106

LIPSCHITZIAN STABILITY IN OPTIMIZATION: THE ROLE OF NONSMOOTH ANALYSIS

R.T. Rockafellar
Department of Mathematics, University of Washington, Seattle, WA 98195, USA

ABSTRACT

The motivations of nonsmooth analysis are discussed. Applications are given to the sensitivity of optimal values, the interpretation of Lagrange multipliers, and the stability of constraint systems under perturbation.

INTRODUCTION

It has been recognized for some time that the tools of classical analysis are not adequate for a satisfactory treatment of problems of optimization. These tools work for the characterization of locally optimal solutions to problems where a smooth (i.e. continuously differentiable) function is minimized or maximized subject to finitely many smooth equality constraints. They also serve in the study of perturbations of such constraints, namely through the implicit function theorem and its consequences. As soon as inequality constraints are encountered, however, they begin to fail. One-sided derivative conditions start to replace two-sided conditions. Tangent cones replace tangent subspaces. Convexity and convexification emerge as more natural than linearity and linearization.

In problems where inequality constraints actually predominate over equations, as is typical in most modern applications of optimization, a qualitative change occurs. No longer is there any simple way of recognizing which constraints are active in a neighborhood of a given point of the feasible set, such as there would be if the set were a cube or simplex, say. The boundary of the feasible set defies easy description and may best be thought of as a nonsmooth hypersurface. It does not take long to realize too that the graphs of many of the objective functions which naturally arise are nonsmooth in a similar way. This is the motivation for much of the effort that has gone into

* Research supported in part by a grant from the National Science Foundation at the University of Washington, Seattle.

Published in: Nondifferential Optimization: Motivations and Applications, (V.F. Demyanov and D. Pallaschke, eds.), Lect. Notes in Econ. and Math. Systems 255, Springer-Verlog, Berlin, pp. 55-73.

introducing and developing various concepts of "tangent cone", "normal cone", "directional derivative" and "generalized gradient". These concepts have changed the face of optimization theory and given birth to a new subject, ***nonsmooth analysis***, which is affecting other areas of mathematics as well.

An important aim of nonsmooth analysis is the formulation of generalized necessary or sufficient conditions for optimality. This in turn receives impetus from research in numerical methods of optimization that involve nonsmooth functions generated by decomposition, exact penalty representations, and the like. The idea essentially is to provide tests that either establish (near) optimality (perhaps stationarity) of the point already attained or generate a feasible direction of improvement for moving to a better point.

Nonsmooth analysis also has other important aims, however, which should not be overlooked. These include the study of sensitivity and stability with respect to perturbations of objective and constraints. In an optimization problem that depends on a parameter vector v, how do variations in v affect the optimal value, the optimal solution set, and the feasible solution set? Can anything be said about rates of change?

This is where Lipschitzian properties take on special significance. They are intermediate between continuity and differentiability and correspond to *bounds* on possible rates of change, rather than rates themselves, which may not exist, at least in the classical sense. Like convexity properties they can be passed along through various constructions where true differentiability, even if one-sided, would be lost. Furthermore, they can be formulated in geometric terms that suit the study multifunctions (set-valued mappings), a subject of great importance in optimization theory but for which classical notions are almost entirely lacking.

It is in this light that the directional derivatives and subgradients introduced by F.H. Clarke [1] [2] should be judged. Clarke's theory emphasizes Lipschitzian properties and sturdily combines convex analysis and classical smooth analysis in a single framework. At the present stage of development, thanks to the efforts of many individuals, it has already had strong effects on almost every area of optimization, from nonlinear programming to the calculus of variations, and also on mathematical questions beyond the domain of optimization per se.

This is not to say, however, that Clarke's derivatives and subgradients are the only ones that henceforth need to be considered. Special situations certainly do require special insights. In particular, there are cases where special one-sided first and second derivatives that are more finely tuned than Clarke's are worth introducing. Significant and useful results can be obtained in such manner. But such results are likely to be relatively limited in scope.

The power and generality of the kind of nonsmooth analysis that is based on Clarke's ideas can be credited to the following features, in summary:

(a) Applicability to a huge class of functions and other objects, such as sets and multifunctions.

(b) Emphasis on geometric constructions and interpretations.

(c) Reduction to classical analysis in the presence of smoothness and to convex analysis in the presence of convexity.

(d) Unified formulation of optimality conditions for a wide variety of problems.

(e) Comprehensive calculus of subgradients and normal vectors which makes possible an effective specialization to particular cases.

(f) Coverage of sensitivity and stability questions and their relationship to Lagrange multipliers.

(g) Focus on local properties of a "uniform" character, which are less likely to be upset by slight perturbations, for instance in the study of directions of descent.

(h) Versatility in infinite as well as finite-dimensional spaces and in treating the integral functionals and differential inclusions that arise in optimal control, stochastic programming, and elsewhere.

In this paper we aim at putting this theory in a natural perspective, first by discussing its foundations in analysis and geometry and the way that Lipschitzian properties come to occupy the stage. Then we survey the results that have been obtained recently on sensitivity and stability. Such results are not yet familiar to many researchers who concentrate on optimality conditions and their use in algorithms. Nevertheless they say much that bears on numerical matters, and they demonstrate well the sort of challenge that nonsmooth analysis is now able to meet.

1. ORIGINS OF SUBGRADIENT IDEAS

In order to gain a foothold on this new territory, it is best to begin by thinking about functions $f: R^n \to R$ that are not necessarily smooth but have strong one-sided directional derivatives in the sense of

$$f'(x;h) = \lim_{\substack{t \downarrow 0 \\ h' \to h}} \frac{f(x+th') - f(x)}{t} \tag{1.1}$$

Examples are (finite) convex functions [3] and *subsmooth* functions, the latter being by definition representable locally as

$$f(x) = \max_{s \in S} f_s(x), \tag{1.2}$$

where S is a compact space (e.g., a finite, discrete index set) and $\{f_s \mid s \in S\}$ is a family of smooth functions whose values and derivatives depend continuously on s and x jointly. Subsmooth functions were introduced in [4]; all smooth functions and all finite convex functions on R^n are in particular subsmooth.

The formula given here for $f'(x;h)$ differs from the more common one in the literature, where the limit $h' \to h$ is omitted (weak one-sided directional derivative). It corresponds in spirit to true (strong) differentiability rather than weak differentiability. Indeed, under the assumption that $f'(x,h)$ exists for all h (as in (1.1)), one has f differentiable at x if and only if $f'(x;h)$ is linear in h. Then the one-sided limit $t \downarrow 0$ is actually realizable as a two-sided limit $t \to 0$.

The classical concept of *gradient* arises from the duality between linear functions on R^n and vectors in R^n. To say that $f'(x;h)$ is linear in h is to say that there is a vector $y \in R^n$ with

$$f'(x;h) = y \cdot h \quad \text{for all} \quad h. \tag{1.3}$$

This y is called the gradient of f at x and is denoted by $\nabla f(x)$.

In a similar way the modern concept of *subgradient* arises from the duality between sublinear functions on R^n and convex subsets in R^n. A function l is said to be *sublinear* if it satisfies

$$l(\lambda_1 h_1 + \ldots + \lambda_m h_m) \leq \lambda_1 l(h_1) + \ldots + \lambda_m l(h_m) \tag{1.4}$$

$$\text{when} \quad \lambda_1 \geq 0, \cdots, \lambda_m \geq 0.$$

It is known from convex analysis [3, §13] that the finite sublinear functions l on R^n are precisely the support functions of the nonempty compact subsets Y of R^n: each l corresponds to a unique Y by the formula

$$l(h) = \max_{y \in Y} y \cdot h \quad \text{for all} \quad h. \tag{1.5}$$

Linearity can be identified with the case where Y consists of just a single vector y.

It turns out that when f is convex, and more generally when f is subsmooth [4], the derivative $f'(x,h)$ is always sublinear in h. Hence there is a nonempty compact subset Y of R^n, uniquely determined, such that

$$f'(x;h) = \max_{y \in Y} y \cdot h \quad \text{for all} \quad h. \tag{1.6}$$

This set Y is denoted by $\partial f(x)$, and its elements y are called subgradients of f at x. With respect to any local representation (1.4), one has

$$Y = \mathrm{co}\{\nabla f_s(x) \mid s \in S_x\}, \text{ where } S_x = \operatorname*{argmax}_{s \in S} f_s(x) \tag{1.7}$$

(co = convex hull), but the set $Y = \partial f(x)$ is of course by its definition independent of the representation used.

In the case of f convex [3, §23] one can define subgradients at x equivalently as the vectors y such that

$$f(x') \geq f(x) + y \cdot (x' - x) \text{ for all } x'. \tag{1.8}$$

For f subsmooth this generalizes to

$$f(x') \geq f(x) + y \cdot (x' - x) + o(|x' - x|), \tag{1.9}$$

but caution must be exercised here about further generalization to functions f that are not subsmooth. Although the vectors y satisfying (1.9) do always form a closed convex set Y at x, regardless of the nature of f, this set Y does not yield an extension of formula (1.6), nor does it correspond in general to a robust concept of directional derivative that can be used as a substitute for $f'(x;h)$ in (1.6). For a number of years, this is where subgradient theory came to a halt.

A way around the impasse was discovered by Clarke in his thesis in 1973. Clarke took up the study of functions $f : R^n \to R$ that are *locally Lipschitzian* in the sense of the difference quotient

$$|f(x'') - f(x')| \,/\, |x'' - x'| \tag{1.10}$$

being bounded on some neighborhood of each point x. This class of functions is of intrinsic value for several reasons. First, it includes all subsmooth functions and consequently all smooth functions and all finite convex functions; it also includes all finite concave functions and all finite saddle functions (which are convex in one vector argument and concave in another; see [3, §35]). Second, it is preserved under taking linear combinations, pointwise maxima and minima of collections of functions (with certain mild assumptions), integration and other operations of obvious importance in optimization. Third, it exhibits properties that are closely related to differentiability. The local boundedness of the difference quotient (1.10) is such a property itself. In fact when f is locally Lipschitzian, the gradient $\nabla f(x)$ exists for all but a negligible set of points x in R^n (the classical theorem of Rademacher, see [5]).

Clarke discovered that when f is locally Lipschitzian, the special derivative expression

$$f^{\circ}(x;h) = \limsup_{\substack{t \downarrow 0 \\ h' \to h \\ x' \to x}} \frac{f(x'+th') - f(x')}{t} \tag{1.11}$$

is always a finite sublinear function of h. Hence there exists a unique nonempty compact convex set Y such that

$$f^{\circ}(x;h) = \max_{y \in Y} y \cdot h \quad \text{for all } h. \tag{1.12}$$

Moreover

$$f^{\circ}(x;h) = f'(x;h) \quad \text{for all } h \text{ when } f \text{ is subsmooth.} \tag{1.13}$$

Thus in denoting this set Y by $\partial f(x)$ and calling its elements subgradients, one arrives at a natural extension of nonsmooth analysis to the class of all locally Lipschitzian functions. Many powerful formulas and rules have been established for calculating or estimating $\partial f(x)$ in this broad context, but it is not our aim to go into them here; see [2] and [6], for instance.

It should be mentioned that Clarke himself did not incorporate the limit $h' \to h$ into the definition of $f^{\circ}(x;h)$, but because of the Lipschitzian property the value obtained for $f^{\circ}(x;h)$ is the same either way. By writing the formula with $h' \to h$ one is able to see more clearly the relationship between $f^{\circ}(x;h)$ and $f'(x;h)$ and also to prepare the ground for further extensions to functions f that are merely lower semicontinuous rather than Lipschitzian. (For such functions one writes $x' \to_f x$ in place of $x' \to x$ to indicate that x is to be approached by x' only in such a way that $f(x') \to f(x)$. More will be said about this later.)

Some people, having gone along with the developments up until this point, begin to balk at the "coarse" nature of the Clarke derivative $f^{\circ}(x;h)$ in certain cases where f is *not* subsmooth and nevertheless is being *minimized*. For example, if $f(x) = -|x| + |x|^2$ one has $f^{\circ}(0;h) = |h|$, whereas $f'(0;h)$ exists too but $f'(0;h) = -|h|$. Thus f' reveals that every $h \neq 0$ gives a direction of descent from 0, in the sense of yielding $f'(0;h) < 0$, but f° reveals no such thing, inasmuch as $f^{\circ}(0;h) > 0$. Because of this it is feared that f° does not embody as much information as f' and therefore may not be entirely suitable for the statement of necessary conditions for a minimum, let alone for employment in algorithms of descent.

Clearly f° cannot replace f' in every situation where the two may differ, nor has this ever been suggested. But even in face of this caveat there are arguments to be made in favor of f° that may help to illuminate its nature and the supporting motivation. The Clarke derivative f° is oriented towards minimization problems, in contrast to f', which is neutral between minimization and maximization. In addition, it emphasizes a certain uniformity. A vector h with $f^\circ(x;h) < 0$ provides a descent direction in a strong *stable* sense: there is an $\varepsilon > 0$ such that for all x' near x, h' near h, and positive t near 0, one has

$$f(x' + th') < f(x') - t\varepsilon.$$

A vector h with $f'(x;h) < 0$, on the other hand, provides descent only from x; at points x' arbitrarily near to x it may give a direction of ascent instead. This instability is not without numerical consequences, since x might be replaced by x' due to round-off.

An algorithm that relied on finding an h with $f'(x;h) < 0$ in cases where $f^\circ(x;h) \geq 0$ for all h (such an x is said to be *substationary* point) seems unlikely to be very robust. Anyway, it must be realized that in executing a method of descent there is very little chance of actually arriving along the way at a point x that is substationary but not a local minimizer. One is easily convinced from examples that such a mishap can only be the consequence of an unfortunate choice of the starting point and disappears under the slightest perturbation. The situation resembles that of cycling in the simplex method.

Furthermore it must be understood that because of the orientation of the definition of f° towards minimization, there is no justice in holding the notion of substationarity up to any interpretation other than the following: a substationary point is either a point where a local *minimum* is attained or one where progress towards a local minimum is "confused". Sometimes, for instance, one hears cited as a failing of f° that f' is able to distinguish between a local minimum and a local maximum in having $f'(x;h) \geq 0$ for all h in the first case, but $f'(x;h) \leq 0$ for all h in the second, whereas $f^\circ(x;h) \geq 0$ for all h in both cases. But this is unfair. A one-sided orientation in nonsmooth analysis is merely a reflection of the fact that in virtually all applications of optimization, there is unambiguous interest in either maximization or minimization, but not both. For theoretical purposes it might as well be minimization.

Certainly the idea that a first-order concept of derivative, such as we are dealing with here, is obliged to provide conditions that distinguish effectively between a local minimum and a local maximum is out of line for other reasons. Classical analysis makes no attempt in that direction, without second derivatives. Presumably, second derivative concepts in nonsmooth analysis will eventually furnish the appropriate distinctions, cf. Chaney [7].

A final note on the question of f° versus f' is the reminder that $f^\circ(x;h)$ is defined for any locally Lipschitzian function f and even more generally, whereas $f'(x;h)$ is only defined for functions f in a narrower class.

An important goal of nonsmooth analysis is not only to make full use of Lipschitz continuity when it is present, but also to provide criteria for Lipschitz continuity in cases where it cannot be known *a priori*, along with corresponding estimates for the local Lipschitz constant. For this purpose, it is necessary to extend subgradient theory to functions that might not be locally Lipschitzian or even continuous everywhere, but merely lower semicontinuous. Fundamental examples of such functions in optimization are the so-called *marginal* functions, which give the minimum value in a parameterized problem as a function of the parameters. Such functions can even take on $\pm\infty$.

Experience with convex analysis and its applications shows further the desirability of being able to treat the indicator functions of sets, which play an essential role in the passage between analysis and geometry.

In fact, the ideas that have been described so far can be extended in a powerful, consistent manner to the class of all lower semicontinuous functions $f: R^n \to \bar{R}$, where $\bar{R} = [-\infty, \infty]$ (extended real number system). There are two complementary ways of doing this, with the same result. In the continuation of the analytic approach we have been following until now, a more subtle directional derivative formula

$$f^\uparrow(x;h) = \lim_{\varepsilon \downarrow 0} \left[\limsup_{\substack{t \downarrow 0 \\ x' \to_f x}} \left[\inf_{|h'-h| \le \varepsilon} \frac{f(x'+th')-f(x')}{t} \right] \right] \tag{1.14}$$

is introduced and shown to agree with $f^\circ(x;h)$ whenever f is locally Lipschitzian and indeed whenever $f^\circ(x;h)$ (in the extended definition with $x' \to_f x$, as mentioned earlier) is not $+\infty$. Moreover $f^\uparrow(x;h)$ is proved always to be a lower semicontinuous, sublinear function of h (extended-real-valued). From convex analysis, then, it follows that either $f^\uparrow(x;0) = -\infty$ or there is a nonempty closed convex set $Y \subset R^n$, uniquely determined, with

$$f^\uparrow(x;h) = \sup_{y \in Y} y \cdot h \quad \text{for all } h. \tag{1.15}$$

This is the approach followed in Rockafellar [8], [9]. One then arrives at the corresponding geometric concepts by taking f to be the indicator δ_C of a closed set C. For any $x \in C$, the function $h \mapsto \delta_C^\uparrow(x;h)$ is itself the indicator of a certain closed set

$T_C(x)$ which happens always to be a convex cone; this is the Clarke *tangent* cone to C at x. The subgradient set

$$N_C(x) = \partial\delta_C(x), \tag{1.16}$$

on the other hand, is a closed convex set too, the Clarke *normal* cone to C to x. The two cones are polar to each other:

$$N_C(x) = T_C(x)^{\circ}, \quad T_C(x) = N_C(x)^{\circ}. \tag{1.17}$$

In a more geometric approach to the desired extension, the tangent cone $T_C(x)$ and normal cone $N_C(x)$ can first be defined in a direct manner that accords with the polarity relations (1.16). Then for an arbitrary lower semicontinuous function $f: R^n \to \bar{R}$ and point x at which f is finite, one can focus on $T_E(x, f(x))$ and $N_E(x, f(x))$, where E is the epigraph of f (a closed subset of R^{n+1}). The cone $T_E(x, f(x))$ is itself the epigraph of a certain function, namely the subderivative $h \mapsto f^{\uparrow}(x;h)$, whereas the cone $N_E(x, f(x))$ provides the subgradients:

$$\partial f(x) = \{y \in R^n \mid (y, -1) \in N_E(x, f(x))\}. \tag{1.18}$$

The polarity between $T_E(x, f(x))$ and $N_E(x, f(x))$ yields the subderivative-subgradient relation (1.14). (Clarke's original extension of ∂f to lower semicontinuous functions [1] followed this geometric approach in defining normal cones directly and then invoking (1.17) as a definition for subgradients. He did not focus much on tangent cones, however, or pursue the idea that $T_E(x, f(x))$ might correspond to a related concept of directional derivative.)

The details of these equivalent forms of extension need not occupy us here. The main thing to understand is that they yield a basic criterion for Lipschitzian continuity, as follows.

THEOREM 1 (Rockafellar [10]). *For a lower semicontinuous function $f: R^n \to \bar{R}$ actually to be Lipschitzian on some neighborhood of the point x, it is sufficient (as well as necessary) that the subgradient set $\partial f(x)$ be nonempty and bounded. Then one has*

$$\limsup_{\substack{x' \to x \\ x'' \to x}} \frac{|f(x'') - f(x')|}{|x'' - x'|} = \max_{y \in \partial f(x)} |y|. \tag{1.19}$$

This criterion can be applied without exact knowledge of $\partial f(x)$ but only an estimate that $\phi \neq \partial f(x) \subset Y$ for some set Y. If Y is bounded, one may conclude that f is locally Lipschitzian around x. If it is known that $|y| < \lambda$ for all $y \in Y$, one has from (1.19)

$$|f(x'') - f(x')| \leq \lambda |x'' - x'| \text{ for } x' \text{ and } x'' \text{ near } x.$$

2. LAGRANGE MULTIPLIERS AND SENSITIVITY

Many ways have been found for deriving optimality conditions for problems with constraints, but not all of them provide full information about the Lagrange multipliers that are obtained. The test of a good method is that it should lead to some sort of interpretation of the multiplier vectors in terms of sensitivity or generalized rates of change of the optimal value in the problem with respect to perturbations. Until quite recently, a satisfactory interpretation along such lines was available only for convex programming and special cases of smooth nonlinear programming. Now, however, there are general results that apply to all kinds of problems, at least in R^n. These results demonstrate well the power of the new nonsmooth analysis and are not matched by anything achieved by other techniques.

Let us first consider a nonlinear programming problem in its canonical parameterization:

$$(P_u) \qquad \text{minimize } g(x) \text{ subject to } x \in K \text{ and}$$
$$g_i(x) + u_i \leq 0 \text{ for } i = 1, \ldots, s,$$
$$= 0 \text{ for } i = s+1, \ldots, m,$$

where $g, g_1, \ldots, g_m$ are locally Lipschitzian functions on R^n and K is a closed subset of R^n; the u_i's are parameters and form a vector $u \in R^m$. By analogy with what is known in particular cases of (P_u), one can formulate the potential *optimality condition* on a feasible solution x, namely that

$$0 \in \partial g(x) + \textstyle\sum_{i=1}^m y_i \partial g_i(x) + N_K(x) \text{ with} \tag{2.1}$$

$$y_i \geq 0 \text{ and } y_i[g_i(x) + u_i] = 0 \text{ for } i = 1, \ldots, s,$$

and a corresponding *constraint qualification* at x:

the only vector $y=(y_1, \ldots, y_m)$ satisfying the version (2.2)

of (2.1) in which the term $\partial g(x)$ is omitted is $y=0$.

In *smooth programming*, where the functions $g, g_1, \ldots, g_m$ are all continuously differentiable and there is no abstract constraint $x \in K$, the first relation in (2.1) reduces to the gradient equation

$$0 = \nabla g(x) + \sum_{i=1}^{m} y_i \nabla g_i(x),$$

and one gets the classical Kuhn-Tucker conditions. The constraint qualification is then equivalent (by duality) to the well known one of Mangasarian and Fromovitz.

In *convex programming*, where $g, g_1, \ldots, g_s$ are (finite) convex functions, $g_{s+1}, \ldots, g_m$ are affine, and K is a convex set, condition (2.1) is always sufficient for optimality. Under the constraint qualification (2.2), which in the absence of equality constraints reduces to the Slater condition, it is also necessary for optimality.

For the general case of (P_u) one has the following rule about necessity.

THEOREM 2 (Clarke [11]). *Suppose x is a locally optimal solution to (P_u) at which the constraint qualification (2.2) is satisfied. Then there is a multiplier vector y such that the optimality condition (2.1) is satisfied.*

This is not the sharpest result that may be stated, although it is perhaps the simplest. Clarke's paper [11] puts a potentially smaller set in place of $N_K(x)$ and provides along side of (2.2) a less stringent constraint qualification in terms of "calmness" of (P_u) with respect to perturbations of u. Hiriart-Urruty [12] and Rockafellar [13] contribute some alternative ways of writing the subgradient relations. For our purposes here, let it suffice to mention that Theorem 2 remains true when the optimality condition (2.1) is given in the slightly sharper and more elegant form:

$$0 \in \partial g(x) + y\,\partial G(x) + N_K(x) \text{ with } y \in N_C(G(x)+u), \tag{2.3}$$

where $G(x) = (g_1(x), \ldots, g_m(x))$ and

$$C = \{w \in R^m \mid w_i \leq 0 \text{ for } i=1,\ldots,s \text{ and } w_i=0 \text{ for } i=s+1,\ldots,m\}. \tag{2.4}$$

The notation $\partial G(x)$ refers to Clarke's generalized Jacobian [2] for the mapping G; one has

$$y\,\partial G(x) = \partial(\textstyle\sum_{i=1}^{m} y_i g_i)(x). \tag{2.5}$$

Theorem 2 has the shining virtue of combining the necessary conditions for smooth programming and the ones for convex programming into a single statement. Moreover it covers subsmooth programming and much more, and it allows for an abstract constraint in the form of $x \in K$ for an arbitrary closed set K. Formulas for calculating the normal cone $N_K(x)$ in particular cases can then be used to achieve additional specializations.

What Theorem 2 does *not* do is provide any interpretation for the multipliers y_i. In order to arrive at such an interpretation, it is necessary to look more closely at the properties of the marginal function

$$p(u) = \text{optimal value (infimum) in}(P_u). \tag{2.6}$$

This is an extended-real-valued function on R^m which is lower semicontinuous when the following mild *inf-boundedness condition* is fulfilled:

For each $\bar{u} \in R^m$, $\alpha \in R$ and $\varepsilon > 0$, the set of all $x \in K$ (2.7)

satisfying $g(x) \le \alpha$, $g_i(x) \le \bar{u}_i + \varepsilon$ for $i = 1,\dots,s$, and

$\bar{u}_i - \varepsilon \le g_i(x) \le \bar{u}_i + \varepsilon$ for $i = s+1,\dots,m$, is bounded in R^n.

This condition also implies that for each u with $p(u) < \infty$ (i.e. with the constraints of (P_u) consistent), the set of all (globally) optimal solutions to (P_u) is nonempty and compact.

In order to state the main general result, we let

$Y(u)$ = set of all multiplier vectors y that satisfy (2.1) (2.8)

for some optimal solution x to (P_u).

THEOREM 3 (Rockafellar [13]). *Suppose the inf-boundedness condition* (2.7) *is satisfied. Let u be such that the constraints of (P_u) are consistent and every optimal solution x to (P_u) satisfies the constraint qualification* (2.2). *Then $\partial p(u)$ is a nonempty compact set with*

$$\partial p(u) \subset \text{co}\, Y(u) \text{ and } \text{ext}\, \partial p(u) \subset Y(u). \tag{2.9}$$

(where "ext" denotes extreme points). In particular p is locally Lipschitzian around u with

$$p^\circ(u;h) \leq \sup_{y\in Y(u)} y\cdot h \quad \textit{for all } h. \tag{2.10}$$

Indeed, any λ satisfying $|y| < \lambda$ for all $y \in Y(u)$ serves as a local Lipschitz constant:

$$|p(u'')-p(u')| \leq \lambda |u''-u'| \quad \textit{when } u' \textit{ and } u'' \textit{ are near } u. \tag{2.11}$$

For smooth programming, this result was first proved by Gauvin [14]. He demonstrated further that when (P_u) has a unique optimal solution x, for which there is a unique multiplier vector y, so that $Y(u) = \{y\}$, then actually p is differentiable at u with $\nabla p(u) = y$. For convex programming one knows (see [3]) that $\partial p(u) = Y(u)$ always (under our inf-boundedness assumption) and consequently

$$p'(u;h) = \max_{y\in Y(u)} y\cdot h. \tag{2.12}$$

Minimax formulas that give $p'(u;h)$ in certain cases of smooth programming where $Y(u)$ is not just a singleton can be for example found in Demyanov and Malozemov [15] and Rockafellar [16]. Aside from such special cases there are no formulas known for $p'(u;h)$. Nevertheless, Theorem 3 does provide an estimate, because $p'(u;h) \leq p^\circ(u;h)$ whenever $p'(u;h)$ exists. (It is interesting to note in this connection that because p is Lipschitzian around u by Theorem 3, it is actually differentiable almost everywhere around u by Rademacher's theorem.)

Theorem 3 has recently been broadened in [6] to include more general kinds of perturbations. Consider the parameterized problem

(Q_v) minimize $f(v,x)$ over all x satisfying
$F(v,x) \in C$ and $(v,x) \in D$,

where v is a parameter vector in R^d, the functions $f: R^d \times R^n \to R$ and $F: R^d \times R^n \to R^m$ are locally Lipschitzian, and the sets $C \subset R^m$ and $D \subset R^d \times R^n$ are closed. Here C could be the cone in (2.4), in which event the constraint $F(v,x) \in C$ would reduce to

$$f_i(v,x) \leq 0 \text{ for } i=1,\ldots,s,$$
$$= 0 \text{ for } i=s+1,\ldots,m,$$

but this choice of C is not required. The condition $(v,x) \in D$ may equivalently be written as $x \in \Gamma(v)$, where Γ is the closed multifunction whose graph is D. It represents therefore an abstract constraint that can vary with v. A fixed abstract constraint $x \in K$ corresponds to $\Gamma(v)=K$, $D=R^d \times K$.

In this more general setting the appropriate optimality condition for a feasible solution x to (Q_v) is

$$(z,0) \in \partial f(v,x) + y\,\partial F(v,x) + N_D(v,x) \tag{2.13}$$

for some y and z with $y \in N_C(F(v,x))$,

and the constraint qualification is

the only vector pair (y,z) satisfying the version of (2.13) in which the term $\partial f(v,x)$ is omitted is $(y,z)=(0,0)$. (2.14)

THEOREM 4 (Rockafellar [6, §8]). *Suppose that x is a locally optimal solution to (Q_v) at which the constraint qualification* (2.14) *is satisfied. Then there is a multiplier pair (y,z) such that the optimality condition* (2.13) *is satisfied.*

Theorem 4 reduces to the version of Theorem 2 having (2.3) in place of (2.1) when (Q_v) is taken to be of the form (P_u), namely when $f(v,x)\equiv g(x)$, $F(v,x)=G(x)+v$, $D=R^m \times K$ $(R^m=R^d)$, and C is the cone in (2.4).

For the corresponding version of Theorem 3 in terms of the marginal function

$$q(v) = \text{optimal value in } (Q_v), \tag{2.15}$$

we take inf-boundedness to mean:

For each $\bar{v} \in R^d$, $\alpha \in R$ and $\varepsilon>0$, the set of all x satisfying for some v with $|v-\bar{v}| \le \varepsilon$ the constraints $F(v,x) \in C$, $(v,x) \in D$, and having $f(v,x) \le \alpha$, is bounded in R^n. (2.16)

Again, this property ensures that q is lower semicontinuous, and that for every v for which the constraints of (Q_v) are consistent, the set of optimal solutions to (Q_v) is nonempty and compact. Let

$$Z(v) = \text{set of all vectors } z \text{ that satisfy the multiplier condition (2.13) for some optimal solution } x \text{ to } (Q_v) \text{ and vector } y. \tag{2.17}$$

THEOREM 5 (Rockafellar [6, §8]). *Suppose the inf-boundedness condition* (2.16) *is satisfied. Let v be such that the constraints of (Q_v) are consistent and every optimal solution x to (Q_v) satisfies the constraint qualification* (2.14). *Then $\partial q(v)$ is a nonempty compact set with*

$$\partial q(v) \subset \text{co}\, Z(v) \ \ \textit{and} \ \ \text{ext}\, \partial q(v) \subset Z(v). \tag{2.18}$$

In particular q is locally Lipschitzian around v with

$$q^{\circ}(v;h) \leq \sup_{z \in Z(v)} z \cdot h \ \ \textit{for all } h. \tag{2.19}$$

Any λ satisfying $|z| < \lambda$ for all $z \in Z(v)$ serves as a local Lipschitz constant:

$$|q(v'') - q(v')| \leq \lambda |v'' - v'| \ \ \textit{when } v' \textit{ and } v'' \textit{ are near } v. \tag{2.20}$$

The generality of the constraint structure in Theorem 5 will make possible in the next section an application to the study of multifunctions.

3. STABILITY OF CONSTRAINT SYSTEMS

The sensitivity results that have just been presented are concerned with what happens to the optimal value in a problem when parameters vary. It turns out, though, that they can be applied to the study of what happens to the feasible solution set and the optimal solution set. In order to explain this and indicate the main results, we must consider the kind of Lipschitzian property that pertains to multifunctions (set-valued mappings) and the way that this can be characterized in terms of an associated distance function.

Let $\Gamma: R^d \rightrightarrows R^n$ be a closed-valued multifunction, i.e. $\Gamma(v)$ is for each $v \in R^d$ a closed subset of R^n, possibly empty. The motivating examples are, first, $\Gamma(v)$ taken to be the set of all feasible solutions to the parameterized optimization problem (Q_v) above, and second, $\Gamma(v)$ taken to be the set of all optimal solutions to (Q_v).

One says that $\Gamma(v)$ is *locally Lipschitzian* around v if for all v' and v'' in some neighborhood of v one has $\Gamma(v')$ and $\Gamma(v'')$ nonempty and bounded with

$$\Gamma(v) = \{x \mid F(v,x) \in C \text{ and } (v,x) \in D\}. \tag{3.4}$$

Suppose for a given v that Γ is locally bounded at v, and that $\Gamma(v)$ is nonempty with the constraint qualification (2.14) satisfied by every $x \in \Gamma(v)$. Then Γ is locally Lipschitzian around v.

COROLLARY. *Let $\Gamma: R^d \rightrightarrows R^n$ be any multifunction whose graph $D = \{(v,x) \mid x \in \Gamma(v)\}$ is closed. Suppose for a given v that Γ is locally bounded at v, and that $\Gamma(v)$ is nonempty with the following condition satisfied for every $x \in \Gamma(v)$:*

$$\text{the only vector } z \text{ with } (z,0) \in N_D(v,x) \text{ is } z = 0. \tag{3.5}$$

Then Γ is locally Lipschitzian around v.

The corollary is just the case of the theorem where the constraint $F(v,x) \in C$ is trivialized. It corresponds closely to a result of Aubin [17], according to which Γ is "pseudo-Lipschitzian" relative to the particular pair (v,x) with $x \in \Gamma(v)$ if

$$\text{the projection of the tangent cone } T_D(v,x) \subset R^d \times R^n \text{ on } R^d \text{ is all of } R^d. \tag{3.6}$$

Conditions (3.5) and (3.6) are equivalent to each other by the duality between $N_D(v,x)$ and $T_D(v,x)$. The "pseudo-Lipschitzian" property of Aubin, which will not be defined here, is a suitable localization of Lipschitz continuity which facilitates the treatment of multifunctions Γ with $\Gamma(v)$ unbounded, as is highly desirable for other purposes in optimization theory (for instance the treatment of epigraphs dependent on a parameter vector v). As a matter of fact, the results in Rockafellar [18] build on this concept of Aubin and are not limited to locally bounded multifunctions. Only a special case has been presented in the present paper.

This topic is also connected with interesting ideas that Aubin has pursued towards a differential theory of multifunctions. Aubin defines the multifunction whose graph is the Clarke tangent cone $T_D(v,x)$, where D is the graph of Γ, to be the ***derivative*** of Γ at v relative to the point $x \in \Gamma(v)$. In denoting this derivative multifunction by $\Gamma'_{v,x}$, we have, because $T_D(v,x)$ is a closed convex cone, that $\Gamma'_{v,x}$ is a ***closed convex process*** from R^d to R^n in the sense of convex analysis [3, §39]. Convex processes are very much akin to linear transformations, and there is quite a ***convex algebra*** for them (see [3, §39], [19], and [20]). In particular, $\Gamma'_{v,x}$ has an ***adjoint*** $\Gamma'^{*}_{v,x}: R^n \rightrightarrows R^d$, which turns out in this case to be the closed convex process with

$$\text{gph } \Gamma'^{\bullet}_{v,x} = \{(w,z) \mid (z,-w) \in N_D(v,x)\}.$$

In these terms Aubin's condition (3.6) can be written as dom $\Gamma'_{v,x} = R^d$, whereas the dual condition (3.5) is $\Gamma'^{\bullet}_{v,x}(0) = \{0\}$. The latter is equivalent to $\Gamma'^{\bullet}_{v,x}$ being locally bounded at the origin.

There is too much in this vein for us to bring forth here, but the few facts we have cited may serve to indicate some new directions in which nonsmooth analysis is now going. We may soon have a highly developed apparatus that can be applied to the study of all kinds of multifunctions and thereby to subdifferential multifunctions in particular.

For example, as an aid in the analysis of the stability of optimal solutions and multiplier vectors in problem (Q_v), one can take up the study of the Lipschitzian properties of the multifunction

$\Gamma(v)$ = set of all (x,y,z) such that x is feasible in (Q_v)

and the optimality condition (2.13) is satisfied.

Some results on such lines are given in Aubin [17] and Rockafellar [21].

REFERENCES

[1] F.H. Clarke, "Generalized gradients and applications", Trans. Amer. Soc. 205 (1975), pp.247-262.

[2] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley-Interscience, New York, 1983.

[3] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton NJ, 1970.

[4] R.T. Rockafellar, "Favorable classes of Lipschitz continuous functions in subgradient optimization", *Processes in Nondifferentiable Optimization*, E. Nurminski (ed.), IIASA Collaborative Proceeding Series, International Institute for Applied Systems Analysis, Laxenburg, Austria, 1982, pp. 125-143.

[5] S. Saks, *Theory of the Integral*, Monografie Matematyczne Ser., no. 7, 1937; 2nd rev.ed. Dover Press, New York, 1964.

[6] R.T. Rockafellar, "Extensions of subgradient calculus with applications to optimization", J. Nonlinear Anal., to appear in 1985.

[7] R.W. Chaney, Math. Oper. Res. 9 (1984).

[8] R.T. Rockafellar, "Generalized directional derivatives and subgradients of nonconvex functions", Canadian J. Math. *32* (1980), pp. 157-180.

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

[10] R.T. Rockafellar, "Clarke's tangent cones and the boundaries of closed sets in R^n", J. Nonlin. Anal. 3 (1970), pp.145-154.

[11] F.H. Clarke, "A new approach to Lagrange multipliers", Math. Oper. Res. 1 (1976), pp. 165-174.

[12] J-B Hiriart-Urruty, "Refinements of necessary optimality conditions in nondifferentiable programming, I," Appl. Math. Opt. 5 (1979), pp.63-82.

[13] R.T. Rockafellar, "Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming", Math. Prog. Study 17 (1982), 28-66.

[14] J. Gauvin, "The generalized gradient of a marginal function in mathematical programming problem", Math. Oper. Res. 4 (1979), pp.458-463.

[15] V.F. Demyanov and V.N Malozemov, "On the theory of nonlinear minimax problems", Russ. Math. Surv. 26 (1971), 57-115.

[16] R.T. Rockafellar,"Directional differentiability of the optimal value in a nonlinear programming problem", Math. Prog. Studies 21 (1984), pp. 213-226.

[17] J.P. Aubin, "Lipschitz behavior of solutions to convex minimization problems", Math. Oper. Res. 9 (1984), pp. 87-111.

[18] R.T. Rockafellar, "Lipschitzian properties of multifunctions", J. Nonlin. Anal., to appear in 1985.

[19] R.T. Rockafellar, "Convex algebra and duality in dynamic models of production", in Mathematical Models of Economic (J. Los', ed.), North-Holland, 1973, pp.351-378.

[20] R.T. Rockafellar, ***Monotone Processes of Convex and Concave Type***, Memoir no.77, Amer. Math. Soc., Providence RI, 1967.

[21] R.T. Rockafellar, "Maximal monotone relations and the second derivatives of nonsmooth functions", Ann. Inst. H. Poincaré, Analyse Non Linéaire 2 (1985), pp.167-184.