# FRÉCHET DIFFERENTIABILITY OF CONVEX FUNCTIONS

BY

EDGAR ASPLUND

*University of Washington, Seattle, Wash.. U.S.A.*

A continuous convex function of one real variable is differentiable, except perhaps at a countable subset of its interval of continuity. The present paper deals with generalizations of this elementary statement to convex functions which are defined on some Banach space $E$, and continuous in the norm topology, with "differentiable" replaced either by "Fréchet differentiable" or "Gâteaux differentiable". Since for $E = L^\infty(0,1)$ the very norm function $f(x) = \|x\|$ for $x$ in $E$, which is convex and continuous on all of $E$, is nowhere even Gâteaux differentiable (Mazur [13]), this amounts to a classification of the category of all Banach spaces depending upon whether certain differentiability statements hold. Therefore we say that a Banach space is a *strong differentiability space* (SDS) if the following theorem holds for it.

THEOREM. (Strong Differentiability Theorem.) *Every continuous convex function is Fréchet differentiable on a dense $G_\delta$ subset of its domain of continuity.*

If the following statement holds for a Banach space, we call it a *weak differentiability space* (WDS):

THEOREM. (Weak Differentiability Theorem.) *Every continuous convex function is Gâteaux differentiable on a dense $G_\delta$ subset of its domain of continuity.*

Some general remarks on these definitions are in order here. First, by a continuous convex function $f$ on the Banach space $E$, we mean a function which is defined and convex on all of $E$, with values in $(-\infty, \infty]$, and finite valued and continuous at least at some point of $E$. Then the set of all points of $E$ where $f$ is finite valued and continuous is a non-empty, open, and convex subset of $E$ which we call the *domain of continuity* of $f$. It is equal to the interior of dom $f$, the *effective domain* of $f$, defined by

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

For convenience, we assume that the values of $f$ on the boundary of dom$f$ (which does not concern us here) are defined so that $f$ is lower semicontinuous on all of $E$.

A second (and obvious) remark is that every SDS is a WDS. The converse is false: Phelps [16] has shown that $l_1$ is not an SDS, but it follows from Mazur [13] (cf. Theorem 2: $l_1$ is separable) that it is a WDS. Also, it is clear that a space that is isomorphic to an SDS (WDS) is itself an SDS (WDS). Actually, if $E$ is an SDS (WDS) and $F=T(E)$ for some continuous linear transformation $T$ (alternatively: $F$ is isomorphic to a quotient space of $E$), then $F$ is also an SDS (WDS), cf. Proposition 4.

Our third remark concerns the character of the exceptional set, on which the continuous convex function $f$ does *not* have to be differentiable. In the elementary case $E=R$ this set is characterized by the word "countable", whereas in the strong and weak differentiability theorems stated it is a "meager (i.e. first Baire category) $F_\sigma$" set. "Countable" implies "meager $F_\sigma$", and the converse is false even for $E=R$ as shown by the Cantor set. Thus there is undoubtedly room for some improvement in the statements of the strong and weak differentiability theorems. On the other hand, they have the functorial property expressed by Proposition 4 and quoted above, and in the case of the strong differentiability theorem there is also the fact that if a convex function is Fréchet differentiable on a dense subset of some open set, then the set of all points in this open set where the function is Fréchet differentiable, is a $G_\delta$ set (Lemma 6).

After these general remarks we are now ready to present the main results of this paper.

THEOREM 1. *If $E$ is a Banach space which can be given an equivalent norm, such that the corresponding dual norm in $E^*$ is locally uniformly rotund, then $E$ is a strong differentiability space. In particular, every Banach space which has a separable dual space, and every reflexive Banach space admitting an equivalent Fréchet differentiable norm, is a strong differentiability space.*

THEOREM 2. *If $E$ is a Banach space which can be given an equivalent norm, such that the corresponding dual norm in $E^*$ is rotund, then $E$ is a weak differentiability space. In particular, every separable and every reflexive Banach space—in fact, every weakly compactly generated Banach space—is a weak differentiability space.*

The first parts of both these theorems are proved in Section 2, after some introductory material (which has been considerably generalized in [5]) is presented in Section 1. The second parts of the two theorems are renorming statements, which in the case of our Theorem 2 is due to Amir and Lindenstrauss [1, Thm. 3]. The second part of Theorem 1 is proved in Section 3 of this paper.

Historically, the first result of the above kind for infinite dimensional spaces was the "smoothness theorem" of Mazur [13], which states essentially that every separable Banach space is a WDS. Lindenstrauss [11] proved that a reflexive Banach space admitting an equivalent Fréchet differentiable norm is an SDS. Actually, in his formulation "dense $G_\delta$" is replaced by "dense", which in view of Lemma 6 below is no restriction. Lindenstrauss' methods are rather different from ours.

By duality, the differentiability theorems for a Banach space $E$ correspond to "variational" theorems on the dual space $E^*$. We state here just the one corresponding to the strong differentiability theorem, because it seems more interesting with regard to applications.

THEOREM 3. *If $E$ is an SDS and $F$ is a norm lower semicontinuous (but not necessarily convex) functional on the dual space $E^*$, with values in $(-\infty, \infty]$, such that*

$$\inf\{F(y) - \langle x,y\rangle : y \in E^*\} > -\infty \tag{1}$$

*for all $x$ in some open neighborhood $U$ of 0 in $E$, then there is a dense $G_\delta$ subset $G \subset U$ such that to each $x$ in $G$ there is a unique solution $y(x)$ in $E^*$ to the variational problem*

$$\inf\{F(y) - \langle x, y\rangle : y \in E^*\} = F(y(x)) - \langle x, y(x)\rangle. \tag{2}$$

*Moreover, $y(x)$ has the property that if the sequence $\{y_n\} \subset E$ satisfies*

$$\lim\{F(y_n) - \langle x, y_n\rangle\} = F(y(x)) - \langle x, y(x)\rangle$$

*then $y_n \to y(x)$ in norm. Also, $x \to y(x)$ is a norm to norm continuous function on $G$, and if $H$ is a dense subset of $U$ on which it is possible to define a norm to norm continuous function $y(x)$ satisfying (2), then $H \subset G$.*

This result is proved in Section 4, and applications are given to generalize certain kinds of variational problems in Hilbert space introduced by Moreau [14] and called "proximal mappings" by him. These include as special cases the problems of finding nearest and farthest points in closed, generally non-convex subsets of a Hilbert space. In a different setting, using rather unrelated methods, the farthest point problem has been treated by Edelstein [8], and with methods more related to those of this paper, by the author [2].

Many functionals $F$ encountered in classical variational calculus are of the form

$$F(y) = \int_a^b f(t, y(t), \dot{y}(t))\, dt,$$

where $f$ is a positive valued function, and in such case the lower semicontinuity of $F$ follows, generally speaking, from Fatou's lemma. We will not elaborate here on the significance of

Theorem 3 for the problems of classical variational calculus, but let us just say that it is a statement of an unconventional kind insofar as it does not say anything about necessary or sufficient conditions for the existence of solutions of the "individual variational problem" $\inf \{F(y): y \in E^*\} = F(y(0))$ but is a statement on a whole family of "linearly perturbed" problems (2), with $x$ as a perturbation parameter. It says then, that if the variational question is meaningful in the sense of (1) for all $x$ in some neighborhood of 0 in $E$, then for many such $x$ a solution of (2) does indeed exist, in a strong sense.

## 1. Fréchet differentiability and duality

We will use the notations and concepts introduced by Fenchel for convex functions. For statements not proved here we refer in general to Brøndsted [6] or Moreau [15].

Given a function $f$ on a Banach space $E$ (or $E^*$), with values in $(-\infty, \infty]$, we will always assume that it is lower semicontinuous in the norm topology. The function $f^*$ defined on $E^*$ by

$$f^*(y) = \sup \{\langle x, y \rangle - f(x): x \in E\} \quad \text{for all } y \text{ in } E^*,$$

is called the *dual function* of $f$. The function $f^*$ is lower semicontinuous in the weak* as well as in the norm topology, convex, and with values in $(-\infty, \infty]$ if $f \not\equiv +\infty$, which we will also assume.

If $f$ is defined on $E^*$ with values in $(-\infty, \infty]$, then $f^*$ is defined correspondingly on $E$, and in either case $f^{**} = (f^*)^*$ is the largest weak (resp. weak*) lower semicontinuous convex minorant of $f$. Hence $f^{**} = f$ in the important case when $f$ itself is convex (and, if defined on $E^*$, weak* lower semicontinuous). If $f = f^{**}$ and $g = g^{**}$, then obviously $f \leqslant g$ if and only if $f^* \geqslant g^*$.

The above concepts and statements are appropriate generalizations of counterparts in a more primitive theory of convex functions on the positive halfline. We state and prove here one result from this theory. Let $\Gamma$ denote the class of convex and lower semicontinuous (i.e.: left continuous) functions $\gamma$ from $[0, \infty)$ to $[0, \infty]$, with $\gamma(0) = 0$. The operation $\gamma \to \gamma^*$ defined by

$$\gamma^*(x) = \sup \{tx - \gamma(t): t \geqslant 0\} \quad \text{for all } x \geqslant 0$$

is an order-inverting involution on $\Gamma$. Put

$$\Gamma_U = \{\gamma \in \Gamma: \gamma(t) > 0 \text{ if } t > 0\},$$

$$\Gamma_L = \{\gamma \in \Gamma: \lim_{t \to 0} \gamma(t)/t = 0\}.$$

We then have the following lemma.

LEMMA 1. *$\gamma$ is in $\Gamma_U$ if and only if $\gamma^*$ is in $\Gamma_L$.*

*Proof.* Suppose that $\gamma$ is in $\Gamma \setminus \Gamma_U$. Then $\gamma(\varepsilon)=0$ for some $\varepsilon>0$. Define the function $\gamma_1$ in $\Gamma$ by

$$\gamma_1(x)=0 \text{ if } 0\leqslant x\leqslant \varepsilon, \quad \gamma_1(x)=+\infty \text{ if } x>\varepsilon.$$

Clearly
$$\gamma_1^*(x)=\varepsilon x \quad \text{for all } x\geqslant 0.$$

Now $\gamma\leqslant\gamma_1$, so that $\gamma^*\geqslant\gamma_1^*$. But this means that $\gamma^*$ is in $\Gamma\setminus\Gamma_L$. Hence $\gamma$ is in $\Gamma_U$ if $\gamma^*$ is in $\Gamma_L$, and the converse follows by reversing the argument.

Lemma 1 is used together with the following consequence of Hölder's inequality.

LEMMA 2. *For any $\gamma$ in $\Gamma$, the function $\gamma(\|x\|)$ defined on $E$ and the function $\gamma^*(\|y\|)$ defined on $E$ with the corresponding norm, form a pair of mutually dual functions.*

We can now state and prove the main result of this section.

PROPOSITION 1. *The following three statements about a function $f$ (norm lower semicontinuous, but not necessarily convex) and its dual function $f^*$, relative to an element $b$ of $E$ and an element $a$ of $E^*$, are equivalent.*

(i) *for some $\gamma\in\Gamma_U$, $f(x)\geqslant f(b)+\langle x-b, a\rangle+\gamma(\|x-b\|)$ for all $x\in E$.*

(ii) *for some $\gamma^*\in\Gamma_L$, $f^*(y)\leqslant f^*(a)+\langle b, y-a\rangle+\gamma^*(\|y-a\|)$ for all $y\in E^*$.*

(iii) dom $f^*=\{y\in E^*: f^*(y)<\infty\}$ *is radial at $a$; and if* $\lim\,(\langle x_n, a\rangle - f(x_n))=f^*(a)$ *then $x_n\to b$ in norm.*

*In all three cases the relation $f^*(a)=\langle b, a\rangle-f(b)$ holds. The roles of $E$ and $E^*$ in these statements may be reversed.*

*Proof.* In view of Lemma 1 and Lemma 2, it is a straightforward application of the definition of the dual function (cf. Brøndsted [6], Theorem 4.1) to check that (i) implies (ii). Applying the same argument to (ii), one gets

$$f^{**}(x)\geqslant\langle b, a\rangle-f^*(a)+\langle x-b, a\rangle+\gamma(\|x-b\|) \quad \text{for all } x\in E. \tag{3}$$

Since $f\geqslant f^{**}$, (i) would follow if one could show that $f^*(a)=\langle b, a\rangle-f(b)$. We will do so, but first we show that (iii) follows. If $\{x_n\}\subset E$ satisfies

$$f^*(a)-(\langle x_n, a\rangle-f(x_n))=\varepsilon_n, \ \lim \varepsilon_n=0$$

then
$$f^{**}(x_n)\leqslant\langle b, a\rangle-f^*(a)+\langle x_n-b, a\rangle+\varepsilon_n,$$

and hence by (3)
$$\|x_n-b\|\leqslant\gamma^{-1}(\varepsilon_n)\to 0 \text{ as } n\to\infty,$$

since $\varepsilon_n \geqslant 0$ for all $n$ and $\gamma$ is in $\Gamma_U$. This proves that (ii) implies (iii), because the first part of (iii) is a consequence of (ii) and the definition of $\Gamma_L$. Also, we have shown that, supposing that any one of (i), (ii), or (iii) holds, then there exists a sequence $\{x_n\} \subset E$ tending to $b$ in norm such that

$$\lim (\langle x_n, a\rangle - f(x_n)) = f^*(a).$$

Using the lower semicontinuity of $f$, we deduce

$$\langle b, a\rangle - f(b) \geqslant f^*(a).$$

But the opposite inequality follows by the definition of $f$, hence

$$f^*(a) = \langle b, a\rangle - f(b) \tag{4}$$

is a consequence of any one of (i), (ii), or (iii). Thus (i) follows from (ii), as asserted. It remains to prove that (iii) implies (ii). By the first statement of (iii), the set

$$D = \{y \in E^* : f^*(y) \leqslant f^*(a) + 1\}$$

is radial at $a$, i.e.

$$\bigcup_{n=1}^{\infty} n(D - a) = E.$$

Since $D$ is also closed and convex, it follows from Baire's category theorem that $D$ is in fact a neighborhood of $a$. Consider the function $\gamma^*$ defined by

$$\gamma^*(t) = \sup \{f^*(a + tz) - f^*(a) - \langle b, tz\rangle : z \in E^*, \ \|z\| = 1\}, \quad \text{for } t \geqslant 0.$$

Because $f^*$ is bounded on a neighborhood of $a$, $\gamma^*$ is bounded near $t = 0$. Furthermore, $\gamma^*(0) = 0$ and, using (4),

$$\gamma^*(t) = \sup \{f(b) + f^*(a + tz) - \langle b, a + tz\rangle : \|z\| = 1\} \geqslant 0 \quad \text{for all } t \geqslant 0.$$

As $\gamma^*$ is obviously a convex and lower semicontinuous function, it follows that $\gamma^*$ is in $\Gamma$, and, except for the fact that we have still to prove that $\gamma^* \in \Gamma_L$, (ii) holds for this function. We know that for some $\varepsilon > 0$ and $M < \infty$

$$\gamma^* \leqslant \gamma_1 + M, \text{ or, equivalently } \gamma(x) \geqslant \varepsilon x - M \text{ for all } x \geqslant 0,$$

where $\gamma_1$ is the function used in the proof of Lemma 1. Assume now that $\gamma^* \notin \Gamma_L$, i.e. that for some sequence $\{y_n\} \subset E^*$ such that $y_n \to a$ in norm, and some $\delta > 0$ we have (with $y_n \neq a$ for all $n$)

$$f(b) + f^*(y_n) - \langle b, y_n\rangle \geqslant \delta \|y_n - a\|.$$

By definition, we may then find a sequence $\{b_n\} \subset E$ such that

$$\langle b_n, y_n\rangle - f(b_n) \geqslant f^*(y_n) - \delta \|y_n - a\|/2.$$

Hence $$\langle b_n, a\rangle - f(b_n) \geqslant \langle b, a\rangle - f(b) + \langle b_n - b, a - y_n\rangle + \delta\|y_n - a\|/2.$$

If we knew that $\{b_n\}$ were bounded, this would imply a contradiction proving that $y^*$ is in $\Gamma_L$, because the above relation would then imply that $\{b_n\}$ satisfies the conditions of case (iii), and $b_n \to b$ in norm would follow. Then for some $n$ one would get $\|b_n - b\| < \delta/2$ implying

$$\langle b_n, a\rangle - f(b_n) > \langle b, a\rangle - f(b) = f^*(a)$$

which is impossible. What we know now about $y$ enables us to use (i) in the form

$$f(b_n) \geqslant f(b) + \langle b_n - b, a\rangle + \varepsilon\|b_n - b\| - M.$$

Together with the relation defining $\{b_n\}$ this yields

$$\varepsilon\|b_n - b\| \leqslant M - (f(b) + f^*(y_n) - \langle b, y_n\rangle) + \langle b_n - b, y_n - a\rangle + \delta\|y_n - a\|/2,$$

proving that $\{b_n\}$ is bounded, for

$$\|b_n - b\| \leqslant 2(M + \delta\varepsilon/4)/\varepsilon \quad \text{if } \|y_n - a\| \leqslant \varepsilon/2.$$

Here ends the proof of Proposition 1.

Statement (ii) in Proposition 1 is usually formulated as "$f^*$ is *Fréchet* (or *strongly*) *differentiable* at $a$ in $E^*$ with *Fréchet differential* $b$ in $E^*$. Correspondingly, we will reformulate (i) as "$f$ is *strongly rotund* at $b$ in $E$ with respect to $a$ in $E^*$".

Clearly, Proposition 1 contains all that is needed to deduce Theorem 3, except for its last sentence, from Theorem 1. Proposition 1 can be considered as a careful reformulation of certain older statements by Šmulyan [17], and it will also be used in the next section to derive the first part of Theorem 1. We note here the following Corollary of Proposition 1.

COROLLARY 1. *If $f^*$ defined on $E^*$ is Fréchet differentiable at $a$ with Fréchet differential $b$ in $E^{**}$, then $b$ is actually in $E$.*

*Proof.* The function $f$, initially defined only on $E$, is extended to all of $E^{**}$ by $f(x) = \infty$ if $x \in E^{**} \setminus E$, and Proposition 1 is applied to $E^{**}$ and $E^*$. Since by (iii) $b$ is the limit in norm of a sequence of elements in dom $f \subset E$, and $E$ is norm closed in $E^{**}$, it follows that $b$ is in $E$.

## 2. Proof of the differentiability statements

We assume that the convex function $f$ is defined on the Banach space $E$ and that the domain of continuity of $f$ is a non-empty open convex subset $C \subset E$. We will see that we may restrict ourselves to consider such functions $f$ that in addition satisfy a Lipschitz condition, and $C = E$. Let the convex functions $g_n$ be defined on $E$ by

$$g_n(x) = n\|x\| + n \quad \text{for all } x \text{ in } E,$$

and the functions $f_n$ by

$$f_n = f \wedge g_n \quad n = 1, 2, \ldots$$

meaning the greatest convex minorant of the functions $f$ and $g_n$. The relation between $f$ and the $f_n$ is given by this lemma.

LEMMA 3. *For each $x$ in $C$ there is a number $\delta > 0$ and an integer $N$ such that*

$$f_n(y) = f(y) \quad \text{for } \|y - x\| \leqslant \delta \text{ and } n \geqslant N.$$

*Proof.* By continuity, there is a $\delta > 0$ such that

$$f(x) - 1 \leqslant f(z) \leqslant f(x) + 1 \quad \text{for } \|z - x\| \leqslant 2\delta. \tag{5}$$

Suppose that $y$ satisfies $\|y - x\| \leqslant \delta$. There is then (at least) one exact affine minorant $\varphi$ of $f$ at $y$,

$$\varphi(z) = f(y) + \langle z - y, a \rangle \leqslant f(z) \quad \text{for all } z \text{ in } E.$$

Here $a$ is an element of $E^*$, and the continuity relation (5) together with $\|y - x\| \leqslant \delta$ shows that $\|a\| \leqslant 2/\delta$. Consider the following upper estimate for $\varphi(z)$.

$$\varphi(z) = f(y) + \langle x - y, a \rangle - f(x) + f(x) - \langle x, a \rangle + \langle z, a \rangle \leqslant f(x) + 2\|x\|/\delta + 2\|z\|/\delta.$$

Let $N$ be an integer greater than both $2/\delta$ and $f(x) + 2\|x\|/\delta$. Then

$$\varphi \leqslant g_n \quad \text{for } n \geqslant N.$$

In this case, $\varphi$ is a convex minorant of both $f$ and $g_n$, thus

$$f_n(y) \leqslant f(y) = \varphi(y) \leqslant f_n(y),$$

proving Lemma 3.

The functions $f_n$ are defined on all of $E$, and it follows from their definition that

$$\|f_n(x) - f_n(y)\| \leqslant n\|x - y\| \quad \text{for all } x \text{ and } y \text{ in } E,$$

i.e. $f_n$ satisfies a Lipschitz condition. If we can prove the first part of Theorem 1 (Theorem 2) for such functions, then we can get a sequence of dense $G_\delta$ subsets of $E$, and the intersection of all of these is by the Baire category theorem still a dense $G_\delta$ subset of $E$. Finally, take the intersection of $C$ with this last dense $G_\delta$ set; the result is a dense $G_\delta$ subset of $C$, and by Lemma 3 $f$ is Fréchet (Gâteaux) differentiable at all points of this subset. Thus there is no loss of generality to assume that $f$ is Lipschitz, hence continuous on all of $E$.

Suppose now that $\|x\|_1$, for $x$ in $E$, is an equivalent norm on $E$, whose corresponding dual norm we denote by $\|y\|_1$, for $y$ in $E^*$. By Lemma 2 we can define a pair of mutually dual functions $h$ and $h^*$ by

$$h(x)=\tfrac{1}{2}\|x\|_1^2 \text{ for all } x \text{ in } E, \quad h^*(y)=\tfrac{1}{2}\|y\|_1^2 \text{ for all } y \text{ in } E^*. \tag{6}$$

Since $f$ is Lipschitz, the open sets $F_n$ defined by

$$F_n=\left\{x: f(x)-\frac{1}{p}\left(h(p(x+y))-\frac{1}{n}\right)>\sup\left(f(z)-\frac{1}{p}h(p(z+y)): z\in E\right) \text{ for some } y\in E \text{ and } p>0\right\}$$

are dense in $E$, thus by Baire's theorem

$$F=\bigcap_{n=1}^{\infty} F_n$$

is a dense $G_\delta$ subset of $E$. We will show that under suitable hypotheses on $h$ (or rather: on $h^*$) the function $f$ is Fréchet (Gâteaux) differentiable at each point of $F$.

By translation, it suffices to prove that $f$ is Fréchet (Gâteaux) differentiable at 0, assuming this point to be in $F$. Furthermore, we may assume that

$$f(0)=0, \quad f(x)\geqslant\langle x, a\rangle \text{ for all } x \text{ in } E, \text{ with some fixed } a \text{ in } E^*.$$

By duality, this is equivalent with

$$f^*(a)=0, \quad f^*(y)\geqslant 0 \text{ for all } y \text{ in } E.$$

Considering this, the statement $0\in F$ shows that there is a sequence $\{y_n\}\subset E$ and a sequence $\{p_n\}\subset(0,\infty)$ such that

$$f(z)<\frac{1}{p_n}\left(h(p_n(z+y_n))-h(p_n y_n)+\frac{1}{n}\right) \quad \text{for all } z \text{ in } E.$$

By duality this implies, denoting $p_n y_n$ by $b_n$,

$$p_n f^*(x)\geqslant-\langle b_n, x\rangle+h^*(x)+h(b_n)-\frac{1}{n} \quad \text{for all } x \text{ in } E^*.$$

In particular, for $x=a$,

$$0\geqslant-\langle b_n, a\rangle+h^*(a)+h(b_n)-\frac{1}{n}.$$

Adding the last two inequalities to the inequality

$$0\geqslant 2(\langle b_n,(x+a)/2\rangle-h(b_n)-h^*((x+a)/2))$$

which derives from the definition of a dual function, we get

$$p_n f^*(x) \geq h^*(x) + h^*(a) - 2h^*((x+a)/2) - \frac{2}{n} \quad \text{for all } x \text{ in } E^*. \tag{7}$$

Up till now we have not used any special properties of the equivalent norm $\|y\|_1$ on $E^*$. Suppose now that this norm is rotund. Then it follows from (6) and (7) that if $x \neq a$, $f^*(x) > 0$. But for $f$ *not* to be Gâteaux differentiable at 0 in $E$ means that for some $x \neq a$,

$$f(z) \geq \langle z, x \rangle \quad \text{for all } z \text{ in } E$$

which by duality implies $f^*(x) = 0$. Hence we have proved the first part of Theorem 2.

In the corresponding part of Theorem 1 we may assume that $\|y\|_1$ is a locally uniformly rotund norm on $E^*$. We will see that the conditions of case (iii) in Proposition 1 are satisfied, with $f$, $E$ interchanged with $f^*$, $E^*$ respectively—a permissible arrangement since $f$ is convex—and with 0 in $E$, $a$ in $E^*$ instead of $a$ in $E^*$, $b$ in $E$ respectively. Thus, suppose that

$$\lim f^*(x_k) = -f(0) = 0$$

for some sequence $\{x_k\} \subset E^*$. Given $\varepsilon > 0$, choose first $n$ so large that $2/n < \varepsilon/2$, and afterwards choose $N$ so that

$$f^*(x_k) \leq \varepsilon/2p_n \quad \text{if } k \geq N$$

Together with (6) and (7) this shows that

$$\tfrac{1}{2}\|x_k\|_1^2 + \tfrac{1}{2}\|a\|_1^2 - \|(x_k + a)/2\|_1^2 < \varepsilon \quad \text{if } k \geq N.$$

Since the equivalent norm is locally uniformly rotund it follows in a straightforward way—the details are in [3]—that $x_n \to a$ in norm. This, by Proposition 1, proves that $f$ is Fréchet differentiable at 0 in $E$, because dom $f = E$ as $f$ is supposed to be Lipschitz. Thus the first part of Theorem 1 is established.

## 3. Proof of the renorming statements of Theorem 1

The following result was communicated orally by Professor R. R. Phelps to the author. With his permission, we reproduce it here.

PROPOSITION 2. *Let $E$ be a reflexive Banach space which has an equivalent Fréchet differentiable norm. Then there exists another equivalent norm for $E$, such that the corresponding dual norm on $E^*$ is locally uniformly rotund.*

*Proof.* Let $\|x\|_1$ be the equivalent, Fréchet differentiable norm on $E$ and define $g$ by

$$g(x) = \tfrac{1}{2}\|x\|_1^2 \quad \text{for all } x \text{ in } E. \tag{8}$$

By the deep Theorem 1 of Lindenstrauss [12], there is a continuous linear injection $T$ of $E^*$ into $c_0(I)$, where $I$ is some index set, uncountable if $E$ is nonseparable. On $c_0(I)$, Day [7] has constructed a functional $p$, which is well known to be an equivalent, locally uniformly rotund norm. We now define the convex function $h^*$ on $E^*$ by

$$h^*(y)=g^*(y)+p^2(Ty) \quad \text{for all } y \text{ in } E^*, \tag{9}$$

and we will show that $\sqrt{2h^*(y)}$ is an equivalent, locally uniformly rotund norm for $E^*$. This proves Proposition 2, for $\sqrt{2h^*(y)}$ is the dual norm corresponding to the equivalent norm $\sqrt{2h(x)}$ on $E$, where $h=(h^*)^*$. It is evident that $\sqrt{2h^*(y)}$ is norm equivalent, since $T$ is continuous. Therefore, it remains to show that if $a$ is a fixed element of $E^*$ and $\{x_n\}\subset E$ is such that

$$\lim \{h^*(x_n)+h^*(a)-2h^*((x_n+a)/2)\}=0 \tag{10}$$

then $x_n \to a$ in norm.

Because of (9) and convexity, the relation (10) splits into two:

$$\lim \{g^*(x_n)+g^*(a)-2g^*((x_n+a)/2)\}=0 \tag{11}$$

and

$$\lim \{p^2(Tx_n)+p^2(Ta)-2p^2((Tx_n+Ta)/2)\}=0. \tag{12}$$

In fact, the contents of each curly bracket are nonnegative. Since $p$ is locally uniformly rotund, (12) implies that $Tx_n\to Ta$ in the norm of $c_0(I)$. As a preliminary, we claim that $x_n\to a$ in the weak topology of $E^*$. Since (10) shows that $\{x_n\}$ is bounded, this amounts to showing that if a subsequence is weakly convergent to an element $b$ in $E^*$, then $b=a$. Reindexing if necessary, we may suppose that $x_n\to b$, weakly. By the continuity of $T$ one gets $Tb=Ta$, hence $b=a$ from the fact that $T$ is one to one.

Having established the weak convergence of $x_n$ to $a$, we now use relation (11) together with the strong rotundity of $g^*$ at $a$ to deduce that

$$\limsup \{2g^*(a)+\langle c, x_n-a\rangle+y(\|x_n-a\|)-2g^*((x_n+a)/2)\}\leqslant 0$$

with some $c$ in $E$ (the existence of $c$ follows from reflexivity) and some $\gamma$ in $\Gamma_U$. Now $g^*$ is, by its definition, weakly lower semicontinuous, therefore

$$\lim \gamma(\|x_n-a\|)\leqslant 0$$

follows, and from it that $x_n\to a$ in norm, since $\gamma$ is in $\Gamma_U$. Proposition 2 is proved.

To complete the proof of Theorem 1 it remains to prove the following statement.

PROPOSITION 3. *Let $E$ be a separable Banach space whose dual space $E^*$ is also separable Then there exists an equivalent norm for $E$ such that the corresponding dual norm on $E^*$ is locally uniformly rotund.*

*Proof.* We will use the following lemma, proved by Klee [10] as a quite nontrivial extension of results of Kadeč [9] that are in themselves rather deep.

LEMMA 4. (Klee [10], Corollary 1.5.) *If $E$ is as in Proposition 3, then there exists in equivalent norm $\|x\|_1$ for $x$ in $E$, such that the corresponding dual norm $\|y\|_1$ for $y$ in $E^*$ is strictly convex and has the property that $\{y_n\} \subset E^*$, $y \in E^*$, $\|y_n\| \to \|y\|$, and $y_n \to y$ in the weak* topology implies $y_n \to y$ in norm.*

We use Klee's norm to define $g$ by (8) and $h^*$ by (9). Note that in this case there is an elementary continuous linear injection $T$ of $E^*$ into $c_0$, given by

$$Ty = \{\langle x_n, y\rangle\},$$

where $\{x_n\}$ is a sequence with a dense span in $E$, such that $\|x_n\| \to 0$. Clearly, $T$ is also continuous from the weak* toplogy of $E^*$ to the weak topology of $c_0$, for if $\alpha = \{\alpha_n\}$ is a continuous linear functional on $c_0$ (i.e. $\Sigma |\alpha_n| < \infty$) then

$$Ty = \langle \Sigma \alpha_n x_n, y\rangle$$

is a weak* continuous linear functional on $E^*$. Therefore, $h^*$ is weak* lower semicontinuous, so that it is the dual of the function $h = h^{**}$, defined on $E$ (this was obvious in the case of Proposition 2).

Suppose now that a sequence $\{x_n\} \subset E^*$ satisfies (10). Then (11) and (12) follows, and as before that $x_n \to a$, this time in the weak* topology on $E^*$. But then we can apply Lemma 4, because (11) gives

$$(1/4)\ (\|x_n\|_1 - \|a\|_1)^2 \leq g^*(x_n) + g^*(a) - 2g^*((x_n + a)/2) \to 0$$

in view of the fact that

$$g^*(y) = \tfrac{1}{2}\|y\|_1^2 \quad \text{for all } y \text{ in } E^*.$$

Hence $x_n \to a$ in Klee's norm, and of course in any equivalent norm on $E^*$. Thus the norm $\|x\|_2 = \sqrt{2h(x)}$ defined for $x$ in $E$ satisfies the conditions of Proposition 3.

## 4. Variational results

We begin this section by completing the proof of Theorem 3. As remarked before, Theorem 1 and Proposition 1 together prove all except the last paragraph of Theorem 3. Throughout the proof we suppose that $f$ is a convex function with a non-empty domain of continuity $C \subset E$, and that $G \subset C$ is the set of all points at which $f$ is Fréchet differentiable.

The Fréchet differential (an element of $E^*$) of $f$ at $x$ in $G$ will be denoted by $Df(x)$. The following three lemmas complete the proof of Theorem 3 and clarify other parts of the situation as well.

LEMMA 5. *The mapping $x \to Df(x)$ of $G$ into $E^*$ is norm to norm continuous.*

LEMMA 6. *If $G$ is dense in $C$, then $G$ is a dense $G_\delta$ set.*

LEMMA 7. *If it is possible to define a norm to norm continuous function $x \to y(x)$ from a dense subset $H \subset C$ into $E^*$, satisfying*

$$f(x) + f^*(y(x)) = \langle x, y(x) \rangle \quad \textit{for all } x \textit{ in } H,$$

*then $H \subset G$ and $y(x) = Df(x)$ for all $x$ in $H$.*

*Remarks.* From the nature of the proofs below it is clear that the roles of $f$, $E$ and $f^*$, $E^*$ in these three lemmas can be reversed. Lemma 5 is a special case of [5, Thm. 4].

*Proof of Lemma 5.* Let the sequence $\{x_n\} \subset G$ converge in norm to $x$ in $G$. Because of the relation

$$f(u) \geqslant f(x_n) + \langle u - x_n, Df(x_n) \rangle \quad \text{for all } u \text{ in } E$$

and the continuity of $f$ on $C$, the sequence $\{Df(x_n)\} \subset E$ must be norm bounded. By the Fréchet differentiability of $f$ at $x$, there is a $\gamma$ in $\Gamma_L$ such that

$$f(x_n) \geqslant f(x) + \langle x_n - x, Df(x) \rangle + \gamma(\|x_n - x\|).$$

Also, the definition of Fréchet differentiability (cf. the remark after the proof of Proposition 1) and that of dual function show that, for any $u$ in $G$,

$$f(u) = \langle u, Df(u) \rangle - f^*(Df(u)) \geqslant \langle u, Df(z) \rangle - f^*(Df(z)) \quad \text{for all } z \text{ in } G.$$

Take the last two relations and combine them:

$$f(x) \geqslant \langle x, Df(x_n) \rangle - f^*(Df(x_n)) \geqslant \langle x - x_n, Df(x_n) - Df(x) \rangle + f(x) + \gamma(\|x_n - x\|).$$

The convergence $x_n \to x$ and the boundedness of $\{Df(x_n)\}$ now shows that the limit of the second part above as $n \to \infty$ is $f(x)$. Thus we may apply Proposition 1 to $f^*$ and $f = f^{**}$, concluding that $Df(x_n) \to Df(x)$ in norm.

*Proof of Lemma 6.* Let $G_n$ be the set of all $x$ in $C$, such that for some $\delta > 0$, the ball $B(x, \delta)$ with center at $x$ and radius $\delta$ satisfies the relation

$$\sup\left\{ \|Df(y) - Df(z)\| : y, z \in G \cap B(x, \delta) \right\} < \frac{1}{n}.$$

By Lemma 5,
$$G \subset \bigcap_{n=1}^{\infty} G_n = F,$$

and so it suffices to prove that $f$ is Fréchet differentiable for each $x$ in $F$, for the sets $G_n$ are obviously open. Since $E^*$ is complete, there is for each each $x$ in $F$ a unique element $Df(x)$ in $E^*$ such that for each $n$ there is a $\delta > 0$ with the property

$$\|Df(y) - Df(x)\| \leqslant \frac{1}{n} \quad \text{for all } y \text{ in } G \cap B(x, \delta).$$

Consequently

$$f(y) \leqslant f(x) - \langle x - y, \; Df(y)\rangle \leqslant f(x) + \langle y - x, \; Df(x)\rangle + \frac{1}{n}\|y - x\|$$

for all $y$ in $G \cap B(x, \delta)$. This proves that $f$ is Fréchet differentiable at $x$ with Fréchet differential $Df(x)$, since $G$ is dense in $C$ by hypothesis.

*Proof of Lemma 7.* The following relation holds for all $x$, $z$ in $H$:

$$f(z) \leqslant f(x) - \langle x - z, y(z)\rangle = f(x) + \langle z - x, y(x)\rangle + \langle z - x, y(z) - y(x)\rangle.$$

Let $x$ be fixed and $z$ vary over $H$. There exists then a non-decreasing function $\varepsilon(t) > 0$ for $t > 0$, finite valued for at least some $t > 0$, and with right limit zero at 0, such that

$$y(z) - y(x) \leqslant \varepsilon(\|z - x\|) \quad \text{for all } z \text{ in } H.$$

Let $\gamma(t)$ be the largest minorant in $\Gamma$ of the function $t\varepsilon(t)$. Obviously, $\gamma$ is in $\Gamma_L$ and

$$f(z) \leqslant f(x) + \langle z - x, y(x)\rangle + \gamma(\|z - x\|) \quad \text{for all } z \text{ in } H.$$

But $H$ is dense in $C$, hence the above relation holds by continuity for all $z$ in $C$, proving that $f$ is Fréchet differentiable at $x$ with Fréchet differential $y(x)$.

The remainder of this section will be devoted to applying Theorem 3 to generalize Moreau's concept of proximal mapping in a Hilbert space. Thus $E$ will be a real Hilbert space, and by "misuse of language" we put $E^* = E$, with $\langle x, y\rangle$ denoting the inner product of $x$ and $y$ in $E$. Define $h$ by

$$h(x) = \tfrac{1}{2}\langle x, x\rangle = \tfrac{1}{2}\|x\|^2 \quad \text{for all } x \text{ in } E.$$

The function $h$ is convex, continuous, and Fréchet differentiable, and $Dh(x) = x$ for all $x$ in $E$. Now let $g$ be any $(-\infty, \infty]$-valued, lower semicontinuous functional defined on $E$, such that

$$F = h + g$$

satisfies condition (1) of Theorem 3. The resulting solution $y(x)$ to the variational problem (2) is then denoted by

$$\operatorname{prox}_g(x) = y(x) \ (= DF^*(x)) \quad \text{for all } x \text{ in } G,$$

and (2) can be rewritten

$$\tfrac{1}{2}\|\operatorname{prox}_g(x) - x\|^2 + g(\operatorname{prox}_g(x)) = \inf\{\tfrac{1}{2}\|y-x\|^2 + g(y) : y \in E\}. \tag{13}$$

This corresponds to the definition given by Moreau in the case when $g$ is a *convex* function. In that case, the mapping $\operatorname{prox}_g$ is defined on all of $E$ and is a contraction, *a fortiori* continuous. In our more general case, it need not be possible to define $\operatorname{prox}_g$ continuously on all of $E$, and if it is, $\operatorname{prox}_g$ does not have to be a contraction. To illustrate the last assertion, put $g = -h/2$; then $F^* = 2h$ so that $\operatorname{prox}_g(x) = 2x$ for all $x$ in $E$.

To give an example of an essentially non-continuous proximal mapping, let $K$ be a norm closed, non-convex subset of $E$ and define $g$ by

$$g(x) = 0 \text{ if } x \text{ is in } K, \quad g(x) = \infty \text{ otherwise.}$$

Since $K$ is closed, $g$ is lower semicontinuous. From (13) we see that $\operatorname{prox}_g(x)$ is in this case the nearest point in $K$ to $x$ in $G$, i.e. $\operatorname{prox}_g$ is just the metric projection from $G$ onto $K$. By Theorem 3, $G$ is a dense $G_\delta$ subset of the entire Hilbert space $E$, and by Lemma 7, $K \subset G$, since the projection is obviously defined and norm to norm continuous on $K$. But if it is possible to define the projection continuously on all of $E$, then by [4], $K$ must be convex, even if continuity is interpreted in the sense "from norm to weak topology".

Finally, we give an example of a proximal mapping which strangely enough turns out to define an antiprojection, i.e. a mapping into farthest points of a given set. Suppose that $K$ is norm closed and bounded in $E$, and define $g$ by

$$g(x) = -\|x\|^2 \text{ if } -x \text{ is in } K, \quad g(x) = \infty \text{ otherwise.}$$

Rewriting (13) for this case, we get that $-\operatorname{prox}_g(x) \in K$ and

$$\tfrac{1}{2}\|-\operatorname{prox}_g(x) - x\|^2 = \sup\{\tfrac{1}{2}\|y-x\|^2 : y \in K\},$$

i.e. $x \to -\operatorname{prox}_g(x)$ is the mapping that takes each point $x$ in $G$ into the unique point in $K$ that lies farthest from $x$. It also follows from Theorem 3 that this mapping is norm to norm continuous, and that if

$$\{y_n\} \subset K \quad \text{and} \quad \lim \|y_n - x\| = \sup\{\|y-x\| : y \in K\},$$

then $y_n \to \operatorname{prox}_g(x)$ in norm. One is wont to think of such properties as characteristic of nearest point mappings onto closed convex sets. We have here shown that they hold much more generally, provided one excludes some meager $F_\sigma$ set depending on the case.

## 5. Additional remarks

The following property of strong (weak) differentiability spaces was mentioned in the introduction.

PROPOSITION 4. *If $E$ is an* SDS (WDS) *and $T$ is a continuous linear mapping of $E$ onto the Banach space $F$, then $F$ is also an* SDS (WDS).

*Proof.* Let $f$ be a convex function with domain of continuity $C \subset F$. Then $fT$ is convex on $E$ with domain of continuity $T^{-1}(C)$, hence by hypothesis Fréchet (Gâteaux) differentiable on some dense $G_\delta$ subset of $T^{-1}(C)$ which we may without risk denote by $T^{-1}(G)$, with some $G \subset C$, for $fT$ is constant on each coset $x + T^{-1}(0)$. By the open mapping theorem, $G$ is a dense $G_\delta$ subset of $C$, and it remains to show that $f$ is Fréchet (Gâteaux) differentiable on $G$. Given $x$ in $T^{-1}(G)$ we have in the Fréchet case

$$fT(z) \leqslant fT(x) + \langle z-x, a\rangle + \gamma(\|z-x\|) \quad \text{for all } z \text{ in } E, \tag{14}$$

with some $\gamma$ in $\Gamma_L$. Since $fT(z) = fT(x)$ if $z \in x + T^{-1}(0)$, it follows that $a$ annihilates $T^{-1}(0)$, hence $a = T^*b$ for some $b$ in $F$. The operator $T$ can be regarded as an isomorphism of $E/T^{-1}(0)$ onto $F$ in which sense it has an inverse $T^{-1}$ with norm $\|T^{-1}\| < \infty$. From (14) follows

$$f(Tz) \leqslant f(Tx) + \langle Tz - Tx, b\rangle + \gamma(\|T^{-1}\| \, \|Tz - Tx\|) \quad \text{for all } z \text{ in } E,$$

proving the Fréchet differentiability of $f$ at $Tx \in G$, since $T$ is onto. The completion of the proof in the case of Gâteaux differentiability is obvious.

Our second remark concerns an unpublished result of Errett Bishop, arrived at by methods quite different from those in the present paper.

PROPOSITION 5. (Bishop.) *If $E$ is a Banach space with a separable dual space, and $K$ is a weak* compact convex subset of $E^*$, then $K$ is the weak* closed convex hull of those of its points that are strongly exposed by functionals from $E$.*

*Proof.* Let $F$ be defined by

$$F(x) = 0 \text{ if } x \text{ is in } K, \quad F(x) = \infty \text{ otherwise.}$$

Since $F$ is convex and weak* lower semicontinuous, $F = (F^*)^*$, with $F^*$ defined on $E$. Because $K$ is bounded Theorem 3 shows that there is a dense $G_\delta$ set $G \subset E$ such that for each $x$ in $G$ there is a $y(x)$ in $K \subset E^*$ with the property that if a sequence $\{y_n\} \subset K$ satisfies

$$\lim \langle x, y_n\rangle = \langle x, y(x)\rangle$$

then $y_n \to y(x)$ in norm, i.e. $y(x)$ in $K$ is "strongly exposed" by $x$ in $E$.

Since $G$ is dense in $E$, it follows that

$$F^*(z) = \sup \{\langle z, y(x)\rangle : x \in G\} \quad \text{for all } z \text{ in } E.$$

The dual of this statement is that $K$ is the weak* closed convex hull of the set $\{y(x): x \in G\}$. That completes the proof of Bishop's result, which is of course also valid if $E$ is any strong differentiability space.

## References


[1]. Amir, D. & Lindenstrauss, J., The structure of weakly compact sets in Banach spaces. *Ann of Math.* To appear.

[2]. Asplund, E., Farthest points in reflexive locally uniformly rotund Banach spaces. *Israel J. Math.*, 4 (1966), 213–216.

[3]. —— Averaged norms. *Israel J. Math.*, 5 (1967), 227–233.

[4]. —— Chebyshev sets in Hilbert space. To appear.

[5]. Asplund, E. & Rockafellar, R. T., Gradients of convex functions. *Trans. Amer. Math. Soc.* To appear.

[6]. Brøndsted, A., Conjugate convex functions in topological vector spaces. *Mat.-Fys. Medd. Danska Vid. Selsk.*, 34 (1964), no. 2.

[7]. Day, M. M., Strict convexity and smoothness of normed spaces. *Trans. Amer. Math. Soc.*, 78 (1955), 516–528.

[8]. Edelstein, M., Farthest points of sets in uniformly convex Banach spaces. *Israel J. Math.*, 4 (1966), 171–176.

[9]. Kadeč, M. I., On weak and norm convergence. [Russian.] *Dokl. Akad. Nauk SSSR (N.S.)*, 122 (1958), 13–16.

[10]. Klee, V., Mappings into normed linear spaces. *Fund. Math.*, 49 (1960), 25–34.

[11]. Lindenstrauss, J., On operators which attain their norm. *Israel J. Math.*, 1 (1963), 139–148.

[12]. —— On nonseparable reflexive Banach spaces. *Bull. Amer. Math. Soc.*, 72 (1966), 967–970.

[13]. Mazur, S., Über konvexe Mengen in linearen normierten Räumen. *Studia Math.*, 4 (1933), 70–84.

[14]. Moreau, J. J., Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93 (1965), 273–299.

[15]. —— Fonctionnelles convexes. Multilith notes, *Collège de France* 1966–1967, p. 108.

[16]. Phelps, R. R., Representation theorems for bounded convex sets. *Proc. Amer. Math. Soc.* 11 (1960), 976–983.

[17]. Šmulyan, V. L., Sur la dérivabilité de la norme dans l'espace de Banach. *Dokl. Akad. Nauk SSSR (N.S.)* 27 (1940), 643–648.


*Received May 13, 1967*