

# GRADIENTS OF CONVEX FUNCTIONS

BY

E. ASPLUND AND R. T. ROCKAFELLAR<sup>(1)</sup>

1. **Introduction.** This paper is concerned with relationship between three notions: the *differentiability* of a convex function  $f$ , the *rotundity* of the convex function  $g$  conjugate to  $f$  and the *continuity* of the subdifferential mapping  $\partial f$  (which reduces to the gradient mapping  $\nabla f$  where  $f$  is differentiable). These notions are considered in the context of various admissible topologies on paired vector spaces.

When  $f$  is the norm  $\|\cdot\|$  on a Banach space  $X$ , or  $f=(1/2)\|\cdot\|^2$ , our results are comparable to the theorems of Šmulyan [19], Cudia [6] and others about the relationship between the differentiability of  $\|\cdot\|$ , the rotundity of the dual unit ball in  $X^*$  and the continuity of the spherical or extended spherical mappings from  $X$  to  $X^*$ ; see Asplund [2]. Our results also contain as a special case some recent results of Lescarret [9] on the strong continuity of gradient mappings in Banach spaces. They are stronger than, but do not quite contain, the theorems of Moreau [12], [13] about the upper semicontinuity of  $\partial f$  and  $\partial_\lambda f$  (see the remark following Proposition 5).

We would like to thank Professor J. J. Moreau for some very helpful suggestions with regard to an earlier version of this paper.

2. **Basic definitions.** Throughout this paper,  $X$  and  $Y$  will denote vector spaces over the real number system  $R$  paired by a bilinear form  $\langle \cdot, \cdot \rangle$ , with respect to which  $X$  distinguishes the points of  $Y$  and  $Y$  distinguishes the points of  $X$ . We denote by  $w(X, Y)$  and  $s(X, Y)$ , respectively, the weak and strong topologies induced on  $X$  by  $Y$ ; similarly  $w(Y, X)$  and  $s(Y, X)$  on  $Y$ . Differentiability properties in the space  $X$  will be shown to be dual to rotundity properties in the space  $Y$ .

Let  $f$  be an extended-real-valued function on  $X$  (i.e. an everywhere-defined function with values in  $R \cup \{\pm\infty\}$ ). Let  $A$  be any nonempty subset of  $X$ . We shall say that  $f$  is *A-differentiable* at a given  $x \in X$  if  $f$  is finite at  $x$  and there exists a  $y \in Y$  such that

$$(2.1) \quad \limsup_{\lambda \downarrow 0, u \in A} \left| \frac{f(x + \lambda u) - f(x)}{\lambda} - \langle u, y \rangle \right| = 0.$$

---

Received by the editors November 21, 1967 and, in revised form, April 24, 1968, and August 13, 1968.

<sup>(1)</sup> The research of the first author was supported in part by the National Science Foundation under grant GP-7536, while that of the second author was supported in part by the Air Force Office of Scientific Research under grant AF-AFOSR-1202-67.

In this event  $y$  is called an  $A$ -gradient of  $f$  at  $x$ , and  $y$  satisfies in particular

$$(2.2) \quad f'(x; u) = \langle u, y \rangle, \quad \forall u \in A,$$

where  $f'(x; u)$  is the *one-sided directional derivative* of  $f$  at  $x$  with respect to  $u$ .

If  $\mathcal{A}$  is a collection of nonempty subsets of  $X$ , we shall say that  $f$  is  $\mathcal{A}$ -differentiable at  $x$  when there exists a  $y$  such that  $y$  is an  $A$ -gradient of  $f$  at  $x$  for every  $A \in \mathcal{A}$ .

If  $\mathcal{A}$  consists of all singleton subsets of  $X$  (subsets consisting of a single point),  $\mathcal{A}$ -differentiability is called *Gâteaux differentiability*. If  $X$  is a normed linear space,  $Y = X^*$  (the dual of  $X$ , in the canonical pairing, of course) and  $\mathcal{A}$  consists of just the unit ball of  $X$ ,  $\mathcal{A}$ -differentiability is *Fréchet differentiability*.

In what follows,  $\mathcal{A}$  will always denote a collection of nonempty subsets of  $X$  such that

- (a) each  $A \in \mathcal{A}$  is  $w(X, Y)$ -bounded,
- (b)  $\bigcup \{A \mid A \in \mathcal{A}\}$  generates  $X$  algebraically,
- (c)  $A \in \mathcal{A}$  implies  $-A \in \mathcal{A}$ .

Assumptions (a) and (b) guarantee that a locally convex Hausdorff topology is induced on  $Y$  by uniform convergence of the linear functionals  $\langle \cdot, y \rangle$  on subsets in  $\mathcal{A}$ ; this topology on  $Y$  will be denoted by  $\mathcal{T}$ . Assumption (b) implies by (2.2) that, for a given  $x \in X$ , there is *at most one*  $y \in Y$  such that  $y$  is an  $A$ -gradient of  $f$  at  $x$  for every  $A \in \mathcal{A}$ . The unique  $y$ , if it exists, is called the  $\mathcal{A}$ -gradient of  $f$  at  $x$ . Assumption (c) implies that, for an  $\mathcal{A}$ -gradient  $y$ , the limits in (2.1) can be taken in the two-sided sense as  $\lambda \rightarrow 0$ , rather than merely as  $\lambda \downarrow 0$ .

We shall mostly be interested in the case where  $f$  is a *convex* function on  $X$ , i.e. where the epigraph

$$\text{epi } f = \{(x, \mu) \mid x \in X, \mu \in R, \mu \geq f(x)\}$$

is a convex set in  $X \oplus R$ . If  $f(x) > -\infty$  for every  $x$  and  $f(x) < +\infty$  for at least one  $x$ , we say that  $f$  is a *proper* convex function. We denote by  $\text{dom } f$  the *effective domain* of  $f$ , which is the convex set

$$\{x \in X \mid f(x) < +\infty\}.$$

A vector  $y \in Y$  is said to be a *subgradient* of  $f$  at  $x$  if

$$(2.3) \quad f(x+u) \geq f(x) + \langle u, y \rangle, \quad \forall u \in X.$$

The set of subgradients of  $f$  at  $x$  is denoted by  $\partial f(x)$ , and the multivalued mapping  $\partial f: x \rightarrow f(x)$  from  $X$  to  $Y$  is called the *subdifferential* of  $f$ . The definition implies that  $\partial f(x)$  is a (possibly empty)  $w(Y, X)$ -closed convex subset of  $Y$  for each  $x$ . Assuming  $f$  is a convex function finite at  $x$ , one has  $y \in \partial f(x)$  if and only if

$$(2.4) \quad \langle u, y \rangle \leq f'(x; u), \quad \forall u \in X.$$

For the general theory of convex functions and subgradients, we refer the reader to [4], [11], [13], [17].

We shall denote by  $\text{rad } f$  the set of points at which the convex set  $\text{dom } f$  is *radial*, i.e. the set of all  $x \in \text{dom } f$  such that every half-line emanating from  $x$  contains points of  $\text{dom } f$  besides  $x$ .

**PROPOSITION 1.** *Let  $f$  be a convex function on  $X$ , and let  $x$  be a point where  $f$  is  $\mathcal{A}$ -differentiable. Then  $f$  is proper and  $x \in \text{rad } f$ . Moreover,  $\partial f(x)$  consists of a single vector  $y$ , the  $\mathcal{A}$ -gradient of  $f$  at  $x$ . This  $y$  is in particular the Gâteaux gradient of  $f$  at  $x$ .*

**Proof.** Let  $U = \bigcup \{A \mid A \in \mathcal{A}\}$ ; by our assumption (b) above, every vector in  $X$  can be expressed as a linear combination of vectors in  $U$ . The  $\mathcal{A}$ -gradient  $y$  satisfies

$$(2.5) \quad \langle u, y \rangle = f'(x; u) = -f'(x; -u)$$

for every  $u \in U$ . Since  $f$  is convex, it follows, as is well known (e.g. see [17, Theorem 25.2]), that (2.5) holds for every  $u$  which is a linear combination of vectors  $u_1, \dots, u_k$  in  $U$ . Thus (2.5) holds for every  $u \in X$ , so that  $\text{dom } f$  is radial at  $x$ , and  $y$  is the Gâteaux gradient. Then  $y$  is the unique subgradient at  $x$ , since it is the only vector satisfying (2.4), and  $f$  is proper by (2.3). This finishes the proof.

Whenever  $\partial f(x)$  consists of exactly one element  $y$  (as in the case of Proposition 1), we shall denote this  $y$  by  $\nabla f(x)$ . The mapping  $\nabla f: x \rightarrow \nabla f(x)$  will be called the *gradient* mapping associated with  $f$ . (The domain of  $\nabla f$  is thus the set of points  $x \in X$  such that  $\partial f(x)$  is nonempty and contains no more than one element, whereas the range of  $\nabla f$  is a subset of  $Y$ .)

Rotundity properties will now be defined. Let  $g$  be an extended real-valued-function on  $Y$ , and let  $B$  be a subset of  $Y$ . We shall say that  $g$  is *B-rotund* at a given  $y \in Y$  relative to a given  $x \in X$  if  $g(y)$  is finite and, for every  $\epsilon > 0$ , there exists a  $\delta > 0$  such that

$$(2.6) \quad \{v \mid g(y+v) - g(y) - \langle x, v \rangle \leq \delta\} \subset \epsilon B.$$

We shall say that  $g$  is  *$\mathcal{T}$ -rotund* at  $y$  relative to  $x$  if  $g$  is  $B$ -rotund for every  $\mathcal{T}$ -neighborhood  $B$  of the origin. Since  $\mathcal{T}$  is a Hausdorff topology on  $Y$ , this condition implies that

$$(2.7) \quad g(y+v) > g(y) + \langle x, v \rangle, \quad \forall v \neq 0,$$

and hence in particular that  $x \in \partial g(y)$ .

The concept of  $\mathcal{T}$ -rotundity is useful in the study of various extremum problems. An extended-real-valued function  $h$  on  $Y$  is said to attain its minimum over a subset  $C$  of  $Y$  at  $y$  *strongly with respect to  $\mathcal{T}$*  if ( $h$  attains its infimum over  $C$  at the point  $y$  and) every sequence  $y_1, y_2, \dots$ , in  $C$  with

$$\lim_{k \rightarrow \infty} h(y_k) = \inf_C h$$

is  $\mathcal{F}$ -convergent to  $y$ . (Here  $\mathcal{F}$ -convergence of sequences actually implies  $\mathcal{F}$ -convergence of all nets.) It is easy to see that  $g$  is  $\mathcal{F}$ -rotund at  $y$  relative to  $x$  if and only if the infimum of the function  $g - \langle x, \cdot \rangle$  over  $Y$  is finite and attained at  $y$  strongly with respect to  $\mathcal{F}$ . Thus, for example, when  $g$  is the indicator of a subset  $C$  of  $Y$ , i.e.

$$(2.8) \quad \begin{aligned} g(y) &= 0 && \text{if } x \in C, \\ &= +\infty && \text{if } x \notin C, \end{aligned}$$

$g$  is  $\mathcal{F}$ -rotund at  $y$  relative to  $x$  if and only if the supremum of the linear functional  $\langle x, \cdot \rangle$  over  $C$  is finite and attained at  $y$  strongly with respect to  $\mathcal{F}$ .

The following result will enable us to concentrate our investigation of rotundity on the case of convex functions.

**PROPOSITION 2.** *Let  $g$  be an extended-real-valued function on  $Y$ , and let  $x$  and  $y$  be elements of  $X$  and  $Y$ , respectively, such that  $g(y)$  is finite. Let  $\bar{g}$  be the  $w(Y, X)$ -lower semicontinuous convex hull of  $g$  (i.e. the pointwise supremum of the collection of all  $w(Y, X)$ -lower semicontinuous convex functions on  $Y$  majorized by  $g$ ). Suppose that there exists a  $\mathcal{F}$ -bounded convex set  $C$  containing the origin in  $Y$  such that*

$$(2.9) \quad \liminf_{\lambda \rightarrow +\infty} \left[ (1/\lambda) \inf_{v \in \lambda C} \{g(y+v) - g(y) - \langle x, v \rangle\} \right] > 0.$$

*Then  $g$  is  $\mathcal{F}$ -rotund at  $y$  relative to  $x$  if and only if  $g$  is  $\mathcal{F}$ -lower semicontinuous at  $y$  and  $\bar{g}$  is  $\mathcal{F}$ -rotund at  $y$  relative to  $x$ . In this event,  $\bar{g}$  is necessarily proper and  $\bar{g}(y) = g(y)$ .*

**Proof.** Replacing  $g$  if necessary by

$$h(v) = g(y+v) - g(y) - \langle x, v \rangle,$$

we can reduce the proof to the case where

$$(2.10) \quad x = 0, \quad y = 0, \quad g(y) = g(0) = 0.$$

Assume that  $g$  is  $\mathcal{F}$ -rotund at 0 relative to 0. Let  $B$  be any closed convex symmetric  $\mathcal{F}$ -neighborhood of 0 in  $Y$ , and let  $p$  be the gauge of  $B$ , i.e.

$$p(v) = \inf \{ \lambda \geq 0 \mid v \in \lambda B \}.$$

Define  $j$  on  $[0, +\infty)$  by

$$j(\lambda) = \inf \{ g(v) \mid p(v) \geq \lambda \}.$$

Clearly  $j$  is a nondecreasing function such that

$$(2.11) \quad j(p(v)) \leq g(v), \quad \forall v \in Y,$$

and by (2.9) we have

$$(2.12) \quad \liminf_{\lambda \rightarrow +\infty} j(\lambda)/\lambda > 0.$$

The  $\mathcal{F}$ -rotundity of  $g$  implies that

$$(2.13) \quad j(0) = 0, \quad \text{but } j(\lambda) > 0 \text{ for every } \lambda > 0.$$

Let  $\bar{j}$  be the lower semicontinuous convex hull of  $j$  on  $[0, +\infty)$ , i.e. the pointwise supremum of the collection of all lower semicontinuous convex functions on  $[0, +\infty)$  majorized by  $j$ . It follows easily from (2.12) and (2.13) that

$$(2.14) \quad \bar{j}(0) = 0, \quad \text{but } \bar{j}(\lambda) > 0 \text{ for every } \lambda > 0,$$

and this implies by the convexity of  $\bar{j}$  that  $\bar{j}$  is increasing. The function

$$(2.15) \quad k(v) = \bar{j}(p(v))$$

is convex and  $w(Y, X)$ -lower semicontinuous on  $Y$ , and  $k \leq g$  by (2.11). Therefore  $k \leq \bar{g} \leq g$ . Since

$$0 = k(0) \leq g(0) = 0,$$

we have  $\bar{g}(0) = g(0) = 0$ , implying that  $g$  is  $w(Y, X)$ -lower semicontinuous at 0, and hence in particular  $\mathcal{F}$ -lower semicontinuous at 0. For every  $\delta > 0$ , we also have

$$(2.16) \quad \{v \mid \bar{g}(v) \leq \delta\} \subset \{v \mid k(v) \leq \delta\}.$$

Given any  $\varepsilon > 0$ , there exists, by (2.14) and the monotonicity of  $\bar{j}$ , some  $\delta > 0$  such that  $\bar{j}(\lambda) \leq \delta$  implies  $\lambda \leq \varepsilon$ . Then  $k(v) \leq \delta$  implies  $p(v) \leq \varepsilon$ , so that by (2.16)

$$\{v \mid \bar{g}(v) \leq \delta\} \subset \{v \mid p(v) \leq \varepsilon\} = \varepsilon B.$$

This shows that  $\bar{g}$  is  $B$ -rotund at 0 relative to 0, and since  $\mathcal{F}$  has a local base consisting of  $w(Y, X)$ -closed convex symmetric sets like  $B$  we may conclude that  $\bar{g}$  is  $\mathcal{F}$ -rotund at 0 relative to 0. In particular  $\bar{g}$  majorizes the constant function 0, so that  $\bar{g}$  is a proper convex function.

On the other hand, assume that  $g$  is  $\mathcal{F}$ -lower semicontinuous at 0, and that  $\bar{g}$  is  $\mathcal{F}$ -rotund at 0 relative to 0. We have

$$(2.17) \quad \inf_{v \in Y} g(v) = \inf_{v \in Y} \bar{g}(v) = \bar{g}(0) \in R$$

(because  $\bar{g} \leq g$  and the constant functions majorized by  $g$  are by definition also majorized by  $\bar{g}$ ), and for each  $\delta > 0$  we have

$$(2.18) \quad \{v \mid \bar{g}(v) \leq \bar{g}(0) + \delta\} \supset \{v \mid g(v) \leq \bar{g}(0) + \delta\} \neq \emptyset.$$

In view of the  $\mathcal{F}$ -rotundity of  $\bar{g}$  at 0 relative to 0, the sets in (2.18) can be made to lie in any given  $\mathcal{F}$ -neighborhood of 0 by choosing  $\delta$  sufficiently small. This implies furthermore, by the  $\mathcal{F}$ -lower semicontinuity of  $g$  at 0, that  $\bar{g}(0) = g(0) = 0$ . Therefore  $g$  is  $\mathcal{F}$ -rotund at 0.

**REMARK.** Condition (2.9) in Proposition 2 is satisfied trivially when the set of points where  $g$  does not have the value  $+\infty$  is  $\mathcal{F}$ -bounded.

**3. Duality between differentiability and rotundity.** The *conjugate* of an extended-real-valued function  $g$  on  $Y$  (with respect to the pairing between  $X$  and  $Y$ ) is the function  $f$  on  $X$  defined by

$$(3.1) \quad f(x) = \sup \{ \langle x, y \rangle - g(y) \mid y \in Y \}.$$

As is known,  $f$  is a  $w(X, Y)$ -lower semicontinuous convex function on  $X$ . If  $g$  is a  $w(Y, X)$ -lower semicontinuous proper convex function, then  $f$  is proper and  $g$  is in turn the conjugate of  $f$ , i.e. one has

$$(3.2) \quad g(y) = \sup \{ \langle x, y \rangle - f(x) \mid x \in X \}.$$

In this case,

$$(3.3) \quad y \in \partial f(x) \Leftrightarrow x \in \partial g(y) \Leftrightarrow \langle x, y \rangle - f(x) - g(y) = 0.$$

The following fundamental result about polars of level sets of conjugate convex functions is implicitly contained in a paper of Moreau [10]. As usual, we denote by  $C^0$  the polar of a set  $C \subset Y$ , i.e.

$$C^0 = \{ u \in X \mid \langle u, v \rangle \leq 1, \quad \forall v \in C \}.$$

**PROPOSITION 3.** *Let  $f$  and  $g$  be proper convex functions conjugate to each other on  $X$  and  $Y$ , respectively. Let  $x \in X$  and  $y \in Y$  be such that the (nonpositive) quantity*

$$\alpha = \inf_{v \in Y} \{ g(y+v) - g(y) - \langle x, v \rangle \} = \langle x, y \rangle - f(x) - g(y)$$

*is finite. Then, for any  $\delta > 0$ ,*

$$(3.4) \quad \{ v \mid g(y+v) - g(y) - \langle x, v \rangle \leq \delta \}^0 \subset \delta^{-1} \{ u \mid f(x+u) - f(x) - \langle u, y \rangle \leq \delta \} \\ \subset 2 \{ v \mid g(y+v) - g(y) - \langle x, v \rangle \leq \alpha + \delta \}^0.$$

**Proof.** The proper convex functions

$$f_0(z) = [f(x + \delta z) - f(x) - \langle \delta z, y \rangle - \delta] / \delta, \\ g_0(v) = [g(y+v) - g(y) - \langle x, v \rangle - \alpha + \delta] / \delta,$$

are conjugate to each other (as can be verified by direct calculation), and they satisfy

$$(3.5) \quad 1 = \inf g_0 \leq g_0(0) < +\infty,$$

$$(3.6) \quad -1 = f_0(0) \geq \inf f_0 > -\infty.$$

In terms of these functions we can rewrite (3.4) as

$$(3.7) \quad \{ v \mid g_0(v) \leq g_0(0) + 1 \}^0 \subset C \subset 2 \{ v \mid g_0(v) \leq 2 \}^0,$$

where

$$C = \{ z \mid f_0(z) \leq 0 \}.$$

Here  $C$  is a  $w(X, Y)$ -closed convex subset of  $X$  containing the origin, so that  $C^{00} = C$ . Thus (3.7) (and Proposition 3) will follow if we show that

$$(3.8) \quad \{v \mid g_0(v) \leq g_0(0) + 1\} \supset C^0 \supset (1/2)\{v \mid g_0(v) \leq 2\}.$$

Let  $h$  denote the support function of  $C$  on  $Y$ , i.e.

$$h(v) = \sup \{\langle z, v \rangle \mid z \in C\}.$$

According to [14, Corollary 4B],  $h$  is related to  $g_0$  by the formula

$$(3.9) \quad h(v) = \inf_{\lambda > 0} \lambda g_0(\lambda^{-1}v).$$

In particular  $h \leq g_0$  by (3.9), so that

$$\{v \mid g_0(v) \leq 2\} \subset \{v \mid h(v) \leq 2\} = 2\{v \mid h(v) \leq 1\} = 2C^0.$$

This establishes the right half of (3.8). To establish the left half of (3.8), it suffices to show that

$$(3.10) \quad \{v \mid h(v) < 1\} \subset \{v \mid g_0(v) \leq g_0(0) + 1\},$$

since the  $w(Y, X)$ -closure of the set on the left in (3.10) is

$$\{v \mid h(v) \leq 1\} = C^0,$$

whereas the set on the right in (3.10) is  $w(Y, X)$ -closed. Given any  $v$  such that  $h(v) < 1$ , there exists by (3.9) some  $\lambda > 0$  such that  $\lambda g_0(\lambda^{-1}v) < 1$ . Since  $g_0(\lambda^{-1}v) \geq 1$  by (3.5), this  $\lambda$  must actually satisfy  $0 < \lambda < 1$ . The convexity of  $g_0$  implies then that

$$g_0(v) \leq (1 - \lambda)g_0(0) + \lambda g_0(\lambda^{-1}v) < g_0(0) + 1.$$

Thus (3.10) holds, and the proof of Proposition 3 is complete.

The fundamental duality between differentiability and rotundity can now be proved by means of Proposition 3.

**PROPOSITION 4.** *Let  $f$  and  $g$  be proper convex functions conjugate to each other on  $X$  and  $Y$ , respectively. Let  $A$  be any nonempty subset of  $X$ , and let  $B$  be the polar of  $A$  in  $Y$ . Let  $x \in X$  and  $y \in \partial f(x)$ . Then  $y$  is an  $A$ -gradient of  $f$  at  $x$  if and only if  $g$  is  $B$ -rotund at  $y$  relative to  $x$ .*

**Proof.** By definition, since  $y \in \partial f(x)$  and (2.3) holds,  $y$  is an  $A$ -gradient at  $x$  if and only if, for every  $\varepsilon > 0$ , there exists a  $\mu > 0$  such that

$$(3.7) \quad \sup_{u \in A} \left\{ \frac{f(x + \lambda u) - f(x)}{\lambda} - \langle u, y \rangle \right\} \leq \varepsilon, \quad 0 < \lambda \leq \mu.$$

Since  $[f(x + \lambda u) - f(x)]/\lambda$  is a nondecreasing function of  $\lambda > 0$  by the convexity of  $f$ , we can write (3.7) as

$$\varepsilon^{-1}A \subset \varepsilon^{-1}\mu^{-1}\{u \mid f(x + u) - f(x) - \langle u, y \rangle \leq \mu\varepsilon\}.$$

It follows from Proposition 3 (with  $\alpha=0$ , as indicated by (3.3)) that  $y$  is an  $A$ -gradient at  $x$  if and only if, for every  $\varepsilon > 0$ , there exists a  $\delta > 0$  such that

$$(3.8) \quad \varepsilon^{-1}A \subset \{v \mid g(y+v) - g(y) - \langle x, v \rangle \leq \delta\}^0.$$

Taking polars, we can express (3.8) equivalently as

$$\varepsilon B \supset \{v \mid g(y+v) - g(y) - \langle x, v \rangle \leq \delta\}.$$

Thus  $y$  is an  $A$ -gradient if and only if  $g$  is  $B$ -rotund at  $y$  relative to  $x$ .

**THEOREM 1.** *Let  $f$  and  $g$  be proper convex functions conjugate to each other on  $X$  and  $Y$ , respectively. Then  $f$  is  $\mathcal{A}$ -differentiable at  $x$  with  $y = \nabla f(x)$  if and only if  $g$  is  $\mathcal{T}$ -rotund at  $y$  relative to  $x$ .*

**Proof.** If  $y$  is the  $\mathcal{A}$ -gradient of  $f$  at  $x$ , we have  $y \in \partial f(x)$  in particular, by Proposition 1. On the other hand, if  $g$  is  $\mathcal{T}$ -rotund at  $y$  relative to  $x$  we have  $x \in \partial g(y)$ , and hence  $y \in \partial f(x)$ . Thus in either case Proposition 4 is applicable, and it follows that  $f$  is  $\mathcal{A}$ -differentiable at  $x$  with  $\mathcal{A}$ -gradient  $y$  if and only if  $g$  is  $B$ -rotund at  $y$  relative to  $x$  for every  $B \in \mathcal{B}$ , where  $\mathcal{B}$  consists of all the polars of sets in  $\mathcal{A}$ . The latter condition is equivalent to  $g$  being  $\mathcal{T}$ -rotund at  $y$  relative to  $x$ , since the sets of the form

$$B = \lambda_1 B_1 \cap \cdots \cap \lambda_m B_m, \quad \lambda_i > 0, \quad B_i \in \mathcal{B},$$

are a local base for  $\mathcal{T}$ , and since  $g$  is  $B$ -rotund for such a  $B$  if and only if  $g$  is  $B_i$ -rotund for  $i=1, \dots, m$ .

**COROLLARY 1.** *Let  $f$  and  $g$  be proper convex functions conjugate to each other on  $X$  and  $Y$ , respectively. Then  $f$  is Gâteaux differentiable at  $x$  with  $y = \nabla f(x)$  if and only if  $g$  is  $w(Y, X)$ -rotund at  $y$  relative to  $x$ .*

*In particular,  $f$  is Gâteaux differentiable at 0 with  $y = \nabla f(0)$  if and only if  $g$  attains its infimum over  $Y$  at  $y$  strongly with respect to the  $w(Y, X)$ -topology.*

**Proof.** Take  $\mathcal{A}$  to be the collection of all singleton subsets of  $X$ .

**COROLLARY 2.** *Let  $X$  be a normed linear space, and let  $Y = X^*$  (in the canonical pairing). Let  $f$  and  $g$  be proper convex functions conjugate to each other on  $X$  and  $Y$ , respectively. Then  $f$  is Fréchet differentiable at  $x$  with  $y = \nabla f(x)$  if and only if  $g$  is norm rotund (i.e. rotund with respect to the norm topology) at  $y$  relative to  $x$ .*

*In particular,  $f$  is Fréchet differentiable at 0 with  $y = \nabla f(0)$  if and only if  $g$  attains its infimum over  $Y$  at  $y$  strongly with respect to the norm topology.*

**Proof.** Let  $\mathcal{A}$  consist of just the unit ball in  $X$ .

**COROLLARY 3.** *Let  $f$  be any  $w(X, Y)$ -lower semicontinuous proper convex function on  $X$ . If  $f$  is  $\mathcal{A}$ -differentiable at  $x$ , then  $f$  is actually  $\mathcal{A}'$ -differentiable at  $x$ , where  $\mathcal{A}'$  consists of all the nonempty  $\mathcal{T}$ -equicontinuous subsets of  $X$ .*

**Proof.** The topology induced on  $Y$  by uniform convergence of linear functionals on subsets of  $\mathcal{A}'$  is the same as  $\mathcal{T}$ . Hence the  $\mathcal{A}'$ -differentiability of  $f$  at  $x$  is the same as the  $\mathcal{A}$ -differentiability of  $f$  at  $x$  by Theorem 1.

Corollary 3 implies in particular that (if  $f$  is any  $w(X, Y)$ -lower semicontinuous proper convex function on  $X$ )  $f$  is Gâteaux differentiable at  $x$  if and only if  $f$  is actually  $\mathcal{T}$ -differentiable at  $x$ , where  $\mathcal{T}$  is the collection of all nonempty finite-dimensional bounded subsets of  $X$ .

**COROLLARY 4.** *Let  $f$  be a  $w(X, Y)$ -lower semicontinuous proper convex function on  $X$ . Suppose there exists an  $\mathcal{A}$ -bounded linear functional  $k$  on  $X$  such that, for a given  $x \in X$ ,  $f(x)$  is finite and*

$$\limsup_{\lambda \downarrow 0} \sup_{u \in A} \left| \frac{f(x + \lambda u) - f(x)}{\lambda} - k(u) \right| = 0, \quad \forall A \in \mathcal{A}.$$

*If  $Y$  is  $\mathcal{T}$ -complete, then there exists a  $y \in Y$  such that  $k(u) = \langle u, y \rangle$  for every  $u \in X$ , so that  $f$  is actually  $\mathcal{A}$ -differentiable at  $x$ .*

**Proof.** Let  $Z$  be the space of all  $\mathcal{A}$ -bounded linear functionals on  $X$ . Under the canonical pairing between  $X$  and  $Z$  (with respect to which the sets in  $\mathcal{A}$  are all  $w(X, Z)$ -bounded), we may regard  $Y$  as a  $w(Z, X)$ -dense subspace of  $Z$  which by  $\mathcal{T}$ -completeness is  $\mathcal{T}$ -closed (the  $\mathcal{T}$ -topology on  $Z$  being, of course, the topology of uniform convergence on the sets in  $\mathcal{A}$ ). Let  $g$  be the convex function on  $Z$  conjugate to  $f$ . Then  $f$  is the conjugate of  $g$  with respect to the pairing between  $X$  and  $Z$ , and at the same time  $f$  is the conjugate of the restriction of  $g$  to  $Y$  with respect to the pairing between  $X$  and  $Y$ . The infimum of  $g - \langle x, \cdot \rangle$  over  $Z$  is thus the same as the infimum of  $g - \langle x, \cdot \rangle$  over  $Y$ , namely  $-f(x)$ . Let  $y_1, y_2, \dots$  be a sequence in  $Y$  such that  $g(y_k) - \langle y_k, x \rangle$  decreases to this infimum. By Theorem 1,  $g$  is  $\mathcal{T}$ -rotund relative to  $x$  at the point of  $Z$  corresponding to  $k$ , so that  $y_k$   $\mathcal{T}$ -converges to this point. Since  $Y$  is  $\mathcal{T}$ -closed in  $Z$ , this point must actually belong to  $Y$ .

**COROLLARY 5.** *Let  $Y$  be a Banach space, and let  $X = Y^*$ . Let  $f$  be a  $w(X, Y)$ -lower semicontinuous proper convex function on  $X$ . If  $f$  is Fréchet differentiable at  $x$ , the Fréchet gradient actually belongs to  $Y$  (rather than merely to  $Y^{**}$ ). Thus  $f$  is Fréchet differentiable at  $x$  if and only if the function  $g - \langle x, \cdot \rangle$  (where  $g$  is the conjugate of  $f$ ) attains its infimum over  $Y$  strongly with respect to the norm topology.*

**Proof.** Apply Theorem 1 and the preceding corollary in the case where  $\mathcal{A}$  consists of just the unit ball of  $X$ .

In the case of Corollary 5 where  $f$  is the norm on  $X = Y^*$ ,  $g$  is the indicator of the unit ball  $C$  of  $Y$  as in (2.8), and one gets a classical result of Šmul'yan [19]: the norm on  $Y^*$  is Fréchet differentiable at  $x$  if and only if the linear functional  $\langle x, \cdot \rangle$  attains its supremum over  $C$  strongly with respect to the norm topology.

Theorem 1 can be extended by means of Proposition 2 to cases where  $g$  is not convex or everywhere  $w(Y, X)$ -lower semicontinuous.

**COROLLARY 6.** *Let  $g$  be an extended-real-valued function on  $Y$ , and let  $f$  be the conjugate of  $g$  on  $X$ . Let  $x$  and  $y$  be elements of  $X$  and  $Y$ , respectively, such that  $g$  is finite and  $\mathcal{T}$ -lower semicontinuous at  $y$ . Suppose there exists a  $\mathcal{T}$ -bounded convex set  $C$  containing the origin in  $Y$  such that*

$$\liminf_{\lambda \rightarrow \infty} \left[ (1/\lambda) \inf_{v \in \lambda C} \{g(y+v) - g(y) - \langle x, v \rangle\} \right] > 0$$

(which is true in particular if  $g$  has the value  $+\infty$  everywhere outside of a certain  $\mathcal{T}$ -bounded subset of  $Y$ ). Then  $f$  is  $\mathcal{A}$ -differentiable at  $x$  with  $y = \nabla f(x)$  if and only if  $g$  is  $\mathcal{T}$ -rotund at  $y$  relative to  $x$ .

**Proof.** Let  $\bar{g}$  be the  $w(Y, X)$ -lower semicontinuous convex hull of  $g$ . Of course,  $f$  is also the conjugate of  $\bar{g}$ , and if either  $f$  or  $\bar{g}$  is proper  $\bar{g}$  is in turn the conjugate of  $f$ . The  $\mathcal{A}$ -differentiability of  $f$  at  $x$  implies the properness of  $f$  by Proposition 1, whereas the  $\mathcal{T}$ -rotundity of  $g$  at  $y$  relative to  $x$  implies the properness of  $\bar{g}$  by Proposition 2. The result is thus immediate from Theorem 1 and Proposition 2.

**4. Differential continuity.** We shall now explore the relationship between the differentiability of a  $w(X, Y)$ -lower semicontinuous proper convex function  $f$  on  $X$  and the continuity of the subdifferential mapping  $\partial f: x \rightarrow \partial f(x)$ . More generally, we shall consider continuity properties of the mapping

$$(x, \lambda) \rightarrow \partial_\lambda f(x), \quad x \in X, \lambda \in R,$$

where  $\partial_\lambda f(x)$  is the set of vectors  $y \in Y$  such that

$$f(x+u) \geq (f(x) - \lambda) + \langle u, y \rangle, \quad \forall u \in X.$$

Note that  $\partial f$  is the restriction of the latter mapping to  $\lambda=0$ .

One has

$$\partial_\lambda f(x) = \{y \mid g(y) - \langle x, y \rangle \leq \beta + \lambda\},$$

where  $g$  is the conjugate of  $f$  and

$$\beta = \inf_{y \in Y} \{g(y) - \langle x, y \rangle\} = -f(x).$$

Thus  $\partial_\lambda f(x)$  is always a  $w(Y, X)$ -closed convex subset of  $Y$ , nonempty in particular when  $x \in \text{dom } f$  and  $\lambda > 0$  (but empty when  $x \notin \text{dom } f$  or  $\lambda < 0$ ).

Throughout this section,  $\mathcal{S}$  will denote a topology on  $X$  which is a locally convex Hausdorff topology having a local base consisting of  $w(X, Y)$ -closed (convex) sets (polars of certain  $w(Y, X)$ -bounded subsets of  $Y$ ).

If the convex set  $\text{dom } f$  has a nonempty interior with respect to  $\mathcal{S}$ , this interior is of course  $\text{rad } f$  (as defined just prior to Proposition 1). If  $\text{dom } f$  happens to be closed with respect to  $w(X, Y)$ , then  $\text{rad } f$  is by definition the  $s(X, Y)$ -interior of

$\text{dom } f$ . Actually, even when  $\text{dom } f$  is not  $w(X, Y)$ -closed,  $\text{rad } f$  is the  $s(X, Y)$ -interior of  $\text{dom } f$  by the convexity and  $w(X, Y)$ -lower semicontinuity of  $f$  (see Corollary 1 below).

The following results will be needed.

**THEOREM 2 (MOREAU [10, 13]).** *Let  $f$  be a  $w(X, Y)$ -lower semicontinuous proper convex function on  $X$ .*

(a) *Suppose that  $f$  has a finite upper bound in some  $\mathcal{S}$ -neighborhood of some point of  $X$ . Then  $\partial_\lambda f(x)$  is an  $\mathcal{S}$ -equicontinuous subset of  $Y$  for every  $x \in \text{rad } f$  and  $\lambda \in \mathbb{R}$ . In fact, given any  $x \in \text{rad } f$  and any  $\mu \in \mathbb{R}$ , there exists an  $\mathcal{S}$ -neighborhood  $U$  of the origin such that the sets  $\partial_\lambda f(x+u)$  are uniformly  $\mathcal{S}$ -equicontinuous for  $u \in U$  and  $\lambda \leq \mu$ , i.e. such that*

$$(4.1) \quad \bigcup \{ \partial_\lambda f(x+u) \mid u \in U, \lambda \leq \mu \}$$

*is  $\mathcal{S}$ -equicontinuous.*

(b) *On the other hand, suppose that, for some  $x \in X$  and some  $\lambda > 0$ ,  $\partial_\lambda f(x)$  is nonempty and  $\mathcal{S}$ -equicontinuous. Then  $x \in \text{rad } f$ , and  $f$  is  $\mathcal{S}$ -continuous throughout  $\text{rad } f$ , so that (a) applies and  $\text{rad } f$  is the  $\mathcal{S}$ -interior of  $\text{dom } f$ .*

**Proof.** This can be deduced from Proposition 3 and the well-known fact [3, Chapter II, p. 92] that a finite convex function on an open convex set  $W$  is continuous throughout  $W$  if it is merely bounded above in a neighborhood of some point of  $W$ . Proposition 3 says that the inclusions

$$(4.2) \quad [\partial_\lambda f(x) - y]^0 \subset C \subset 2[\partial_\mu f(x) - y]^0$$

hold for

$$(4.3) \quad C = \mu^{-1} \{ u \mid f(x+u) - f(x) - \langle u, y \rangle \leq \mu \}$$

whenever  $\lambda, \mu, x$  and  $y$  are such that  $\mu > 0, f(x) < +\infty, g(y) < +\infty$  and

$$(4.4) \quad \lambda - \mu = f(x) + g(y) - \langle x, y \rangle.$$

Suppose that  $f$  has a finite upper bound in some  $\mathcal{S}$ -neighborhood of some point of  $X$ , as hypothesized in (a). Then  $\text{dom } f$  has a nonempty  $\mathcal{S}$ -interior, which must be  $\text{rad } f$ , and  $f$  is  $\mathcal{S}$ -continuous throughout this interior. Fix any  $y \in Y$  such that  $g(y)$  is finite. Since  $g$  is conjugate to  $f$ , we have

$$g(y) \geq \langle x, y \rangle - f(x)$$

for every  $x \in X$ ; thus the linear function  $\langle \cdot, y \rangle$  on  $X$  is majorized by a function of the form  $f + \text{const}$ , implying that  $\langle \cdot, y \rangle$  has a finite upper bound in some  $\mathcal{S}$ -neighborhood of some point of  $X$  and hence that  $\langle \cdot, y \rangle$  is  $\mathcal{S}$ -continuous. The expression

$$f(x+u) - f(x) - \langle u, y \rangle$$

in (4.3) therefore depends  $\mathcal{S}$ -continuously on  $x$  and  $u$ , as long as  $x \in \text{rad } f$  and  $x+u \in \text{rad } f$ . It follows that, given any  $x \in \text{rad } f$  and any real number  $\mu > 0$ , there exist  $\mathcal{S}$ -neighborhoods  $U$  and  $V$  of the origin in  $X$  such that

$$f(x'+u)-f(x')-\langle u, y \rangle \leq \mu$$

whenever  $x'-x \in U$  and  $u \in \mu V$ , or in other words

$$\mu^{-1}\{u \mid f(x'+u)-f(x')-\langle u, y \rangle \leq \mu\} \supset V$$

whenever  $x'-x \in U$ . By the second inclusion in (4.2), we then have

$$V \subset 2[\partial_\mu f(x')-y]^0$$

whenever  $x'-x \in U$ . Thus

$$\partial_\lambda f(x+u) \subset \partial_\mu f(x+u) \subset 2V^0+y$$

for every  $u \in U$  and  $\lambda \leq \mu$ . Since  $V$  is an  $\mathcal{S}$ -neighborhood of the origin, and the linear function  $\langle \cdot, y \rangle$  is  $\mathcal{S}$ -continuous, the set  $2V^0+y$  is  $\mathcal{S}$ -equicontinuous in  $Y$  and (a) is established.

To prove (b), suppose now that  $x \in X$  and  $\lambda > 0$  are such that  $\partial_\lambda f(x)$  is nonempty and  $\mathcal{S}$ -equicontinuous. Then  $f(x) < +\infty$ . Choose any  $y \in \partial_\lambda f(x)$  such that actually

$$f(x)+g(y)-\langle x, y \rangle < \lambda$$

(as is possible by (3.1), since  $f$  is the conjugate of  $g$ ), and define  $\mu$  by (4.4). Then (4.2) holds. Since  $y$  belongs to the  $\mathcal{S}$ -equicontinuous set  $\partial_\lambda f(x)$ , the linear function  $\langle \cdot, y \rangle$  is  $\mathcal{S}$ -continuous and the translate  $\partial_\lambda f(x)-y$  is again an  $\mathcal{S}$ -equicontinuous set. The polar of  $\partial_\lambda f(x)-y$  in  $X$  is therefore an  $\mathcal{S}$ -neighborhood of the origin. Then, by the first inclusion in (4.2), the convex function

$$h(u) = f(x+u)-f(x)-\langle u, y \rangle$$

is bounded above in some  $\mathcal{S}$ -neighborhood of the origin and hence is  $\mathcal{S}$ -continuous at the origin. This implies that  $f$  is  $\mathcal{S}$ -continuous at  $x$ , and the conclusion of (b) is immediate.

**COROLLARY 1** (CF. ROCKAFELLAR [14]). *Let  $f$  be a  $w(X, Y)$ -lower semicontinuous proper convex function on  $X$ . Then  $f$  is  $s(X, Y)$ -continuous throughout  $\text{rad } f$  (so that  $\text{rad } f$  is in particular the  $s(X, Y)$ -interior of  $\text{dom } f$ ). Moreover*

(a)  $\partial_\lambda f(x)$  is a  $w(Y, X)$ -bounded subset of  $Y$  for every  $x \in \text{rad } f$  and  $\lambda \in R$ . In fact, given any  $x \in \text{rad } f$  and any  $\mu \in R$ , there exists an  $s(X, Y)$ -neighborhood  $U$  of the origin such that the sets  $\partial_\lambda f(x+u)$  are uniformly  $w(Y, X)$ -bounded for  $u \in U$  and  $\lambda \leq \mu$ , i.e.

$$\bigcup \{\partial_\lambda f(x+u) \mid u \in U, \lambda \leq \mu\}$$

is  $w(Y, X)$ -bounded.

(b) *On the other hand, if  $x$  is such that  $\partial_\lambda f(x)$  is nonempty and  $w(Y, X)$ -bounded for some  $\lambda > 0$ , then  $x \in \text{rad } f$ .*

**Proof.** For any  $x \in \text{rad } f$ , the  $w(X, Y)$ -closed convex set

$$\{u \mid f(x+u) \leq f(x)+1\}$$

is radial at the origin, and hence it is by definition an  $s(X, Y)$ -neighborhood of the origin. It follows from the continuity fact cited at the beginning of the proof of Theorem 2 that  $f$  is  $s(X, Y)$ -continuous throughout  $\text{rad } f$ . The corollary is then obtained by specializing  $\mathcal{S}$  to  $s(X, Y)$  in Theorem 2.

**COROLLARY 2 (MOREAU [10], [13]).** *Let  $f$  be a  $w(X, Y)$ -lower semicontinuous proper convex function on  $X$ , such that  $f$  has a finite upper bound in some  $\mathcal{S}$ -neighborhood of some point of  $X$ . Assume that  $\mathcal{S}$  is compatible with the duality between  $X$  and  $Y$ . Then  $\partial f(x)$  is a nonempty  $w(Y, X)$ -compact (convex) subset of  $Y$  for every  $x \in \text{rad } X$ . In fact, given any  $x \in \text{rad } f$ , there exists an  $\mathcal{S}$ -neighborhood  $U$  of the origin such that the set*

$$\bigcup \{\partial f(x+u) \mid u \in U\}$$

*is relatively  $w(Y, X)$ -compact.*

**Proof.** Since  $\mathcal{S}$  is compatible with the duality,  $\mathcal{S}$ -equicontinuous sets are relatively  $w(X, Y)$ -compact. Then, for  $x \in \text{rad } f$ ,  $\partial f(x)$  is nonempty, because  $\partial f(x)$  is the intersection of the nonempty  $w(Y, X)$ -closed  $\mathcal{S}$ -equicontinuous sets  $\partial_\lambda f(x)$ ,  $\lambda > 0$ .

It can be shown, incidentally, that under the hypothesis of Corollary 2 one has

$$(4.5) \quad f'(x; u) = \max \{\langle u, y \rangle \mid y \in \partial f(x)\}$$

for every  $x \in \text{rad } f$ . In this event, of course,  $f$  is Gâteaux-differentiable at  $x$  if  $x \in \text{rad } f$  and  $\partial f(x)$  contains just one vector  $y$ . (See Moreau [10].)

We shall denote by  $\bar{Y}$  the collection of all nonempty  $w(Y, X)$ -closed bounded convex subsets of  $Y$ . According to Corollary 1,  $\partial_\lambda f(x)$  is an element of  $\bar{Y}$  for every  $x \in \text{rad } f$  and  $\lambda > 0$ .

By the  $\mathcal{T}$ -topology on  $\bar{Y}$ , we shall mean the topology in which, for each  $C \in \bar{Y}$ , the sets of the form

$$\{D \in \bar{Y} \mid D \subset C+B \text{ and } C \subset D+B\}$$

constitute a fundamental system of neighborhoods of  $C$  as  $B$  ranges over all  $\mathcal{T}$ -neighborhoods of the origin in  $Y$ . Of course, when  $Y$  is regarded as a subset of  $\bar{Y}$ , the relative  $\mathcal{T}$ -topology induced on  $Y$  is the same as the  $\mathcal{T}$ -topology already present on  $Y$ .

The  $\mathcal{T}$ -topology on  $\bar{Y}$  can also be obtained in a dual way. For each  $C \in \bar{Y}$ , let  $h_C$  denote the support function of  $C$  on  $X$ , i.e.

$$h_C(x) = \sup_{y \in C} \langle x, y \rangle.$$

As is well known [7], the correspondence  $C \leftrightarrow h_C$  is one-to-one between  $\bar{Y}$  and the collection of all finite  $w(X, Y)$ -lower semicontinuous positively homogeneous convex functions on  $X$ , and it preserves addition and nonnegative scalar multiplication. (This is a natural extension of the correspondence between points of  $Y$  and linear functions on  $X$ .) It is not hard to show that the  $\mathcal{T}$ -topology on  $\bar{Y}$  is the same as the topology on  $\bar{Y}$  induced by uniform convergence of the support functions  $h_C$  on all  $\mathcal{T}$ -equicontinuous subsets of  $X$ .

Note that a sequence of sets  $D_1, D_2, \dots$  in  $\bar{Y}$   $\mathcal{T}$ -converges to a point  $y \in Y$  (regarded as a singleton set in  $\bar{Y}$ ) if and only if, for every  $\mathcal{T}$ -neighborhood  $B$  of the origin in  $Y$ , there exists an integer  $m$  such that  $D_k \subset y + B$  for all  $k > m$ . Thus the conjugate  $g$  of  $f$  is  $\mathcal{T}$ -rotund at  $y$  relative to  $x$  if and only if the nonempty  $w(Y, X)$ -closed convex sets

$$y + \{v \mid g(y+v) - \langle x, v \rangle \leq g(y) + \lambda\}, \quad \lambda > 0,$$

are  $w(Y, X)$ -bounded and  $\mathcal{T}$ -converge in  $\bar{Y}$  to  $y$  as  $\lambda \downarrow 0$ . Since (3.3) holds and

$$\partial_\lambda f(x) = y + \{v \mid g(y+v) - \langle x, v \rangle \leq \langle x, y \rangle - f(x) + \lambda\},$$

the conclusion of Theorem 1 can therefore be stated as follows:  *$f$  is  $\mathcal{A}$ -differentiable at  $x$  with  $y = \nabla f(x)$  if and only if the sets  $\partial_\lambda f(x)$  are nonempty and bounded for  $\lambda > 0$  (i.e.  $x \in \text{rad } f$ ) and  $\partial_\lambda f(x)$   $\mathcal{T}$ -converges in  $\bar{Y}$  to  $y$  as  $\lambda \downarrow 0$ .*

**PROPOSITION 5.** *Let  $f$  be a  $w(X, Y)$ -lower semicontinuous proper convex function on  $X$ , such that  $f$  has a finite upper bound in some  $\mathcal{S}$ -neighborhood of some point of  $X$ . Assume that all the sets in  $\mathcal{A}$  are  $\mathcal{S}$ -bounded (which is true in particular if either  $\mathcal{S}$  or  $\mathcal{T}$  is compatible with the duality between  $X$  and  $Y$ ). The mapping*

$$(x, \lambda) \rightarrow \partial_\lambda f(x)$$

*is then continuous from  $(\text{rad } f) \times (0, +\infty)$  in the  $\mathcal{S}$ -topology (i.e. the product of the  $\mathcal{S}$ -topology on  $\text{rad } f$  and the ordinary topology on the real interval  $(0, +\infty)$ ) to  $\bar{Y}$  in the  $\mathcal{T}$ -topology.*

**Proof.** Choose any  $x \in \text{rad } f$ , any  $\delta > 0$ , and any  $w(Y, X)$ -closed convex  $\mathcal{T}$ -neighborhood  $B$  of the origin in  $Y$ . Let  $U$  be an  $\mathcal{S}$ -neighborhood of the origin in  $X$  such that, for  $\mu = 2\delta$ , the set in (4.1), which we shall denote by  $M$ , is  $\mathcal{T}$ -equicontinuous. Since the sets in  $\mathcal{A}$  are all  $\mathcal{S}$ -bounded in  $X$ ,  $\mathcal{S}$ -equicontinuous sets are all  $\mathcal{T}$ -bounded in  $Y$ . Thus we can find a real number  $\rho \geq 2\delta$  such that

$$(4.6) \quad M - M \subset \rho B.$$

Choose an  $\varepsilon > 0$  such that

$$(4.7) \quad 2\varepsilon < \delta/(\rho + 1).$$

Since  $f$  is  $\mathcal{S}$ -continuous at  $x$  (Theorem 2), there exists an  $\mathcal{S}$ -neighborhood  $W$  of the origin,  $W \subset U$ , such that

$$(4.8) \quad |f(x+u_1) - f(x+u_2)| \leq \varepsilon/2, \quad \forall u_1 \in W, \forall u_2 \in W,$$

and

$$(4.9) \quad |\langle w, y \rangle| \leq \varepsilon/4, \quad \forall w \in W, \forall y \in M.$$

We shall demonstrate that

$$\partial_\lambda f(x+u) \subset \partial_\delta f(x) + B \quad \text{and} \quad \partial_\delta f(x) \subset \partial_\lambda f(x+u) + B$$

for arbitrary  $u \in W$  and  $\lambda \in [\delta - \varepsilon, \delta + \varepsilon]$ , and this will establish the desired continuity at the point  $(x, \delta)$ .

In what follows,  $\alpha$  and  $\beta$  denote real numbers such that  $\alpha < \beta$ . For any  $u_1$  and  $u_2$  in  $(\text{dom } f) - x$ ,  $\partial_\alpha f(x+u_1)$  consists of the elements  $y \in Y$  such that, for every  $z \in X$ ,

$$\begin{aligned} f(z) &\geq f(x+u_1) + \langle z-x-u_1, y \rangle - \alpha \\ &= f(x+u_2) + \langle z-x-u_2, y \rangle \\ &\quad - [\alpha + f(x+u_2) - f(x+u_1) + \langle u_2-u_1, y \rangle]. \end{aligned}$$

If  $u_1 \in W$ ,  $u_2 \in W$ ,  $\alpha \leq 2\delta$  and  $\beta - \alpha \geq \varepsilon$ , every  $y \in \partial_\alpha f(x+u_1)$  belongs to  $M$ , so that by (4.8) and (4.9)

$$\alpha + f(x+u_2) - f(x+u_1) + \langle u_2-u_1, y \rangle \leq \alpha + (\varepsilon/2) + (\varepsilon/4) + (\varepsilon/4) \leq \beta,$$

and hence

$$\partial_\alpha f(x+u_1) \subset \partial_\beta f(x+u_2).$$

It follows in particular that

$$(4.10) \quad \partial_{\lambda-\varepsilon} f(x) \subset \partial_\lambda f(x+u) \subset \partial_{\lambda+\varepsilon} f(x), \quad \forall u \in W, \forall \lambda \leq 2\delta.$$

On the other hand, for any  $\theta > 1$  we have

$$\alpha = (1/\theta)\gamma + [1 - (1/\theta)]\beta \quad \text{for } \gamma = \beta + \theta(\alpha - \beta) < \alpha,$$

so that

$$(4.11) \quad (1/\theta)\partial_\gamma f(x) + [1 - (1/\theta)]\partial_\beta f(x) \subset \partial_\alpha f(x).$$

The latter inclusion follows from the fact that

$$\partial_\lambda f(x) = \{y \mid h(y) \leq \lambda\},$$

where  $h$  is a certain proper convex function on  $Y$ , namely

$$h(y) = f(x) + g(y) - \langle x, y \rangle$$

( $g$  being the conjugate of  $f$ ). If  $0 < \alpha < \beta \leq 2\delta$ , and  $\theta > 1$  is such that  $\gamma > 0$ , i.e.

$$(4.12) \quad 1 < \theta < \beta/(\beta - \alpha),$$

we have by (4.11) and (4.6)

$$\begin{aligned}\partial_\beta f(x) &\subset [1 - (1/\theta)]^{-1}[\partial_\alpha f(x) - (1/\theta)\partial_\gamma f(x)] \\ &= \partial_\alpha f(x) + [1/(\theta - 1)][\partial_\alpha f(x) - \partial_\gamma f(x)] \\ &\subset \partial_\alpha f(x) + [\rho/(\theta - 1)]B.\end{aligned}$$

This calculation uses the fact that, since  $\partial_\alpha f(x)$  is a convex set, one has

$$\partial_\alpha f(x) = [1 - (1/\theta)]\partial_\alpha f(x) + (1/\theta)\partial_\alpha f(x).$$

If  $\alpha = \delta - 2\varepsilon$  and  $\beta = \delta$ , (4.12) is satisfied for  $\theta = \rho + 1$  by virtue of (4.7), and consequently

$$(4.13) \quad \partial_\delta f(x) \subset \partial_{\delta - 2\varepsilon} f(x) + B.$$

Similarly, if  $\alpha = \delta$  and  $\beta = \delta + 2\varepsilon$ , (4.12) is satisfied for  $\theta = \rho + 1$ , so that

$$(4.14) \quad \partial_{\delta + 2\varepsilon} f(x) \subset \partial_\delta f(x) + B.$$

For any  $u \in W$  and  $\lambda \in [\delta - \varepsilon, \delta + \varepsilon]$ , we have  $\lambda < 2\delta$  (since  $\varepsilon < \delta$ ) and

$$\partial_{\delta - 2\varepsilon} f(x) \subset \partial_{\lambda - \varepsilon} f(x) \subset \partial_{\lambda + \varepsilon} f(x) \subset \partial_{\delta + 2\varepsilon} f(x),$$

so that

$$\partial_\delta f(x) \subset \partial_\lambda f(x + u) + B$$

by (4.10) and (4.13), while

$$\partial_\lambda f(x + u) \subset \partial_\delta f(x) + B$$

by (4.10) and (4.14). This completes the proof of Proposition 5.

**COROLLARY** *Let  $f$  be a  $w(X, Y)$ -lower semicontinuous proper convex function on  $X$ . Assume either that  $X$  is a barrelled space in some topology such that  $Y = X^*$ , or that  $Y$  is a barrelled space in some topology such that  $X = Y^*$ . Then the mapping  $(x, \lambda) \rightarrow \partial_\lambda f(x)$  is continuous from  $(\text{rad } f) \times (0, +\infty)$  in the  $s(X, Y)$ -topology to  $\bar{Y}$  in the  $s(Y, X)$ -topology.*

**Proof.** Invoke Corollary 1 of Theorem 2 along with the fact that, in a barrelled space or its dual, the weakly bounded sets are the same as the strongly bounded sets [3, Chapter III].

Proposition 5 and the results below should be compared with a recent theorem of Moreau [13, 11d]. This result says that (under the same hypothesis about  $f$  as in Proposition 5) if  $\mathcal{S}$  is compatible with the duality between  $X$  and  $Y$  and  $\mathcal{T}$  is the topology of uniform convergence on  $\mathcal{S}$ -compact subsets of  $X$ , then the mapping  $(x, \lambda) \rightarrow \partial_\lambda f(x)$  is *upper semicontinuous* as a multivalued mapping from  $(\text{rad } f) \times R$  in the  $\mathcal{S}$ -topology to  $Y$  in the  $\mathcal{T}$ -topology. In this case,  $\partial f$  is in particular  $(\mathcal{S}, \mathcal{T})$ -upper semicontinuous from  $\text{rad } f$  to  $Y$ , so that the gradient mapping  $\nabla f$  is  $(\mathcal{S}, \mathcal{T})$ -continuous (and consequently  $(\mathcal{S}, w(Y, X))$ -continuous) where it exists.

If  $f$  is  $\mathcal{A}$ -differentiable at  $x$  and the hypothesis of Proposition 5 is satisfied, it follows from Proposition 5 and the observation preceding it that  $\partial_\lambda f(z)$   $\mathcal{T}$ -converges to  $\nabla f(x)$  in  $\bar{Y}$  as  $z$   $\mathcal{S}$ -converges to  $x$  and  $\lambda \downarrow 0$ . This implies in turn that  $\partial f(z)$   $\mathcal{T}$ -converges to  $\nabla f(x)$  as  $z$   $\mathcal{S}$ -converges to  $x$  in the set

$$D = \{z \in \text{rad } f \mid \partial f(z) \neq \emptyset\}.$$

Thus the mapping  $\partial f$  from  $D$  to  $\bar{Y}$  is  $(\mathcal{S}, \mathcal{T})$ -continuous at  $x$ . When  $\mathcal{S}$  is compatible with the duality between  $X$  and  $Y$ , we have  $D = \text{rad } f$  by Corollary 2 of Theorem 3, and a stronger result may be stated.

**THEOREM 3.** *Let  $f$  be a  $w(X, Y)$ -lower semicontinuous proper convex function on  $X$ , such that  $f$  has a finite upper bound in some  $\mathcal{S}$ -neighborhood of some point of  $X$ . Assume that  $\mathcal{S}$  is compatible with the duality between  $X$  and  $Y$  (see the remark below). Then, in order that the mapping  $x \rightarrow \partial f(x)$  from  $\text{rad } f$  to  $\bar{Y}$  be  $(\mathcal{S}, \mathcal{T})$ -continuous at  $x$ , it is necessary and sufficient that  $f$  be  $\mathcal{A}$ -differentiable at  $x$ .*

**Proof.** The sufficiency of the condition has just been explained. To prove the necessity, fix any  $x \in \text{rad } f$  at which the mapping in question is  $(\mathcal{S}, \mathcal{T})$ -continuous. It is enough to show that, given any  $\varepsilon > 0$  and any  $A \in \mathcal{A}$ , there exists a  $\lambda > 0$  such that

$$(4.15) \quad [f(x + \lambda u) - f(x)]/\lambda - \langle u, y \rangle \leq 2\varepsilon, \quad \forall u \in A, \forall y \in \partial f(x).$$

Let  $B = A^0 \cap (-A)^0$ . Then  $B$  is a symmetric convex  $\mathcal{T}$ -neighborhood of the origin in  $Y$ , and

$$(4.16) \quad |\langle u, v \rangle| \leq 1, \quad \forall u \in A, \forall v \in B.$$

Let  $U$  be an  $\mathcal{S}$ -neighborhood of  $x$  such that

$$(4.17) \quad \partial f(x + w) \subset \partial f(x) + \varepsilon B, \quad \forall w \in U,$$

$$(4.18) \quad \partial f(x) \subset \partial f(x + w) + \varepsilon B, \quad \forall w \in U.$$

Since  $\mathcal{S}$  is compatible with the duality, the set  $A$ , being  $w(X, Y)$ -bounded by assumption, is actually  $\mathcal{S}$ -bounded, and we can choose a  $\lambda > 0$  such that  $\lambda A \subset U$ . Let  $u \in A$ . If  $v \in \partial f(x + \lambda u)$ , we have by definition

$$f(x) \geq f(x + \lambda u) + \langle x - (x + \lambda u), v \rangle,$$

and consequently

$$(4.19) \quad [f(x + \lambda u) - f(x)]/\lambda - \langle u, y \rangle \leq \langle u, v \rangle - \langle u, y \rangle$$

for any  $y \in Y$ . When  $y \in \partial f(x)$ , the left side of (4.19) is nonnegative, so that  $\langle u, y \rangle \leq \langle u, v \rangle$ . Since the latter holds for any  $y \in \partial f(x)$  and  $v \in \partial f(x + \lambda u)$ , we must have

$$(4.20) \quad \sup \{\langle u, w \rangle \mid w \in \partial f(x)\} \leq \inf \{\langle u, w \rangle \mid w \in \partial f(x + \lambda u)\}.$$

On the other hand, since  $\lambda u \in \lambda A \subset U$ , for any  $v \in \partial f(x + \lambda u)$  we have

$$(4.21) \quad \begin{aligned} \langle u, v \rangle &\leq \sup \{ \langle u, w \rangle \mid w \in \partial f(x + \lambda u) \} \\ &\leq \sup \{ \langle u, w \rangle \mid w \in \partial f(x) \} + \varepsilon \end{aligned}$$

by (4.17) and (4.16), while for any  $y \in \partial f(x)$  we have

$$(4.22) \quad \begin{aligned} \langle u, y \rangle &\geq \inf \{ \langle u, w \rangle \mid w \in \partial f(x) \} \\ &\geq \inf \{ \langle u, w \rangle \mid w \in \partial f(x + \lambda u) \} - \varepsilon \end{aligned}$$

by (4.18) and (4.16). Combining (4.20), (4.21) and (4.22), we see that, for any  $v \in \partial f(x + \lambda u)$  and  $y \in \partial f(x)$ ,

$$\begin{aligned} \langle u, v \rangle &\leq \inf \{ \langle u, w \rangle \mid w \in \partial f(x + \lambda u) \} + \varepsilon \\ &\leq \langle u, y \rangle + 2\varepsilon \end{aligned}$$

in (4.19), and the inequality (4.15) holds as desired.

**REMARK.** The compatibility assumption on  $\mathcal{S}$  in Theorem 3 was used only to ensure that all the sets in  $\mathcal{A}$  are  $\mathcal{S}$ -bounded, and that  $\partial f(z) \neq \emptyset$  for every  $z$  in some  $\mathcal{S}$ -neighborhood of the point  $x \in \text{rad } f$  where continuity was in question. Therefore, the conclusion of Theorem 3 remains valid when the latter conditions are satisfied, even if  $\mathcal{S}$  is not compatible with the duality between  $X$  and  $Y$ .

**COROLLARY 1.** *Let  $f$  be a  $w(X, Y)$ -lower semicontinuous proper convex function on  $X$ . Suppose that  $X$  is a barrelled space in some topology such that  $Y = X^*$ . In order that  $f$  be Gâteaux differentiable at  $x$ , it is necessary and sufficient that  $x \in \text{rad } f$  and that the mapping  $\partial f$  from  $\text{rad } f$  to  $\bar{Y}$  be continuous at  $x$  from the  $s(X, Y)$ -topology to the  $w(Y, X)$ -topology.*

**Proof.** Take  $\mathcal{S} = s(X, Y)$  and  $\mathcal{T} = w(Y, X)$ . The hypothesis implies that  $\mathcal{S}$  is compatible with the duality between  $X$  and  $Y$  [3, Chapter III]. Moreover,  $f$  is  $\mathcal{S}$ -continuous on  $\text{rad } f$  by Corollary 1 of Theorem 2, so Theorem 3 is applicable. (If  $\text{rad } f$  were empty,  $f$  could not be Gâteaux differentiable anywhere in view of Proposition 1, so that the present corollary would be vacuous.)

**COROLLARY 2.** *Let  $X$  be a normed linear space with  $Y = X^*$ , and let  $f$  be a proper convex function on  $X$  lower semicontinuous with respect to the norm topology. In order that  $f$  be Fréchet differentiable at  $x$ , it is necessary and sufficient that  $x \in \text{rad } f$  and that the mapping  $\partial f$  from  $\text{rad } f$  to  $\bar{Y}$  be continuous at  $x$  from the norm topology to the norm topology.*

**Proof.** We note first that  $f$  is actually  $w(X, Y)$ -lower semicontinuous, since lower semicontinuity depends only on the closedness of the convex level sets  $\{x \mid f(x) \leq \alpha\}$ , and the closed convex sets in  $X$  are the same in all topologies compatible with the duality between  $X$  and  $Y$ . Apply Theorem 3 as in the proof of the preceding corollary, but with  $\mathcal{T} = s(Y, X)$ .

**COROLLARY 3.** *Let  $X$  be a normed linear space with  $Y = X^*$ , and let  $f$  be a  $w(X, Y)$ -lower semicontinuous proper convex function on  $X$ . Let  $U$  be a nonempty subset of  $X$  open with respect to the norm topology. Suppose that, for each  $x \in U$ ,  $\partial f(x)$  consists of a single element of  $Y$  (denoted by  $\nabla f(x)$ ). Then  $f$  is Gâteaux differentiable throughout  $U$ , and the gradient mapping  $\nabla f$  from  $U$  into  $Y$  is continuous from the norm topology to the  $w(Y, X)$ -topology. In order that  $f$  actually be Fréchet differentiable throughout  $U$ , it is necessary and sufficient that  $\nabla f$  be norm-to-norm continuous from  $U$  into  $Y$ .*

**Proof.** Since  $\partial f(x) = \emptyset$  for  $x \notin \text{dom } f$ ,  $U$  must be contained in  $\text{dom } f$ , and hence in  $\text{rad } f$ . By Corollary 1 of Theorem 2,  $f$  is norm continuous throughout  $U$ , so that (4.5) holds, implying that  $\nabla f(x)$  is the Gâteaux gradient of  $f$  at  $x$ . The result now follows from Corollary 1 and Corollary 2 above.

In view of the known monotonicity properties of  $\partial f$  [15], Theorem 3 and its corollaries suggest that, for a continuous convex function, differentiability is to be expected in a large subset of the domain of continuity. Such results are classical in the finite-dimensional case, and they have also been proved for certain classes of Banach spaces. For details, see Asplund [2].

**5. Gradient homeomorphisms.** In this section, we shall characterize certain cases where the gradient mapping  $\nabla f$  is a homeomorphism between subsets of  $X$  and  $Y$  in the strong topologies.

**THEOREM 4.** *Suppose that the strong topology  $s(X, Y)$  on  $X$  is compatible with the duality between  $X$  and  $Y$ , and that  $Y$  is  $s(Y, X)$ -complete and let  $f$  be a proper convex function on  $X$ . In order that  $\nabla f$  be a homeomorphism from (all of)  $X$  onto  $Y$  with respect to topologies  $s(X, Y)$  and  $s(Y, X)$ , it is necessary and sufficient that  $s(X, Y)$  be normable in such a way that*

- (a)  $f$  is Fréchet differentiable at every  $x \in X$ ,
- (b)  $f$  is norm rotund relative to  $\nabla f(x)$  at every  $x \in X$ ,
- (c)  $f - \langle \cdot, y \rangle$  attains its minimum on  $X$  for every  $y \in Y$ .

*In this event,  $X$  is a reflexive Banach space with respect to the norm in question, with  $Y = X^*$ . Moreover, the convex function  $g$  on  $Y$  conjugate to  $f$  likewise satisfies conditions (a), (b) and (c), and the gradient mapping  $\nabla g$  is the inverse of  $\nabla f$ . The conjugate of  $g$  is in turn  $f$ .*

**Proof.** The existence of  $\nabla f(x)$  for every  $x$  implies that  $f$  is the pointwise supremum of the affine functions of the form

$$z \rightarrow f(x) + \langle z - x, \nabla f(x) \rangle, \quad x \in X,$$

so that  $f$  is  $w(X, Y)$ -lower semicontinuous, and  $f$  is in turn the conjugate of  $g$ .

Assume that  $\nabla f$  is a homeomorphism as specified. Then by (3.3),  $\nabla g$  is the inverse of  $\nabla f$ , so that  $\nabla g$  is a homeomorphism from  $Y$  onto  $X$  with respect to the strong topologies. Since the domains of  $\nabla f$  and  $\nabla g$  are contained in  $\text{dom } f$  and

dom  $g$ , respectively,  $f$  and  $g$  must be finite everywhere. Corollary 1 of Theorem 2 implies that  $f$  and  $g$  are everywhere continuous with respect to the strong topologies, and that all the sets of the form  $\partial_\lambda f(x)$  and  $\partial_\lambda g(y)$  are weakly bounded. In particular,  $f$  is  $s(X, Y)$ -continuous at 0, and for  $\lambda = f(0) + g(0) + 1$  the set

$$\partial_\lambda g(0) = \{x \mid f(x) \leq f(0) + 1\}$$

is  $w(X, Y)$ -bounded (implying  $s(X, Y)$ -bounded, because  $s(X, Y)$  is compatible with the duality), so that  $X$  contains  $s(X, Y)$ -neighborhoods of the origin which are  $s(X, Y)$ -bounded. It follows that  $s(X, Y)$  is normable. We can suppose therefore that  $X$  is a normed linear space with  $Y = X^*$  (the strong topologies on  $X$  and  $Y$  being the norm topologies). The norm-to-norm continuity of  $\nabla f$  implies by Theorem 3 that (a) holds. Similarly, the norm-to-norm continuity of  $\nabla g$  implies by Theorem 3 and the remark following it that  $g$  is Fréchet differentiable throughout  $Y$ , and hence by Theorem 1 that  $f$  is norm rotund at  $\nabla g(y)$  relative to  $y$  for every  $y \in Y$ . This yields conditions (b) and (c), because  $x = \nabla g(y)$  if and only if  $y = \nabla f(x)$ .

Conversely, suppose  $s(X, Y)$  is normable in such a way that (a), (b) and (c) are satisfied. By (a) and Theorem 3,  $\partial f$  reduces to the single-valued mapping  $\nabla f$ , and this mapping is norm-to-norm continuous from all of  $X$  into  $Y$ . By (c), the range of  $\nabla f$  is all of  $Y$ . By (b) and Theorem 1,  $g$  is Fréchet differentiable at  $y = \nabla f(x)$  for each  $x \in X$  (and hence at each  $y \in Y$ ). Furthermore, the Fréchet gradient mapping  $\nabla g$ , which must be the inverse of  $\nabla f$  in view of (3.3), is norm-to-norm continuous by Theorem 3. Thus  $\nabla f$  is a homeomorphism from  $X$  onto  $Y$  with respect to the norm topologies.

An incomplete normed linear space cannot be homeomorphic to a Banach space, as has been proved by Klee [8], so that  $X$ , being homeomorphic to its dual  $Y$ , must be complete. It remains only to show that  $X^{**} = X$ . Let  $h$  be the conjugate of  $g$  on  $X^{**}$  with respect to the canonical pairing between  $X^{**}$  and  $X^* = Y$ . Since  $g$  is in particular norm rotund at  $\nabla f(0)$  relative to 0 (by virtue of  $f$  being Fréchet differentiable at 0), it follows by the duality between  $h$  and  $g$  that  $h$  is Fréchet differentiable at 0. Hence, by Proposition 1, dom  $h$  has a nonempty interior in  $X^{**}$ . Let  $D$  be the set of points in  $X^{**}$  where  $h$  has a subgradient coming from  $X^*$ . According to Brøndsted-Rockafellar [5, Theorem 2],  $D$  is dense in dom  $h$  in the norm topology. But  $D$  is just the range of  $\nabla g$ , which is  $X$ , because under the pairing between  $X^{**}$  and  $X^*$  we have  $y \in \partial h(x)$  if and only if  $x \in \partial g(y)$ ; cf. (3.3). This shows that the nonempty interior of dom  $h$  in  $X^{**}$  is contained in  $X$ , and consequently that  $X^{**} = X$ , proving Theorem 4.

In Theorem 4, the finite convex functions  $f$  and  $g$  can be expressed in terms of each other by the formulas

$$(5.1) \quad g(y) = \langle (\nabla f)^{-1}(y), y \rangle - f((\nabla f)^{-1}(y)),$$

$$(5.2) \quad f(x) = \langle x, (\nabla g)^{-1}(x) \rangle - g((\nabla g)^{-1}(x)).$$

In other words, each is the *Legendre transform* of the other. In Theorem 5, we shall further extend the theory of Legendre transforms of convex functions (see [16] for the finite-dimensional case) to certain extended-real-valued functions on reflexive Banach spaces. The following proposition is needed as a preliminary.

**PROPOSITION 6.** *Let  $X$  be a Banach space with  $Y=X^*$ , and let  $f$  be a proper convex function on  $X$  lower semicontinuous with respect to the norm topology. Suppose there exists a subset  $W$  of  $X$ , open in the norm topology, such that  $W \cap \text{dom } f \neq \emptyset$  and*

$$\bigcup \{ \partial f(x) \mid x \in W \}$$

*is bounded in  $Y$ . Then  $\text{rad } f \neq \emptyset$ , and  $\partial f(x)$  is nonempty and bounded if and only if  $x \in \text{rad } f$ .*

**Proof.** As noted already in the proof of Corollary 2 to Theorem 3, lower semicontinuity with respect to the norm topology implies  $w(X, Y)$ -lower semicontinuity. The lemma of Brøndsted-Rockafellar [5] says that, for any  $x \in X$  and any positive  $\varepsilon$  and  $\lambda$ , one has

$$(5.3) \quad \partial_\varepsilon f(x) \subset \bigcup \{ \partial f(x + \lambda u) \mid u \in U \} + (\varepsilon/\lambda)B,$$

where  $U$  is the unit ball in  $X$  and  $B$  is the dual unit ball in  $Y$ . If  $x \in W \cap \text{dom } f$ , and  $\lambda > 0$  is chosen so small that  $x + \lambda U \subset W$ , the right side of (5.3) is a bounded subset of  $Y$  by the hypothesis, implying that  $\partial_\varepsilon f(x)$  is bounded (as well as nonempty). Then  $x \in \text{rad } f$  by Corollary 1 of Theorem 2, so that  $\text{rad } f \neq \emptyset$ . On the other hand, suppose  $\text{rad } f \neq \emptyset$ . Then  $\text{rad } f$  is the interior of  $\text{dom } f$  with respect to  $s(X, Y)$ , which is the norm topology. Moreover,  $f$  is continuous on  $\text{rad } f$  with respect to the norm topology (Corollary 1 of Theorem 2), and  $\partial f(x)$  is nonempty and bounded for every  $x \in \text{rad } f$  (Corollary 2 of Theorem 2).

Now let  $x$  be a point of  $X$  such that  $x \notin \text{rad } f$  but  $\partial f(x) \neq \emptyset$ ; we shall show that  $\partial f(x)$  is unbounded, and this will complete the proof of Proposition 6. Since  $x$  is not an interior point of the convex set  $\text{dom } f$  (and this interior  $\text{rad } f$  is nonempty),  $x$  can be separated from  $\text{dom } f$ , i.e. there exists a nonzero  $v \in Y$  such that

$$\langle x, v \rangle \geq \langle z, v \rangle, \quad \forall z \in \text{dom } f.$$

For any  $y \in \partial f(x)$  and any  $z \in \text{dom } f$ , we then have

$$\begin{aligned} f(z) &\geq f(x) + \langle z - x, y \rangle \\ &\geq f(x) + \langle z - x, y + \lambda v \rangle, \quad \forall \lambda \geq 0. \end{aligned}$$

This also holds trivially when  $z \notin \text{dom } f$  (since then  $f(z) = +\infty$ ), and hence it holds for every  $z \in X$ . Therefore

$$y + \lambda v \in \partial f(x), \quad \forall y \in \partial f(x), \forall \lambda \geq 0,$$

and since  $v \neq 0$  it follows that  $\partial f(x)$  is not bounded.

**THEOREM 5.** *Let  $X$  be a reflexive Banach space with  $Y = X^*$ , and let  $f$  be a proper convex function on  $X$  which is lower semicontinuous with respect to the norm topology. Let*

$$D = \{x \in X \mid \partial f(x) \neq \emptyset\}.$$

*In order that  $\partial f$  be a one-to-one mapping from  $D$  into  $\bar{Y}$  which is a homeomorphism with respect to the norm topologies, it is necessary and sufficient that the following conditions be satisfied (in which case  $\partial f$  reduces to the gradient mapping  $\nabla f$  and  $D = \text{rad } f$ ):*

- (a)  $\text{rad } f \neq \emptyset$ ,
- (b)  $f$  is Fréchet differentiable throughout  $\text{rad } f$ ,
- (c)  $f$  is rotund relative to  $\nabla f(x)$  in the norm topology at each  $x \in \text{rad } f$ ,
- (d) whenever  $z \in \text{rad } f$  and  $x$  is a point of  $\text{dom } f$  not in  $\text{rad } f$ , one has

$$\lim_{\lambda \downarrow 0} \langle z - x, \nabla f((1 - \lambda)x + \lambda z) \rangle = -\infty.$$

*In this event the convex function  $g$  conjugate to  $f$  on  $Y$  likewise satisfies conditions (a), (b), (c) and (d). Furthermore, the domain and range of the mapping  $\nabla f$  are the open convex sets  $\text{rad } f$  and  $\text{rad } g$ , respectively, and the inverse of  $\nabla f$  is  $\nabla g$ .*

**Proof.** As noted in the proof of Corollary 2 of Theorem 3,  $f$  is actually  $w(X, Y)$ -lower semicontinuous. Suppose that  $\nabla f$  is a homeomorphism from  $D$  into  $\bar{Y}$  with respect to the norm topologies. Then  $D = \text{rad } f \neq \emptyset$  by Proposition 6. The continuity of  $\nabla f$  implies by Corollary 2 of Theorem 3 that  $f$  is Fréchet differentiable throughout  $\text{rad } f$ , so that  $\partial f$  reduces to  $\nabla f$ . Thus  $\nabla f$  is a one-to-one mapping of  $\text{rad } f$  into  $Y$  which is a norm-to-norm homeomorphism. By (3.3),  $\partial g$  reduces to  $\nabla g$ , and  $\nabla g$  is the inverse of  $\nabla f$  (consequently a homeomorphism). It follows by applying the preceding argument to  $g$  in place of  $f$  (which is permissible because  $X$  is reflexive) that

$$(5.4) \quad \{y \mid \partial g(y) \neq \emptyset\} = \text{rad } g \neq \emptyset,$$

and that  $\nabla g$  is a norm-to-norm homeomorphism of  $\text{rad } g$  onto  $\text{rad } f$ . Hence  $g$  is Fréchet differentiable at every  $y \in \text{rad } g$  by Corollary 2 of Theorem 3. By Corollary 2 to Theorem 1,  $f$  is then norm rotund at  $\nabla g(y)$  relative to  $y$  for every  $y \in \text{rad } g$ . Thus conditions (a), (b) and (c) hold, along with

$$(d') \quad D \subset \text{rad } f.$$

Conversely, suppose that (a), (b), (c) and (d') hold. Then  $\partial f$  reduces to  $\nabla f$ , which is by Corollary 2 of Theorem 3 a norm-to-norm continuous mapping of  $D$  into  $Y$ . The range of  $\nabla f$  is  $\{y \mid \partial g(y) \neq \emptyset\}$  by (3.3), and for each  $y$  in this range there exists by condition (c) some  $x$  such that  $f$  is norm rotund at  $x$  relative to  $y$ . This implies by Theorem 1 that  $g$  is Fréchet differentiable at  $y$ . The inverse of  $\nabla f$  is then  $\nabla g$  by (3.3), and  $\nabla g$  is norm-to-norm continuous by Corollary 2 of Theorem 3. In other words,  $\nabla f$  is a norm-to-norm homeomorphism.

To complete the proof of Theorem 5, it suffices to show that (d') is equivalent to (d), assuming that (a) and (b) hold. Of course, (d') is equivalent under (a) to the condition that  $\partial f(x) = \emptyset$  for every  $x$  which is a point of  $\text{dom } f$  not in  $\text{rad } f$ . Let  $x$  be a particular point of  $\text{dom } f$  not in  $\text{rad } f$ , and let  $z \in \text{rad } f$ . For  $0 < \lambda \leq 1$ , the point  $(1-\lambda)x + \lambda z$  belongs to  $\text{rad } f$ , since  $\text{rad } f$  is the interior of the convex set  $\text{dom } f$  in the norm topology when  $\text{rad } f \neq \emptyset$ . The restriction of  $f$  to the line segment joining  $x$  and  $z$  is a lower semicontinuous convex function, so the directional derivative

$$f'((1-\lambda)x + \lambda z; z-x) = \langle z-x, \nabla f((1-\lambda)x + \lambda z) \rangle$$

decreases to  $f'(x; z-x)$  as  $\lambda$  decreases to 0. We must show that  $\partial f(x) \neq \emptyset$  if and only if  $f'(x; z-x) > -\infty$ . One direction is easy: if  $\partial f(x)$  contains an element  $y$ , we have

$$(5.5) \quad \langle z-x, y \rangle \leq f'(x; z-x) = \inf_{\lambda > 0} \frac{f(x + \lambda(z-x)) - f(x)}{\lambda},$$

and hence  $f'(x; z-x) > -\infty$ . On the other hand, suppose  $f'(x; z-x) > -\infty$ . The function  $h = f'(x; \cdot)$  is, of course, convex and positively homogeneous on  $X$ , and it satisfies

$$(5.6) \quad h(w-x) \leq f(w) - f(x), \quad \forall w \in X.$$

Since  $f$  is norm continuous on  $\text{rad } f$  by Corollary 1 of Theorem 2,  $h$  is finitely bounded above on a (norm) neighborhood of the point  $z-x$ . Hence, by the fact cited to prove Theorem 2,  $h$  is not only finite but norm continuous at  $z-x$ . This implies that  $h$  majorizes at least one continuous affine function (see Brøndsted [4]), and since  $h$  is positively homogeneous the affine function can be taken to be linear. Thus there exists a  $y \in Y$  such that

$$\langle u, y \rangle \leq h(u) \leq f(x+u) - f(x), \quad \forall u \in X.$$

Thus  $y$  belongs to  $\partial f(x)$ , so  $\partial f(x) \neq \emptyset$  and the proof is complete.

The functions  $f$  and  $g$  in Theorem 5 are given in terms of each other by the Legendre transformation formulas (5.1) and (5.2) for  $x \in \text{rad } f$  and  $y \in \text{rad } g$ . Thus Theorem 5 yields a certain one-to-one Legendre correspondence between certain pairs  $(f, C)$ , where  $C$  is an open convex set in  $X$  and  $f$  is a Fréchet differentiable convex function on  $C$ , and pairs  $(g, D)$ , where  $D$  is an open convex set in  $Y$  and  $g$  is a Fréchet differentiable convex function on  $D$ . This correspondence can be described directly, i.e. without mentioning conjugacy, much as in [16]; we leave the details to the reader.

In the case where the gradient homeomorphism is required to map all of  $X$  onto  $Y$ , Theorem 5 yields a result which is comparable to Theorem 4 but stronger due to the assumption of reflexivity.

**COROLLARY** *Let  $X$  be a reflexive Banach space with  $Y = X^*$ , and let  $f$  be a proper convex function on  $X$ . In order that the gradient mapping  $\nabla f$  be a norm-to-norm homeomorphism from  $X$  onto  $Y$ , it is necessary and sufficient that the following conditions hold:*

- (a)  *$f$  is Fréchet differentiable at every  $x \in X$ ,*
- (b)  *$f$  is norm rotund relative to  $\nabla f(x)$  at every  $x \in X$ ,*
- (c)  *$\lim_{\lambda \rightarrow +\infty} f(\lambda x)/\lambda = +\infty$  for every nonzero  $x \in X$ .*

*In this event the conjugate  $g$  of  $f$  likewise satisfies conditions (a), (b) and (c), and  $\nabla g$  is the inverse of  $\nabla f$ .*

**Proof.** The existence of  $\nabla f(x)$  for every  $x$  implies, as explained at the beginning of the proof of Theorem 4, that  $f$  is  $w(X, Y)$ -lower semicontinuous. Conditions (a) and (b) of the corollary are necessary and sufficient, in view of Theorem 5, for  $\nabla f$  to be a norm-to-norm homeomorphism from  $X$  into  $Y$ . The range of  $\nabla f$  is then  $\text{rad } g$  by Theorem 5, and  $\text{rad } g$  is the interior of the convex set  $\text{dom } g$ . Thus the range of  $\nabla f$  is all of  $Y$  if and only if  $\text{dom } g$  is  $w(Y, X)$ -dense in  $Y$ , which is equivalent to condition (c) according to Rockafellar [14, Theorem 5B].

#### REFERENCES

1. E. Asplund, *Averaged norms*, Israel J. Math. **5** (1967), 227–233.
2. ———, *Fréchet differentiability of convex functions*, Acta Math. **121** (1968), 31–47.
3. N. Bourbaki, *Espaces vectoriels topologiques*, Hermann, Paris, 1953 and 1955.
4. A. Brøndsted, *Conjugate convex functions in topological vector spaces*, Mat.-Fys. Medd. Danske. Vid. Selsk. **34** (1964), 1–26.
5. A. Brøndsted and R. T. Rockafellar, *On the subdifferentiability of convex functions*, Proc. Amer. Math. Soc. **16** (1965), 605–611.
6. D. F. Cudia, *The geometry of Banach spaces. Smoothness*, Trans. Amer. Math. Soc. **110** (1964), 284–314.
7. L. Hörmander, *Sur la fonction d'appui des ensembles convexes dans une espace localement convexe*, Ark. Mat. **3** (1954), 181–186.
8. V. L. Klee, *Invariant metrics in groups (solution of a problem of Banach)*, Proc. Amer. Math. Soc. **3** (1952), 484–487.
9. C. Lescarret, *Applications 'prox' dans une espace de Banach*, C.R. Acad. Sci. Paris **265** (1967), 676–678.
10. J. J. Moreau, *Sur la fonction polaire d'une fonction semi-continue supérieurement*, C.R. Acad. Sci. Paris **258** (1964), 1128–1131.
11. ———, *Sous-différentiabilité*, Proceedings of the Colloquium on Convexity, Copenhagen, 1965, W. Fenchel (Editor), Copenhagen, 1967, pp. 185–201.
12. ———, *Semi-continuité du sous-gradient d'une fonctionnelle*, C.R. Acad. Sci. Paris **260** (1965), 1067–1070.
13. ———, *Fonctionnelles convexes*, mimeographed lecture notes, Seminaire "Equations aux dérivées partielles," Collège de France, 1966.
14. R. T. Rockafellar, *Level sets and continuity of conjugate convex functions*, Trans. Amer. Math. Soc. **123** (1966), 46–63.
15. ———, *Characterization of the subdifferentials of convex functions*, Pacific J. Math. **17** (1966), 81–90. A correction to the proof of Theorems 3 and 4 of this paper is given in [18].

16. R. T. Rockafellar, *Conjugates and Legendre transforms of convex functions*, *Canad. J. Math.* **19** (1967), 200–205.
17. ———, *Convex analysis*, Princeton Univ. Press, Princeton, N.J., 1969.
18. ———, *On the maximal monotonicity of subdifferential mappings*, (to appear).
19. V. L. Šmulyan, *Sur la derivabilité de la norme dans l'espace de Banach*, *Dokl. Akad. Nauk SSSR* **27** (1940), 643–648. (Russian)

UNIVERSITY OF WASHINGTON,  
SEATTLE, WASHINGTON