Subdifferentials of convex functions

Jordan Bell
jordan.bell@gmail.com
Department of Mathematics, University of Toronto

April 21, 2014

Whenever we speak about a vector space in this note we mean a vector space over $\mathbb{R}$. If X is a topological vector space then we denote by X^* the set of all continuous linear maps $X \to \mathbb{R}$. X^* is called the *dual space of* X, and is itself a vector space.[1]

1 Definition of subdifferential

If X is a topological vector space, $f : X \to [-\infty, \infty]$ is a function, $x \in X$, and $\lambda \in X^*$, then we say that λ is a *subgradient of* f *at* x if[2]

$$f(y) \geq f(x) + \lambda(y - x), \qquad y \in X.$$

The *subdifferential of* f *at* x is the set of all subgradients of f at x and is denoted by $\partial f(x)$. Thus ∂f is a function from X to the power set of X^*, i.e. $\partial f : X \to 2^{X^*}$. If $\partial f(x) \neq \emptyset$, we say that f is *subdifferentiable at* x.

It is immediate that if there is some y such that $f(y) = -\infty$, then

$$\partial f(x) = \begin{cases} X^* & f(x) = -\infty \\ \emptyset & f(x) > -\infty \end{cases}, \qquad x \in X.$$

Thus, little is lost if we prove statements about subdifferentials of functions that do not take the value $-\infty$.

Theorem 1. *If X is a topological vector space, $f : X \to [-\infty, \infty]$ is a function and $x \in X$, then $\partial f(x)$ is a convex subset of X^*.*

[1] In this note, we are following the presentation of some results in Charalambos D. Aliprantis and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third ed., chapter 7. Three other sources for material on subdifferentials are: Jean-Paul Penot, *Calulus Without Derivatives*, chapter 3; Viorel Barbu and Teodor Precupanu, *Convexity and Optimization in Banach Spaces*, fourth ed., §2.2, pp. 82–125; and Jean-Pierre Aubin, *Optima and Equilibria: An Introduction to Nonlinear Analysis*, second ed., chapter 4, pp. 57–73.

[2] $\infty + \infty = \infty$, $-\infty - \infty = -\infty$, and $\infty - \infty$ is nonsense; if $a \in \mathbb{R}$, then $a - \infty = -\infty$ and $a + \infty = \infty$.

Proof. If $\lambda_1, \lambda_2 \in \partial f(x)$ and $0 \le t \le 1$, then of course $(1-t)\lambda_1 + t\lambda_2 \in X^*$. For any $y \in X$ we have

$$\begin{aligned} f(y) &= (1-t)f(y) + tf(y) \\ &\ge (1-t)f(x) + (1-t)\lambda_1(y-x) + tf(x) + t\lambda_2(y-x) \\ &= f(x) + \big((1-t)\lambda_1 + t\lambda_2\big)(y-x), \end{aligned}$$

showing that $(1-t)\lambda_1 + t\lambda_2 \in \partial f(x)$ and thus that $\partial f(x)$ is convex. □

To say that $0 \in \partial f(x)$ is equivalent to saying that $f(y) \ge f(x)$ for all $y \in X$ and so $f(x) = \inf_{y \in X} f(y)$. This can be said in the following way.

Lemma 2. *If X is a topological vector space and $f : X \to [-\infty, \infty]$ is a function, then x is a minimizer of f if and only if $0 \in \partial f(x)$.*

2 Convex functions

If X is a set and $f : X \to [-\infty, \infty]$ is a function, then the *epigraph of f* is the set

$$\operatorname{epi} f = \{(x, \alpha) \in X \times \mathbb{R} : \alpha \ge f(x)\},$$

and the *effective domain of f* is the set

$$\operatorname{dom} f = \{x \in X : f(x) < \infty\}.$$

To say that $x \in \operatorname{dom} f$ is equivalent to saying that there is some $\alpha \in \mathbb{R}$ such that $(x, \alpha) \in \operatorname{epi} f$. We say that f is *finite* if $-\infty < f(x) < \infty$ for all $x \in X$.

If X is a vector space and $f : X \to [-\infty, \infty]$ is a function, then we say that f is *convex* if $\operatorname{epi} f$ is a convex subset of the vector space $X \times \mathbb{R}$.

If X is a set and $f : X \to [-\infty, \infty]$ is a function, we say that f is *proper* if it does not take only the value ∞ and never takes the value $-\infty$. It is unusual to talk merely about proper functions rather than proper convex functions; we do so to make clear how convexity is used in the results we prove.

3 Weak-* topology

Let X be a topological vector space and for $x \in X$ define $e_x : X^* \to \mathbb{R}$ by $e_x \lambda = \lambda x$. The *weak-* topology* on X^* is the initial topology for the set of functions $\{e_x : x \in X\}$, that is, the coarsest topology on X^* such that for each $x \in X$, the function $e_x : X^* \to \mathbb{R}$ is continuous.

Lemma 3. *If X is a topological vector space, τ_1 is the weak-* topology on X^*, and τ_2 is the subspace topology on X^* inherited from $\mathbb{R}^X$ with the product topology, then $\tau_1 = \tau_2$.*

Proof. Let $\lambda_i \in X^*$ converge in τ_1 to $\lambda \in X^*$. For each $x \in X$, the function $e_x : X^* \to \mathbb{R}$ is τ_1 continuous, so $e_x\lambda_i \to e_x\lambda$, i.e. $\lambda_i x \to \lambda x$. But for $f_i \in \mathbb{R}^X$ to converge to $f \in \mathbb{R}^X$ means that for each x, we have $f_i(x) \to f(x)$. Thus λ_i converges to λ in τ_2. This shows that $\tau_2 \subseteq \tau_1$.

Let $x \in X$, and let $\lambda_i \in X^*$ converge in τ_2 to $\lambda \in X^*$. We then have $e_x\lambda_i = \lambda_i x \to \lambda x = e_x\lambda$; since λ_i was an arbitrary net that converges in τ_2, this shows that e_x is τ_2 continuous. Thus, we have shown that for each $x \in X$, the function e_x is τ_2 continuous. But τ_1 is the coarsest topology for which e_x is continuous for all $x \in X$, so we obtain $\tau_1 \subseteq \tau_2$. □

In other words, the weak-* topology on X^* is the topology of pointwise convergence. We now prove that at each point in the effective domain of a proper function on a topological vector space, the subdifferential is a weak-* closed subset of the dual space.[3]

Theorem 4. *If X is a topological vector space, $f : X \to (-\infty, \infty]$ is a proper function, and $x \in \operatorname{dom} f$, then $\partial f(x)$ is a weak-* closed subset of X^*.*

Proof. If $\lambda \in \partial f(x)$, then for all $y \in X$ we have

$$f(y) \geq f(x) + \lambda(y - x),$$

so, for any $v \in X$, using $y = v + x$,

$$f(v + x) \geq f(x) + \lambda v,$$

or,

$$\lambda v \leq f(x + v) - f(x);$$

this makes sense because $f(x)$ is finite. On the other hand, let $\lambda \in X^*$. If $\lambda v \leq f(x + v) - f(x)$ for all $v \in X$, then $\lambda(v - x) \leq f(v) - f(x)$, i.e. $f(v) \geq f(x) + \lambda(v - x)$, and so $\lambda \in \partial f(x)$. Therefore

$$\partial f(x) = \bigcap_{v \in X} \{\lambda \in X^* : \lambda v \leq f(x + v) - f(x)\}. \tag{1}$$

Defining $e_v : X^* \to \mathbb{R}$ for $v \in X$ by $e_v\lambda = \lambda v$, for each $v \in X$ we have

$$e_v^{-1}(-\infty, f(x + v) - f(x)] = \{\lambda \in X^* : \lambda v \leq f(x + v) - f(x)\}.$$

Because e_v is continuous, this inverse image is a closed subset of X^*. Therefore, each of the sets in the intersection (1) is a closed subset of X^*, and so $\partial f(x)$ is a closed subset of X^*. □

[3]cf. Charalambos D. Aliprantis and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third ed., p. 265, Theorem 7.13.

4 Support points

If X is set, A is a subset of X, and $f : X \to [-\infty, \infty]$ is a function, we say that $x \in X$ is a *minimizer of f over A* if

$$f(x) = \inf_{y \in A} f(y),$$

and that x is a *maximizer of f over A* if

$$f(x) = \sup_{y \in A} f(y).$$

If A is a nonempty subset of a topological vector space X and $x \in A$, we say that x is a *support point of A* if there is some nonzero $\lambda \in X^*$ for which x is a minimizer or a maximizer of λ over A. Moreover, x is a minimizer of λ over A if and only if x is a maximizer of $-\lambda$ over A. Thus, if we know that x is a support point of a set A, then we have at our disposal both that x is a minimizer of some nonzero element of X^* over A and that x is a maximizer of some nonzero element of X^* over A.

If x is a support point of A and A is not contained in the hyperplane $\{y \in X : \lambda y = \lambda x\}$, we say that A is *properly supported at x*. To say that A is not contained in the set $\{y \in X : \lambda y = \lambda x\}$ is equivalent to saying that there is some $y \in A$ such that $\lambda y \neq \lambda x$.

In the following lemma, we show that the support points of a set A are contained in the boundary ∂A of the set.

Lemma 5. *If X is a topological vector space, A is a subset of X, and x is a support point of A, then $x \in \partial A$.*

Proof. Because x is a support point of A there is some nonzero $\lambda \in X^*$ for which x is a maximizer of λ over A:

$$\lambda x = \sup_{y \in A} \lambda y.$$

As λ is nonzero there is some $y \in X$ with $\lambda y > \lambda x$. For any $t > 0$,

$$(1-t)\lambda x + t\lambda y = \lambda((1-t)x + ty) = (1-t)\lambda x + t\lambda y > (1-t)\lambda x + t\lambda x = \lambda x,$$

hence if $t > 0$ then $(1-t)\lambda x + ty \notin A$. But $(1-t)x + ty \to x$ as $t \to 0$ and $x \in A$, showing that $x \in \partial A$. □

The following lemma gives conditions under which a boundary point of a set is a proper support point of the set.[4]

Lemma 6. *If X is a topological vector space, C is a convex subset of X that has nonempty interior, and $x \in C \cap \partial C$, then C is properly supported at x.*

[4] Charalambos D. Aliprantis and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third ed., p. 259, Lemma 7.7.

Proof. The Hahn-Banach separation theorem[5] tells us that if A and B are disjoint nonempty convex subsets of X and A is open then there is some $\lambda \in X^*$ and some $t \in \mathbb{R}$ such that

$$\lambda a < t \leq \lambda b, \qquad a \in A, b \in B.$$

Check that the interior of a convex set in a topological vector space is convex, and hence that we can apply the Hahn-Banach separation theorem to $\{x\}$ and C°: as x belongs to the boundary of C it does not belong to the interior of C, so $\{x\}$ and C° are disjoint nonempty convex sets. Thus, there is some $\lambda \in X^*$ and some $t \in \mathbb{R}$ such that $\lambda y < t \leq \lambda x$ for all $y \in C^\circ$, from which it follows that $\lambda x \leq \lambda y$ for all $y \in C$, and $\lambda \neq 0$ because of the strict inequality for the interior. As $x \in C$, this means that x is a maximizer of λ over C, and as $\lambda \neq 0$ this means that x is a support point of C. But C° is nonempty and if $y \in C^\circ$ then $\lambda x < \lambda y$, hence x is a proper support point of C. □

5 Subdifferentials of convex functions

If $f : X \to (-\infty, \infty]$ is a proper function then there is some $y \in X$ for which $f(y) < \infty$, and for f to have a subgradient λ at x demands that $f(y) \geq f(x) + \lambda(y - x)$, and hence that $f(x) < \infty$. Therefore, if f is a proper function then the set of x at which f is subdifferentiable is a subset of $\operatorname{dom} f$.

We now prove conditions under which a function is subdifferentiable at a point, i.e., under which the subdifferential at that point is nonempty.[6]

Theorem 7. *If X is a topological vector space, $f : X \to (-\infty, \infty]$ is a proper convex function, x is an interior point of* $\operatorname{dom} f$*, and f is continuous at x, then f has a subgradient at x.*

Proof. Because f is convex, the set $\operatorname{dom} f$ is convex, and the interior of a convex set in a topological vector space is convex so $(\operatorname{dom} f)^\circ$ is convex. f is proper so it does not take the value $-\infty$, and on $\operatorname{dom} f$ it does not take the value ∞, hence f is finite on $\operatorname{dom} f$. But for a finite convex function on an open convex set in a topological vector space, being continuous at a point is equivalent to being continuous on the set, and is also equivalent to being bounded above on an open neighborhood of the point.[7] Therefore, f is continuous on $(\operatorname{dom} f)^\circ$ and is bounded above on some open neighborhood V of x contained in $(\operatorname{dom} f)^\circ$, say $f(y) \leq M$ for all $y \in V$. $V \times (M, \infty)$ is an open subset of $X \times \mathbb{R}$, and is contained in $\operatorname{epi} f$. This shows that $\operatorname{epi} f$ has nonempty interior. Since $f(x) < \infty$, if $\epsilon > 0$ then $(x, f(x) - \epsilon) \notin \operatorname{epi} f$, and since $f(x) > -\infty$ we have $(x, f(x)) \in \operatorname{epi} f$, and therefore $(x, f(x)) \in \operatorname{epi} f \cap \partial(\operatorname{epi} f)$. We can now apply Lemma 6: $\operatorname{epi} f$ is a convex subset of the topological vector space $X \times \mathbb{R}$ with nonempty interior and

[5] Gert K. Pedersen, *Analysis Now*, revised printing, p. 65, Theorem 2.4.7.

[6] Charalambos D. Aliprantis and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third ed., p. 265, Theorem 7.12.

[7] Charalambos D. Aliprantis and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third ed., p. 188, Theorem 5.43.

$(x, f(x)) \in \operatorname{epi} f \cap \partial(\operatorname{epi} f)$, so $\operatorname{epi} f$ is properly supported at $(x, f(x))$. That is, Lemma 6 shows that there is some $\Lambda \in (X \times \mathbb{R})^*$ such that

$$\Lambda(x, f(x)) = \sup_{(y,\alpha) \in \operatorname{epi} f} \Lambda(y, \alpha),$$

and there is some $(y, \alpha) \in \operatorname{epi} f$ for which $\Lambda(x, f(x)) > \Lambda(y, \alpha)$. Now, there is some $\lambda \in X^*$ and some $\beta \in \mathbb{R}^* = \mathbb{R}$ such that $\Lambda(y, \alpha) = \lambda y + \beta\alpha$ for all $(y, \alpha) \in X \times \mathbb{R}$. Thus, there is some nonzero $\lambda \in X^*$ and some $\beta \in \mathbb{R}$ such that

$$\lambda x + \beta f(x) = \sup_{(y,\alpha) \in \operatorname{epi} f} \lambda y + \beta\alpha.$$

If $\beta > 0$ then the right-hand side would be ∞ while the left-hand side is constant and $< \infty$, so $\beta \leq 0$. Suppose by contradiction that $\beta = 0$. Then $\lambda x \geq \lambda y$ for all $y \in \operatorname{dom} f$, and as $\lambda \neq 0$ this means that x is a support point of $\operatorname{dom} f$, and then by Lemma 5 we have that $x \in \partial(\operatorname{dom} f)$, contradicting $x \in (\operatorname{dom} f)^\circ$. Hence $\beta < 0$, so

$$\lambda x + \beta f(x) \geq \lambda y + \beta f(y), \qquad y \in \operatorname{dom} f,$$

i.e.,

$$f(y) \geq f(x) - \frac{\lambda}{\beta}(y - x), \qquad y \in \operatorname{dom} f.$$

Furthermore, if $y \notin \operatorname{dom} f$ then $f(y) = \infty$, for which the above inequality is true. Therefore, $f(y) \geq f(x) - \frac{\lambda}{\beta}(y - x)$ for all $y \in X$, showing that $-\frac{\lambda}{\beta}$ is a subgradient of f at x. □

6 Directional derivatives

Lemma 8. *If X is a vector space, $f : X \to (-\infty, \infty]$ is a proper convex function, $x \in \operatorname{dom} f$, $v \in X$, and $0 < h' < h$, then*

$$\frac{f(x + h'v) - f(x)}{h'} \leq \frac{f(x + hv) - f(x)}{h}.$$

Proof. We have

$$x + h'v = \frac{h'}{h}(x + hv) + \frac{h - h'}{h} x,$$

and because f is convex this gives

$$f(x + h'v) \leq \frac{h'}{h} f(x + hv) + \frac{h - h'}{h} f(x),$$

i.e.

$$f(x + h'v) - f(x) \leq \frac{h'}{h}(f(x + hv) - f(x)).$$

Dividing by h',

$$\frac{f(x+h'v)-f(x)}{h'} \leq \frac{f(x+hv)-f(x)}{h}.$$

□

If $f : X \to (-\infty, \infty]$ is a proper convex function, $x \in \operatorname{dom} f$, and $v \in X$, then the above lemma shows that

$$h \mapsto \frac{f(x+hv)-f(x)}{h}$$

is an increasing function $(0, \infty) \to (-\infty, \infty]$, and therefore that

$$\lim_{h \to 0^+} \frac{f(x+hv)-f(x)}{h}$$

exists; it belongs to $[-\infty, \infty]$, and if there is at least one $h > 0$ for which $f(x+hv) < \infty$ then the limit will be $< \infty$. We define the *one-sided directional derivative of* f *at* x to be the function $d^+ f(x) : X \to [-\infty, \infty]$ defined by[8]

$$d^+ f(x)v = \lim_{h \to 0^+} \frac{f(x+hv)-f(x)}{h}, \qquad v \in X.$$

Lemma 9. *If X is a topological vector space, $f : X \to (-\infty, \infty]$ is a proper convex function, $x \in (\operatorname{dom} f)^\circ$, f is continuous at x, and $v \in X$, then $-\infty < d^+ f(x)v < \infty$.*

Proof. Because $x \in (\operatorname{dom} f)^\circ$, there is some $h > 0$ for which $x + hv \in \operatorname{dom} f$ and hence for which $f(x+hv) < \infty$. This implies that $d^+ f(x)v < \infty$.

Let $h > 0$. By Theorem 7, the subdifferential $\partial f(x)$ is nonempty, i.e. there is some $\lambda \in X^*$ for which $f(y) \geq f(x) + \lambda(y-x)$ for all $y \in X$. Thus, for all $v \in X$ we have, with $y = x + hv$,

$$f(x+hv) \geq f(x) + \lambda(hv),$$

i.e.,

$$\lambda v \leq \frac{f(x+hv)-f(x)}{h}.$$

Since this difference quotient is bounded below by λv, its limit as $h \to 0^+$ is $> -\infty$, and therefore $d^+ f(x)v > -\infty$. □

[8] We are following the notation of Charalambos D. Aliprantis and Kim C. Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third ed., p. 266.