Can. J. Math. Vol. **45** (6) 1993, pp. 1167–1183.

SUBGRADIENT CRITERIA FOR MONOTONICITY, THE LIPSCHITZ CONDITION, AND CONVEXITY

F. H. CLARKE, R. J. STERN AND P. R. WOLENSKI

ABSTRACT Let $f\ H \longrightarrow (-\infty, \infty]$ be lower semicontinuous, where H is a real Hilbert space An approach based upon nonsmooth analysis and optimization is used in order to characterize monotonicity of f with respect to a cone, as well as Lipschitz behavior and constancy The results, which involve hypotheses on the proximal subgradient $\partial^{\pi}f$, specialize on the real line to yield classical characterizations of these properties in terms of the Dini derivate They also give new extensions of these results to the multidimensional case A new proof of a known characterization of convexity in terms of proximal subgradient monotonicity is also given

1. **Introduction.** The calculus of Dini derivates has played an important role in the analysis of functions of a real variable, as evidenced by such works as those of Boas [1], Hobson [8], McShane [10], Riesz-Nagy [13], and Saks [15]. Prominent among the topics to which it has been applied are those of monotonicity and Lipschitz behavior. While these issues retain all their interest for functions of several variables, the corresponding results in derivate terms seem to a great extent undeveloped in this context.

The purpose of this article is to extend to several dimensions, and to Hilbert spaces as well, the classical Dini criteria for the functional properties mentioned above. Our approach is a novel one in this connection, and is inspired by the development of nonsmooth analysis, in which optimization and analysis have always gone hand-in-hand. It serves to unify in an efficient way a number of results, some of which are known and several of which are new.

A feature of the approach is that the results are most naturally couched in terms of the "proximal subgradients" of nonsmooth analysis; the Dini derivate-type versions are immediate corollaries. The presentation, however, is entirely self-contained.

Throughout the paper, H is a real Hilbert space. Suppose that $f: H \longrightarrow (-\infty, \infty]$ is lower semicontinuous and $x \in H$ is such that $f(x) < \infty$. An element $\xi \in H$ is said to be a *proximal subgradient* of f at x provided that there exists $\sigma > 0$ such that

$$f(y) - f(x) + \sigma\|y - x\|^2 \geq \langle \xi, y - x \rangle \tag{1.1}$$

for all y near x. The set of all proximal subgradients of f at x (which could be empty)

The first author was supported in part by the Natural Sciences Engineering Research Council of Canada
The first author was supported in part by le fonds FCAR du Quebec
Received by the editors August 24, 1992
AMS subject classification 26B05
Key words and phrases proximal subgradient, lower Dini derivate, cone monotonicity, Lipschitz behavior, constancy, convexity

is denoted by $\partial^\pi f(x)$. If $f(x) = \infty$, then it is said that $\partial^\pi f(x) = \phi$, by convention. (The terminology and notation is that of [5].)

In order to motivate the type of results we seek, as well as preview our basic proof technique, consider the following: Let $H = R^n$ and assume that $f: R^n \longrightarrow (-\infty, \infty]$ is bounded below and finite at least at one point. Suppose that there exists $C \geq 0$ such that for every $x \in R^n$ one has

$$\xi \in \partial^\pi f(x) \Longrightarrow \|\xi\| \leq C.$$[1]

Does it follow that f is Lipschitz of rank C? The answer is "yes", and a proof can be derived from some known (and rather deep) results from nonsmooth analysis. (It should be mentioned that the proof alluded to is not operative in a Hilbert space setting. This is because a key element in the argument is the equivalence of Lipschitz behavior of f near x and boundedness of the Clarke generalized gradient $\partial f(x)$, a fact which need not hold in an infinite dimensional setting.) Here we sketch a new proof that is not only more direct, but also lends itself to an infinite dimensional generalization: Suppose $x, y \in R^n$ and $\varepsilon > 0$. Define $g_\varepsilon: R^n \longrightarrow [0, \infty)$ by

$$g_\varepsilon(z) := (C + \varepsilon)\|z - y\|.$$

Then $f + g_\varepsilon$ is lower semicontinuous and is readily seen to attain its minimum on R^n, say at z_ε. If $z_\varepsilon \neq y$, then g is differentiable at z_ε and

$$\|g'(z_\varepsilon)\| = C + \varepsilon. \tag{1.2}$$

Since z_ε is a minimizer, it follows directly from the definition of the proximal subgradient that

$$0 \in \partial^\pi (f + g_\varepsilon)(z_\varepsilon). \tag{1.3}$$

By a proximal calculus fact (see Lemma 2.2 below) one can show that (1.3) implies

$$-g'_\varepsilon(z_\varepsilon) \in \partial^\pi f(z_\varepsilon). \tag{1.4}$$

But we have assumed that the proximal subgradients of f are bounded by C, and thus (1.2) and (1.4) cannot both hold. Therefore $z_\varepsilon = y$, and we have

$$\begin{aligned} f(y) &= f(y) + g_\varepsilon(y) \\ &\leq f(x) + g_\varepsilon(x) \\ &= f(x) + (C + \varepsilon)\|y - x\|. \end{aligned}$$

The inequality above holds because y is the minimizer of $f + g_\varepsilon$. Upon switching the roles of x and y and letting $\varepsilon \longrightarrow 0$, it follows that f is Lipschitz of order C.

The argument in the previous paragraph fails in infinite dimensions only in the step where a minimizer of $f + g_\varepsilon$ is guaranteed to exist. It is at this point that we can invoke

[1] Of course, if $\partial^\pi f(x) = \phi$, then the implication holds vacuously, this logical convention will be in effect throughout the paper

the smooth variational principle of Borwein and Preiss [2], which asserts that a slightly perturbed function does admit a minimizer. The smoothness of the perturbation is the key feature which allows the proximal calculus lemma used above in (1.4) to still be operational. (For example, Ekeland's variational principle does not lead to useful information here.) We obtain in this way the Hilbert space version of the Lipschitz criterion above (see Theorem 3.6 below).

The case $C = 0$ deserves special mention. In this case, the conclusion reduces to f being constant. Clarke and Redheffer [7] have recently given a simple proof of this fact, but only in finite dimensions. However, the basic idea in [7] is behind much of the present paper, and can be summarized as follows: If g is C^{1+} and $f + g$ has a minimum at z, then

$$-g'(z) \in \partial^\pi f(z). \tag{1.5}$$

Under given hypotheses on $\partial^\pi f$, one then makes judicious choices of g in conjunction with (1.5), so as to deduce properties of f.

It should be noted that the device of introducing a smooth function g so that (1.5) holds is the only mechanism we use to generate proximal subgradients. In fact, other than the density of the set of points

$$\operatorname{dom}(\partial^\pi f) := \{x : \partial^\pi f(x) \neq \phi\}$$

in the set

$$\operatorname{dom}(f) := \{x : f(x) < \infty\},$$

(see Borwein and Preiss [2, Theorem 3.1] and Theorem 2.4 below), little is known about the "size" of $\operatorname{dom}(\partial^\pi f)$. This contrasts, for example, with theorems on the size of the set of points where a Lipschitz function is differentiable (*e.g.* Rademacher's Theorem; see also Preiss [12]), or with results on Darboux-like properties of Dini derivates (see *e.g.* Bruckner [3, Chapter 11]).

There are other subgradients that one might wish to consider as well, but we next show that the proximal subgradient is the best possible for the goals of this paper. Suppose that $m\colon [0,\infty) \longrightarrow [0,\infty)$ is a *modulus function*; that is, m is continuous, nondecreasing, and $m(0) = 0$. We say that $\xi \in H$ is an *m-subgradient* of f at x provided that there exists $\sigma > 0$ such that

$$f(y) - f(x) + \sigma\|y - x\| m(\|y - x\|) \geq \langle \xi, y - x\rangle$$

for all y near x. We denote the set of such ξ by $\partial^m f(x)$. Evidently, $\partial^\pi f(x) \subset \partial^m f(x)$ for all modulus functions m which satisfy

$$\liminf_{t\downarrow 0} \frac{m(t)}{t} > 0. \tag{1.6}$$

On the other hand, if m is a modulus function such that (1.6) fails to hold, then possibly $\partial^m f(x) = \phi$ for every x, as it is for example with $f(x) = -\|x\|^2$. Consequently, ∂^π is

https://doi.org/10.4153/CJM-1993-065-x Published online by Cambridge University Press

the smallest (in terms of graph inclusion) among all m-subgradient maps ∂^m for which $\mathrm{dom}(\partial^m f)$ is dense in $\mathrm{dom}(f)$ for all lower semicontinuous functions f Furthermore, it follows that all our results involving conditions on the proximal subgradient $\partial^\pi f(x)$ have, as corollaries, corresponding ones in terms of the presubgradient $\hat{\partial} f(x)$ and the generalized gradient $\partial f(x)$ (see [5]), since both of these contain $\partial^\pi f(x)$ (and may be nonempty even when $\partial^\pi f(x) = \phi$) We shall not make these explicit in what follows

There is an extensive literature on the utilization of differential-type properties of functions in order to characterize their behavior See *e g* Saks [15], Boas [1], and Bruckner [3] Most of the literature is in dimension $n = 1$, the higher dimensional analogues are then usually reduced to this case An important and often-featured notion of generalized (directional) derivative is the *lower Dini derivate* $\underline{D}f$, which for $x \in \mathrm{dom}(f)$ is defined by

$$\underline{D}f(x, v) = \liminf_{t \downarrow 0} \frac{f(x + tv) - f(x)}{t}$$

There are several classical theorems with hypotheses involving $\underline{D}f$ which we generalize and/or derive as simple corollaries of our results involving hypotheses on $\partial^\pi f$ In order to give the flavor of one such result (appearing in Boas [1, p 128]), let $n = 1$ Then if f is continuous, one has

$$\underline{D}f(x, 1) \le 0 \quad \forall x \in R \Longleftrightarrow f \text{ nonincreasing on } R \tag{1 7}$$

We will see that (1 7) follows from Corollary 3 4 below, which is itself a consequence of our more general (cone) monotonicity result, Theorem 3 2 But actually, Theorem 3 2 brings something new even to the case $n = 1$, since

(a) only lower semicontinuity is assumed, f may be extended-valued, and

(b) the condition which characterizes f nonincreasing, namely

$$\xi \le 0 \quad \forall \xi \in \partial^\pi f(x) \tag{1 8}$$

only needs to be verified on the set $\mathrm{dom}(\partial^\pi f)$

The next section consists of preliminary material The results of §3 include the aforementioned monotonicity result, Theorem 3 2, and its consequences involving the characterization of Lipschitz and constant behavior in terms of proximal subgradients and the lower Dini derivate Then in §4 we employ our general technique in order to provide a new proof (and extensions to Hilbert space) of a result due to Poliquin [11], in which convexity of a function is characterized in terms of proximal subgradient monotonicity

2. **Preliminaries.** Our methods will utilize the following class of functions

DEFINITION 2 1 Suppose that U is an open subset of H and that $g\ U \longrightarrow R$ Then g is said to be C^{1+} on U provided that g is Fréchet differentiable on U, with the Frechet derivative map $x \longrightarrow g'(x)$ locally Lipschitz in x on U

The following lemma gives a proximal calculus sum rule, which says that the classical calculus sum rule holds whenever one of the functions is C^{1+}

https://doi.org/10.4153/CJM-1993-065-x Published online by Cambridge University Press

LEMMA 2.2. *Let U be an open subset of H. Suppose that $f: U \longrightarrow (-\infty, \infty]$ is lower semicontinuous, and let $x \in H$. Suppose further that g is C^{1+} on an open neighborhood of x. Then*

$$\xi \in \partial^{\pi}(f+g)(x) \Longrightarrow \xi - g'(x) \in \partial^{\pi} f(x). \tag{2.1}$$

PROOF. By the Mean Value Theorem and the Lipschitz assumption on g', there exists $M > 0$ such that

$$g(y) - g(x) \leq \langle g'(x), y - x\rangle + M\|y - x\|^2 \tag{2.2}$$

for all y near x. Since $\xi \in \partial^{\pi}(f+g)(x)$, we have

$$f(y) + g(y) - f(x) - g(x) + \sigma\|y - x\|^2 \geq \langle \xi, y - x\rangle \tag{2.3}$$

for some $\sigma > 0$ and all y near x. Upon combining (2.2) and (2.3), one arrives at

$$f(y) - f(x) + (M + \sigma)\|y - x\|^2 \geq \langle \xi - g'(x), y - x\rangle,$$

which says that $\xi - g'(x) \in \partial^{\pi} f(x)$. ∎

Lemma 2.2 provides a simple device for generating proximal subgradients. Suppose that f is lower semicontinuous, that g is C^{1+} near x, and assume that $f+g$ has a minimum at x. Since $0 \in \partial^{\pi}(f+g)(x)$ (a direct consequence of the definition), we obtain from Lemma 2.2 that $-g'(x) \in \partial^{\pi} f(x)$. In finite dimensions, this mechanism can be used directly to considerable effect. In infinite dimensions, however, minimizers of a lower semicontinuous function may no longer exist, and at first glance it might appear that this procedure is no longer useful. However, a theorem of Borwein and Preiss [2] provides a powerful tool for generating minimizers of a slightly perturbed function. We next state this theorem as it applies in Hilbert space.

THEOREM 2.3 (BORWEIN-PREISS). *Assume that $f: H \longrightarrow (-\infty, \infty]$ is lower semicontinuous and bounded below. Suppose that $\varepsilon > 0$ and $x_0 \in H$ are such that*

$$f(x_0) < \inf_{x \in H} f(x) + \varepsilon. \tag{2.4}$$

Then for all $\lambda > 0$ there exist $w \in H$ and $z \in H$ such that

(a) $\|x_0 - w\| < \lambda$, $\|w - z\| < \lambda$,

(b)

$$f(w) < \inf_{x \in H} f(x) + \varepsilon,$$

(c) $f(w) + (\frac{\varepsilon}{\lambda^2})\|w - z\|^2 < f(x) + (\frac{\varepsilon}{\lambda^2})\|x - z\|^2 \quad \forall x \in H,\ x \neq w.$

As an illustration of our basic technique, we offer a simple proof that a lower semicontinuous function on a Hilbert space possesses proximal subgradients on a set of points which is dense in its domain. The finite dimensional version of this result is relatively straightforward In infinite dimensions, Borwein and Preiss [2, Theorem 3 1] proved the result based on (the Banach space version of) Theorem 2 3 The present proof is given in order to illustrate the general proof technique which we will employ in the sequel (The open unit ball in H, centered at 0, is denoted by B)

THEOREM 2 4 *Let U be an open subset of H, and let f $U \longrightarrow (-\infty, \infty]$ be lower semicontinuous Then* $\mathrm{dom}(\partial^\pi f)$ *is dense in* $\mathrm{dom}(f)$

PROOF Let $x_0 \in \mathrm{dom}(f)$ Then there exists $\delta > 0$ such that f is bounded below on $\{x_0 + \delta B\} \subset U$ Define g_δ $H \longrightarrow [0, \infty]$ by

$$g_\delta(x) = \begin{cases} \frac{1}{\delta^2 - \|x - x_0\|^2} & \text{if } \|x - x_0\| < \delta \\ \infty & \text{otherwise} \end{cases}$$

If we interpret $(f + g_\delta)(x)$ to be ∞ for $x \notin \mathrm{dom}(g_\delta) = x_0 + \delta B$, we see that $f + g_\delta$ is an extended real-valued function which is lower semicontinuous and bounded below on H We now apply the Borwein-Preiss theorem Upon letting $\lambda = \varepsilon = 1$ in the theorem and writing $w = x_\delta,\ z = z_\delta$, we see that the function $\theta_\delta(x) = \|x - z_\delta\|^2$ is such that $f + g_\delta + \theta_\delta$ is minimized at x_δ, which is clearly in $\{x_0 + \delta B\}$ Since $0 \in \partial^\pi(f + g + \theta_\delta)(x_\delta)$, Lemma 2 2 implies that $-g'(x_\delta) - \theta'_\delta(x_\delta) \in \partial^\pi f(x_\delta)$ Now the fact that δ may be chosen arbitrarily small concludes the proof ∎

3. Monotonicity, Lipschitz behavior, and constancy.

A set $K \subset H$ is a *cone* if $x \in K$, $t \geq 0$ imply that $tx \in K$ The (negative) *polar* of a cone K is the set

$$K^* = \{y \in H \quad \langle x, y \rangle \leq 0 \quad \forall x \in K\}$$

DEFINITION 3 1 Suppose that U is an open convex subset of H, and let $K \subset H$ be a cone A function f $U \longrightarrow (-\infty, \infty]$ is said to be *K-nonincreasing* on U if

$$x, y \in U, \quad y \in x + K \Longrightarrow f(y) \leq f(x)$$

In the following new result, a characterization of the K-nonincreasing property is given in terms of the proximal subgradient

THEOREM 3 2 *Suppose that $K \subset H$ is a cone, $U \subset H$ is open and convex and f $U \longrightarrow (-\infty, \infty]$ is lower semicontinuous Then f is K-nonincreasing on U if and only if*

$$\partial^\pi f(x) \subset K^* \quad \forall x \in U \tag{3 1}$$

PROOF In order to prove the necessity of (3 1), let $x \in U$ and suppose that $\xi \in \partial^\pi f(x)$ Let $z \in K$ By assumption, we have that $f(x + tz) \leq f(x)$ for all $t > 0$ sufficiently small By the definition of $\partial^\pi f(x)$, there exists $\sigma > 0$ such that for all small $t > 0$ one has

$$0 \geq f(x + tz) - f(x) \geq t\langle \xi, z \rangle - \sigma t^2 \|z\|^2 \tag{3 2}$$

Dividing (3.2) by t and letting $t \longrightarrow 0$ leads to $\langle \xi, z \rangle \leq 0$. Since $z \in K$ is arbitrary, we conclude that $\xi \in K^*$; that is, (3.1) holds.

Let us now assume that (3.1) holds, and prove that f is K-nonincreasing on U. We shall assume that $K \neq \{0\}$, since the result is trivial otherwise. By translation of the data, it suffices to show that if $0, x \in K \cap U$, then $f(x) \leq f(0)$. Obviously, we only need to consider the case $f(0) < \infty$.

Let $0 \neq x \in K \cap U$, and note that

$$K^* \subset \{\xi \in H : \langle \xi, x \rangle \leq 0\}. \tag{3.3}$$

Let

$$P := \{p \in H : \langle p, x \rangle = 0\}$$

and consider the orthogonal decomposition

$$H = \operatorname{span}\{x\} \oplus P.$$

Our notation in regard to this decomposition will be somewhat flexible: If $q \in H$ decomposes as $q = tx + y$, then we write $q = (tx, y)$. Let $\varepsilon, \delta > 0$. For $(tx, y) \in H$, define

$$g_{\varepsilon,\delta}(tx, y) := \begin{cases} \frac{(t-1)^2}{1+\varepsilon-t} + \frac{\|y\|^2}{\delta^2 - \|y\|^2} & \text{if } 1 \leq t < 1 + \varepsilon,\ \|y\| < \delta \\ \frac{\delta^2(t-1)^2}{\delta+t} + \frac{\|y\|^2}{\delta^2-\|y\|^2} & \text{if } -\delta < t < 1,\ \|y\| < \delta \\ \infty & \text{otherwise}. \end{cases}$$

Then $g_{\varepsilon,\delta}$ is C^{1+} near any point where it is finite-valued, $g_{\varepsilon,\delta}(x) = 0$, and $g_{\varepsilon,\delta}(0) = \delta$. Also, it is readily verified that for all $0 < \delta, \eta < 1$, there exists $\eta' > 0$ (independent of $\varepsilon > 0$) such that

$$\langle g'_{\varepsilon,\delta}(tx, y), x \rangle \geq -\eta' \Longrightarrow t \geq 1 - \eta \tag{3.4}$$

whenever $\|y\| < \delta$.

Since f is lower semicontinuous and since U is open and convex, a straightforward covering argument implies the existence of an open bounded set $V \subset \bar{V} \subset U$ (with the bar denoting closure) such that V contains the (compact) set

$$\{tx : 0 \leq t \leq 1\},$$

and such that f is bounded below on V. Now assume that $\varepsilon, \delta > 0$ are taken sufficiently small so as to ensure that $\operatorname{dom}(g_{\varepsilon,\delta}) \subset V$. Upon regarding $(f + g_{\varepsilon,\delta})(x) = \infty$ for $x \notin \operatorname{dom}(g_{\varepsilon,\delta})$, we see that the function $f + g_{\varepsilon,\delta}$ thusly modified, is lower semicontinuous on H, bounded below on V, and has nonempty domain, since f is finite at $0 \in V$. Now, for each fixed $\delta > 0$, we apply the Borwein-Preiss theorem to $f + g_{\varepsilon,\delta}$. Specifically, for $\lambda = \varepsilon^{1/4}$, we write $x_{\varepsilon,\delta} = w$, $z_{\varepsilon,\delta} = z$, and take

$$\theta_{\varepsilon,\delta}(x) = \varepsilon^{1/2}\|x - z_{\varepsilon,\delta}\|^2.$$

https://doi.org/10.4153/CJM-1993-065-x Published online by Cambridge University Press

Here $x_{\varepsilon,\delta} \in \mathrm{dom}(g_{\varepsilon,\delta})$ is a minimizer of $f + g_{\varepsilon,\delta} + \theta_{\varepsilon,\delta}$ over H. Note that we have

$$\lim_{\varepsilon\to 0}\theta_{\varepsilon,\delta}(x_{\varepsilon,\delta}) = \lim_{\varepsilon\to 0}\|\theta'_{\varepsilon,\delta}(x_{\varepsilon,\delta})\| = 0 \tag{3.5}$$

uniformly in $\delta > 0$.

Lemma 2.2 implies that

$$-g'_{\varepsilon,\delta}(x_{\varepsilon,\delta}) - \theta'_{\varepsilon,\delta}(x_{\varepsilon,\delta}) \in \partial^{\pi} f(x_{\varepsilon,\delta}). \tag{3.6}$$

In view of (3.3), (3.6) and the assumption that (3.1) holds, one arrives at

$$\langle g'_{\varepsilon,\delta}(x_{\varepsilon,\delta}) + \theta'_{\varepsilon,\delta}(x_{\varepsilon,\delta}), x\rangle \geq 0. \tag{3.7}$$

Now let $\{\delta_i\}$ and $\{\eta_i\}$ be sequences of positive numbers approaching 0. By (3.4), for each i we can choose η'_i independently of ε so that

$$\Big\langle g'_{\varepsilon,\delta_i}\big((tx, y)\big), x\Big\rangle \geq -\eta'_i \Longrightarrow t \geq 1 - \eta_i \tag{3.8}$$

for all y such that $\|y\| < \delta_i$. In light of (3.5) and (3.7), we may choose $\varepsilon_i \downarrow 0$ and y_i, $\|y_i\| < \delta_i$, so that

$$\Big\langle g'_{\varepsilon_i,\delta_i}\big((t_{\varepsilon_i,\delta_i}x, y_i)\big), x\Big\rangle \geq -\eta'_i. \tag{3.9}$$

For notational convenience, let us now replace the subscript pair $(\varepsilon_i, \delta_i)$ simply by i, and write $x_i = (t_{\varepsilon_i,\delta_i}x, y_i)$. Then from (3.8) and (3.9), we have $t_i \geq 1 - \eta_i$. By recalling where g_i is finite, we must also have $t_i \leq 1 + \varepsilon_i$; therefore $x_i \longrightarrow x$. From the lower semicontinuity of f, the fact that x_i minimizes $f + g_i + \theta_i$ over H, and the convergence $\theta_i(0) \longrightarrow 0$ as $i \longrightarrow \infty$, we obtain

$$\begin{aligned} f(x) &\leq \liminf_{i\to\infty} f(x_i) \\ &\leq \liminf_{i\to\infty}\{f(x_i) + g_i(x_i) + \theta_i(x_i)\} \\ &\leq \liminf_{i\to\infty}\{f(0) + g_i(0) + \theta_i(0)\} \\ &= \liminf_{i\to\infty}\{f(0) + \delta_i + \theta_i(0)\} \\ &= f(0). \end{aligned}$$

This completes the proof. ∎

COROLLARY 3.3. *Under the assumptions of Theorem 3.2, f is K-nonincreasing on U if and only if f is $\overline{\mathrm{co}}(K)$-nonincreasing on U, where $\overline{\mathrm{co}}$ denotes closed convex hull.*

PROOF. This follows directly from the theorem, upon noting that

$$\{\overline{\mathrm{co}}(K)\}^* = K^*.$$

∎

The following corollary of Theorem 3.2 provides a characterization of cone monotonicity in terms of the lower Dini derivate.

COROLLARY 3.4. *Suppose that* $K \subset H$ *is a cone,* $U \subset H$ *is open and convex, and that* $f\colon U \longrightarrow (-\infty, \infty]$ *is lower semicontinuous. Then* f *is* K*-nonincreasing on* U *if and only if*

$$\underline{D}f(x; v) \leq 0 \quad \forall x \in U \cap \operatorname{dom}(f), \quad \forall v \in K. \tag{3.10}$$

PROOF. The necessity of (3.10) is clear from the definitions. Now assume that (3.10) holds. Also, let $x \in U \cap \operatorname{dom}(f)$ and suppose $\xi \in \partial^{\pi} f(x)$. Then there exists $\sigma \geq 0$ such that for each $v \in K$ we have

$$\frac{f(x+tv) - f(x)}{t} + t\sigma \|v\|^2 \geq \langle \xi, v \rangle \tag{3.11}$$

for all sufficiently small positive t. It follows that

$$0 \geq \underline{D}f(x; v) \geq \langle \xi, v \rangle.$$

From the arbitrariness of v, we obtain $\xi \in K^*$. The result follows by applying Theorem 3.2. ∎

Another Dini derivate characterization of cone monotonicity is given next. In spite of the simplicity of the proof, it is apparently new, even for $H = R^n$. Furthermore, unlike Corollary 3.4 (which it generalizes), its proof cannot be reduced to one dimensional arguments. (The conical hull of a set $\Lambda \subset H$ is denoted $\operatorname{cone}(\Lambda)$.)

COROLLARY 3.5. *Let the assumptions of Corollary 3.4 hold. Then a necessary and sufficient condition for* f *to be* K*-nonincreasing is that for each* $x \in U \cap \operatorname{dom}(f)$ *there exist a set* $\Lambda_x \subset H$ *such that* $K \subset \overline{\operatorname{co}}\big(\operatorname{cone}(\Lambda_x)\big)$ *and*

$$\underline{D}f(x; v) \leq 0 \quad \forall v \in \Lambda_x. \tag{3.12}$$

PROOF. The necessity is immediate, since we can take $\Lambda_x = K$. To prove the sufficiency, let $x \in U \cap \operatorname{dom}(f)$ and $\xi \in \partial^{\pi} f(x)$. Then there exists $\sigma > 0$ such that for each $v \in \Lambda_x$, (3.11) holds for all small $t > 0$. It follows that $\langle \xi, v \rangle \leq 0$ for all $v \in \Lambda_x$, which implies that $\xi \in K^*$. One now invokes Theorem 3.2. ∎

We will now turn our attention towards deriving subgradient and Dini derivate criteria for Lipschitz behavior and constancy. The following result appears to be new in the infinite dimensional case, although closely related ones have been given by Treiman [16]. See also Rockafellar [14] for an early result in finite dimensions.

THEOREM 3.6. *Let* U *be an open convex subset of* H*, and assume that* $f\colon U \longrightarrow (-\infty, \infty]$ *is lower semicontinuous, where* $\operatorname{dom}(f) \neq \phi$*. Let* $C \geq 0$*. Then* f *is Lipschitz of rank* C *on* U *if and only if*

$$\sup\{\|\xi\| : \xi \in \partial^{\pi} f(x)\} \leq C \quad \forall x \in U. \tag{3.13}$$

We are going to provide two proofs of this result The first proof makes direct use of Theorem 3 2 The second proof is independent, but has in common with the proof of Theorem 3 2 that it uses the general method of applying Theorem 2 3 in order to "approximately" minimize $f+g$, where g is appropriately chosen Also, the second proof can be readily extended to some Banach (not necessarily Hilbert) space settings

PROOF 1 First assume that f is Lipschitz on U of rank C Let $x \in U \cap \mathrm{dom}(f)$ and $\xi \in \partial^{\pi} f(x)$ Then there exists $\sigma \in R$ such that for each given unit vector $v \in H$, (3 11) holds for all $t > 0$ sufficiently small Letting $t \longrightarrow 0$, we obtain $C \geq \langle \xi, v \rangle$ Upon considering $v = \xi / \|\xi\|$, we obtain $\|\xi\| \leq C$, that is, (3 13) holds, and the "only if" part of the theorem is proven

We now turn to the "if" part of the result Let $x_0 \in U \cap \mathrm{dom}(f)$ We wish to show that when (3 13) holds, one has

$$f(y) - f(x_0) \leq C\|y - x_0\| \quad \forall y \in U \tag{3 14}$$

This will complete the proof, because the roles of x_0 and y are clearly interchangeable

By translation of the data, we can assume that $x_0 = 0$ Let $0 \neq y \in U$, and define the closed convex cone

$$K = \{ty \quad t \geq 0\}$$

We introduce the function $\tilde{f} \;\; U \longrightarrow (-\infty, \infty]$ given by

$$\tilde{f}(x) = f(x) - C\frac{\langle x, y\rangle}{\|y\|}$$

If $\tilde{\xi} \in \partial^{\pi}\tilde{f}(x)$, then by Lemma 2 2 one has

$$\tilde{\xi} = \xi - \frac{Cy}{\|y\|},$$

where $\xi \in \partial^{\pi} f(x)$ Then $\tilde{\xi} \in K^*$, since

$$\langle \tilde{\xi}, y\rangle = \langle \xi, y\rangle - C\|y\| \leq 0,$$

by virtue of the Cauchy-Schwartz inequality and the assumption that (3 13) holds By Theorem 3 2, this implies that $\tilde{f}$ is K-nonincreasing on U, which yields (3 14) with $x_0 = 0$ This completes the proof ■

PROOF 2 The proof of the "only if" part is the same as that given above We now wish to prove the "if" part Let $x_0 \in U$ be such that $f(x_0) < \infty$, and assume that (3 13) holds Let $N(x_0) = \{x_0 + 3\delta B\} \subseteq U$, where $\delta > 0$ Since f is lower semicontinuous, we may take δ sufficiently small to ensure that f is bounded below on $N(x_0)$ Let $M > C$ Choose $y \in \{x_0 + \delta B\}$, and define $g \;\; H \longrightarrow [0, \infty]$ by

$$g(x) = \begin{cases} M\|x - y\| & \text{if } \|x - y\| \leq \delta \\ M\|x - y\| + \frac{(\|x - y\| - \delta)^2}{2\delta - \|x - y\|} & \text{if } \delta \leq \|x - y\| \leq 2\delta \\ \infty & \text{otherwise} \end{cases}$$

https://doi.org/10.4153/CJM-1993-065-x Published online by Cambridge University Press

As in the proof of Theorem 3.2, we obtain that $f+g$ is lower semicontinuous and bounded below on H, upon taking $f+g$ to be ∞ on the complement of dom(g). Let $\varepsilon > 0$ be given. We now apply Theorem 2.3 (with $\lambda = \varepsilon^{1/4}$, as in the proof of Theorem 3.2) to the function $f+g$, and conclude that there exists a minimizer x_ε of the function $f+g+\theta_\varepsilon$ over $\{y+2\delta B\}$, where $\theta_\varepsilon = \varepsilon^{1/2}\|x-z_\varepsilon\|^2$. Now note that g is C^{1+} on the set $\{y+2\delta B\}\backslash\{y\}$, and that for x in this set we have $\|g'(x)\| \geq M > C$. Suppose that $x_\varepsilon \neq y$. Then from Lemma 2.2 we obtain

$$-g'(x_\varepsilon) - \theta'_\varepsilon(x_\varepsilon) \in \partial^\pi f(x_\varepsilon). \tag{3.15}$$

Condition (3.13) therefore implies

$$\|g'(x_\varepsilon)\| \leq C + \|\theta'_\varepsilon(x_\varepsilon)\|. \tag{3.16}$$

A contradiction results upon choosing ε so small that $\|\theta'_\varepsilon(x_\varepsilon)\| < M - C$. We conclude that $x_\varepsilon = y$, and obtain

$$(f+g+\theta_\varepsilon)(y) \leq (f+g+\theta_\varepsilon)(x) \quad \forall x \in \{x_0 + \delta B\}.$$

Upon letting $\varepsilon \longrightarrow 0$ and noting that $g(y) = 0$, one arrives at

$$f(y) \leq (f+g)(x) \leq f(x) + M\|x-y\| \quad \forall x \in \{x_0 + \delta B\}.$$

Since the roles of x and y may be reversed in the preceding statement, we can let $M \longrightarrow C$ and conclude that f is Lipschitz of rank C on $\{x_0 + \delta B\}$. Hence we have shown that f is Lipschitz of rank C near any point where f is finite.

It remains to extend this local Lipschitz condition to all of U. Let x and y be points in U, with $x \in$ dom(f), and consider the (compact) line segment $[x, y]$. We first claim that f is finite on the entire segment $[x, y]$. Suppose not. Then $0 < t^* < 1$, where

$$t^* := \sup\bigl\{t \in (0,1] : f\bigl(x + t(y-x)\bigr) < \infty\bigr\}.$$

Let $0 < t' < t^*$, and consider the segment $[0, t']$. Then a finite covering argument and the local Lipschitz property already verified combine to yield

$$f\bigl(x + t'(y-x)\bigr) \leq f(x) + C\|y-x\|.$$

Now, upon letting $t' \uparrow t^*$ and recalling that f is lower semicontinuous, we see that

$$f\bigl(x + t^*(y-x)\bigr) < \infty.$$

But then f is Lipschitz (and hence finite) near $x + t^*(y-x)$, and the definition of t^* is therefore violated. Hence f is finite on the entire segment $[x, y]$. A further finite covering argument then shows that

$$f(y) \leq f(x) + C\|y-x\|.$$

Upon reversing the roles of x and y, we obtain that f is Lipschitz on U. This completes the proof. ∎

In the following corollary, the local Lipschitz property is characterized in terms of the lower Dini derivate. We omit the proof.

COROLLARY 3 7 *Let $U \subset H$ be open and convex, and assume that f $U \longrightarrow (-\infty, \infty]$ is lower semicontinuous Then f is Lipschitz of rank C on U if and only if for every $x \in U \cap \mathrm{dom}(f)$ there exists a set $\Lambda_x \subset H$ such that* $\mathrm{cone}(\Lambda_x) = H$ *and*

$$\underline{D}f(x, v) \leq C\|v\| \quad \forall v \in \Lambda_x \tag{3 17}$$

To develop a weaker derivate criterion for Lipschitz behavior, we introduce the following notion A subset Λ of H is said to be a *bounding set* of rank r (for $r > 0$) provided that

$$r\langle \xi, v\rangle \leq 1 \quad \forall v \in \Lambda \Rightarrow \|\xi\| \leq 1$$

Note that the unit ball (open or closed) is a bounding set of rank 1

COROLLARY 3 8 *Let $U \subset H$ be open and convex, and assume that f $U \longrightarrow (-\infty, \infty]$ is lower semicontinuous Then f is Lipschitz of rank C on U if and only if for every $x \in U \cap \mathrm{dom}(f)$ there exist $r_x > 0$ and a bounding set Λ_x of rank r_x satisfying*

$$\underline{D}f(x, v) \leq C/r_x \quad \forall v \in \Lambda_x \tag{3 18}$$

PROOF The necessity is evident if f is Lipschitz of rank C, take $r_x = 1$ and Λ_x equal to the unit ball, (3 18) follows For the converse, assume that $C > 0$ and let $\xi \in \partial^\pi f(x)$ We derive (as in Corollary 3 5) that for any $v \in \Lambda_x$,

$$\langle \xi, v\rangle \leq \underline{D}f(x, v) \leq C/r_x,$$

or equivalently,

$$r_x\langle C^{-1}\xi, v\rangle \leq 1 \quad \forall v \in \Lambda_x$$

Since Λ_x is a bounding set of rank r_x, we deduce that $\|\xi\| \leq C$, and Theorem 3 6 implies that f is Lipschitz of rank C on U The case $C = 0$ follows from a limiting argument ∎

A result on constancy is given next, see Clarke [4, §6] and Clarke and Redheffer [7] for a finite dimensional version

COROLLARY 3 9 *Assume that $U \subset H$ is open, and let f $U \longrightarrow (-\infty, \infty]$ be lower semicontinuous Then the following hold*

(a) f is locally constant on U if and only if $\partial^\pi f(x) \subset \{0\}$ $\forall x \in U$

(b) f is locally constant on U if and only if for each $x \in U \cap \mathrm{dom}(f)$ there exists a set $\Lambda_x \subset H$, with $\mathrm{cone}(\Lambda_x) = H$, *such that $\underline{D}f(x, v) \leq 0$ $\forall v \in \Lambda_x$*

(c) Assume further that U is connected Then in parts (a) and (b) one may replace "locally constant on U" with "constant on U"

PROOF. Part (a) is a consequence of either Theorem 3.2 (take $K = H$) or Theorem 3.6 (take $C = 0$), while part (b) follows from either Corollary 3.5 (take $K = H$) or Corollary 3.7 (with $C = 0$). To prove part (c), let $x_0 \in U, f(x_0) = c$, and consider the set

$$S := \{x \in U : f(x) = c\}.$$

It follows that S is both open and closed with respect to U, and the connectedness assumption implies that $S = U$. ∎

REMARK 3.10. All our results involving Dini derivates go through with trivial modifications if we use the alternate definition

$$\underline{D}f(x; v) = \liminf_{\substack{t \downarrow 0 \\ u \to v}} \frac{f(x+tu) - f(x)}{t}.$$

(There is no distinction when f is Lipschitz near x.) This less classical derivate yields strengthened versions of all our conclusions involving directional derivatives.

4. **Convexity and proximal subgradient monotonicity.** We shall now apply our methods in order to prove a Hilbert space version of a result established by Poliquin [11] in a finite dimensional setting. A Banach space variant of the result (in terms of generalized, rather than proximal, subgradients) for locally Lipschitz functions is to be found in Clarke [6] (Proposition 2.2.9).

THEOREM 4.1. *Let $f: U \longrightarrow (-\infty, \infty]$ be lower semicontinuous, where U is an open convex subset of H. Then f is convex if and only if its proximal subgradient map is monotone that is, for every pair $x_1, x_2 \in U \cap \operatorname{dom}(\partial^\pi f)$ we have*

$$\xi_1 \in \partial^\pi f(x_1),\ \xi_2 \in \partial^\pi f(x_2) \Longleftrightarrow \langle x_2 - x_1, \xi_2 - \xi_1 \rangle \geq 0. \tag{4.1}$$

PROOF. The necessity of the monotonicity condition is well-known, since for a convex function, the proximal and classical subgradient of convex analysis are one and the same. For the sufficiency part of the argument, we shall proceed by supposing to the contrary that f is not convex. Failure of convexity implies the existence of points $a, b \in U \cap \operatorname{dom}(f)$ and a scalar $\lambda \in (0, 1)$ such that the point $c = \lambda a + (1 - \lambda b)$ satisfies

$$f(c) > \lambda f(a) + (1 - \lambda) f(b).$$

First assume that $f(c) < \infty$. For ease of notation later, let us take $c = 0$. Now replace f by $f - h$, where h is an affine function such that $h(a) = f(a)$ and $h(b) = f(b)$. We may do so without loss of generality, since this replacement affects neither convexity nor proximal subgradient monotonicity. Hence we are assuming the existence of $a, b \in U$ such that $f(a) = f(b) = 0$, where $0 \in U$ is contained in the line segment (a, b), and $f(0) > 0$.

Let us orthogonally decompose the space as

$$H = \operatorname{span}\{a\} \oplus M.$$

https://doi.org/10.4153/CJM-1993-065-x Published online by Cambridge University Press

With respect to this decompositon, we shall express a given vector $x = ta + y$ uniquely as $x = (ta, y)$. Also, let $k > 0$ be such that $b = -ka$.

For any given $\delta > 0$, define a function $g_\delta \colon H \longrightarrow [0, \infty]$ by

$$g_\delta(x) = g_\delta(ta, y) = \begin{cases} \frac{t^2\delta^2}{(1+\delta)^2 - t^2} + \frac{\|y\|^2}{\delta^2 - \|y\|^2} & \text{if } 0 \leq t < 1 + \delta,\ \|y\| < \delta \\ \frac{t^2\delta^2}{(k+\delta)^2 - t^2} + \frac{\|y\|^2}{\delta^2 - \|y\|^2} & \text{if } 0 \geq t > -k - \delta,\ \|y\| < \delta \\ \infty & \text{otherwise.} \end{cases}$$

Note that this function is convex, and is of class C^{1+} on the convex open set where it is finite.

For given $\gamma > 0$, define

$$f_\gamma(x) := f(x) + \frac{\gamma}{2}\|x\|^2.$$

Condition (4.1) and Lemma 2.2 together imply that for every pair $x_1, x_2 \in U$ we have

$$\xi_1 \in \partial^\pi f_\gamma(x_1), \quad \xi_2 \in \partial^\pi f_\gamma(x_2) \Longrightarrow \langle x_2 - x_1, \xi_2 - \xi_1 \rangle \geq \gamma \|x_2 - x_1\|^2. \tag{4.2}$$

By lower semicontinuity of f, a covering argument (as in the proof of Theorem 3.2) shows that there exists a bounded open set $S \subseteq \bar{S} \subseteq U$ containing $[a, b]$, such that f (and therefore f_γ) is bounded below on S. We now fix $\gamma > 0$ and apply the Borwein-Preiss theorem as before. For all sufficiently small $\delta > 0$ there exists a function $\theta^+_{\delta,\gamma}$ and a point $x^+_{\delta,\gamma}$ at which the minimum of the lower semicontinuous function $f_\gamma + g^+_\delta + \theta^+_{\delta,\gamma}$ is attained over S, where we have defined

$$g^+_\delta(x) = \begin{cases} g_\delta(x) = g_\delta(ta, y) & \text{if } t \geq 0 \\ \infty & \text{otherwise.} \end{cases}$$

The lower semicontinuity assumption on f implies that there exists $r > 0$ such that

$$\|x\| < r \Longrightarrow f(x) > \frac{f(0)}{2}. \tag{4.3}$$

Consider a sequence $\delta_\imath \longrightarrow 0$ as $i \longrightarrow \infty$. Upon noting that

$$g^+_{\delta_\imath}(a) = g^+_{\delta_\imath}(1a, 0) \longrightarrow 0 \text{ as } \delta_\imath \longrightarrow 0, \tag{4.4}$$

we can fix γ at a sufficiently small value, and arrange for $\theta^+_\imath := \theta^+_{\delta_\imath,\gamma}$ to be such that

$$(f_\gamma + g^+_{\delta_\imath} + \theta^+_\imath)(z^+_\imath) < \frac{f(0)}{2} \tag{4.5}$$

for all sufficiently large i, where we have denoted corresponding minimizers of $f_\gamma + g^+_{\delta_\imath} + \theta^+_\imath$ by

$$z^+_\imath = (t^+_\imath a, y^+_\imath).$$

We conclude from (4.3) and (4.5) that for each such i,

$$\|z^+_\imath\| > r. \tag{4.6}$$

Obviously,

$$(4.7) \qquad 0 \leq t_i^+ < 1 + \delta_i, \quad \|y_i^+\| < \delta_i.$$

Therefore for sufficiently large i we have

$$(4.8) \qquad -(g_{\delta_i}^+)'(z_i^+) - (\theta_i^+)'(z_i^+) \in \partial^\pi f(z_i^+).$$

One can now repeat the above discussion with b replacing a, and arrive at a point z_i^- analagous to z_i^+; in particular, (4.6)–(4.8) hold with the obvious modifications. We have

$$\langle -(g_{\delta_i}^+)'(z_i^+) - (\theta_i^+)'(z_i^+) - [-(g_{\delta_i}^-)'(z_i^-) - (\theta_i^-)'(z_i^-)], z_i^+ - z_i^- \rangle = A_i + B_i,$$

where

$$A_i := \langle -(g_{\delta_i})'(z_i^+) + (g_{\delta_i})'(z_i^-), z_i^+ - z_i^- \rangle,$$
$$B_i := \langle -(\theta_i^+)'(z_i^+) + (\theta_i^-)'(z_i^-), z_i^+ - z_i^- \rangle.$$

It is our goal to show that for sufficiently large i,

$$(4.9) \qquad A_i + B_i < \gamma \|z_i^+ - z_i^-\|^2,$$

In view of (4.8) and (4.2), this will provide the requisite contradiction. Since $b = -ka$, it readily follows from (4.6)–(4.7) (and the analogs for z_i^-) that for large i we have

$$(4.10) \qquad \|z_i^+ - z_i^-\|^2 > r^2.$$

Hence (4.9) will follow if we can bound each of the terms A_i and B_i above by $\gamma r^2/2$, for all large i. (Recall that the quantity $\gamma > 0$ has been fixed earlier in the argument.) The convexity of g_{δ_i} implies monotonicity of its Fréchet derivative, yielding $A_i \leq 0$. As for B_i, note that Theorem 2.3 (as we have applied it in our previous results) allows for a priori bounding of this term as required, since the minimizers z_i^+ and z_i^- are all contained in the bounded open set S. This completes the proof in case $f(0)$ is finite. In case $f(0) = \infty$, we can replace the quantity $f(0)/2$ at the point (4.3) in the preceding argument by any fixed positive number, and the same proof applies. ∎

REMARK 4.2. The intention throughout this paper has been to characterize certain properties of lower semicontinuous functions via conditions imposed on the proximal subgradients. While classical characterizations of monotonicity and Lipschitz properties relied on Dini derivate conditions, we have shown that the derivate results are straightforward corollaries of characterizations based on the more geometrical concept of proximal subgradient. A further advantage of a proximal subgradient condition is that its verification is required only at points where the subgradient is nonempty, whereas a derivate condition (generally) makes assumptions about all points. On the other hand, there are classical theorems on the line, concerning monotonicity, which allow for a countable set of exceptional points at which no hypothesis is made. (See for example [1, p. 128].) The approach put forth in this article can be adapted to recover such results on the line, but the larger question of systematically strengthening our multidimensional results by allowing exceptional sets of some kind, is essentially open.

REMARK 4.3. It has been pointed out to us by Philip Loewen and John Rowland that Theorem 3.2 can be derived from the generalized mean value theorem of Zagrodny [17]. New criteria for Lipschitz behavior have been derived via this route by Loewen [9]. We thank both Loewen and Rowland for these remarks and for some useful comments on an earlier version of this article. We also thank Jim Zhu for helpful discussions.

REFERENCES

1. R P Boas, *A Primer of Real Functions*, Carus Mathematical Monographs **13**, Mathematical Association of America, Rahway, 1960
2. J M Borwein and D Preiss, *A smooth variational principle with applications to subdifferentiability and to differentiability of convex functions*, Trans Amer Math Soc **303**(1987), 517–527
3. A M Bruckner, *Differentiation of Real Functions*, Springer-Verlag **659**, Berlin, 1978
4. F H Clarke, *An indirect method in the calculus of variations*, Trans Amer Math Soc , in press
5. ———, *Methods of Dynamic and Nonsmooth Optimization*, CBMS-NSF Regional Conference Series in Applied Mathematics **57**, S I A M , Philadelphia, 1989
6. ———, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, 1983, Republished as Classics in Applied Mathematics **5**, S I A M , Philadelphia, 1990
7. F H Clarke and R M Redheffer, *The proximal subgradient and constancy*, Canad Math Bull **36**(1993) 30–32
8. E W Hobson, *Functions of a Real Variable and the Theory of Fourier s Series*, Cambridge University Press, Cambridge, 1907
9. P Loewen, *Optimal Control via Nonsmooth Analysis*, CRM Lecture Notes Series, Amer Math Soc , Summer School on Control, CRM, Universite de Montréal, (1992), Amer Math Soc , Providence 1993
10. E J McShane, *Integration*, Princeton University Press, Princeton, 1944
11. R A Poliquin, *Subgradient monotonicity and convex functions*, Nonlinear Analysis **15**(1990), 305–317
12. D Preiss, *Differentiability of Lipschitz functions on Banach spaces*, J Funct Anal **91**(1990), 312–345
13. F Riesz and B Sz Nagy, *Functional Analysis*, Ungar, New York, 1955
14. R T Rockafellar, *Clarke's tangent cones and boundaries of closed sets in R^n*, Nonlinear Analysis **3**(1979), 145–154
15. S Saks, *Theory of the Integral*, Hafner, New York, 1937, L C Young, trans Reprinted by Dover Press 1964

16. J S Treiman, *Generalized gradients, Lipschitz behavior and directional derivatives*, Can J Math **37**(1985), 1074–1084
17. D Zagrodny, *Approximate mean value theorem for upper derivatives*, Nonlinear Analysis **12**(1988), 1413–1428

Centre de recherches mathématiques
Université de Montreal
Montreal, Québec
H3C 3J7

Department of Mathematics and Statistics
Concordia University
Montreal, Quebec
H4B 1R6

Department of Mathematics
Louisiana State University
Baton Rouge, Louisiana 70803
U S A