Lecture Notes in Mathematics

Edited by A. Dold and B. Eckmann

1364

Robert R. Phelps

Convex Functions, Monotone Operators and Differentiability

Springer-Verlag

Lecture Notes in Mathematics

Edited by A. Dold and B. Eckmann

1364

Robert R. Phelps

Convex Functions, Monotone Operators and Differentiability

Springer-Verlag
Berlin Heidelberg New York London Paris Tokyo

Author

Robert R. Phelps
Department of Mathematics GN-50, University of Washington
Seattle, WA 98195, USA

Mathematics Subject Classification (1980): 46B20, 46B22, 47H05, 49A29, 49A51, 52A07

ISBN 3-540-50735-3 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-50735-3 Springer-Verlag New York Berlin Heidelberg

Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr.
2146/3140-543210

PREFACE

These notes had their genesis in a widely distributed but unpublished set of notes *Differentiability of convex functions on Banach spaces* which I wrote in 1977-78 for a graduate course at University College London (UCL). Those notes were largely incorporated into J. Giles' 1982 Pitman Lecture Notes *Convex analysis with application to differentiation of convex functions*. In the course of doing so, he reorganized the material somewhat and took advantage of any simpler proofs available at that time. I have not hesitated to return the compliment by using a few of those improvements. At my invitation, R. Bourgin has also incorporated material from the UCL notes in his extremely comprehensive 1983 Springer Lecture Notes *Geometric aspects of convex sets with the Radon-Nikodym property*. The present notes do not overlap too greatly with theirs, partly because of a substantially changed emphasis and partly because I am able to use results or proofs that have come to light since 1983.

Except for some subsequent revisions and modest additions, this material was covered in a graduate course at the University of Washington in Winter Quarter of 1988. The students in my class all had a good background in functional analysis, but there is not a great deal needed to read these notes, since they are largely self-contained; in particular, no background in convex functions is required. The main tool is the separation theorem (a.k.a. the Hahn-Banach theorem); like the standard advice given in mountaineering classes (concerning the all-important bowline for tying oneself into the end of the climbing rope), you should be able to employ it using only one hand while standing blindfolded in a cold shower.

These notes have been influenced very considerably by frequent conversations with Isaac Namioka (who has an almost notorious instinct for simplifying proofs) as well as occasional conversations with Terry Rockafellar; I am grateful to them both. I am also grateful to Jon Borwein, Marian Fabian and Simon Fitzpatrick, each of whom sent me useful suggestions based on a preliminary version.

Robert R. Phelps
October 5, 1988
Seattle, Washington

Production note: I typed these notes on a Macintosh using MacWrite 4.5 (with the *Princeton* Font 2.0) for the text and MacPaint for the drawings. The non-mathematical portions (such as the present page) were done in the *New York* font and all of it was printed on an Apple LaserWriter II.

INTRODUCTION

The study of the differentiability properties of convex functions on infinite dimensional spaces has continued on and off for over fifty years. There are a couple of obvious reaons for this. Aside from the intrinsic interest of investigating the many consequences implicit in something as simple as convexity, there is the satisfaction (for this author, at least) in discovering that a number of apparently disparate mathematical topics (extreme points - rather, strongly exposed points - of noncompact convex sets, monotone operators, perturbed optimization of real-valued functions, differentiability of vector-valued measures) are in fact closely intertwined, with differentiability of convex functions forming a common thread.

Starting in Section 1 with the definition of convex functions and a fundamental differentiability property in the one-dimensional case [right-hand and left-hand derivatives always exist], we get quickly to the first infinite dimensional result, Mazur's intriguing 1933 theorem: A continuous convex function on a separable Banach space has a dense G_δ set of points where it is (Gateaux) differentiable. In order to go beyond Mazur's theorem, some time is spent in studying the subdifferential of a convex function f; this is a set-valued map from the space to its dual whose image at each point x consists of all plausible candidates for the derivative of f at x. [The function f is Gateaux differentiable precisely when the subdifferential is single-valued, and it is Fréchet differentiable precisely when its subdifferential is single-valued and norm-to-norm continuous.]

Since a subdifferential is a special case of a monotone operator, Section 2 starts with a detailed look at monotone operators. These objects are of independent origin, having been extensively studied in the sixties and early seventies by numerous mathematicians (with major contributions from H. Brezis, F. Browder and G. J. Minty) in connection with nonlinear partial differential equations and other aspects of nonlinear analysis. (See, for instance, [Bre] or [Pa-Sb]). Also in the sixties, an in-depth study of monotone operators in fairly general spaces was carried out by R. T. Rockafellar, who established a number of fundamental properties, such as their local boundedness. He also gave an elegant characterization of those monotone operators which are the subdifferentials of convex functions, a theorem which is much easier to state than to prove (and which is not proved in full generality until Section 3). [The connection between monotone operators and derivatives of convex functions is readily apparent on the real line, since monotone operators coincide in that case with monotone nondecreasing functions, as do the right-hand derivatives of convex functions of one variable.]

In 1968, E. Asplund extended Mazur's theorem in two ways: He found more general spaces in which the same conclusion holds, and he studied a less general class of Banach spaces (now called Asplund spaces) in which a stronger conclusion holds. (Namely, he replaced the Gateaux derivative by the stronger Fréchet derivative.) Asplund used an ingenious combination of analytic and geometric techniques to prove some of the basic theorems in the subject. Roughly ten years later, P. Kenderov (as well as R. Robert and S. Fitzpatrick) proved some general continuity theorems for monotone operators which, when applied to subdifferentials, yield Asplund's results as special cases. In Section 2 we follow this approach, incorporating recent work by D. Preiss and L. Zajicek to obtain the major differentiability theorems.

The results of Section 2 all involve continuous convex functions defined on open convex sets. For many applications, it is more suitable to consider lower semicontinuous convex functions, even those which are extended real valued (possibly equal to $+\infty$). (For instance, in many optimization problems one finds just such a function in the form of the supremum of an infinite family of affine continuous functions.) Lower semicontinuous convex functions also yield a natural way to translate results about closed convex sets into results about convex functions and vice versa. (For instance, the set of points on or above the graph of such a convex function - its epigraph - forms a closed convex set). In Section 3 one will find some classical results (various versions and extensions of the Bishop-Phelps theorems) which, among other things, guarantee that subdifferentials still exist for lower semicontinuous convex functions. A nonconvex version of this type of theorem is I. Ekeland's variational principle, which asserts that a lower semicontinuous function which nearly attains its minimum at a point x admits arbitrarily small perturbations (by translates of the norm) which do attain a minimum, at points near x. This result, while simple to state and prove, has been shown by Ekeland [Ek] to have an extraordinarily wide variety of applications, in areas such as optimization, mathematical programming, control theory, nonlinear semigroups and global analysis.

In Section 4, a variational principle is established which uses differentiable perturbations; this recent result is due to J. Borwein and D. Preiss. Some deep theorems about differentiability of convex functions fall out as fairly easy corollaries, and it is reasonable to expect future useful applications.

Section 5 describes the duality between Asplund spaces and spaces with the Radon-Nikodym property (RNP). These are Banach spaces for which a Radon-Nikodym-type differentiation theorem is valid for vector measures with values in the space. Spaces with the RNP have an interesting history, starting in the late sixties with the introduction by M. Rieffel of a geometric property (dentability) which turned out to characterize the RNP and which has led to a number of other characterizations in terms of the extreme points (or strongly exposed points) of bounded closed convex subsets of the space. A truly beautiful result in this area is the fact that a Banach space is an Asplund space if and only if its dual has the RNP. (Superb expositions of the RNP may be found in the books by J. Diestel and J. J. Uhl [Di-U] and R. Bourgin [Bou].) In Section 5, the RNP is defined in terms of dentability, and a number of basic results are obtained using more recent (and simpler) proofs than are used in the above monographs. One will also find there J. Bourgain's proof of C. Stegall's perturbed optimization theorem for semicontinuous functions on spaces with the RNP; this yields as a corollary the theorem that in such spaces every bounded closed convex set is the closed convex hull of its strongly exposed points.

The notion of perturbed optimization has been moving closer to center stage, since it not only provides a more general format for stating previously known theorems, but also permits the formulation of more general results. The idea is simple: One starts with a real-valued function f which is, say, lower semicontinuous and bounded below on a nice set, and shows that there exist arbitrarily small perturbations g such that $f + g$ attains a minimum on the set. The perturbations g might be restrictions of continuous linear functionals of small norm, or perhaps Lipschitz functions of small Lipschitz norm. Moreover, for really nice sets, the perturbed function attains a strong minimum: Every mimimizing sequence converges.

The brief Section 6 is devoted to the class of Banach spaces in which every continuous convex function is Gateaux differentiable in a dense set of points (dropping the previous condition that the set need be a G_δ). Some evidence is presented that this is perhaps the "right" class to study.

Even more general than monotone operators is a class of set valued maps (from a metric space, say, to a dual Banach space) which are upper semicontinuous and take on weak* compact convex values, the so-called usco maps. In Section 7, some interesting connections between monotone operators and usco maps are described, culminating in a topological proof of one of P. Kenderov's continuity theorems.

CONTENTS

1. Convex functions on real Banach spaces.

The letter E will always denote a real Banach space, D will be a nonempty open convex subset of E and f will be a convex function on D. That is, $f:D \to R$ satisfies

$$f[tx + (1 - t)y] \leq tf(x) + (1 - t)f(y)$$

whenever $x, y \in D$ and $0 < t < 1$. If equality always holds, f is said to be affine. A function $f:D \to R$ is said to be concave if $-f$ is convex. We will be studying the differentiability properties of such functions, assuming, in the beginning, that they are continuous.

1.1 Examples.

(a) The norm function $f(x) = \|x\|$ is an obvious example. More generally, if C is a nonempty convex subset of E, then the distance function

$$d_C(x) = \inf\{\|x - y\|: y \in C\}, \qquad x \in E,$$

is continuous and convex on $D = E$. (Note that $d_C(x) = \|x\|$ if $C = \{0\}$.)

(b) The supremum of any family of convex functions is convex on the set where it is finite. In particular, if A is a nonempty bounded subset of E, then the farthest distance function $x \to \sup\{\|x - y\|: y \in A\}$ is continuous and convex on $D = E$.

(c) The norm function is also generalized by sublinear functionals, that is, functions $p:E \to R$ which satisfy

$$p(x + y) \leq p(x) + p(y) \text{ and } p(tx) = tp(x) \text{ whenever } t \geq 0.$$

Obviously, the supremum of a finite family of linear functionals is sublinear. A sublinear functional p is continuous if and only if there exists $M > 0$ such that $p(x) \leq M\|x\|$ for all x.

(d) The Minkowski gauge functional is another generalization of the norm function: Suppose that C is a convex subset of E, with $0 \in \text{int } C$. Define

$$p_C(x) = \inf\{ \lambda > 0: x \in \lambda C\}, \qquad x \in E.$$

The functional p_C is sublinear and nonnegative. Moreover, $p_C(x) = 0$ if and only if $R^+x \subset C$, and $\text{bdry } C = \{x: p_C(x) = 1\}$; in fact

$$\text{int } C = \{x: p_C(x) < 1\} \subset C \subset \{x: p_C(x) \leq 1\} = \overline{C}.$$

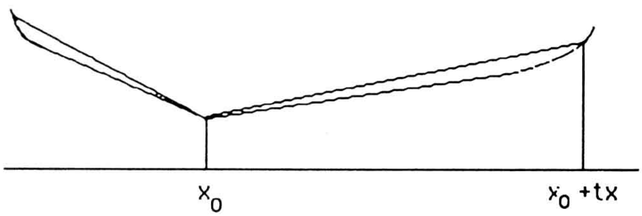
There exists $M > 0$ such that $p_C(x) \leq M\|x\|$ for all x (take $M = 1/r$, where the ball of radius r centered at 0 is contained in C), hence p_C is necessarily continuous. Conversely, any positive-homogeneous, subadditive, nonnegative and continuous functional p on E is of the form p_C, simply take $C = \{x: p(x) \leq 1\}$. These functionals fail to be seminorms if and only if C is not symmetric with respect to 0, that is, if and only if there exists x with $p_C(x) \neq p_C(-x)$.

1.2 Lemma. If $x_0 \in D$, then for each $x \in E$ the "right hand" directional derivative

(1) $$d^+f(x_0)(x) = \lim_{t \to 0+} \frac{f(x_0 + tx) - f(x_0)}{t}$$

exists and defines a sublinear functional on E.

Proof. Note that since D is open, $f(x_0 + tx)$ is defined for sufficiently small $t > 0$. The picture below shows why $d^+f(x_0)$ exists; the difference quotient is nonincreasing as $t \to 0^+$, and bounded below, by the corresponding difference quotient from the left.



To prove this, we can assume that $x_0 = 0$ and $f(x_0) = 0$. If $0 < t < s$, then by convexity

$$f(tx) \leq \frac{t}{s} f(sx) + \frac{(s - t)}{s} f(0) = \frac{t}{s} f(sx),$$

which proves monotonicity. Applying this to $-x$ in place of x, we see that

$$-[f(x_0 - tx) - f(x_0)]/t$$

is non<u>decreasing</u> as $t \to 0^+$. Moreover, by convexity again, for $t > 0$

$$2f(x_0) \leq f(x_0 - 2tx) + f(x_0 + 2tx), \qquad \text{so that}$$

$$\frac{-[f(x_0 - 2tx) - f(x_0)]}{2t} \leq \frac{[f(x_0 + 2tx) - f(x_0)]}{2t}$$

which shows that the right side is bounded below and the left is bounded above. Thus, both limits exist; the left one is $-d^+f(x_0)(-x)$ and we obviously have

$$-d^+f(x_0)(-x) \le d^+f(x_0)(x).$$

It is also obvious that $d^+f(x)$ is positively homogeneous. To see that it is subadditive, use convexity again to show that for $t > 0$,

$$\frac{2[f(x + t(u + v)) - f(x)]}{2t} \le \frac{f(x + 2tu) - f(x)}{2t} + \frac{f(x + 2tv) - f(x)}{2t}$$

and take limits as $t \to 0^+$.

1.3 Definition. The convex function f is said to be Gateaux differentiable at $x_0 \in D$ provided the limit

$$df(x_0)(x) = \lim_{t \to 0} \frac{f(x_0 + tx) - f(x_0)}{t}$$

exists for each $x \in E$. The function $df(x_0)$ is called the Gateaux derivative (or Gateaux differential) of f at x_0.

It is immediate from this definition that f is Gateaux differentiable at x_0 if and only if $-d^+f(x_0)(-x) = d^+f(x_0)(x)$ for each $x \in E$. Since a sublinear functional p is linear if and only if $p(-x) = -p(x)$ for all x, this shows that f is Gateaux differentiable at x_0 if and only if

$$x \to d^+f(x_0)(x)$$

is linear in x; in particular, if this is true, then $df(x_0)$ is a linear functional on E.

1.4 Examples.

(a) If f is a linear functional on E (not necessarily continuous), then $df(x_0)(x) = f(x)$ for all x_0 and x. For an example of a discontinuous linear functional on a normed linear space, let $f(x) = x'(0)$, for x in the space of all polynomials on $[-1, 1]$ with supremum norm. (It is easy to construct a sequence of polynomials x_n converging uniformly to 0 such that $x_n'(0) = 1$ for all n.) Thus, $x \to df(x_0)(x)$ need not be continuous.

(b) The norm $\|x\|_1 = \sum |x_n|$ in ℓ^1 is Gateaux differentiable precisely at those points $x = (x_n)$ for which $x_n \neq 0$ for all n. In this case, the Gateaux differential is the bounded sequence $(\operatorname{sgn} x_n) \in \ell^\infty$. The norm in $\ell^1(\Gamma)$ (Γ uncountable) is not Gateaux differentiable at any point.

Proof. If $x \in \ell^1$ and $x_n = 0$ for some n, let $\delta_n = (0, 0, \ldots, 0, 1, 0, \ldots)$ be the sequence whose only nonzero term is a 1 in the n-th place. It follows that $\|x + t\delta_n\|_1 - \|x\|_1 = |t|$, so dividing both sides by t shows that the (two-sided) limit as $t \to 0$ does not exist. [This observation shows how to prove the second assertion, since any element of $\ell^1(\Gamma)$ vanishes at all but a countable number of members of Γ.] Suppose, on the other hand, that for every n, $x_n \neq 0$, that $\varepsilon > 0$ and that $y \in \ell^1$. We can choose $N > 0$ such that $\sum_{n>N}|y_n| < \varepsilon/2$. For sufficiently small $\delta > 0$ we have

$$\operatorname{sgn}(x_n + ty_n) = \operatorname{sgn} x_n \quad \text{if} \quad 1 \leq n \leq N, \quad |t| < \delta.$$

Consequently,

$$|t^{-1}(\|x + ty\|_1 - \|x\|_1) - \sum y_n \operatorname{sgn} x_n| \leq$$

$$|\sum_{n=1}^{N} t^{-1}\{|x_n + ty_n| - |x_n| - ty_n \operatorname{sgn} x_n\}| + 2\sum_{n>N}|y_n| < \varepsilon$$

provided $|t| < \delta$.

If f is a <u>continuous</u> convex function which is Gateaux differentiable at a point, then its differential is a continuous linear functional. This is a consequence of the following basic result.

1.5 Notation. If $x \in E$ and $r > 0$, the closed ball centered at x is denoted by $B(x; r) = \{y \in E: \|x - y\| \leq r\}$.

1.6 Proposition. If the convex function f is continuous at $x_0 \in D$, then it is <u>locally Lipschitzian</u> at x_0, that is, there exist $M > 0$ and $\delta > 0$ such that $B(x_0; \delta) \subset D$ and

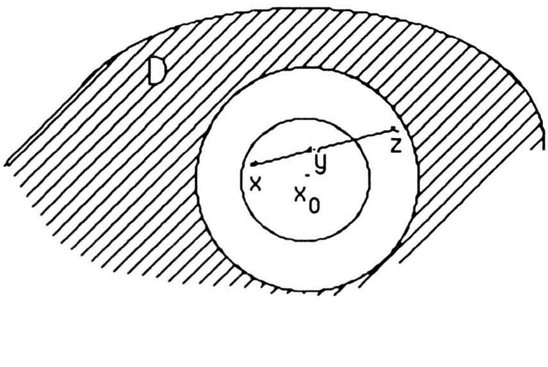
$$|f(x) - f(y)| \leq M\|x - y\|$$

whenever $x, y \in B(x_0; \delta)$.

Proof. Since f is continuous at x_0, it is locally bounded there; that is, there exist $M_1 > 0$ and $\delta > 0$ such that $|f| \leq M_1$ on $B(x_0; 2\delta) \subset D$. If x, y are distinct points of $B(x_0; \delta)$, let $\alpha = \|x - y\|$ and let

$$z = y + (\delta/\alpha)(y - x);$$

see the sketch below.



Note that $z \in B(x_0; 2\delta)$. Since $y = [\alpha/(\alpha+\delta)]z + [\delta/(\alpha+\delta)]x$ is a convex combination (lying in $B(x_0; 2\delta)$), we have

$$f(y) \leq [\alpha/(\alpha+\delta)]f(z) + [\delta/(\alpha+\delta)]f(x) \quad \text{so}$$

$$f(y) - f(x) \leq [\alpha/(\alpha+\delta)]\{f(z) - f(x)\} \leq (\alpha/\delta)\cdot 2M_1 = (2M_1/\delta)\|x - y\|.$$

Interchanging x and y gives the desired result, with $M = 2M_1/\delta$.

Note that we only used local boundedness of f; hence the latter property is <u>equivalent</u> to continuity for convex functions.

1.7 Corollary. If the convex function f is continuous at $x_0 \in D$, then $d^+f(x_0)$ is a continuous sublinear functional on E, and hence $df(x_0)$ (when it exists) is a continuous linear functional.

Proof. Given $x_0 \in D$ there exists a neighborhood B of x_0 and $M > 0$ such that, if $x \in E$, then

$$f(x_0 + tx) - f(x_0) \leq Mt\|x\|$$

provided $t > 0$ is sufficiently small that $x_0 + tx \in B$. Thus, for all points $x \in E$, we have $d^+f(x_0)(x) \leq M\|x\|$, which implies that $d^+f(x_0)$ is continuous.

1.8 Proposition. The continuous convex function f is Gateaux differentiable at $x_0 \in D$ if and only if there exists a unique functional x^* in E^* satisfying

$$(*) \qquad \langle x^*, x - x_0\rangle \leq f(x) - f(x_0), \qquad x \in D,$$

or equivalently

$$(**) \qquad \langle x^*, y\rangle \leq d^+f(x_0)(y), \qquad y \in E.$$

Proof. We first show that (*) and (**) are equivalent. If x^* satisfies (*), then for any $y \in E$ we have $x_0 + ty \in D$ for sufficiently small $t > 0$ hence $t\langle x^*, y\rangle = \langle x^*, (x_0 + ty) - x_0\rangle \leq f(x_0 + ty) - f(x_0)$ which implies that x^* satisfies (**). Conversely, if x^* satisfies (**) and $x \in D$, let $y = x - x_0$; then $x_0 + t(x - x_0) \in D$ if $0 < t \leq 1$ so

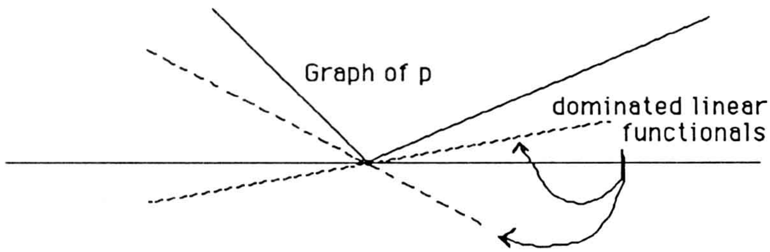
$$\langle x^*, x - x_0\rangle \leq d^+f(x_0)(x - x_0) \leq t^{-1}[f(x_0 + t(x - x_0)) - f(x_0)].$$

Setting $t = 1$ yields (*).

If $df(x_0)$ exists, then $df(x_0)(x - x_0) \leq f(x) - f(x_0)$ as above, so $df(x_0)$ satisfies (*). Moreover, if x^* satisfies (*), then it satisfies (**); linearity of $d^+f(x_0) = df(x_0)$ implies that $x^* = df(x_0)$.

Conversely, suppose that x^* is the unique element of E^* satisfying (*), hence the unique element satisfying (**). We now apply the general fact that if a continuous sublinear functional p majorizes exactly one linear functional, then p is itself linear. Indeed, if p is not linear, then it dominates many linear functionals (see the sketch below); the proof is an easy consequence of the Hahn-Banach theorem: If $-p(-x) < p(x)$, find p-dominated extensions of the linear functionals

$$f_1(rx) = rp(x) \quad \text{and} \quad f_2(rx) = -rp(-x).$$

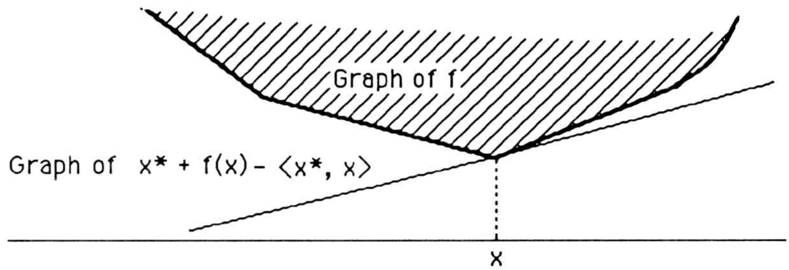


The functionals x^* which satisfy (*) play an important role in the study of convex functions, so they are singled out for special attention.

1.9 Definition. If f is a convex function defined on the convex set C and $x \in C$, we define the <u>subdifferential of</u> f <u>at</u> x to be the set $\partial f(x)$ of all $x^* \in E^*$ satisfying

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

Note that this is the same as saying that the affine function $x^* + \alpha$, where $\alpha = f(x) - \langle x^*, x\rangle$, is dominated by f and is equal to it at $y = x$, as indicated in the sketch.



The Hahn-Banach argument we used above shows quickly that if f is continuous at x_0, then $\partial f(x_0)$ is nonempty: $d^+f(x_0)$ is continuous and sublinear, so (as above) there exists x^* such that $\langle x^*, y\rangle \leq d^+f(x_0)(y)$ for all $y \in E$. Using the fact that the right-hand difference quotients for $d^+f(x_0)$ are decreasing, replacing y by $y - x_0$ and letting $t = 1$, we get

$$\langle x^*, y - x_0\rangle \leq d^+f(x_0)(y - x_0) \leq f((y - x_0) + x_0) - f(x_0) \quad \text{for all } y \in C.$$

As we will see later, it is still possible to have $\partial f(x_0)$ nonempty for certain convex f which are <u>not</u> continuous at x_0.

1.10 Exercise. Prove that for any convex function f the set $\partial f(x_0)$ (possibly empty!) is convex and weak* closed. (Note that a continuous convex f is Gateaux differentiable at x_0 if and only if $\partial f(x_0)$ is a singleton.)

1.11 Proposition. If the convex function f is continuous at $x_0 \in D$, then $\partial f(x_0)$ is a nonempty, convex and weak* compact subset of E^*. Moreover, the map $x \to \partial f(x)$ is locally bounded at x_0, that is, there exist $M > 0$ and and neighborhood U of x_0 in D such that $\|x^*\| \leq M$ whenever $x \in U$ and $x^* \in \partial f(x)$.

Proof. The fact that $\partial f(x_0)$ is nonempty, weak* closed and convex follows from the preceding remarks and Exercise 1.10. The fact that it is weak* compact will follow from Alaoglu's theorem, once we have shown the local boundedness property. Since, by Proposition 1.6, f is locally Lipschitzian at x_0, there exist $M > 0$ and a neighborhood U of x_0 such that

$$|f(y) - f(x)| \leq M\|y - x\| \quad \text{whenever } x, y \in U.$$

If $x \in U$ and $x^* \in \partial f(x)$, then for all $y \in U$ we have

$$\langle x^*, y - x\rangle \leq f(y) - f(x) \leq M\|y - x\|,$$

which implies that $\|x^*\| \leq M$.

1.12 Definitions. Supposed that E and F are normed linear spaces, that U is a nonempty open subset of E and that $\varphi: U \to F$ is a continuous function. We can extend the definition of Gateaux differentiability as follows: Say that φ is <u>Gateaux differentiable</u> at the point $x_0 \in U$ provided there exists a continuous linear map from E to F (denoted by $d\varphi(x_0)$) such that

$$(\#) \qquad d\varphi(x_0)(x) = \lim_{t\to 0^+} t^{-1}\{\varphi(x_0 + tx) - \varphi(x_0)\} \quad \text{for each } x \in E.$$

Another way of stating this is to say that φ has directional derivatives at x_0 in every direction x and the resulting function of x is continuous and linear.

We say that φ is <u>Frechet differentiable</u> at $x_0 \in U$ provided there exists a continuous linear map from E to F (denoted by $\varphi'(x_0)$) such that for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$(\#\#) \qquad \|\varphi(x_0 + x) - \varphi(x_0) - \varphi'(x_0)(x)\| \leq \varepsilon\|x\| \quad \text{whenever} \quad \|x\| < \delta.$$

We call $\varphi'(x_0)$ (which is easily seen below to be unique) the <u>Frechet differential</u> (or <u>Frechet derivative</u>) of φ.

For the moment, we will be dealing with <u>real</u>-valued continuous functions, so Gateaux and Frechet derivatives will be continuous linear maps from E into R, that is, elements of E^*.

1.13 Facts.

(a) If f is a continuous function which is Frechet differentiable at x_0, then it is Gateaux differentiable there and $\varphi'(x_0) = d\varphi(x_0)$. To see this, replace x in $(\#\#)$ by tx, fix x and let $t \to 0^+$. Since limits are unique, the operator $d\varphi(x_0)$ is uniquely determined, hence $\varphi'(x_0)$ is unique.

(b) Note that φ is Frechet differentiable at x_0 if it is Gateaux differentiable there and if the limit in $(\#)$ exists <u>uniformly</u> for $\|x\| \leq 1$ as $t \to 0^+$.

1.14 Examples.

(a) The norm in ℓ^1 is not Frechet differentiable at any point.

Proof. By Example 1.4(b), we need only consider a point $x = (x_n)$ for which $x_n \neq 0$ for all n. Given such an x, for each $m \geq 1$ let

$$y^m = (0, 0, \ldots, 0, -2x_m, -2x_{m+1}, -2x_{m+2}, \ldots).$$

Then $\|y^m\|_1 \to 0$ as $m \to \infty$. Of course, the sequence $(\text{sgn } x_n)$ is our only candidate for the Frechet differential. But

$$\left| \|x + y^m\|_1 - \|x\|_1 - \sum (y^m)_n \text{ sgn } x_n \right| = \left| \sum_{n \geq m} (-2|x_n|) \right| = \|y^m\|_1.$$

(b) The square of the norm in Hilbert space H is everywhere Frechet differentiable. By the chain rule, the norm is therefore differentiable at every point other than the origin.

Proof. If $x, y \in H$, then $\|x + y\|^2 - \|x\|^2 - 2(x, y) = \|y\|^2$; it follows readily that $y \to 2(x, y)$ is the Frechet derivative of $\|\cdot\|^2$ at x.

(c) There exists an equivalent norm on ℓ^1 which is Gateaux differentiable at every point (except the origin), but is nowhere Frechet differentiable. This striking example will be easy to prove after we have developed a few tools in later sections, so it will be postponed until Section 5 (following Theorem 5.12).

(d) In Hilbert space H let C be a nonempty closed convex set and denote by P the Lipschitz continuous nearest point mapping (or metric projection) of H onto C; that is, for all $x \in H$, $P(x)$ is the unique point satisfying

$$\|x - P(x)\| = \inf\{\|x - y\|: y \in C\}.$$

Recall that $\|P(x) - P(y)\| \leq \|x - y\|$ for all $x, y \in H$. Define f on H by $f(x) = (1/2)[\|x\|^2 - \|x - P(x)\|^2]$; then f is continuous, convex and everywhere Frechet differentiable, with $f'(x) = P(x)$ for all x.

Proof. Since

$$2f(x) = \|x\|^2 - \inf\{\|x - y\|^2: y \in C\} = \sup\{2\langle x, y\rangle - \|y\|^2: y \in C\},$$

f is the supremum of affine functions, hence is convex (and it is clearly continuous). To see the differentiability property, fix $x \in H$; then for any $y \in H$ we have

$$\|(x + y) - P(x + y)\| \leq \|(x + y) - P(x)\|, \text{ so } f(x + y) - f(x) - \langle P(x), y\rangle \geq 0.$$

On the other hand, since $\|x - P(x)\| \leq \|x - P(x + y)\|$ we get

$$f(x + y) - f(x) - \langle P(x), y\rangle \leq \langle y, P(x + y) - P(x)\rangle \leq$$
$$\leq \|y\|\cdot\|P(x + y) - P(x)\| \leq \|y\|^2,$$

which proves the differentiability assertion.

1.15 Exercises.

(a) Prove that for continuous convex functions in finite dimensional spaces, Gateaux differentiability implies Frechet differentiability. (Hint: Use the Fact 1.13 (b), the local Lipschitz property and compactness of the unit ball