PACIFIC JOURNAL OF MATHEMATICS
Vol. 17, No. 3, 1966

CHARACTERIZATION OF THE SUBDIFFERENTIALS OF CONVEX FUNCTIONS

R. T. ROCKAFELLAR

Each lower semi-continuous proper convex function f on a Banach space E defines a certain multivalued mapping ∂f from E to E^* called the subdifferential of f. It is shown here that the mappings arising this way are precisely the ones whose graphs are maximal "cyclically monotone" relations on $E \times E^*$, and that each of these is also a maximal monotone relation. Furthermore, it is proved that ∂f determines f uniquely up to an additive constant. These facts generally fail to hold when E is not a Banach space. The proofs depend on establishing a new result which relates the directional derivatives of f to the existence of approximate subgradients.

Let E be a 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

$$(1.1) \qquad f(\lambda x + (1-\lambda)y) \leqq \lambda f(x) + (1+\lambda)f(y)$$

for all x and y in E when $0 < \lambda < 1$. A *subgradient* of f at $x \in E$ is an $x^* \in E^*$ such that

$$f(y) \geqq f(x) + \langle y - x, x^* \rangle \quad \text{for all} \quad y \in E .$$

(This says that $f(x)$ is finite and that the graph of the affine function $h(y) = f(x) + \langle y - x, x^* \rangle$ is a "nonvertical" supporting hyperplane at $(x, f(x))$ to the epigraph of f, which is the convex subset of $E \oplus R$ consisting of all the points lying above the graph of f.) For each $x \in E$, we denote by $\partial f(x)$ the set of all subgradients of f at x, which is a weak* closed convex set in E^*. If $\partial f(x) \neq \emptyset$, f is said to be *subdifferentiable* at x. The *subdifferential* of f is the multivalued mapping (relation) ∂f which assigns the set $\partial f(x)$ to each x.

The notion of subdifferentiability has been developed recently in [3], [6], [8], [11]. Much of the work has concerned the existence of subgradients. It is known, for example, that f is subdifferentiable wherever it is finite and continuous (see [6] or [8]). Results in [3] show among other things that, if E is a Banach space and f is lower semi-continuous (l.s.c.), then the set of points where f is subdiffer-

Received January 7, 1965. This work was supported in part by the Air Force Office of Scientific Research, through a grant at The University of Texas.

entiable is dense in the effective domain of f (which is the convex set of all x such that $f(x) < +\infty$). The present paper, however, will be concerned instead with general properties of the subdifferentials of l.s.c. proper convex functions, as relations on $E \times E^*$.

Some facts are already known about the global nature of subdifferentials. The most remarkable from a geometric point of view is the following.

THEOREM (Moreau [11]). *Let f be a l.s.c. proper convex function on a real Hilbert space H. Then the addition mapping $(x, x^*) \to x + x^*$ from $H \times H$ to H maps the graph of ∂f homeomorphically onto H.*

Among the many interesting consequences of this theorem is the arcwise connectedness of the set of points where f is subdifferentiable. Whether this arcwise connectedness is present when E is not a Hilbert space, is an open question. In the case where f is everywhere finite and Gateaux differentiable, ∂f reduces to the ordinary single-valued gradient mapping ∇f. Then Moreau's theorem says that the equation

$$x + \nabla f(x) = u \tag{1.2}$$

has a unique solution x for each $u \in H$, and that this solution depends continuously on u.

Similar results have been arrived at independently through the study of monotone relations. A relation ρ on $E \times E^*$ is said to be *monotone* if

$$\langle y - x, y^* - x^* \rangle \geqq 0$$

holds whenever $x^* \in \rho(x)$ and $y^* \in \rho(y)$. A *maximal* monotone relation is one whose graph is not properly contained in the graph of another monotone relation. The following theorem invites comparison with the one of Moreau.

THEOREM (Minty [7]). *Let ρ be a maximal monotone relation on $H \times H$, where H is a real Hilbert space. Then the addition mapping $(x, x^*) \to x + x^*$ from $H \times H$ to H maps the graph of ρ homeomorphically onto H.*

We shall prove in § 5 that, if E is a Banach space, the subdifferential ∂f of each l.s.c. proper convex function f on E is a maximal monotone relation ρ. In particular, Moreau's theorem can be viewed as a special case of Minty's theorem. The maximal monotonicity of ∂f has previously been verified by Minty [6] in the case where f is also everywhere finite and continuous. Special connections between convexity and monotonicity have also been noted by Kachurovskii [5]. We are indebted to the referee for bringing this latter paper to our attention.

Not every monotone relation arises from a convex function. For instance, every positive semi-definite linear mapping ρ on a real Hilbert space is a (single-valued) monotone relation, but such a mapping is the subdifferential ∂f of a proper convex function if and only if it is also self-adjoint. In general, one is led to ask for properties which characterize the relations which are subdifferentials.

In §2 we shall show that a given relation on $E \times E^*$ is embedded in the subdifferential of some proper convex function on E if and only if it is "cyclically monotone" in a certain sense. When E is a Banach space, the subdifferentials of the l.s.c. proper convex functions on E turn out to be precisely the maximal cyclically monotone relations on $E \times E^*$. This will be proved in § 4.

Our Banach space theorems depend heavily on the fundamental existence lemma for subgradients in [3]. They also make use of a new result in § 3 which describes the directional derivatives of f in terms of the "approximate subgradients" introduced in [3].

2. Embedding problem. Let ρ be a relation on $E \times E^*$. When is ρ embedded in a subdifferential, i.e., when does there exist a proper convex function f such that $\rho(x) \subseteqq \partial f(x)$ for all x? This can also be viewed as a kind of "integration" problem: given a set of pairs $\{(x_i, x_i^*), i \in I\}$ in $E \times E^*$ (namely, the graph of ρ), one seeks a proper convex function f satisfying the "differential" conditions

$$f(y) \geqq f(x_i) + \langle y - x_i, x_i^* \rangle, i \in I ,$$

for all $y \in E$.

There is a simple necessary condition which ρ must satisfy if the embedding problem is to have a solution. Indeed, if f is a proper convex function on E and if

$$x_i^* \in \rho(x_i) \subseteqq \partial f(x_i) , \qquad i = 0, 1, \cdots, n ,$$

then

$$\infty > f(x_j) \geqq f(x_i) + \langle x_j - x_i, x_i^* \rangle$$

for all i and j and hence

$$(2.1) \qquad 0 \geqq \langle x_0 - x_n, x_n^* \rangle + \cdots + \langle x_2 - x_1, x_1^* \rangle + \langle x_1 - x_0, x_0^* \rangle .$$

A relation ρ which satisfies (2.1) for every set of $n + 1$ pairs in its graph will be called *monotone of degree n*. A relation which is monotone of degrees n for all $n > 0$ will simply be called *cyclically monotone*. The condition can also be stated as follows: ρ is cyclically monotone if and only if

$$\sum \langle x_i, x_i^* \rangle \geqq \sum \langle x_{\sigma(i)}, x_i^* \rangle$$

for every finite set of points in the graph of ρ and every permutation σ.

Monotonicity of degree n implies monotonicity of degree m for all $m \leqq n$. Note that ρ is monotone of degree 1 if and only if it is a *monotone relation*. Thus every cyclically monotone relation is monotone. In the one-dimensional case, the converse is also true: every monotone relation is cyclically monotone. It is easy to see, however, that this is false when $E = R^n$ with $n > 1$. The following conjecture does seem plausible, though: if $E = R^n$, then each relation on $E \times E^*$ which is monotone of degree n is actually monotone of all degrees, i.e., is cyclically monotone. We have not seriously investigated this question.

The cyclic monotonicity condition can be viewed heuristically as a discrete substitute for two classical conditions: that a smooth convex function has a positive semi-definite second differential, and that all circuit integrals of an integrable vector field must vanish.

THEOREM 1. *Let E be a topological vector space, and let ρ be a relation on $E \times E^*$. In order that there exist a proper convex function f on E such that $\partial f \supseteq \rho$, it is necessary and sufficient that ρ be cyclically monotone.*

Proof. The necessity of the condition was demonstrated in the preceding remarks. To prove the sufficiency, we suppose, therefore, that ρ is a cyclically monotone relation. There is no loss of generality if we also suppose ρ is nonempty and fix some $x_0 \in E$ and $x_0^* \in E^*$ with $x_0^* \in \rho(x_0)$. For each $x \in E$, let

$$f(x) = \sup \{\langle x - x_n, x_n^* \rangle + \cdots + \langle x_1 - x_0, x_0^* \rangle\} ,$$

where $x_i^* \in \rho(x_i)$ for $i = 1, \cdots, n$ and the supremum is taken over all possible finite sets of such pairs (x_i, x_i^*). We shall show that f is a proper convex function with $\partial f \supseteq \rho$.

Note first that f is a supremum of a nonempty collection of affine functions, one for each choice of $(x_1, x_1^*), \cdots, (x_n, x_n^*)$. Hence $f(x) > -\infty$ for all x and the convexity condition (1.1) is satisfied. Furthermore,

$$f(x_0) = 0$$

because ρ is cyclically monotone. Hence f is a proper convex function.

Next, choose any $\bar{x}$ and $\bar{x}^*$ with $\bar{x}^* \in \rho(\bar{x})$. We shall show that $\bar{x}^* \in \partial f(\bar{x})$. It is enough to show that, for each $\alpha < f(\bar{x})$, we have

$$f(x) \geqq \alpha + \langle x - \bar{x}, \bar{x}^* \rangle \quad \text{for all} \quad x . \tag{2.2}$$

Given $\alpha < f(\bar{x})$, we can choose (by the definition of f) pairs (x_i, x_i^*) such that $x_i^* \in \rho(x_i)$ for $i = 1, \cdots, k$, and

$$(2.3) \qquad \alpha < \langle \bar{x} - x_k, x_k^* \rangle + \cdots + \langle x_1 - x_0, x_0^* \rangle .$$

Let $x_{k+1} = \bar{x}$ and $x_{k+1}^* = \bar{x}^*$. Then

$$f(x) \geqq \langle x - x_{k+1}, x_{k+1}^* \rangle + \langle x_{k+1} - x_k, x_k^* \rangle + \cdots + \langle x_1 - x_0, x_0^* \rangle$$

for all x by the definition of f. This implies (2.2) via (2.3). Therefore $\partial f \supseteqq \rho$ and the theorem has been proved.

COROLLARY 1. *If ρ is a maximal cyclically monotone relation, then $\rho = \partial f$ for some proper convex function f.*

By a *maximal* cyclically monotone relation, we of course mean one whose graph is not properly contained in the graph of any other cyclically monotone relation. Thus Corollary 1 follows immediately from Theorem 1.

It is easy to prove, using Zorn's lemma, that every cyclically monotone relation is embedded in a maximal one. Since every subdifferential is cyclically monotone, we may, therefore, also state:

COROLLARY 2. *If f_1 is any proper convex function on E, then there exists a proper convex function f_2 on E such that $\partial f_1 \subseteqq \partial f_2$ and ∂f_2 is a maximal cyclically monotone relation.*

Each maximal cyclically monotone relation is actually the subdifferential of some *lower semi-continuous* proper convex function. Indeed, it is easy to see that, if f is any proper convex function with a non-empty subdifferential, then the function g defined by

$$(2.4) \qquad g(x) = \liminf_{y \to x} f(y) \quad \text{for all} \quad x$$

is a l.s.c. proper convex function with

$$(2.5) \qquad \partial g(x) \supseteqq \partial f(x) \quad \text{for all} \quad x .$$

If ∂f is maximal, we must, therefore, have $\partial f = \partial g$.

When E is a Banach space, the converse is also true: the subdifferential of *every* l.s.c. proper convex function is a maximal cyclically monotone relation. This fact will be proved in § 4. The situation is not necessarily so simple when E is not a Banach space. There exists, for instance, a (reflexive) Frechet Montel space on which there is a l.s.c. proper convex function whose subdifferential is *empty* and hence certainly not maximal. (See [3].) The same counterexample shows that, in general, two l.s.c. proper convex functions can have the same subdifferential and yet differ by more than just an additive constant. (Take a function with empty subdifferential, and compare it with its translates.) It will be demonstrated in § 4 that this difficulty never arises in Banach spaces.

One might define a proper convex function to be *maximal* if it is l.s.c. and its subdifferential is a maximal cyclically monotone relation. Although not every l.s.c. proper convex function on E need be maximal when E is not a Banach space, many maximal ones always do exist by Corollary 2 and the above remarks. It would be interesting to know whether the class of maximal functions in the non-Banach space case possesses any significant properties as a whole. For instance, do such functions satisfy conditions (A) and (B) in [3]?

3. Approximate subgradients and directional derivatives. In this section we shall establish a new result about the directional derivatives of a convex function. This result will be crucial in the proving of our Banach space theorems in § 4 and § 5.

We shall assume in the following that E is locally convex and Hausdorff, and that f is a lower semi-continuous proper convex function on E.

The *conjugate* of f is the function f^* on E^* defined by

$$f^*(x^*) = \sup \{\langle x, x^*\rangle - f(x) \mid x \in E\} \tag{3.1}$$

for each $x^* \in E^*$. It is known that f^* is a proper convex function on E^*, l.s.c. in the weak* as well as the strong topology. Furthermore, the conjugate f^{**} of f^* on E^{**} coincides with f on E (considered as a subspace of E^{**}), i.e.

$$f(x) = \sup \{\langle x, x^*\rangle - f^*(x^*) \mid x^* \in E^*\} \tag{3.2}$$

for all $x \in E$. This duality will be needed later. The reader interested in the theory of conjugate convex functions should consult [2], [4], [9], [11], [13], [14].

For each $\varepsilon > 0$, we define an "approximate subdifferential relation" $\partial_\varepsilon f$ as in [3] by letting $\partial_\varepsilon f(x)$ be the set of $x^* \in E^*$ such that

$$f(y) \geqq [f(x) - \varepsilon] + \langle y - x, x^*\rangle \quad \text{for all} \quad y \in E\,. \tag{3.3}$$

If $f(x)$ is finite, $\partial_\varepsilon f(x)$ is a *nonempty* weak* closed convex subset of E^* for each $\varepsilon > 0$, and $\partial_\varepsilon f(x)$ decreases to $\partial f(x)$ as $\varepsilon \downarrow 0$. The sense in which $\partial_\varepsilon f$ approximates ∂f is explained by the following result, proved by A. Brøndsted and the author. We restate in here for convenience, since it will be applied both in § 4 and § 5.

LEMMA ([3]). *Let E be a Banach space, and let f be a l.s.c. proper convex function on E. Let $x \in E$, $x^* \in E^*$, and $\varepsilon > 0$ be such that $x^* \in \partial_\varepsilon f(x)$. Select any λ with $0 < \lambda < \infty$. Then there exist $\bar{x} \in E$ and $\bar{x}^* \in E^*$ such that*

$$\|\bar{x} - x\| \leqq \lambda, \; \|\bar{x}^* - x^*\| \leqq \varepsilon/\lambda, \; \bar{x}^* \in \partial f(\bar{x}) .$$

Now let x be any point at which f is finite. It is a classical fact (e.g. see [1]) that *directional derivative*

$$f'(x; y) = \lim_{\lambda \downarrow 0} [f(x + \lambda y) - f(x)]/\lambda \tag{3.4}$$

exists for all $y \in E$ (although it may be infinite) because the difference quotient decreases as $\lambda \downarrow 0$. Furthermore, $f'(x; \cdot)$ is positively homogeneous, i.e.

$$f(x; \lambda y) = \lambda f(x; y) \quad \text{for all} \quad \lambda > 0 \quad \text{and} \quad y \in E ,$$

and it is a proper convex function on E provided it is greater than $-\infty$ for each y.

There is an elementary relationship between subgradients and directional derivatives. Namely, if $f(x)$ is finite one evidently has

$$x^* \in \partial f(x) \Leftrightarrow f'(x; y) \geqq \langle y, x^* \rangle \quad \text{for all} \quad y . \tag{3.5}$$

Given any nonempty weak* closed convex set C^* in E^*, let us denote the *support function* of C^* on E by $\sigma(C^*; \cdot)$. Thus

$$\sigma(C^*; y) = \sup \{\langle y, x^* \rangle \mid x^* \in C^*\}$$

for each $y \in E$, and $\sigma(C^*; \cdot)$ is a positively homogeneous l.s.c. proper convex function on E. Formula (3.5) says that, if $\partial f(x) \neq \emptyset$,

$$f'(x; y) \geqq \sigma(\partial f(x); y) \text{ for all } y .$$

But equality need not always hold, not even in the finite dimensional case, although it may be shown from (3.5) using the theory of conjugate convex functions that

$$\sigma(\partial f(x); y) = \liminf_{z \to y} f'(x; z)$$

for all y when $\partial f(x)$ is *nonempty*.

The following theorem says that, just as $\partial f(x)$ is the intersection of the $\partial_\varepsilon f(x)$ for $\varepsilon > 0$, so is $f'(x; \cdot)$ the infimum of the (l.s.c.) support functions of the $\partial_\varepsilon f(x)$ for $\varepsilon > 0$, even when $\partial f(x)$ is empty. This throws new light on the discontinuities of the directional derivative function. In connection with the lemma stated above, it will also provide us with a powerful means translating facts about the directional derivatives of f into facts about its subdifferential.

THEOREM 2. *Let E be a locally convex Hausdorff topological vector space, and let f be a l.s.c. proper convex function on E. Let x be a point at which f is finite. Then, for all $y \in E$,*

$$\sigma(\partial_\varepsilon f(x); y) \downarrow f'(x; y) \text{ as } \varepsilon \downarrow 0 . \tag{3.6}$$

Proof. Suppose we are given a nonempty weak* closed convex set C^* in E^* of the form

$$C^* = \{x^* \mid f^*(x^*) - \langle x, x^* \rangle \leqq \beta\} , \tag{3.7}$$

where f^* is the conjugate of f and

$$\infty > \beta > \inf \{f^*(x^*) - \langle x, x^* \rangle \mid x^* \in E^*\} > -\infty . \tag{3.8}$$

The author showed in [13] that the support function of C^* could then be calculated by the formula.

$$\sigma(C^* \mid y) = \inf_{\lambda > 0} [f(x + \lambda y) + \beta]/\lambda .$$

By the definition of $\partial_\varepsilon f$ and f^*, $x^* \in \partial_\varepsilon f(x)$ if and only if

$$f(x) - \varepsilon - \langle x, x^* \rangle \leqq \inf_y \{f(y) - \langle y, x^* \rangle\} = -f^*(x^*) .$$

Thus $\partial_\varepsilon f(x) = C^*$ in (3.7), where

$$\beta = \varepsilon - f(x)$$

satisfies (3.8) by (3.2). Thus

$$\sigma(\partial_\varepsilon f(x); y) = \inf_{\lambda > 0} [f(x + \lambda y) - f(x) + \varepsilon]/\lambda \tag{3.9}$$

for each $\varepsilon > 0$. But

$$f'(x; y) = \inf_{\lambda > 0} [f(x + \lambda y) - f(x)]/\lambda , \tag{3.10}$$

since the difference quotient in (3.4) decreases as $\lambda \downarrow 0$. Formula (3.6) is obvious from (3.9) and (3.10).

REMARK. It is easy to deduce from the symmetric formulas (3.1) and (3.2) that

$$x^* \in \partial f(x) \Longleftrightarrow x \in \partial f^*(x^*) , \tag{3.11}$$

$$x^* \in \partial_\varepsilon f(x) \Longleftrightarrow x \in \partial_\varepsilon f^*(x^*) . \tag{3.12}$$

In fact the conditions in (3.11) and (3.12) are equivalent respectively to

$$f(x) + f^*(x^*) = \langle x, x^* \rangle , \tag{3.11$'$}$$

$$f(x) + f^*(x^*) \leqq \langle x, x^* \rangle + \varepsilon \, . \tag{3.12'}$$

The dual version of the above lemma, in which unstarred elements are everywhere interchanged with their starred counterparts, is therefore valid. This is true despite the fact that E may not be reflexive. The dual of Theorem 2 must likewise be valid, by (3.12) and the fact that the support function formula quoted in the proof of Theorem 2 was established in a symmetric context in [13]. The duals of the Lemma and of Theorem 2 will both be needed below in the proof of Theorem 3.

4. Characterization problem. We shall now show that, in the Banach space case, the subdifferentials of the l.s.c. functions are completely characterized as the maximal cyclically monotone ones. Counter-examples described in § 2 show that this can fail outside of Banach spaces. Note the interesting resemblance between our result and the fundamental theorem of calculus.

THEOREM 3. *Let E be a Banach space and let ρ be a relation on $E \times E^*$. In order that there exist a l.s.c. proper convex function f on E such that*

$$\partial f = \rho \, ,$$

it is both necessary and sufficient that ρ be a maximal cyclically monotone relation. Moreover, the solution f is then unique up to an arbitrary additive constant.

Proof. The sufficiency of the condition was demonstrated in general in §2 (see Corollary 1 and the remark following Corollary 2). To prove its necessity, assume that f is l.s.c. proper convex. By Corollary 2 in §2 (and the remark following it), there exists a l.s.c. proper convex function g such that

$$\partial g \supseteqq \partial f \tag{4.1}$$

and ∂g is a *maximal* cyclically monotone relation. We shall show that (4.1) implies

$$g = f + \text{const.}$$

This will complete the proof of the theorem.

Fix any $z \in E$ such that $\partial f(z) \neq \emptyset$. This is possible, since the lemma in §3 trivially implies ∂f is not empty. By (4.1) we also have $\partial g(z) \neq \emptyset$, so both $f(z)$ and $g(z)$ are finite. It will be shown first that

$$f(x) - f(z) \leqq g(x) - g(z) \quad \text{for all } x \, . \tag{4.2}$$

Fix any $x \in E$ and any $\alpha < f(x) - f(z)$. Let

$$Q(\lambda) = f(z + \lambda(x - z)) \text{ for all } \lambda .$$

Then Q is a l.s.c. proper convex function on the real line, with $Q(0)$ finite and

$$Q(1) - Q(0) = f(x) - f(z) > \alpha .$$

The set of λ in $[0, 1]$, for which $Q(\lambda) < \infty$, is an interval containing 0, and the right derivative $Q'_+(\lambda)$ is well-defined and nondecreasing on this interval. Indeed, by the classical one-dimensional theory of convex functions, there must exist λ_i such that

$$\begin{gathered} 0 \leqq \lambda_0 < \lambda_1 < \cdots < \lambda_n < 1 \\ Q(\lambda_i) < \infty,\ Q'_+(\lambda_i) > -\infty , \\ (\lambda_1 - \lambda_0)Q'_+(\lambda_0) + \cdots + (\lambda_{n+1} - \lambda_n)Q'_+(\lambda_n) > \alpha , \end{gathered} \tag{4.3}$$

where $\lambda_{n+1} = 1$ and λ_0 can be chosen arbitrarily small. Let

$$x_i = z + \lambda_i(x - z) , \qquad i = 0, 1, \cdots, n . \tag{4.4}$$

In terms of the directional derivatives of f, (4.3) says

$$\begin{gathered} f(x_i) < \infty, f'(x_i; x_{i+1} - x_i) > -\infty , \\ f'(x_0; x_1 - x_0) + \cdots + f'(x_n; x_{n+1} - x_n) > \alpha , \end{gathered} \tag{4.5}$$

where $x_{n+1} = x$ and x_0 can be chosen arbitrarily close to z. We may now choose $\alpha_i \in R$ and $\delta > 0$ such that

$$\begin{gathered} f'(x_i; x_{i+1} - x_i) > \alpha_i \qquad \text{for } i = 0, 1, \cdots, n , \\ \alpha_0 + \alpha_1 + \cdots + \alpha_n > \alpha + (n + 1)\delta . \end{gathered} \tag{4.6}$$

By Theorem 2, for any $\varepsilon > 0$ we can find $x_i^* \in E^*$ for $i = 0, 1, \cdots, n$, such that

$$x_i^* \in \partial_\varepsilon f(x_i) \quad \text{and} \quad \langle x_{i+1} - x_i, x_i^* \rangle > \alpha_i .$$

Applying the Lemma in § 3 with $\lambda^2 = \varepsilon$, we then get $\bar{x}_i$ and $\bar{x}_i^*$ for $i = 0, 1, \cdots, n$ such that

$$\| \bar{x}_i - x_i \| \leqq \varepsilon^{1/2}, \| \bar{x}_i^* - x_i^* \| \leqq \varepsilon^{1/2}, \bar{x}_i^* \in \partial f(\bar{x}_i) .$$

By choosing ε sufficiently small, we can ensure that also

$$\langle \bar{x}_{i+1} - \bar{x}_i, \bar{x}_i^* \rangle > \alpha_i - \delta \qquad \text{for } i = 0, 1, \cdots, n , \tag{4.7}$$

where we have set $\bar{x}_{n+1} = x_{n+1} = x$. At the same time, by also choosing x_0 sufficiently near to z, we can ensure that $\bar{x}_0$ is as near as we please to z. Now (4.7) implies by the definition of α_i and δ in (4.6) that

$$(4.8) \qquad \langle x - \bar{x}_n, \bar{x}_n^* \rangle + \cdots + \langle \bar{x}_1 - \bar{x}_0, \bar{x}_0^* \rangle > \alpha \ .$$

But

$$\bar{x}_i^* \in \partial f(\bar{x}_i) \subseteqq \partial g(\bar{x}_i)$$

by (4.1), so (4.8) implies

$$\alpha < [g(x) - g(\bar{x}_n)] + \cdots + [g(\bar{x}_1) - g(\bar{x}_0)] = g(x) - g(\bar{x}_0) \ .$$

Furthermore, since each neighborhood of z contains some $\bar{x}_0$ for which this last inequality is true, we have

$$g(x) - \alpha \geqq \liminf_{y \to z} g(y) = g(z)$$

because g is l.s.c. We have shown this for an arbitrary $\alpha < f(x) - f(z)$, so we may conclude (4.2) holds as desired.

To prove the inequality complementary to (4.2), we invoke duality to see that

$$(4.9) \qquad f^*(x^*) - f^*(z^*) \leqq g^*(x^*) - g^*(z^*) \ \text{for all } x^* \ ,$$

where f^* and g^* are the conjugates of f and g, and z^* is any fixed element of E^* with

$$(4.10) \qquad z^* \in \partial f(z) \subseteqq \partial g(z) \ ,$$

z as before. The proof of (4.9) is completely parallel to that of (4.2). It is valid, even though E might not be reflexive, because of the duality explained following the proof of Theorem 2. (In particular, $\partial g^* \supseteqq \partial f^*$ by (4.1) and (3.11).) Now (4.10) implies

$$f(z) + f^*(z^*) = \langle z, z^* \rangle \ ,$$
$$g(z) + g^*(z^*) = \langle z, z^* \rangle \ ,$$

by the definition of the subgradients and conjugate functions, as already in (3.11′). Substituting in (4.9), we get

$$f^*(x^*) + f(z) \leqq g^*(x^*) + g(z) \ \text{for all } x^* \ .$$

Therefore, for all $x \in E$,

$$f(z) + \inf_{x^*}\{f^*(x^*) - \langle x, x^* \rangle\} \leqq g(z) + \inf_{x^*}\{g^*(x^*) - \langle x, x^* \rangle\} \ .$$

But this is the same as

$$(4.11) \qquad f(z) - f(x) \leqq g(z) - g(x) \ \text{for all } x$$

by formula (3.2) for the conjugate of the conjugate function in § 3. In view of (4.2), we must actually have equality in (4.11). Thus g differs from f by at most a constant, as we wanted to prove.

5. Maximal monotone relations. Every cyclically monotone relation is monotone, as pointed out in § 2. It does not follow from this, however, that every *maximal* cyclically monotone relation is a *maximal* monotone relation. We do not know whether or not this is true in general, but the following theorem implies (via Theorem 3) that it is true in Banach spaces.

THEOREM 4. *Let E be a Banach space, and let f be a l.s.c. proper convex function on E. Then ∂f is a maximal monotone relation on $E \times E^*$.*

Proof. Suppose that $z \in E$ and $z^* \in E^*$ have the property that

$$\langle x - z, x^* - z^* \rangle \geqq 0 \quad \text{whenever} \quad x^* \in \partial f(x) . \tag{5.1}$$

We must show that then

$$z^* \in \partial f(z) . \tag{5.2}$$

Actually, replacing f by

$$h(x) = f(z + x) - \langle x, z^* \rangle$$

if necessary, we may assume that $z = 0$ and $z^* = 0$. Thus it is enough to prove the following fact: if

$$0 \notin \partial f(0) \tag{5.3}$$

then there exists some $\bar{x}$ and $\bar{x}^*$ such that

$$\bar{x}^* \in \partial f(\bar{x}) \quad \text{and} \quad \langle \bar{x}, \bar{x}^* \rangle < 0 . \tag{5.4}$$

Now (5.3) implies by definition that $f(0)$ is not the minimum of f on E. Thus there exists some x_0 with

$$f(0) > f(x_0) .$$

Let $Q(\lambda) = f(\lambda x_0)$ for all $\lambda \in R$. Then Q is a l.s.c. proper convex function on the real line and $Q(0) > Q(1)$. Hence, by the well-known theory of one-dimensional convex functions, there exists some λ_0 such that

$$0 < \lambda_0 \leqq 1, Q(\lambda_0) < \infty, Q'_-(\lambda_0) < 0 , \tag{5.5}$$

where Q'_- is the left derivative of Q. In terms of f and its directional derivatives, (5.5) says

$$f(\lambda_0 x_0) < \infty, -f'(\lambda_0 x_0; -x_0) < 0 .$$

Thus for $x = \lambda_0 x_0$ we have

$$f(x) < \infty, f'(x; -x) > 0 . \tag{5.6}$$

Choose any $\varepsilon > 0$. Then by (5.6) and Theorem 2 there exists some x^* with

$$x^* \in \partial_\varepsilon f(x) \quad \text{and} \quad \langle -x, x^* \rangle > 0 .$$

Applying the lemma in §3 with $\lambda^2 = \varepsilon$, we can get an $\bar{x}$ and $\bar{x}^*$ with

$$\| \bar{x} - x \| \leqq \varepsilon^{1/2}, \| \bar{x}^* - x^* \| \leqq \varepsilon^{1/2}, \bar{x}^* \in \partial f(\bar{x}) .$$

Since $\langle x, x^* \rangle < 0$, we will also have $\langle \bar{x}, \bar{x}^* \rangle < 0$ when ε is sufficiently small. This shows that (5.4) can be satisfied, thereby completing the proof of the theorem.

REMARK. The preceding proof closely resembles Minty's proof of his Theorem 2 in [6], except towards the end. The difference is that in the special case treated by Minty f was necessarily subdifferentiable at every point, whereas here an approximation argument based on the difficult results in § 3 was required before any conclusion could be reached about the subgradients of f. A much simpler proof valid for Hilbert spaces has been given by Moreau [11].

REFERENCES

1. T. Bonnesen and W. Fenchel, *Theorie der Konvexen Korper* (Springer, Berlin, 1934).
2. A. Brøndsted, *Conjugate convex functions in topological vector spaces*, Mat. Fys. Medd. Dansk. Vid. Selsk. **34**, No. 2, 1964.
3. A. Brøndsted and R. T. Rockafellar, *On the subdifferentiability of convex functions*, Bull. Amer. Math. Soc. **16** (1965), 605-611.
4. W. Fenchel, *On conjugate convex functions*, Canad. J. Math. **1** (1949), 73-77.
5. R. I. Kachurovskii, *On monotone operators and convex functionals*, Uspekhi **15**, No. 4 (1960), 213-215 (Russian).
6. G. J. Minty, *On the monotonicity of the gradient of a convex function*, Pacific J. Math. **14** (1964), 243-247.
7. ———, *Monotone (nonlinear) operators in Hilbert space*, Duke Math. J. **29** (1962), 341-346.
8. ———, *Fonctionelles sous-différentiables*, C. R. Acad. Sci. **257** (Dec., 1963), 4117-4119.
9. ———, *Sur la fonction polaire d'une fonction continue supérieurment*, C. R. Acad. Sci. **258** (Jan., 1964).
10. ———, *Semi-continuité du sous-gradient d'une fonctionelle*, C. R. Acad. Sci. **260** (1965), 1067-1070.
11. ———, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France **93** (1965), 273-299.
12. R. T. Rockafellar, *Convex Functions and Dual Extremum Problems*, doctoral dissertation (multilith, Harvard, 1963).
13. R. T. Rockafellar, *Level sets and continuity of conjugate convex functions*, to

appear in Trans. Amer. Math. Soc..
14. ———, *Extension of Fenchel's duality theorem for convex functions*, Duke Math. J. **33** (1966), 81–90.

PRINCETON UNIVERSITY