From: Math. Programming: The State of the Art (Bonn, 1982), A. Bachem et al., eds., Springer-Verlag, 1983.

96

Generalized Subgradients in Mathematical Programming

R. T. Rockafellar*
University of Washington, Department of Mathematics, C138 Padelford Hall, GN-50, Seattle, WA 98195, USA

Abstract. Mathematical programming problems, and the techniques used in solving them, naturally involve functions that may well fail to be differentiable. Such functions often have "subdifferential" properties of a sort not covered in classical analysis, but which provide much information about local behavior. This paper outlines the fundamentals of a recently developed theory of generalized directional derivatives and subgradients.

Introduction

The theory of subgradients of convex functions is by now widely known and has found numerous applications in mathematical programming. It serves in the characterization of optimality conditions, sensitivity analysis, and the design and validation of algorithms. What is not so widely known is the surprisingly powerful extension of this theory to nonconvex functions. This has been achieved in full measure only in the last few years, following breakthroughs made by F. H. Clarke [13].

Generalized subgradients have already been applied to optimal control problems by Clarke [16], [17], [18], [19], [20], [21], and Thibault [57], to Lagrange multiplier theory and sensitivity analysis by Aubin [2], Aubin and Clarke [4], [5], Auslender [6], [7], Chaney [9], Clarke [22], Gauvin [26], Gauvin and Dubeau [27], Hiriart-Urruty [30], [31], [32], [34], Pomerol [42], Rockafellar [47], [48], [50], [52], [53], Mifflin, Nguyen and Strodiot [40], to nonlinear programming algorithms by Chaney and Goldstein [10], Goldstein [28], and Mifflin [38], [39], to stochastic programming by Hiriart-Urruty [33], and to game theory by Aubin [3]. Many more applications are surely possible and will be made when more researchers have gained familiarity with the concepts and results.

The purpose of the present article is to help matters along by reviewing the important role of nonsmooth, not necessarily convex functions in mathematical programming and describing briefly the central facts about such functions. Of course, much has to be left out. The reader can find further details in the lecture notes [49] and the many articles which are cited. Only the finite-dimensional case will be discussed, but the references often contain infinite-dimensional generalizations.

* Research supported in part by the Air Force Office of Scientific Research, United States Air Force, under grant no. F4960-82-K-0012.

1. The Role of Nonsmooth Functions

A real-valued function is said to be *smooth* if it is of class C^1, i.e. continuously differentiable. The functions which appear in classical problems in physics and engineering typically are smooth , but not so in economics and other areas where the operation of maximization and minimization are basic. Such operations give rise to quantities whose dependence on certain variables or parameters may not even be continuous, much less differentiable. Often these quantities nevertheless do exhibit some kind of generalized differentiability behavior, and this can be important in being able to work with them.

Let us look at a standard mathematical programming problem of the form

$$(P)\qquad \text{minimize } f_0(x) \text{ over all } x\in C \text{ satisfying } f_i(x)\begin{cases}\leq 0 & \text{for } i=1,\dots,s\\ =0 & \text{for } i=s+1,\dots,m,\end{cases}$$

where C is a subset of $\mathbb{R}^n$ and each f_i is a real-valued function on C, $i=0,1,\dots,m$. Nonsmoothness can occur in connection with (P) not only because of the way the constraint and objective functions may be expressed, but also as an inescapable feature of auxiliary problems that may be set up as an aid to solving (P). For example, duality, penalty methods and decomposition techniques often require consideration of nonsmooth functions.

It is common in operations research to see problems in which the objective function f_0 is "piecewise linear":

$$(1.1)\qquad f_0(x)=\max_{t=1,\dots,r}\varphi_t(x),$$

where each φ_t is affine (i.e. linear + constant). Then the graph of f_0 can have "creases" and "corners" where differentiability fails. Sometimes f_0 is given as the supremum in an optimization problem in which x is a parameter vector:

$$(1.2)\qquad f_0(x)=\sup_{t\in T}\varphi_t(x).$$

Here T could be a subset of some space $\mathbb{R}^d$, defined by a further system of constraints, or it could be any abstract set. With $T=\{1,\dots,r\}$, we revert from (1.2) to (1.1). Such an f_0 may fail to be smooth, but if the functions φ_t are all affine, f_0 is at least *convex*. Indeed, this is almost a characterization of convexity: a real-valued function f_0 on $\mathbb{R}^n$ is convex if and only if it can be expressed in the form (1.2) for some collection of affine functions h_t. (Convexity relative only to a convex subset C or $\mathbb{R}^n$ can be characterized similarly when semicontinuity properties are present; see [43, § 12].)

Thus, problems where merely the convexity of f_0 is a natural assumption, as in many economic applications, can well involve nonsmoothness. On the other hand, smooth convex functions do exist, so a formula of type (1.2) does not *preclude* smoothness of f_0.

How can one tell in a particular instance of (1.2) with smooth functions φ_t whether f_0 is smooth or not? More generally, what partial differentiability properties of f_0 can be deduced from (1.2)? Such questions are not covered by traditional mathematics, but they have attracted attention in optimization theory. Some answers will be provided below.

Much of what has just been said about f_0 also applies to the constraint functions in (P), at least for the inequality constraints. Such a function f_i could be expressed as in (1.1) or (1.2); in particular, f_i might be convex without being smooth. Of course, if

$$f_i(x) = \sup_{t\in T_i} \varphi_{it}(x), \tag{1.3}$$

with smooth φ_{it}'s, the single constraint $f_i(x) \leqq 0$ is equivalent to a system of smooth constraints:

$$\varphi_{it}(x) \leqq 0 \quad \text{for every } t \in T_i. \tag{1.4}$$

In this sense it may seem artificial to be worried about f_i being nonsmooth. Why not just write (1.4) rather than $f_i(x) \leqq 0$?

Actually, the thinking can go just as well in the other direction. The set T_i may be infinite, or if finite, very large. Thus it may not be practical to treat (1.4) in full. Lumping (1.4) together as a single condition $f_i(x) \leqq 0$ is a form of constraint *aggregation*. If enough is known about the behavior of formulas of type (1.3), it may be possible to treat f_i directly, generating only the particular φ_{it}'s needed locally at any time.

Note that a function of form (1.2) might have $+\infty$ as a value, unless restrictions are imposed on T and the way that $\varphi_t(x)$ depends on t. Minimizing such an extended-real-valued function is definitely a matter of interest too, for instance in connection with duality-based methods.

Duality is a major source of nonsmooth optimization problems. The ordinary dual associated with (P) is

$$\text{(D)} \qquad \text{maximize } g(y) \text{ over all } y \in Y$$

where

$$Y = \{y = (y_1, \ldots, y_m) \in R^m \mid y_i \geqq 0 \quad \text{for } i = 1, \ldots, s\}, \tag{1.5}$$

$$g(y) = \inf_{x\in C} \varphi_x(y) \quad \text{for } \varphi_x(y) = f_0(x) + \sum_{i=1}^m y_i f_i(x). \tag{1.6}$$

Except for the reversal of maximization and minimization, we can identify this as an instance of a problem of type (P) where the objective function is represented as in (1.2) but might not be finite everywhere. Since $\varphi_x(y)$ is affine in y, g is a concave function, quite possibly nonsmooth.

Several important methods for solving (P) proceed by way of (D), despite nonsmoothness. In Dantzig-Wolfe decomposition, a cutting plane algorithm is used to maximize g. If C happens to be a convex polyhedron and the functions f_i are affine, g is piecewise linear and the maximization of g can be formulated simply as a problem in linear programming. The number of "pieces" of g can be so enormous, however, that this approach is impractical. More hope lies in treating g as a nonsmooth function which nevertheless has useful "subdifferential" properties, as will be discussed later. This is also the case in integer pro-

gramming methods which solve (D) partially in order to obtain a lower bound for the minimum in (P) (cf. Held, Wolfe and Crowder [29]).

The dual of a "geometric" programming problem in the sense of Duffin, Peterson and Zener [25] provides another example. This consists of maximizing, subject to linear constraints, a certain finite, concave function of nonnegative variables which has a formula in terms of logarithms that looks harmless enough. The function fails, however, to be differentiable at boundary points of its domain.

Exact penalty methods, which replace (P) by a single minimization problem with no constraints, or at least simpler constraints, also lead to nonsmoothness. Under mild assumptions on (P), it will be true that for $r > 0$ sufficiently large the optimal solutions to (P) are the same as those to the problem

$$\text{minimize } f(x) \text{ over all } x \in C, \text{ with}$$

$$f(x) = f_0(x) + r\left(\sum_{i=1}^{s} [f_i(x)]_+ + \sum_{i=s+1}^{m} |f_i(x)|\right), \tag{1.7}$$

where

$$[\alpha]_+ = \max\{\alpha, 0\}, \tag{1.8}$$

and of course $|\alpha| = \max\{\alpha, -\alpha\}$. This type of penalty function has been considered by Zangwill [62], Petrzykowski [41], Howe [36], Conn [12] and Chung [11], for instance. If the functions f_i are smooth, the f in (1.7) can be represented as in (1.1) with smooth φ_t's. Thus f is piecewise smooth, not everywhere differentiable. Another exact penalty approach, based on the (quadratic-type) augmented Lagrangian for (P) (see [55]), preserves first-order smoothness of the f_i's but can create discontinuities in second derivatives.

So far we have been discussing nonsmoothness of the kind which arises from representations (1.2), (1.3). This is relatively easy to deal with, but a trickier kind of nonsmoothness is encountered in parametric programming and the study of Lagrange multipliers. Suppose that in place of (P) we have a problem which depends on a parameter vector $v = (v_1, \ldots, v_d) \in R^d$:

$$(P_v) \qquad \begin{aligned} &\text{minimize } f_0(v, x) \text{ over all } x \in C(v) \subset R^n \\ &\text{such that } f_i(v, x) \begin{cases} \geqq 0 & \text{for } i = 1, \ldots, s, \\ = 0 & \text{for } i = s+1, \ldots, m. \end{cases} \end{aligned}$$

Let $p(v)$ denote the optimal value in (P_v). Then p is a function on R^d whose values can in general be not just real numbers but $-\infty$ and $+\infty$ (the latter for v such that (P_v) is infeasible). What can be said about generalized differentiability properties of p?

Actually, p can be represented by

$$p(v) = \inf_{x\in R^n} \varphi_x(v), \tag{1.9}$$

where

$$\varphi_x(v) = \begin{cases} f_0(v, x) & \text{if } x \text{ is a feasible solution to } (P_v), \\ \infty & \text{otherwise.} \end{cases} \tag{1.10}$$

Thus p is the pointwise infimum of a collection of functions on R^d. These functions φ_x are not smooth, though, even if the f_i's are smooth, because of the jump to $+\infty$ when the parameter vector v shifts into a range where the feasibility of a given x is lost. For this reason the nonsmoothness of p is harder to analyze, yet strong results have been obtained (cf. [48] and its references). Generally speaking, directional derivative properties of p at v are closely related to the possible Lagrange multiplier vectors associated with optimality conditions for (P_v).

The simplest case is the one where $f_i(v,x)$ is convex jointly in (v,x), and the set of (x,v) satisfying $x\in C(v)$ is convex. Then p is a convex function. Even so, p may have nonfinite values.

The importance of p in sensitivity analysis is clear: we may need to know the rate of change of the optimal value $p(v)$, in some sense, as v varies. The components of v may be economic quantities subject to fluctuation or control. In Benders decomposition, (P_v) is just a subproblem: the real task consists of minimizing in x and v jointly, subject to the given conditions. The residual problem, or master problem, connected with this formulation, is that of minimizing $p(v)$ over all $v\in R^n$. Obviously, generalized differential properties of p can have much to do with success in this task.

Finally, we wish to point out that in the broad picture of optimization theory there are other nonsmooth functions worthy of consideration. Prime examples are the *indicators* ψ_C of sets $C\subset R^n$:

$$\psi_C(x)=\begin{cases}0 & \text{if } x\in C,\\ \infty & \text{if } x\notin C.\end{cases} \tag{1.11}$$

When C is a convex set, ψ_C is a convex function. "Tangential" properties of C correspond to "differential" properties of ψ_C. The study of these provides a bridge between geometry and analysis that leads to a deeper understanding of many topics, such as the characterization of optimal solutions.

2. Useful Classes of Functions

As we have observed, it is not enough just to consider real-valued functions: the values $+\infty$ and $-\infty$ sometimes need to be admitted. We shall speak of a function $f: R^n\to[-\infty,\infty]$ as *proper* if $f(x)>-\infty$ for all x, and $f(x)<\infty$ for at least one x. The set

$$\operatorname{dom} f=\{x\in R^n \mid f(x)<\infty\} \tag{2.1}$$

is called the *effective domain* of f, and

$$\operatorname{epi} f=\{(x,\alpha)\in R^{n+1}\mid f(x)\le\alpha\} \tag{2.2}$$

the *epigraph* of f. For most purposes, there is little generality lost in concentrating on the case where f is *lower semicontinuous*:

$$f(x)=\liminf_{x'\to x} f(x') \quad \text{for all } x\in R^n. \tag{2.3}$$

This property holds if and only if epi f is a closed set.

Convex analysis [43] serves as a guide in developing a theory of differentiability in such a general context. The function f is convex if and only if epi f is a convex set. We shall see below that amazingly much of the theory of one-sided directional derivatives and subgradients of convex functions has a natural extension to the class of all lower semicontinuous proper functions on R^n. Incidentally, it is easy to give conditions on the parameterized problem (P_v) in § 1 which ensure that the corresponding optimal value function p is lower semicontinuous and proper; see [48].

Although a function f may be extended-real-valued in the large, we are often interested in situations where f is finite in a neighborhood of a certain point x and has stronger properties in such a neighborhood as well. Among the most important properties to be considered in such a context is Lipschitz continuity: f is *Lipschitzian* on an open set U if f is real-valued on U and there exists a number $\lambda\ge 0$ such that

$$|f(x')-f(x)|\le\lambda|x'-x| \quad \text{when } x\in U,\ x'\in U. \tag{2.4}$$

(Here $|\cdot|$ stands for the Euclidean norm on R^n.) This condition can also be put in the form

$$[f(x+th)-f(x)]/t\le\lambda|h| \quad \text{when } x\in U,\ x+th\in U,\ t>0. \tag{2.5}$$

Thus it expresses a bound on difference quotients.

We shall say f is *Lipschitzian around* $\bar{x}$ if it is Lipschitz continuous in some neighborhood of $\bar{x}$, and that f is *locally Lipschitzian* on an open set U if f is Lipschitzian around each $x\in U$. The distinction between this and simply being Lipschitzian on U is that the modulus λ in (2.4) need not be valid for all of U, but may change from one neighborhood to another. Two classical results on local Lipschitz continuity may be cited.

Theorem 2.1 [43, § 10]. *A convex function is locally Lipschitzian on any open subset of R^n where its values are finite.*

Theorem 2.2 (Rademacher; cf. [56]). *A locally Lipschitzian function f on an open set $U\subset R^n$ is differentiable except at a negligible set of points in U.*

A *negligible* set is a set of measure zero in the Lebesgue sense: for any $\varepsilon>0$, it can be covered by a sequence of balls whose total n-dimensional volume does not exceed ε.

A more subtle form of Lipschitz continuity that will be important to us below is the following. Suppose f is a lower semicontinuous, proper function on R^n, and let $\bar{x}\in\operatorname{dom} f$. We say that f is *directionally Lipschitzian at $\bar{x}$ with respect to the vector $\bar{h}$* if

$$\begin{gathered}\text{there exist } \varepsilon>0 \text{ and } \lambda\ge 0 \text{ such that}\\ [f(x+th)-f(x)]/t\le\lambda|h| \text{ when } f(x)\le f(\bar{x})+\varepsilon,\\ |x-\bar{x}|\le\varepsilon,\ |h-\bar{h}|\le\varepsilon,\ 0<t<\varepsilon.\end{gathered} \tag{2.6}$$

If $\bar{h}=0$, this reduces to f being Lipschitzian around $\bar{x}$.

We say simply that f is *directionally Lipschitzian at* $\bar{x}$ if (2.6) holds for at least one $\bar{h}$. Note that f need not be finite on a neighborhood of $\bar{x}$ for this condition to hold. For example, if f is *convex*, then f is directionally Lipschitzian at x with respect to any h such that $x+th \in \operatorname{int}(\operatorname{dom} f)$ for $t>0$ sufficiently small [46]. Thus if $\operatorname{int}[\operatorname{dom} f] \neq \emptyset$, f is directionally Lipschitzian at every $x \in \operatorname{dom} f$.

Another example: if f is a *nondecreasing* function on R^n in the sense that

$$f(x) \leqq f(x') \quad \text{when } x'-x \in R^n_+, \tag{2.7}$$

then f is directionally Lipschitzian at every $x \in \operatorname{dom} f$. (Consider $h \in \operatorname{int} R^n_+$.)

From a geometrical point of view, f is directionally Lipschitzian at x if and only if epi f has "Lipschitzian boundary" in a neighborhood of $(\bar{x}, f(x))$; see [44].

Moving to properties stronger than Lipschitz continuity but still short of actual smoothness, we come upon two highly significant classes of functions already suggested by the discussion in § 1. Let us say that f is *subsmooth (or lower-$\mathcal{C}^1$) around* x if there is an open neighborhood X of x and a representation

$$f(x) = \max_{t \in T} \varphi_t(x) \quad \text{for all } x \in X, \tag{2.8}$$

where T is a compact topological space, each φ_t is of class $\mathcal{C}^1$, and the values of φ_t and its first partial derivatives are continuous not only with respect to $x \in X$ but $(t, x) \in T \times X$. (In particular, T could be any finite index set in the discrete topology. Then the continuity requirements in t are trivial: f is just expressible locally as the pointwise maximum of a finite collection of smooth functions as in (1.1).) We shall say f is *subsmooth on* U, an open set in R^n, if f is subsmooth around every $x \in U$. (The representation (2.8) may be different for different regions of U.) Obviously every smooth function is also subsmooth (take T to be a singleton).

In a similar vein, we shall call f *strongly subsmooth (of order r) on U* (or *lower-$\mathcal{C}^r$*), if in the local representations (2.8) around points of U the functions φ_t are actually of class $\mathcal{C}^r$ with $2 \leqq r \leqq \infty$, and their partial derivatives up through order r depend continuously on (t, x).

Theorem 2.3 [13]. *If f is subsmooth (or strongly subsmooth) on an open set $U \subset R^n$, then f is locally Lipschitzian on U.*

Theorem 2.4 [51]. *The classes of strongly subsmooth functions of order r on U, for $2 \leqq r \leqq \infty$, all coincide, so that one can speak simply of a single class of strongly subsmooth functions on U without reference to any particular r. There do exist subsmooth functions which are not strongly subsmooth, however.*

Theorem 2.5 [51]. *A real-valued function f on an open set $U \subset R^n$ is strongly subsmooth on U if and only if in some open convex neighborhood X of each $x \in U$ there is a representation:*

$$f(x) = g(x) + h(x) \quad \text{for all } x \in X, \quad \textit{with } g \textit{ convex}, \ h \textit{ of class } \mathcal{C}^2. \tag{2.9}$$

Then there exist such representations with h actually of class $\mathcal{C}^\infty$, in fact with $h(x) = -\rho|x|^2$, $\rho > 0$.

Corollary 2.6. *A convex function is strongly subsmooth on any open subset of R^n where its values are finite.*

One other class of functions of great importance in optimization deserves mention: the *saddle functions*. Suppose $f(y, z)$ is convex in $y \in Y$ and concave in $z \in Z$, where $Y \times Z \subset R^n$ is convex. Then f is locally Lipschitzian on the interior of $Y \times Z$ [43, Theorem 35.1] and has many other properties, such as the existence of one-sided directional derivatives [43, Theorem 35.6], but f is not subsmooth. More generally, one could consider the class of all functions expressible locally as linear combinations of such saddle functions along with convex and concave functions. No abstract characterization of this class is known.

3. Sublinear Functions Representing Convex Sets

In the classical approach to differentiability, one seeks to approximate a function f around a point x by a *linear* function l:

$$f(x+h) - f(x) = l(h) + o(|h|). \tag{3.1}$$

This l expresses directional derivatives of f at x with respect to various vectors h. Next, one uses the duality between linear functions l on R^n and vectors $y \in R^n$ to define the *gradient* of f at x: there is a unique y such that

$$l(h) = y \cdot h \quad \text{for all } h \in R^n. \tag{3.2}$$

This y is the gradient $\nabla f(x)$.

We cannot limit ourselves merely to linear functions l as approximations, in trying to capture the generalized differentiability properties of functions f belonging to the various classes mentioned in § 2. A broader duality correspondence than the one between vectors and linear functions is therefore required. The correspondence about to be described replaces vectors $y \in R^n$ by closed convex sets $Y \subset R^n$.

A function $l: R^n \to [-\infty, +\infty]$ is said to be *sublinear* if l is convex, positively homogeneous ($l(\lambda h) = \lambda l(h)$ for $\lambda > 0$), and $l(0) < \infty$. These conditions mean that epi l is a closed convex cone in R^{n+1} containing the origin. Every linear function is in particular sublinear.

Theorem 3.1 [43, § 13]. *Let l be a sublinear function on R^n which is lower semicontinuous. Then either l is proper, with $l(0) = 0$, or l has no values other than $+\infty$ and $-\infty$. In the latter case, the set of points h where $l(h) = -\infty$ is a closed convex cone containing 0.*

Theorem 3.2 [43, § 13]. *There is a one-to-one correspondence between proper, lower semicontinuous, sublinear functions l on R^n and the nonempty, closed, convex subsets Y of R^n, given by*

$$l(h) = \sup_{y \in Y} y \cdot h \quad \textit{for all } h \in R^n \tag{3.3}$$

$$Y = \{y \in R^n \mid y \cdot h \leqq l(h) \quad \textit{for all } h \in R^n\}. \tag{3.4}$$

The special case of singleton sets Y yields the classical correspondence between linear functions and vectors.

The Euclidean norm $l(h)=|h|$ is an example of a sublinear function which is not linear. It corresponds to the ball $Y=\{y \mid |y|\leq 1\}$. The sublinear function

$$l(h)=\max\{a_1\cdot h,\ldots,a_m\cdot h\} \tag{3.5}$$

corresponds to the polytope

$$Y=\mathrm{co}\{a_1,\ldots,a_m\}. \tag{3.6}$$

The function

$$l(h)=\begin{cases}0 & \text{if } h\in K\\ \infty & \text{if } h\notin K,\end{cases} \tag{3.7}$$

where K is a closed convex cone containing 0, corresponds to the polar cone $Y=K^{\circ}$.

A finite sublinear function on R^n, being convex, is continuous (cf. Theorem 2.1), hence certainly lower semicontinuous. This gives the following special case of Theorem 3.2.

Corollary 3.3 [43]. *Formulas* (3.3), (3.4), *give a one-to-one correspondence between the finite sublinear functions l on R^n and the nonempty compact convex subsets Y of R^n.*

The duality between finiteness of l and boundedness of Y extends to a more detailed relationship. Here we recall that since Y is a closed convex set, if there is no halfspace having a nonempty bounded intersection with Y, then there is a vector $z\neq 0$ such that the line $\{y+tz \mid t\in R\}$ is included in Y for every $y\in Y$ [43, § 13]. In the latter case, Y is just a bundle of parallel lines.

Theorem 3.4 [43, § 13]. *Under the correspondence in Theorem* 3.2, *one has* $h\in \mathrm{int}(\mathrm{dom}\, l)$ *if and only if for some $\beta\in R$ the set $\{y\in Y \mid y\cdot h\geq\beta\}$ is bounded and nonempty. Thus the convex cone* $\mathrm{dom}\, l$ *has nonempty interior if and only if Y cannot be expressed as a bundle of parallel lines.*

The pattern we ultimately wish to follow is that of defining for a given function f and point x a kind of generalized directional derivative which is a *lower semicontinuous sublinear function* l of the direction vector h. The elements y of the corresponding closed convex set Y will be the "subgradients" of f at x.

4. Contingent Derivatives

For a function $f: R^n\to[-\infty,\infty]$ and a point x where f is finite, the ordinary one-sided *directional derivative of f at x with respect to h* is

$$f'(x;h)=\lim_{t\downarrow 0}\frac{f(x+th)-f(x)}{t}, \tag{4.1}$$

if it exists (possible as $+\infty$ or $-\infty$). This concept of a derivative has its uses, but it is inadequate for dealing with functions that are not necessarily continuous at x. One of the drawbacks is that difference quotients are only considered along rays emanating from x; no allowance is made for "curvilinear behavior".

As an illustration, consider the indicator function of the curve $C=\{x=(x_1,x_2)\in R^2 \mid x_2=x_1^3\}$:

$$f(x_1,x_2)=\begin{cases}0 & \text{if } x_2=x_1^3\\ \infty & \text{if } x_2\neq x_1^3.\end{cases} \tag{4.2}$$

Obviously $f'(0,0;h_1,h_2)=\infty$ for all nonzero vectors $h=(h_1,h_2)$. There is nothing to distinguish the vectors tangent ot the curve C at $(0,0)$, namely the ones with $h_2=0$, from the others. Although the function (4.2) may be seen as an extreme case, it reflects a lack of responsiveness of ordinary directional derivatives that is rather widespread.

Certainly in handling optimal value functions like the p introduced towards the end of § 1, it is essential to consider more subtle limits of difference quotients. Instead of just sequences of points $x^k=x+t_k h$ with $t_k\downarrow 0$, one at least needs to look at the possibility of $x^k=x+t_k h^k$ with $t_k\downarrow 0$ and $h^k\to h$. (A sequence $\{x^k\}$ which can be expressed in the latter form with $h\neq 0$ is said to *converge to x in the direction of h.*)

A useful concept in this regard is that of the *contingent derivative* (also called the *lower Hadamard derivative*) of f at x with respect to h:

$$f^{\bullet}(x;h)=\liminf_{\substack{h'\to h\\ t\downarrow 0}}\frac{f(x+th')-f(x)}{t} = \lim_{\varepsilon\downarrow 0}\left[\inf_{|h'-h|\leq\varepsilon}\frac{f(x+th')-f(x)}{t}\right]. \tag{4.3}$$

In contrast to the ordinary derivative $f'(x;h)$, this kind of limit always exists, of course. In example (4.2) one gets

$$f^{\bullet}(0,0;h_1,h_2)=\begin{cases}0 & \text{if } h_2=0,\\ \infty & \text{if } h_2\neq 0.\end{cases}$$

Strong geometric motivation for the contingent derivative is provided by its epigraphical interpretation. To explain this, we recall that the *contingent cone* to a closed set $C\subset R^n$ at a point $x\in C$ is the set

$$K_C(x)=\{h \mid \exists t_k\downarrow 0,\ h^k\to h, \text{ with } x+t_k h^k\in C\}. \tag{4.4}$$

Thus $K_C(x)$ consists of the zero vector and all vectors $h\neq 0$ giving directions in which sequences in C can converge to x. It is elementary that $K_C(x)$ is closed, and that $K_C(x)$ truly is a *cone*, i.e. $\lambda h\in K_C(x)$ when $h\in K_C(x)$ and $\lambda>0$.

The contingent cone was introduced by Bouligand in the 1930's (cf. [56]). In the mathematical programming literature, it is often called the "tangent cone", but we prefer to reserve that term for another concept to be described in § 6 and keep to the classical usage. The following fact is easily deduced from the definitions.

Theorem 4.1. *For the contingent derivative function $l(h) = f^{\#}(x;h)$, epi l is the contingent cone to epi f at the point $(x, f(x))$.*

Corollary 4.2. *The contingent derivative $f^{\#}(x;h)$ is a lower semicontinuous, positively homogeneous function of h with $f^{\#}(x;0) < \infty$.*

Since contingent cones to convex sets are convex, we also have the following.

Corollary 4.3. *Suppose f is a proper convex function, and let $x \in \operatorname{dom} f$. Then the contingent derivative function $l(h) = f^{\#}(x;h)$ is not only lower semicontinuous but sublinear.*

The convex case also exhibits a close relationship between ordinary directional derivatives and contingent derivatives.

Theorem 4.4 [43]. *Suppose f is a proper convex function on R^n, and let x dom f. Then $f'(x;h)$ exists for every h (possibly as $+\infty$ or $-\infty$) and*

$$f^{\#}(x;h) = \liminf_{h' \to h} f'(x;h'). \tag{4.5}$$

In fact $f^{\#}(x;h) = f'(x;h)$ for every h such that $x + th \in \operatorname{int}(\operatorname{dom} f)$ for some $t \to 0$.

The sublinearity in Corollary 4.3 is the basis of the well known subgradient theory of convex analysis [43], elements of which will be reviewed in §7. Unfortunately, contingent cones to nonconvex sets generally are *not* convex. By the same token, the contingent derivative function $l(h) = f^{\#}(x;h)$ is generally *not* convex when f is not convex, except for certain important cases which will be noted in §6. The contingent derivative does not, therefore, lead to a robust theory of subgradients in terms of the duality correspondence in §3.

What is needed is a concept of directional derivative possessed of an inherent convexity. Such a concept has been furnished by Clarke [13] for locally Lipschitzian functions. We shall present it in the next section in an extended form which to a certain extent is useful also in connection with functions which are not locally Lipschitzian. The full concept of derivative needed in treating such general functions, the so-called "subderivative", will not be discussed until §6, however.

5. Clarke Derivatives

For a lower semicontinuous function f and a point x where f is finite, we define the *extended Clarke derivative* of f at x with respect to the vector h to be

$$f^{\circ}(x;h) = \limsup_{\substack{x' \to_f x \\ h' \to h \\ t \downarrow 0}} \frac{f(x'+th') - f(x')}{t} \tag{5.1}$$

where the notation is used that

$$x' \to_f x \iff \begin{cases} x' \to x \\ f(x') \to f(x). \end{cases} \tag{5.2}$$

If f happens to be continuous at x, one has $f(x') \to f(x)$ when $x' \to x$, so the extra notation is unnecessary. The case where f is not necessarily continuous at x is of definite interest, however, as the following fact (immediate from the definition) well indicates.

Theorem 5.1. *Let f be a lower semicontinuous, proper function on R^n and let $x \in \operatorname{dom} f$. Then for a vector h one has $f^{\circ}(x;h) < \infty$ if and only if f is directionally Lipschitzian at x with respect to h.*

Clarke's original definition of $f^{\circ}(x;h)$ in [13] applied only to the case where f actually is Lipschitzian around x. The formula then takes on a simpler form.

Theorem 5.2. *Suppose f is Lipschitzian around x. Then*

$$f^{\circ}(x;h) = \limsup_{\substack{x' \to x \\ t \downarrow 0}} \frac{f(x'+th) - f(x')}{t} \tag{5.3}$$

A striking feature of the extended Clarke derivative is its inherent convexity.

Theorem 5.3. *Let f be a lower semicontinuous, proper function on R^n, and let $x \in \operatorname{dom} f$. Then the extended Clarke derivative function $l(h) = f^{\circ}(x;h)$ is convex and positively homogeneous, but not necessarily lower semicontinuous (in fact* dom l *is an open convex cone).*

If f is Lipschitzian around x, a property equivalent to having $\infty > l(0) = f^{\circ}(x;0)$, then l is not only sublinear but finite everywhere (hence continuous).

This result too is simply a consequence of the definitions. We take note now of a very important case where extended Clarke derivatives, contingent derivatives and ordinary directional derivatives all have the same value.

Theorem 5.4. *Suppose f is subsmooth around x. Then $f'(x;h)$ exists as a real number (finite) for every $h \in R^n$, and*

$$f'(x;h) = f^{\#}(x;h) = f^{\circ}(x;h). \tag{5.4}$$

In fact for any local representation of f as in (2.8) (with φ_t smooth, and $\varphi_t'(x;h)$ continuous jointly in t and x for each h) one has

$$f'(x;h) = \max_{t \in T_x} \varphi_t'(x;h) \quad \text{for all } h \in R^n, \tag{5.5}$$

where

$$T_x = \{t \in T \mid \varphi_t(x) = f(x)\} = \arg\max_{t \in T} \varphi_t(x). \tag{5.6}$$

Formula (5.5) was first proved by Danskin [24], and the equation $f'(x;h) = f^{\circ}(x;h)$ was established by Clarke [13]. The proof of (5.5) shows that

$f^*(x; h)$ is given by the same maximum, in fact directional derivatives exist in the Hadamard sense:

$$\lim_{\substack{h' \to h \\ t \downarrow 0}} \frac{f(x + th') - f(x)}{t} = \max_{i \in I_x} \varphi_i'(x; h) \quad \text{for all } h \in R^n \tag{5.7}$$

A subgradient version of (5.5) will be given in Theorem 7.3 and Corollary 7.4. It is good to note, however, that (5.4) associates with each x a finite, sublinear function l that depends only on f and not on any particular max representation (2.8).

For functions which are not locally Lipschitzian or even directionally Lipschitzian, the extended Clarke derivative conveys little information. Thus for the function in (4.2), one has $f^\circ(0, 0; h_1, h_2) = \infty$ for all (h_1, h_2). Indeed, if f is not directionally Lipschitzian at x, one necessarily has $f^\circ(x; h) = \infty$ for all h by Theorem 5.1.

6. Subderivatives and the Clarke Tangent Cone

A type of directional derivative will now be described which is able to serve as the foundation for a theory of subgradients of very general functions. It agrees with the contingent derivative and extended Clarke derivative in the important cases where those derivatives yield a lower semicontinuous, sublinear function l of the direction vector h. But, it yields such an l no matter what the circumstances with the other derivatives.

Let f be lower semicontinuous and proper, and let $x \in \operatorname{dom} f$. The *subderivative* of f at x with respect to h is defined to be

$$f^\uparrow(x; h) = \lim_{\varepsilon \downarrow 0} \left[\limsup_{\substack{x' \to_f x \\ t \downarrow 0}} \left[\inf_{|h' - h| \le \varepsilon} \frac{f(x' + th') - f(x')}{t} \right] \right], \tag{6.1}$$

where again the notation (5.2) is used as a shorthand. This rather complicated limit is a sort of amalgam of the ones used in defining $f^*(x; h)$ and $f^\circ(x; h)$; cf. (4.3) and (5.1). It was first given by Rockafellar [45], but as we shall see below, it is closely related to a geometrical notion of Clarke [13]. The initial difficulties in appreciating the nature of $f^\uparrow(x; h)$ are far outweighed by its remarkable properties. A convincing case can be made for this derivative as the natural one to consider for general functions f. Clearly

$$f^*(x; h) \le f^\uparrow(x; h) \le f^\circ(x; h) \quad \text{for all } h. \tag{6.2}$$

Theorem 6.1 [45]. *Let f be a lower semicontinuous, proper function on R^n, and let $x \in \operatorname{dom} f$. Then the subderivative function $l(h) = f^\uparrow(x; h)$ is lower semicontinuous and sublinear.*

Just as the contingent derivative corresponds to a contingent cone to the epigraph of f (cf. Theorem 4.1), the subderivative corresponds to another geometrical concept. For a closed set $C \subset R^n$, the *Clarke tangent cone* at a point $x \in C$ is defined to be

$$T_C(x) = \{h \mid \forall x^k \to x \text{ in } C,\ t_k \downarrow 0,\ \exists h^k \to h \text{ with } x^k + t_k h^k \in C\}. \tag{6.3}$$

This is always a *closed convex cone containing* 0, a surprising fact in view of the lack of any convexity assumptions whatsoever on C. For a direct proof, see [44]. The cone $T_C(x)$ was originally defined by Clarke [13] in a more circuitous manner, but formula (6.3) turned out to be implicit in one of his results (see [13, Prop. 3.7]). We shall say more about the properties of this cone in Theorem 6.8. Obviously the Clarke tangent cone is a subset of the contingent cone in §4:

$$T_C(x) \subset K_C(x). \tag{6.4}$$

Theorem 6.2 [45]. *Let f be a lower semicontinuous function on R^n, and let $x \in \operatorname{dom} f$. Let $l(h) = f^\uparrow(x; h)$. Then the epigraph of l is the Clarke tangent cone to the epigraph of f at $(x, f(x))$.*

The relationship between subderivatives and the Clarke derivatives of the preceding section is quite simple.

Theorem 6.3 [46]. *Let f be a lower semicontinuous, proper function on R^n, and let $x \in \operatorname{dom} f$. Then*

$$\operatorname{int}\{h \in R^n \mid f^\uparrow(x; h) < \infty\} = \{h \in R^n \mid f^\circ(x; h) < \infty\}, \tag{6.5}$$

and on this open convex cone one has $f^\uparrow(x; h) = f^\circ(x; h)$.

Corollary 6.4. *Let f be a lower semicontinuous, proper function on R^n, and let $x \in \operatorname{dom} f$. Then one has the epigraphical relationship*

$$\operatorname{int}\{(h, \beta) \in R^{n+1} \mid \beta \ge f^\uparrow(x; h)\} = \{(h, \beta) \in R^{n+1} \mid \beta > f^\circ(x; h)\}. \tag{6.6}$$

Corollary 6.5. *Let f be a lower semicontinuous, proper function on R^n, and let $x \in \operatorname{dom} f$. Let $\bar{h}$ be an element of the cone (6.5). Then*

$$f^\uparrow(x; h) = \lim_{\varepsilon \downarrow 0} f^\circ(x; (1 - \varepsilon)h + \varepsilon \bar{h}) \quad \textit{for all } h \in R^n. \tag{6.7}$$

Corollaries 6.4 and 6.5 follow from Theorem 6.3 by way of basic facts about the closures and interiors of epigraphs of convex functions [43, Lemma 7.3 and Corollary 7.5.1]. Of course, the cones (6.5) and (6.6) are nonempty only when f is directionally Lipschitzian at x; cf. Theorem 5.1. Thus in the directionally Lipschitzian case, $f^\uparrow$ can be constructed from the simpler function f° by taking $f^\uparrow(x; h) = f^\circ(x; h)$ on the cone (6.5) and limits (6.7) at boundary points, but when f is not directionally Lipschitzian, no help can be obtained from f° at all.

The relationship between subderivatives and contingent derivatives is not as easy to describe, although it does turn out that subderivatives can be expressed as certain limits of contingent derivatives. First of all, let us introduce the terminology: f is *subdifferentially regular* at x if $f^\uparrow(x; h) = f^*(x; h)$ for all h.

Theorem 6.6 [13], [46]. *If f is a lower semicontinuous and convex on a neighborhood of x (a point where f is finite), or if f is subsmooth around x, then f is subdifferentially regular at x.*

Theorem 6.7 [46]. *If f is a lower semicontinuous, proper function on R^n and subdifferentially regular at the point $x\in \operatorname{dom} f$, then for every h in the cone (6.5) the ordinary directional derivative $f'(x;h)$ exists, and $f'(x;h)=f^{\uparrow}(x;h)$.*

Theorem 6.7 provides a natural generalization of Theorems 4.4 and 5.3, which correspond to the cases in Theorem 6.6. An example of a subdifferentially regular function not covered by Theorem 6.6 is given by (4.2); this function is not covered by Theorem 6.7 either, since it is not directionally Lipschitzian; the set (6.5) is empty for every $x\in \operatorname{dom} f$.

To gain deeper understanding of the cases where f is subdifferentially regular, we can appeal to a result of Cornet [23] about the relationship between the Clarke tangent cone and the contingent cone.

Theorem 6.8 [23]. *Let C be a closed set in R^n and let $x\in C$. Then*

(6.8) $$T_C(x)=\liminf_{\substack{x'\to x\\ x'\in C}} K_C(x') := \{h \mid \forall x^k\to x \text{ with } x^k\in C,\ \exists h^k\to h \text{ with } h^k\in K_C(x^k)\}.$$

Thus $T_C(x)=K_C(x)$ if and only if the multifunction $K_C: x\mapsto K_C(x)$ is lower semicontinuous at x relative to C.

This can be applied to the epigraphs of $f^{\uparrow}(x;\cdot)$ and $f^{\#}(x;\cdot)$ due to Theorems 4.1 and 6.2. The limit in (6.8) can be expressed in function terms using a fact in [46, Proposition 1]. We then obtain the following formula for $f^{\uparrow}$ in terms of $f^{\#}$.

Theorem 6.9. *Let f be lower semicontinuous, and let x be a point where f is finite. Then for all h.*

(6.9) $$f^{\uparrow}(x;h)=\lim_{\varepsilon\downarrow 0}\left[\limsup_{x'\to_f x}\left[\inf_{\|h'-h\|\le\varepsilon} f^{\#}(x';h')\right]\right].$$

This has force whether or not f is directionally Lipschitzian at x.

Corollary 6.10. *Let f be lower semicontinuous, and let x be a point where f is finite. Then f is subdifferentially regular at x if and only if for every sequence $x^k\to x$ with $f(x^k)\to f(x)$ and every h, there is a sequence $h^k\to h$ with*

$$\limsup_{k\to\infty} f^{\#}(x^k;h^k)\le f^{\#}(x;h).$$

Incidentally, in infinite-dimensional spaces Theorem 6.8 generally fails (cf. Treiman [61]).

7. Generalized Subgradients

The sublinearity and lower semicontinuity of the subderivative function in Theorem 6.1 make it possible to invoke the duality correspondence in § 3. For

an arbitrary lower semicontinuous function f on R^n and point x where f is finite, we define

(7.1) $$\partial f(x)=\{y\in R^n \mid y\cdot h\le f^{\uparrow}(x;h) \text{ for all } h\in R^n\}.$$

The vectors $y\in\partial f(x)$ are called *subgradients* (or *generalized gradients*) of f at x. This terminology is totally in harmony with other usage, e.g. in convex analysis, as will be seen in a moment. From Theorems 3.1, 3.2 and 6.1 we immediately see the exact connection between subderivatives and subgradients.

Theorem 7.1. *Let f be a lower semicontinuous, proper function on R^n, and let $x\in\operatorname{dom} f$. Then $\partial f(x)$ is a closed convex set. One has $\partial f(x)\neq\emptyset$ if and only if $f^{\uparrow}(x;0)>-\infty$, in which case*

(7.2) $$f^{\uparrow}(x;h)=\sup_{y\in\partial f(x)} y\cdot h \quad \text{for all } h.$$

Before drawing some general facts from these formulas, we look at some particular classes of functions.

Theorem 7.2. *Suppose f is convex, finite at x, and lower semicontinuous on a neighborhood of x. Then*

(7.3) $$\begin{aligned}\partial f(x)&=\{y \mid y\cdot h\le f'(x;h) \text{ for all } h\}\\ &=\{y \mid f(x+th)\ge f(x)+y\cdot th \text{ for all } t>0 \text{ and } h\}.\end{aligned}$$

The first equation is valid by Theorem 4.4 and the subdifferential regularity asserted in Theorem 6.6. The second equation then holds, because

(7.4) $$f'(x;h)=\inf_{t>0}\frac{f(x+th)-f(x)}{t}$$

in the convex case. The second expression in (7.3) (where one could just as well take $t=1$) is the definition customarily used for the subgradients of a convex function (cf. [43, § 23]). Thus Theorem 7.2 lays to rest any doubts about the present terminology versus terminology already established in convex analysis.

Theorem 7.3. *Suppose f is subsmooth around x. Then for any local representation of f as in (2.8) (with the gradient $\nabla\varphi_t(x)$ depending continuously not just on x but on (t,x)) one has*

(7.5) $$\partial f(x)=\operatorname{co}\{\nabla\varphi_t(x) \mid t\in T_x\},$$

where T_x is the set in (5.6) and "co" stands for convex hull.

Theorem 7.3 follows at once from the subdifferential regularity of subsmooth functions (Theorem 6.6) and the derivative equations in Theorem 5.4. The set Y on the right side of (7.5) is compact and convex, and the formula in (5.5) asserts that the sublinear function corresponding to this Y as in Theorem 3.2 is $l=f'(x;\cdot)$.

Corollary 7.4. *Suppose $f(x) = \max\{\varphi_1(x), \ldots, \varphi_r(x)\}$, where each φ_i is smooth. Then $\partial f(x)$ is the polytope generated by the gradients $\nabla\varphi_i(x)$ of the functions φ_i which are active at x, in the sense that $\varphi_i(x) = f(x)$.*

Now we look at Lipschitzian cases. The next result combines Theorem 6.3 with Theorem 3.4 for $l(h) = f^{\uparrow}(x; h)$.

Theorem 7.5. *Let f be a lower semicontinuous, proper function on R^n, and let $x \in \operatorname{dom} f$. Then f is directionally Lipschitzian at x if and only if $\partial f(x)$ is nonempty but not expressible simply as a bundle of parallel lines. In that case one has*

$$\partial f(x) = \{y \mid y \cdot h \leq f^{\circ}(x; h) \text{ for all } h\}. \tag{7.6}$$

Theorem 7.6 [44]. *Let f be a lower semicontinuous, proper function on R^n, and let $x \in \operatorname{dom} f$. Then f is Lipschitzian around x if and only if $\partial f(x)$ is nonempty and bounded.*

Of course (7.5) holds too when f is Lipschitzian around x, since that corresponds to f being directionally Lipschitzian with respect to $h = 0$. The necessity of the condition in Theorem 7.6 was observed by Clarke in his original paper [13]. Clarke also furnished (7.6) as a characterization of subgradients of locally Lipschitzian functions (see [14]), but his definition of $\partial f(x)$ for general lower semicontinuous functions, although equivalent to the one presented here (stemming from Rockafellar [45]) was rather circuitous. Not having the concept of subderivative at his disposal, he started with a special formula for subgradients of locally Lipschitzian functions (see Theorem 8.5) and used it to develop his notion of tangent cone by a dual method. Then he defined subgradients in general by a geometric version of formula (7.1) which corresponds to the epigraph relationship in Theorem 6.2.

In the locally Lipschitzian case there is a generalized mean value theorem which serves as a further characterization of the sets $\partial f(x)$; see Lebourg [37].

8. Relationship with Differentiability

A convex function f is differentiable at x if and only if $\partial f(x)$ consists of a single vector y, namely $y = \nabla f(x)$; cf. [43, Theorem 25.1]. What is the situation for nonconvex functions? Something more than ordinary differentiability is involved.

Recall that f is *differentiable* at x in the classical sense if and only if f is finite on a neighborhood of x and there exists a vector y such that

$$\lim_{\substack{h' \to h \\ t \downarrow 0}} \frac{f(x + th') - f(x)}{t} = y \cdot h \quad \text{for all } h. \tag{8.1}$$

Then y is called the *gradient* of f at x and denoted by $\nabla f(x)$. The concept of *strict* differentiability of f at x is less well known; it means that

$$\lim_{\substack{x' \to x \\ h' \to h \\ t \downarrow 0}} \frac{f(x' + th') - f(x')}{t} = y \cdot h \quad \text{for all } h. \tag{8.2}$$

This is a localization of continuous differentiability: f is smooth on U (an open set in R^n) if and only if f is strictly differentiable at every $x \in U$.

Theorem 8.1. *Let f be lower semicontinuous and let x be a point where f is finite. Then f is strictly differentiable at x if and only if $\partial f(x)$ consists of a single vector y, namely the gradient $\nabla f(x)$. In this event f must be Lipschitzian around x.*

Clarke [13] proved this fact under the assumption that f is locally Lipschitzian. The general case follows from Clarke's result and Theorem 7.6.

Subdifferentially regular functions have an especially strong property in this regard.

Theorem 8.2 [51]. *Suppose f is locally Lipschitzian and subdifferentially regular on the open set $U \subset R^n$, as is true in particular if f is subsmooth on U. Then f is strictly differentiable wherever it is differentiable. Thus except for a negligible set of points x in U, the convex set $\partial f(x)$ reduces to a single vector.*

The final assertion is based on Rademacher's theorem (Theorem 2.2). Finite convex functions are in particular subsmooth (Corollary 2.6), so Theorem 8.2 explains the fact mentioned at the beginning of this section.

Corollary 8.3. *Suppose f is subsmooth around x and has a local representation* (2.8) *such that there is only one $t \in T$ with φ_t equal to f at x. Then f is strictly differentiable at x with gradient $\nabla f(x) = \nabla\varphi_t(x)$.*

The case in the corollary is obtained from Theorem 7.3. Note that this answers the question raised in § 1 about when a function expressed as a maximum of a collection of smooth functions φ_t as in (2.8) can actually be smooth. It is smooth if the representation satisfies the assumptions in the definition of subsmoothness in § 2, and if the maximum is attained for each x by a unique t. The latter is true, for instance, if T is a convex set and $\varphi_t(x)$ is strictly concave in t for each x.

Strongly subsmooth functions, as defined in § 2, have an even nicer property. Recall that f is *twice-differentiable* at x if it is finite in a neighborhood of x and there exist $y \in R^n$ and $H \in R^{n \times n}$ such that

$$\lim_{\substack{h' \to h \\ t \downarrow 0}} \frac{f(x + th') - f(x) - y \cdot th'}{t} = \tfrac{1}{2} h \cdot Hh \quad \text{for all } h. \tag{8.3}$$

(This form of the definition does not require f to be once differentiable on a neighborhood of x.) A classical theorem of Alexandroff says that finite convex functions are twice differentiable almost everywhere. This translates by way of Theorem 2.5 into the following.

Theorem 8.4 [51]. *A strongly subsmooth function f on an open set $U \subset R^n$ is twice differentiable except on a negligible subset of U.*

Since subsmooth functions are quite common in optimization the preceding results can be applied in many situations.

Locally Lipschitzian functions which are not subdifferentially regular do not necessarily have $\partial f(x)$ consisting just of $\nabla f(x)$ at points where $\nabla f(x)$ exists, since f may be differentiable but not strictly differentiable at such points. An example exists of a locally Lipschitzian function f which is *nowhere* strictly differentiable [51], although f must be differentiable almost everywhere by Theorem 2.2. Nevertheless $\partial f(x)$ can by constructed from knowledge of $\nabla f(x')$ at points near to x, as the next theorem shows.

Theorem 8.5 [13], [14]. *Suppose f is Lipschitzian around x. Then*

(8.4) $\quad \partial f(x)=\mathrm{co}\{y \mid \exists x^k \to x$ *with f differentiable at x^k and* $\nabla f(x^k)\to y\}$.

This is the formula originally used by Clarke [13] for subgradients of locally Lipschitzian functions.

A generalization of Theorem 8.5 to arbitrary lower semicontinuous functions f has been furnished by Rockafellar [52], [49]. In such a setting one must consider not sequences of gradients $\nabla f(x^k)$, but "lower semigradients" or "proximal subgradients", and the notion of convex hull must be broadened to include "points at infinity".

9. Subdifferential Calculus

Not much could be accomplished with generalized subgradients if there were no rules for calculating or estimating them, beyond the definition itself. Such rules do exist in the convex case [43], and many of them have now been extended. We can only mention a few here.

A formula for the subgradients of the pointwise maximum of a collection of smooth functions φ_t, $t\in T$, has already been given in Theorem 7.3 and Corollary 7.4. This has been generalized to certain collections of nonsmooth functions φ_t by Clarke [13], [14].

An especially important operation to consider is that of addition of functions.

Theorem 9.1 [46]. *Let f_1 and f_2 be lower semicontinuous functions on R^n, and let x be a point where both f_1 and f_2 are finite. Suppose there exists an $h\in R^n$ such that $f_1^\uparrow(x;h)<\infty$ and f_2 is directionally Lipschitzian at x with respect to h. Then*

$$\partial(f_1+f_2)(x)\subset\partial f_1(x)+\partial f_2(x). \tag{9.1}$$

Moreover equality holds in (9.1) *if f_1 and f_2 are subdifferentially regular at x.*

Corollary 9.2. *Suppose f_1 is finite at x and lower semicontinuous around x, and f_2 is Lipschitzian around x. Then the conclusions of the theorem are valid.*

The corollary is the case of the theorem where $h=0$. A number of applications of Theorem 9.2 have an indicator function in place of either f_1 or f_2. The following fact then comes into play.

Theorem 9.3 [13]. *Let f be the indicator ψ_C of a closed set $C\subset R^n$, and let $x\in C$. Then*

$$\partial f(x)=N_C(x),\ \textit{where}\ N_C(x)=T_C(x)^\circ\ \textit{(polar)}. \tag{9.2}$$

The polar $N_C(x)$ of the Clarke tangent cone $T_C(x)$ is the *Clarke normal cone* to C at x, and its elements are called *normal vectors*. A nonzero normal vector exists at x if and only if x is a boundary point of C (cf. Rockafellar [44]). For direct expressions of the normal cone as the convex hull of limits of more special kinds of normals at special points, see Clarke [13] and Treiman [61].

Theorem 9.1 is only a sample of the kind of calculus that can be carried out. The chain rule too has its generalizations; see Clarke [14], Rockafellar [44], [54]. Further rules are listed in [49].

As far as mathematical programming is concerned, the question of how to estimate the subgradients of an optimal value function p, as described toward the end of § 1, is highly significant, and much effort has been expended on it (see [48] and its references). We must content ourselves here with indicating what the answer is in a special case.

Let us consider the problem

$$(P_u)\qquad \begin{array}{l}\text{minimize } f_0(x) \text{ subject to}\\ \qquad f_i(x)+u_i\le 0 \text{ for } i=1,\dots,m,\end{array}$$

where $u=(u_1,\dots,u_m)$ and the functions $f_0, f_1,\dots,f_m$ are locally Lipschitzian on R^n. Suppose that for every $u\in R^n$ and $\alpha\in R$ the set of feasible solutions to (P_u) with $f_0(x)\le\alpha$ is bounded (maybe empty). Let $p(u)$ denote the optimal value in (P_u). Then p is a lower semicontinuous, proper function on R^m (convex actually if every f_i is convex, but that is not the situation we want to restrict ourselves to at the moment). Since p is nondecreasing in u, it is also directionally Lipschitzian throughout its effective domain (cf. end of § 2).

Theorem 9.4. *Let $p(u)$ be the optimal value in problem (P_u) under the above assumptions. Fix any u such that $p(u)$ is finite, and let $X(u)$ denote the corresponding set of optimal solutions. For each $x\in X(u)$, let $K(u,x)$ denote the set of all Lagrange multiplier vectors $y\in R^m$ satisfying the generalized Kuhn-Tucker conditions*

$$\begin{array}{l} y_i\ge 0 \text{ and } y_i f_i(x)=0 \text{ for } i=1,\dots,m,\\ 0\in\partial f_0(x)+y_1\partial f_1(x)+\dots+y_m\partial f_m(x).\end{array} \tag{9.3}$$

Similarly let $K_0(u,x)$ be the set of vectors y which would satisfy (9.3) *if the term $\partial f_0(x)$ were omitted. If $K_0(u,x)$ consists of just the zero vector for every $x\in X(u)$, then p is Lipschitzian around u and*

$$\partial p(u)\subset \mathrm{co}\Big[\bigcup_{x\in X(u)} K(x,y)\Big]\quad \textit{(compact)}, \tag{9.4}$$

$$p^\uparrow(u;h)=p^\circ(u;h)\le \max_{x\in X(u)}\ \max_{y\in K(u,x)} y\cdot h \quad \textit{for all } h. \tag{9.5}$$

For a stability condition ensuring that $\partial p(u)=K(x,u)$ in (9.4) see Pomerol [42]. A more abstract analysis of $\partial p(u)$ in the case of functions p of the form

$$p(u)=\inf_x f(u,x),$$

with f lower semicontinuous on $R^m\times R^n$ and extended-real-valued, is carried out in [50].

References

1. J. P. Aubin, "Gradients généralisés de Clarke", *Annales des Sciences Mathématiques Québec* 2 (1978) 197–252.
2. J. P. Aubin, "Further properties of Lagrange multipliers in nonsmooth optimization", *Applied Mathematics and Optimization* 6 (1980) 79–80.
3. J. P. Aubin, "Locally Lipschitzian cooperative games", *Journal of Mathematical Economics* 8 (1981) 241–262.
4. J. P. Aubin and F. H. Clarke, "Multiplicateurs de Lagrange en optimisation non convexe et applications", *Comptes Rendus de l'Academie des Sciences Paris* 285 (1977) 451–454.
5. J. P. Aubin and F. H. Clarke, "Shadow prices and duality for a class of optimal control problems, *SIAM Journal on Control and Optimization* 17 (1979) 567–586.
6. A. Auslender, "Stabilité différentiable en programmation non convexe nondifférentiable", *Comptes Rendus de l'Academie des Sciences Paris* 286.
7. A. Auslender, "Differential stability in nonconvex and nondifferentiable programming", *Mathematical Programming Study* 10 (1978).
8. A. Auslender, "Minimisation de fonctions localement lipschitziennes: Applications a la programmation mi-convex, mi-differentiable", in: O. L. Mangasarian et al., ed., Nonlinear Programming 3 (Academic Press, New York, 1978) 429–460.
9. R. W. Chaney, "Second-order sufficiency conditions for nondifferentiable programming problems", *SIAM Journal on Control and Optimization* 20 (1982) 20–33.
10. R. W. Chaney and A. A. Goldstein, " An extension of the method of subgradients", in C. Lemarechal and R. Mifflin, eds., Nonsmooth Optimization (Pergamon Press, New York, 1978).
11. S. M. Chung, "Exact penalty algorithms for nonlinear programming", in: O. L. Mangasarian et al., ed., Nonlinear Programming 3 (Academic Press, New York, 1978) 197–224.
12. A. R. Conn, "Constrained optimization using a nondifferentiable penalty function", *SIAM Journal on Numerical Analysis* 10 (1973) 760–784.
13. F. H. Clarke, "Generalized gradients and applications", *Transactions of the American Mathematical Society* 205 (1975) 247–262.
14. F. H. Clarke, "Generalized gradients of Lipschitz functionals", *Advances in Mathematics* 40 (1981) 52–67.
15. F. H. Clarke, "On the inverse function theorem", *Pacific Journal of Mathematics* 67 (1976) 97–102.
16. F. H. Clarke, "The Euler-Lagrange differential inclusion", *Journal of Differential Equations* 19 (1975) 80–90.
17. F. H. Clarke, "La condition hamiltonienne d'optimalité", *Comptes Rendus de l'Academie des Sciences Paris* 280 (1975) 1205–1207.
18. F. H. Clarke, "Optimal solutions to differential inclusions", *Journal of Optimization Theory and Applications* 19 (1976) 469–478.
19. F. H. Clarke, "The generalized problem of Bolza", *SIAM Journal on Control and Optimization* 14 (1976) 682–699.
20. F. H. Clarke, "The maximum principle under minimal hypotheses", *SIAM Journal on Control and Optimization* 14 (1976) 1078–1091.
21. F. H. Clarke, "Inequality constraints in the calculus of variations", *Canadian Journal of Mathematics* 39 (1977) 528–540.
22. F. H. Clarke, "A new approach to Lagrange multipliers", *Mathematics of Operations Research* 1 (1976) 165–174.
23. B. Cornet, "Regularity properties of normal and tangent cones", forthcoming.

24. J. M. Danskin, The theory of max-min and its application to weapons allocations problems (Springer-Verlag, New York, 1967).
25. R. J. Duffin, E. L. Peterson and C. Zener, Geometric programming (Wiley, New York, 1967).
26. J. Gauvin, "The generalized gradient of a marginal function in mathematical programming", *Mathematics of Operations Research* 4 (1979) 458–463.
27. J. Gauvin and F. Dubeau, "Differential properties of the marginal value function in mathematical programming", *Mathematical Programming Studies*, forthcoming.
28. A. A. Goldstein, "Optimization of Lipschitz continuous functions", *Mathematical Programming* 13 (1977) 14–22.
29. M. Held, P. Wolfe and H. P. Crowder, "Validation of subgradient optimization", *Mathematical Programming* 6 (1974) 62–88.
30. J.-B. Hiriart-Urruty, "Tangent cones, generalized gradients and mathematical programming in Banach spaces", *Mathematics of Operations Research* 4 (1979) 79–97.
31. J.-B. Hiriart-Urruty, "On optimality conditions in nondifferentiable programming", *Mathematical Programming*, forthcoming.
32. J.-B. Hiriart-Urruty, "Gradients generalisés de fonctions marginales", *SIAM Journal on Control and Optimization* 16 (1978) 301–316.
33. J.-B. Hiriart-Urruty, "Conditions nécessaires d'optimalité pour un programme stochastique avec recours", *SIAM Journal on Control and Optimization* 16 (1978) 317–329.
34. J.-B. Hiriart-Urruty, "Refinements of necessary optimality conditions in nondifferentiable programming, I", *Applied Mathematics and Optimization* 5 (1979) 63–82.
35. J.-B. Hiriart-Urruty, "Extension of Lipschitz functions", *Journal of Mathematical Analysis and Applications* 77 (1980) 539–554.
36. S. Howe, "New conditions for the exactness of a simple penalty function", *SIAM Journal on Control and Optimization* 11 (1973) 378–381.
37. G. Lebourg, "Valeur moyenne pour gradient généralisé", *Comptes Rendus de l'Academie des Sciences Paris* 281 (1975) 795–797.
38. R. Mifflin, "Semismooth and semiconvex functions in constrained optimization", *SIAM Journal on Control and Optimization* 15 (1977) 959–972.
39. R. Mifflin, "An algorithm for constrained optimization with semismooth functions", *Mathematics of Operations Research* 2 (1977) 191–207.
40. R. Mifflin, V. H. Nguyen and J. J. Strodiot, "On conditions to have bounded multipliers in locally Lipschitz programming", *Mathematical Programming* 18 (1980) 100–106.
41. T. Petrzykowski, "An exact potential method for constrained maxima", *SIAM Journal on Numerical Analysis* 6 (1969) 299–304.
42. C. Pomerol, "The Lagrange multiplier set and the generalized gradient set of the marginal function of a differentiable program in a Banach space", *Journal of Optimization Theory and Applications*, forthcoming.
43. R. T. Rockafellar, Convex analysis (Princeton University Press, Princeton, NJ, 1970).
44. R. T. Rockafellar, "Clarke's tangent cones and the boundaries of closed sets in R^n", *Nonlinear Analysis: Theory, Methods and Applications* 3 (1979) 145–154.
45. R. T. Rockafellar, "Generalized directional derivatives and subgradients of nonconvex functions", *Canadian Journal of Mathematics* 32 (1980) 257–280.
46. R. T. Rockafellar, "Directionally Lipschitzian functions and subdifferential calculus", *Proceedings of the London Mathematical Society* 39 (1979) 331–355.
47. R. T. Rockafellar, "Proximal subgradients, marginal values and augmented Lagrangians in nonconvex optimization", *Mathematics of Operations Research* 6 (1981) 424–436.

48. R. T. Rockafellar, "Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming", *Mathematical Programming Study* 17 (1982) 28–66.
49. R. T. Rockafellar, "The theory of subgradients and its applications to problems of optimization: convex and nonconvex functions" (Heldermann-Verlag, West-Berlin, 1981).
50. R. T. Rockafellar, "Augmented Lagrangians and marginal values in parameterized optimization problems", in: A. Wierzbicki, ed., Generalized Lagrangian methods in optimization (Pergamon Press, New York, 1982).
51. R. T. Rockafellar, "Favorable classes of Lipschitz continuous functions in subgradient optimization", in: E. Nurminski, ed., Nondifferentiable Optimization (Pergamon Press, New York, 1982).
52. R. T. Rockafellar, "Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization", *Mathematics of Operations Research* 6 (1981) 427–437.
53. R. T. Rockafellar, "Marginal values and second-order necessary conditions for optimality", *Mathematical Programming*, to appear.
54. R. T. Rockafellar, "Directional differentiability of the optimal value function in a nonlinear programming problem", *Mathematical Programming Studies*, to appear.
55. R. T. Rockafellar, "Augmented Lagrange multiplier functions and duality in nonconvex programming", *SIAM Journal on Control* 12 (1974) 268–285.
56. S. Saks, "Theory of the integral" (Hafner Publishing Co., New York, 1937).
57. L. Thibault, "Problème de Bolza dans un espace de Banach séparable", *Compes de l'Academie des Sciences Paris* 282 (1976) 1303–1306.
58. L. Thibault, "Subdifferentials of compactly Lipschitzian vector-valued functions", *Annali di Matematica Pura e Applicata* 125 (1980) 157–192.
59. L. Thibault, "On generalized differentials and subdifferentials of Lipschitz vector-valued functions", *Nonlinear Analysis: Theory, Methods and Applications*, forthcoming.
60. L. Thibault, "Subdifferentials of nonconvex vector-valued functions", *Journal of Mathematical Analysis and Applications*, forthcoming.
61. J. Treiman, "Characterization of Clarke's tangent and normal cones in finite and infinite dimensions", *Nonlinear Analysis: Theory, Methods and Applications*, forthcoming.
62. W. I. Zangwill, "Nonlinear programming via penalty functions", *Management Science* 13 (1961) 34–358.