92

from: Progress in Nondifferentiable Optimization (E. Nurminski, ed.), IIASA Collaborative Proceedings Series, 1982. (Published by the Institute for Applied Systems Analysis, Laxenburg, Austria.)

FAVORABLE CLASSES OF LIPSCHITZ-CONTINUOUS FUNCTIONS IN SUBGRADIENT OPTIMIZATION

R. Tyrrell Rockafellar
Department of Mathematics
University of Washington
Seattle, Washington
USA

1. INTRODUCTION

A function $f : R^n \to R$ is said to be *locally Lipschitzian* if for each $x \in R^n$ there is a neighborhood X of x such that, for some $\lambda \geq 0$,

$$|f(x") - f(x')| \leq \lambda |x" - x'| \quad \text{for all } x' \in X,\ x" \in X \tag{1.1}$$

Examples include continuously differentiable functions, convex functions, concave functions, saddle functions and any linear combination or pointwise maximum of a finite collection of such functions.

Clarke (1975 and 1980), has shown that when f is locally Lipschitzian, the generalized directional derivative

$$f^\circ(x;v) = \limsup_{\substack{x' \to x \\ t \downarrow 0}} \frac{f(x'+tv) - f(x')}{t} \tag{1.2}$$

is for each x a finite, sublinear (i.e., convex and positively homogeneous) function of v. From this it follows by classical convex analysis that the set

$$\partial f(x) = \{y \in R^n | y \cdot v \leq f^\circ(x;v) \text{ for all } v \in R^n\} \tag{1.3}$$

is nonempty, convex, compact, and satisfies

$$f^{\circ}(x;v) = \max\{y\cdot v \mid y \in \partial f(x)\} \qquad \text{for all } v \in R^n. \qquad (1.4)$$

The elements of $\partial f(x)$ are what Clarke called "generalized gradients" of f at x, but we shall call them *subgradients*. As Clarke has shown, they are the usual subgradients of convex analysis when f is convex or concave (or for that matter when f is a saddle function). When f is continuously differentiable, $\partial f(x)$ reduces to the singleton $\{\nabla f(x)\}$.

In subgradient optimization, interest centers on methods for minimizing f that are based on being able to generate for each x at least one (but not necessarily every) $y \in \partial f(x)$, or perhaps just an approximation of such a vector y. One of the main hopes is that by generating a number of subgradients at various points in some neighborhood of x, the behavior of f around x can roughly be assessed. In the case of a convex function f this is not just wishful thinking, and a number of algorithms, especially those of bundle type (e.g., Lemarechal 1975 and Wolfe 1975) rely on such an approach. In the nonconvex case, however, there is the possibility, without further assumptions on f than local Lipschitz continuity, that the multifunction $\partial f : x \to \partial f(x)$ may be rather bizarrely disassociated from f. An example given at the end of this section has f locally Lipschitzian, yet such that there exist many other locally Lipschitzian functions g, not merely differing from f by an additive constant, for which $\partial g(x) = \partial f(x)$ for all x. Subgradients alone cannot discriminate between the properties of these different functions and therefore cannot be effective in determining their local minima.

Besides the need for conditions that imply a close connection between the behavior of f and the nature of ∂f, it is essential to ensure that ∂f has adequate continuity properties for the construction of "approximate" subgradients and in order to prove the convergence of various algorithms involving subgradients. The key seems to lie in postulating the existence of the ordinary directional derivatives

$$f'(x;v) = \lim_{t \downarrow 0} \frac{f(x+tv) - f(x)}{t} \qquad (1.5)$$

and some sort of relationship between them and ∂f. Mifflin (1977a and 1977b), most notably has worked in this direction.

In the present article we study the relationship between f' and ∂f for several special classes of locally Lipschitzian functions that suggest themselves as particularly amenable to computation. First we give some new results about continuity properties of f' when f belongs to the rather large class of functions that are "subdifferentially regular". Next we pass to functions f that are *lower*-C^k for some k, $1 \le k \le \infty$, in the following sense: for each point $\bar{x} \in R^n$ there is for some open neighborhood X of $\bar{x}$ a representation

$$f(x) = \max_{s \in S} F(x,s) \qquad \text{for all } x \in X, \qquad (1.6)$$

where S is a compact topological space and $F : X \times S \to R$ is a function which has partial derivatives up to order k with respect to x and which along with all these derivatives is continuous not just in x, but jointly in $(x,s) \in X \times S$. We review the strong results obtained by Spingarn (forthcoming) for lower-C^1 functions, which greatly illuminate the properties treated by Mifflin (1977b), and we go on to show that for $k \ge 2$ the classes of lower-C^k functions all coincide and have a simple characterization.

Before proceeding with this, let us review some of the existence properties of f' and continuity properties of ∂f that are possessed by any locally Lipschitzian function. This will be useful partly for background but also to provide contrast between such properties, which are not adequate for purposes of subgradient optimization, and the refinements of them that will be featured later.

Local Lipschitz continuity of a function $f : R^n \to R$ implies by a classical theorem of Rademacher (see Stein 1970) that for almost every $x \in R^n$, ∇f is differentiable at x, and moreover that the gradient mapping ∇f, on the set where it exists, is locally bounded.

Given any $x \in R^n$, a point where f may or not happen to be differentiable, there will in particular be in every neighborhood of x a dense set of points x' where $\nabla f(x')$ exists, and for any sequence of such points converging to x, the corresponding sequence of gradients will be bounded and have cluster points, each of which is, of course, the limit of some convergent subsequence. Clarke demonstrated in Clarke (1975) that $\partial f(x)$ is the convex hull of all such possible limits:

$$\partial f(x) = \text{co}\{\lim \nabla f(x') \mid x' \to x,\ f \text{ differentiable at } x'\}. \quad (1.7)$$

Two immediate consequences (also derivable straight from properties of $f^{\circ}(x;v)$ without use of Rademacher's theorem) are first that ∂f is *locally bounded*: for every x one has that

$$\bigcup_{x' \in X} \partial f(x') \text{ is bounded for some neighborhood } X \text{ of } x, \quad (1.8)$$

and second that ∂f is *upper semicontinuous* in the strong sense:

$$\text{for any } \varepsilon > 0 \text{ there is a } \delta > 0 \text{ such that}$$
$$\partial f(x') \subset \partial f(x) + \varepsilon B \text{ whenever } |x' - x| \leq \delta \quad , \quad (1.9)$$

where

$$B = \text{closed unit Euclidean ball} = \{x \mid |x| \leq 1\} \quad . \quad (1.10)$$

The case where $\partial f(x)$ consists of a single vector y is the one where f is *strictly* differentiable at x with $\nabla f(x) = y$, which by definition means

$$\lim_{\substack{x' \to x \\ t \downarrow 0}} \frac{f(x' + tv) - f(x')}{t} = y \cdot v \quad \text{for all } v \in R^n . \quad (1.11)$$

This is pointed out in Clarke (1975). From (1.7) it is clear that this property occurs if and only if x belongs to the domain of ∇f, and ∇f is continuous at x relative to its domain.

We conclude this introduction with an illustration of the abysmal extent to which ∂f could in general, without assumptions beyond local Lipschitz continuity, fail to agree with ∇f on the domain of ∇f and thereby lose contact with the local properties of f.

<u>Counterexample</u>

There is a Lipschitzian function $f : R^n \to R$ such that

$$\partial f(x) = [-1,1]^n \qquad \text{for all } x \in R^n \quad . \quad (1.12)$$

To construct f, start with a measurable subset A of R such that for every nonempty open interval $I \subset R$, both $\text{mes}\{A \cap I\} > 0$ and $\text{mes}\{A \setminus I\} > 0$. (Such sets do exist and are described in most texts on Lebesgue measure.) Define $h : R \to R$ by

$$h(t) = \int_0^t \theta(\tau)\, d(\tau) , \qquad \text{where } \theta(t) = \begin{cases} 1 & \text{if } t \in A , \\ -1 & \text{if } t \in A . \end{cases}$$

Since $\|\theta\|_\infty = 1$, h is Lipschitzian on R with Lipschitz constant $\lambda = 1$. Hence $h'(t)$ exists for almost every t, and $|h'(t)| \leq 1$. In fact $h' = \theta$ almost everywhere, from which it follows by the choice of A that the sets $\{t \mid h'(t) = 1\}$ and $\{t \mid h'(t) = -1\}$ are both dense in R. Now let

$$f(x) = \sum_{i=1}^{n} h(x_i) \qquad \text{for } x = (x_1, \ldots, x_n) \quad .$$

Then f is Lipschitzian on R^n with gradient

$$\nabla f(x) = (h'(x_1), \ldots, h'(x_n))$$

existing if and only if $h'(x_i)$ exists for $i = 1, \ldots, n$. Therefore $\nabla f(x) \in [-1,1]^n$ whenever $\nabla f(x)$ exists, and for each of the corner points e of $[-1,1]^n$ the set $\{x \mid \nabla f(x) = e\}$ is dense in R^n. Formula (1.7) implies then that (1.12) holds.

Note that every translate $g(x) = f(x - a)$ has $\partial g \equiv \partial f$, because ∂f is constant, and yet $g - f$ may be far from constant.

2. SUBDIFFERENTIALLY REGULAR FUNCTIONS

A locally Lipschitzian function $f : R^n \to R$ is *subdifferentially regular* if for every $x \in R^n$ and $v \in R^n$ the ordinary directional derivative (1.5) exists and coincides with the generalized one in (1.2):

$$f'(x;v) = f^{\circ}(x;v) \qquad \text{for all } x, v.$$

Then in particular $f'(x;v)$ is a finite, subadditive function of v; this property in itself has been termed the *quasidifferentiability* of f at x by Pshenichnyi (1971).

THEOREM 1. (Clarke 1975). *If f is convex or lower-C^k on R^n for some $k \geq 1$, then f is not only locally Lipschitzian but subdifferentially regular.*

Clarke did not study lower-C^k functions as such but proved in Clarke (1975) a general theorem about the subgradients of "max functions" represented as in (1.6) with $F(x,s)$ not necessarily differentiable in x. His theorem says in the case of lower-C^k functions that

$$\partial f(x) = \operatorname{co} \{\nabla_x F(x,s) \mid s \in I(x)\} \tag{2.1}$$

where

$$I(x) = \arg \max_{s \in S} F(x,s) \quad . \tag{2.2}$$

It follows from this, (1.4), and the definition of subdifferential regularity, that

$$f'(x;v) = \max \{\nabla_x F(x,s) \cdot v \mid s \in I(x)\} \tag{2.3}$$

for lower-C^1 functions, a well known fact proved earlier by Danskin (1967).

The reader should bear in mind, however, that Theorem 1 says considerably more in the case of lower-C^k functions than just this.

By asserting the equality of f' and f°, it implies powerful things about the semicontinuity of f' and strict differentiability of f. We underline this with the new result which follows.

THEOREM 2. *For a function $f : R^n \to R$, the following are equivalent:*

(a) f is locally Lipschitzian and subdifferentially regular;

(b) $f'(x;v)$ exists finitely for all x, v, and is upper semicontinuous in x.

Proof.

(a) $\Rightarrow$ (b). This is the easy implication; since $f' = f^{\circ}$ under subdifferential regularity, we need only apply (1.4) and (1.3).

(b) $\Rightarrow$ (a). For any x' and v the function $Q(t) = f(x' + tv)$ has both left and right derivatives at every t by virtue of (b):

$$Q'_+(t) = f'(x' + tv ; v), \qquad Q'_-(t) = -f'(x' + tv ; -v) \tag{2.4}$$

Moreover, the upper semicontinuity in (b) implies that for any fixed x and v there is a convex neighborhood X of x and a constant $\lambda \geq 0$ such that

$$f'(x' + tv ; v) \leq \lambda \text{ and } -f'(x' + tv ; v) \geq -\lambda \text{ when } x' + tv \in X \tag{2.5}$$

Since Q has right and left derivatives everywhere and these are locally bounded, it is the integral of these derivatives (cf. Saks (1937)):

$$Q(t_1) - Q(t_0) = \int_{t_0}^{t_1} Q'_+(\tau)d\tau = \int_{t_0}^{t_1} Q'_-(\tau)d\tau$$

From this and (2.5) it follows that

$$|f(x' + tv) - f(x')| \leq \lambda t \text{ when } x' \in X, \ x' + tv \in X \ .$$

Thus the local Lipschitz property (1.1) holds as long as $x'' - x'$ is some multiple of a fixed v. To complete the argument, consider not just one v but a basis $v_1, \ldots, v_n$ for R^n.

Each $x \in R^n$ has convex neighborhoods X_i and constants $\lambda_i \geq 0$ such that

$$|f(x'+tv_i)-f(x')| \leq \lambda_i t \quad \text{when } x' \in X_i ,\ x'+tv_i \in X_i \quad (2.6)$$

Then there is a still smaller neighborhood X of x and a constant $\alpha \geq 0$ such that for $x' \in X$ and $x'' \in X$ one has

$$x'' = x' + t_1 v_1 + \ldots + t_n v_n$$

with x' and $x'+t_1v_1 \in X_1$, $x'+t_1v_1$ and $(x'+tv_1)+tv_2 \in X_2$, and so forth, and

$$|t_1| + \ldots + |t_n| \leq \alpha|x''-x'| \quad .$$

Then by (2.6)

$$|f(x'')-f(x')| \leq |f(x'+t_1v_1)-f(x')| + |f(x'+tv_1+tv_2)-f(x'+tv_1)| + \ldots$$

$$\leq \lambda_1 t_1 + \lambda_2 t_2 + \ldots + \lambda_n t_n$$

$$\leq (\lambda_1+\lambda_2+\ldots+\lambda_n)\alpha|x''-x'| \quad .$$

In other words, f satisfies the Lipschitz condition (1.1) with $\lambda = (\lambda_1 + \ldots + \lambda_n)\alpha$. Thus f is locally Lipschitzian.

We argue next that $f'(x;v) \leq f^\circ(x;v)$ for all x,v by (1.2), and therefore via (1.7) that

$$f^\circ(x;v) = \limsup_{x' \to x} f'(x';v) \quad . \qquad (2.7)$$

The "lim sup" in (2.7) is just $f'(x';v)$ under (b), so we conclude that $f'(x;v) = f^\circ(x;v)$. Thus (b) does imply (a), and the proof of Theorem 2 is complete. □

COROLLARY 1. *Suppose f is locally Lipschitzian and subdifferentially regular on R^n and let D be the set of all points where f happens to be differentiable. Then at each $x \in D$, f is in fact strictly differentiable. Furthermore, the gradient mapping is continuous relative to D.*

COROLLARY 2. *If f is locally Lipschitzian and subdifferentially regular on R^n, then ∂f is actually single-valued at almost every $x \in R^n$.*

These corollaries are immediate from the facts about differentiability of f that were cited in §1 in connection with formula (1.7). The properties they assert have long been known for convex functions but have not heretofore been pointed out as properties of all lower-C^k functions. They hold for such functions by virtue of Theorem 1.

COROLLARY 3. *Suppose f is locally Lipschitzian and subdifferentially regular on R^n. If g is another locally Lipschitzian function on R^n such that $\partial g = \partial f$, then g = f + const.*

Proof. By Corollary 2, ∂g is single-valued almost everywhere. Recalling that g is strictly differentiable wherever ∂g is single-valued, we see that at almost every $x \in R^n$ the function h = g - f is strictly differentiable with $\nabla h(x) = \nabla g(x) - \nabla f(x) = 0$. Since h is locally Lipschitzian, the fact that $\nabla h(x) = 0$ for almost all x implies h is a constant function. □

COROLLARY 4. *Suppose f is locally Lipschitzian and subdifferentially regular on R^n. Then for every continuously differentiable mapping $\xi: R \to R^n$, the function $Q(t) = f(\xi(t))$ has right and left derivatives $Q'_+(t)$ and $Q'_-(t)$ everywhere, and these satisfy*

$$Q'_+(t) = \limsup_{\tau \to t} Q'_+(\tau) = \limsup_{\tau \to t} Q'_-(\tau) \quad ,$$
$$\qquad (2.8)$$
$$Q'_-(t) = \liminf_{\tau \to t} Q'_+(\tau) = \liminf_{\tau \to t} Q'_-(\tau) \quad .$$

Proof. The function Q is itself locally Lipschitzian and subdifferentiably regular (cf. Clarke 1980). Apply Theorem 2 to Q, noting that $Q'_+(t) = Q'(t;1) = Q^\circ(t;1)$ and $Q'_-(t) = -Q'(t;-1) = -Q^\circ(t;-1)$, and hence also $\partial Q(t) = [Q'_-(t), Q'_+(t)]$. The reason $Q'_+(\tau)$ and $Q'_-(\tau)$ can appear interchangeably in (2.8) is that by specialization of (1.7) to Q, as well as the characterizations of Q'_+ and Q'_- just mentioned, one has

$$Q'_+(\tau) = \limsup_{\tau'\to\tau} Q'(\tau') \quad , \quad Q'_-(\tau) = \liminf_{\tau'\to\tau} Q'(\tau') \quad ,$$

where the limits in this case are over the values τ' where $Q'(\tau')$ exists. □

3. LOWER-C^1 FUNCTIONS AND SUBMONOTONICITY

The multifunction $\partial f : R^n \rightrightarrows R^n$ is said to be *monotone* if

$$(x'-x'')\cdot(y'-y'') \geq 0 \quad \text{whenever } y' \in \partial f(x'),\ y'' \in \partial f(x'') \tag{3.1}$$

This is an important property of long standing in nonlinear analysis, and we shall deal with it in §4. In this section our aim is to review results of Spingarn (forthcoming) on two generalizations of monotonicity and their connection with subdifferentially regular functions and lower-C^1 functions. The generalized properties are as follows: ∂f is *submonotone* if

$$\liminf_{\substack{x'\to x\\ y'\in\partial f(x')}} \frac{(x'-x)\cdot(y'-y)}{|x'-x|} \geq 0 \ , \quad \forall x,\ \forall y \in \partial f(x) \quad , \tag{3.2}$$

and it is *strictly submonotone* if

$$\liminf_{\substack{x'\to x\\ x''\to x\\ y'\in\partial f(x')\\ y''\in\partial f(x'')}} \frac{(x''-x')\cdot(y''-y')}{|x''-x'|} \geq 0 \ , \quad \forall x \quad . \tag{3.3}$$

To state the results, we adopt Spingarn's notation:

$$\partial f(x)_v = \{y \in \partial f(x) \mid (y'-y)\cdot v \leq 0,\ \forall y' \in \partial f(x)\} \tag{3.4}$$

Thus $\partial f(x)_v$ is a certain face of the compact convex set $\partial f(x)$, the one consisting of all the points y at which v is a normal vector. Let us also recall the notion of *semismoothness* of f introduced by Mifflin (1977): this means that

$$\begin{gathered}\text{whenever } x^j \to x,\ v^j \to v,\ t_j \downarrow 0,\ y^j \to y, \text{ with}\\ y^j \in \partial f(x^j + t_j v^j), \text{ then one has } y\cdot v = f'(x;v) \quad .\end{gathered} \tag{3.5}$$

THEOREM 3 (Spingarn (forthcoming)). *The following properties of a locally Lipschitzian function* $f: R^n \to R$ *are equivalent:*

(a) *f is both subdifferentially regular and semismooth;*

(b) *∂f is submonotone;*

(c) *∂f is directionally upper semicontinuous in the sense that for every* $x \in R^n$, $v \in R^n$ *and* $\varepsilon > 0$, *there is a* $\delta > 0$ *such that*

$$\partial f(x + tv') \subset \partial f(x)_v + \varepsilon B \quad \text{when } |v'-v| < \delta \text{ and } 0 < t < \delta \ . \tag{3.6}$$

THEOREM 4 (Spingarn (forthcoming)). *The following properties of a locally Lipschitzian function* $f: R^n \to R$ *are equivalent:*

(a) *f is lower* C^1;

(b) *∂f is strictly submonotone;*

(c) *∂f is strictly directionally upper semicontinuous in the sense that for every* $x \in R^n$, $v \in R^n$ *and* $\varepsilon > 0$, *there is a* $\delta > 0$ *that*

$$(y''-y')\cdot v' \geq -\varepsilon \text{ when } |x'-x| < \delta,\ |v'-v| < \delta,\ 0 < t < \delta,$$
$$y' \in \partial f(x') \text{ and } y'' \in \partial f(x'+tv'). \qquad (3.7)$$

Spingarn has further given a number of valuable counter-examples in his forthcoming paper. These demonstrate that

$$\partial f \text{ submonotone} \not\Rightarrow \partial f \text{ strictly submonotone}, \qquad (3.8)$$

$$f \text{ subdifferentially regular} \not\Rightarrow f \text{ lower-}C^1, \qquad (3.9)$$

$$f \text{ quasidifferentiable and semismooth} \not\Rightarrow f \text{ subdifferentially regular.} \qquad (3.10)$$

Comparing Theorems 3 and 4, we see that lower-C^1 functions have distinctly sharper properties than the ones of quasidifferentiability and semismoothness on which Mifflin, for instance, based his minimization algorithm (1977a). In perhaps the majority of applications of subgradient optimization the functions are actually lower-C^1, or even lower-C^∞. This suggests the possibility of developing improved algorithms which take advantage of the sharper properties. With this goal in mind, we explore in the next section what additional characteristics are enjoyed by lower-C^k functions for $k > 1$.

4. LOWER-C^2 FUNCTIONS AND HYPOMONOTONICITY

The properties of lower-C^k functions for $k \geq 2$ turn out, rather surprisingly, to be in close correspondence with properties of convex functions It is crucial, therefore, that we first take a look at the latter. We will have an opportunity at the same time to verify that convex functions are special examples of lower-C^∞ functions. The reader may have thought of this as obvious, because a convex function can be represented as a maximum of affine (linear-plus-a-constant) functions, which certainly are C^∞. The catch is, however, that a representation must be constructed in terms of affine functions which depend *continuously* on a parameter s ranging over a *compact* set, if the definition of lower-C^∞ is to be satisfied.

We make use now of the concept of monotonicity of ∂f defined at the beginning of §3.

THEOREM 5. *For a locally Lipschitzian function* $f: R^n \to R$, *the following properties are equivalent:*

(a) f *is convex ;*

(b) ∂f *is montone ;*

(c) for each $\bar{x} \in R^n$ *there is a neighborhood* X *of* $\bar{x}$ *and a representation of* f *as in* (1.6) *with* S *a compact topological space,* $F(x,s)$ *affine in* x *and continuous in* s .

Proof. (a) $\Rightarrow$ (c). In terms of the conjugate f^* of the convex function f, we have the formula

$$f(x) = \max_{y \in R^n} \{y \cdot x - f^*(y)\} \text{ for all } x, \qquad (4.1)$$

where the maximum is attained at y if and only if $y \in \partial f(x)$ (see Rockafellar 1970, §23). Any $\bar{x}$ has a compact neighborhood X on which ∂f is bounded. The set

$$S = \{(y,\beta) \in R^{n+1} \mid \exists x \in X \text{ with } y \in \partial f(x), \beta = y \cdot x - f(x)\}$$

is then compact, and we have as a special case of (4.1)

$$f(x) = \max_{(y,\beta) \in S} \{y \cdot x - \beta\} .$$

This is a representation of the desired type with $s = (y,\beta)$, $F(x,s) = y \cdot x - \beta$.

(c) $\Rightarrow$ (a). The representations in (c) imply certainly that f is convex relative to some neighborhood of each point. Thus for any fixed x and v the function $Q(t) = f(x+tv)$ has left and right derivatives Q'_- and Q'_+ which are nondecreasing in some neighborhood of each t. These derivatives are then nondecreasing relative to $t \in (-\infty,\infty)$, and it follows from this that

Q is a convex function on $(-\infty,\infty)$ (cf. Rockafellar 1970, §24) Since this is true for every x and v, we are able to conclude that f itself is convex.

(a) $\Rightarrow$ (b). This is well-known (cf. Rockafellar 1970, §24).

(b) $\Rightarrow$ (a). A direct argument could be given, but we may as well take advantage of Theorem 3. Monotonicity of ∂f trivially implies submonotonicity, so we know from Theorem 3 that f is subdifferentially regular. Fixing any x and v, we have by the monotonicity of ∂f that

$$((x+t''v)-(x+t'v))\cdot(y''-y')\geq 0 \quad \text{when}$$

$$t'<t,\ v'\in\partial f(x+t'v),\ y''\in\partial f(x+t''v)\ .$$

This implies

$$\sup_{y'\in\partial f(x+t'v)} y'\cdot v \leq \inf_{y''\in\partial f(x+t''v)} y''\cdot v = -\sup_{y''\in\partial f(x+t''v)} [-y''\cdot v],$$

or equivalently (by 1.4) and subdifferential regularity)

$$f'(x+t'v;v)\leq -f'(x+t''v;-v) \quad \text{when } t'\leq t'' \ . \tag{4.2}$$

Since also

$$-f'(x';-v)\leq f'(x';v) \text{ for all } x',v,$$

by the sublinearity of $f'(x';\cdot)$, (4.2) tells us that the function $Q(t)=f(x+tv)$ has left and right derivatives which are everywhere nondecreasing in $t\in(-\infty,\infty)$. Again as in the argument that (c) implies (a), we conclude from this fact that f is convex on R^n. □

COROLLARY 5. *Every convex function* $f:R^n\to R$ *is in particular lower-*C^∞.

Proof. In the representation in (c) we must have $F(x,s)=a(s)\cdot x-\alpha(s)$ for certain $a(s)\in R^n$ and $\alpha(s)\in R$ that depend continuously on x. This is the only way that $F(x,s)$ can be affine in x and continuous in s. Then, of course, $F(x,s)$ has partial derivatives of all orders with respect to x, and these are all continuous in (x,s). □

Let us now define two notions parallel to Spingarn's submonotonicity and strict submonotonicity: ∂f is *hypomonotone* if

$$\liminf_{\substack{x'\to x\\ y'\in\partial f(x')}} \frac{(x'-x)\cdot(y'-y)}{|x'-x|^2} > -\infty \text{ for all } x \text{ and } y\in\partial f(x) \tag{4.3}$$

and *strictly hypomonotone* if

$$\liminf_{\substack{x'\to x\\ x''\to x\\ y'\in\partial f(x')\\ y''\in\partial f(x'')}} \frac{(x''-x')\cdot(y''-y')}{|x''-x'|^2} > -\infty \quad \text{for all } x\ . \tag{4.4}$$

Clearly hypomonotone implies submonotone, and strictly hypomonotone implies strictly submonotone. We have little to say here about hypomonotonicity itself, but the importance of strict hypomonotonicity is demonstrated by the following result.

THEOREM 6. *For a locally Lipschitzian function* f *on* R^n, *the following properties are equivalent:*

(a) f *is lower-*C^2 ;

(b) ∂f *is strictly hypomonotone* ;

(c) For every $\bar{x}\in R^n$ *there is a convex neighborhood* X *of* $\bar{x}$ *on which* f *has a representation*

$$f=g-h \text{ on } X \text{ with } g \text{ convex, } h \text{ quadratic convex.} \tag{4.5}$$

(d) For every $\bar{x}\in R^n$ *there is a neighborhood* X *of* $\bar{x}$ *and a representation of* f *as in* (1.6) *with* S *a compact topological space,* $F(x,s)$ *quadratic in* x *and continuous in* s.

Proof.

(a) $\Rightarrow$ (c). Choose any $\bar{x}$ and consider on some neighborhood X of $\bar{x}$ a representation (1.6) of f as in the definition of f being lower-C^2: F(x,s) has second partial derivatives in x, and these are continuous with respect to (x,s). Shrink X if necessary so that it becomes a compact convex neighborhood of $\bar{x}$. The Hessian matrix $\nabla_x^2 F(x,s)$ depends continuously on (x,s) ranging over a compact set $X \times S$, so we have

$$\min_{\substack{(x,s) \in X \times S \\ |v| = 1}} v \cdot \nabla_x^2 F(x,s) v > -\infty .$$

Denote this minimum by $-\rho$ and let

$$G(x,s) = F(x,s) + (\rho/2)|x|^2 . \tag{4.6}$$

Then

$$v \cdot \nabla_x^2 G(x,s) v = v \cdot [\nabla^2 F(x,s) + \rho I] v \geq 0 \tag{4.7}$$

for all $(x,s) \in X \times S$ when $|v| = 1$ and hence also in fact for all $v \in R^n$, because both sides of (4.7) are homogeneous of degree 2 with respect to v. Thus $\nabla_x^2 G(x,s)$ is a positive semidefinite matrix for each $(x,s) \in X \times S$, and G(x,s) is therefore a convex function of $x \in X$ for each $s \in S$. The function

$$g(x) = \max_{s \in S} G(x,s)$$

is accordingly convex, and we have from (4.6) and (1.6) that (4.5) holds for this and $h(x) = (\rho/2)|x|^2$.

(c) $\Rightarrow$ (d). Given a representation as in (c), we can translate it into one as in (d) simply by plugging in a representation of g of the type described in Theorem 5(c).

(d) $\Rightarrow$ (a). Any representation of type (d) is a special case of the kind of representation in the definition of f being lower-C^2 (in fact lower-C^∞); if a quadratic function of x depends

continuously on s, so must all its coefficients in any expansion as a polynomial of degree 2.

(c) $\Rightarrow$ (b). Starting from (4.5) we argue that $\partial f(x) = \partial g(x) - \partial h(x)$ (cf. Clarke 1980, §3, and Rockafellar 1979, p.345), where ∂g happens to be monotone (Theorem 5) and ∂h is actually a linear transformation: $y \in \partial f(x)$ if and only if $y = Ax$, where A is symmetric and positive semidefinite. For $y' \in \partial f(x')$, $y'' \in \partial f(x'')$ we have $y' + Ax' \in \partial g(x')$ and $y'' + Ax'' \in \partial g(x'')$, so from the monotonicity of ∂g it follows that

$$\begin{aligned} 0 &\leq (x' - x'') \cdot ([y' + Ax'] - [y'' + Ax'']) \\ &= (x' - x'') \cdot (y' - y'') + (x' - x'') \cdot A(x' - x'') . \end{aligned} \tag{4.8}$$

Choosing $\rho > 0$ large enough that

$$v \cdot Av \leq \rho |v|^2 \quad \text{for all } v \in R^n$$

we obtain from (4.8) that

$$(x'' - x') \cdot (y'' - y') \geq \rho |x'' - x'|^2 \quad \text{when } x' \in X,\ x'' \in X,\ y' \in \partial f(x'),\ y'' \in \partial f(x'') . \tag{4.9}$$

Certainly (4.4) holds then for $x = \bar{x}$, and since $\bar{x}$ was an arbitrary point of R^n we conclude that ∂f is hypomonotone.

(b) $\Rightarrow$ (c). We are assuming (4.4), so for any $\bar{x}$ we know we can find a convex neighborhood X of $\bar{x}$ and a $\rho > 0$ such that (4.9) holds. Let $g(x) = f(x) + (\rho/2)|x|^2$, so that $\partial g = \partial f + \rho I$ (cf. Clarke 1980, §3, and Rockafellar 1979, p. 345). Then by (4.9), ∂g is monotone on X, and it follows that g is convex on X (cf. Theorem 5; the argument in Theorem 5 is in terms of functions on all of R^n, but it is easily relativized to convex subsets of R^n). Thus (4.5) holds for this g and $h(x) = (\rho/2)|x|^2$. □

COROLLARY 6. *If a function* $f:R^n \to R$ *is lower-*C^2*, it is actually lower-*C^∞*. Thus for* $2 \leq k \leq \infty$ *the classes of lower-*C^k *functions all coincide.*

Proof. As noted in the proof that (d) $\Rightarrow$ (a), any representation of the kind in (d) actually fits the definition of f being lower-C^∞.

COROLLARY 7. *Let* $f:R^n \to R$ *be lower-*C^2*. Then at almost every* $x \in R^n$*, f is twice-differentiable in the sense that there is a quadratic function q for which one has*

$$f(x') = q(x') + o(|x'-x|^2) \quad .$$

Proof. This is a classical property of convex functions (cf. Alexandroff 1939), and it carries over to general lower-C^2 functions via the representation in (c).

Counterexample

Since the lower-C^k functions are all the same for $k \geq 2$, it might be wondered if the lower-C^1 functions are really any different either. But here is an example of a lower-C^1 function that is not lower-C^2. Let $f(x) = -|x|^{3/2}$ on R. Then f is of class C^1, hence in particular a lower-C^2, and there would exist by characterization (d) in Theorem 6 numbers α, β, γ, such that

$$f(x) \geq \alpha + \beta x + \gamma x^2 \quad \text{for all } x \text{ near } 0, \text{ with equality when } x = 0 .$$

Then $\alpha = f(0) = 0$ and $-|x|^{3/2} \geq \beta x + \gamma x^2$, from which it follows on dividing by $|x|$ and taking the limits $x \downarrow 0$ and $x \uparrow 0$ that $\beta = 0$. Thus γ would have to be such that $-|x|^{3/2} \geq \gamma|x|^2$ for all x sufficiently near 0, and this is impossible. Therefore f is not lower-C^2.

REFERENCES

Alexandroff, A.D. 1939. Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it. Leningrad State Univ. Ann., Math. Ser. 6:3-35 (Russian).

Clarke, F.H. 1975. Generalized gradients and applications. Trans. Amer. Math. Soc. 205:247-262.

Clarke, F.H. 1980. Generalized gradients of Lipschitz continuous functionals. Advances in Math.

Danskin, J.M. 1967. The Theory of Max-Min and its Applications to Weapons Allocations Problems. Springer-Verlag. New York.

Lemarechal, C. 1975. An extension of Davidon methods to nondifferentiable problems. Math. Programming Study 3: Nondifferentiable Optimization, 95-109. North Holland.

Mifflin, R. 1977a An algorithm for constrained optimization with semismooth functions. Math. of Op. Research 2: 191-207.

Mifflin, R. 1977b Semismooth and semiconvex functions in constrained optimization. SIAM J. Control Opt. 15: 959-972.

Pshenichnyi, B.N. 1971. Necessary Conditions for an Extremum. Marcel Dekker. New York.

Rockafellar, R.T. 1970. Convex Analysis. Princeton Univ. Press.

Rockafellar, R.T. 1977. Directionally Lipschitzian functions and subdifferential calculus. Proc. London Math. Soc. 39: 331-355.

Saks, S. 1937. Theory of the Integral. Hafner. New York.

Spingarn, J.E. (forthcoming) Submonotone subdifferentials of Lipschitz functions. Trans.Amer.Math.Soc.

Stein, E.M. 1970. Singular Integrals and Differentiability Properties of Functions. Princeton Univ. Press.

Wolfe, P. 1975. A method of conjugate gradients for minimizing nondifferentiable functions. Math. Programming Study 3: Nondifferentiable Optimization, 145-173. North Holland.