Journal of Convex Analysis
Volume 2 (1995), No.1/2, 117–144

Proximal Smoothness and the Lower–C^2 Property

F. H. Clarke[1,2]
Centre de recherches mathématiques, Université de Montréal, Montréal, Québec H3C 3J7, Canada.

R. J. Stern[1,2]
Department of Mathematics and Statistics, Concordia University, Montreal, Quebec H4B 1R6, Canada.

P. R. Wolenski
Department of Mathematics, Louisiana State University, Baton Rouge, Louisiana 70803, USA.

Dedicated to R. T. Rockafellar on his 60th Birthday

A subset X of a real Hilbert space H is said to be *proximally smooth* provided that the function $d_X : H \to R$ (the distance to X) is continuously differentiable on an open tube U around X. It is proven that this property is equivalent to d_X having a nonempty proximal subgradient at every point of U, and that the (Gâteaux = Fréchet) derivative is locally Lipschitz on U. The Lipschitz behavior of the derivative is a consequence of the fact that under proximal smoothness, the metric projection onto X is single valued and Lipschitz on U. Alternate characterizations of proximal smoothness are given as well, in terms of properties of the proximal normal cone multifunction on X and on nearby closed neighborhoods of X. In case X is weakly closed, the list of equivalences is extended to include each point of U admitting a unique closest point in X. Further specializations are given in finite dimensions. In that setting, we discuss properties of locally Lipschitz real valued functions whose epigraphs are proximally smooth in a local sense. It is demonstrated that this function class coincides with the *lower–C^2* functions studied by Rockafellar.

Keywords : Hilbert space, distance function, proximal smoothness, Gâteaux, strict and Fréchet derivatives, closest points, nonsmooth analysis, optimization, epigraph, lower–C^2 function.

1991 Mathematics Subject Classification: 26B05

1. Introduction

In the core theory of nonsmooth analysis, the major analytic concepts for functions (various generalized notions of directional derivative and gradient) have geometric counterparts for sets (corresponding generalizations of the classical ideas of tangency and normality). This duality, which has always played an important role in the development of the subject,

[1] Research supported in part by the Natural Sciences Engineering Research Council of Canada.
[2] Research supported in part by le fonds FCAR du Québec.

ISSN 0944-6532 / $ 2.50

is generally achieved by rephrasing analytic properties of a function f in terms of geometric properties of epi(f), the epigraph of f. An indispensible tool employed in forging this link has been the distance function. The present article proceeds in this vein. In a general Hilbert space setting, we shall employ proximal analysis in order to characterize *proximal smoothness* of a general closed set X, where this connotes continuous differentiability of d_X (the distance function to X) on an open tube U around X. We then proceed to establish analytic properties of the class of functions $f : R^n \to R$ for which epi(f) is proximally smooth in a local sense. It transpires that this function class corresponds precisely to one considered by R. T. Rockafellar in [18]: f is said to be *lower–C^2* provided that for each point $y \in R^n$ there exists an open neighborhood N_y of y so that locally f has the representation

$$f(x) = \max_{s\in S} F(x,s) \quad \forall x \in N_y, \tag{1.1}$$

where S is a compact set in some topological space, and $F : N_y \times S \to R$ is a function which has second partial derivatives with respect to x and which, along with these derivatives, is jointly continuous in $(x, s) \in N_y \times S$. Thus f is the value function of a family of maximization problems parametrized by x and satisfying specific continuity and smoothness requirements. Rockafellar derived several equivalent conditions which characterize a locally Lipschitz function f being lower–C^2, including a type of monotonicity in terms of the Clarke subdifferential of f, the expressibility of f as a difference of convex and quadratically convex functions, as well as the fact (referred to in section 5 below) that F in (1.1) may without loss of generality be taken to be of a particular form in which it is quadratic in x and continuous in s. In [18], the distinctions between the class of lower–C^2 functions and the analogously defined class of lower–C^1 functions were made clear. Properties of the latter class are well–known; see e.g. Danskin [9], Clarke [8], [5], Rockafellar [19] and Spingarn [20]. As was pointed out in [18], however, the lower–C^k classes all coincide for $2 \le k \le \infty$. Let us also mention that the lower–C^2 property has appeared elsewhere in the literature as well, along with geometric versions, under such names as "generalized convexity", "weak convexity", "convexity up to a square", "F–convexity", and so on. We refer the reader to Vial [21] for a comprehensive treatment of this subject as well as for references and historical comments. See also Hiriart-Urruty [12] in this regard.

The plan of the present article is as follows: Section 2 contains the requisite background in nonsmooth analysis, including recent results of Clarke, Ledyaev and Wolenski [6] on proximal subdifferentiability of the distance function in Hilbert space. Section 3 contains certain general results on the distance function which are of independent interest, and are required in section 4, where we present our main characterizations of proximal smoothness, with the only requirement on X being closedness. Notable among the properties equivalent to proximal smoothness is that d_X possess a nonempty proximal subdifferential at every point of the tube U. It will also be shown that the derivative is then necessarily locally Lipschitz on U. The Lipschitz behavior of the derivative will be seen to follow from the fact that for X proximally smooth, the metric projection onto X is single valued and Lipschitz on U. We shall also obtain geometric characterizations which involve the behavior of the proximal normal cone multifunction on X, as well as on nearby approximations of X. Under the additional assumption that X is weakly closed, the list of equivalences is extended to include the existence of a unique closest point in X for every $u \in U$, and certain specializations to the finite dimensional case are included as well. In section 5 we demonstrate that a locally Lipschitz function's being lower–C^2 is in fact a local form of proximal smoothness of its epigraph.

Prior to proceeding, we pause for a heuristic digression which serves to motivate some of the subsequent results. Consider a closed set $X \subseteq R^n$. A useful pedagogical device for describing the existence of proximal normals at a point $x \in \mathrm{bdry}(X)$ is the "paintability" of x; that is, the existence a closed ball (that can be thought of as a paintroller, if one prefers) which touches X only at x. We can then go on to say that X is "paintable" provided that, outfitted with an infinite family of paintrollers of arbitrarily small radii, we can paint every point in the boundary of X. But what if we had our disposal only a *single* paintroller; that is, only one radius r was available? Our results will show that paintability of X under this restriction is tantamount to what we have above termed proximal smoothness of X. An outcome of this is that if one paints the boundary of X with a single roller, and allows the roller to assume all possible positions at each point being painted, the locus of the roller's center (which is the boundary of a closed r–neighborhood of X) is a C^1 manifold; see part 4 of Corollary 4.15 below.

2. Preliminaries in nonsmooth analysis

In this section we present a concise overview of basic material from nonsmooth analysis; other required facts will be mentioned later. We cite Clarke [8], [5], [4], and Loewen [14] as general references and guides to the literature.

Throughout this article, we will let X denote a closed subset of a real Hilbert space H, which is equipped with an inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $\| \cdot \|$. The distance of a point $u \in H$ to X is given by

$$d_X(u) := \inf\{\|u - x\| : x \in X\}.$$

We require the fact that the distance function $d_X : H \to R$ is globally Lipschitz of rank 1. The (possibly empty) set of closest points to u in X is denoted

$$\mathrm{proj}_X(u) := \{x \in X : \|u - x\| = d_X(u)\}.$$

The multifunction proj_X is referred to as the *metric projection* onto X. If $u \notin X$ and $x \in \mathrm{proj}_X(u)$, then we say that the vector $u - x$ is a *perpendicular* to X at x. The set of all nonnegative multiples of such perpendiculars is denoted $N_X^P(x)$, and is referred to as the *proximal normal cone* (or P–normal cone) to X at x. One can show that $\zeta \in N_X^P(x)$ if and only if there exists $M > 0$ such that the following *proximal normal inequality* holds:

$$M\|y - x\|^2 \geq \langle \zeta, y - x \rangle \quad \forall y \in X.$$

(In general, M will depend on x.) Thus the P–normal cone is convex (but may not be closed even in finite dimensions). If $x \in \mathrm{int}(X)$ or no perpendiculars to X exist at x, then by convention, we set $N_X^P(x) = \{0\}$. Let $x \in X$ and $0 \neq \zeta \in N_X^P(x)$. If one has

$$X \cap \mathrm{int}\left\{x + r\left(\frac{\zeta}{\|\zeta\|} + \mathrm{cl}(B)\right)\right\} = \phi \quad \forall y \in X,$$

then we shall say that ζ is *realized by an r-ball.* This expresses the fact that open ball of radius r centered at $x + r\frac{\zeta}{\|\zeta\|}$ has empty intersection with X. It will be useful for us to note that this condition can be rephrased as

$$\left\| x + r\frac{\zeta}{\|\zeta\|} - y \right\| \geq r \quad \forall y \in X.$$

Upon squaring and expanding, this in turn becomes

$$\frac{1}{2r}\|y - x\|^2 \geq \left\langle \frac{\zeta}{\|\zeta\|}, y - x \right\rangle \quad \forall y \in X. \tag{2.1}$$

Let $f : U \to (-\infty, \infty]$ be a lower semicontinuous extended real valued function on an open set $U \subseteq H$, and denote its epigraph by

$$\mathrm{epi}(f) := \{(x, y) : x \in \mathrm{dom}(f),\ y \geq f(x)\},$$

where $\mathrm{dom}(f)$ is the set of points in U where f is finite. The lower semicontinuity assumption is equivalent to *epi*(f) being closed, a fact which is relevant in regard to the next definition: $\zeta \in H$ is said to be a *proximal subgradient* (or P–subgradient) of f at $x \in U$ provided that

$$(\zeta, -1) \in N^P_{\mathrm{epi}(f)}(x, f(x)).$$

The set of all such vectors is called the P–*subdifferential* of f at x, and is denoted $\partial_P f(x)$. Similarly, the P–*superdifferential* of an upper semicontinuous function f at x is defined as

$$\partial^P f(x) := -\partial_P(-f)(x),$$

with the members of this set referred to as P–*supergradients*. One can prove that $\zeta \in \partial_P f(x)$ if and only if there exist positive numbers σ and γ such that the following *proximal subgradient inequality* is satisfied:

$$f(y) - f(x) + \sigma\|y - x\|^2 \geq \langle \zeta, y - x \rangle \quad \forall y \in x + \gamma B,$$

where B denotes the open unit ball in H. We shall require the fact that $\partial_P f(x) \neq \phi$ for all x in a dense subset of $\mathrm{dom}(f)$; for an optimization based proof, see Clarke, Stern and Wolenski [7].

The following result on proximal subdifferentiability of the distance function to a closed subset of a Hilbert space, and the existence of nearest points, appears in Clarke, Ledyaev and Wolenski [6], and is a consequence of more general variational principles. Related work on the existence of closest points in more general Banach space settings (as well as further references on this topic) may be found in Borwein and Fitzpatrick [2]; see also Borwein and Giles [3] and Ioffe [13].

Theorem 2.1. *Let $u \notin X$, and suppose that $\partial_P d_X(u) \neq \phi$. Then the following hold:*
(1) $\mathrm{proj}_X(u)$ *is a singleton, say* $\{x\}$, *and* $\partial_P d_X(u)$ *is the singleton* $\{\zeta\}$, *where*

$$\zeta := \frac{u - x}{d_X(u)}.$$

(2) *If u_i is a sequence in X such that $u_i \to u$, and $\{\zeta_i\} = \partial_P d_X(u_i)$, then $\zeta_i \to \zeta$.*

Since d_X has a nonempty P–subdifferential on a dense subset of $H\backslash X$, the preceding theorem readily yields the conclusion that $\mathrm{proj}_X(u) \neq \phi$ for all u in a dense subset of $H\backslash X$, the complement of X. (See also Lau [14] for a Banach space version.) It readily follows that $N_X^P(x) \neq \{0\}$ for a dense set of $x \in \mathrm{bdry}(X)$.

The *limiting normal cone* (or L–normal cone) to X at $x \in X$ is defined to be the set

$$N_X^L(x) := \{\zeta : \zeta_i \xrightarrow{w} \zeta,\ \zeta_i \in N_X^P(x_i),\ x_i \to x\},$$

where "$\xrightarrow{w}$" denotes weak convergence; this set is nonempty and weakly closed. The *convexified normal cone* (or C–normal cone) to X at x is defined as

$$N_X^C(x) := \mathrm{clco}[N_X^L(x)],$$

the closure of the convex hull of the L–normal cone at x.

The normal cones defined above give rise to corresponding nonempty subdifferential sets for f as one would expect:

$$\partial_L f(x) := \{\zeta : (\zeta, -1) \in N^L_{\mathrm{epi}(f)}(x, f(x)).$$

$$\partial_C f(x) := \{\zeta : (\zeta, -1) \in N^C_{\mathrm{epi}(f)}(x, f(x)).$$

These are referred to as the *limiting subdifferential* (or L–subdifferential) and *convexified subdifferential* (or C–subdifferential) of f at x, respectively. It is not difficult to show directly that

$$\partial_L f(x) = \{\zeta : \zeta_i \xrightarrow{w} \zeta,\ \zeta_i \in \partial_P f(x_i),\ x_i \to x,\ f(x_i) \to f(x)\}.$$

Should it exist, the *directional derivative* of f at x in the direction v is the quantity

$$f'(x; v) := \lim_{t \downarrow 0} \frac{f(x + tv) - f(x)}{t}.$$

If there exists $\zeta \in H$ such that $f'(x; v) = \langle \zeta, v\rangle$ for every v, then ζ is unique, and we say that $\zeta := f'(x)$ is the *Gâteaux derivative* of f at x. When f has a Gâteaux derivative at x and additionally the convergence defining $f'(x; v)$ is uniform on bounded sets of v, then $f'(x)$ is called the *Fréchet derivative.* If f has a Gâteaux derivative at x, then

$$\partial_P f(x) \subseteq \{f'(x)\}.$$

In other words, if f is Gâteaux differentiable at x and has a nonempty P–subdifferential at x, then this reduces to the singleton $\{f'(x)\}$.

Let us now assume that f is Lipschitz of rank K on an open set U. By a result in [7], this is equivalent to $\|\zeta\| \leq K$ for every $\zeta \in \partial_P f(x)$, for each $x \in U$ where the P–subdifferential is nonempty. In the Lipschitz case, $\partial_C f(x)$ is weakly compact, and one has the relation

$$\partial_C f(x) = \mathrm{clco}[\partial_L f(x)].$$

Furthermore, the Gâteaux derivative satisfies has

$$f'(x) \in \partial_C f(x)$$

at points of differentiability. (If $H = R^n$, then Gâteaux and Fréchet differentiabilities coincide at x; this is false if the Lipschitz assumption is removed.) A Lipschitz function f is said to be *strictly differentiable* at x provided that $\partial_L f(x)$ (or equivalently $\partial_C f(x)$) reduces to a singleton ζ. Should this occur, then necessarily ζ is the Gâteaux derivative $f'(x)$. We say that f is *continuously differentiable* at a point $x_o \in H$ provided that the Gâteaux derivative exists on a ball around x_o, and the mapping $x \to f'(x)$ is continuous on that ball. It is not difficult to show that this is equivalent to continuous Fréchet differentiability; hence there is no ambiguity about what is meant by a function being continuously differentiable (C^1). We will require the fact that continuous differentiability of f at x implies strict differentiability at x. (On the other hand, even a C^1 function $f : R \to R$ can have $\partial_P f(x) = \phi$ "often", in either the sense of measure or category, as is demonstrated by an example in [6].)

For f Lipschitz near x and given $v \in H$, one defines the *generalized directional derivative* of f at x in the direction v as

$$f^o(x; v) := \limsup_{\substack{t \downarrow 0 \\ y \to x}} \frac{f(y + tv) - f(y)}{t}.$$

For any v, one has

$$f^o(x; v) = \max\{\langle \zeta, v\rangle : \zeta \in \partial_C f(x)\}.$$

Furthermore, $\zeta \in \partial_C f(x)$ if and only if

$$\langle \zeta, v\rangle \leq f^o(x; v) \quad \forall v \in H.$$

We say that f (still assumed Lipschitz near x) is *regular* at x if the ordinary directional derivative $f'(x; v)$ exists and $f'(x; v) = f^o(x; v)$ for every $v \in H$. We shall require the fact that regularity at x is guaranteed when f is continuously differentiable at x, or if f is strictly differentiable at x. A closed set $X \subset H$ is said to be regular at $x \in X$ provided that

$$\limsup_{\substack{y \to x \\ y \in X}} \left\langle \zeta, \frac{y - x}{\|y - x\|} \right\rangle \leq 0 \quad \forall \zeta \in N_X^C(x).$$

One can show that this definition of regularity agrees with the usual one given via tangency notions ([8], [5]) and that a sufficient condition for this property is that $N_X^P(x) = N_X^C(x)$. Also, regularity of a Lipschitz function f at x is equivalent to regularity of its epigraph at $(x, f(x))$, and so a sufficient condition for this is $\partial_P f(x) = \partial_C f(x)$.

3. General results

In this section, we shall collect certain preparatory results, which are also of some independent interest. The following result complements Theorem 2.1.

Theorem 3.1. *Let $u \notin X$ be such that* $\mathrm{proj}_X(u) \neq \phi$. *Then the following hold:*
(1) $\partial^P d_X(u) \neq \phi$.
(2) *If $\partial_P d_X(u) \neq \phi$, then d_X is Fréchet differentiable at u and*

$$\partial_P d_X(u) = \partial^P d_X(u) = \{d'_X(u)\} = \left\{ \frac{u-x}{d_X(u)} \right\},$$

where $\mathrm{proj}_X(u) = \{x\}$.

Proof. Consider the function

$$f_x(w) := \|w - x\|.$$

For $0 < \delta < d_X(u)$, it is continuously differentiable on the ball $u + \delta B$, and there exists $K > 0$ such that for each $x \in X$, the mapping

$$w \to f'_x(w) = \frac{w-x}{\|w-x\|}$$

is Lipschitz of rank K on $u + \delta B$. By the mean value theorem, for each $x \in X$ and $w \in u + \delta B$,

$$f_x(w) - f_x(u) = \langle f'_x(q), w - u \rangle$$

for some q contained in the line segment (u, w). The Cauchy–Schwarz inequality now implies that for each such u and w one has

$$f_x(w) - f_x(u) \leq \langle f'_x(u), w - u \rangle + K\|w - u\|^2.$$

Let $\tilde{x} \in \mathrm{proj}_X(u)$. Then for each $w \in u + \delta B$ we have

$$d_X(w) - d_X(u) - K\|w - u\|^2 \leq \langle f'_{\tilde{x}}(u), w - u \rangle, \tag{3.1}$$

and therefore $f'_{\tilde{x}}(u) \in \partial^P d_X(u)$.
Now let $\zeta \in \partial_P d_X(u)$. Then there exist positive numbers σ and γ such that

$$d_X(w) - d_X(u) + \sigma\|w - u\|^2 \geq \langle \zeta, w - u \rangle \quad \forall w \in u + \gamma B, \tag{3.2}$$

where $\gamma < \delta$. The inequalities (3.1) and (3.2) together imply that $\zeta = f'_{\tilde{x}}(u)$ is the Fréchet derivative of d_X at u. □

Remark 3.2. Since the P–subdifferential of the distance function is nonempty on a dense subset of $H \backslash X$, Theorems 2.1 and 3.1 together yield the fact that d_X is Fréchet differentiable on a dense subset of $H \backslash X$. In view of the Lipschitz nature of d_X, this conclusion by itself could also have been reached via Rademacher's theorem in finite dimensions, or via a result of Preiss [15] in a setting more general than Hilbert space.

We shall have frequent occasion to employ the closed r–neighborhoods of X ($r \geq 0$) defined via

$$X(r) := \{u \in H : d_X(u) \leq r\}.$$

Two useful elmentary facts about $X(r)$ are now summarized.

Lemma 3.3.
(1) *Suppose that* $x \in \mathrm{proj}_X(x+\delta\zeta)$, *where* $\delta > 0$ *and* $\|\zeta\| = 1$. *Then*

$$0 \le r \le s < \delta \Longrightarrow x + r\zeta \in \mathrm{proj}_{X(r)}(x+s\zeta).$$

(2) *Let* $\zeta \in N^P_{X(r)}(u)$ *and* $\|\zeta\| = 1$. *Then there exists* $\delta > 0$ *such that* $d_X(u+\delta\zeta) = r+\delta$.

Proof. Upon noting that $x+r\zeta \in X(r)$, the first assertion will follow upon showing that $x+r\zeta \in \mathrm{proj}_{X(r)}(x+\delta\zeta)$. To this end, we only need to verify that $d_{X(r)}(x+\delta\zeta) \ge \delta - r$. To see this, observe that if $d_{X(r)}(x+\delta\zeta) < \delta - r$, then $d_X(x+\delta\zeta) < \delta$, violating $d_X(x+\delta\zeta) = \delta$.

In order to prove the second assertion of the lemma, denote $z = u + \delta\zeta$, where $\delta > 0$ is chosen small enough to ensure that $d_{X(r)}(z) = \|z - u\| = \delta$. Then $d_X(z) \le r + \delta$. Now consider any $x \in X$. One has $d_X(x) = 0$, and therefore the Intermediate Value Theorem implies that there exists a point y in the segment $[x, z]$ such that $d_X(y) = r$; that is, $y \in X(r)$. But then

$$\|z - x\| = \|y - x\| + \|z - y\| \ge r + \delta,$$

since $\|y - x\| \ge r$ and $\|z - y\| \ge \delta$. Hence $d_X(z) \ge r + \delta$, and therefore equality holds. $\square$

Consider $u \in H$ such that $d_X(u) = r > 0$, and suppose that $\zeta \in \partial_P d_X(u)$. Then the proximal subgradient inequality tells us that that there exist positive numbers σ and γ such that

$$d_X(w) - d_X(u) + \sigma\|w-u\|^2 \ge \langle \zeta, w - u\rangle \quad \forall w \in u + \gamma B,$$

and so

$$\sigma\|w-u\|^2 \ge \langle \zeta, w-u\rangle \quad \forall w \in \{u + \gamma \mathrm{cl}(B)\} \cap X(r).$$

Therefore ζ is a P–normal to the set $\{u + \gamma\mathrm{cl}(B)\} \cap X(r)$ at u, and therefore also $\zeta \in N^P_{X(r)}(u)$. Since $\|\zeta\| = 1$, we have shown that

$$\partial_P d_X(u) \subseteq N^P_{X(r)}(u) \cap \{\zeta \in H : \|\zeta\| = 1\}.$$

The reverse containment is true as well, but does not follow as readily. This is taken up next.

Theorem 3.4. *Suppose that* $d_X(u) = r > 0$. *Then*

$$\partial_P d_X(u) = N^P_{X(r)}(u) \cap \{\zeta \in H : \|\zeta\| = 1\}. \tag{3.3}$$

Proof. In view of the preceding discussion, we only need to show that the right side of (3.3) is contained in the left. Suppose that $\zeta \in N^P_{X(r)}(u)$ and that $\|\zeta\| = 1$. Then, by part 2 of Lemma 3.3, there exists $\delta > 0$ such that

$$d_X(u+\delta\zeta) = r + \delta \tag{3.4}$$

and

$$\{y \in H : \|y - u - \delta\zeta\| < r + \delta\} \cap X = \phi. \tag{3.5}$$

Then since d_X has Lipschitz rank 1, for all u' one has

$$d_X(u') \geq r + \delta - \|u' - u - \delta\zeta\|. \tag{3.6}$$

Now, $\zeta \in \partial_P d_X(u)$ if and only if there exists $\sigma > 0$such that

$$d_X(u') - \langle \zeta, u' - u \rangle + \sigma \|u' - u\|^2 \geq d_X(u) = r \tag{3.7}$$

for all u' near u. From (3.6) it then follows that (3.7) holds provided that

$$\delta + \sigma \|u' - u\|^2 \geq \langle \zeta, u' - u \rangle + \|u' - u - \delta\zeta\| \tag{3.8}$$

for all u' sufficiently near u. Now let

$$\begin{aligned} a &:= \|u' - u\| \\ b &:= \|u' - u - \delta\zeta\| \\ \theta &:= \arccos\left(\frac{\langle \zeta, u' - u \rangle}{\|u' - u\|}\right). \end{aligned}$$

Then (3.8) is equivalent to

$$\delta + \sigma a^2 \geq a\cos(\theta) + b. \tag{3.9}$$

By the law of cosines,

$$b^2 = a^2 + \delta^2 - 2\delta a \cos(\theta),$$

and so

$$a\cos(\theta) = \frac{a^2 + \delta^2 - b^2}{2\delta}. \tag{3.10}$$

First suppose that $\delta \leq b$. Then (3.10) implies

$$\begin{aligned} a\cos(\theta) &= \frac{a^2}{2\delta} + \frac{(\delta + b)(\delta - b)}{2\delta} \\ &\leq \frac{a^2}{2\delta} + \delta - b. \end{aligned}$$

From this it follows that (3.9) holds with $\sigma = \frac{1}{2\delta}$. Now suppose that $\delta > b$. Then (3.10) implies that $\cos(\theta) \geq 0$ and

$$\frac{2\delta}{\delta + b} a\cos(\theta) = \frac{a^2}{\delta + b} + \delta - b. \tag{3.11}$$

Since

$$\frac{2\delta}{\delta + b} \geq 1,$$

we have

$$\begin{aligned} a\cos(\theta) + b &\leq \frac{2\delta}{\delta + b} a\cos(\theta) + b \\ &= \frac{a^2}{\delta + b} + \delta \\ &\leq \frac{a^2}{\delta} + \delta, \end{aligned}$$

which is (3.9) with $\sigma = \frac{1}{\delta}$. □

In view of part 1 of Lemma 3.3, we have the following.

Corollary 3.5. *Let $\zeta \in N_X^P(x)$ and $\|\zeta\| = 1$. Then*

$$\partial_P d_X(x + r\zeta) = \{\zeta\} \quad \forall r \in (0, 1].$$

We shall also require the following result.

Proposition 3.6.
(1) *Suppose that $u \notin X$, $\mathrm{proj}_X(u) \neq \phi$, and that d_X is Gâteaux differentiable at u. Then $\|d'_X(u)\| = 1$.*
(2) *Let $u \notin X$, and d_X be continuously differentiable at u. Then u admits a unique closest point $x \in X$, and*

$$d'_X(u) = \frac{u - x}{d_X(u)}.$$

Proof. In order to prove part 1, let $x \in X$ be a closest point to u, and let $v = x - u$. Then for small positive t one has $d_X(u + tv) = (1 - t)d_X(u)$. Consequently

$$\begin{aligned} d'_X(u; v) &= \lim_{t \downarrow 0} \frac{(1 - t)d_X(u) - d_X(u)}{t} \\ &= \lim_{t \downarrow 0} \frac{-t d_X(u)}{t} \\ &= -d_X(u) = -\|x - u\| = \langle d'_X(u), x - u \rangle. \end{aligned}$$

Since the distance function is Lipschitz of rank 1, we have $\|d'_X(u)\| \leq 1$. The Cauchy–Schwarz inequality therefore implies $\|d'_X(u)\| = 1$.

We now turn to part 2. The continuous differentiability assumption implies strict differentiablity at u; that is $\partial_L d_X(u) = \{d'_X(u)\}$. Then $d'_X(u)$ is the weak limit of a sequence ζ_i such that $\partial_P d(u_i) = \{\zeta_i\}$, where $u_i \to u$. But since Theorem 3.1 implies that $\zeta_i = d'_X(u_i)$ and d'_X is continuous, it follows that $d'_X(u)$ is actually the strong limit of ζ_i. Now note that

$$\zeta_i = \frac{u_i - x_i}{d_X(u_i)},$$

where $\mathrm{proj}_X(u_i) = \{x_i\}$. We then have

$$x := u - d_X(u) d'_X(u) = \lim x_i$$

and

$$d'_X(u) = \frac{u - x}{d_X(u)}.$$

Since the norm of this expression is 1 and $x \in X$, it follows that $x \in \mathrm{proj}_X(u)$. It remains to show that this x is the unique element in $\mathrm{proj}_X(u)$. Suppose that $x' \in \mathrm{proj}_X(u)$. Denote

$$u_\delta := u - \delta \frac{u - x'}{d_X(u)}.$$

Then

$$x' \in \mathrm{proj}_X(u_\delta) \quad \forall \delta \in (0, d_X(u)).$$

It follows that

$$\frac{u - x'}{d_X(u)} \in N^P_{X(d_X(u)-\delta)}(u_\delta).$$

Then by Theorem 3.4,

$$\left\{\frac{u - x'}{d_X(u)}\right\} = \partial_P d_X(u_\delta) = \{d'_X(u_{\delta)}\}.$$

Now letting $\delta \downarrow 0$, continuity of the derivative implies

$$d'_X(u) = \frac{u - x'}{d_X(u)};$$

that is, $x' = x$. □

For any $u \notin X$, we have

$$-d_X(u) = \sup\{-\|u - x\| : x \in X\},$$

and therefore, near u, the negative of the distance function is the pointwise supremum of a family of C^1 functions parametrized by x. However, attainment is not guaranteed (unless weak closedness of X or finite dimensionality of H is assumed). For this reason, known results on pointwise maximum functions (see Clarke [8], [5]) are not directly applicable in order to prove regularity of $-d_X$ on $H\backslash X$. However, the special nature of the distance function allows for an independent proof. This, and a bit more, is given in the next result.

Theorem 3.7. *Let $u \notin X$. Then the following hold:*
(1) *$-d_X$ is regular at u.*
(2) *d_X is regular at u if and only if d_X is strictly differentiable at u.*

Proof. Let $\zeta \in \partial_L d_X(u)$. Then there exist sequences u_i and ζ_i such that $u_i \to u$, $\zeta_i \xrightarrow{w} \zeta$, and

$$\partial_P d_X(u_i) = \{\zeta_i\} = \left\{\frac{u_i - x_i}{d_X(u_i)}\right\},$$

where $\mathrm{proj}_X(u_i) = \{x_i\}$. Now, for any $v \in H$,

$$\begin{aligned} d_X(u_i + tv) - d_X(u_i) &\leq \|u_i + tv - x_i\| - \|u_i - x_i\| \\ &= \frac{\|u_i + tv - x_i\|^2 - \|u_i - x_i\|^2}{\|u_i + tv - x_i\| + \|u_i - x_i\|} \\ &= \frac{2t\langle u_i - x_i, v\rangle + t^2\|v\|^2}{\|u_i + tv - x_i\| + \|u_i - x_i\|}. \end{aligned}$$

Therefore

$$\begin{aligned}\limsup_{t\downarrow 0}\frac{d_X(u+tv)-d_X(u)}{t} &= \limsup_{t\downarrow 0}\left[\lim\frac{d_X(u_i+tv)-d_X(u_i)}{t}\right]\\ &\le \limsup_{t\downarrow 0}\left[\lim\frac{2\langle u_i-x_i,v\rangle+t\|v\|^2}{\|u_i+tv-x_i\|+\|u_i-x_i\|}\right]\\ &= \langle\zeta,v\rangle.\end{aligned}$$

Since the C–subdifferential satisfies

$$\partial_C d_X(u)=\mathrm{clco}[\partial_L d_X(u)]=-\partial_C(-d_X)=(u),$$

it follows from the preceding that for all $v\in H$,

$$\begin{aligned}\liminf_{t\downarrow 0}\frac{-d_X(u+tv)+d_X(u)}{t} &\ge \sup\{\langle-\zeta,v\rangle:\zeta\in\partial_L(d_X)(u)\}\\ &= \max\{\langle-\zeta,v\rangle:\zeta\in\partial_C(d_X)(u)\}\\ &= \max\{\langle\zeta,v\rangle:\zeta\in\partial_C(-d_X)(u)\}\\ &= (-d_X)^o(u;v)\\ &= \limsup_{\substack{t\downarrow 0\\ w\to u}}\frac{-d_X(w+tv)+d_X(w)}{t}\\ &\ge \limsup_{t\downarrow 0}\frac{-d_X(u+tv)+d_X(u)}{t}.\end{aligned}$$

From this we obtain

$$(-d_X)^o(u;v)=\lim_{t\downarrow 0}\frac{-d_X(u+tv)+d_X(u)}{t},$$

which verifies the regularity of $-d_X$ at u.
As for the second part of the theorem's statement, we first recall that strict differentiability of d_X at u implies regularity. Now suppose that both d_X and $-d_X$ are regular at u, and let $\zeta\in\partial_C d_X(u)$. Then for all $v\in H$ one has

$$\langle\zeta,v\rangle\le d_X^o(u;v)=d'_X(u;v).$$

Since also $-\zeta\in\partial_C(-d_X)(u)$, we have

$$\langle-\zeta,v\rangle\le(-d_X)^o(u;v)=-d'_X(u;v).$$

It follows that

$$\langle\zeta,v\rangle=d_X^o(u;v)\quad\forall v\in H,$$

which implies that $\partial_C d_X(u)$ is the singleton $\{\zeta\}$. □

4. Proximally smooth sets

It is our primary aim in the present section to characterize when a closed set $X \subseteq H$ is *proximally smooth*; that is, when there exists $r > 0$ such that the distance function is continuously differentiable on a "tube" of the form

$$U(r) := \{u \in H : 0 < d_X(u) < r\}.$$

Let us note at the outset that it is possible for a closed nonconvex set to be proximally smooth, and that continuous differentiability of the distance function on the tube $U(r)$ does not imply the property on $U(r')$ for $r' > r$; both phenomena are easily illustrated by examples involving C^2 manifolds in finite dimensions.

The following is our main result characterizing proximal smoothness. Among other facts, it asserts the following one: $X \subseteq H$ is proximally smooth if and only if the P–subdifferential $\partial_P d_X(u)$ is nonempty for each $u \in U(r)$. We shall denote

$$Y(r) := \{u \in H : d_X(u) \geq r\}.$$

Theorem 4.1. *For given $r > 0$, the following are equivalent:*
(a) *d_X is continuously differentiable on $U(r)$.*
(b) $\text{proj}_X(u) \neq \phi \quad \forall u \in U(r)$ *and the Gâteaux derivative $d'_X(u)$ exists.*
(c) $\text{proj}_X(u) \neq \phi \quad \forall u \in U(r)$, *and for every $r' \in (0, r)$, one has*

$$d_X(u) + d_{Y(r')}(u) = r' \quad \forall u \in U(r').$$

(d) $\text{proj}_X(u) \neq \phi \quad \forall u \in U(r)$, *and every nonzero* P*–normal to X is realized by an r-ball.*
(e) *For every $r' \in (0, r)$ and $u \in H$ such that $d_X(u) = r'$, one has $N^P_{X(r')}(u) \neq \{0\}$.*
(f) $\partial_P d_X(u) \neq \phi \quad \forall u \in U(r)$.

Remark 4.2.
(1) $U(r) = \phi$ if and only if $X = H$, in which case all the statements are vacuously true; consequently H is proximally smooth.
(2) Since the distance function to the empty set is identically $+\infty$ by convention, the formula in part (c) implies that $Y(r') \neq \phi$ if X is a proximally smooth proper subset of H.

Proof of the theorem:
(a) $\Longrightarrow$ (b):

This follows from part 2 of Proposition 3.6.

(b) $\Longrightarrow$ (c):

By (b), $\text{proj}_X(u) \neq \phi$ for every $u \in U(r)$. It remains to show that for any given $r' \in (0, r)$,

$$d_X(u) + d_{Y(r')}(u) = r' \quad \forall u \in U(r'). \tag{4.1}$$

Fix $r' \in (0, r)$ and let $u \in U(r')$. We first claim that

$$d_X(u) + d_{Y(r')}(u) \geq r'. \tag{4.2}$$

For this purpose, we can without loss of generality assume that u admits a closest point in $Y(r')$, since this is so for a dense set of u in $U(r')$. Let $y \in \mathrm{proj}_{Y(r')}(u)$. Then, if x is a closest point in X to u,

$$r' = d_X(y) \le \|y - x\| \le \|u - x\| + \|y - u\| = d_X(u) + d_{Y(r')}(u),$$

which gives (4.2).

We now must show that

$$d_X(u) + d_{Y(r')}(u) \le r' \quad \forall r' \in (0, r). \tag{4.3}$$

Introduce the function

$$V(\alpha) := d_{Y(\alpha)}(u), \quad 0 < \alpha < r.$$

It is useful to express V as a value function:

$$V(\alpha) = \left\{ \begin{array}{rl} \inf\{\|u - z\| : d_X(z) \ge \alpha\} & \text{if } Y(\alpha) \neq \phi \\ +\infty & \text{otherwise} \end{array} \right\}, \quad 0 < \alpha < r. \tag{4.4}$$

Claim 4.3. *V is finite, nondecreasing, and Lipschitz of rank 1 on $(0, r)$.*

The nondecreasing property is immediate. Prior to proving the rest of the claim, let us observe that it implies (4.3) holds. Indeed, Claim 4.3 implies that for any $r' \in (0, r)$, we have

$$V(r') - V(d_X(u)) \le r' - d_X(u).$$

Then, since $V(r') = d_{Y(r')}(u)$ and $V(d_X(u)) = 0$, (4.3) immediately follows.

By part 1 of Proposition 3.6, $\|d'_X(u)\| = 1$. Denoting $\zeta = d'_X(u)$, one has

$$\begin{aligned} d_X(u + t\zeta) &= d_X(u) + \langle \zeta, t\zeta \rangle + o(t) \\ &= d_X(u) + t + o(t). \end{aligned}$$

Therefore for any sufficiently small $\varepsilon > 0$, there exists a point z_ε such that $\|z_\varepsilon - u\| < 2\varepsilon$ and $d_X(z_\varepsilon) > d_X(u) + \varepsilon$. From this it follows that

$$V(d_X(u) + \varepsilon) < 2\varepsilon,$$

and therefore

$$\lim_{\alpha \downarrow d_X(u)} V(\alpha) = 0. \tag{4.5}$$

Since $V(\alpha) = 0$ on $(0, d_X(u)]$, (4.5) implies that proving Claim 4.3 reduces to verifying the following.

Claim 4.4. *V is Lipschitz of rank 1 on $(d_X(u), r)$.*

Let us first make note of the fact that

$$V(\alpha) = \inf\{\|u - z\| : d_X(z) = \alpha\} \quad \forall \alpha \in (d_X(u), r). \tag{4.6}$$

To see this, consider a point z such that $d_X(z) > \alpha > d_X(u)$. Then there exists a point y in the segment (u, z) such that $d_X(y) = \alpha$. Since $\|u - y\| < \|u - z\|$, it follows that points z such that $d_X(z) > \alpha$ can be discarded in the definition of V given by (4.4), for α in the interval $(d_X(u), r)$. This verifies (4.6).

For $\alpha \in (d_X(u), r)$, consider the lower semicontinuous function defined by

$$W(\alpha) := \lim_{\alpha' \uparrow \alpha} V(\alpha').$$

In order to prove Claim 4.4, it suffices to show that W is Lipschitz of rank 1 on $(d_X(u), r)$. Let $\zeta \in \partial_P W(\alpha_o)$, where $\alpha_o \in (d_X(u), r)$. Note that since W is nondecreasing. Then, referring to [7], it follows that $\zeta \geq 0$, a fact required below. By another result in [7], the required Lipschitz behavior of W will follow upon verifying that $\|\zeta\| \leq 1$.

In view of the proximal subgradient inequality, there exist positive numbers σ and γ such that

$$W(\alpha) - W(\alpha_o) + \sigma(\alpha - \alpha_o)^2 \geq \zeta(\alpha - \alpha_o) \tag{4.7}$$

whenever $|\alpha - \alpha_o| < \gamma$. Then, since $V(\alpha) \geq W(\alpha)$ on $(d_X(u), r)$, for points z such that $d_X(z) = \alpha \in (d_X(u), r)$, one has

$$\|u - z\| + \sigma(d_X(z) - \alpha_o)^2 - \zeta d_X(z) \geq W(\alpha_o) - \zeta\alpha_o \tag{4.8}$$

if $|d_X(z) - \alpha_o| < \gamma$. Now let $\varepsilon \in (0, \frac{\gamma}{2})$, and choose z_o to satisfy $0 < \alpha_o - d_X(z_o) < \varepsilon$ and $|\|u - z_o\| - W(\alpha_o)| < \varepsilon$. Then, if $|d_X(z) - d_X(z_o)| < \frac{\gamma}{2}$, we have $|d_X(z) - \alpha_o| < \gamma$. Bearing in mind that $\zeta \geq 0$, we have

$$\begin{aligned}
&\|u - z\| + \sigma(d_X(z) - d_X(z_o)^2 - \zeta d_X(z) \\
&\quad = \|u - z\| + \sigma(d_X(z) - \alpha_o)^2 - \zeta d_X(z) + \sigma[(d_X(z) - d_X(z_o))^2 - (d_X(z) - \alpha_o)^2] \\
&\quad \geq W(\alpha_o) - \zeta\alpha_o + \sigma[(d_X(z) - d_X(z_o))^2 - (d_X(z) - \alpha_o)^2] \\
&\quad \geq \|u - z_o\| - \zeta d_X(z_o) - \varepsilon + \zeta(d_X(z_o) - \alpha_o) + \sigma[(d_X(z) - d_X(z_o))^2 - (d_X(z) - \alpha_o)^2] \\
&\quad \geq \|u - z_o\| - \zeta d_X(z_o) - \varepsilon - \varepsilon\zeta,
\end{aligned}$$

where in the last inequality we used the fact that

$$(d_X(z) - d_X(z_o))^2 - (d_X(z) - \alpha_o)^2 \geq 0,$$

as a straightforward calculation shows. Let us denote $\delta := \varepsilon + \varepsilon\zeta$. Then z_o is δ–optimal for minimizing the function

$$z \longrightarrow \|u - z\| + \sigma(d_X(z) - d_X(z_o))^2 - \zeta d_X(z)$$

over the set

$$S := \{z \in H : \|z - z_o\| \leq \frac{\gamma}{2}\},$$

since

$$z \in S \Longrightarrow |d_X(z) - d_X(z_o)| \leq \frac{\gamma}{2}.$$

By virtue of Ekeland's Theorem ([10], [5]), there exists z' such that $\|z' - z_o\| < \sqrt{\delta}$ and such that z' minimizes the function

$$h(z) := \|u - z\| + \sigma(d_X(z) - d_X(z_o))^2 - \zeta d_X(z) + \sqrt{\delta}\|z - z'\|$$

over S. We may choose ε sufficiently small to ensure that z' is in the interior of S; that is, such that $\sqrt{\delta} < \frac{\gamma}{2}$. Then a necessary condition for optimality of z' is

$$0 \in \partial_C h(z').$$

The calculus of the C–subdifferential ([5]) then yields

$$0 = w_1 + 2\sigma\sqrt{\delta}d'_X(z') - \zeta d'_X(z') + \sqrt{\delta}w_2,$$

where $\|w_i\| \leq 1$, $i = 1, 2$. Since $\|d'_X(u)\| = 1$ (by part 1 of Proposition 3.6) and δ is arbitrarily small, we conclude that $\|\zeta\| \leq 1$, as required.
(c) $\Longrightarrow$ (d):

In view of formula (2.1), it suffices to show that if $r' \in (0, r)$, then every nonzero P–normal to X is realized by an r'–ball. Fix $r' \in (0, r)$, let $x \in X$, $\zeta \in N_X^P(x)$, and $\|\zeta\| = 1$. Then there exists $\delta \in (0, r')$ such that

$$\text{proj}_X(x + r''\zeta) = \{x\} \quad \forall r'' \in (0, \delta).$$

For any $r'' \in (0, \delta)$, part 1 of Lemma 3.3 implies that

$$x + r''\zeta \in \text{proj}_{X(r'')}(x + \delta\zeta).$$

Therefore

$$\zeta \in N^P_{X(r'')}(x + r''\zeta).$$

In view of Theorem 3.4, we have $\partial_P d_X(x+r''\zeta) = \{\zeta\} =$. Then Theorem 3.1 implies that

$$\partial_P d_X(x + r''\zeta) = \partial^P d_X(x + r''\zeta) = \{\zeta\}.$$

Since the formula in (c) is an identity in u (locally), we obtain

$$\partial_P d_{Y(r')}(x + r''\zeta) = \{-\zeta\},$$

and so $\text{proj}_{Y(r')} = x+r''\zeta+\alpha\zeta$ for some $\alpha >= 0$. But since (c) holds and $d_X(x+r''\zeta) = r''$, it follows that $\alpha == r' - r''$, and therefore $x + r'\zeta$ is in the boundary of $Y(r')$. Hence $d_X(x + r'\zeta) = r'$, which shows that ζ is realized by an r'–ball.
(d) $\Longrightarrow$ (e):

Suppose that $d_X(u) = r'$ where $0 < r' < r$, and let $x \in \text{proj}_X(u)$. Then

$$\zeta := \frac{u - x}{r'} \in N_X^P(x),$$

and $\|\zeta\| = 1$. Hence (d) implies that

$$\text{proj}_X(u + \delta\zeta) = \{x\} \quad \forall \delta \in (0, r - r').$$

Therefore part 1 of Lemma 3.3 gives

$$u \in \text{proj}_{X(r')}(u + \delta\zeta) \quad \forall \delta \in (0, r - r'),$$

and so $N^P_{X(r')}(u) \neq \{0\}$, as required.

(e) $\Longrightarrow$ (f):

This follows immediately from Theorem 3.4.

(f) $\Longrightarrow$ (a):

By Theorems 2.1 and 3.1, for any $u \in U(r)$, one has

$$\partial_P d_X(u) = \{d'_X(u)\} = \left\{ \frac{u - x}{d_X(u)} \right\},$$

where now d'_X denotes the Fréchet derivative. The required continuity of the derivative follows from part 2 of Theorem 2.1. This completes the proof of the theorem. □

Remark 4.5. Note that if X is weakly closed, as when $H = R^n$, then in conditions (b)–(d), the requirement that $\text{proj}_X(u)$ be nonempty is automatically satisfied.

When the conditions of Theorem 4.1 hold for a given r, we shall say that X is *proximally smooth of radius* r. If X is convex, then it is weakly closed, and it follows that $\text{proj}_X(u) \neq \phi$ for all u. Furthermore, classical normals coincide with P–normals, and so (2.1) holds with 0 left hand side for every $x \in X$, and therefore for every $r > 0$. Hence X is proximally smooth of radius r for every $r > 0$. The following corollary asserts the converse.

Corollary 4.6. *A closed set $X \subseteq H$ is convex if and only if it is proximally smooth of radius r for every $r > 0$.*

Proof. The "only if" part has already been explained, so let us assume that X is proximally smooth for every $r > 0$. Let $u \notin X$. The proximal smoothness assumption implies that u admits a closest point in X, say $x =$. Furthermore, for every $r > 0$, we have

$$\frac{1}{2r}\|y - x\|^2 \geq \left\langle \frac{\zeta}{\|\zeta\|}, y - x \right\rangle \quad \forall y \in X,$$

where $\zeta := u - x$. But then

$$0 \geq \langle \zeta, y - x \rangle \quad \forall y \in X.$$

Since $\langle \zeta, u - x \rangle > 0$, we have shown that for each point $u \notin X$, there exists a hyperplane separating u from X. This is equivalent to convexity of X. □

A further result in the same vein is the following. It follows directly from Corollary 4.6 and Theorem 4.1.

Corollary 4.7. *A closed set $X \subseteq H$ is convex if and only if it satisfies the following two conditions for every $u \in H \backslash X$:*
(1) $\mathrm{proj}_X(u) \neq \phi$.
(2) *The Gâteaux derivative* $d'_X(u)$ *exists.*

The next result asserts that proximal smoothness of a closed set $X \subseteq H$ implies Lipschitz behavior of both the (necessarily single valued) metric projection proj_X and the (Fréchet = Gâteaux) derivative d'_X near X. In particular, the distance function to a closed set X has the feature that continuity of d'_X on a tube $U(r)$ implies that d'_X must in fact be locally Lipschitz on $U(r)$.

Theorem 4.8. *Let X be proximally smooth, with $r > 0$ as in* Theorem 4.1. *Then the following hold:*
(1) *Let $r' \in (0, r)$. Then* proj_X *is Lipschitz of rank $\frac{r}{r-r'}$ on $U(r')$.*
(2) *For every u and w in $U(r)$ one has*

$$\langle u - w, \mathrm{proj}_X(u) - \mathrm{proj}_X(w) \rangle \geq 0. \tag{4.9}$$

(3) *d'_X is locally Lipschitz on $U(r)$.*

Proof. Let u and w be points in the tube $U(r')$. Then

$$\left\langle \frac{u - \mathrm{proj}_X(u)}{d_X(u)}, \mathrm{proj}_X(w) - \mathrm{proj}_X(u) \right\rangle \leq \frac{1}{2r} \|\mathrm{proj}_X(w) - \mathrm{proj}_X(u)\|^2 \tag{4.10}$$

and

$$\left\langle \frac{w - \mathrm{proj}_X(w)}{d_X(w)}, \mathrm{proj}_X(u) - \mathrm{proj}_X(w) \right\rangle \leq \frac{1}{2r} \|\mathrm{proj}_X(w) - \mathrm{proj}_X(u)\|^2. \tag{4.11}$$

Since $d_X(u) \leq r'$ and $d_X(w) \leq r'$, (4.10) and (4.11) yield

$$\langle u - \mathrm{proj}_X(u), \mathrm{proj}_X(w) - \mathrm{proj}_X(u) \rangle \leq \frac{r'}{2r} \|\mathrm{proj}_X(w) - \mathrm{proj}_X(u)\|^2 \tag{4.12}$$

and

$$\langle \mathrm{proj}_X(w) - w, \mathrm{proj}_X(w) - \mathrm{proj}_X(u) \rangle \leq \frac{r'}{2r} \|\mathrm{proj}_X(w) - \mathrm{proj}_X(u)\|^2, \tag{4.13}$$

respectively. Upon adding the inequalities (4.12) and (4.13), we obtain

$$\langle u - w + \mathrm{proj}_X(w) - \mathrm{proj}_X(u), \mathrm{proj}_X(w) - \mathrm{proj}_X(u) \rangle \leq \frac{r'}{r} \|\mathrm{proj}_X(w) - \mathrm{proj}_X(u)\|^2,$$

and therefore

$$\langle u - w, \mathrm{proj}_X(w) - \mathrm{proj}_X(u) \rangle \leq \frac{r' - r}{r} \|\mathrm{proj}_X(w) - \mathrm{proj}_X(u)\|^2. \tag{4.14}$$

Consequently,

$$\|\mathrm{proj}_X(w) - \mathrm{proj}_X(u)\|^2 \leq \frac{r}{r-r'}\langle w-u, \mathrm{proj}_X(w) - \mathrm{proj}_X(u)\rangle.$$

An application of the Cauchy–Schwarz inequality then yields

$$\|\mathrm{proj}_X(w) - \mathrm{proj}_X(u)\| \leq \frac{r}{r-r'}\|w-u\|. \tag{4.15}$$

Since u and w were arbitrary points in $U(r')$, part 1 of the assertion follows. Part 2 is a consequence of (4.14), while part 3 follows readily from part 1 and the fact that the formula

$$d'_X(u) = \frac{u - \mathrm{proj}_X(u)}{d_X(u)}$$

holds for every $u \in U(r)$. □

Remark 4.9. The above results have some bearing on the issue of extending a multifunction defined on X and which is Lipschitz there (in the Hausdorff sense), beyond X. In the single–valued case, this is well understood; global extensions exist and require no smoothness properties of X, even in infinite dimensions; see Hiriart–Urruty [11]. The multivalued case is more delicate, however. When X is proximally smooth, a local extension to $X(r')$ is provided by $\tilde{F}(u) = F(\mathrm{proj}_X(u))$, in view of Theorem 4.8; furthermore, the extension has Lipschitz rank arbitrarily close to that of F on X, depending upon how small we choose r'. Responding to a conjecture of the authors, Azé and Horvath have recently shown that in finite dimensions, global Lipschitz extensions do exist [1], but whose ranks increase as a function of the dimension; the infinite dimensional case remains unsettled.

In Theorem 4.8, if X is additionally assumed to be convex, then r may be taken arbitrarily large, and (4.10), (4.11) hold with 0 on the right hand sides. We then readily obtain the following.

Corollary 4.10. *Let X be a closed and convex subset of H. Then on H, proj_X is Lipschitz of rank* 1, (4.9) *holds, and d'_X is locally Lipschitz.*

The consequences of Theorem 4.1 can be strengthened under the additional hypothesis that X be weakly closed.

Theorem 4.11. *Assume that X is weakly closed and let $r > 0$. Then conditions* (a)–(f) *of* Theorem 4.1 *are equivalent to*
(g) $\mathrm{proj}_X(u)$ *is a singleton for every $u \in U(r)$.*

Proof. Since the other conditions imply (g), we only need to prove that (g) implies (a). To this end, we shall posit (g), and show that d_X has a continuously varying strict derivative.

Given $u \in U(r)$, let $\zeta \in \partial_L d_X(u)$. Then

$$\frac{u_i - x_i}{d_X(u_i)} \xrightarrow{w} \zeta, \tag{4.16}$$

where u_i is a sequence in X converging to u, and x_i is the unique closest point in X to u_i. The weak closedness of X then implies that

$$x_i \xrightarrow{w} x := u - d_X(u)\zeta \in X, \tag{4.17}$$

and so

$$\zeta = \frac{u - x}{d_X(u)}. \tag{4.18}$$

Since (4.16) implies $\|\zeta\| \leq 1$, we conclude that $x \in \operatorname{proj}_X(u)$. Since this x is unique, it follows that ζ is the sole element in $\partial_L d_X(u)$. Since $u \in U(r)$ was arbitrary, d_X is strictly differentiable on $U(r)$. We denote this derivative by d'_X. Letting u_i be a sequence in $U(r)$ such that $u_i \to u \in U(r)$, we wish to show that $d'_X(u_i) \to d'_X(u)$. We have that for each i,

$$d'_X(u_i) = \frac{u_i - x_i}{d_X(u_i)},$$

where $\operatorname{proj}_X(u_i) = \{x_i\}$. Likewise,

$$d'_X(u) = \frac{u - x}{d_X(u)},$$

where $\operatorname{proj}_X(u) = \{x\}$. Therefore we need to show that $x_i \to x$. In fact, it is enough to show that $x_i \xrightarrow{w} x$, for then we would have

$$\frac{u_i - x_i}{d_X(u_i)} \xrightarrow{w} \frac{u - x}{d_X(u)},$$

and since both sides of this expression are of norm 1, it would follow that the convergence is strong, implying x_i converges to x strongly as well, as required. Now, by way of contradiction, suppose that x was not the weak limit of x_i. Since this sequence is norm bounded, by Alaoglu's theorem there exists a subsequence so that $x_{i'} \xrightarrow{w} \hat{x} \neq x$, where $\hat{x} \in X$ by the weak closedness assumption. But since

$$\frac{u_{i'} - x_{i'}}{d_X(u_{i'})} \xrightarrow{w} \frac{u - \hat{x}}{d_X(u)}$$

and each term on the left is of norm 1, we have

$$\left\| \frac{u - \hat{x}}{d_X(u)} \right\| \leq 1,$$

and so the norm in fact equals 1. Consequently $\hat{x} \in \operatorname{proj}_X(u)$, which violates the uniqueness of x as the closest point in X to u. □

In a similar spirit to Corollary 4.6, we then have the following result.

Corollary 4.12. *A weakly closed set $X \subseteq H$ is convex if and only if* $\operatorname{proj}_X(u) =$ *is a singleton for every $u \notin X$.*

Remark 4.13.

(1) A longstanding conjecture (still open to our knowledge) would assert the statement of Corollary 4.12 with the a priori assumption of weak closedness replaced by strong closedness.

(2) Theorems 4.8 and 4.11 together imply that if $X \subseteq H$ is weakly closed, then the metric projection proj_X is single valued on $U(r)$ if and only if it is Lipschitz on $U(r')$ for any $r' \in (0, r)$, with the Lipschitz rank being $\frac{r}{r-r'}$.

(3) It is interesting to note that it is possible, even in finite dimensions, for a point $u \notin X$ to have a unique closest point in X, while $\partial_P d_X(u) = \phi$. Let $X = \mathrm{epi}(f)$ where $f : R \to R$ is given by

$$f(x) = \sqrt{1 - x^2} - 1 + x^6.$$

The point $(0, -1)$ admits the origin as a unique closest point in X. However, every point of the form $(0, -y)$ with $y > 1$ admits multiple closest points. Consequently, $N^P_{X(1)}(0, -1) = \{0\}$, and Theorem 3.4 implies that $\partial_P d_X(0, -1) = \phi$.

We now specialize still further, to the finite dimensional case.

Corollary 4.14. *Suppose that $H = R^n$ and let $r > 0$. Then conditions* (a)–(g) *in* Theorems 4.1, 4.11 *are equivalent to each of the following:*

(h) *The Gâteaux derivative $d'_X(u)$ exists for each $u \in U(r)$.*

(i) *d_X is strictly differentiable on $U(r)$.*

(j) *d_X is regular on $U(r)$.*

Proof. The equivalence of (g),(h) and (i) follows readily from results in Clarke [8]; see also Chapter 2 of Clarke [5]. Then the equivalence of (j) to these conditions follows from Theorem 3.7. □

Some geometric consequences of proximal smoothness in finite dimensions are summarized in the next corollary.

Corollary 4.15. *Let $H = R^n$ and assume that $r > 0$ is such that the equivalent conditions* (a)–(f) *hold. Then the following hold:*

(1) *For every $x \in X$, one has*

$$N^P_X(x) = N^L_X(x) = N^C_X(x).$$

(2) *For each $r' \in (0, r)$ and each u such that $d_X(u) = r'$, one has*

$$N^P_{X(r')}(u) = N^L_{X(r')}(u) = N^C_{X(r')}(u) = \{\alpha(u - x) : \alpha \geq 0\},$$

where x is the unique closest point to u in X.

(3) *X is regular, as is each set $X(r')$ for $0 < r' < r$.*

(4) *For each $r' \in (0, r)$, the boundary of $X(r')$ is a C^1–manifold.*

Proof. Part 1 will follow upon verifying that $N^L_X(x) \subseteq N^P_X(x)$ for each $x \in \mathrm{bdry}(X)$. To this end, let us assume that $0 \neq \zeta \in N^L_X(x)$ for a boundary point x. Then for a sequence of boundary points $x_i \to x$ there is a corresponding sequence $\zeta_i \in N_X^P(x_i)$ such that $\zeta_i \to \zeta$. In view of (2.1), for each i we have

$$\frac{1}{2r}\|y - x_i\|^2 \geq \left\langle \frac{\zeta_i}{\|\zeta_i\|}, y - x_i \right\rangle \quad \forall y \in X.$$

Then

$$\frac{1}{2r}\|y - x\|^2 \geq \left\langle \frac{\zeta}{\|\zeta\|}, y - x_i \right\rangle \quad \forall y \in X,$$

which shows that $\zeta \in N_X^P(x)$.

In order to verify part 2, first note that in view of condition (d) of Theorem 4.1, part 1 of Lemma 3.3 implies that for each $r' \in (0, r)$, every nonzero P–normal to $X(r')$ is realized by an $(r - r')$–ball. Hence each $X(r')$ is itself proximally smooth, and so part 1 and Theorem 2.1 yield part 2 of the assertion.
Part 3 follows from parts 1 and 2, while. Part 4 follows from condition (a) of Theorem 4.1 (continuous differentiability of d_X on $U(r)$) and the fact that on $U(r)$, $\|d'_X(u)\| = 1$ (by part 1 of Proposition 3.6). □

Remark 4.16. Suppose that $H = R^n$.

(1) Part 1 of the preceding corollary implies that if X is proximally smooth, then

$$N_X^P(x) \neq \{0\} \quad \forall x \in \text{bdry}(X), \tag{4.19}$$

since the C–normal cone to a boundary point of a closed subset of R^n is always nonzero. (In a general Hilbert space setting, however, one can have $N_X^P(x) = \{0\}$ for a boundary point x, even for closed and convex X.)

(2) In general, the C–normal cone is not a closed multifunction, unless one makes extra hypotheses on X. One such condition is that X be *epi–Lipschitz*; that is, locally, X is the image under an isometry of the epigraph of a Lipschitz function. (See Rockafellar [17] and section 7.3 of Clarke [5].) The preceding corollary shows that a different condition implying the closedness of the multifunction N_X^C—and N_X^P too—is that X be proximally smooth; this is a consequence of the closedness of the multifunction N_X^L. Let us note, however, that (4.19) alone does not guarantee the P–normal cone mapping's closedness. Consider for example $X = \text{epi}(f)$ for the function $f : R \to R$ given by

$$f(x) = \begin{cases} -|x|^{3/2} & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$$

Here (4.19) holds, but one can readily confirm that $N_X^P(0, 0)$ is not closed.

(3) The following example shows that closedness of the multifunction N_X^P is not sufficient for proximal smoothness of an epi–Lipschitz set X. Choose sequences r_n and x_n such that $r_n > 0$, $r_n \downarrow 0$, $x_n \in (0, 1)$, and such that the intervals $(x_n - r_n, x_n + r_n)$ are all disjoint and contained in $(0, 1)$. Then set

$$a_n := \frac{\varepsilon_n r_n}{\sqrt{1 + \varepsilon_n^2}},$$

where $\varepsilon_n > 0$ and $\varepsilon_n \downarrow 0$. The intervals $(x_n - a_n, x_n + a_n)$ are all disjoint and contained in $(0, 1)$ too. Now take $X \subseteq R^2$ to be the epigraph of the following function:

$$f(x) = \begin{cases} \sqrt{r_n^2 - (x - x_n)^2} - \sqrt{r_n^2 - a_n^2} & \text{if } \|x - x_n\| \le a_n \\ 0 & \text{otherwise} \end{cases}$$

Then in $(x_n - a_n, x_n + a_n)$, we have

$$f'(x) = \frac{-(x - x_n)}{\sqrt{r_n^2 - (x - x_n)^2}}.$$

Hence $|f'(x)| \le \varepsilon_n$ whenever $|x - x_n| \le a_n$, implying that f is Lipschitz. Then, upon noting that

$$N_X^P(0, 0) = \{(0, y) : y \le 0\},$$

it is readily checked that the multifunction N_X^P is closed. However, X is not proximally smooth, since near the origin, the circles employed in the construction have arbitrarily small radii.

5. Connection to the lower–C^2 property for Lipschitz functions

In this section we continue to assume that $H = R^n$. Let us introduce a basic hypothesis:

(A) (a) $U \subseteq R^n$ is open, convex and bounded.
    (b) $f : U \to R$ is Lipschitz.

Under these assumptions, we shall say that f is *lower–C^2 on U* provided that there exist $\sigma > 0$, a compact set S (in some topological space V) and continuous functions $b : S \to R^n$, $c : S \to R$, such that

$$f(x) = \max_{s \in S} \left\{ -\sigma \|x\|^2 + \langle b(s), x \rangle + c(s) \right\} \quad \forall x \in U. \tag{5.1}$$

In this case we shall also say that f is σ–lower–C^2 on U. This is in fact a local form of the lower–C^2 property, as can be seen from the proof of Theorem 6 in Rockafellar [17], who investigated this class of functions as one having favorable properties for optimization. The following characterization is required as well; it leads to the connection between the lower–C^2 property and proximal smoothness.

Theorem 5.1. *Let hypothesis* (A) *hold. Then the following are equivalent for a given* $\sigma > 0$*:*

(a) *f is σ–lower–C^2 on U.*

(b) *For each $x \in U$, there exists $\zeta \in R^n$ such that*

$$f(y) \ge -\sigma \|y - x\|^2 + \langle \zeta, y - x \rangle + f(x) \quad \forall y \in U. \tag{5.2}$$

(c) *For each $x \in U$,* (5.2) *holds for all $\zeta \in \partial_P f(x)$.*

Proof.
(a) $\Longrightarrow$ (b):
Let $x \in U$, and let $\hat{s} \in S$ be such that

$$f(x) = -\sigma\|x\|^2 + \langle b(\hat{s}), x\rangle + c(\hat{s}). \tag{5.3}$$

Since

$$f(y) \geq -\sigma\|y\|^2 + \langle b(\hat{s}), y\rangle + c(\hat{s}) \quad \forall y \in U,$$

a simple calculation shows that (5.3) yields (5.2) with

$$\zeta = -2\sigma x + b(\hat{s}).$$

(b) $\Longrightarrow$ (c):
Let $x_0 \in U$ and $\zeta_0 \in \partial_P f(x_0)$. Then there exists $\alpha > 0$ so that

$$f(y) \geq p(y) := -\alpha\|y - x_0\|^2 + \langle\zeta_0, y - x_0\rangle + f(x_0) \tag{5.4}$$

for all y sufficiently near x_0. Since hypothesis (A) implies that f is bounded below on U, it readily follows that α can be taken large enough to ensure that the quadratic p satisfies

$$f(y) > p(y) \quad \forall y \in U,\ y \neq x_0. \tag{5.5}$$

Let

$$\tilde{\alpha} = \inf\{\alpha > 0 : (5.5) \text{ holds}\}.$$

If $\tilde{\alpha} \leq \sigma$, then (5.2) holds and there is nothing further to show. So we shall suppose that $\tilde{\alpha} > \sigma$, and let $0 < \varepsilon < \tilde{\alpha} - \sigma$. Also, let the quadratic p be as in (5.4), but with $\alpha = \tilde{\alpha}$. Note that

$$f(y) \geq p(y) \quad \forall y \in U.$$

Since $\tilde{\alpha}$ is an infimum, there exists $x_1 \in U$ different from x_0 such that

$$p(x_1) + \varepsilon\|x_1 - x_0\|^2 > f(x_1). \tag{5.6}$$

For $0 \leq t \leq 1$, let

$$x_t = tx_1 + (1-t)x_0.$$

By the present assumption that (b) holds, there exists $\zeta_t \in R^n$ and $c_t \in R$ so that

$$q_t(y) := -\sigma\|y - x_0\|^2 + \langle\zeta_t, y - x_0\rangle + c_t$$

satisfies

$$q_t(x_t) = f(x_t)$$

and

$$q_t(y) \leq f(y) \quad \forall y \in U.$$

That is, we have chosen ζ_t so that $\zeta_t - 2\sigma(x_t - x_0) \in \partial_P f(x_t)$ and c_t so that $c_t = q_t(x_0)$. Let $\overline{p}$ and $\overline{q}_t$ denote the restrictions of p and q_t to the line segment $[x_0, x_1]$; that is, for $s \in [0,1]$ we set

$$\overline{p}(s) = -\tilde{\alpha}s^2\|x_1 - x_0\|^2 + s\langle\zeta_0, x_1 - x_0\rangle + f(x_0)$$

and

$$\overline{q}_t(s) = -\sigma s^2 \|x_1 - x_0\|^2 + s\langle \zeta_t, x_1 - x_0\rangle + c_t.$$

If we define

$$w_t(s) := \overline{q}_t(s) - \overline{p}(s) - \varepsilon s^2 \|x_1 - x_0\|^2,$$

then

$$w_t(0) = c_t - f(x_0) \le 0,$$
$$w_t(1) \le f(x_1) - p(x_1) - \varepsilon \|x_1 - x_0\|^2 \le 0,$$

and

$$\frac{d^2}{ds^2} w_t(s) = 2(-\sigma + \tilde{\alpha} - \varepsilon)\|x_1 - x_0\| < 0.$$

Thus w_t does not attain its maximum in $(0,1)$, and therefore must satisfy $w_t(s) \le 0$ for all $s \in [0,1]$; that is

$$\overline{q}_t(s) \le \overline{p}(s) + \varepsilon s^2 \|x_1 - x_0\|^2 \quad \forall t \in (0,1), \quad \forall s \in [0,1]. \tag{5.7}$$

Upon noting that $\overline{p}(t) \le f(x_t) = q_t(x_t) = \overline{q}_t(t)$ and letting $s = t$ in (5.7), we obtain

$$\overline{p}(t) \le \overline{q}_t(t) \le \overline{p}(t) + \varepsilon t^2 \|x_1 - x_0\|^2. \tag{5.8}$$

Hence $c_t \to f(x_0)$ as $t \downarrow 0$. Since $c_t = \overline{q}_t(0) \le f(x_0)$, we also see from the first inequality in (5.8) that

$$\limsup_{t \downarrow 0} \langle \zeta_t, x_1 - x_0\rangle \ge \langle \zeta_0, x_1 - x_0\rangle. \tag{5.9}$$

Thus we have

$$\begin{aligned}
f(x_1) &\ge \limsup_{t\downarrow 0} \overline{q}_t(1) \\
&= \limsup_{t\downarrow 0} [-\sigma \|x_1 - x_0\|^2 + \langle \zeta_t, x_1 - x_0\rangle + c_t] \\
&> -(\tilde{\alpha} - \varepsilon)\|x_1 - x_0\|^2 + \langle \zeta_0, x_1 - x_0\rangle + f(x_0) \\
&= p(x_1) + \varepsilon \|x_1 - x_0\|^2.
\end{aligned}$$

This violates (5.6), and therefore $\tilde{\alpha} \le \sigma$.

(c) $\Longrightarrow$ (a):

Let $x \in U$ and $\zeta \in \partial_P f(x)$ be such that (5.2) holds. We then have

$$f(y) = \max_{(x,\zeta)\in\Gamma} \left\{-\sigma \|y - x\|^2 + \langle \zeta, y - x\rangle + f(x)\right\} \quad \forall y \in U,$$

where

$$\Gamma := \{(x,\zeta) : x \in U,\ \zeta \in \partial_P f(x)\};$$

this set is nonempty by the density of P–subgradients. Denote the Lipschitz rank of f on U by K. Then

$$\sup\{\|\zeta\| : \zeta \in \partial_P f(x),\ x \in U\} \le K.$$

Therefore cl(Γ) is compact, and

$$f(y) = \max_{(x,\zeta)\in \mathrm{cl}(\Gamma)} \left\{ -\sigma\|y-x\|^2 + \langle \zeta, y-x\rangle + f(x) \right\} \quad \forall y \in U.$$

This is readily seen to be a representation of f which is of the desired form (5.1). □

In order to establish the relationship between the lower–C^2 property for functions and the proximal smoothness of sets, it is convenient to extend th= e definition of proximal smoothness to sets that are merely locally closed. (Note that this property is adequate for defining $N_X^P(x)$.) We say that the locally clos= ed set X is proximally smooth of radius $r > 0$ if formula (2.1) holds for all $x \in \mathrm{bdry}(X)$ and all nonzero $\zeta \in N_X^P(x)$. This is precisely condition (d) of Theorem 4.1, and hence if X is closed, w= e recover the prior sense of proximal smoothness. In general, however, there = is a qualitative difference, especially as regards the existence of closest points; for example, the open ball B is locally closed and proximally smooth, but no point outside B admits a closest point in B.

The connection between the lower–C^2 property for Lipschitz functions and proximal smoothness of their epigraphs is given in the following result.

Theorem 5.2. *Let hypothesis* (A) *hold. Then the following are equivalent:*
(a) *f is lower–C^2 on U.*
(b) *There exists $\sigma > 0$ such that for every $x \in U$, one has*

$$\sigma\|(y,\alpha) - (x, f(x))\|^2 \geq \left\langle \frac{(\zeta,-1)}{\|(\zeta,-1)\|}, (y,\alpha) - (x, f(x)) \right\rangle \tag{5.10}$$

for all $\zeta \in \partial_P f(x)$, $y \in U$, $\alpha \geq f(y)$.
(c) epi(f) *is proximally smooth.*
Furthermore, if f is lower–C^2 on U, then one has

$$\partial_P f(x) = \partial_L f(x) = \partial_C f(x) \quad \forall x \in U, \tag{5.11}$$

and in particular, f is regular on U.

Proof. We first will prove that (a) and (b) are equivalent. Suppose that (a) holds, and let f be σ–lower–C^2 on U. We will show that (b) holds. For each $x \in U$ and $\zeta \in \partial_P f(x)$, the inequality (5.2) holds. This implies that

$$\alpha - f(x) + \sigma\left[\|y-x\|^2 + (\alpha - f(y))^2\right] \geq \langle \zeta, y-x\rangle \quad \forall y \in U, \quad \forall \alpha \geq f(y).$$

It follows that

$$\sigma\|(y,\alpha) - (x, f(x))\|^2 \geq \langle (\zeta,-1), (y,\alpha) - (x, f(x))\rangle \tag{5.12}$$

for all $\zeta \in \partial_P f(x)$, $y \in U$, $\alpha \geq f(y)$, and therefore (5.10) holds, and (b) follows.

Now suppose that (b) holds. Let $x \in U$ and $\zeta \in \partial_P f(x)$. Continuing to denote the Lipschitz rank of f on U by K, we observe that (b) implies

$$(\sigma\sqrt{K^2+1})\|(y,\alpha) - (x, f(x))\|^2 \geq \langle (\zeta,-1), (y,\alpha) - (x, f(x))\rangle$$

for all $\zeta \in \partial_P f(x),\ y \in U,\ \alpha \geq f(y)$. The last inequality can be rewritten as

$$(\sigma\sqrt{K^2+1})\left[\|y-x\|^2 + \|f(y)-f(x)\|^2\right] \geq \langle \zeta, y-x \rangle - f(y) + f(x).$$

Since

$$\|f(y)-f(x)\|^2 \leq K^2\|y-x\|^2,$$

we therefore have

$$f(y) \geq -\sigma(1+K^2)^{\frac{3}{2}}\|y-x\|^2 + \langle \zeta, y-x\rangle + f(x) \quad \forall y \in U.$$

In view of the fact that $x \in U$ and $\zeta \in \partial_P f(x)$ were arbitrary, we have shown that f is $\sigma(1+K^2)^{\frac{3}{2}}$–lower–$C^2$ on U; that is, condition (a) holds and is equivalent to (b).

Since f is Lipschitz on U, for each $x \in U$ the proximal normal cone $N^P_{\mathrm{epi}(f)}(x, f(x))$ is generated by vectors of the form $(\zeta, -1)$ as ζ varies in the proximal subdifferential $\partial_P f(x)$. Consequently, (5.10) is equivalent to condition (d) of Theorem 4.1, with $X = \mathrm{epi}(f)$, which is locally closed.

The "furthermore" part of the assertion follows from Corollary 4.15. □

Remark 5.3. The equivalences provided by Theorems 5.1 and 5.2 go through with minor modifications if R^n is replaced by a general real Hilbert space H. At the outset, one then requires weak continuity of the functions $b(\cdot)$ and $c(\cdot)$, as well as weak compactness of the set S.

Acknowledgment. The authors benefited from useful discussions with Yu. S. Ledyaev.

References

[1] D. Azé and C. D. Horvath: A remark on extension of Lipschiz multifunctions. (to appear).

[2] J. M. Borwein and S. Fitzpatrick: Existence of nearest points in Banach spaces. Can. J. Math., 51:702–720, 1989.

[3] J. M. Borwein and J. R. Giles: The proximal normal formula in Banach spaces. Trans. Amer. Math. Soc., 302:371–381, 1987.

[4] F. H. Clarke: Methods of Dynamic and Nonsmooth Optimization. Volume 57 of CBMS-NSF Regional Conference Series in Applied Mathematics, S.I.A.M., Philadelphia, 1989.

[5] F. H. Clarke: Optimization and Nonsmooth Analysis. Wiley-Interscience, New York, 1983. Republished as Vol. 5 of Classics in Applied Mathematics, S.I.A.M., Philadelphia, 1990.

[6] F. H. Clarke, Yu. S. Ledyaev, and P. R. Wolenski: Proximal analysis and minimization principles. J. Math. Anal. Appl. to appear.

[7] F. H. Clarke, R. J. Stern, and P. R. Wolenski: Subgradient criteria for monotonicity, the Lipschitz condition, and convexity. Can. J. Math., 45:1167–1183, 1993.

[8] F.H. Clarke: Generalized gradients and applications. Trans. Amer. Math. Soc., 205:247–262, 1975.

[9] J. M. Danskin: The Theory of Max Min. Springer Verlag, New York, 1967.

[10] I. Ekeland: On the variational principle. J. Math. Anal. Appl., 47:324–353, 1974.

[11] J. B. Hiriart–Urruty: Extension of Lipschitz functions. J. Math. Anal. Appl., 77:539–554, 1980.

[12] J. B. Hiriart–Urruty: Generalized differentiability, duality and optimization for problems dealing with differences of convex functions. Lecture Notes in Economics and Mathematical Systems, 256:37–70, 1984.

[13] A. Ioffe: Approximate subdifferentials and applications II. Mathematika, 33:111–128, 1986.

[14] K. S. Lau: Almost Chebychev subsets in relexive Banach spaces. Indiana Univ. Math. J., 2:791–795, 1978.

[15] P. Loewen: Optimal Control via Nonsmooth Analysis. CRM Proceedings and Lecture Notes, Amer. Math. Soc., 1993.

[16] D. Preiss: Differentiability of Lipschitz functions on Banach spaces. J. Funct. Anal, 91:312–345, 1990.

[17] R. T. Rockafellar: Clarke's tangent cones and boundaries of closed sets in R^n. Nonlinear Analysis, 3:145–154, 1979.

[18] R. T. Rockafellar: Favorable classes of Lipschitz continuous functions in subgradient optimization. in Nondifferentiable Optimization, E. Nurminski, Ed., Permagon Press, New York, 1982.

[19] R. T. Rockafellar: The Theory of Subgradients and its Applications to Problems of Optimization: Convex and Nonconvex Functions. Heldermann Verlag, Berlin, 1981.

[20] J. E. Spingarn: Submonotone subdifferentials of Lipschitz functions. Trans. Amer. Math. Soc., 264:77–89, 1981.

[21] J.-P. Vial: Strong and weak convexity of sets and functions. Math. Ops. Res., 8:231–259, 1983.