PROTO-DERIVATIVES AND THE GEOMETRY OF SOLUTION MAPPINGS IN NONLINEAR PROGRAMMING

A. B. Levy[1] **and R. T. Rockafellar**[2]

Abstract: We quantify the sensitivity of KKT pairs associated with a parameterized family of nonlinear programming problems. Our approach involves proto-derivatives, which are generalized derivatives appropriate even in cases when the KKT pairs are not unique; we investigate what the theory of such derivatives yields in the special case when the KKT pairs are unique (locally). We demonstrate that the graph of the KKT multifunction is just a reoriented graph of a Lipschitz mapping, and use proto-differentiability to show that the graph of the KKT multifunction actually has the stronger property of being a reorientation of the graph of a B-differentiable mapping. Our results indicate that proto-derivatives provide the same kind of information for possibly set-valued mappings (like the KKT multifunction) that B-derivatives provide for single-valued mappings.

Key words: Proto-derivative; sensitivity analysis; nonlinear programming

September, 1995

[1] Department of Mathematics, Bowdoin College, Brunswick, ME 04011 USA
[2] Department of Mathematics, University of Washington, Seattle, WA 98195 USA

1. INTRODUCTION

A nonlinear programming problem in $x \in I\!R^n$ as parameterized by $w \in I\!R^d$ can be formulated in terms of a set $X \subseteq I\!R^n$, functions f_i on $I\!R^n \times I\!R^d \to I\!R$ for $i = 0, 1, \ldots, m$, and a convex cone $K \subseteq I\!R^m$:

$$\text{minimize } f_0(w, x) \text{ over all } x \in C(w) \tag{1}$$

where

$$C(w) := \{x \in X : F(w, x) \in K\} \quad \text{for} \quad F(w, x) = \big(f_1(w, x), \ldots, f_m(w, x)\big). \tag{2}$$

It will be supposed here that X and K are closed and every f_i is of class $\mathcal{C}^2$.

An important issue in the study of such a problem is the response of solutions to changes in the parameters, but this faces the difficulty that there can be several (or many, or no) optimal solutions x associated with a particular w. The solution mapping $w \mapsto x$ is necessarily then a *multifunction* (set-valued) whose "rates of change" cannot be captured by traditional methodology. The same goes for the mapping $w \mapsto (x, y)$ that assigns to each w the pairs (x, y) in which x satisfies a first-order optimality condition with y as multiplier vector.

Because of this difficulty, much of the theory of rates of change in the dependence of solutions on parameters has centered so far on circumstances in which the multifunction under investigation happens to be single-valued in a local sense. An adequate conceptual framework is provided then by relatively elementary notions of Lipschitz continuity and one-sided differentiability. It is not necessary to appeal to the more challenging notions of such kind that have been devised for handling multifunctions in general. But by narrowing the scope to mappings that are essentially single-valued, a significant part of the overall picture could be lost.

The multifunctions that associate parameter elements w with solutions x or quasi-solution pairs (x, y) tend to be very special within the realm of multifunctions. Although not necessarily single-valued, their graphs reflect many of the geometric features of the graphs of single-valued mappings. Our aim in this paper is to develop such geometry and its consequences, with emphasis on "proto-differentiability" as providing the natural counterpart to the type of one-sided differentiability, namely B-differentiability, that has come to prevail in studies of the single-valued setting. For this purpose we direct our attention mainly to quasi-solution pairs (x, y), as follows.

The generalized Karush-Kuhn-Tucker (KKT) conditions for (1)–(2) at a feasible solution x, as established in Rockafellar [25], concern a multiplier vector $y = (y_1, \ldots, y_m)$ such that

$$y \in N_K\big(F(w,x)\big), \qquad -\big[\nabla_x f_0(w,x) + y_1 \nabla_x f_1(w,x) + \cdots + y_m \nabla_x f_m(w,x)\big] \in N_X(x), \quad (3)$$

where $N_K\big(F(w,x)\big)$ is the normal cone to K at $F(w,x)$ and $N_X(x)$ is the normal cone to X at x. The first normal cone can be taken in the sense of convex analysis, but unless X happens to be convex the second should be interpreted in the broader sense of limits of proximal norm vectors (unconvexified); cf. [25]. Such a multiplier vector exists when x is locally optimal and satisfies the constraint qualification that there is no vector $y \neq 0$ such that

$$y \in N_K\big(F(w,x)\big), \qquad -\big[y_1 \nabla_x f_1(w,x) + \cdots + y_m \nabla f_m(w,x)\big] \in N_X(x). \qquad (4)$$

(This is given in Theorem 4.2 of [25] for the case of $K = I\!R^s_- \times I\!R^{m-s}$; the version for general K can be obtained by applying Theorem 10.1 of that paper to the indicator function δ_K.)

When $X = I\!R^n$ and $K = I\!R^s_- \times I\!R^{m-s}$, which we refer to as the *conventional case* of our problem, conditions (3) and (4) reduce to the classical KKT conditions and the dual form of the Mangasarian-Fromovitz constraint qualification. Other choices of the cone K allow for coverage of positive-definite programming, for instance. In taking X to be a proper subset of $I\!R^n$ one can incorporate nonnegativity requirements, upper and lower bounds on variables, and indeed other constraints of any kind without having to introduce additional multipliers whose perturbations might need to be coped with. In the *convex case* of our problem, i.e., when X is convex, $f_0(w,x)$ is convex in x, and $F(w,x)$ is convex in x with respect to the partial ordering induced by K, the generalized KKT conditions (3) on (x,y) are sufficient for x to be a globally optimal solution to (1)–(2).

We are interested in the KKT multifunction that assigns to each w the pairs (x,y) satisfying (3). In the conventional case, at least, much is already known about the properties of this multifunction; see for example [14], [13], [16], [2], [3], [1], [4]). Robinson's property of strong regularity provides a useful criterion for localized single-valuedness and Lipschitz continuity, even B-differentiability [16]. Closely tied to such results is the analysis of parametric dependence not only on w but on certain "canonical" parameters as well. The role of these canonical elements is to ensure that the parameterization is sufficiently rich. They are also in fact the key to the graphical geometry that is the topic here.

The *primal canonical perturbation* vector $u \in I\!R^m$ shifts $F(w,x)$ to $F(w,x) + u$, while the *dual canonical perturbation* vector $v \in I\!R^n$ shifts $f_0(w,x)$ to $f_0(w,x) - \langle v, x\rangle$. In bringing these vectors in, we adopt the format of a nonlinear programming problem parameterized by $(u,v,w) \in I\!R^m \times I\!R^n \times I\!R^d$:

$$\text{minimize } \; f_0(w,x) - \langle v, x\rangle \; \text{ over all } \; x \in C(u,w) \qquad (5)$$

where

$$C(u,w) := \big\{ x \in X : F(w,x) + u \in K \big\} \quad \text{for} \quad F(w,x) = \big(f_1(w,x), \ldots, f_m(w,x)\big). \quad (6)$$

The generalized KKT conditions on x and a multiplier vector y then have the form:

$$y \in N_K\big(F(w,x)+u\big) \quad \text{and} \quad v - \nabla_x f_0(w,x) - \nabla_x F(w,x)^\top y \in N_X(x). \qquad (7)$$

The multifunction we specifically look at in this format is

$$S_{\text{KKT}} : (u,v,w) \mapsto \big\{ (x,y) : \text{ (7) holds } \big\}. \qquad (8)$$

Obviously, the parameters represented by u and v could notationally be built into the specification of w, but there are advantages to making them explicit in this manner.

Our approach is to study the geometry and generalized differentiability of the multifunction S_{KKT} in (8) without restricting our focus only to situations where S_{KKT} exhibits single-valuedness. We identify properties that automatically turn into Lipschitz continuity or B-differentiability under single-valuedness, but have important content even without that. The issue of ascertaining single-valuedness is thus posed not as a prerequisite to the analysis of "rates of change," but as a separate matter that can be taken up subsequent to such analysis.

We begin in Section 2 by demonstrating under mild assumptions on X (convexity would suffice but is not required) that the graph of S_{KKT} is a Lipschitz manifold in the sense introduced by Rockafellar [18] in work with subgradient multifunctions. The dimension of this manifold is $m+n+d$, the same as that of the domain space for S_{KKT}, so this shows very powerfully the "functionlike" nature of S_{KKT}, even though its values need not be singletons everywhere. The Lipschitz aspect of the manifold furnishes preliminary insight into limitations on the effects of perturbations. When S_{KKT} does happen to be single-valued locally, the graphical geometry entails that it must be locally Lipschitz continuous too. A Lagrangian version of the generalized KKT conditions is used as a stepping stone to these results.

In moving on to generalized differentiability in Section 3, we rely on "proto-derivatives," which are distinguished by their being defined in terms of set convergence of graphs. In heuristic terms, a multifunction is proto-differentiable at a point in its graph, if the image of its graph near the point "stabilizes" as closer and closer views are examined. It was noted in [7] that a single-valued mapping is B-differentiable if and only if it is continuous and proto-differentiable with single-valued proto-derivative mapping (in which event the proto-derivatives are the same as the B-derivatives). We augment this here by showing that a single-valued locally Lipschitz continuous mapping is B-differentiable if and only if it is proto-differentiable; there is no need in this case to assume single-valuedness of the proto-derivative multifunction. As a geometric extension of this fact, we prove that the graph a proto-differentiable multifunction is locally a Lipschitz manifold if and only if it is locally a B-differentiable Lipschitz manifold.

In application to sensitivity analysis, we prove that if the sets X and K are "fully amenable" (as holds in particular if they are polyhedral convex, but allows also for cases where X is not necessarily convex), then the multifunction $S_{\rm KKT}$ in (8) is proto-differentiable. This has the consequence that the graph of $S_{\rm KKT}$ is a B-differentiable Lipschitz manifold, a powerful property not previously observed.

Although new insights into graphical geometry are presented in this paper, the methodology of proto-differentiability has already been found to be effective in the study of multifunctions much more general than the one treated here; cf. our earlier work in [6]. On the other hand, many questions remain unanswered, especially concerning the suppression of some of the elements involved in a setting like (8). In [7] we obtained proto-derivatives for quasi-solution multifunctions associated with first-order optimality conditions for similar minimization problems, but in the form $(v, w) \mapsto x$, without explicit dependence on primal perturbations u or pairing with multiplier elements y. Whether or not the graph of such a multifunction can be seen locally as a Lipschitz manifold is still unclear, however. As for the suppression of v, results of Levy [5] give approximations at least of "outer" proto-derivatives, which are weaker than true proto-derivatives, but again the graphical geometry is not yet satisfactorily understood.

Of course, when some localization of the multifunction $S_{\rm KKT}$ in (8) is single-valued and Lipschitz continuous, or B-differentiable, there is no difficulty whatever retaining these properties in suppressing some of the elements. Even the submapping $w \mapsto x$ will be single-valued and Lipschitz continuous, or B-differentiable, in that case.

2. GRAPHS AS LIPSCHITZ MANIFOLDS

The notation $S : I\!R^k \rightrightarrows I\!R^l$ indicates a multifunction, or set-valued mapping, that assigns to each $w \in I\!R^k$ a set of vectors $z \in I\!R^l$. The *graph* of S is $\operatorname{gph} S = \{(w, z) : z \in S(w)\}$. For some choices of w the set $S(w)$ may be empty, or it may just be a singleton. The *effective domain* of S is $\operatorname{dom} S = \{w : S(w) \neq \emptyset\}$. As a special case, S might be *single-valued* on a set D, meaning that $S(w)$ is a singleton for every $w \in D$ (and hence in particular that $D \subseteq \operatorname{dom} S$). A weaker concept is that of S being *locally single-valued around* $(\bar{w}, \bar{z})$, a point of $\operatorname{gph} S$, which refers to the existence of neighborhoods W of $\bar{w}$ and Z of $\bar{z}$ such that the "submapping" $w \in W \mapsto S(w) \cap Z$ is single-valued.

When such single-valuedness occurs, one can go on to ask whether the submapping in question is actually continuous or even Lipschitz continuous. If so, there are obvious consequences for the geometry of $\operatorname{gph} S$. But the same graphical geometry would persist if a smooth one-to-one mapping of $I\!R^{k+l}$ onto itself were applied to $\operatorname{gph} S$, regardless of whether the image of $\operatorname{gph} S$ were again locally the graph of a single-valued mapping. Putting this in reverse, one can have situations where $\operatorname{gph} S$ is not locally single-valued around $(\bar{w}, \bar{z})$, and yet it corresponds under a certain nonlinear local transformation, smooth in both directions, to the graph of a single-valued, Lipschitz continuous mapping. The following concept assists in formalizing the idea.

Definition 2.1. *A subset M of $I\!R^N$ is locally a Lipschitz manifold of dimension s around the point $\bar{u} \in M$ in the sense of* [18] *if, under a smooth change of coordinates around $\bar{u}$, it can be identified locally with the graph of a Lipschitz continuous mapping of dimension s, or in other words, if there is an open neighborhood U of $\bar{u}$ in $I\!R^N$ and a one-to-one mapping Φ of U onto an open set in $I\!R^s \times I\!R^{N-s}$ with both Φ and Φ^{-1} continuously differentiable, such that $\Phi(M \cap U)$ is the graph of some Lipschitz continuous mapping $H : O \to I\!R^{N-s}$ for an open set O in $I\!R^s$.*

In particular, the graph of any locally Lipschitz continuous mapping $G : I\!R^k \to I\!R^l$ (single-valued) is locally a Lipschitz manifold of dimension k in $I\!R^{k+l}$; for this one can take Φ to be the identity. The same is true then of G^{-1}, which need not be a single-valued mapping.

Indeed, if a multifunction $S : I\!R^k \rightrightarrows I\!R^l$ is such that $\operatorname{gph} S$ is a Lipschitz manifold of dimension s around $(\bar{w}, \bar{z})$, and if the multifunction $S' : I\!R^{k'} \rightrightarrows I\!R^{l'}$ is such that $\operatorname{gph} S'$ corresponds to $\operatorname{gph} S$ under a one-to-one transformation which is continuously differentiable in both directions and associates $(\bar{w}, \bar{z})$ with $(\bar{w}', \bar{z}')$ (with $k + l = k' + l'$), then $\operatorname{gph} S'$ too is locally a Lipschitz manifold of dimension s around $(\bar{z}', \bar{w}')$. (Here the transformation could merely be local.) The following elementary facts illustrate this. (By the *inverse* of

S is meant the multifunction $S^{-1} : z \mapsto \{w : z \in S(w)\}$.)

Proposition 2.2. *If the graph of $S : I\!R^k \rightrightarrows I\!R^l$ is locally a Lipschitz manifold of dimension s around $(\bar{w}, \bar{z})$, then the graph of the inverse $S^{-1} : I\!R^l \rightrightarrows I\!R^k$ is likewise locally a Lipschitz manifold of dimension s around $(\bar{z}, \bar{w})$.*

Proposition 2.3. *If the graph of $S : I\!R^k \rightrightarrows I\!R^l$ is locally a Lipschitz manifold of dimension s around $(\bar{w}, \bar{z})$, and if the single-valued mapping $G : I\!R^k \to I\!R^l$ is of class $\mathcal{C}^1$, then the graph of $S + G$ is locally a Lipschitz manifold of dimension s around $\big(\bar{w}, \bar{z} + G(\bar{z})\big)$.*

Proof. The mapping $\Phi : (w, z) \mapsto \big(w, z + G(z)\big)$ carries one graph onto the other. This is a $\mathcal{C}^1$ mapping for which the inverse, $\Phi^{-1} : (w, z) \mapsto \big(w,\ z - G(z)\big)$, is $\mathcal{C}^1$ as well. □

Especially important is the next example, which concerns a key class of mappings which can well fail to be single-valued.

Proposition 2.4 (Minty [8])**.** *For any maximal monotone multifunction $T : I\!R^n \rightrightarrows I\!R^n$, the set* gph T *is locally a Lipschitz manifold of dimension n around all of its points.*

The subgradient multifunctions ∂f of proper, lower semicontinuous, convex functions $f : I\!R^n \to \overline{I\!R}$ (the extended reals) are known to be maximal monotone, cf. [17], and the same is therefore true for the graphs of normal cone multifunctions $x \mapsto N_X(x)$ when X is a closed, convex set (and $N_X(x)$ is taken to be the empty set for $x \notin X$), inasmuch as $N_X = \partial \delta_X$.

Corollary 2.5. *For any proper, lsc, convex function $f : I\!R^n : \overline{I\!R}$ the graph of the subgradient multifunction $\partial f : I\!R^n \rightrightarrows I\!R^n$ is locally a Lipschitz manifold of dimension n around all of its points. Likewise, for any nonempty, closed, convex set $X \subseteq I\!R^n$ the graph of the normal cone multifunction $N_X : I\!R^n \rightrightarrows I\!R^n$ is locally a Lipschitz manifold of dimension n around all of its points.*

The scope of this geometric property can be greatly be extended now beyond the bounds of convex analysis on the basis of recent results of Poliquin and Rockafellar [11]. The subgradients then are "limiting proximal subgradients." Recall that a function $f : I\!R^n \to \overline{I\!R}$ is *amenable* at a point $\bar{x}$ if there is a local representation $f(x) = g\big(G(x)\big)$ in which G is a $\mathcal{C}^1$ mapping into a space $I\!R^m$, the function $g : I\!R^m \to \overline{I\!R}$ is a proper, lower semicontinuous and convex, the point $G(\bar{x})$ lies in dom g, and the constraint qualification is satisfied that there is no nonzero vector y in the normal cone $N_{\operatorname{dom} g}\big(G(\bar{x})\big)$ for which the gradient of the mapping $x \mapsto \langle y, G(x) \rangle$ vanishes. It is *strongly amenable* if there is such a representation with G of class $\mathcal{C}^2$ rather than just of class $\mathcal{C}^1$.

The category of strongly amenable functions includes all $\mathcal{C}^2$ functions f, all proper, lsc, convex functions f, and more. For instance, it includes all functions of the form $f = h + \delta_D$ in which h is the pointwise max of a finite collection of $\mathcal{C}^2$ functions on $I\!R^n$ and D is a subset of $I\!R^n$ specified by finitely many equality and inequality constraints for $\mathcal{C}^2$ functions—provided only that, at the point $\bar{x}$ where the amenability is to be tested, the Mangasarian-Fromovitz constraint qualification is fulfilled; see [19] and [10] (the terminology of amenability comes from the latter). Any function of the kind just described is in fact "fully amenable," a higher property which will be of interest in Section 3 and will explained there when it is needed.

Theorem 2.6. *If a function $f : I\!R^n \to \overline{I\!R}$ is strongly amenable at $\bar{x}$, then for any subgradient $\bar{v} \in \partial f(\bar{x})$ the graph of the subgradient multifunction $\partial f : I\!R^n \rightrightarrows I\!R^n$ is locally a Lipschitz manifold of dimension n around $(\bar{x}, \bar{v})$.*

Proof. Proposition 2.5 of Poliquin and Rockafellar [11] establishes that f is "prox-regular" and "subdifferentially continuous" at $\bar{x}$. Theorem 3.2 of the same paper demonstrates, among other things, that when these properties hold and $\bar{v} \in \partial f(\bar{x})$ there is a neighborhood of $(\bar{x}, \bar{v})$ with respect to which the graph of ∂f is maximal submonotone, i.e., such that the multifunction $T = \partial f + \lambda I$ is maximal monotone for some $\lambda > 0$. Then the graph of T is, by a localization of Minty's theorem cited above, locally a Lipschitz manifold of dimension n around $(\bar{x}, \bar{v} + \lambda\bar{x})$. Since $\partial f = T - \lambda I$, we conclude from Proposition 2.3 that the graph of ∂f itself is locally a Lipschitz manifold of dimension n around $(\bar{x}, \bar{v})$. □

A set D is defined to be *amenable* or *strongly amenable* at a point $\bar{x}$ if its indicator function has these properties. In particular, any closed, convex set D is strongly amenable at any of its points, and so too is any set D specified locally by finitely many $\mathcal{C}^2$ constraints in such a manner that the Mangasarian-Fromovitz constraint qualification holds.

Corollary 2.7. *If a set $X \subseteq I\!R^n$ is strongly amenable at one of its points $\bar{x}$, then for any normal vector $\bar{v} \in N_X(\bar{x})$ the graph of the multifunction N_X is locally a Lipschitz manifold of dimension n around $(\bar{x}, \bar{v})$.*

Proof. This specializes Theorem 2.6 to the case of $f = \delta_X$. □

These results have laid the foundation for proving the following fact about the geometry of the general multifunction S_{KKT} introduced above.

Theorem 2.8. *Let $(\bar{x}, \bar{y}) \in S_{\text{KKT}}(\bar{u}, \bar{v}, \bar{w})$ and suppose the set X is strongly amenable at the point $\bar{x}$. Then the graph of S_{KKT} is locally a Lipschitz manifold of dimension $m+n+d$ around $(\bar{u}, \bar{v}, \bar{w}; \bar{x}, \bar{y})$.*

Proof. The relation $y \in N_K\big(F(w,x)+u\big)$ in (7) can be expressed in terms of the polar cone $Y = K^*$ as $F(w,x)+u \in N_Y(y)$, cf. [17], Corollary 23.5.4. Let $D = I\!R^d \times X \times Y$, so that

$$N_D(w,x,y) = \big\{(z',v',u') : z' = 0 \in I\!R^d,\ v' \in N_X(x),\ u' \in N_Y(y)\big\}. \tag{9}$$

Further, define the mapping $G : I\!R^d \times I\!R^n \times I\!R^m$ by

$$G(w,x,y) = \big(0,\ \nabla f_0(w,x) + \nabla_x F(w,x)^\top y, -F(w,x)\big).$$

In this notation, the KKT conditions (7) come down to $(0,v,u) \in N_D(w,x,y)+G(w,x,y)$. We thus have

$$(w,x,y;z,v,u) \in \operatorname{gph}(N_D+G) \quad\Longleftrightarrow\quad \begin{cases} (u,v,w;x,y) \in \operatorname{gph} S_{\mathrm{KKT}}, \\ z = 0. \end{cases} \tag{10}$$

Therefore, the assertion that $\operatorname{gph} S_{\mathrm{KKT}}$ is locally a Lipschitz manifold of dimension $m+n+d$ around $(\bar u,\bar v,\bar w;\bar x,\bar y)$ is equivalent to the assertion that $\operatorname{gph}(N_D+G)$ has this property around $(\bar w,\bar x,\bar y;0,\bar v,\bar u)$.

Because Y is a closed, convex cone, we know that D is strongly amenable at $(\bar w,\bar x,\bar y)$ whenever the points $\bar x \in X$ and $\bar y \in Y$ are such that X is strongly amenable at $\bar x$. Then by Corollary 2.7 the graph of the normal cone multifunction N_D is locally a Lipschitz manifold of dimension $d+n+m$ around $(\bar w,\bar x,\bar y,0,v',u')$ for any vectors $v' \in N_X(\bar x)$ and $u' \in N_Y(\bar y)$. On the other hand, the mapping G is of class $\mathcal{C}^1$ (by our blanket assumption that f_0 and the component functions f_i of F are of class $\mathcal{C}^2$). It follows then by Proposition 2.3 that the graph of N_D+G is locally a Lipschitz manifold of dimension $d+n+m$ around any of its points $(\bar w,\bar x,\bar y,\bar z,\bar v,\bar u)$ such that X is strongly amenable at $\bar x$. This gives what we need. □

3. GRAPHICAL DIFFERENTIABILITY

A mapping $G : I\!R^k \to I\!R^l$ (single-valued) is *B-differentiable* at $\bar{w}$ as defined by Robinson [15] if the difference quotient mappings

$$\Delta_t G(\bar{w}) : \omega \mapsto \big[G(\bar{w} + t\omega) - G(\bar{w})\big]/t \quad \text{for} \quad t > 0$$

converge pointwise as $t \searrow 0$ to a continuous mapping $H : I\!R^k \to I\!R^l$ and do so uniformly on bounded sets. Then $H(\omega)$ is the one-sided directional derivative of G relative to ω.

Classical differentiability is the case where $H(\omega)$ is linear in ω. The definition of B-differentiability always implies at least that $H(0) = 0$ and that H is positively homogeneous, i.e., $H(\lambda\omega) = \lambda H(\omega)$ for $\lambda > 0$ (hence also for $\lambda = 0$), but it allows $H(\omega)$ to be nonlinear in ω in other respects. We will denote H by $DG(\bar{w})$ and call this the B-derivative mapping for G at $\bar{w}$.

Of course, when G is Lipschitz continuous around $\bar{w}$ the same holds uniformly for the mappings $\Delta_t G(\bar{w})$, and mere pointwise convergence of these mappings to H as $t \searrow 0$ implies that H is globally Lipschitz continuous and that the convergence is uniform on bounded sets. In this case, therefore, B-differentiability is automatic simply from the existence of $\lim_{t \searrow 0} \big[G(\bar{w} + t\omega) - G(\bar{w})\big]/t$ for every ω.

Definition 3.1. *A subset M of $I\!R^N$ is locally a B-differentiable Lipschitz manifold of dimension s near the point $\bar{u} \in M$ if, under a smooth change of coordinates around $\bar{u}$, it can be identified locally with the graph of a Lipschitz continuous mapping of dimension s that happens also to be B-differentiable, or in other words, if there is an open neighborhood U of $\bar{u}$ in $I\!R^N$ and a one-to-one mapping Φ of U onto an open set in $I\!R^s \times I\!R^{N-s}$ with both Φ and Φ^{-1} continuously differentiable, such that $\Phi(M \cap U)$ is the graph of some B-differentiable, Lipschitz continuous mapping $H : O \to I\!R^{N-s}$ for an open set O in $I\!R^s$.*

The graph of any B-differentiable, locally Lipschitz continuous mapping $G : I\!R^k \to I\!R^l$ is locally a B-differentiable Lipschitz manifold in $I\!R^{k+l}$. So too is the graph of G^{-1} (generally just a multifunction). Following the same geometric patterns as in the preceding section, we arrive at the principle that if a multifunction $S : I\!R^k \rightrightarrows I\!R^l$ is such that $\operatorname{gph} S$ is locally a B-differentiable Lipschitz manifold of dimension s around $(\bar{w}, \bar{z})$, and if the multifunction $S' : I\!R^{k'} \rightrightarrows I\!R^{l'}$ is such that $\operatorname{gph} S'$ corresponds to $\operatorname{gph} S$ under a one-to-one transformation which is continuously differentiable in both directions and associates $(\bar{w}, \bar{z})$ with $(\bar{w}', \bar{z}')$, then $\operatorname{gph} S'$ too is a B-differentiable Lipschitz manifold of dimension s around $(\bar{z}', \bar{w}')$.

Facts analogous to Propositions 2.2 and 2.3 can be stated at once.

Proposition 3.2. *If the graph of $S : I\!R^k \rightrightarrows I\!R^l$ is locally a B-differentiable Lipschitz manifold of dimension s around $(\bar{w}, \bar{z})$, then the graph of the inverse $S^{-1} : I\!R^l \rightrightarrows I\!R^k$ is likewise locally a B-differentiable Lipschitz manifold of dimension s around $(\bar{z}, \bar{w})$.*

Proposition 3.3. *If the graph of $S : I\!R^k \rightrightarrows I\!R^l$ is locally a B-differentiable Lipschitz manifold of dimension s around $(\bar{w}, \bar{z})$, and if the single-valued mapping $G : I\!R^k \to I\!R^l$ is of class $\mathcal{C}^1$, then the graph of the multifunction $S + G$ is locally a B-differentiable Lipschitz manifold of dimension s around $\big(\bar{w}, \bar{z} + G(\bar{z})\big)$.*

We wish to take advantage of such geometry in treating various multifunctions that arise in optimization. For this we are led to a concept of generalized differentiability called proto-differentiability, which was introduced in [21]. Proto-differentiability is distinguished from other differentiability notions through its utilization of set convergence of graphs.

Consider any multifunction $S : I\!R^k \rightrightarrows I\!R^l$ and any pair $(\bar{w}, \bar{z}) \in \operatorname{gph} S$. For each $t > 0$ one can form the difference quotient multifunction

$$\Delta_t S(\bar{w}|\bar{z}) : \omega \mapsto \big[S(\bar{w} + t\omega) - \bar{z}\big]/t \quad \text{for} \quad t > 0.$$

(When $\bar{z}$ happens to be the sole element of $S(\bar{w})$, the notation $\Delta_t S(\bar{w})$ suffices.) Instead of asking the difference quotient multifunctions $\Delta_t S(\bar{w}|\bar{z})$ to converge in some kind of pointwise sense as $t \searrow 0$, proto-differentiability asks that they converge graphically, i.e., that their graphs converge as subsets of $I\!R^k \times I\!R^l$ to the graph of some multifunction $H : I\!R^k \rightrightarrows I\!R^l$. Then H is the *proto-derivative* multifunction at $\bar{w}$ for $\bar{z}$; the notation we will use for this multifunction H is $DS(\bar{w}|\bar{z})$. It associates with each $\omega \in I\!R^k$ some (possibly empty) subset of $I\!R^l$.

The concept of Painlevé-Kuratowski set convergence underlies the formation of these graphical limits. It refers to a kind of approximation described from two sides as follows. The *inner set limit* of a parameterized family of sets $\{C_t\}_{t>0}$ in $I\!R^N$ is the set of points η such that for *every* sequence $t_k \searrow 0$ there is a sequence of points $\eta_k \in C_{t_k}$ with $\eta_k \to \eta$. The *outer set limit* of the family is the set of points η such that for *some* sequence $t_k \searrow 0$ there is a sequence of points $\eta_k \in C_{t_k}$ with $\eta_k \to \eta$. When the inner and outer set limits coincide, the common set C is the *limit* as $t \searrow 0$.

In our framework, this is applied to sets that are the graphs of multifunctions. For a multifunction $S : I\!R^k \rightrightarrows I\!R^l$ and any pair $(\bar{w}, \bar{z})$ in $\operatorname{gph} S$, the graph of the difference quotient mapping $\Delta_t S(\bar{w}|\bar{z})$ is $t^{-1}\big[\operatorname{gph} S - (\bar{w}, \bar{z})\big]$. The multifunction $D^+S(\bar{w}|\bar{z}) : I\!R^k \rightrightarrows I\!R^l$ having as its graph the outer limit of the sets $\operatorname{gph} \Delta_t S(\bar{w}|\bar{z})$ as $t \searrow 0$ is called the *outer graphical derivative* of S at $\bar{w}$ for $\bar{z}$. In parallel, the multifunction $D^-S(\bar{w}|\bar{z}) : I\!R^k \rightrightarrows I\!R^l$

having as its graph the inner limit of these sets is the *inner graphical derivative*. Proto-differentiability of S at $\bar{w}$ for $\bar{z}$ is the case where the outer and inner derivatives agree, the common mapping being then the *proto-derivative*: $DS(\bar{w}|\bar{z}) = D^+S(\bar{w}|\bar{z}) = D^-S(\bar{w}|\bar{z})$, cf. Rockafellar [24]. (Again, if $\bar{z}$ happens to be the only element of $S(\bar{w})$, the notation can be simplified to $DS(\bar{w})$.)

The following result comes from [7] and clarifies the relationship in the single-valued case between proto-differentiability and B-differentiability.

Proposition 3.4 ([7], Prop. 2.2)**.** *Let O be an open neighborhood of a point $\bar{w} \in I\!R^k$ and consider a single-valued mapping $G : O \to I\!R^l$. Then G is B-differentiable at $\bar{w}$ if and only if G is continuous at $\bar{w}$ and (as a special case of a multifunction that happens to be single-valued) is proto-differentiable at $\bar{w}$ with $DG(\bar{w})$ single-valued, in which event one has the local expansion*

$$G(\bar{w} + t\omega) \;=\; G(\bar{w}) + tDG(\bar{w})(\omega) + o\big(t|\omega|\big) \quad \text{for} \quad t > 0.$$

When G happens to be Lipschitz continuous around $\bar{w}$, an even stronger result holds.

Proposition 3.5. *In the setting of Proposition 3.4, suppose that G is also Lipschitz continuous around $\bar{w}$. Then G is B-differentiable at $\bar{w}$ if and only if G is proto-differentiable at $\bar{w}$, in which event the proto-derivative multifunction reduces to the B-derivative mapping.*

Proof. Necessity follows immediately from Proposition 3.4. For sufficiency, we apply [22] (Theorem 4.3), noting that for the single-valued mapping G local Lipschitz continuity around $\bar{w}$ is equivalent to "pseudo-Lipschitz" continuity at $\bar{w}$ as a multifunction. (In [22], B-differentiability is equivalent to a property called "semi-differentiability.") □

These results mean that proto-differentiability extends to multifunctions, just in the manner that might be wished, the notion of one-sided directional differentiability deemed most appropriate in the sensitivity analysis of single-valued mappings, smooth or nonsmooth. The question of whether a certain mapping is single-valued or not can be dealt with as a separate issue, which need not be resolved before progress can be made on quantitative stability of solutions.

Theorem 3.6. *For a multifunction $S : I\!R^k \rightrightarrows I\!R^l$, let $(\bar{w}, \bar{z})$ be a point around which the graph of S is locally a Lipschitz manifold of dimension s. In order that the graph of S be locally in fact a B-differentiable Lipschitz manifold around $(\bar{w}, \bar{z})$, it is necessary and sufficient that, for all (w, z) within some neighborhood of $(\bar{w}, \bar{z})$ relative to* $\operatorname{gph} S$, *the mapping S should be proto-differentiable at w for z.*

Proof. A transformation Φ as in Definitions 2.1 and 3.1 allows us to pass to the framework where S is replaced by a single-valued Lipschitz continuous mapping G. Then Proposition 3.5 can be brought into play, and the result is immediate. □

The route to applying this result to optimality conditions lies in the second-order variational analysis of subgradient multifunctions and a further form of "amenability." Recall that a convex function $g : I\!R^m \to \overline{I\!R}$ is *piecewise linear-quadratic* if the set $\operatorname{dom} g$ is polyhedral and can be represented as the union of finitely many polyhedral convex sets, relative to each of which g has a formula as a polynomial function of degree no more than 2. A function $f : I\!R^n \to \overline{I\!R}$ is called *fully amenable* at $\bar{x}$ if it satisfies the earlier definition of being strongly amenable at $\bar{x}$ and does so with the additional condition that the convex function g in that representation is piecewise linear-quadratic.

Functions f of this type were first studied by Rockafellar [19] for their second-order "epi-derivatives." The connection between such generalized second derivatives and proto-derivatives of the corresponding subgradient multifunctions ∂f was established for convex functions by Rockafellar [23] and for arbitrary fully amenable functions by Poliquin [9]. The following result was obtained in particular.

Theorem 3.7 ([9]). *Suppose that the function $f : I\!R^n \to \overline{I\!R}$ is fully amenable at $\bar{x}$, and let $\bar{v} \in \partial f(\bar{x})$. Then for all pairs (x, v) in some neighborhood of $(\bar{x}, \bar{v})$ relative to $\operatorname{gph} \partial f$, the subgradient multifunction ∂f is proto-differentiable at x for v.*

We can now deduce from this an important geometric property of the graphs of subgradient multifunctions.

Theorem 3.8. *If the function $f : I\!R^n \to \overline{I\!R}$ is fully amenable at $\bar{x}$, then for any $\bar{v} \in \partial f(\bar{x})$ the graph of the subgradient multifunction $\partial f : I\!R^n \rightrightarrows I\!R^n$ is locally a B-differentiable Lipschitz manifold around $(\bar{x}, \bar{v})$.*

Proof. It is merely necessary to combine Theorem 3.7 with Theorems 2.6 and 3.6. □

There is no need here to discuss the large class of functions that are fully amenable (see [10]), because attention here is turned toward application to KKT conditions. The case of indicator functions is therefore the main one to consider.

A set $X \subseteq I\!R^n$ is called *fully amenable* at one of its points $\bar{x}$ if the indicator function δ_X is fully amenable at $\bar{x}$.

Corollary 3.9. *If a set $X \subseteq I\!R^n$ is fully amenable at $\bar{x}$, then for any $\bar{v} \in N_X(\bar{x})$ the graph of the normal cone multifunction $N_X : I\!R^n \rightrightarrows I\!R^n$ is locally a B-differentiable Lipschitz manifold around $(\bar{x}, \bar{v})$.*

The operational meaning of full amenability of a set can be elucidated as follows.

Proposition 3.10. *A set $X \subseteq I\!R^n$ is fully amenable at one of its points $\bar{x}$ if and only if there exists a local representation of X around $\bar{x}$ by a finite family of $\mathcal{C}^2$ constraints (equations, inequalities or a mixture) such that the Mangasarian-Fromovitz constraint qualification is satisfied at $\bar{x}$.*

Proof. An indicator function δ_D is convex and piecewise linear-quadratic if and only if D is a polyhedral convex set. Through its definition, therefore, full amenability of X at $\bar{x}$ means the existence of a neighborhood V of $\bar{x}$ yielding a representation $X \cap V = G^{-1}(D) \cap V$ for some polyhedral convex set D and $\mathcal{C}^2$ mapping G with the property that no nonzero vector $y \in N_D\big(G(\bar{x})\big)$ has $\nabla G(\bar{x})^\top y = 0$. The set D can be expressed by a system of finitely many linear equations and inequalities, and moreover this can be accomplished in such a manner that the system satisfies the Mangasarian-Fromovitz constraint qualification at $G(\bar{x})$. Namely, we can write $D = G_0^{-1}(D_0)$ for a set D_0 of the form $I\!R^q_+ \times I\!R^r$ and an affine mapping G_0 with the property that no nonzero vector $y_0 \in N_{D_0}\big(G_0(G(\bar{x}))\big)$ has $\nabla G_0(G(\bar{x}))^\top y_0 = 0$. Then the local representation $X = [G_0 \circ G]^{-1}(D_0)$ corresponds to a standard system of $\mathcal{C}^2$ constraints around $\bar{x}$ such that the Mangasarian-Fromovitz constraint qualification is satisfied at $\bar{x}$.

Proposition 3.11. *A polyhedral convex set $X \subseteq I\!R^n$ is fully amenable at all of its points.*

Proof. This is obvious from the comment made at the beginning of the proof of the preceding proposition. The mapping G in the definition of full amenability can in this case be taken to be the identity. □

Our principle geometric result about the KKT multifunction defined in (8) can now be stated and proved.

Theorem 3.12. *Let $(\bar{x}, \bar{y}) \in S_{\mathrm{KKT}}(\bar{u}, \bar{v}, \bar{w})$ and suppose the set X is fully amenable at the point $\bar{x}$. Suppose also that the cone K is fully amenable at $F(\bar{w}, \bar{x}) + \bar{u}$. Then the graph of S_{KKT} is locally a B-differentiable Lipschitz manifold of dimension $m + n + d$ around the point $(\bar{u}, \bar{v}, \bar{w}; \bar{x}, \bar{y})$.*

Proof. The argument closely follows the lines of the one employed for the related result in Theorem 2.8. We introduce D and G in exactly the same manner as there and use them to represent the graph of S_{KKT} as in (10). The challenge becomes that of demonstrating that the graph of $N_D + G$ is locally a B-differentiable Lipschitz manifold around $(\bar{w}, \bar{x}, \bar{y}; 0, \bar{v}, \bar{u})$. By virtue of Proposition 3.3, this can be accomplished by demonstrating that the graph of N_D is locally a B-differentiable Lipschitz manifold around the point

$$(\bar{w}, \bar{x}, \bar{y}; 0, \bar{v}', \bar{u}') \quad \text{for} \quad \bar{v}' = \bar{v} - \nabla_x f_0(\bar{w}, \bar{x}) - \nabla_x F(\bar{w}, \bar{x})^\top \bar{y}, \quad \bar{u}' = F(\bar{w}, \bar{x}) + \bar{u}.$$

The product structure exhibited for the graph of N_D in (9) brings this down to verifying that the graph of N_X is a B-differentiable Lipschitz manifold around the point $(\bar{x}, \bar{v}')$, whereas the graph of N_Y is such a manifold around $(\bar{y}, \bar{u}')$.

The required property of the graph of N_X follows from Corollary 3.9 through the full amenability assumed for X. The corresponding assumption for K likewise tells us that the graph of N_K is locally a B-differentiable Lipschitz manifold around $(\bar{u}', \bar{y})$. But $N_Y = N_K^{-1}$ through the polarity between Y and K, so by Proposition 3.2 the graph of N_Y is locally a B-differentiable Lipschitz manifold around $(\bar{y}, \bar{u}')$, as desired. □

Corollary 3.13. *The conclusion of Theorem 3.12 holds in particular when both X and K are polyhedral convex sets, or when both X and K can be represented by systems of finitely many $\mathcal{C}^2$ constraints for which the Mangasarian-Fromovitz constraint qualification is fulfilled at $\bar{x}$ and $F(\bar{w}, \bar{x}) + \bar{u}$, respectively.*

Proof. This specializes to the criteria for full amenability in Propositions 3.10 and 3.11. □

References

1. A.L. Dontchev, *Implicit function theorems for generalized equations*, to appear, Mathematical Programming, 1995.
2. A.L. Dontchev and W.W. Hager, *On Robinson's implicit function theorem*, Set-Valued Analysis and Differential Inclusions, Birkhäuser, 1991.
3. ———, *Implicit functions, Lipschitz maps, and stability in optimization*, to appear, Mathematics of Operations Research, 1995.
4. J. Kyparisis, *Parametric variational inequalities with multivalued solution sets*, Mathematics of Operations Research 17 (1992), pp. 341–364.
5. A.B. Levy, *Implicit set-valued mapping theorems and the sensitivity analysis of variational conditions*, preprint, 1995.
6. A.B. Levy and R.T. Rockafellar, *Sensitivity analysis of solutions to generalized equations*, Transactions of the American Mathematical Society 345 (1994), pp. 661–671.
7. ———, *Sensitivity of solutions in nonlinear programming problems with nonunique multipliers*, accepted for publication in Nonsmooth Optimization, (D. Du, L. Qi and R. Womersley, eds.), World Scientific Publishers, 1995.

8. G.J. Minty, *Monotone (nonlinear) operators in Hilbert space*, Duke Math. Journal 29 (1962), pp. 341-346.

9. R.A. Poliquin, *Proto-differentiation of subgradient set-valued mappings*, Canadian Journal of Mathematics 42 (1990), pp. 520-532.

10. R.A. Poliquin and R.T. Rockafellar, *Amenable functions in optimization*, Nonsmooth Optimization Methods and Applications (F. Giannessi, ed.), Gordon and Breach, 1992, pp. 338-353.

11. ———, *Prox-regular functions in variational analysis*, to appear in Transactions of the American Mathematical Society.

12. S.M. Robinson, *Generalized equations and their solutions, part I: Basic theory*, Mathematical Programming Study 10 (1979), pp. 128–141.

13. ———, *Strongly regular generalized equations*, Mathematics of Operations Research 5 (1980), pp. 43–62.

14. ———, *Generalized equations and their solutions, part ii: Applications to nonlinear programming*, Mathematical Programming Study 19 (1982), pp. 200–221.

15. ———, *Local structure of feasible sets in nonlinear programming, part iii: Stability and sensitivity*, Mathematical Programming Study 30 (1987), pp. 45–66.

16. ———, *An implicit-function theorem for a class of nonsmooth functions*, Mathematics of Operations Research 16 (1991), pp. 292–309.

17. R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

18. ———, *Maximal monotone relations and the second derivatives of nonsmooth functions*, Annales de l'Institût Henri Poincaré - Analyse non linéaire 2 (1985), pp. 167–184.

19. ———, *First and second-order epi-differentiability in nonlinear programming*, Transactions of the American Mathematical Society 307 (1988), pp. 75–108.

20. ———, *Perturbation of generalized Kuhn-Tucker points in finite-dimensional optimization*, Nonsmooth Analysis and Related Topics (F.H. Clarke et al., eds.), Plenum Press, 1989, pp. 393–402.

21. ———, *Proto-differentiability of set-valued mappings and its applications in optimization*, Analyse Non Linéaire (H. Attouch, J. P. Aubin, F.H. Clarke, and I. Ekeland, eds.), Gauthier-Villars, 1989, pp. 449–482.

22. ———, *Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives*, Mathematics of Operations Research 14 (1989), pp. 462–484.

23. ______, *Generalized second derivatives of convex functions and saddle functions*, Transactions of the American Mathematical Society 322 (1990), pp. 51-77.

24. ______, *Nonsmooth analysis and parametric optimization*, Methods of Nonconvex Analysis (A. Cellina, ed.), Lecture Notes in Mathematics, vol. 1446, Springer-Verlag, 1990, pp. 137–151.

25. ______, *Lagrange multipliers and optimality*, SIAM Review 35 (1993), pp. 183–238.