GENERALIZED GRADIENTS AND APPLICATIONS

BY

FRANK H. CLARKE

ABSTRACT. A theory of generalized gradients for a general class of functions is developed, as well as a corresponding theory of normals to arbitrary closed sets. It is shown how these concepts subsume the usual gradients and normals of smooth functions and manifolds, and the subdifferentials and normals of convex analysis. A theorem is proved concerning the differentiability properties of a function of the form $\max\{g(x, u): u \in U\}$. This result unifies and extends some theorems of Danskin and others. The results are then applied to obtain a characterization of flow-invariant sets which yields theorems of Bony and Brezis as corollaries.

Introduction. Some of the most important recent advances in optimization have come about as a result of the systematic replacement of smoothness assumptions by convexity. This is exemplified by the work of Rockafellar [12], [13], which has extended the boundaries of treatable problems and has in addition led to new techniques for dealing with problems of a familiar nature.

It is natural to ask whether analogous results can be proven without either smoothness or convexity. A general theory of necessary conditions for such problems has been obtained [3] and the results described in [4]. The conditions are expressed, in part, by means of generalized "gradients". They subsume the results of the smooth and convex cases and they yield, among other things, significant extensions of the Pontryagin maximum principle of optimal control theory.

We describe in this article the generalized theory of gradients and some of its consequences. As mentioned, the principal application of this theory has been to variational problems, but the two main applications given here concern the differential properties of max functions (§2) and flow-invariant sets (§4).

1. **Locally Lipschitz functions.** Throughout this section, we shall be dealing with a function $f: R^n \longrightarrow R$ which will be assumed locally Lipschitz. That is, for each bounded subset B of R^n there exists a constant K such that

$$|f(x_1) - f(x_2)| \leqslant K|x_1 - x_2|$$

for all points x_1 and x_2 of B. (The vertical bars denote the Euclidean norm.) It is known [14] that such a function has at almost all points x a derivative (gradient), which we denote $\nabla f(x)$. It is easily verified that the function ∇f is bounded on bounded subsets of its domain of definition.

(1.1) DEFINITION. The *generalized gradient* of f at x, denoted $\partial f(x)$, is the convex hull of the set of limits of the form $\lim \nabla f(x + h_i)$, where $h_i \longrightarrow$ 0 as $i \longrightarrow \infty$. (An extension of this definition to functions not almost everywhere differentiable will be given later.)

It follows that $\partial f(x)$ is a nonempty convex compact set [11, Theorem 17.2]. The mapping ∂f from R^n to the subsets of R^n (we use the term *multifunction* from R^n to R^n) is also seen to be *upper semicontinuous*: if v_i and x_i are sequences tending to v and x respectively, and if v_i belongs to $\partial f(x_i)$ for each i, then v belongs to $\partial f(x)$.

The *epigraph* of a function f, denoted epi (f), is the set $\{(x, s) : f(x) \leqslant s\}$. We shall say f is *convex* if epi (f) is a convex set (see [11] for other equivalent definitions). The vector p is a *subgradient* of f at x if, for all v in R^n,

$$f(x + v) - f(x) \geqslant v \cdot p,$$

where the right side denotes the usual inner product.

(1.2) PROPOSITION. *If f is convex, then $\partial f(x)$ is the set of subgradients of f at x.*

The proof is immediate from convex analysis [11, Theorem 25.6].

(1.3) DEFINITION. The generalized directional derivative of f at x in the v direction, denoted $f^\circ(x; v)$, is given by

$$f^\circ(x; v) = \limsup_{h \to 0; \delta \downarrow 0} [f(x + h + \delta v) - f(x + h)]/\delta.$$

(As with (1.1), this definition is appropriate only for the Lipschitz functions we are considering at present.)

We shall denote the usual one-sided directional derivative of f at x by $f'(x; v)$. Thus

$$f'(x; v) = \lim_{\delta \downarrow 0} [f(x + \delta v) - f(x)]/\delta,$$

whenever this limit exists.

(1.4) PROPOSITION. $f^\circ(x; v) = \max\{\zeta \cdot v : \zeta \in \partial f(x)\}$. *(That is, $f^\circ(x; \cdot)$ is the support function of $\partial f(x)$.)*

PROOF. Define the function $g : R^n \times R^n \longrightarrow R$ by

$$(1.5) \qquad\qquad g(x, v) = \max\{\zeta \cdot v : \zeta \in \partial f(x)\}.$$

It is immediate from (1.1) that

(1.6) $$g(x, v) = \limsup_{h \to 0} v \cdot \nabla f(x + h),$$

where this is finite. It follows that for suitable sequences h_i and δ_i tending to 0,

$$g(x, v) = \lim_{i \to \infty} [f(x + h_i + \delta_i v) - f(x + h_i)]/\delta_i,$$

whence $g \leqslant f^\circ$ by definition. We need only prove that, for any positive ϵ,

(1.7) $$f^\circ(x; v) \leqslant g(x, v) + \epsilon,$$

which we proceed to do. We may assume v is nonzero. Let $g(x, v) = \alpha$. It follows from (1.6) that for h in some ball of radius $\kappa > 0$ around 0, when $\nabla f(x + h)$ exists,

(1.8) $$v \cdot \nabla f(x + h) \leqslant \alpha + \epsilon.$$

Let a subset S of R^n have measure 0, and consider the set of lines in R^n parallel to a given vector. It is a consequence of Fubini's Theorem that almost all of these lines meet S in a set of 0 one-dimensional measure. If we apply this fact in the case where S is the set of points at which ∇f fails to exist, we deduce that for almost all h the function $\delta \to f(x + h + \delta v)$ is differentiable a.e., its derivative being $v \cdot \nabla f(x + h + \delta v)$. We have then

$$f(x + h + \delta v) = f(x + h) + \int_0^\delta v \cdot \nabla f(x + h + tv)\, dt.$$

In view of (1.8), we see that for almost all h such that $|h| < \kappa/2$, and for all $\delta < \kappa/|v|2$,

$$[f(x + h + \delta v) - f(x + h)]/\delta = (1/\delta) \int_0^\delta v \cdot \nabla f(x + h + tv)\, dt \leqslant \alpha + \epsilon.$$

Since f is continuous, this must then hold for all such h without exception. The required inequality (1.7) follows from (1.3).   Q.E.D.

(1.9) COROLLARY. $f^\circ(x; v)$ *is a convex function of* v.

PROOF. A supremum of linear functions is a convex function [11, Theorem 5.5].

(1.10) COROLLARY. *If for some point* ζ *of* R^n *and for all* v *in* R^n *we have*

$$\zeta \cdot v \leqslant \limsup_{\delta \downarrow 0} [f(x + \delta v) - f(x)]/\delta,$$

then ζ *belongs to* $\partial f(x)$.

PROOF. It follows from the above hypothesis that, for all v in R^n,

$\zeta \cdot v \leqslant f^\circ(x; v)$. The conclusion then follows from (1.4) combined with [11, Theorem 13.1]. Q.E.D.

A slight modification of the proof of (1.4) yields:

(1.11) PROPOSITION. *Let E be a subset of R^n of measure 0, and let ∂f_E and f_E° be defined as in (1.1) and (1.3) respectively, except that in those definitions the points $x + h_i$, $x + h$, and $x + h + \delta v$ are constrained to the complement of E. Then*

$$f^\circ(x; v) = f_E^\circ(x; v) = \max\{\zeta \cdot v : \zeta \in \partial f_E(x)\},$$

and

$$\partial f_E(x) = \partial f(x).$$

(1.12) PROPOSITION. $\partial(f_1 + f_2)(x) \subset \partial f_1(x) + \partial f_2(x)$.

PROOF. Let E be the set of points at which either f_1 or f_2 fails to be differentiable. Since any sequence of the form $\nabla(f_1 + f_2)(x + h_i)$, where $x + h_i$ lies in the complement of E, admits a subsequence such that $\nabla f_1(x + h_{n_i})$ and $\nabla f_2(x + h_{n_i})$ both exist and converge, the result follows from (1.1) and (1.11). Q. E. D.

The function $f(x) = x^2 \sin(1/x)$ is differentiable at 0, but it is not difficult to establish that $\partial f(0)$ contains more than the point $\nabla f(0) = 0$. In fact, $\partial f(0) = [-1, 1]$. By taking $f_1 = f$ and $f_2 = -f$ in (1.12), we see that strict inclusion may occur.

The reason that $\partial f(0)$ does not reduce to $\nabla f(0)$ in the above case, as we now see, lies in the discontinuity of the gradient.

(1.13) PROPOSITION. *The following are equivalent:*
(a) $\partial f(x) = \{\zeta\}$, *a singleton.*
(b) $\nabla f(x)$ *exists, $\nabla f(x) = \zeta$, and ∇f is continuous at x relative to the set upon which it exists.*

PROOF. That (b) implies (a) is evident from (1.1). Assume that (a) holds. Define $g: [-1, 1] \times B(1, 0) \to R$ by

$$g(\delta, v) = [f(x + \delta v) - f(x)]/\delta - \zeta \cdot v \quad \text{for } \delta \neq 0,$$
$$= 0 \qquad\qquad\qquad\qquad\qquad \text{for } \delta = 0,$$

where $B(1, 0)$ is the closed unit ball about 0 in R^n. Then (1.4) implies

$$\limsup_{u \to v; \delta \downarrow 0} g(\delta, u) = 0 = g(0, v).$$

The same argument with $-f$ and $-\zeta$ replacing f and ζ shows that the same inequality holds for the lim inf also, and hence that g is continuous (and

consequently uniformly continuous) on the compact set $[0, 1] \times B(1, 0)$. Because $g(-\delta, v) = -g(\delta, -v)$ and $g(0, \cdot) \equiv 0$, this uniform continuity holds on $[-1, 1] \times B(1, 0)$. It follows by definition that $\nabla f(x)$ exists and equals ζ.

The last assertion in (b) follows from the previously noted fact that ∇f is bounded in a neighborhood of x and the fact that any limit of the form in (1.1) must converge to ζ if (a) holds.   Q.E.D.

We may use (1.13) to deduce the following characterization of C^1 functions:

COROLLARY. *Let* B *be an open subset of* R^n. *The following are equivalent*:

(a) f *is* C^1 *on* B.

(b) *For each* x *in* B, *the following limit exists for each* v *in* R^n:

$$\lim_{h \to 0; \delta \to 0} [f(x + h + \delta v) - f(x + h)]/\delta.$$

(c) $\partial f(x)$ *is a singleton for each* x *in* B.

2. Max functions. We now characterize the generalized gradient of a type of function important in optimization.

(2.1) THEOREM. *Let* U *be a sequentially compact space, and let* $g: R^n \times U \longrightarrow R$ *have the following properties*:

(a) $g(x, u)$ *is u. s. c. in* (x, u).

(b) g *is locally Lipschitz in* x, *uniformly for* u *in* U.

(c) $g_x^\circ(x, u; \cdot) = g_x'(x, u; \cdot)$, *the derivatives being with respect to* x.

(d) $\partial_x g(x, u)$ *is u. s. c. in* (x, u).

Then, if we let $f(x) = \max\{g(x, u): u \in U\}$,

(1) f *is locally Lipschitz*.

(2) $f'(x; v)$ *exists*.

(3) $f'(x; v) = f^\circ(x; v) = \max\{\zeta \cdot v : \zeta \in \partial_x g(x, u), u \in M(x)\}$, *where* $M(x) = \{u \in U: g(x, u) = f(x)\}$.

(4) $\partial f(x)$ *is the convex hull of* $\{\partial_x g(x, u): u \in M(x)\}$.

REMARK. The differential properties of the function f have been the subject of much investigation (see [5], [8] for surveys and references). In a suitable setting, hypotheses (b), (c), and (d) will follow if g is convex in x (by (1.2), [11, Theorems 23.4 and 24.5]), and we thereby generalize some results of other authors. The conditions on g also hold when $\nabla_x g(x, u)$ exists and is continuous in (x, u), and we thereby obtain a generalization of a theorem originally due to Danskin [6] (see also [9, Theorem 3.3]).

PROOF. Notice first that because g is u. s. c. and U is sequentially

compact, the use of "max" rather than "sup" is justified; for the same reasons $M(x)$ is nonempty. Let K be a Lipschitz constant for $g(x, u)$ (in x) for x in a bounded set and for all u in U. Let x_1 and x_2 be points in the bounded set, and choose any $\bar{u}$ in $M(x_1)$. Then

$$f(x_1) = g(x_1, \bar{u}) \leqslant g(x_2, \bar{u}) + K|x_1 - x_2| \leqslant f(x_2) + K|x_1 - x_2|.$$

Since we may similarly obtain the same inequality with x_1 and x_2 switched, (1) follows.

We now prove (4). We have, for any u in $M(x)$ and v in R^n,

$$g_x^o(x, u; v) = g_x'(x, u; v) = \lim_{\delta \downarrow 0} [g(x + \delta v, u) - g(x, u)]/\delta$$
$$\leqslant \lim_{\delta \downarrow 0} \sup [f(x + \delta v) - f(x)]/\delta.$$

By (1.4) we then have this last expression greater or equal to $\zeta \cdot v$ for any ζ in $\partial_x g(x, u)$, whence any such ζ belongs to $\partial f(x)$ by (1.10). Consequently we derive

(2.2) $\text{co}\{\partial_x g(x, u) : u \in M(x)\} \subset \partial f(x),$

where "co" denotes "convex hull." Now let $\bar{x}$ be a point where $\nabla f(\bar{x})$ exists. Then, for any u in $M(\bar{x})$,

$$v \cdot \nabla f(\bar{x}) = \lim_{\delta \downarrow 0} [f(\bar{x} + \delta v) - f(\bar{x})]/\delta$$
$$\geqslant \lim_{\delta \downarrow 0} \sup [g(\bar{x} + \delta v, u) - g(\bar{x}, u)]/\delta$$
$$= g_x'(\bar{x}, u; v) = g_x^o(\bar{x}, u; v) = \max\{\zeta \cdot v : \zeta \in \partial_x g(\bar{x}, u)\}.$$

Since this holds for any v in R^n, we deduce by [11, Theorem 13.1]

(2.3) $\partial_x g(\bar{x}, u) = \{\nabla f(\bar{x})\}.$

Now suppose that we have a sequence of points $x + h_i$ where $\nabla f(x + h_i)$ exists and h_i tends to 0. Let u_i belong to $M(x + h_i)$. We may suppose that u_i converges to some u_0 in U, since U is sequentially compact. It follows readily that u_0 belongs to $M(x)$. But from (2.3) it follows that if $\lim_{i \to \infty} \nabla f(x + h_i)$ exists, it must be in the set $\partial_x g(x, u_0)$, in view of hypothesis (d). Consequently we deduce from the above and (1.1)

$$\partial f(x) \subset \text{co}\{\partial_x g(x, u) : u \in M(x)\},$$

and this combined with (2.2) yields (4).

Because of (4), given any v, we may choose u in $M(x)$ and ζ in $\partial_x g(x, u)$ such that $f^o(x; v) = \zeta \cdot v$. We have then

$$\liminf_{\delta \downarrow 0} [f(x + \delta v) - f(x)]/\delta \geqslant \liminf_{\delta \downarrow 0} [g(x + \delta v, u) - g(x, u)]/\delta$$

$$= g'_x(x, u; v) = g^{\circ}_x(x, u; v) \geqslant \zeta \cdot v = f^{\circ}(x; v).$$

But the reverse inequality always holds by (1.3), and hence

$$\liminf_{\delta \downarrow 0} [f(x + \delta v) - f(x)]/\delta = \limsup_{\delta \downarrow 0} [f(x + \delta v) - f(x)]/\delta$$

and both equal $f^{\circ}(x; v)$. This proves (2) and (3) with the exception of the expression for $f^{\circ}(x; v)$ in (3), which is an easy consequence of (1.5).   Q.E.D.

Theorem 2.1 shows that $f'(x; v)$ is convex in v. Such functions are termed quasidifferentiable, and can serve as the natural setting for very general problems in optimization [9]. It is not difficult to show that if a function g has $g'(x; v)$ convex in v and u.s.c. in x, then g satisfies $g^{\circ}(x; v) = g'(x; v)$ (hypothesis (c) in (2.1)). This is the case for smooth, convex, and max functions.

We shall now use (2.1) to characterize the generalized gradient of the function

$$d_E(x) = d(x, E) = \min\{|x - e| : e \in E\},$$

where E is a nonempty closed subset of R^n. The function d_E is Lipschitz by the argument that proved assertion (1) of Theorem (2.1).

(2.4) PROPOSITION. *Suppose $\nabla d_E(x)$ exists and is different from* 0. *Then*

(1) *x belongs to the complement of E.*

(2) *There exists a unique point e in E closest to x.*

(3) *$\nabla d_E(x) = (x - e)/|x - e|$.*

PROOF. If x is in E, then for any v in R^n we have

$$v \cdot \nabla d_E(x) = \lim_{\delta \downarrow 0} [d_E(x + \delta v) - d_E(x)]/\delta = \lim_{\delta \downarrow 0} d_E(x + \delta v)/\delta \geqslant 0,$$

whence $\nabla d_E(x) = 0$, contrary to assumption. This proves (1).

For (2) and (3) we need only note that $g(y, u) = |y - u|$ is C^1 for y near x and u in E, and apply Theorem 2.1 to the function

$$-d_E(y) = \max\{-|y - u| : u \in E\}.$$

Q.E.D.

(2.5) COROLLARY. *Let e belong to E. Then*

$$\partial d_E(e) = \mathrm{co}\{0, \lim (x_i - e_i)/|x_i - e_i|\},$$

where we consider all sequences x_i, e_i such that x_i is not in E and has closest point e_i in E, and $x_i \to e$ as $i \to \infty$.

PROOF. We have the left side contained in the right by (2.4) and (1.1). The opposite inclusion requires only that 0 belong to $\partial d_E(e)$. This follows from (1.10) and the following observation, for any v in R^n:

$$\limsup_{\delta \downarrow 0} \, [d_E(e + \delta v) - d_E(e)]/\delta = \limsup_{\delta \downarrow 0} d_E(e + \delta v)/\delta \geqslant 0.$$

Q.E.D.

3. **Normals to sets.** Let E be a closed nonempty subset of R^n, and let e be a point in E.

(3.1) DEFINITION. The *cone of normals* to E at e, denoted $N_E(e)$, is the closure of the set

$$\{p \in R^n : sp \in \partial d_E(e) \text{ for some } s \text{ in } (0, \infty)\}.$$

We say p is *normal* to E at e if p belongs to $N_E(e)$. It is easy to see that $N_E(e)$ is a closed convex cone.

(3.2) PROPOSITION.

$$N_E(e) = \text{cl co}\left\{\lim_{i \to \infty} s_i(x_i - e_i)\right\},$$

where we consider all sequences with $s_i \geqslant 0$, x_i converging to e, and having closest point e_i in E.

PROOF. Let $p = \lim s_i(x_i - e_i)$. We may assume $(x_i - e_i)/|x_i - e_i|$ converges to a limit ζ. Then $\zeta \in \partial d_E(e)$ by (2.5), and

$$p = [\lim s_i |x_i - e_i|] \, [\lim (x_i - e_i)/|x_i - e_i|] = |p|\zeta,$$

whence $p \in N_E(e)$. This shows that the right side in (3.2) is contained in the left, and the opposite inclusion follows from (2.5) and (3.1).   Q.E.D.

(3.3) PROPOSITION. (1) *Let E be a C^1 manifold. Then $N_E(e)$ coincides with the usual space of normals to E at e.*

(2) *Let E be convex. Then $N_E(e)$ coincides with the cone of normals to E at e in the sense of convex analysis.*

PROOF. We first prove (1).

Let x_0 be such that $\nabla d_E(x_0)$ exists and is nonzero. It follows from (2.4) that x_0 lies in the complement of E and that x_0 has a unique nearest point e_0 in E. In a neighborhood of e_0, E has a representation in local coordinates of the form

$$E = \{e \in R^n : e = f(v), \ v \text{ in an open subset } V \text{ of } R^m\},$$

where $f: R^m \longrightarrow R^n$ is C^1, $e_0 = f(v_0)$ for some v_0 in V, and the Jacobian $f_x(v_0)$ has maximal rank (E then has dimension m). Recall that the normal

space at e_0 is the space orthogonal to that spanned by the m columns of $f_x(v_0)$. Because $|f(v) - x_0|^2$ has a local minimum at $v = v_0$, we deduce

$$(f(v_0) - x_0)f_x(v_0) = 0,$$

whence we see that $(x_0 - e_0)/|x_0 - e_0|$ (which equals $\nabla d_E(x_0)$ by (2.4)) lies in the usual normal space at e_0. Now suppose that x_i and e_i are sequences of points such as these, where x_i tends to e and $\nabla d_E(x_i)$ converges. The limit of the sequence $\nabla d_E(x_i)$ must then belong to the usual normal space at e, since this space is spanned by vector functions continuous in e. We conclude that $\partial d_E(e)$ is contained in the usual normal space at e, and hence that $N_E(e)$ is also.

It will suffice to prove the opposite inclusion by showing that an arbitrary unit vector p in the usual normal space at e is contained in $\partial d_E(e)$, in view of (3.1). We assume $e = 0$ and that, near 0, E has the representation $E = \{e \in R^n : F(e) = 0\}$, where $F : R^n \rightarrow R^{n-m}$ is such that $F_x(0)$ has maximal rank. Recall that in this setting the usual normal space at 0 is spanned by the vectors $\nabla F^i(0)$, $1 \leqslant i \leqslant n - m$, and hence $p = \lambda F_x(0)$ for some vector λ in R^{n-m}. Let δ_i decrease to 0, and let $\delta_i p$ have closest point e_i in E. Then, because 0 is in E,

$$(3.4) \qquad\qquad |\delta_i p - e_i| \leqslant \delta_i,$$

and consequently we may assume that e_i/δ_i converges to a limit in R^n. From the fact that

$$F(x) = F(0) + xF_x(0) + o(x),$$

we have $e_i F_x(0) = -o(e_i)$, since $F(0) = 0$. Squaring (3.4) and substituting gives

$$(3.5) \qquad\qquad 2\delta_i \lambda \cdot o(e_i) + |e_i|^2 \leqslant 0,$$

where we have also replaced p by $\lambda F_x(0)$.

Now if any e_i is zero, then $\delta_i p$ has nearest point 0 in E, and $p \in \partial d_E(0)$ by (2.5). Hence we may assume the e_i are different from 0. Dividing (3.5) by $\delta_i |e_i|$ and taking limits, we find that $\lim e_i / \delta_i = 0$.

Because e_i/δ_i converges to 0, we have $\delta_i p - e_i \neq 0$ for i large. But then we see that the vectors $(\delta_i p - e_i)/|\delta_i p - e_i|$ belong to $\partial d_E(e_i)$ by (2.5) and that they converge to p. Since ∂d_E is u.s.c., $p \in \partial d_E(0)$ and we are done.

PROOF OF (2). Recall that a vector p is normal to E at e_0 in the sense of convex analysis iff, for all e in E, $p \cdot (e_0 - e) \geqslant 0$. It is well known that this holds iff $p + e_0$ has closest point e_0 in E. This fact combined with (3.2) gives the desired equivalence.   Q.E.D.

(3.6) DEFINITION. The *tangent cone* to E at e, denoted $T_E(e)$, is the cone dual to $N_E(e)$:
$$T_E(e) = \{\zeta : \zeta \cdot v \leqslant 0 \text{ for all } v \text{ in } N_E(e)\}.$$

It follows from Proposition (1.4) that v is tangent to E at e (i.e. $v \in T_E(e)$) iff $d_E^0(e; v) = 0$. The following result will be useful later, and gives an alternate characterization of tangents.

(3.7) PROPOSITION. *Let e_0 belong to E, and let v be a point of R^n. The following are equivalent*:
(1) $v \in T_E(e_0)$.
(2) $\lim_{e \to e_0; e \in E} \lim \inf_{\delta \downarrow 0} d_E(e + \delta v)/\delta = 0$.

PROOF. By the above remark, we have $d_E^0(e; v) = 0$ if (1) holds, and hence

(3.8) $$\lim_{h \to 0; \delta \downarrow 0} \sup [d_E(e_0 + h + \delta v) - d_E(e_0 + h)]/\delta = 0.$$

It is easy to see that the expression occurring in (2), with lim sup in place of the limit there is no greater than that in (3.8), and yet is nonnegative. The existence of the limit and equality (2) follow.

Suppose now that (2) holds. Let x be a point not in E where $\nabla d_E(x)$ exists. Then if e is the point of E closest to x,

(3.9) $$|x - e| = d_E(x) \leqslant |x - e - \delta v| + d_E(e + \delta v),$$

for any positive δ, since d_E is Lipschitz with constant 1. Squaring both sides in (3.9) and rearranging, we derive

(3.10) $$v \cdot \frac{(x - e)}{|x - e|} \leqslant \frac{\delta v \cdot v}{2|x - e|} + \delta^{-1} d_E(e + \delta v) \left[\frac{d_E(e + \delta v)}{2|x - e|} + \frac{|x - e - \delta v|}{|x - e|} \right].$$

We may obtain a bound on the last term in this expression:

$$\frac{d_E(e + \delta v)}{2|x - e|} + \frac{|x - e - \delta v|}{|x - e|}$$

(3.11) $$\leqslant (2|x - e|)^{-1}\{d_E(x) + |x - e - \delta v| + 2|x - e| + \delta |v|\}$$

$$\leqslant (2|x - e|)^{-1}\{4d_E(x) + 2\delta |v|\} \leqslant 2 + \frac{\delta |v|}{|x - e|}.$$

For any positive ϵ, we may choose δ_0 such that

$$\frac{\delta_0}{|x - e|} < \epsilon, \qquad \frac{d_E(e + \delta_0 v)}{\delta_0} \leqslant \lim_{\delta \downarrow 0} \inf \frac{d_E(e + \delta v)}{\delta} + \epsilon.$$

Substituting into (3.10), and using (3.11) and the equality

$$(x - e)/|x - e| = \nabla d_E(x)$$

(by (2.4)), we arrive at

$$v \cdot \nabla d_E(x) \leqslant \epsilon |v|^2 + (2 + \epsilon |v|) \left[\lim_{\delta \downarrow 0} \inf d_E(e + \delta v)/\delta + \epsilon \right].$$

Hence we must have

$$v \cdot \nabla d_E(x) \leqslant 2 \lim_{\delta \downarrow 0} \inf d_E(e + \delta v)/\delta,$$

and since the right-hand side goes to 0 as e goes to e_0, we conclude from (1.1) that $v \cdot \zeta \leqslant 0$ for every element ζ of $\partial d_E(e_0)$. This implies (1). Q.E.D.

We shall apply the above notions to flow-invariance in the next section. We proceed to relate them now to the generalized gradients defined in §1. The first result generalizes the familiar calculus result that $(df/dx, -1)$ is a vector normal to the graph of f.

(3.12) LEMMA. *Let* $f: R^n \longrightarrow R$ *be given, and suppose* $\nabla f(x_0)$ *exists. If* E *is the epigraph of* f *(see §1) then*

$$(\nabla f(x_0), -1) \in N_E(x_0, f(x_0)).$$

PROOF. Let us assume $x_0 = 0$ and $f(x_0) = 0$, and put $\nabla f(0) = \zeta$. We shall be done if we establish

(3.13) $(\zeta, -1)/|(\zeta, -1)| \in \partial d_E(0, 0)$.

Let $g(x) = \zeta \cdot x$, and let d_G be the distance function corresponding to epi(g). It is not difficult to prove that for a fixed (v, s) in $R^n \times R$ we have

$$d_E(\delta (v, s)) \geqslant d_G(\delta (v, s)) + o(\delta).$$

Consequently we deduce

(3.14) $\lim_{\delta \downarrow 0} \sup [d_E(\delta(v, s)) - d_E(0, 0)]/\delta \geqslant \lim_{\delta \downarrow 0} \sup d_G(\delta(v, s))/\delta.$

But

$$d_G(\delta(v, s))/\delta = 0 \geqslant (\zeta, -1) \cdot (v, s) \qquad \text{if } s \geqslant \zeta \cdot v.$$
$$= (v, s) \cdot (\zeta, -1)/|(\zeta, -1)| \quad \text{otherwise.}$$

In any case, the left side of (3.14) is no less than

$$(v, s) \cdot (\zeta, -1)/|(\zeta, -1)|,$$

and (3.13) follows from (1.10). Q.E.D.

(3.15) LEMMA. *Let* $f: R^n \longrightarrow R$ *be locally Lipschitz, and let* $E = \text{epi}(f)$. *If the point* $(y, s) \in R^n \times R$ *is not in* E *and has closest point* (x, t) *in* E, *then* $(y - x) \in (t - s)\partial f(x)$.

PROOF. There is a circle about (y, s) with (x, t) on its boundary and no points of E in its interior. This implies

$$(3.16) \qquad (y - x, s - t) \cdot (z - x, f(z) - t) \leqslant o(z - x).$$

Let v be any point of R^n. We put $z = x + \delta v$ in (3.16) and derive (using $t \geqslant s$ and $t \geqslant f(x)$)

$$\limsup_{\delta \downarrow 0} (t - s) (f(x + \delta v) - f(x))/\delta \geqslant v \cdot (y - x).$$

This yields the desired result by (1.10).   Q.E.D.

(3.17) PROPOSITION. *Let f be locally Lipschitz and let $E = \mathrm{epi}(f)$. The following are equivalent*:

(1) $\zeta \in \partial f(x)$.

(2) $(\zeta, - 1) \in N_E(x, f(x))$.

PROOF. That (1) implies (2) follows from (1.1) and an appeal to Lemma (3.12). The reverse implication is a consequence of Lemma (3.15) and Proposition (3.2), which states that N_E is generated by points of the type appearing in (3.15).   Q.E.D.

Suppose now that $f: R^n \longrightarrow (\infty, \infty]$ is l.s.c. Then $E = \mathrm{epi}(f)$ is a closed subset of $R^n \times R$.

(3.18) DEFINITION. Let f be finite at x. We define

$$\partial f(x) = \{p : (p, - 1) \in N_E(x, f(x))\},$$

and we call $\partial f(x)$ the *generalized gradient* of f at x.

NOTE. Proposition (3.17) assures that (3.18) is a consistent extension of Definition (1.1) to non-Lipschitz functions. In this more general case, ∂f can be empty, and need not be compact. An instance of the former behavior is the function $- |x|^{1/2}$ at 0, and the negative of this function provides an example of the latter case. Consequently, the relation $\partial (- f(x)) = - \partial f(x)$, which is obvious in the Lipschitz case, may fail. We see from (3.12) that whether f is Lipschitz or not, $\partial f(x)$ always contains $\nabla f(x)$ when this exists.

It is natural to wonder whether one may find a suitable extension of Definition (1.3) to the non-Lipschitz case, so that an analogue of (1.4) is valid. It is in fact possible to carry this out, but it is best deferred, along with a thorough study of the properties of these extended constructions. This is because there exists a way of defining certain tangent cones and gradients on an arbitrary locally convex linear topological space, which reduce on R^n to the ones given here. As well as providing a more convenient setting for deriving general results, this method yields also alternate definitions on R^n independent of which equivalent

norm is used. Thus our apparent dependence on the Euclidean norm in this paper may be removed. Whichever approach is used however, we require the results proven here for R^n.

We conclude this section by considering the generalized gradients of two types of functions extremely important in optimization: convex and indicator functions. The first result removes the Lipschitz hypothesis from Proposition (1.2).

(3.19) PROPOSITION. *Let* $f: R^n \rightarrow (-\infty, \infty]$ *be l. s. c. and convex, and let* f *be finite at* x. *Then* $\partial f(x)$ *is the set of subgradients of* f *at* x.

PROOF. p is a subgradient at x iff $(p, -1)$ is normal to epi(f) at $(x, f(x))$ in the sense of convex analysis. This is equivalent by (3.3) to $(p, -1)$ being normal in the sense of (3.1) to epi(f) at $(x, f(x))$, which in turn is equivalent by (3.18) to the inclusion $p \in \partial f(x)$. Q.E.D.

(3.20) DEFINITION. Let E be a closed subset of R^n. The *indicator function* of E is defined by

$$\delta(x \mid E) = 0 \quad \text{if } x \in E,$$
$$= +\infty \quad \text{otherwise.}$$

Note that $\delta(\cdot \mid E)$ is l. s. c.

(3.21) PROPOSITION. *Let* x *lie in* E. *Then* $\partial \delta(x \mid E) = N_E(x)$.

PROOF. The epigraph of $\delta(\cdot \mid E)$ is $E \times [0, \infty)$. By (3.2) and (3.18), we need only show that p is normal to E at x iff $(p, -1)$ is normal to $E \times [0, \infty)$ at $(x, 0)$. This follows readily from (3.2) along with the following two observations: if a point (v, s) in $R^n \times R$ has closest point (e, t) in $E \times [0, \infty)$, then v has closest point e in E; if v has closest point e in E, then $(v, -1)$ has closest point $(e, 0)$ in $E \times [0, \infty)$. Q.E.D.

4. **Flow-invariant sets.** Let X be a multifunction from R^n to R^n. That is, for each $x \in R^n$, $X(x)$ is a subset of R^n. We shall assume that $X(x)$ is nonempty and compact. A *trajectory* for X will mean an absolutely continuous function $x: [0, 1] \rightarrow R^n$ such that

(4.1) $\dot{x}(t) \in X(x(t))$ a.e.

where "a.e." denotes "for almost all t in $[0, 1]$" and $\dot{x}(t)$ is the derivative of x, which exists a.e. The interval $[0, 1]$ is chosen here just for convenience. Relation (4.1) is often referred to as a differential inclusion or a generalized differential equation.

(4.2) DEFINITION. X is *Lipschitz* if there exists a constant K with the following property: given any x_1, x_2 in R^n and a point v_1 in $X(x_1)$, there

exists v_2 in $X(x_2)$ such that $|v_1 - v_2| \leqslant K|x_1 - x_2|$.

This is equivalent to saying that X is Lipschitz in the Hausdorff metric on closed sets. If X is actually a function, (4.2) describes a Lipschitz condition on X.

(4.3) DEFINITION. The closed subset F of R^n is *flow-invariant* for X if any trajectory x for X with initial value in F remains in F. That is, for any trajectory x for X, $x(0) \in F$ implies $x(t) \in F$ for $t \geqslant 0$.

The following theorem demonstrates that the notion of tangency defined in §3 is precisely the right one to characterize flow-invariant sets.

(4.4) THEOREM. *Let X be a Lipschitz multifunction, and let F be a nonempty closed subset of R^n. The following are equivalent:*

(1) *F is flow-invariant for X.*

(2) *For each x in F, $X(x)$ is tangent to F at x (that is, $X(x) \subset T_F(x)$, the latter having been defined in (3.6)).*

PROOF. Suppose first that (2) holds. Let x be a trajectory for X, and let $x(0)$ lie in F. Define $f: [0, 1] \longrightarrow [0, \infty)$ by $f(t) = d_F(x(t))$. We need only prove that f is 0 on $[0, 1]$.

Since $\dot{x}(t)$ is easily seen to be bounded, it follows that x and hence f is Lipschitz on $[0, 1]$. Let t be any point where $f'(t)$ and $\dot{x}(t)$ exist and such that $\dot{x}(t)$ lies in $X(x(t))$ (almost all points in $[0, 1]$ have this property). Then (the following limits are taken as δ goes to 0)

$$
\begin{aligned}
f'(t) &= \lim[d_F(x(t + \delta)) - d_F(x(t))]/\delta \\
&= \lim[d_F(x(t) + \delta\dot{x}(t) + o(\delta)) - d_F(x(t))]/\delta \\
&= \lim[d_F(x(t) + \delta\dot{x}(t)) - d_F(x(t))]/\delta \\
&\leqslant d_F^0(x(t); \dot{x}(t)) \quad \text{(by (1.3))}.
\end{aligned}
$$
(4.5)

Assume for the moment that $x(t)$ does not lie in F. By (1.4) and (2.4) we see that

(4.6) $d_F^0(x(t); \dot{x}(t)) = \dot{x}(t)(x(t) - y)/|x(t) - y|$,

where y is a point in F closest to $x(t)$. By hypothesis, there exists a point v in $X(y)$ such that

$$|v - \dot{x}(t)| \leqslant K|y - x(t)| = Kd_F(x(t)).$$

We then have

(4.7) $\dot{x}(t) \cdot \dfrac{x(t) - y}{|x(t) - y|} \leqslant v \cdot \dfrac{x(t) - y}{|x(t) - y|} + K|x(t) - y|.$

The first term on the right side of this last expression is nonpositive since

$X(y)$ is tangent to F at y (note that $x(t) - y \in N_F(y)$ by (3.2) and v lies in $X(y)$). Substituting (4.6) and (4.7) into (4.5) we arrive at

$$(4.8) \qquad\qquad f'(t) \leqslant Kf(t).$$

If $x(t)$ does lie in F, then $d_F^0(x(t); \dot{x}(t))$ is 0 by hypothesis (since $\dot{x}(t) \in X(x(t))$) and in this case also (4.5) leads to (4.8), which consequently holds a.e. A standard argument then establishes that $f \equiv 0$, and (1) follows.

Let us now posit (1). Let y_0 be any point of F. We shall show that $X(y_0)$ is tangent to F at y_0. Choose any $v_0 \in X(y_0)$.

Let y be an arbitrary point of F. There exists a point v in $X(y)$ satisfying $|v - v_0| \leqslant K|y - y_0|$, and also a trajectory x for X such that $x(0) = y$ and $\dot{x}(0)$ exists with value v [7, Theorem 5]. It follows from flow-invariance that $x(t)$ lies in F for $0 \leqslant t \leqslant 1$. Hence (the lim sup's are as δ goes to 0)

$$(4.9) \qquad \begin{aligned} \lim \sup d_F(y + \delta v)/\delta &= \lim \sup d_F(x(0) + \delta \dot{x}(0))/\delta \\ &= \lim \sup d_F(x(\delta))/\delta = 0. \end{aligned}$$

Consequently,

$$\lim_{y \to y_0; y \in F} \sup \; \lim_{\delta \downarrow 0} \inf d(y + \delta v_0)/\delta$$

$$\leqslant \lim_{y \to y_0; y \in F} \sup \; \lim_{\delta \downarrow 0} \inf [d(y + \delta v)/\delta + |v - v_0|] \leqslant 0,$$

by (4.9). We conclude from (3.7) that $v_0 \in T_F(y_0)$.   Q.E.D.

(4.10) COROLLARY (BONY [1]). *Let X be a locally Lipschitz function and suppose that whenever a point z has closest point y in F we have*

$$(4.11) \qquad\qquad (z - y) \cdot X(y) \leqslant 0.$$

Then F is flow-invariant for X.

PROOF. It is clear from (3.2) that (4.11) implies (2) of the theorem. Q.E.D.

(4.12) COROLLARY (BREZIS [2]). *Let X be a locally Lipschitz function. Then F is flow-invariant for X iff, for each $y \in F$,*

$$(4.13) \qquad\qquad \lim_{\delta \downarrow 0} d_F(y + \delta X(y))/\delta = 0.$$

PROOF. Our proof of the theorem showed that flow-invariance implies (4.9), and that this in turn implies (2). Since (1) and (2) are in fact equivalent by the theorem, each is equivalent to (4.9), which is in turn equivalent to (4.13). Q.E.D.

REMARK. Redheffer [10] has given simplified proofs and extended

versions of the original theorems of Bony and Brezis. He replaces the Lipschitz hypothesis on X by a weaker one involving "uniqueness functions". The proof of Theorem (4.4) could be modified to treat this case.

REFERENCES

1. J. M. Bony, *Principe du maximum, inégalité de Harnack et unicité du problème de Cauchy pour les opérateurs elliptiques dégénérés*, Ann. Inst. Fourier (Grenoble) 19 (1969), 277–304.   MR 41 #7486.

2. H. Brezis, *On a characterization of flow-invariant sets*, Comm. Pure Appl. Math. 23 (1970), 261–263.   MR 41 #2161.

3. F. H. Clarke, *Necessary conditions for nonsmooth problems in optimal control and the calculus of variations*, Thesis, University of Washington, 1973.

4. ———, *Necessary conditions for nonsmooth variational problems* (Proc. Fourteenth Biennial Sem. Canad. Math. Congr., 1974), Springer-Verlag, New York (to appear).

5. V. F. Dem'janov and V. N. Malozemov, *The theory of nonlinear minimax problems*, Uspehi Mat. Nauk 26 (159) (1971), 53–104 = Russian Math. Surveys 26 (1971), 57–115. MR 45 #6435.

6. J. M. Danskin, Jr., *The theory of max-min and its application to weapons allocation problems*, Springer-Verlag, New York, 1967.   MR 37 #3843.

7. A. F. Filippov, *Classical solutions of differential equations with multivalued right-hand side*, SIAM J. Control 5 (1967), 609–621.   MR 36 #4047.

8. W. Hogan, *Directional derivatives for extremal-value functions with applications to the completely convex case*, Operations Res. 21 (1973), 188–209.

9. B. N. Pšeničnyĭ, *Necessary conditions for an extremum*, Izdat. "Nauka", Moscow, 1969; English transl., Pure and Appl. Math., vol. 4, Dekker, New York, 1971.   MR 43 #2584; #2585.

10. R. M. Redheffer, *The theorems of Bony and Brezis on flow-invariant sets*, Amer. Math. Monthly 79 (1972), 740–747.   MR 46 #2166.

11. R. T. Rockafellar, *Convex analysis*, Princeton Math. Ser., no. 28., Princeton Univ. Press, Princeton, N. J., 1970.   MR 43 #445.

12. ———, *Conjugate convex functions in optimal control and the calculus of variations*, J. Math. Anal. Appl. 32 (1970), 174–222.   MR 42 #929.

13. ———, *Existence and duality theorems for convex problems of Bolza*, Trans. Amer. Math. Soc. 159 (1971), 1–40.   MR 43 #7995.

14. E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Math. Ser., no. 30, Princeton Univ. Press, Princeton, N. J. 1970.   MR 44 #7280.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF BRITISH COLUMBIA, VANCOUVER, BRITISH COLUMBIA, CANADA

Current address: U. E. R. Mathématiques de la Décision, Université de Paris IX (Dauphine), 75775 Paris 16, France