Ec 181 KC Border
Convex Analysis and Economic Theory AY 2019–2020

Topic 14: Subgradients

14.1 Subgradients

We have seen in Theorem 13.3.3 that a regular convex function $f\colon \boldsymbol{R}^{\mathrm{m}} \to \boldsymbol{R}$ is the supremum of the affine functions that it dominates. Suppose this supremum is attained as a maximum for some affine function at the point x. That is, assume that

$$g\colon y \mapsto p \cdot y - \beta$$

is an affine function that satisfies

$$g \leqq f \text{ and } g(x) = f(x). \tag{M}$$

See Figure 14.1.1 or Figure 14.1.2. There are a number of equivalent statements that summarize this relation.

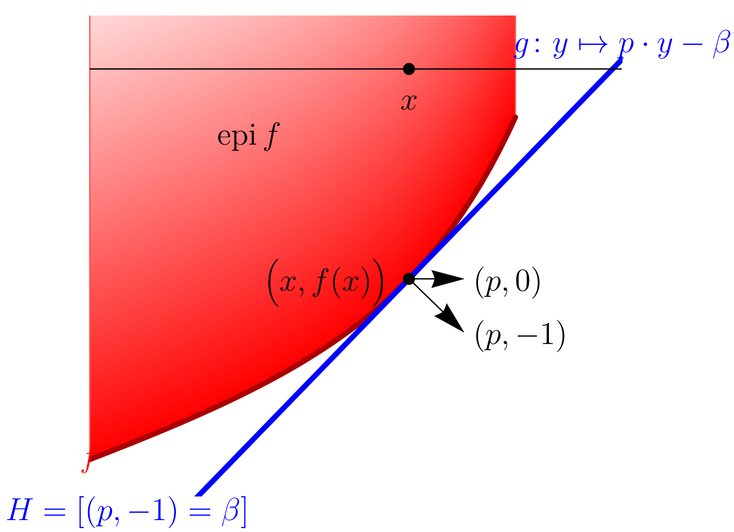

Figure 14.1.1. The affine function $g\colon y \mapsto p \cdot y - \beta$ satisfies $g \leqq f$ and $g(x) = f(x)$. Equivalently, the hyperplane $H = \{(y, \alpha) \in X \times \boldsymbol{R} : (p, -1) \cdot (y, \alpha) = \beta\}$ supports epi f at the point $\big(x, f(x)\big)$, which maximizes $(p, -1)$ over epi f and the maximum value is β.

14.1.1 Proposition *Let $f\colon X \to \boldsymbol{R}$ be a proper convex function. The following statements are equivalent.*

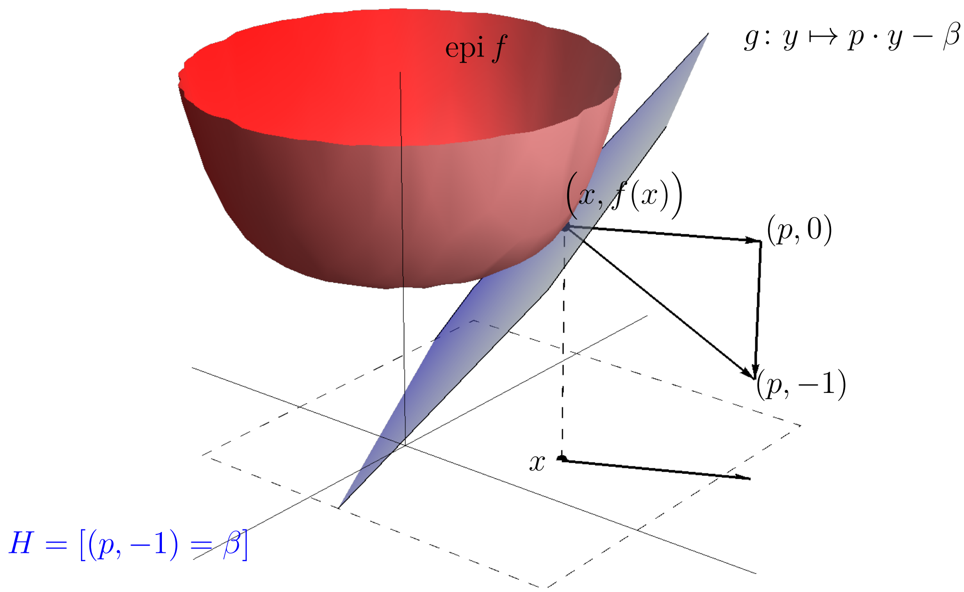

Figure 14.1.2. The affine function $g\colon y \mapsto p \cdot y - \beta$ satisfies $g \leqq f$ and $g(x) = f(x)$. Equivalently, The hyperplane $H = \{(y, \alpha) \in X \times \boldsymbol{R} : (p, -1) \cdot (y, \alpha) = \beta\}$ supports epi f at the point $\big(x, f(x)\big)$, which maximizes $(p, -1)$ over epi f and the maximum values is β.

1. *The affine function* $g\colon y \mapsto p \cdot y - \beta$ *satisfies* (**M**).

2. $\beta = p \cdot x - f(x) = \max_y p \cdot y - f(y)$.

3. *The hyperplane* $H = \{(y,\alpha) \in X \times \boldsymbol{R} : (p,-1)\cdot(y,\alpha) = \beta\}$ *supports* epi f *at the point* $\big(x, f(x)\big)$ *as a maximizer and the maximum value is* $\beta = p\cdot x - f(x)$. *Or in other words,*
$$\beta = \pi_{\operatorname{epi} f}\Big((p,-1)\Big),$$
where $\pi_{\operatorname{epi} f}$ *is the profit (support) function of the epigraph of* f.

4. $\beta = p \cdot x - f(x)$ *and*
$$(\,\forall y\,)\ [\, f(x) + p \cdot (y - x) \leqslant f(y)\,]. \tag{S}$$

Proof: The proof is trivial, but sufficiently important that I'll write it out. Statement (1) can be written as

$$(\,\forall y\,)\ [\, p \cdot y - \beta \leqslant f(y)\,]\ \&\ p \cdot x - \beta = f(x).$$

This rearranges to become

$$(\,\forall y\,)\ [\, \beta \geqslant p \cdot y - f(y)\,]\ \&\ \beta = p \cdot x - f(x),$$

so eliminating β gives

$$(\,\forall y\,)\ [\, p \cdot x - f(x) \geqslant p \cdot y - f(y)\,] \tag{1}$$

which is equivalent to statement (2). It can also be rewritten as

$$(\,\forall y\,)\ \Big[\, (p,-1) \cdot \Big(x, f(x)\Big) \geqslant (p,-1) \cdot \Big(y, f(y)\Big)\,\Big]$$

which is equivalent to

$$(\,\forall y\,)\ \ (\,\forall \alpha \geqslant f(y)\,)\ \Big[\, (p,-1) \cdot \Big(x, f(x)\Big) \geqslant (p,-1) \cdot (y, \alpha)\,\Big],$$

which is (3). Return now to (1), and rearrange it to get

$$(\,\forall y\,)\ [\, f(x) + p \cdot (y - x) \leqslant f(y)\,],$$

which is just statement (4). ∎

14.1.2 Definition *Relation* (**S**) *above is called the* ***subgradient inequality*** *for* f *at* x. *If a vector* p *satisfies the subgradient inequality for* f *at* x, *it is called a* ***subgradient of*** $\boldsymbol{f}$ ***at*** $\boldsymbol{x}$. *The set of subgradients of* f *at* x *is called the* ***subdifferential*** *of* f *at* x, *and is denoted* $\partial f(x)$. *A function* f *is* ***subdifferentiable at*** $\boldsymbol{x}$ *if the subdifferential* $\partial f(x)$ *is nonempty.*

There is nothing in the subgradient inequality that requires f to be convex, so we can refer to the subgradient of any function. But the following result shows that the concept is most useful for convex functions. Unless I mention otherwise, from now on I shall only talk about subgradients of convex functions.

14.1.3 Proposition *If C is a nonempty convex subset of $\boldsymbol{R}^{\mathrm{n}}$ and $f\colon C \to \boldsymbol{R}$ has a subgradient at each point of C, then f is convex.*

Moreover, if for each $x \in C$, there is a subgradient p that satisfies the subgradient inequality with strict inequality, that is,

$$f(y) > f(x) + p \cdot (y - x) \quad \text{for all } y \neq x,\, y \in C,$$

then f is strictly convex. Conversely, if f is strictly convex and has a subgradient, then the subgradient inequality is strict (except when $y = x$).

Proof: Let $x, y \in C$, $x \neq y$, let $0 < \lambda < 1$, and let p be a subgradient at $z = (1-\lambda)x + \lambda y$. By the subgradient inequality, $f(x) \geqslant f(z) + p \cdot (x - z)$ and $f(y) \geqslant f(z) + p \cdot (y - z)$, so

$$(1-\lambda)f(x) + \lambda f(y) \geqslant (1-\lambda)f(z) + \lambda f(z) + p \cdot \Big[(1-\lambda)(x-z) + \lambda(y-z\Big] = f(z).$$

That is, f is convex.

The proof of strict convexity is the same, replacing $\geqslant$ by $>$.

For the converse, assume that f is strictly convex, and that p is a subgradient of f at x. Let $y \neq x$ and let $z = (x+y)/2$. Then by strict convexity and the subgradient inequality we have

$$\tfrac{1}{2}f(x) + \tfrac{1}{2}f(y) > f(z) \geqslant f(x) + p \cdot (z - x).$$

But $z - x = (y - x)/2$, so subtracting $f(x)/2$ from the outer inequality gives

$$\tfrac{1}{2}f(y) > \tfrac{1}{2}f(x) + \tfrac{1}{2}p \cdot (y - x).$$

Draw some pictures!

Multiplying by 2 gives the strict subgradient inequality

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

∎

We now mention a few properties of the subdifferential of a convex function.

14.1.4 Lemma *The subdifferential $\partial f(x)$ of a convex function is a closed convex (possibly empty) set.*

If f is a proper convex function and f is subdifferentiable at x, then $x \in \operatorname{dom} f$.

If there exists some point x at which a convex function f is finite and subdifferentiable, then f is proper.

Proof: The subdifferential $\partial f(x)$ is the intersection of closed half-spaces

$$\partial f(x) = \bigcap_y \{p : p \cdot (y-x) \leqslant f(y) - f(x)\}$$

and so closed and convex.

Assume that f is proper. Then there exists some $y \in \operatorname{dom} f$, so $f(y)$ is finite. By the subgradient inequality at y, we have $f(x) \leqslant f(y) - p \cdot (y-x) < \infty$, so $x \in \operatorname{dom} f$.

If f is subdifferentiable at x and $f(x)$ is finite, then for every y, we have $f(y) \geqslant f(x) + p \cdot (y-x) > -\infty$, so f is proper. ∎

We shall not be very interested in subgradients of improper functions, but by definition the improper constant convex functions ∞ and $-\infty$ are everywhere subdifferentiable and every p is a subgradient.

14.1.5 Proposition *A proper convex function on $\boldsymbol{R}^{\mathrm{m}}$ is subdifferentiable at each point of the relative interior of its effective domain.*

Proof: Let f be a proper convex function, and let x belong to $\operatorname{ri}\operatorname{dom} f$. Observe that $\big(x, f(x)\big)$ belongs to the epigraph of f, but not to its relative interior. Since the epigraph is convex, the Supporting Hyperplane Theorem 8.4.4 asserts that there is a nonzero $(p, \lambda) \in \boldsymbol{R}^{\mathrm{m}} \times \boldsymbol{R}$ properly supporting the epigraph at $\big(x, f(x)\big)$ as a maximizer. That is,

$$p \cdot x + \lambda f(x) \geqslant p \cdot y + \lambda \alpha \qquad \text{for all } y \in \operatorname{dom} f \text{ and all } \alpha \geqslant f(y). \tag{2}$$

I claim that $\lambda < 0$: Choosing $y = x$ in (2) implies $\lambda f(x) \geqslant \lambda \alpha$ for $\alpha \geqslant f(x)$ so $\lambda \leqslant 0$. Suppose momentarily that $\lambda = 0$. Since x belongs to the relative interior of $\operatorname{dom} f$, for any z in $\operatorname{dom} f$ there is some $\varepsilon > 0$ such that $x \pm \varepsilon(x - z)$ belong to $\operatorname{dom} f$. Then (2) (with $y = x \pm \varepsilon(x-z)$) implies $p \cdot (x - z) = 0$. Thus $(p, 0) \cdot (z, \alpha) = (p, 0) \cdot \big(x, f(x)\big)$ for all $(z, \alpha) \in \operatorname{epi} f$. But this contradicts the properness of the support at $\big(x, f(x)\big)$. Therefore $\lambda < 0$.

Dividing (p, λ) by $-\lambda > 0$ implies that $\big((-1/\lambda)p, -1\big)$ also supports the epigraph as a maximizer, so $(-1/\lambda)p$ is a subgradient by Proposition 14.1.1. ∎

Non-subdifferentiability may occur on the boundary of the domain.

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

In the infinite-dimensional case, A. Brøndsted and R. T. Rockafellar [3] give an example of a lower semicontinuous proper convex function defined on the Fréchet space $\boldsymbol{R}^{\mathbb{N}}$ that is nowhere subdifferentiable. Their example is based on the set in Klee [5].

14.1.1 Supergradients

There is, of course, a similar concept for concave functions. Let $f\colon \boldsymbol{R}^{\mathrm{m}} \to \boldsymbol{R}^{\sharp}$ be a concave function. A vector p is a **supergradient of $\boldsymbol{f}$ at $\boldsymbol{x}$** if it satisfies the **supergradient inequality**

$$(\,\forall y\,)\ [\,f(x) + p\cdot(y-x) \geqslant f(y)\,]. \tag{S$'$}$$

The set of supergradients of f at x is called the **superdifferential** of f at x, and is also denoted $\partial f(x)$. If the superdifferential is nonempty at x, we say that f is **superdifferentiable** at x. Rockafellar [6, p. 308] uses the term subgradient to mean both subgradient and supergradient, and subdifferential to mean both subdifferential and superdifferential, but suggests that the above terminology as being more appropriate, so I shall use it.[1] The definitions are potentially inconsistent for affine functions, which are both concave and convex, but thanks to the following result it all works out.

14.1.7 Lemma *The affine function $f\colon x \mapsto p\cdot x - \beta$ satisfies $\partial f(x) = \{p\}$, whether f is viewed as concave or convex.*

Proof: Clearly p satisfies both the supergradient and subgradient inequalities. Now suppose q satisfies the supergradient inequality $p\cdot x - \beta + q\cdot(y-x) \geqslant p\cdot y - \beta$ for all y. Pick any v and set $y = x+v$ and conclude $q\cdot v \geqslant p\cdot v$, and do the same for $-v$. This shows that $(p-q)\cdot v = 0$ for all v, so $q = p$. Thus p is the unique solution of the supergradient inequality. Ditto for the subgradient inequality. ∎

It is clear that if f is either concave or convex, then

$$\partial(-f)(x) = -\partial f(x),$$

where ∂ indicates the superdifferential when preceding a concave function and the subdifferential when preceding a convex function.

14.1.2 Sub/supergradients and extrema

An immediate consequence of the definition is the following result, which we shall see later can be interpreted as a kind of "first order condition" for a minimum.

14.1.8 Lemma *A proper convex function f is minimized at $x \in \operatorname{dom} f$ if and only if $0 \in \partial f(x)$.*

A proper concave function f is maximized at $x \in \operatorname{dom} f$ if and only if $0 \in \partial f(x)$.

The proof follows immediately by setting $p = 0$ in the subgradient inequality. This result also shows that a proper convex function f is subdifferentiable at any minimizer, even if it is not an interior point.

[1] Borwein and Zhu [2, p. 294] also adopt the terms supergradient and superdifferential, albeit in a more general framework.

14.1.3 The gradient is a subgradient

According to Proposition 14.1.1, when p is a subgradient at x, then x maximizes $p \cdot y - f(y)$. If f is differentiable at x, the first order condition for this maximum is that $p = f'(x)$, so the gradient of f is a subgradient. In fact, if $\partial f(x)$ is a singleton, then f is differentiable at x and $\partial f(x) = \{f'(x)\}$, see Theorem 18.3.1 below.

The following generalizations of Theorems 6.1.6 and 6.1.7 provide a useful way to characterize the convexity of differentiable functions on $\boldsymbol{R}^{\mathrm{m}}$.

14.1.9 Theorem *Suppose $f \colon \boldsymbol{R}^{\mathrm{m}} \to \boldsymbol{R}^{\sharp}$ is a proper convex function, and is differentiable at a point $x \in \operatorname{int} \operatorname{dom} f$. Then the gradient vector $f'(x)$ is a subgradient of f at x.*

Proof: If $y \notin \operatorname{dom} f$, then $f(y) = \infty$, so the subgradient inequality holds. So let $y \in \operatorname{dom} f$. Rewrite the definition of convexity as

$$f\Big(x + \lambda(y - x)\Big) \leqslant f(x) + \lambda\Big(f(y) - f(x)\Big).$$

Rearranging and dividing by $\lambda > 0$,

$$\frac{f\Big(x + \lambda(y - x)\Big) - f(x)}{\lambda} \leqslant f(y) - f(x).$$

Letting $\lambda \downarrow 0$, the left hand side converges to $f'(x) \cdot (y - x)$, and we see that $f'(x)$ satisfies the subgradient inequality. ■

The converse is true as the following argument shows.

14.1.10 Theorem *Let $f \colon \boldsymbol{R}^{\mathrm{m}} \to \boldsymbol{R}^{\sharp}$ be differentiable on a convex open set $U = \operatorname{dom} f$. Suppose that for every x and y in $\operatorname{dom} f$, we have $f(x) + f'(x) \cdot (y - x) \leqslant f(y)$. Then f is convex.*

Proof: For each $x \in U$, define the function h_x by $h_x(y) = f(x) + f'(x) \cdot (y - x)$. Each h_x is affine and so convex, $f \geqslant h_x$ for each $x \in U$, and $f(x) = h_x(x)$. Thus

$$f = \sup_{x \in U} h_x,$$

so by Exercise 1.3.3(5), f is convex. ■

The next result is now immediate.

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

14.1.4 Euler's Theorem for subgradients

A real-valued function f defined on a cone C in a vector space is **homogeneous of degree k** if for every $x \in C$ and $\lambda > 0$,

$$f(\lambda x) = \lambda^k f(x).$$

You may recall Euler's Theorem for Homogeneous Functions, which states that for a differentiable function f that is homogeneous of degree k if and only if $kf(x) = f'(x) \cdot x$ for every x. The following is a version of one half of this theorem in terms of subgradients. It may be found in Hendrickson and Buehler [4], who prove it in a particular infinite-dimensional context.

14.1.12 Theorem (Euler's Theorem for subgradients) *Let C be a convex cone in $\boldsymbol{R}^{\mathrm{n}}$, and let $f \colon C \to \boldsymbol{R}$ be homogeneous of degree k, and let p be a subgradient of f at x. (The function f is not necessarily convex.) Then*

$$kf(x) = p \cdot x.$$

Proof: Homogeneity and the subgradient inequality imply that for $\lambda > 0$, we have

$$\lambda^k f(x) = f(\lambda x) \geqslant f(x) + p \cdot (\lambda x - x),$$

so

$$(\lambda^k - 1)f(x) \geqslant (\lambda - 1)p \cdot x.$$

For $\lambda \neq 1$ division gives

$$\frac{\lambda^k - 1}{\lambda - 1} f(x) \geqslant p \cdot x \text{ for } \lambda > 1 \quad \text{and} \quad \frac{\lambda^k - 1}{\lambda - 1} f(x) \leqslant p \cdot x \text{ for } \lambda < 1. \tag{3}$$

By l'Hôpital's Rule,

$$\lim_{\lambda \to 1} \frac{\lambda^k - 1}{\lambda - 1} = \lim_{\lambda \to 1} \frac{k\lambda^{k-1}}{1} = k,$$

so (3) implies $kf(x) = p \cdot x$. ∎

14.2 Jensen's Inequality

14.2.1 Theorem *Let $f \colon \boldsymbol{R} \to \boldsymbol{R}^\sharp$ be a convex function and let X be a random variable taking values in* $\operatorname{dom} f$ *and satisfying $\boldsymbol{E}\,|X| < \infty$. Then*

$$f(\boldsymbol{E}\,X) \leqslant \boldsymbol{E}\,f(X).$$

Proof: The result is immediate if X is degenerate (constant). If X is not degenerate, then $\boldsymbol{E}\,X$ belongs to the interior of the convex hull of the range of X, which in turn belongs to $\operatorname{int}\operatorname{dom} f$. By Corollary 14.1.5 f is subdifferentiable at $\boldsymbol{E}\,X$. Let p belong to $\partial f(\boldsymbol{E}\,X)$. Evaluate the subgradient inequality at $\boldsymbol{E}\,X$:

$$f(\boldsymbol{E}\,X) + p(X - \boldsymbol{E}\,X) \leqslant f(X) \quad \text{for all values of } X,$$

so take expectations to get

$$f(\boldsymbol{E}\,X) + p\Big(\underbrace{\boldsymbol{E}(X - \boldsymbol{E}\,X)}_{=0}\Big) \leqslant \boldsymbol{E}\,f(X),$$

which is Jensen's Inequality. ∎

References

[1] J. M. Borwein and Q. J. Zhu. 1999. A survey of subdifferential calculus with applications. *Nonlinear Analysis* 38(6):687–773.
DOI: 10.1016/S0362-546X(98)00142-4

[2] ——— . 2005. *Techniques of variational analysis*. Number 20 in CMS Books in Mathematics/Ouvrages de Mathèmatiques de la SMC. New York: Springer Science+Business Media.

[3] A. Brøndsted and R. T. Rockafellar. 1965. On the subdifferentiability of convex functions. *Proceedings of the American Mathematical Society* 16:605–611.
`http://www.jstor.org/stable/2033889`

[4] A. D. Hendrickson and R. J. Buehler. 1971. Proper scores for probability forecasters. *Annals of Mathematical Statistics* 42:1916–1921.
`http://www.jstor.org/stable/2240117`

[5] V. L. Klee, Jr. 1963. On a question of Bishop and Phelps. *American Journal of Mathematics* 85:95–98. `http://www.jstor.org/stable/2373188`

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