ON THE SUBDIFFERENTIABILITY OF CONVEX FUNCTIONS

A. BRØNDSTED AND R. T. ROCKAFELLAR[1]

1. Introduction. Let E be a locally convex Hausdorff topological vector space over the real numbers R with dual E^*. Let f be a proper convex function on E, i.e., an everywhere-defined function with values in $]-\infty, \infty]$, not identically $+\infty$, such that

(1.1) $f(\lambda x + (1 - \lambda)y) \leqq \lambda f(x) + (1 - \lambda)f(y)$ if $x \in E$, $y \in E$,

$$0 < \lambda < 1.$$

A vector $x^* \in E^*$ is called a *subgradient* of f at $x \in E$ if

(1.2) $f(y) \geqq f(x) + (y - x, x^*)$ for all $y \in E$.

(Thus the subgradients of f correspond to the nonvertical supporting hyperplanes to the convex set consisting of all the points of $E \oplus R$ lying above the graph of f.) The set of subgradients of f at x is denoted by $\partial f(x)$. If $\partial f(x)$ is not empty, f is said to be *subdifferentiable* at x. If f actually had a gradient $x^* = \nabla f(x)$ at x in the sense of Gateaux (or Frechet), one would in particular have $\partial f(x) = \{\nabla f(x)\}$ (see Moreau [5, p. 20]).

It is immediate from (1.2) that $\partial f(x)$ is a weak* closed convex set in E^* for each $x \in E$, and that the effective domain

$$\operatorname{dom} \partial f = \{x \mid \partial f(x) \neq \varnothing\}$$

of the subgradient mapping $\partial f: x \to \partial f(x)$, i.e., the set of points where f is subdifferentiable, is contained in the effective domain of f, which is the convex set

$$\operatorname{dom} f = \{x \mid f(x) < \infty\}.$$

One would like to know when dom ∂f is dense in dom f. This is certainly true whenever

(A) $f(y) = \liminf\limits_{x \to y} f(x)$ for all y, $\bar{f}(x) = \begin{cases} f(x) & \text{if } x \in \operatorname{dom} \partial f, \\ +\infty & \text{otherwise.} \end{cases}$

Condition (A) says dom ∂f actually has a dense intersection with every (convex) set of the form $\{y \mid f(y) < (y, y^*) - \mu\}$, $y^* \in E^*$, $\mu \in R$. One may also ask whether f is the supremum of the supporting affine

functions determined by its subgradients, i.e.

(B) $f(y) = \sup\{f(\bar{x}) + (y - \bar{x}, \bar{x}^*) \mid \bar{x} \in \operatorname{dom} \partial f, \bar{x}^* \in \partial f(\bar{x})\}$ for all $y \in E$.

Conditions (A) and (B) would each require that f be lower semi-continuous (l.s.c.).

Moreau [5] has shown that if a bicompact convex function (see below) exists on E, then, conversely, (A') and (B) are true for *every* l.s.c. proper convex function f on E, where (A') is the weaker version of (A) obtained when $\bar{f}$ is replaced by its convex envelope, which is the largest convex function majorized by $\bar{f}$. (Condition (A') says solely that f is completely determined by its values at the points where it is subdifferentiable, i.e. that f is the l.s.c. convex envelope of $\bar{f}$.) He has also shown that a bicompact convex function does exist if E is a reflexive Banach space.

We shall prove here that, in fact, a bicompact convex function exists *only* if E is a reflexive Banach space in the Mackey topology $\tau(E, E^*)$. But the l.s.c. proper convex functions on E are the same for all topologies compatible with the duality between E and E^*. In other words, when proving theorems about conditions like (A') and (B), one can assume without loss of generality that the given topology on E coincides with its Mackey topology. Thus Moreau's general theorem about (A') and (B) turns out to be equivalent to his special theorem for reflexive Banach spaces.

The latter result can be deduced more directly using arguments like those in the recent paper of Bishop and Phelps [1] on the existence of support points of convex sets. Indeed, we shall prove in this way that (A) and (B) hold for every l.s.c. proper convex function f if E is *any* Banach space. One cannot hope to extend this theorem much further, however, in view of certain convex sets ingeniously constructed by Klee [3], [4]. These will enable us to display a l.s.c. proper convex function f on a reflexive Frechet space (or on a pre-Hilbert space if so desired) such that f is *nowhere* subdifferentiable. Of course, (A) and (B) fail in the worst way for such an f.

It would be interesting to know whether every space for which (A) and (B) always hold is necessarily a Banach space "in disguise" (i.e. in its Mackey topology) as the counter-examples seem to suggest. We have not been able to answer this question.

2. **The existence of bicompact convex functions.** Throughout this section, let f be a l.s.c. proper convex function on E. The *conjugate* of f is the function f^* on E^* defined by

(2.1) $f^*(x^*) = \sup\{(x, x^*) - f(x) \mid x \in E\}$ for each $x^* \in E^*$.

It is known that f^* is again a proper convex function on E^*, l.s.c. in the weak* topology (and hence *a fortiori* in the strong topology on E^*). Moreover, the conjugate f^{**} of f^* coincides with f on E (considered as a subspace of E^{**}), i.e.

(2.2) $f(x) = \sup\{(x, x^*) - f^*(x^*) \mid x^* \in E^*\}$ for each $x \in E$.

(This notion of conjugacy, originally due to Fenchel, is developed in infinite-dimensional spaces in [2], [8], [9].) Notice that the statements

(2.3) $x^* \in \partial f(x)$, $x \in \partial f^*(x^*)$, $(x, x^*) = f(x) + f^*(x^*)$,

are equivalent by (2.1), (2.2) and the definition of a subgradient.

Moreau [5] defines f to be *bicompact* if $\{x \mid f(x) \leq (x, y^*) - \mu\}$ is a weakly compact (convex) subset of E for all $y^* \in E^*$ and $\mu \in R$, and $\{x^* \mid f^*(x^*) \leq (y, x^*) - \mu\}$ is a weak* compact (convex) subset of E^* for all $y \in E$ and $\mu \in R$.

THEOREM 1. *The function f is bicompact if and only if the following three conditions are satisfied:*

(a) *E is a reflexive Banach space in the Mackey topology $\tau(E, E^*)$ (which is the strongest locally convex topology on E yielding the same dual E^*);*

(b) *f is finite throughout E;*

(c) *$\{x \mid f(x) \leq (x, y^*) - \mu\}$ is bounded in E for every $y^* \in E^*$ and $\mu \in R$.*

PROOF. Suppose first that f is bicompact. Then (c) is trivially true, and, for each fixed $y \in E$, $\{x^* \mid f^*(x^*) \leq (y, x^*) - \mu\}$ is weak* compact for all $\mu \in R$. According to a remarkable new theorem of Moreau [7], the latter property is equivalent to f being finite and $\tau(E, E^*)$ continuous at y. Thus (b) holds and f is $\tau(E, E^*)$ continuous throughout E. In particular, $\{x \mid f(x) \leq f(0) + 1\}$ must be a weakly compact neighborhood of 0. The existence of such a neighborhood implies that the $\tau(E, E^*)$ topology is normable, and that the unit sphere for the norm is weakly compact. But E must then be a Banach space under $\tau(E, E^*)$, since the norm completion of the unit sphere is trivially contained in its weak completion. Now (a) follows from the well-known fact that a Banach space is reflexive if and only if its unit sphere is weakly compact.

Conversely, suppose (a), (b) and (c) are satisfied. The convex sets $\{x \mid f(x) \leq (x, y^*) - \mu\}$ are bounded by (c) and closed because f is l.s.c.; they must therefore all be weakly compact by reflexivity. On

the other hand, f is $\tau(E, E^*)$ continuous throughout E by (b), because a l.s.c. convex function on a Banach space is automatically continuous on open sets where it is finite [2, p. 11]. The theorem of Moreau invoked in the first half of the proof now implies the sets $\{x^* | f^*(x^*) \leqq (y, x^*) - \mu\}$ are all weak* compact. Thus f is bicompact.

REMARK. Moreau proved in [5] that the function $f(x) = \|x\|^p$, $p > 1$, is bicompact on any reflexive Banach space. This is also a direct consequence of Theorem 1.

3. **Existence of subgradients.** Let f be l.s.c. proper convex on E. For each $\epsilon > 0$, we may define a set $\partial_\epsilon f(x)$ of "approximate subgradients" of f at x by

$$(3.1) \quad \begin{aligned} \partial_\epsilon f(x) &= \{x^* \mid f(z) \geqq [f(x) - \epsilon] + (z - x, x^*) \text{ for all } z \in E\} \\ &= \{x^* \mid f(x) + f^*(x^*) - (x, x^*) \leqq \epsilon\}. \end{aligned}$$

Since (3.1) represents $\partial_\epsilon f(x)$ as the set of solutions x^* to an infinite system of linear inequalities, $\partial_\epsilon f(x)$ is a weak* closed convex set in E^* for each $\epsilon > 0$. Evidently $\partial_\epsilon f(x)$ decreases as ϵ decreases to 0, and the intersection of the nest of $\partial_\epsilon f(x)$ for $\epsilon > 0$ is just $\partial f(x)$. Also, $\partial_\epsilon f(x)$ is nonempty for $\epsilon > 0$ and $x \in \text{dom } f$ by (2.2). The following lemma, whose proof was suggested by that of the fundamental lemma of Bishop and Phelps in [1], estimates how well $\partial_\epsilon f$ "approximates" ∂f.

LEMMA. *Assume that E is a Banach space and that $x^* \in \partial_\epsilon f(x)$. Then, for any $\lambda > 0$, there exist vectors $\bar{x}$ and $\bar{x}^*$ such that $\|\bar{x} - x\| \leqq \lambda$, $\|\bar{x}^* - x^*\| \leqq \epsilon/\lambda$ and $\bar{x}^* \in \partial f(\bar{x})$.*

PROOF. Define the relation $y \prec z$, for y and z in dom f, to mean that

$$(3.2) \quad (\epsilon/\lambda)\|y - z\| \leqq [f(y) - (y, x^*)] - [f(z) - (z, x^*)].$$

It is obvious that $\prec$ is reflexive and anti-symmetric. Transitivity follows from the subadditivity of the norm. Thus $\prec$ is a partial ordering of the set dom f. By Zorn's Lemma, there exists a maximal totally ordered subset M of $\{z \in \text{dom } f \mid x \prec z\}$. For notational convenience, we shall write $M = \{z_\alpha \mid \alpha \in I\}$, where I is a totally-ordered index set. Since $x^* \in \partial_\epsilon f(x)$, (3.1) and (3.2) require

$$f(z_\alpha) - (z_\alpha, x^*) \geqq f(z_\beta) - (z_\beta, x^*) \geqq f(x) - (x, x^*) - \epsilon > -\infty \text{ when } \alpha < \beta.$$

Therefore

$$(3.3) \quad f(z_\alpha) - (z_\alpha, x^*) \downarrow \rho > -\infty \text{ as } \alpha \uparrow.$$

This implies $\{z_\alpha\}$ is a Cauchy net. Indeed, for any $\delta > 0$ we could

choose α large enough that $f(z_\beta) - (z_\beta, x^*) < \rho + \delta(\epsilon/\lambda)$ for all $\beta > \alpha$. Then $\|z_\alpha - z_\beta\| < \delta$ for all $\beta > \alpha$ by the definition of $\prec$. Inasmuch as E is a Banach space, we may conclude $\{z_\alpha\}$ has a limit $\bar{x} \in E$. The lower semi-continuity of f in (3.2) and (3.3) implies that $\bar{x} \in \text{dom } f$ and $z_\alpha \prec \bar{x}$ for all α. In particular $x \prec \bar{x}$, so that

$$(\epsilon/\lambda)\|\bar{x} - x\| \leqq - [f(\bar{x}) - f(x) - (\bar{x} - x, x^*)] \leqq \epsilon$$

by the definitions of $\prec$ and $\partial_\epsilon f$. Hence $\|\bar{x} - x\| \leqq \lambda$. Furthermore, $\bar{x} \prec z$ can happen only for $z = \bar{x}$, because the totally-ordered set M was maximal. Therefore

$$(\epsilon/\lambda)\|\bar{x} - z\| > [f(\bar{x}) - (\bar{x}, x^*)] - [f(z) - (z, x^*)] \quad \text{for all } z \neq \bar{x}.$$

This means that, in $E \oplus R$, the sets

$$C_1 = \{\langle y, \mu \rangle \mid \mu \geqq h(y) = f(\bar{x} + y) - f(\bar{x}) - (y, x^*)\},$$
$$C_2 = \{\langle y, \mu \rangle \mid \mu < - (\epsilon/\lambda)\|y\|\},$$

have no point in common. But C_1 is a closed convex set, because it is the supergraph of a l.s.c. proper convex function h, and C_2 is an open convex cone. Hence C_1 and C_2 can be separated by a hyperplane in $E \oplus R$. Due to the nature of C_2, we can take this hyperplane to be the graph of a continuous linear function on E, thus there exists some $z^* \in E^*$ such that

$$(3.4) \quad -(\epsilon/\lambda)\|y\| \leqq (y, z^*) \leqq f(\bar{x} + y) - f(\bar{x}) - (y, x^*) \quad \text{for all } y.$$

Set $\bar{x}^* = x^* + z^*$. The left half of (3.4) says $\|\bar{x}^* - x^*\| \leqq \epsilon/\lambda$, and the right half says $\bar{x}^* \in \partial f(\bar{x})$.

4. Main theorem. The Lemma just proved is crucial in the following result.

THEOREM 2. *If E is a Banach space, then conditions* (A) *and* (B) *are satisfied by every l.s.c. proper convex function f on E.*

Moreover, the conjugates f^* of such functions actually satisfy the stronger conditions (A*) *and* (B*) *obtained from* (A) *and* (B) *by restricting attention to the existence of subgradients of f^* belonging to E (not just to E^{**}).*

PROOF. Since f is l.s.c., (A) can be proved by showing that the "lim inf" does not exceed $f(x)$ when $x \in \text{dom } f$. Given any $\delta > 0$, choose any $x^* \in \partial_\epsilon f(x)$, where $\epsilon = \delta/2$. Choose $\lambda > 0$ so small that $\lambda < \delta$ and $\lambda\|x^*\| < \delta/2$. Now let $\bar{x}$ and $\bar{x}^*$ be the vectors whose existence is guaranteed by the Lemma. The three conditions on $\bar{x}$ and $\bar{x}^*$ then yield

$$f(\bar{x}) - f(x) \leqq - (x - \bar{x}, \bar{x}^*) \leqq \|\bar{x} - x\| \|\bar{x}^*\|$$
$$\leqq \lambda(\|x^*\| + \epsilon/\lambda) < \delta/2 + \delta/2 = \delta.$$

Thus $\bar{x} \in \text{dom } \partial f$, $\|\bar{x} - x\| < \delta$ and $f(\bar{x}) < f(x) + \delta$. Since $\delta > 0$ was arbitrary, this yields (A).

Virtually the same argument proves (A*) holds for f^*. This is apparent if, in the wording of the Lemma, we set $\epsilon/\lambda = \lambda^*$, $\lambda = \epsilon/\lambda^*$, and replace the conditions $x^* \in \partial_\epsilon f(x)$, $\bar{x}^* \in \partial f(\bar{x})$, by the equivalent conditions $x \in \partial_\epsilon f^*(x^*)$, $\bar{x} \in \partial f^*(\bar{x}^*)$. (The equivalence is immediate from (2.3) and the symmetry in (3.1)).

The fact that (B) holds for f follows directly from (2.2) and condition (A*) for f^*, because of (2.3). Similarly, (B*) for f^* is a consequence of (A) for f.

REMARK. The Lemma can also be employed, much in the above manner, to derive results of Bishop and Phelps [1]. In this case, one would make use of the one-to-one correspondence between nonempty closed convex sets C in E and their indicator functions δ_C (where δ_C is 0 on C and $+\infty$ outside of C), which are l.s.c. proper convex functions. The conjugate of δ_C is the support function σ_C of C. Hence $x^* \in \partial_\epsilon \delta_C(x)$ if and only if $x \in C$ and $(x, x^*) \geqq \alpha - \epsilon$, where

$$\infty > \alpha = \sigma_C(x^*) = \sup\{(z, x^*) \mid z \in C\}.$$

In particular, the nonzero subgradients of δ_C at x are precisely the vectors x^* defining nontrivial supporting hyperplanes to C at x.

5. **A counterexample.** Klee [3] has constructed a nonempty closed convex set C in a certain reflexive Frechet space E (actually a Montel space), such that C has no support points whatsoever. This C happens to contain various half-lines emanating from the origin, but no whole lines. Under these circumstances, we may construct a function f as follows. Fix any $x_0 \neq 0$ such that $\{\lambda x_0 \mid \lambda \geqq 0\} \subseteq C$. For each x let

$$(5.1) \qquad f(x) = \min\{\lambda \in R \mid x + \lambda x_0 \in C\},$$

where the minimum is understood to be $+\infty$ when no such λ exists. We shall prove that:

The function f is l.s.c. proper convex on E, but it is nowhere subdifferentiable.

Since C contains no whole lines, f does not take on the value $-\infty$. The convexity condition (1.1) is easy to verify. To show lower semicontinuity, we need to observe first that

$$(5.2) \qquad f(x + \mu x_0) = f(x) - \mu \quad \text{for all } x \in E \text{ and } \mu \in R.$$

This implies each level set $\{x \mid f(x) \leq \mu\}$ of f is a translate of C_0 = $\{x \mid f(x) \leq 0\}$. Obviously $C_0 \supseteq C$. The reverse inclusion is easy to deduce from (5.1) using the fact that C is closed and $\{\lambda x_0 \mid \lambda \geq 0\} \subseteq C$. The (sub-) level sets of f are therefore all closed, so f is l.s.c. Finally, suppose for a moment that f had a subgradient x^* at some point x. Then $x \in \mathrm{dom}\, f$ and

$$(5.3) \qquad \infty > (x, x^*) - f(x) \geq (y, x^*) \quad \text{for all } y \in C$$

by (1.2), because f is nonpositive on C by definition. On the other hand, (5.2) and (1.2) imply $(x_0, x^*) = -1$. Hence

$$(5.4) \qquad x^* \neq 0 \quad \text{and} \quad (x, x^*) - f(x) = (x + f(x)x_0, x^*).$$

But (5.3) and (5.4) say that C has a nontrivial supporting hyperplane at the point $x + f(x)x_0$ (a point of C by the definition of f), and we know this to be impossible. Therefore f cannot have a subgradient.

A similar example of a nowhere subdifferentiable f on an (incomplete) inner product space can be constructed, using definition (5.1), from another of Klee's supportless convex sets [4]. In this case $x_0 \neq 0$ is arbitrary, and the lower semi-continuity of f follows easily from the pre-compactness of the set in question.

References

1. Errett Bishop and R. R. Phelps, *The support functionals of a convex set,* Convexity, Proc. Sympos. Pure Math. Vol. 7, Amer. Math. Soc., Providence, R. I., 1962; pp. 27–35.

2. Arne Brøndsted, *Conjugate convex functions in topological vector spaces,* Mat. Fys. Medd. Dan. Vid. Selsk. **34** (1964), no. 2.

3. Victor Klee, *On a question of Bishop and Phelps,* Amer. J. Math. **85** (1963), 95–98.

4. ———, *Extremal structure of convex sets.* II, Math. Z. **69** (1958), 90–104.

5. Jean-Jacques Moreau, *Etude locale d'une fonctionelle convexe,* Faculté des Sciences de Montpellier, Séminaires de Mathématiques, 1963.

6. ———, *Fonctionelles sous-différentiables,* C. R. Acad. Sci. Paris **257** (1963), 4117–4119.

7. ———, *Sur la fonction polaire d'une fonction semi-continue supérieurement,* C. R. Acad. Sci. Paris **258** (1964), 1128–1130.

8. ———, *Fonctions convexes en dualité,* Faculté des Sciences de Montpellier, Séminaires de Mathématiques, 1962.

9. R. T. Rockafellar, *Level sets and continuity of conjugate convex functions,* submitted to Trans. Amer. Math. Soc.

THE UNIVERSITY OF COPENHAGEN, DENMARK AND
    THE UNIVERSITY OF TEXAS