiable on a convex open set $U \subset \boldsymbol{R}^n$. Suppose that for every x and y in C, we have $f(x) + f'(x) \cdot (y - x) \geqslant f(y)$. Then f is concave.*

Proof: For each $x \in C$, define the function h_x by $h_x(y) = f(x) + f'(x) \cdot (y - x)$. Each h_x is concave, $f \leqslant h_x$ for each x, and $f(x) = h_x(x)$. Thus

$$f = \inf_{x \in C} h_x.$$

Now the infimum of a family of concave functions is concave (why?), so f is concave. ∎

Theorem 29 below provides a powerful generalization of this result.

Definition 3 *Let $C \subset \boldsymbol{R}^{\mathrm{m}}$ be a convex set, and let $f\colon C \to \boldsymbol{R}$ be concave. A vector p is a* ***supergradient*** *of f at the point x if for every y it satisfies the* ***supergradient inequality****,*

$$f(x) + p \cdot (y - x) \geqslant f(y).$$

Similarly, if f is convex, then p is a ***subgradient*** *of f at x if*

$$f(x) + p \cdot (y - x) \leqslant f(y)$$

for every y.

For concave f, the set of all supergradients of f at x is called the ***superdifferential*** *of f at x, and is denoted $\partial f(x)$. For convex f this denotes the set of subgradients and is called the* ***subdifferential****.*[1]

Theorem 1 clearly implies that the following.

Lemma 4 *The gradient of a concave function at a point of differentiability is also a supergradient.*

In fact, if $\partial f(x)$ is a singleton, then f is differentiable at x and $\partial f(x) = \{f'(x)\}$, see Theorem 14 below or Rockafellar [6, Theorem 25.1, p. 242].

The superdifferential $\partial f(x)$ of a concave function is obviously a closed convex set (since it is the set of solutions to a system of weak linear inequalities, one for each y). If the superdifferential is nonempty at x, we say that f is superdifferentiable at x. (Similar terminology applies to the subgradient of a convex function.)

Example 5 (A non-superdifferentiable point) Define $f\colon [0,1] \to [0,1]$ by $f(x) = x^{\frac{1}{2}}$. Then f is clearly concave, but $\partial f(0) = \varnothing$, since the supergradient inequality implies $p \cdot x \geqslant f(x) - f(0) = x^{\frac{1}{2}}$, so $p \geqslant (\frac{1}{x})^{\frac{1}{2}}$ for all $0 < x \leqslant 1$. Clearly no real number p fills the bill. □

The problem with this example is that the graph of the function becomes arbitrarily steep as we approach the boundary. This cannot happen for interior points. The proof is closely related to that of Theorem **??**.

Theorem 6 (Superdifferentiability) *A concave function on a nonempty convex set in $\boldsymbol{R}^{\mathrm{n}}$ is superdifferentiable at each interior point.*[2]

Proof: Let f be a concave function on the convex set C, and let x belong to the interior of C. Let S denote the strict subgraph of f, that is,

$$S = \{(y, \alpha) \in C \times \boldsymbol{R} : \alpha < f(y)\}.$$

[1] Rockafellar [6, p. 308] suggests this terminology as being more appropriate than the terminology he actually uses, so I shall use it. He uses the term subgradient to mean both subgradient and supergradient, and subdifferential to mean both subdifferential and superdifferential.

[2] Actually it is superdifferentiable at each point in the relative interior of its domain. For more detailed results, see Rockafellar [6, §23], especially Theorem 23.4.

Since f is concave, the set S is convex. Now note that the point $(x, f(x))$ does not belong to S, so by the Separating Hyperplane Theorem, there is a nonzero $(p, \lambda) \in \boldsymbol{R}^{\mathrm{n}} \times \boldsymbol{R}$ separating the point and S. That is,

$$p \cdot x + \lambda f(x) \geqslant p \cdot y + \lambda \alpha \tag{2}$$

for all $y \in C$ and all $\alpha < f(y)$. By considering very negative values of α, we conclude that $\lambda \geqslant 0$. Suppose momentarily that $\lambda = 0$. Since x belongs to the interior of C, for any z in $\boldsymbol{R}^{\mathrm{n}}$ there is some $\varepsilon > 0$ such that $x \pm \varepsilon z$ belong to C. Then equation (2) implies $p \cdot z = 0$. Since z is arbitrary, $p = 0$, so $(p, \lambda) = 0$, a contradiction. Therefore $\lambda > 0$. Dividing by λ, letting $\alpha \uparrow f(y)$, and rearranging yields

$$f(x) + (-\frac{p}{\lambda}) \cdot (y - x) \geqslant f(y),$$

so that $-\frac{p}{\lambda}$ satisfies the supergradient inequality. ∎

2 Concavity and continuity

Concave functions are continuous at interior points. The only discontinuities can be jumps downward at the boundary of the domain.

Theorem 7 (Local continuity of convex functions) *If a convex function is defined and bounded above on a neighborhood of some point in a tvs, then it is continuous at that point.*

Proof: Let C be a convex set in a tvs, and let $f \colon C \to \boldsymbol{R}$ be convex. We begin by noting the following consequences of convexity. Fix $x \in C$ and suppose z satisfies $x + z \in C$ and $x - z \in C$. Let $\delta \in [0, 1]$. Then $x + \delta z = (1 - \delta)x + \delta(x + z)$, so $f(x + \delta z) \leqslant (1 - \delta)f(x) + \delta f(x + z)$. Rearranging terms yields

$$f(x + \delta z) - f(x) \leqslant \delta \left[f(x + z) - f(x) \right], \tag{3}$$

and replacing z by $-z$ gives

$$f(x - \delta z) - f(x) \leqslant \delta \left[f(x - z) - f(x) \right]. \tag{4}$$

Also, since $x = \frac{1}{2}(x + \delta z) + \frac{1}{2}(x - \delta z)$, we have $f(x) \leqslant \frac{1}{2} f(x + \delta z) + \frac{1}{2} f(x - \delta z)$. Multiplying by two and rearranging terms we obtain

$$f(x) - f(x + \delta z) \leqslant f(x - \delta z) - f(x). \tag{5}$$

Combining (4) and (5) yields

$$f(x) - f(x + \delta z) \leqslant f(x - \delta z) - f(x) \leqslant \delta \left[f(x - z) - f(x) \right].$$

This combined with (3) implies

$$\left| f(x + \delta z) - f(x) \right| \leqslant \delta \max\{ f(x + z) - f(x), f(x - z) - f(x) \}. \tag{6}$$

Now let $\varepsilon > 0$ be given. Since f is bounded above on an open neighborhood of x, there is a circled neighborhood V of zero, and a constant $M \geqslant 0$ such that $x + V \subset C$ and if $y \in x + V$, then $f(y) < f(x) + M$. Choosing $0 < \delta \leqslant 1$ so that $\delta M < \varepsilon$, equation (6) implies that if $y \in x + \delta V$, then $|f(y) - f(x)| < \varepsilon$. Thus f is continuous at x. ∎

Theorem 8 (Global continuity of convex functions) *Let f be a convex function on an open convex set C in $\boldsymbol{R}^{\mathrm{n}}$. The following are equivalent.*

1. *f is continuous on C.*
2. *f is upper semicontinuous on C.*
3. *f is bounded above on a neighborhood of some point in C.*
4. *f is continuous at some point in C.*

Proof: (1) $\implies$ (2) Obvious.

(2) $\implies$ (3) Let $x \in C$. If f is upper semicontinuous and convex, then $\{y \in C : f(y) < f(x) + 1\}$ is a convex open neighborhood of x on which f is bounded.

(3) $\implies$ (4) This is Theorem 7.

(4) $\implies$ (1) Suppose f is continuous at x, and let y be any other point in C. Since scalar multiplication is continuous, $\{\beta \in \boldsymbol{R} : x + \beta(y - x) \in C\}$ includes an open neighborhood of 1. This implies that there is some point z in C such that $y = \lambda x + (1 - \lambda)z$ with $0 < \lambda < 1$.

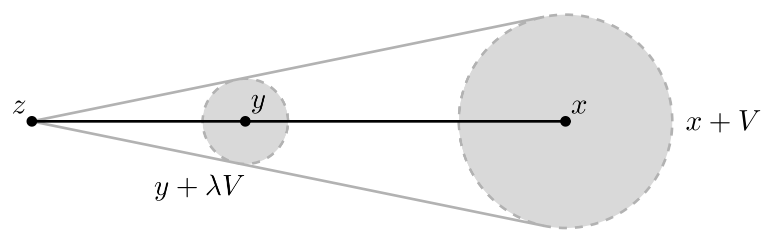

Figure 1. (4) $\implies$ (1).

Also, since f is continuous at x, there is a circled neighborhood V of zero such that $x+V \subset C$ and f is bounded above on $x + V$, say by M. We claim that f is bounded above on $y + \lambda V$. To see this, let $v \in V$. Then $y + \lambda v = \lambda(x + v) + (1 - \lambda)z \in C$. The convexity of f thus implies

$$f(y + \lambda v) \leqslant \lambda f(x + v) + (1 - \lambda)f(z) \leqslant \lambda M + (1 - \lambda)f(z).$$

That is, f is bounded above by $\lambda M + (1 - \lambda)f(z)$ on $y + \lambda V$. By Theorem 7, f is continuous at y. ∎

Theorem 9 *In a finite dimensional vector space, every convex function is continuous on the interior of its domain.*

Proof: Let $f : C \to \boldsymbol{R}$ be a convex function defined on a convex subset C of the Euclidean space $\boldsymbol{R}^{\mathrm{n}}$, and let x be an interior point of C. Then there exist $a, b \in C$ with $a < b$ such that the box $[a, b] = \{y \in \boldsymbol{R}^{\mathrm{n}} : a \leq y \leq b\}$ is a neighborhood of x and satisfies $[a, b] \subset C$. Since $[a, b]$ is the convex hull of a finite set of points of C, the convexity of f implies that f is bounded above on $[a, b]$. So by Theorem 8, f is continuous at x. ∎

3 Concavity and differentiability

We now examine the differentiability of concave functions. We start with the following simple, but fundamental, result for concave functions of one variable, cf. Fenchel [2, 2.16, p. 69], Phelps [5, Theorem 1.16, pp. 9–11], or Royden [7, Proposition 5.17, p. 113].

Lemma 10 *Let f be a concave function defined on some interval I of $\boldsymbol{R}$, and let $I_1 = [x_1, y_1]$ and $I_2 = [x_2, y_2]$ be nondegenerate subintervals of I. That is, $x_1 < y_1$ and $x_2 < y_2$. Assume that I_1 lies to the left of I_2. That is, $x_1 \leqslant x_2$ and $y_1 \leqslant y_2$. Then the slope of the chord over I_1 is greater than the slope of the chord over I_2. In particular,*

$$\frac{f(y_1) - f(x_1)}{y_1 - x_1} \geqslant \frac{f(y_2) - f(x_1)}{y_2 - x_1} \geqslant \frac{f(y_2) - f(x_2)}{y_2 - x_2},$$

Proof: Since $x_1 < y_1 \leqslant y_2$ we can write y_1 as a convex combination of x_1 and y_2, namely

$$y_1 = \frac{y_2 - y_1}{y_2 - x_1} x_1 + \frac{y_1 - x_1}{y_2 - x_1} y_2.$$

By concavity

$$f(y_1) \geqslant \frac{y_2 - y_1}{y_2 - x_1} f(x_1) + \frac{y_1 - x_1}{y_2 - x_1} f(y_2).$$

Subtracting $f(x_1)$ from both sides gives

$$f(y_1) - f(x_1) \geqslant \frac{x_1 - y_1}{y_2 - x_1} f(x_1) + \frac{y_1 - x_1}{y_2 - x_1} f(y_2).$$

Dividing by $y_1 - x_1$ gives

$$\frac{f(y_1) - f(x_1)}{y_1 - x_1} \geqslant \frac{-1}{y_2 - x_1} f(x_1) + \frac{1}{y_2 - x_1} f(y_2) = \frac{f(y_2) - f(x_1)}{y_2 - x_1}.$$

Similarly

$$x_2 = \frac{y_2 - x_2}{y_2 - x_1} x_1 + \frac{x_2 - x_1}{y_2 - x_1} y_2,$$

so

$$\begin{aligned}
f(x_2) &\geqslant \frac{y_2 - x_2}{y_2 - x_1} f(x_1) + \frac{x_2 - x_1}{y_2 - x_1} f(y_2), \\
-f(x_2) &\leqslant \frac{x_2 - y_2}{y_2 - x_1} f(x_1) + \frac{x_1 - x_2}{y_2 - x_1} f(y_2), \\
f(y_2) - f(x_2) &\leqslant \frac{x_2 - y_2}{y_2 - x_1} f(x_1) + \frac{y_2 - x_2}{y_2 - x_1} f(y_2), \\
\frac{f(y_2) - f(x_2)}{y_2 - x_2} &\leqslant \frac{-1}{y_2 - x_1} f(x_1) + \frac{1}{y_2 - x_1} f(y_2) = \frac{f(y_2) - f(x_1)}{y_2 - x_1}.
\end{aligned}$$

Combining these inequalities completes the proof. ∎

This lemma has a number of consequences.

Corollary 11 *Let f be a concave function defined on some interval I of $\boldsymbol{R}$. Then at every interior point x, f has a left-hand derivative $f'(x^-)$ and right-hand derivative $f'(x^+)$. Moreover, $f'(x^-) \geqslant f'(x^+)$, and both $f'(x^-)$ and $f'(x^+)$ are nonincreasing functions. Consequently, there are at most countably many points where $f'(x^-) > f'(x^+)$, that is, where f is nondifferentiable.*

Need to discuss projecting cones here.

Following Fenchel [2] and Rockafellar [6], define the one-sided directional derivative

$$f'(x;v) = \lim_{\lambda \downarrow 0} \frac{f(x+\lambda v) - f(x)}{\lambda},$$

allowing the values ∞ and $-\infty$. (Phelps [5] uses the notation $d^+(x)(v)$.)

Lemma 12 *Let f be concave. Then $\{v : f'(x;v)$ is finite$\}$ is a cone, and $f'(x;\cdot)$ is positively homogeneous and concave on that cone.*

Proof: From

$$\frac{f(x+\alpha\lambda v) - f(x)}{\lambda} = \alpha \frac{f(x+\alpha\lambda v) - f(x)}{\alpha\lambda},$$

we see that $\{v : f'(x;v)$ is finite$\}$ is a cone and that $f'(x;\alpha v) = \alpha f'(x;v)$ for all $\alpha \geqslant 0$. That is, $f'(x;v)$ is positively homogeneous of degree one in v. Furthermore, if f is concave, then

$$f\big(x+\alpha\lambda u+(1-\alpha)\lambda v\big) - f(x) \geqslant \alpha\big(f(x+\lambda u) - f(x)\big) + (1-\alpha)\big(f(x+\lambda v) - f(x)\big),$$

so dividing by λ and taking limits shows that $f'(x;\cdot)$ is concave. ∎

There is an intimate relation between one-sided directional derivatives and the superdifferential, cf. Rockafellar [6, Theorem 32.2, p. 216].

Lemma 13 *Let f be a concave function on the convex set C. Then*

$$p \in \partial f(x) \iff \forall v \ \left(x+v \in C \implies p \cdot v \geqslant f'(x;v)\right).$$

Proof: ($\Longrightarrow$) If $x+v \in C$, then $x+\lambda v \in C$ for $0 \leqslant \lambda \leqslant 1$. So if $p \in \partial f(x)$, then by the supergradient inequality

$$\begin{aligned} f(x) + p \cdot (\lambda v) &\geqslant f(x+\lambda v)) \\ p \cdot (\lambda v) &\geqslant f(x+\lambda v) - f(x) \\ p \cdot v &\geqslant \frac{f(x+\lambda v) - f(x)}{\lambda} \\ p \cdot v &\geqslant f'(x;v). \end{aligned}$$

($\Longleftarrow$) If $p \notin \partial f(x)$, then there is some v such that $x+v \in C$, but the supergradient inequality is violated, that is,

$$f(x) + p \cdot v < f(x+v). \tag{7}$$

By concavity, for $0 < \lambda \leqslant 1$,

$$f(x+\lambda v) = f\big((1-\lambda)x + \lambda(x+v)\big) \geqslant (1-\lambda)f(x) + \lambda f(x+v),$$

and rearranging yields

$$\begin{aligned} f(x+\lambda v) &\geqslant f(x)+\lambda\big[f(x+v)-f(x)\big] \\ f(x+\lambda v)-f(x) &\geqslant \lambda\big[f(x+v)-f(x)\big] \\ \frac{f(x+\lambda v)-f(x)}{\lambda} &\geqslant f(x+v)-f(x) \end{aligned}$$

so by (7)

$$\frac{f(x+\lambda v)-f(x)}{\lambda} \geqslant f(x+v)-f(x) > p\cdot v,$$

so taking limits gives $f'(x;v) > p\cdot v$. The conclusion now follows by contraposition. ∎

The next result may be found in Rockafellar [6, Theorem 25.1, p. 242].

Theorem 14 *Let f be a concave function defined on the convex set $C \subset \boldsymbol{R}^{\mathrm{n}}$. Then f is differentiable at the interior point $x \in C$ if and only if the superdifferential $\partial f(x)$ is singleton, in which case $\partial f(x) = \{f'(x)\}$.*

Proof: ($\Longrightarrow$) Suppose f is differentiable at the interior point x. The for any v, $f'(x;v) = f'(x)\cdot v$, and there is an $\varepsilon > 0$ such that $x+\varepsilon v \in C$. Now the superdifferential $f'(x)$ is nonempty, since by Lemma 4 $f'(x) \in \partial f(x)$, so by Lemma 13, if $p \in \partial f(x)$, then

$$p\cdot\varepsilon v \geqslant f'(x;\varepsilon v) = f'(x)\cdot\varepsilon v.$$

Since this holds for all v, we have $p = f'(x)$.

($\Longleftarrow$) Suppose $f'(x) = \{p\}$. Since x is interior there is an $\alpha > 0$ such that if $v \in \alpha B$, then $x+v \in C$, where B is the unit ball in $\boldsymbol{R}^{\mathrm{n}}$. For such v, the supergradient inequality asserts that $f(x)+p\cdot v \geqslant f(x+v)$. Define the concave function g on B by

$$g(v) = f(x+v)-f(x)-p\cdot v \leqslant 0.$$

Note that $g(0) = 0$, so 0 is a supergradient of g at 0.

In fact, $\partial g(0) = \{0\}$. For if $q \in \partial g(0)$, we have

$$\begin{aligned} g(0)+q\cdot v &\geqslant g(v) \\ 0+q\cdot v &\geqslant f(x+v)-f(x)-p\cdot v \\ f(x)+(p+q)\cdot v &\geqslant f(x+v), \end{aligned}$$

which implies $p+q \in \partial f(x)$, so $q = 0$.

Thus by Lemma 13, for all $v \in \alpha B$, Needs work.

$$0\cdot v \geqslant g'(0;v) = \lim_{\lambda\downarrow 0}\frac{g(\lambda v)}{\lambda}.$$

by ???? g is bounded on B, so for every $\eta > 0$, there is some $\delta > 0$ such that $v \in \delta B$ implies

$$0 \geqslant \frac{g(v)}{\|v\|} \geqslant -\eta.$$

But this asserts that g is differentiable at 0 and $g'(0) = 0$.
****

Now note that if g is differentiable at 0 with $g'(0) = 0$, then f is differentiable at x with $f'(x) = p$. ∎ Elaborate.

The next results may be found in Fenchel [2, Theorems 33–34, pp. 86–87].

Theorem 15 *Let f be a concave function on the open convex set C. For each direction v, $f'(x; v)$ is a lower semicontinuous function of x, and $\{x : f'(x; v) + f'(x; -v) < 0\}$ has Lebesgue measure zero. Thus $f'(x; v) + f'(x; -v) = 0$ almost everywhere, so f has a directional derivative in the direction v almost everywhere. Moreover, when the directional derivative exists, then it is continuous in x.*

Proof: Since f is concave, it is continuous (Theorem t:convex-continuity). Fix v and choose $\lambda_n \downarrow 0$. Then $g_n(x) := \frac{f(x+\lambda_n v)-f(x)}{\lambda_n}$ is continuous and by Lemma 10, $g_n(x) \uparrow f'(x; v)$. Thus a wll knwon result implies that $f'(x; v)$ is lower semicontinuous.
***** ∎

The next fact may be found in Fenchel [2, Theorem 35, p. 87ff], or Katzner [3, Theorems B.5-1 and B.5-2].

Fact 16 *If $f \colon C \subset \boldsymbol{R}^{\mathrm{n}} \to \boldsymbol{R}$ is twice differentiable, then the Hessian H_f is negative semidefinite if and only if f is concave. If H_f is everywhere negative definite, then f is strictly concave.*

****

4 Support functions

The Separating Hyperplane Theorem is the basis for a number of results concerning closed convex sets. Given any set A in $\boldsymbol{R}^{\mathrm{n}}$ its **closed convex hull**, denoted $\overline{\mathrm{co}}\, A$, is the intersection of all closed convex sets that include A. That is,

$$\overline{\mathrm{co}}\, A = \bigcap \{C : A \subset C \text{ and } H \text{ is closed and convex}\}.$$

It is of course the smallest closed convex set that includes A. If A is empty, then it is closed and convex, so $\overline{\mathrm{co}}\, A$ is empty. If A is nonempty, then $\overline{\mathrm{co}}\, A$ is nonempty since $\boldsymbol{R}^{\mathrm{n}}$ itself is closed and convex. Less obvious is the following.

Theorem 17 *Let A be a subset of $\boldsymbol{R}^{\mathrm{n}}$. Then*

$$\overline{\mathrm{co}}\, A = \bigcap \{H : A \subset H \textit{ and } H \textit{ is a closed half space}\}.$$

In particular, a closed convex set is the intersection of all the closed half spaces that include it.

Proof: Clearly $\overline{\text{co}}\, A$ is included in the intersection since every closed half space is also a closed convex set. It is also clear that the result is true for $A = \varnothing$. So assume A, and hence $\overline{\text{co}}\, A$, is nonempty.

It suffices to show that if $x \notin \overline{\text{co}}\, A$, then there is a closed half space that includes $\overline{\text{co}}\, A$ but does not contain x. By the Separating Hyperplane Theorem there is a nonzero p that strongly separates the closed convex set $\overline{\text{co}}\, A$ from the compact set $\{x\}$. But this clearly implies that there is a closed half space of the form $[p \geqslant \alpha]$ that includes $\overline{\text{co}}\, A$, but doesn't contain x. ∎

The **support function** μ_A of a set A is a handy way to summarize all the closed half spaces that included A. It is defined by[3]

$$\mu_A(p) = \inf\{p \cdot x : x \in A\}.$$

We allow for the case that $\mu_A(p) = -\infty$. Note that $\mu_\varnothing$ is the improper concave function $+\infty$. Also note that the infimum may not actually be attained even if it is finite. For instance, consider the closed convex set $A = \{(x, y) \in \boldsymbol{R}^2_{++} : xy \geqslant 1\}$, and let $p = (0, 1)$. Then $\mu_A(p) = 0$ even though $p \cdot (x, y) = y > 0$ for all $(x, y) \in A$. If A is compact, then of course μ_A is always finite, and there is some point in A where the infimum is actually a minimum.

Theorem 17 immediately implies yields the following description of $\overline{\text{co}}\, A$ in terms of μ_A.

Theorem 18 *For any set A in $\boldsymbol{R}^{\mathrm{n}}$,*

$$\overline{\text{co}}\, A = \{x \in \boldsymbol{R}^{\mathrm{n}} : \forall p \in \boldsymbol{R}^{\mathrm{n}} \ p \cdot x \geqslant \mu_A(p)\}.$$

Moreover, $\mu_A = \mu_{\overline{\text{co}}\, A}$.

Proof: Observe that

$$C := \{x \in \boldsymbol{R}^{\mathrm{n}} : \forall p \in \boldsymbol{R}^{\mathrm{n}} \ p \cdot x \geqslant \mu_A(p)\} = \bigcap\{[p \geqslant \mu_A(p)] : p \in \boldsymbol{R}^{\mathrm{n}}\}$$

is an intersection of closed half spaces. By definition, if $x \in A$, then $p \cdot x \geqslant \mu_A(p)$, that is, $A \subset [p \geqslant \mu_A(p)]$. Thus by Theorem 17, $\overline{\text{co}}\, A \subset C$.

For the reverse inclusion, suppose $x \notin \overline{\text{co}}\, A$. By the Separating Hyperplane Theorem there is a nonzero p such $\overline{\text{co}}\, A \subset [p \geqslant \alpha]$ and $p \cdot x < \alpha$. Since $A \subset \overline{\text{co}}\, A$ we have $\mu_A(p) = \inf\{p \cdot x : x \in A\} > p \cdot x$, so $x \notin C$.

To see that $\mu_A = \mu_{\overline{\text{co}}\, A}$ first note that $\mu_A \geqslant \mu_{\overline{\text{co}}\, A}$ since $A \subset \overline{\text{co}}\, A$. The first part of the theorem implies $\mu_{\overline{\text{co}}\, A} \geqslant \mu_A$. ∎

Lemma 19 *The support function μ_A is concave and positively homogeneous of degree 1, that is, $\mu_A(\lambda p) = \mu_A(\lambda p)$ for all p and all $\lambda \geqslant 0$.*

Proof: Each x defines a linear (and therefore concave) function ℓ_x via $\ell_x \colon p \mapsto p \cdot x$. Thus $\mu_A = \inf_{x \in A} \ell_x$ is concave. Homogeneity is obvious. ∎

[3] Fenchel [2] and Roko and I [1, p. 219] define $h_A(p) = \sup\{p \cdot x : x \in A\}$, which makes it convex rather than concave, and $h_A(p) = -\mu_A(-p)$. The definition in these notes follows Mas-Colell, Whinston, and Green [4], and may be more useful to economists.

If the infimum of p is actually achieved at a point in A, we can say more. By Theorem 18 we might as well assume that A is closed and convex.

Theorem 20 *Let C be a closed convex set. Then x is a supergradient of the support function μ_C at p if and only if x belongs to C and minimizes p over C. In other words,*

$$\partial\mu_C(p) = \{x \in C : p \cdot x = \mu_C(p)\}.$$

Proof: Recall that x satisfies the supergradient inequality if and only if for every q

$$\mu_C(p) + x \cdot (q - p) \geqslant \mu_C(q).$$

I first claim that if x does not belong to C, it is not a supergradient of μ_C at p. For if $x \notin C$, then since C is closed and convex, by Theorem 18 there is some q for which $q \cdot x < \mu_C(q)$. Thus for $\lambda > 0$ large enough, $\lambda q \cdot x < \mu_C(\lambda q) + \big(p \cdot x - \mu_C(p)\big)$. Rearranging terms violates the supergradient inequality applied to λq.

Now suppose that x is a supergradient of μ_C at p. Then setting $q = 0$ in the supergradient inequality, we conclude that $\mu_C(p) \geqslant p \cdot x$. But x must belong to C, so the definition of μ_C implies $\mu_C(p) \leqslant p \cdot x$. Thus, $\partial\mu_C(p) \subset \{x \in C : p \cdot x = \mu_C(p)\}$.

Suppose now that x belongs to C and $p \cdot x = \mu_C(p)$, that is, x minimizes p over C. By the definition of μ_C, for any $q \in \boldsymbol{R}^{\mathrm{n}}$, $q \cdot x \geqslant \mu_C(q)$. Now add $\mu_C(p) - p \cdot x = 0$ to the left-hand side of the inequality to obtain the supergradient inequality. Thus $\{x \in C : p \cdot x = \mu_C(p)\} \subset \partial\mu_C(p)$, completing the proof. ∎

Corollary 21 *Let C be a closed convex set. Suppose x belongs to C and strictly minimizes p over C. Then μ_C is differentiable at p and*

$$\mu_C'(p) = x.$$

Proof: This follows from Theorem 20 and Theorem 14. ∎

Example 22 Let's look at $C = \{(x_1, x_2) \in \boldsymbol{R}^2_{++} : x_1 x_2 \geqslant 1\}$. This is a closed convex set and its support function is easily calculated. If $p \notin \boldsymbol{R}^2_+$, then $\mu_C(p) = -\infty$. For $p \geqq 0$, it not hard to see that $\mu_C(p) = 2\sqrt{p_1 p_2}$, which has no supergradient when $p_1 = 0$ or $p_2 = 0$.

(To see this, consider first the case $p \geqq 0$. The Lagrangean for the minimization problem is $p_1 x_1 + p_2 x_2 + \lambda(1 - x_1 x_2)$. By the Lagrange Multiplier Theorem, the first order conditions are $p_1 - \lambda x_1^* = 0$ and $p_2 - \lambda x_2^* = 0$. Thus $x_1^* x_2^* = \frac{p_1 p_2}{\lambda^2}$, so $\lambda = \sqrt{p_1 p_2}$. Thus $x_1^* = \sqrt{\frac{p_1}{p_2}}$ and $x_2^* = \sqrt{\frac{p_2}{p_1}}$ and $\mu_C(p) = p_1 x_1^* + p_2 x_2^* = 2\sqrt{p_1 p_2}$.

Now suppose some $p_i < 0$. For instance, suppose $p_2 < 0$. Then $p \cdot (\varepsilon, \frac{1}{\varepsilon}) \to -\infty$ as $\varepsilon \to 0$, so $\mu_C(p) = -\infty$.) □

5 Maxima of concave functions

Concave functions have two important properties. One is that any local maximum is a global maximum. The other is that first order conditions are necessary and sufficient for a maximum.

Theorem 23 (Concave local maxima are global) *Let $f\colon C \to \boldsymbol{R}$ be a concave function (C convex). If x^* is a local maximizer of f, then it is a global maximizer of f over C.*

Proof: Let x belong to C. Then for small $\lambda > 0$, $f(x^*) \geqslant f\big(\lambda x + (1-\lambda)x^*\big)$. (Why?) By the definition of concavity,

$$f\big(\lambda x + (1-\lambda)x^*\big) \geqslant \lambda f(x) + (1-\lambda)f(x^*).$$

Thus $f(x^*) \geqslant \lambda f(x) + (1-\lambda)f(x^*)$, which implies $f(x^*) \geqslant f(x)$. ∎

Corollary 24 *If f is strictly concave, a local maximum is a strict global maximum.*

Theorem 25 (First order conditions for concave functions) *Suppose f is concave on a convex set $C \subset \boldsymbol{R}^n$. A point x^* in C is a global maximum point of f if and only 0 belongs to the superdifferential $\partial f(x^*)$.*

Proof: Note that x^* is a global maximum point of f if and only if

$$f(x^*) + 0 \cdot (y - x^*) \geqslant f(y)$$

for all y in C, but this is just the supergradient inequality for 0. ∎

In particular, this result shows that f is superdifferentiable at any maximum point, even if it is not an interior point. The next result is immediate.

Corollary 26 *Suppose f is concave on a convex neighborhood $C \subset \boldsymbol{R}^n$ of x^*, and differentiable at x^*. If $f'(x^*) = 0$, then f has a global maximum over C at x^*.*

Note that the conclusion of Theorem 23 does not hold for quasiconcave functions. For instance,

$$f(x) = \begin{cases} 0 & x \leqslant 0 \\ x & x \geqslant 0, \end{cases}$$

has a local maximum at -1, but it is not a global maximum over $\boldsymbol{R}$. However, if f is explicitly quasiconcave, then we have the following.

Theorem 27 (Local maxima of explicitly quasiconcave functions) *Let $f\colon C \to \boldsymbol{R}$ be an explicitly quasiconcave function (C convex). If x^* is a local maximizer of f, then it is a global maximizer of f over C.*

Proof: Let x belong to C and suppose $f(x) > f(x^*)$. Then by the definition of explicit quasiconcavity, for any $1 > \lambda > 0$, $f\big(\lambda x + (1-\lambda)x^*\big) > f(x^*)$. Since $\lambda x + (1-\lambda)x^* \to x^*$ as $\lambda \to 0$ this contradicts the fact that f has a local maximum at x^*. ∎

6 Supergradients and cyclically monotone mappings

Recall that a real function $g\colon X \subset \boldsymbol{R} \to \boldsymbol{R}$ is increasing if $x \geqslant y$ implies $g(x) \geqslant g(y)$. Another way to say this is $\big[g(x) - g(y)\big](x - y) \geqslant 0$ for all x, y. Or equivalently, g is nondecreasing if

$$g(x)(y - x) + g(y)(x - y) \leqslant 0 \quad \text{for all } x, y.$$

More generally, a correspondence $\varphi : X \subset \boldsymbol{R}^{\mathrm{m}} \twoheadrightarrow \boldsymbol{R}^{\mathrm{m}}$ is **monotone increasing** if

$$(p_x - p_y) \cdot (x - y) \geqslant 0 \quad \text{for all } x, y \in X, \text{ and all } p_x \in \varphi(x), p_y \in \varphi(y).$$

We could also write this as $p_x \cdot (y - x) + p_y \cdot (x - y) \leqslant 0$. A mapping φ is **monotone decreasing** if the reverse inequality always holds.

There is a natural generalization of these conditions. A finite sequence $x_0, x_1, \dots, x_n, x_{n+1}$ with $x_{n+1} = x_0$ is sometimes called a **cycle**. A mapping $g\colon U \subset \boldsymbol{R}^{\mathrm{m}} \to \boldsymbol{R}^{\mathrm{m}}$ is called **cyclically monotone increasing** if for every cycle $x_0, x_1, \dots, x_n, x_{n+1} = x_0$ in U, we have

$$g(x_0) \cdot (x_1 - x_0) + g(x_1) \cdot (x_2 - x_1) + \cdots + g(x_n) \cdot (x_0 - x_n) \leqslant 0.$$

If the same sum is always $\geqslant 0$, we shall say that g is **cyclically monotone decreasing**.

More generally, a correspondence $\varphi\colon U \subset \boldsymbol{R}^{\mathrm{m}} \twoheadrightarrow \boldsymbol{R}^{\mathrm{m}}$ is called **cyclically monotone increasing**[4] if for every cycle $(x_0, p_0), (x_1, p_1), \dots, (x_{n+1}, p_{n+1}) = (x_0, p_0)$ in the graph of φ, that is, with $p_i \in \varphi(x_i)$ for all i, we have

$$p_0 \cdot (x_1 - x_0) + p_1 \cdot (x_2 - x_1) + \cdots + p_n \cdot (x_0 - x_n) \leqslant 0.$$

We mention that if $m = 1$ ($\boldsymbol{R}^{\mathrm{m}} = \boldsymbol{R}$) then a function g is cyclically monotone if and only if it is monotone. For $m \geqslant 2$, there are monotone functions that are not cyclically monotone, see Rockafellar [6, p. 240].

The next result is a simple corollary of Theorem 1.

Corollary 28 (Cyclic monotonicity of the derivative) *Let f be concave and differentiable on a convex open set $U \subset \boldsymbol{R}^{\mathrm{m}}$. Then the gradient mapping $x \mapsto f'(x)$ is cyclically monotone decreasing. That is, for any cycle $x_0, x_1, \dots, x_n, x_{n+1}$ in U with $x_{n+1} = x_0$, we have*

$$\sum_{k=0}^{n} f'(x_k) \cdot (x_{k+1} - x_k) \geqslant 0.$$

Proof: By Theorem 1, $f'(x_k) \cdot (x_{k+1} - x_k) \geqslant f(x_{k+1}) - f(x_k)$. Summing both sides gives

$$\sum_{k=0}^{n} f'(x_k) \cdot (x_{k+1} - x_k) \geqslant \sum_{k=0}^{n} \big[f(x_{k+1}) - f(x_k)\big] = 0,$$

where the last equality follows from the fact that $x_{n+1} = x_0$. ∎

[4] Most authors define monotone and cyclically monotone correspondences to be increasing, and do not make a definition for decreasing monotonicity. This is because mathematicians find convex functions (such as norms) to be natural, and as we shall see below there is an important relationship between convex functions and (cyclically) monotone increasing mappings. Economists however find concave functions to be naturally occurring (as in production functions) so it seems natural to introduce a term for (cyclically) monotone decreasing mappings. Just keep in mind that for every statement about convex functions, there is a corresponding one for concave functions derived by replacing f by $-f$.

Note that the gradient of a convex function is cyclically monotone increasing.

The remarkable fact is that the supergradient correspondence is characterized by cyclic monotonicity. The next result is due to Rockafellar, and may be found (in different terminology) in his book [6, Theorem 24.8, p. 238].

Theorem 29 (Rockafellar) *Let $C \subset \boldsymbol{R}^{\mathrm{m}}$ be a nonempty convex set and let $\varphi : C \twoheadrightarrow \boldsymbol{R}^{\mathrm{m}}$ be a correspondence with nonempty values. Then φ is cyclically monotone decreasing if and only if there is an upper semicontinuous concave function $f : C \to \boldsymbol{R}$ satisfying*

$$\varphi(x) \subset \partial f(x) \quad \textit{for every } x \in C.$$

Proof: If $\varphi(x) \subset \partial f(x)$ for a concave f, then the definition of $\partial f(x)$ and the same argument used to prove Corollary 28 show that φ is cyclically monotone decreasing.

For the converse, assume φ is cyclically monotone decreasing. Fix any point x_0 in C and fix $p_0 \in \varphi(x_0)$. Define the function $f : C \to \boldsymbol{R}$ by

$$f(x) = \inf\{p_0 \cdot (x_1 - x_0) + \cdots + p_n \cdot (x - x_n) : p_i \in \varphi(x_i), x_i \in C, i = 1, \ldots, n\}.$$

Now, having fixed $(x_0, p_0), \ldots, (x_n, p_n)$, the sum

$$g(x) = p_0 \cdot (x_1 - x_0) + \cdots + p_n \cdot (x - x_n)$$

is an affine function of x. (The construction of such functions g is illustrated in Figures 2 and 3. Note the dependence on the order of the x_i.) Thus f is the pointwise infimum of a collection of affine functions, so it is concave. Furthermore, each of these functions is continuous, so f is upper semicontinuous. (Exercise: The pointwise infimum f of a family $\{g_i\}$ of upper semicontinuous functions is upper semicontinuous. Hint: $[f \geqslant \alpha] = \bigcap_i [g_i \geqslant \alpha]$.)

Cyclic monotonicity implies that the infimum defining f is finite, that is, $f(x) > -\infty$ for every $x \in C$. To see this, fix some p in $\varphi(x)$. Then by cyclic monotonicity

$$p_0 \cdot (x_1 - x_0) + \cdots + p_n \cdot (x - x_n) + p \cdot (x_0 - x) \geqslant 0.$$

Setting $m(x) = p \cdot (x - x_0)$ and rearranging gives

$$p_0 \cdot (x_1 - x_0) + \cdots + p_n \cdot (x - x_n) \geqslant m(x).$$

Therefore $f(x) \geqslant m(x) > -\infty$ for any x.

We claim that f is the desired function. That is, any x, y in C and any $p \in \varphi(x)$ satisfy the supergradient inequality

$$f(x) + p \cdot (y - x) \geqslant f(y).$$

To see this, let $\varepsilon > 0$ be given. Then by the definition of f, there is a finite sequence $(x_0, p_0), \ldots, (x_n, p_n)$ in the graph of φ with $f(x) + \varepsilon > p_0 \cdot (x_1 - x_0) + \cdots + p_n \cdot (x - x_n) \geqslant f(x)$. Extend this sequence by appending (x, p). Then, again by the definition of f,

$$p_0 \cdot (x_1 - x_0) + \cdots + p_n \cdot (x - x_n) + p \cdot (y - x) \geqslant f(y).$$

Combining these gives

$$f(x) + \varepsilon + p \cdot (y - x) > p_0 \cdot (x_1 - x_0) + \cdots + p_n \cdot (x - x_n) + p \cdot (y - x) \geqslant f(y).$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $f(x) + p \cdot (y - x) \geqslant f(y)$, so indeed $\varphi(x) \subset \partial f(x)$. ∎

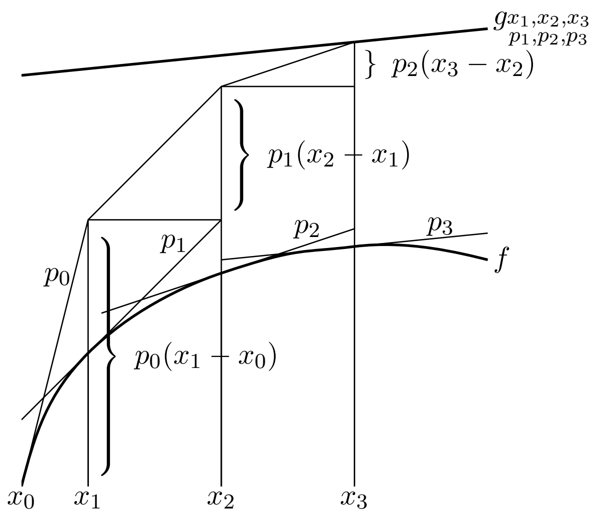

Figure 2. The function $g_{p_1,p_2,p_3}^{x_1,x_2,x_3}(y) = p_0 \cdot (x_1 - x_0) + p_1 \cdot (x_2 - x_1) + p_2 \cdot (x_3 - x_2) + p_3 \cdot (y - x_3)$, where each p_i is taken from $\partial f(x_i)$.

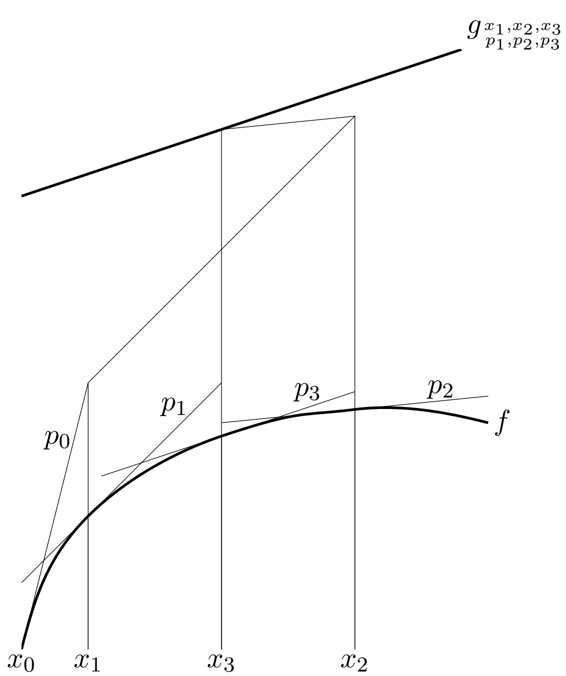

Figure 3. Another version of $g_{p_1,p_2,p_3}^{x_1,x_2,x_3}(y) = p_0 \cdot (x_1 - x_0) + p_1 \cdot (x_2 - x_1) + p_2 \cdot (x_3 - x_2) + p_3 \cdot (y - x_3)$, where the x_i have been reordered.

7 Cyclic monotonicity and second derivatives

From Corollary 28 we know that the gradient of a concave function $f\colon C \to \boldsymbol{R}$, where C is an open convex set in $\boldsymbol{R}^{\mathrm{n}}$, satisfies

$$\sum_{k=0}^{n} f'(x_k) \cdot (x_{k+1} - x_k) \geqslant 0.$$

Specializing this yields

$$f'(x_0) \cdot (x_1 - x_0) + f'(x_1) \cdot (x_0 - x_1) \geqslant 0,$$

which can be rearranged as

$$\big(f'(x_1) - f'(x_0)\big) \cdot (x_1 - x_0) \leqslant 0.$$

Consider a point x in C and choose v so that $x \pm v$ to C. Then by the argument above with $x_0 = x$ and $x_1 = x + v$,

$$\big(f'(x + \lambda v) - f'(x)\big) \cdot (\lambda v) \leqslant 0.$$

Dividing by the positive quantity λ^2 implies

$$v \cdot \frac{\big(f'(x + \lambda v) - f'(x)\big)}{\lambda} \leqslant 0.$$

Define the function $g\colon (-1, 1) \to \boldsymbol{R}$ by

$$g(\lambda) = v \cdot f'(x + \lambda v).$$

In particular, if f is twice differentiable, then by the Chain Rule

$$D^2 f(x)(v, v) = g'(0) = \lim_{\lambda \to 0} v \cdot \frac{g(\lambda) - g(0)}{\lambda} \leqslant 0.$$

Thus the Hessian matrix $f''(x)$ is negative semidefinite, which gives another proof of half of Fact 16.

Now let's return to support functions.

Lemma 30 *Suppose $x(p)$ minimizes $p \cdot x$ over the nonempty set A. Suppose further that it is the unique minimizer of $p \cdot x$ over $\overline{\mathrm{co}}\, A$. If $\frac{\partial^2 \mu_C(p)}{\partial p_i^2}$ exists (or equivalently $\frac{\partial x(p)}{\partial p_i}$ exists), then*

$$\frac{\partial x(p)}{\partial p_i} \leqslant 0.$$

Proof: This follows from Corollary 21 and the discussion above. ∎

This by the way, summarizes almost everything interesting we now about cost minimization.

References

[1] C. D. Aliprantis and K. C. Border. 1999. *Infinite dimensional analysis: A hitchhiker's guide*, 2d. ed. Berlin: Springer–Verlag.

[2] W. Fenchel. 1953. Convex cones, sets, and functions. Lecture notes, Princeton University, Department of Mathematics. From notes taken by D. W. Blackett, Spring 1951.

[3] D. W. Katzner. 1970. *Static demand theory*. London: Macmillan.

[4] A. Mas-Colell, M. D. Whinston, and J. R. Green. 1995. *Microeconomic theory*. Oxford: Oxford University Press.

[5] R. R. Phelps. 1993. *Convex functions, monotone operators and differentiability*, 2d. ed. Number 1364 in Lecture Notes in Mathematics. Berlin: Springer–Verlag.

[6] R. T. Rockafellar. 1970. *Convex analysis*. Number 28 in Princeton Mathematical Series. Princeton: Princeton University Press.

[7] H. L. Royden. 1988. *Real analysis*, 3d. ed. New York: Macmillan.