Ample Parameterization of Variational Inclusions*

A. L. Dontchev[1] and R. T. Rockafellar[2]

Abstract. For a general category of variational inclusions in finite dimensions, a class of parameterizations, called "ample" parameterizations, is identified that is rich enough to provide a full theory of Lipschitz-type properties of solution mappings without the need for resorting to the auxiliary introduction of canonical parameters. Ample parameterizations also support a detailed description of the graphical geometry that underlies generalized differentiation of solutions mappings. A theorem on proto-derivatives is thereby obtained. The case of a variational inequality over a polyhedral convex set is given special treatment along with an application to minimizing a parameterized function over such a set.

Keywords. Variational inequalities, calmness, Aubin continuity, Lipschitzian localizations, graphical derivatives, sensitivity of minimizers, variational analysis.

Revised version, January 2001

---

* This research was undertaken under grant DMS–9803089 from the National Science Foundation.

[1] Mathematical Reviews, Ann Arbor, MI 48107-8604.

[2] Department of Mathematics, University of Washington, Seattle, WA 98195.

1. Introduction

This paper is concerned with implicit-function-like results for parameterized variational inclusions (generalized equations) of the broad form

$$f(w,x)+F(x)\ \ni\ 0, \tag{1.1}$$

where $w\in I\!R^d$ is the parameter, $x\in I\!R^n$ is the solution, $f: I\!R^d\times I\!R^n\to I\!R^m$ is a smooth (i.e., $\mathcal{C}^1$) function, and $F: I\!R^n \rightrightarrows I\!R^m$ is a set-valued mapping with closed graph. The focus is on local properties of the solution mapping

$$S: w\mapsto S(w)=\big\{x\,\big|\,f(w,x)+F(x)\ \ni\ 0\big\} \tag{1.2}$$

at a pair (w_*,x_*) with $x_*\in S(w_*)$. We investigate Lipschitz-type properties such as calmness, Aubin continuity, and Lipschitzian localization, as well as graphical properties connected with generalized differentiation.

It is well understood that in order to make progress in this area the parameterization has to be "rich enough." A standard technique for ensuring such richness is to introduce explicitly, alongside of w, the so-called *canonical parameters* y that correspond to perturbing the right side in (1.1) to

$$f(w,x)+F(x)\ \ni\ y, \tag{1.3}$$

and then to work with extended mapping $\widetilde{S}: I\!R^d\times I\!R^m\to I\!R^n$ given by

$$\widetilde{S}:(w,y)\mapsto \widetilde{S}(w,y)=\big\{x\,\big|\,f(w,x)+F(x)\ \ni\ y\big\}. \tag{1.4}$$

Results obtained for $\widetilde{S}$ can be specialized to S by taking $y=0$. That approach seems inefficient, though, since the extended inclusion in (1.3) could also be written like (1.1):

$$\tilde{f}(\tilde{w},x)+F(x)\ \ni\ 0,\quad \text{where } \tilde{w}=(w,y) \text{ and } \tilde{f}(\tilde{w},x)=f(w,x)-y. \tag{1.5}$$

It would be preferable to capture the needed richness of the parameterization through an assumption on (1.1) itself, moreover in a manner that provides more flexibility by being merely local. We accomplish that here through the following concept.

Definition 1.1 (ample parameterization)**.** *The variational inclusion (1.1) will be called amply parameterized at a pair* $(w_*,x_*)\in \operatorname{gph} S$ *if the partial Jacobian matrix* $\nabla_w f(w_*,x_*)$ *for* f *with respect to* w *at* (w_*,x_*) *has full rank:*

$$\operatorname{rank}\nabla_w f(w_*,x_*)=m,\qquad \nabla_w f(w_*,x_*)\in I\!R^{m\times d}. \tag{1.6}$$

Obviously this condition is fulfilled at every point $(\tilde{w}_*, x_*) = (w_*, y_*, x_*)$ in the graph of the extended mapping $\tilde{S}$ in (1.4), viewed as in (1.5). Hence ample parameterization can always be enforced by passing from S to $\widetilde{S}$, in confirmation of the standard technique.

Supplied with this concept, we begin by studying the relationship between S and an auxiliary mapping S_* at (w_*, x_*) of the general type

$$S_* : y \mapsto S_*(y) = \{x \mid f_*(x) + F(x) \ni y\}, \tag{1.7}$$

where f_* denotes any (smooth) *first-order approximation* to $f(w_*, \cdot)$ at x_* in the sense that

$$f_*(x_*) = f(w_*, x_*) \quad \text{and} \quad \nabla f_*(x_*) = \nabla_x f(w_*, x_*). \tag{1.8}$$

Among the prime candidates for f_* are the simple restriction $f_*(x) = f(w_*, x)$ or its linearization $f_*(x) = f(w_*, x_*) + \nabla_x f(w_*, x_*)(x - x_*)$. Our results, however, depend only on the assumption in (1.7) that (1.8) holds, so in stating them in terms of S_* we achieve a more efficient presentation which emphasizes what is truly essential.

Note that S_* can itself be viewed as a solution mapping in this context, namely one in which there is only a canonical parameterization. Indeed, the choice $f_*(x) = f(w_*, x)$ corresponds to $S_*(y) = \widetilde{S}(w_*, y)$. In comparing properties of S and S_* we continue a long tradition coming from the classical implicit function theorem, where $F = 0$ and the mapping $w \mapsto \{x \mid f(w, x) = 0\}$ is compared to the mapping $y \mapsto \{x \mid f(w_*, x) = y\}$ or its linearization. Our contribution is to develop the comparison definitively not just for one, but for several key properties in our general setting, while employing the concept of ample parameterization to achieve statements that are more succinct and convenient.

Sections 2, 3 and 4 follow this pattern for the properties of calmness, Aubin continuity and Lipschitzian localization, respectively. In each case, under ample parameterization, the property in question holds for S if and only if it holds for S_*. Even without ample parameterization, if the property holds for S_* it must hold for S as well.

In Section 5 we show, again under ample parameterization, that S is graphically Lipschitzian if and only if F is graphically Lipschitzian. Furthermore, we demonstrate in Section 6 that such equivalence carries over to proto-differentiability of S versus that of F, and we obtain a corresponding formula for the proto-derivatives, which reveals that they are given as solutions to an auxiliary variational inclusion.

In Section 7 we specialize to the case of F being the normal cone mapping N_C to a convex set C; that is, the case where (1.1) is a *variational inequality*. We take advantage of the fact that N_C is then graphically Lipschitzian, and when C is polyhedral, N_C is proto-differentiable. From the resulting formula for proto-derivatives, we show that when the derivative mapping is convex-valued the proto-differentiability turns into the stronger property of semi-differentiability.

Finally, in Section 8 we apply our results to an optimization problem with perturbations only in the cost function. We show that the standard second-order sufficient optimality condition is equivalent to the combination of optimality at the reference point and calmness of the stationary point mapping. Moreover the strong second-order sufficient condition is equivalent to the Lipschitzian localization property of the mapping that gives local minimizers. A formula for semi-derivatives of this mapping is also provided.

A separate paper [6] is devoted to applications of these results to the perturbation of saddle points in convex optimization.

Throughout, any norm is denoted by $\|\cdot\|$ and $B_a(x)$ is the closed ball of radius a centered at x. The graph of a set-valued mapping $\Gamma : I\!\!R^p \rightrightarrows I\!\!R^n$ is the set $\operatorname{gph}\Gamma = \big\{(z,x) \in I\!\!R^p \times I\!\!R^n \,\big|\, x \in \Gamma(z)\big\}$ and the inverse of Γ is $\Gamma^{-1} : x \mapsto \big\{z \in I\!\!R^p \,\big|\, x \in \Gamma(z)\big\}$.

2. Calmness

To start, we consider a graphically localized version of the "upper-Lipschitz continuity" property introduced for set-valued mappings by Robinson [21]. For functions, the property goes back earlier to Clarke [1], who called it "calmness," and that is the term we prefer here in line with the recent book [23].

Definition 2.1 (calmness). *A mapping $\Gamma : I\!\!R^p \rightrightarrows I\!\!R^n$ is said to be calm at z_* for isolated x_* when $(z_*, x_*) \in \operatorname{gph}\Gamma$ and there exist neighborhoods U of x_* and V of z_* along with a constant γ such that*

$$\|x - x_*\| \le \gamma\|z - z_*\| \text{ for all } z \in V \text{ and } x \in \Gamma(z) \cap U.$$

This condition implies that $\Gamma(z_*) \cap U = \{x_*\}$, so x_* is an isolated point of $\Gamma(z_*)$, hence the terminology; but calmness can also be defined in a broader sense which reduces to the present one when x_* is isolated point, yet has meaning even when x_* is not isolated (cf. [23; p. 399]). The broader concept will not enter here. For single-valued mappings, there is no difference.

The calmness in Definition 2.1 was formally introduced by Dontchev [3] as the "local upper-Lipschitz property at a point in the graph" of a mapping. Earlier, without giving it a name, Rockafellar [22] characterized it in terms of the graphical derivatives of the set-valued mapping. That result will be applied in Section 6. For recent studies of calmness in the context of mathematical programming, see Klatte [9] and Levy [12].

The following theorem for variational inclusions furnishes a general result of implicit function type for the calmness property.

Theorem 2.2 (criterion for calmness)**.** *The mapping S is calm at w_* for isolated x_* when the mapping S_* is calm at 0 for isolated x_*. Under the ample parameterization condition (1.6), moreover, the two assertions are equivalent.*

We will deduce Theorem 2.2 from another result which we state next.

Theorem 2.3 (calmness in composition)**.** *Consider a mapping $N : \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ of the form $N(w) = \{x \mid x \in M(h(w,x))\}$ where $M : \mathbb{R}^p \rightrightarrows \mathbb{R}^n$ is set-valued and $h : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^p$ is $\mathcal{C}^1$. Let (w_*, x_*) be such that $\nabla_x h(w_*, x_*) = 0$. If M is calm at $z_* = h(w_*, x_*)$ for isolated x_*, then N is calm at w_* for isolated x_*.*

Proof. First, since $\nabla_x h(w_*, x_*) = 0$, we know that for any real $\lambda > 0$ and neighborhoods W of w_*, U of x_* and V of z_* there exist positive reals a, b and c such that the balls $B_a(w_*)$, $B_b(x_*)$, and $B_c(z_*)$ are contained in W, U and V, respectively, and for any fixed $w \in B_a(w_*)$ the function $x \mapsto h(w,x)$ is Lipschitz continuous on $B_b(x_*)$ with a Lipschitz constant λ. Of course, the radii a and b can be chosen arbitrarily small, and then c can be made arbitrary small as well, independently of the initial choice of λ. Let κ be an associated Lipschitz constant of the function $w \mapsto h(w,x)$ on $B_a(w_*)$, independent of $x \in B_b(x_*)$.

Suppose M is calm at z_* for isolated x_* with neighborhoods V' of z_* and U' of x_* and constant γ. Choose

$$0 < \lambda < 1/\gamma. \tag{2.1}$$

By the property of h just mentioned, there exist a, b and c such that $B_b(x_*) \subset U'$ and $B_c(z_*) \subset V'$, and moreover with the property that for any $w \in B_a(w_*)$ the function $h(w, \cdot)$ is Lipschitz continuous on $B_b(x_*)$ with a Lipschitz constant λ. Choose a and b smaller if necessary so that

$$\lambda a + \kappa b \le c. \tag{2.2}$$

Let $w \in B_a(w_*)$ and $x \in N(w) \cap B_b(x_*)$. Then $x \in M(h(w,x)) \cap B_b(x_*)$. Using (2.2) we have $\|h(w,x) - z_*\| = \|h(w,x) - h(w_*, x_*)\| \le \lambda a + \kappa b \le c$. From the calmness of M we then have $\|x - x_*\| \le \gamma \|h(w,x) - z_*\| \le \gamma\lambda\|x - x_*\| + \gamma\kappa\|w - w_*\|$, hence

$$\|x - x_*\| \le \frac{\gamma\kappa}{1 - \lambda\kappa}\|w - w_*\|.$$

This establishes that the mapping N is calm at w_* for x_* with constant $\gamma\kappa/(1 - \lambda\kappa)$. ◻

Theorem 2.3 is a purely metric result and can be formulated in terms only of the constants involved. Accordingly, there is no real need to have $\nabla_x h(w_*, x_*) = 0$ or even to have h be differentiable. All that is required, as seen through the proof, is for h to be

Lipschitz continuous in x with a "sufficiently small" Lipschitz constant. In fact the result can be stated in a context of metric spaces.

In the proof of Theorem 2.2, still ahead, we will also employ the following lemma, where the classical implicit function theorem comes in.

Lemma 2.4 (reparameterization). *Under the ample parameterization condition (1.6), and for a function f_* satisfying the condition (1.8), there exist neighborhoods U, V and W of x_*, $y = 0$ and w_*, respectively, and a $\mathcal{C}^1$ function $\omega : U \times V \to W$ such that*

(i) $y + f(\omega(x, y), x) = f_*(x)$ *for every* $y \in V$ *and* $x \in U$,

(ii) $\omega(x_*, 0) = w_*$ *and* $\nabla_x \omega(x_*, 0) = 0$.

Proof. Let $B := \nabla_w f(w_*, x_*)$; by assumption, this matrix in $I\!R^{m\times d}$ has full row rank m. In terms of the transpose $B^\top$, consider the system of equations

$$\begin{aligned} w - w_* + B^\top z &= 0 \\ y + f(w, x) - f_*(x) &= 0, \end{aligned} \tag{2.3}$$

where (w, z) is the variable and (x, y) is the parameter. Clearly, $(w_*, 0)$ is a solution of (2.3) for the parameter choice $(x_*, 0)$. The Jacobian J at $(w_*, 0, x_*, 0)$ of the function of (w, z) on left side of (2.3) has the form

$$J = \begin{bmatrix} I & B^\top \\ B & 0 \end{bmatrix},$$

where I is the identity. It is well known that when B has full row rank the matrix J is nonsingular. Hence, from the classical implicit function theorem, we conclude that, locally around $(w_*, 0, x_*, 0)$, there exists a $\mathcal{C}^1$ function $\Omega : (x, y) \mapsto (\omega(x, y), \zeta(x, y))$ such that

$$\begin{aligned} \omega(x, y) - w_* + B^\top \zeta(x, y) &= 0 \\ y + f(\omega(x, y), x) - f_*(x) &= 0 \end{aligned} \tag{2.4}$$

with $\Omega(x_*, 0) = (w_*, 0)$. This yields (i) and the first condition in (ii). By differentiating the system we see further that $J\nabla_x \Omega(x, y)$ must vanish locally, and since J is nonsingular this implies that $\nabla_x \Omega(x, y)$ vanishes locally. In particular, then, $\nabla_x \omega(x_*, 0) = 0$. □

Proof of Theorem 2.2. From the definitions of S and S_* in (1.2) and (1.7) we have $x \in S(w)$ if and only if $x \in S_*(y)$ for $y = f_*(x) - f(w, x)$. Thus, we can write

$$S(w) = \{x \mid x \in S_*(f_*(x) - f(w, x))\}. \tag{2.5}$$

By taking $h(w,x) = f_*(x) - f(w,x)$, which has $\nabla_x h(w_*,x_*) = 0$ by virtue of (1.8), we can put this in the framework of Theorem 2.3 with $M = S_*$. This lets us conclude that calmness of S_* implies calmness of S.

Assume now that the ample parameterization condition (1.6) holds and consider a mapping ω as guaranteed in Lemma 2.4 with respect to certain neighborhoods U, V and W. Fix $y \in V$. If $x \in S_*(y) \cap U$ and $w = \omega(x,y)$, then $w \in W$ and $y + f(w,x) = f_*(x)$, hence $x \in S(w) \cap U$. Conversely, if $x \in S(w(x,y)) \cap U$, then clearly $x \in S_*(y) \cap U$. Thus,

$$S_*(y) \cap U = \{x \mid x \in S(\omega(x,y)) \cap U\}. \tag{2.6}$$

Since calmness of S at w_* for isolated x_* is local property of the graph of S relative to the point (w_*,x_*), this holds if and only if the same holds for the truncated mapping $S_U : w \mapsto S(w) \cap U$. That equivalence is valid for S_* as well. Applying Theorem 2.3 now in the context of (2.6) with $h = \omega$, we get the desired equivalence for S versus S_*. □

3. Aubin property

The idea behind the Aubin property, which Aubin called "pseudo-Lipschitz continuity," can be traced back to the original proofs of the Lyusternik and Graves theorems; see [2], [4], [7], [11] and [23] for discussions. This property is known to correspond, with respect to taking inverses of mappings, to "metric regularity," a condition which plays a major role in optimization.

Definition 3.1 (Aubin property). *A mapping $\Gamma : I\!R^p \rightrightarrows I\!R^n$ is said to have the Aubin continuity property at z_* for x_* when $(z_*,x_*) \in \operatorname{gph}\Gamma$ and there exist neighborhoods U of x_* and V of z_* along with a constant γ such that*

$$z', z'' \in V,\ x' \in \Gamma(z') \cap U \quad \Longrightarrow \quad \exists\, x'' \in \Gamma(z'') \text{ with } \|x' - x''\| \le \gamma \|z' - z''\|.$$

Keeping the pattern of the preceding section, we establish a result about the Aubin property that is completely parallel to the one about calmness in the preceding section.

Theorem 3.2 (criterion for Aubin property). *The mapping S has the Aubin property at w_* for x_* when the mapping S_* has the Aubin property at 0 for x_*. Under the ample parameterization condition (1.6), moreover, the two assertions are equivalent.*

Not only is the statement of Theorem 3.2 completely parallel to that of Theorem 2.2, the proofs are parallel as well. The key is a composition rule that can be regarded as a version of the Lyusternik-Graves theorem.

Theorem 3.3 (Aubin property in composition)**.** *Consider a mapping* $N : I\!R^d \rightrightarrows I\!R^n$ *of the form* $N(w) = \{x \mid x \in M(h(w,x))\}$ *where* $M : I\!R^p \rightrightarrows I\!R^n$ *is set-valued with closed graph and* $h : I\!R^d \times I\!R^n \to I\!R^p$ *is* $\mathcal{C}^1$*. Let* (w_*, x_*) *be such that* $\nabla_x h(w_*, x_*) = 0$*. If* M *has the Aubin property at* $z_* = h(w_*, x_*)$ *for* x_**, then* N *has the Aubin property at* w_* *for* x_**.*

Proof. Let the mapping M have the Aubin property at z_* for x_* with neighborhoods V' of z_* and U' of x_* and a constant γ. Let λ satisfy (2.1) and choose the constants a, b and c as in the proof of Theorem 2.3. Choose a smaller if necessary so that

$$\frac{4\gamma\kappa a}{1-\gamma\lambda} \leq b. \tag{3.1}$$

Let $w', w'' \in B_a(w_*)$ and let $x' \in N(w') \cap B_{b/2}(x_*)$. Then $x' \in M(h(w',x')) \cap B_{b/2}(x_*)$. We get from the Aubin property of M the existence of $x_1 \in M(h(w'',x'))$ such that $\|x_1 - x'\| \leq \gamma \|h(w',x') - h(w'',x')\| \leq \gamma\kappa\|w' - w''\|$. Also, through (3.1),

$$\|x_1 - x_*\| \leq \|x_1 - x'\| + \|x' - x_*\| \leq \gamma\kappa\|w' - w''\| + \|x' - x_*\| \leq \gamma\kappa(2a) + \frac{b}{2} \leq b,$$

and consequently $\|h(w'',x_1) - z_*\| \leq \lambda a + \kappa b \leq c$, from (2.2). Hence, from the Aubin property of M, there exists $x_2 \in M(h(w'',x_1))$ such that

$$\|x_2 - x_1\| \leq \gamma\|h(w'',x_1) - h(w'',x')\| \leq \gamma\lambda\|x_1 - x'\| \leq (\gamma\lambda)\gamma\kappa\|w' - w''\|.$$

By induction, we obtain a sequence $x_1, x_2, \ldots, x_k, \ldots$ with $x_k \in M(h(w'',x_{k-1}))$ and $\|x_k - x_{k-1}\| \leq (\gamma\lambda)^{k-1}\gamma\kappa\|w' - w''\|$. Setting $x_0 = x'$ and using (3.1), we get

$$\begin{aligned} \|x_k - x_*\| &\leq \|x_0 - x_*\| + \sum_{j=1}^{k} \|x_j - x_{j-1}\| \\ &\leq \frac{b}{2} + \sum_{j=0}^{k-1} (\gamma\lambda)^j \gamma\kappa\|w' - w''\| \leq \frac{b}{2} + \frac{2a\gamma\kappa}{1-\gamma\lambda} \leq b, \end{aligned}$$

hence $\|h(w'',x_k) - z_*\| \leq \lambda a + \kappa b \leq c$. Then there exists $x_{k+1} \in M(h(w'',x_k))$ such that $\|x_{k+1} - x_k\| \leq \gamma\|h(w'',x_k) - h(w'',x_{k-1})\| \leq \gamma\lambda\|x_k - x_{k-1}\| \leq (\gamma\lambda)^k\gamma\kappa\|w' - w''\|$, and the induction step is complete.

The sequence $\{x_k\}$ is Cauchy, hence convergent to some $x'' \in B_a(x_*) \subset U'$. From the closedness of $\operatorname{gph} M$ that has been assumed, and the continuity of h we deduce that $x'' \in M(h(w'',x'')) \cap U'$, hence, $x'' \in N(w'')$. Furthermore, using the estimate

$$\|x_k - x'\| \leq \sum_{j=1}^{k} \|x_j - x_{j-1}\| \leq \sum_{j=0}^{k-1} (\gamma\lambda)^j \gamma\kappa\|w' - w''\| \leq \frac{\gamma\kappa}{1-\gamma\lambda}\|w' - w''\|$$

we obtain, on passing to the limit with respect to $k \to \infty$, that $\|x'' - x'\| \le \gamma' \|w' - w''\|$. Thus, N has the Aubin property at 0 for x_* with constant $\gamma' = (\gamma\kappa)/(1 - \gamma\lambda)$. □

Proof of Theorem 3.2. Repeat the argument in the proof of Theorem 2.2, simply replacing the composition rule in Theorem 2.3 by the one in Theorem 3.3. □

4. Lipschitzian localization

The Lipschitzian localization property is a looser form of the smooth localization property that appears in the classical implicit function theorem. In the context of variational inequalities, Lipschitzian localization is the property in Robinson's "strong regularity" theorem [20]; see [10], [11], [17] and [23] for more on this subject.

Definition 4.1 (Lipschitzian localization)**.** *A mapping $\Gamma : I\!R^p \rightrightarrows I\!R^n$ is said to have a single-valued Lipschitzian localization at z_* for x_* when $(z_*, x_*) \in \operatorname{gph} \Gamma$ and there exist neighborhoods U of x_* and V of z_* such that the mapping $V \ni z \mapsto \Gamma(z) \cap U$ is single-valued and Lipschitz continuous.*

For this property we have an analog of Theorems 2.2 and 3.2 in the following mode.

Theorem 4.2 (criterion for Lipschitzian localization)**.** *The mapping S has a single-valued Lipschitzian localization at w_* for x_* when the mapping S_* has a single-valued Lipschitzian localization at 0 for x_*. Under the ample parameterization condition (1.6), moreover, the two assertions are equivalent.*

Again we establish this by way of a composition rule.

Theorem 4.3 (Lipschitzian localization in composition)**.** *Consider $N : I\!R^d \rightrightarrows I\!R^n$ of the form $N(w) = \{x \mid x \in M(h(w,x))\}$ where $M : I\!R^p \rightrightarrows I\!R^n$ is set-valued mapping with closed graph and $h : I\!R^d \times I\!R^n \to I\!R^p$ is $\mathcal{C}^1$. Let (w_*, x_*) be such that $\nabla_x h(w_*, x_*) = 0$. If M has a single-valued Lipschitzian localization at $z_* = h(w_*, x_*)$ for x_*, then N has a single-valued Lipschitzian localization at w_* for x_*.*

Proof. Suppose M has a single-valued Lipschitzian localization at z_* for x_* with neighborhoods U and V and a constant γ. In particular then, M has the Aubin property at z_* for x_* with the same constant γ and consequently, as already proved, N has the Aubin property at w_* for x_*. It is sufficient therefore to verify that there exist neighborhoods U' of x_* and W' of w_* such that $N(w) \cap U'$ is a singleton for every $w \in W'$.

Observe that we can choose a neighborhood W of w_* and shrink U if necessary so that the Lipschitz constant λ of the function $h(w, \cdot)$ on U works for any $w \in W$. Suppose that there exist two sequences, x_k^1 and x_k^2, converging to x_* and a sequence w_k converging to w_*, such that $x_k^i \in N(w_k)$, $i = 1, 2$, and $x_k^1 \neq x_k^2$ for a sufficiently large k so that $x_k^i \in U$, $w_k \in W$, and $h(w_k, x_k^i) \in V$. Since $M(h(w_k, x_k^i)) \cap U$ is a singleton for large k, we have $x_k^i = M(h(w_k, x_k^i)) \cap U$, $i = 1, 2$. From the Lipschitz continuity of both $M(\cdot) \cap U$ and $h(w_k, \cdot)$ we finally obtain

$$0 \neq \|x_k^1 - x_k^2\| \leq \gamma \|h(w_k, x_k^1) - h(w_k, x_k^2)\| \leq \gamma\lambda \|x_k^1 - x_k^2\| < \|x_k^1 - x_k^2\|.$$

This contradiction demonstrates that N has the property claimed. □

Proof of Theorem 4.2. Repeat the argument in the proof of Theorem 2.2, simply replacing the composition rule in Theorem 2.3 by the one in Theorem 4.3. □

5. Lipschitzian graphical geometry

Beyond the property of Lipschitzian localization treated in Section 4, there is a more subtle kind of Lipschitzian behavior which is especially common for solution mappings without single-valuedness, but which, unlike the Aubin property of Section 3 or even the calmness property of Section 2, does not revolve around comparing values of the mapping at two different points. Instead, this property centers on Lipschitzian geometry of the graph of the mapping. It has strong implications for generalized differentiability.

Definition 5.1 (graphically Lipschitzian mappings)**.** *A mapping $\Gamma : I\!R^p \rightrightarrows I\!R^n$ is said to be graphically Lipschitzian at z_* for x_*, and of dimension k in this respect, when $(z_*, x_*) \in \operatorname{gph}\Gamma$ and there is a change of coordinates in $I\!R^p \times I\!R^n$ around (z_*, x_*) that is $\mathcal{C}^1$ in both directions, under which $\operatorname{gph}\Gamma$ can be identified locally with the graph in $I\!R^k \times I\!R^{p+n-k}$ of a Lipschitz continuous mapping defined around a point $u_* \in I\!R^k$.*

Background on graphically Lipschitzian mappings can be found in [23]. As a special case, of course, if Γ has a single-valued Lipschitzian localization around $z_* \in I\!R^p$, then Γ is graphically Lipschitzian of dimension p at z_* for $x_* = \Gamma(z_*)$. The point of Definition 5.1, however, is that many mappings of fundamental interest in variational analysis and optimization can fail to be single-valued and Lipschitz continuous and yet possess hidden properties of Lipschitzian character which deserve to be recognized and placed in service.

An important class of graphically Lipschitzian mappings which by no means need to be single-valued and Lipschitz continuous is furnished by the *maximal monotone* mappings $F : I\!R^n \rightrightarrows I\!R^n$; the theory of maximal monotonicity is available in detail in Chapter 12 of [23]. Within this category are the normal cone mappings N_C associated with the nonempty, closed, convex sets C in $I\!R^n$ and more generally the subgradient mappings $\partial\varphi$ associated with the lower semicontinuous, proper, convex functions φ on $\mathbb{R}^n$. A normal cone mapping will be the focus in the next section. When $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone, $\operatorname{gph} F$ is in fact an n-dimensional Lipschitzian manifold in a global sense.

Maximal monotonicity is not the only source of examples. A broad class of normal cone mappings N_C and subgradient mappings $\partial\varphi$ for which the graphical Lipschitzian property prevails without C or φ having to be convex has been developed by Poliquin and Rockafellar [16] under the heading of "prox-regularity" and more specially "strong amenability" (see also 10.24 and 13.46 of [23]). Such sets C and functions φ arise very commonly in optimization. For instance, a set C given by finitely many $\mathcal{C}^2$ equality and inequality constraints is strongly amenable at any point satisfying the Mangasarian-Fromovitz constraint qualification; a function φ is sure to be strongly amenable when it is the sum of the indicator of a strongly amenable set and a function that is $\mathcal{C}^2$ or the max of finitely many $\mathcal{C}^2$ functions. The associated mappings N_C and $\partial\varphi$ then likewise furnish choices of F that are graphically Lipschitzian.

The next theorem shows that, under ample parameterization, graphically Lipschitzian properties of the solution mapping S can be derived from those of F by way of the natural correspondence between the graphs of these mappings:

$$(x, -f(w,x)) \in \operatorname{gph} F \quad \Longleftrightarrow \quad (w,x) \in \operatorname{gph} S. \tag{5.1}$$

Theorem 5.2 (criterion for Lipschitzian geometry)**.** *Under the ample parameterization condition (1.6), the mapping S is graphically Lipschitzian of dimension q at w_* for x_* if and only if the mapping F is graphically Lipschitzian of dimension k at x_* for y_*, where*

$$y_* = -f(w_*, x_*), \qquad q = k + d - m.$$

Proof. Define $Q : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^m$ by

$$Q(w,x) = (x, -f(w,x)). \tag{5.2}$$

Then from (5.1), $\operatorname{gph} S = Q^{-1}(\operatorname{gph} F)$. Under the ample parameterization condition the Jacobian $\nabla Q(w_*, x_*)$ of Q at (w_*, x_*) has full rank $n+m$; in particular this requires $d + n \geq n + m$, i.e., $d - m \geq 0$. Therefore, with respect to a neighborhood O of (w_*, x_*), Q^{-1} has the effect of transforming any graphically Lipschitzian manifold of dimension k in $\mathbb{R}^n \times \mathbb{R}^m$ into one of dimension $k + (d-m)$ in $\mathbb{R}^d \times \mathbb{R}^n$. The equivalence is now immediate. □

Corollary 5.3 (maximal monotonicity). *Under the ample parameterization condition, if F is a maximal monotone mapping, $F : I\!R^n \rightrightarrows I\!R^n$, then S is graphically Lipschitzian of dimension d at w_* for x_*.*

Proof. When F is maximal monotone, it is everywhere graphically Lipschitzian of dimension n (cf. 12.15 of [23]). Then, by virtue of Theorem 5.2, S is graphically Lipschitzian of dimension $n + d - n = d$ at w_* for x_*. □

Corollary 5.4 (strong amenability). *Under the ample parameterization condition, if F is a normal cone mapping N_C or subgradient mapping $\partial\varphi$ for a set C or function φ that is strongly amenable at x_*, then S is graphically Lipschitzian of dimension d at w_* for x_*.*

Proof. Here we rely on the graphically Lipschitzian behavior of such normal cone mappings and subgradient mappings as noted prior to the statement of Theorem 5.2. □

In order to tie Theorem 5.2 in with the patterns of equivalence in the preceding sections, it is worth stating also the following elementary consequence.

Corollary 5.5 (equivalent geometries in approximation). *The mapping S_* is graphically Lipschitzian of dimension k at 0 for x_* if and only if F is graphically Lipschitzian of dimension k at x_* for y_*, where $y_* = -f_*(x_*)$. Thus, under the ample parameterization condition (1.6), S is graphically Lipschitzian of dimension q at w_* for x_* if and only if S_* is graphically Lipschitzian of dimension k at 0 for x_*, where $q = k + d - m$.*

Proof. Theorem 5.2 can be applied to S_* as a special kind of solution mapping, which corresponds to replacing $f(w, x)$ by $g(y, x) = f_*(x) - y$ with y as the new parameter, in $I\!R^m$ instead of $I\!R^d$. For g, the condition of ample parameterization is satisfied trivially at $(0, x_*)$. Moreover, $-g(0, x_*) = -f(w_*, x_*) = y_*$. Therefore, S_* is graphically Lipschitzian of dimension q_* at 0 for x_* if and only if F is graphically Lipschitzian of dimension k at x_* for y_*, the relation between q_* and k being like that between q and k in Theorem 5.2, except that d is replaced by m. Then $q_* = k + m - m = k$.

In combination now with the statement about S and F in Theorem 5.2, this observation yields the claimed relationship between S and S_*. □

6. Generalized differentiation

In the graphical context of Theorem 5.2, there is a powerful geometric notion of generalized differentiation which can be used even though S may only be set-valued. One says that S is *proto-differentiable at* w_* *for* x_* when $x_* \in S(w_*)$ and the difference quotient mappings

$$\Delta_\tau S(w_* \,|\, x_*) : w' \;\mapsto\; \tau^{-1}[S(w_* + \tau w') - x_*], \quad \tau > 0,$$

converge graphically as $\tau \searrow 0$; in other words, there is a mapping $D : I\!R^d \to I\!R^n \times I\!R^m$ such that $\operatorname{gph} \Delta_\tau S(w_* \,|\, x_*)$ converges to $\operatorname{gph} D$ as $\tau \to 0$. Proto-differentiability was introduced in [22], and much about it can be found now also in [23]; see [13] and [14] as well, where special properties in the case of a graphically Lipschitzian mapping are laid out.

Proto-differentiability is closely involved with the tangent cone $T_{\operatorname{gph} S}(w_*, x_*)$ to $\operatorname{gph} S$ at (w_*, x_*). This cone is the graph of the mapping $DS(w_* \,|\, x_*) : I\!R^d \rightrightarrows I\!R^n$ that in general is called the *graphical derivative* of S at w_* for x_*; by definition,

$$x' \in DS(w_* \,|\, x_*)(w') \quad \Longleftrightarrow \quad (w', x') \in T_{\operatorname{gph} S}(w_*, x_*). \tag{6.1}$$

The graphs of the mappings $\Delta_\tau S(w_* \,|\, x_*)$ are the sets $\tau^{-1}[\operatorname{gph} S - (w_*, x_*)]$, which have $T_{\operatorname{gph} S}(w_*, x_*)$ as their outer set limit ("lim sup") as $\tau \searrow 0$. What makes the property of proto-differentiability special is that the outer limit is required to equal the inner set limit ("lim inf") and thus be a true set limit. As translated to the language of tangent cones, proto-differentiability of S at w_* for x_* means that $\operatorname{gph} S$ is *geometrically derivable* at (w_*, x_*). See [23] for more on this subject. Anyway, it is clear that when the graphical limit D in the definition of proto-differentiability exists it has to be $DS(w_* \,|\, x_*)$, although the latter has meaning (and uses) even in the absence of proto-differentiability.

The power of proto-differentiability in the presence of Lipschitzian graphical geometry comes from the tight mode of local approximation it affords, in a manner reminiscent that of classical differentiability. To appreciate this, consider first the case where S happens to be single-valued and Lipschitz continuous around w_*, with x_* the unique element of $S(w_*)$. Proto-differentiability implies then that the mapping $DS(w_* \,|\, x_*)$ (which in this case could simply be denoted by $DS(w_*)$) is likewise single-valued and Lipschitz continuous and

$$S(w) = S(w_*) + DS(w_* \,|\, x_*)(w - w_*) + o(|w - w_*|), \tag{6.2}$$

where $o(t)$ denotes a term such that $o(t)/t \to 0$ as $t \searrow 0$. This is ordinary differentiability precisely when the mapping $DS(w_* \,|\, x_*)$ is, in addition, linear.

In general when S and $DS(w_* \,|\, x_*)$ are single-valued (but $DS(w_* \,|\, x_*)$ might not be linear), we speak of the property in (6.2) as the *semi-differentiability* of S at w_* for $x_* = S(w_*)$. For more discussion of semi-differentiability, see [23].

In moving next to the case where S is not necessarily single-valued and Lipschitz continuous but merely graphically Lipschitzian at (w_*, x_*), it is crucial to observe that although the type of approximation in (6.2) depends strongly on the particular coordinate system on the graph, specifically the decomposition into components w and x, the notion of proto-differentiability does not. Because it is based on set convergence in the graph space, proto-differentiability is preserved under changes of coordinates. Therefore, *proto-differentiability of a graphically Lipschitzian mapping S corresponds to the tight mode of local approximation to* $\operatorname{gph} S$ *as in (6.2), but applied obliquely, to a different coordinate system than the* (w, x) *system.*

Note that the mapping $DS(w_* \,|\, x_*)$ anyway is always *positively homogeneous*, since its graph is a cone; one has $DS(w_* \,|\, x_*)(0) \ni 0$ and $DS(w_* \,|\, x_*)(\lambda w') = \lambda DS(w_* \,|\, x_*)(w')$ for all w' when $\lambda > 0$.

Proto-differentiability has only been described so far in terms of S, but of course the concept also applies to F, and this now comes on stage as well. For a pair $(x_*, y_*) \in \operatorname{gph} F$ we have

$$y' \in DF(x_*, y_*)(x') \iff (x', y') \in T_{\operatorname{gph} F}(x_*, y_*). \tag{6.3}$$

If F happens for example to be single-valued and, at x_*, is differentiable in the usual sense, then F is proto-differentiable at x_* for $y_* = F(x_*)$ with $DF(x_* \,|\, y_*)$ being the usual derivative mapping (for which $DF(x_*)$ is then a simpler notation).

Theorem 6.1 (proto-derivative formula)**.** *Under the ample parameterization condition (1.6), the mapping S is proto-differentiable at w_* for x_* if and only if the mapping F is proto-differentiable at x_* for $y_* = -f(w_*, x_*)$. Then*

$$\begin{aligned} &DS(w_* \,|\, x_*)(w') = \big\{ x' \,\big|\, g(w', x') + G(x') \ni 0 \big\}, \text{ where} \\ &\quad g(w', x') = \nabla_w f(w_*, x_*) w' + \nabla_x f(w_*, x_*) x' \text{ and } G(x') = DF(x_* \,|\, y_*)(x'). \end{aligned} \tag{6.4}$$

Proof. We appeal again to the setup in the proof of Theorem 5.2, where $\operatorname{gph} S = Q^{-1}(\operatorname{gph} F)$ for the mapping Q in (5.2). Because the Jacobian of Q has full rank under ample parameterization, we can determine the tangent cone $T_{\operatorname{gph} S}(w_*, x_*)$ by the general rule of variational analysis given in 6.7 of [23], obtaining

$$T_{\operatorname{gph} S}(w_*, x_*) = \big\{ (w', x') \,\big|\, \nabla Q(w_*, x_*)(w', x') \in T_{\operatorname{gph} F}(x_*, y_*) \big\}. \tag{6.5}$$

This furnishes through the formulas for $DS(w_*, x_*)$ and $DF(x_*, y_*)$ in (6.1) and (6.3) the formula in (6.4). A parallel formula holds for the corresponding "derivable cones" to $\operatorname{gph} S$ and $\operatorname{gph} F$, which are defined with outer set limits replaced by inner set limits. The geometric derivability of gph F at (x_*, y_*) thus corresponds to the geometric derivability of gph S at (w_*, x_*). Hence we have the equivalence between proto-differentiability of S and that of F. □

Corollary 6.2 (derivative criterion for calmness). *Under the ample parameterization condition (1.6) and the assumption that F is proto-differentiable at x_* for $y_* = -f(w_*, x_*)$, the mapping S is calm at w_* for isolated x_* if and only if*

$$\nabla_x f(w_*, x_*)x' + DF(x_* \,|\, y_*)(x') \ni 0 \quad \Longrightarrow \quad x' = 0. \tag{6.6}$$

Proof. According to the characterization of calmness of set-valued mappings developed in [22; Theorem 4.1] in terms of graphical derivatives, S is calm at w_* for isolated x_* if and only if $DS(w_* \,|\, x_*)(0) = \{0\}$. This criterion translates to (6.6) through the derivative formula in Theorem 6.1. □

The especially attractive feature of Theorem 6.1 is that the graphical derivative of the solution mapping S turns out itself to be a solution mapping in our framework, namely one that corresponds to g and G in place of f and F, with w' as the parameter and x' as the solution. A derivative formula in this pattern was originally exhibited in [22] for a variational inequality with canonical perturbations. That case will be elaborated below.

To make the best use of Theorem 6.1 and Corollary 6.2, one needs to recognize situations where F is proto-differentiable. The example of F single-valued and differentiable has already been mentioned. Other examples emerge from the second-order variational analysis of sets and functions that are *fully amenable*, this being a refinement of the strong amenability in [16] that had a role in the preceding section. For the theory of full amenability and the graphical derivative formulas it provides, along with examples, we refer to [23] and restrict ourselves here to recording the following consequence of Theorem 6.1.

Corollary 6.3 (full amenability). *Under the ample parameterization condition (1.6), if $F = N_C$ or $F = \partial\varphi$ for a set C or function φ that is fully amenable at x_*, then S is not only graphically Lipschitzian at w_* for x_* but also proto-differentiable there.*

Proof. The graphically Lipschitzian property is implied by Corollary 5.4, inasmuch as full amenability is a special case of strong amenability. The rest comes out of Theorem 6.1 and the fact, just cited, that F is proto-differentiable at x_* for $y_* \in F(x_*)$ when F is of the form described. □

7. Application to variational inequalities

We concentrate now on the special case where S is the solution mapping for a parameterized variational inequality,

$$S(w) = \{x \mid f(w,x) + N_C(x) \ni 0\} \tag{7.1}$$

with respect to a nonempty convex set $C \subset I\!R^n$ that is *polyhedral*. This choice allows us to obtain a quite detailed picture of the geometry of proto-derivatives of S and to provide a basis for their actual computation. Because of convexity, the vectors y in the normal cone $N_C(x)$ at any $x \in C$ are the ones that satisfy

$$\langle y,\, x' - x\rangle \le 0 \quad \text{for all } x' \in C.$$

Typically in the literature on variational inequalities this condition, with $y = -f(w,x)$, is written in place of the condition $f(w,x) + N_C(x) \ni 0$, but the normal cone version helps to put things into the right framework of set-valued mappings. When $x \notin C$, $N_C(x)$ is interpreted as $\emptyset$.

Our goal is to apply the theory of the preceding sections to $F = N_C$ and make the most of the special properties that follow from C being polyhedral. We say that a mapping is *piecewise polyhedral* when its graph is the union of a collection of finitely many polyhedral (convex) sets. If the mapping is single-valued, this is the same as it being piecewise linear (see 2.48 of [23]). For a vector y, we let $y^\perp = \{u \mid \langle y, u\rangle = 0\}$. This notation is used in the next theorem in defining the cone K_* that is known as the *critical cone* associated with the variational inequality in (7.1) for $w = w_*$ and $x = x_*$.

Theorem 7.1 (proto-derivatives for variational inequalities). *Let $F = N_C$ for a polyhedral convex set $C \subset I\!R^n$ and assume that the ample parameterization condition (1.6) holds. Then S is both graphically Lipschitzian of dimension d and proto-differentiable at w_* for x_*, with its proto-derivatives given by an auxiliary variational inequality, namely*

$$\begin{aligned} &DS(w_* \,|\, x_*)(w') = \{x' \mid g(w',x') + N_{K_*}(x') \ni 0\}, \quad \textit{where} \\ &\quad g(w',x') = \nabla_w f(w_*,x_*)w' + \nabla_x f(w_*,x_*)x' \quad \textit{and} \quad K_* = T_C(x_*) \cap f(w_*,x_*)^\perp. \end{aligned} \tag{7.2}$$

Furthermore, the mapping $DS(w_* \,|\, x_*)$ is itself graphically Lipschitzian of dimension d everywhere and is piecewise polyhedral.

Proof. This mainly constitutes a further specialization of Theorems 5.2 and 6.1 along the lines of Corollaries 5.4 and 6.3. When $F = N_C$ with C polyhedral (and nonempty since by blanket assumption we are working with a pair $(w_*, x_*) \in \operatorname{gph} S$), we have F maximal monotone and everywhere proto-differentiable, with the proto-derivative mapping being itself a normal cone mapping; specifically, $DF(x_* \,|\, y_*) = N_{K_*}$ for $K_* = T_C(x_*) \cap y_*^{\perp}$, which we apply here to $y_* = -f(w_*, x_*)$. (This reduction of $DF(x_* \,|\, y_*)$ to a normal cone mapping depends crucially on C being polyhedral; for details see [21] or the Reduction Lemma in [5].)

Because the tangent cones to a polyhedral set C are themselves polyhedral, the cone K_* is polyhedral and the mapping N_{K_*} is therefore piecewise polyhedral (see [18] or 12.31 of [23]). Recall now the general way that the graph of S corresponded to that of F through a mapping Q as in (5.1) and (5.2). In the context of the auxiliary variational inequality in (7.2), the same holds for $\operatorname{gph} DS(w_* \,|\, x_*)$ versus $\operatorname{gph} N_{K_*}$, and furthermore with a replacement for Q that is a linear mapping. From this it is apparent that $\operatorname{gph} DS(w_* \,|\, x_*)$ inherits the piecewise polyhedrality of $\operatorname{gph} N_{K_*}$. □

A proto-derivative formula akin to the one in Theorem 7.1 was originally established in [22], but in terms of canonical parameters. Here we have extended it in terms of ample parameterization as well as provided new information about the graph of the derivative mapping, its piecewise polyhedrality.

Corollary 7.2 (piecewise linear geometry)**.** *In the setting of Theorem 7.1, the graph of $DS(w_* \,|\, x_*)$ is a piecewise linear manifold of dimension d in the sense of being a Lipschitzian manifold formed as the union of a finite collection of d-dimensional polyhedral sets.*

Proof. Theorem 7.1 reveals that $DS(w_* \,|\, x_*)$ is a mapping of the sort to which Corollary 5.2 applies. Hence $\operatorname{gph} DS(w_* \,|\, x_*)$ is a d-dimensional Lipschitzian manifold, in fact "globally" because this graph is a cone and therefore determined by its properties around the origin. On the other hand, $DS(w_* \,|\, x_*)$ is piecewise polyhedral by Theorem 6.1. That supplies the piecewise linearity of the Lipschitzian mapping underlying the definition of the graphically Lipschitzian property (cf. 12.31 of [23] again). In expressing the graph as the union of a finite collection of polyhedral sets, it can be arranged that none of these sets is included in any of the others, and they must then all be of dimension d. □

Corollary 7.3 (calmness of variational inequalities)**.** *In the setting of Theorem 7.1, the mapping S is calm at w_* for isolated x_* if and only if*

$$\nabla_x f(w_*, x_*)x' + N_{K_*}(x') \ni 0 \quad \Longrightarrow \quad x' = 0.$$

Proof. We get this immediately from Corollary 6.2. □

Especially of interest for proto-differentiability is the case of Theorem 7.1 where S is locally single-valued and Lipschitz continuous. When that holds, the proto-differentiability turns into a stronger property. A critical role in reaching that conclusion can be played by the result in Theorem 4.2, this being an extended version of Robinson's strong regularity theorem [19]. In other work which is closely related, King and Rockafellar in [8] obtained a graphical-derivative characterization of single-valuedness for set-valued mappings with a "subinvertibility" property which in particular can be guaranteed through monotonicity. The next theorem could largely be derived as a specialization of that work, but because of a difference in contexts we find it more expedient and illuminating to proceed directly.

Recall here the concept of *semi-differentiability* that was described for single-valued S and $DS(w_* \,|\, x_*)$ in terms of the approximation in (6.2).

Theorem 7.4 (single-valuedness relations). *Let $F = N_C$ for a polyhedral convex set $C \subset I\!R^n$ and assume that the ample parameterization condition (1.6) holds. Suppose further that S is convex-valued around w_*, in the sense that $S(w)$ is a convex set for all w in some neighborhood of w_*. Then the following properties are equivalent:*

(a) *S is single-valued and Lipschitz continuous on some neighborhood of w_*,*

(b) *$DS(w_* \,|\, x_*)$ is single-valued on some neighborhood of 0 (hence everywhere).*

Moreover, then S is semi-differentiable at w_* for x_*, and $DS(w_* \,|\, x_*)$ is not only Lipschitz continuous and positively homogeneous but also piecewise linear.

Proof. Since S is convex-valued, it is single-valued and Lipschitz continuous around w_* if and only it has a single-valued Lipschitzian localization at w_* for x_*. This is critical because this localization property is all that we are able to relate to $DS(w_* \,|\, x_*)$, inasmuch as $DS(w_* \,|\, x_*)$ depends only on the geometry of $\operatorname{gph} S$ at (w_*, x_*).

The proto-differentiability of S at w_* for x_*, which we know from Theorem 7.1, reduces to the semi-differentiability in (6.2) when S is locally single-valued and Lipschitz continuous, as noted earlier (see [23]). Furthermore from Theorem 7.1 (and Corollary 7.2), the mapping $DS(w_* \,|\, x_*)$, being piecewise polyhedral, must be piecewise linear when it is single-valued (cf. 2.48 and 9.57 of [23]). Thus, (a) implies (b) along with piecewise linear semi-differentiability.

To complete the proof of the theorem, we must show if (b) holds, then S has a single-valued Lipschitzian localization at w_* for x_*. For this purpose we can invoke Theorem 4.2 in order to transform the task into one of showing that an auxiliary mapping S_* has a single-valued Lipschitzian localization at 0 for x_*, where S_* has the form (1.7)–(1.8), as in the earlier parts of this paper, except that now $F = N_C$. We specifically choose the function f_* in (1.8) by $f_*(x) = f(w_*, x_*) + \nabla_x f(w_*, x_*)(x - x_*)$, so that

$$S_*(y) = \{x \mid h(y,x) + N_C(x) \ni 0\}, \quad \text{where} \\ h(y,x) = f(w_*, x_*) + \nabla_x f(w_*, x_*)(x - x_*) - y. \tag{7.3}$$

Because C is polyhedral, the mapping N_C is piecewise polyhedral (cf. 12.31 of [23]), and it follows then, because h is linear, that S_* is piecewise polyhedral. Theorem 7.1 is applicable to S_* in place of S, with minor adjustments of notation. It yields the formula

$$DS_*(0 \,|\, x_*)(y') = \{x' \mid -y' + \nabla_x f(w_*, x_*)x' + N_{K_*}(x') \ni 0\} \tag{7.4}$$

for the same critical cone K_* as in (7.2), along with the information that $DS_*(0 \,|\, x_*)$ is piecewise polyhedral.

Crucial now will be the general fact when a set G is polyhedral its tangent cone $T_G(z)$ at a point $z \in G$ coincides in some neighborhood of the origin with the translated set $G - z$. This obviously carries over to piecewise polyhedral sets G as well. Applying it to $G = \operatorname{gph} S_*$ at $z = (0, x_*)$, and remembering that $DS_*(0 \,|\, x_*)$ is the mapping which has $T_{\operatorname{gph} S_*}(0, x_*)$ as its graph, we see that $\operatorname{gph} S_* - (0, x_*)$ coincides with $\operatorname{gph} DS_*(0 \,|\, x_*)$ in a neighborhood of the origin.

In the light of this, it will suffice for us to demonstrate that $DS_*(0 \,|\, x_*)$ is single-valued when $DS(w_* \,|\, x_*)$ is single-valued, inasmuch as the single-valuedness of $DS_*(0 \,|\, x_*)$ in combination with its piecewise polyhedrality will imply its Lipschitz continuity (again cf. 2.48 and 9.57 of [23]). For arbitrary y', is there one and only one x' satisfying in (7.4) the condition $-y' + \nabla_x f(w_*, x_*)x' + N_{K_*}(x') \ni 0$? Under the ample parameterization condition (1.6), it is possible to write $-y' = \nabla_w f(w_*, x_*)w'$ for some w'. The question then is whether there is one and only one x' satisfying

$$\nabla_w f(w_*, x_*)w' + \nabla_x f(w_*, x_*)x' + N_{K_*}(x') \ni 0.$$

Through our assumption that $DS(w_* \,|\, x_*)$ is single-valued, the answer from formula (7.2) is yes, and we are done. □

Proposition 7.5 (example of convex-valuedness). *In particular, the solution mapping S in (7.1) is convex-valued, as postulated in Theorem 7.4, when $f(w,x)$ is monotone with respect to $x \in C$, in the sense that*

$$\langle f(w,x') - f(w,x''),\, x' - x'' \rangle \geq 0 \quad \text{for } x', x'' \in C.$$

Proof. Under this assumption the variational inequality is of monotone type, in which case its set of solutions is convex, as is well known. □

8. Application to minimization over a polyhedral set

In this section we specialize further to the case of a parameterized variational inequality coming out of a minimization problem with fixed linear constraints. This will provide an illustration also of our results on calmness and show how they are related to second-order conditions for optimality. Applications to primal-dual aspects of convex optimization in a format allowing for constraint perturbations will be found in our forthcoming paper [6].

The basic problem we consider here has the form

$$\text{minimize } \varphi(w,x) \text{ over } x \in C, \tag{8.1}$$

where C is a nonempty *polyhedral* (convex) subset of $\mathbb{R}^n$ and the function $\varphi : \mathbb{R}^d \times \mathbb{R}^n$ is of class $\mathcal{C}^2$. For this problem, parameterized by w, the first-order optimality condition is

$$-\nabla_x \varphi(w,x) \in N_C(x), \tag{8.2}$$

and the points x satisfying it are the "quasi-optimal" solutions called *stationary points.* The mapping from w to such points x has the form

$$S : w \mapsto \big\{ x \,\big|\, \nabla_x \varphi(w,x) + N_C(x) \ni 0 \big\} \tag{8.3}$$

and fits our framework as the case of the general mapping S in (1.2) where $m = n$ and

$$f(w,x) = \nabla_x \varphi(w,x), \qquad F(x) = N_C(x). \tag{8.4}$$

The specialization of F to the normal cone mapping N_C for a polyhedral set C was already the topic in the preceding section, so what is new here is merely the specialization of f to $\nabla_x \varphi$. The assumption that $\varphi \in \mathcal{C}^2$ gives us $f \in \mathcal{C}^1$ as required, with

$$\nabla_w f(w,x) = \nabla^2_{xw} \varphi(w,x) \in \mathbb{R}^{n\times d}, \qquad \nabla_x f(w,x) = \nabla^2_{xx} \varphi(w,x) \in \mathbb{R}^{n\times n}, \tag{8.5}$$

and the ample parameterization condition (1.6) for a pair $(w_*, x_*) \in \operatorname{gph} S$ coming out as

$$\operatorname{rank} \nabla^2_{xw} \varphi(w_*, x_*) = n. \tag{8.6}$$

Furnished with this information, it is easy to apply to the stationary point mapping in (8.3) all the results obtained so far in this paper, in particular the ones in Section 7, in which the critical cone becomes

$$K_* = T_C(x_*) \cap \nabla_x \varphi(w_*, x_*)^\perp. \tag{8.7}$$

Rather than recording the details of that, we aim here at exploring certain connections between second-order optimality and our results on calmness and Aubin property.

Recall that, in partnership with the first-order condition for optimality that we are now placing on our reference element (w_*, x_*) in taking it to belong to the graph of the mapping S in (8.3), the *standard second-order necessary condition* for local optimality is

$$\langle u, \nabla^2_{xx}\varphi(w_*, x_*)u\rangle \geq 0 \quad \text{for all } u \in K_* \tag{8.8}$$

for the critical cone K_* in (8.7), whereas the *standard second-order sufficient condition* is

$$\langle u, \nabla^2_{xx}\varphi(w_*, x_*)u\rangle > 0 \quad \text{for all nonzero } u \in K_*. \tag{8.9}$$

The *strong second-order sufficient condition* for local optimality is

$$\langle u, \nabla^2_{xx}\varphi(w_*, x_*)u\rangle > 0 \quad \text{for all nonzero } u \in K_* - K_*. \tag{8.10}$$

Because K_* is convex, $K_* - K_*$ is the smallest subspace of $I\!R^n$ that includes K_*; it is called the *critical subspace* associated with w_* and x_*.

Theorem 8.1 (calmness of optimal solution mappings)**.** *Under the ample parameterization condition (8.6), the following properties of the stationary point mapping S in (8.3) are equivalent at the reference pair $(w_*, x_*) \in \operatorname{gph} S$:*

(i) *the standard second-order sufficient condition (8.9) holds;*

(ii) *x_* is a local minimizer in problem (8.1) for w_*, and S is calm at w_* for isolated x_*.*

Proof. According to Corollary 7.3 as applied to $f = \nabla_x\varphi$, we have calmness at w_* for isolated x_* if and only if

$$\nabla^2_{xx}\varphi(w_*, x_*)x' + N_{K_*}(x_* \,|\, y_*)(x') \ni 0 \quad \Longrightarrow \quad x' = 0. \tag{8.11}$$

On the other hand, we have available to us the following description of normal vectors to a closed convex cone K in terms of the polar cone K^*, as applied to $K = K_*$:

$$v \in N_{K_*}(u) \quad \Longleftrightarrow \quad u \in K_*, \quad v \in K_*^*, \quad u \perp v \tag{8.12}$$

(cf. 11.4(b) of [23]). Therefore, S is calm at w_* for isolated x_* if and only if

$$u \in K_*, \quad -\nabla^2_{xx}\varphi(w_*, x_*)u \in K_*^*, \quad \langle u, \nabla^2_{xx}\varphi(w_*, x_*)u\rangle = 0 \quad \Longrightarrow \quad u = 0. \tag{8.13}$$

Let (i) hold. Then of course x_* is a local minimizer as described, but is S calm at w_* for x_*? If this were not true, there would exist by (8.11) some $u \neq 0$ satisfying the conditions in (8.13), and that would contradict the inequality $\langle u, \nabla_{xx}\varphi(w_*, x_*)u\rangle > 0$ known from the supposition in (i) that (8.9) is satisfied .

Conversely now, let (ii) hold. Because x_* is a local minimizer, the second-order necessary condition (8.8) must be fulfilled; this can be written as

$$u \in K_* \quad \Longrightarrow \quad -\nabla^2_{xx} g(w_*, x_*)u \in K_*^*.$$

The calmness of S, as identified with (8.13), eliminates the possibility of there being a nonzero $u \in K_*$ such that the inequality in (8.8) fails to be strict. Thus, the necessary condition (8.8) turns into the sufficient condition (8.9), and (i) is satisfied. □

We investigate next, in association with the stationary point mapping S in (8.3), the mapping

$$S_* : y \mapsto \big\{ x \,\big|\, \nabla_x \varphi(w_*, x) + N_C(x) \ni y \big\}, \tag{8.14}$$

which has the form in the general theory of the earlier parts of this paper with $f_*(x) = f(w_*, x) = \nabla_x\varphi(w_*, x)$. From Theorem 3.2 we know that, under the ample parameterization condition (8.6), S has the Aubin property at w_* for x_* if and only if this mapping S_* has that property at 0 for x_*. From Theorem 4.2, likewise under the ample parameterization condition (8.6), S has a single-valued Lipschitzian localization at w_* for x_* if and only if this S_* has such a localization at 0 for x_*.

Something else can be brought into this picture. In [5; Theorem 3] we proved that in a variational inequality like the current one, in which C is polyhedral, the Aubin property and the Lipschitzian localization property are equivalent for S and also for S_*. On the other hand, by a result of Poliquin and Rockafellar [17; Theorem 4.5], the strong second-order sufficient condition (8.10) holds if and only if S_* has the Lipschitzian localization property. By combining these results we arrive at the following characterization.

Theorem 8.2 (Lipschitzian localization of optimal solution mappings). *Under the ample parameterization condition (8.6), the following properties of the stationary point mapping S in (8.3) are equivalent at the reference pair $(w_*, x_*) \in \operatorname{gph} S$:*

(i) *the strong second-order sufficient condition (8.10) holds at (w_*, x_*);*

(ii) *S has a single-valued Lipschitzian localization at w_* for x_* such that, for all $(w, x) \in \operatorname{gph} S$ near (w_*, x_*), x is not only a stationary point but a local minimizer in problem (8.1).*

This can be supplemented by a description of the resulting semi-derivatives of the mapping S.

Theorem 8.3 (perturbations of local minimizers)**.** *In the context of the properties in Theorem 8.2, the mapping S is semi-differentiable at w_*; thus (6.2) holds. Moreover in this case $DS(w_* \,|\, x_*)$ is a piecewise linear mapping such that $DS(w_* \,|\, x_*)(w')$ is the unique solution x' to the variational inequality*

$$\nabla^2_{xw}\varphi(w_*, x_*)w' + \nabla^2_{xx}\varphi(w_*, x_*)x' + N_{K_*}(x') \ni 0, \tag{8.15}$$

or equivalently, the unique optimal solution to the quadratic programming subproblem

$$\text{minimize } \langle x', \nabla^2_{xw}\varphi(w_*, x_*)w'\rangle + \frac{1}{2}\langle x', \nabla^2_{xx}\varphi(w_*, x_*)x'\rangle \text{ over } x' \in K_*. \tag{8.16}$$

Proof. We apply Theorem 7.4 and then get the description of $DS(w_* \,|\, x_*)(w')$ through (8.15) by specializing formula (7.2) of Theorem 7.1. Next, we observe that (8.15) is the first-order optimality condition for the problem in (8.16), and, because of the second-order sufficiency we have at hand, it gives local minimizers. □

References

1. F. H. Clarke, A new approach to Lagrange multipliers, *Math. of Oper. Research* **1** (1976), 165–174.
2. A. L. Dontchev, The Graves theorem revisited, *J. Convex Anal.* **3** (1996), 45–53.
3. A. L. Dontchev, Characterizations of Lipschitz stability in optimization, in *Recent developments in well-posed variational problems*, pp. 95–115, Math. Appl. 331, Kluwer, Dordrecht, 1995.
4. A. L. Dontchev and W. W. Hager, An inverse mapping theorem for set-valued maps, *Proc. Amer. Math. Soc.* **121** (1994), 481–489.
5. A. L. Dontchev and R. T. Rockafellar, Characterizations of strong regularity for variational inequalities over polyhedral convex sets, *SIAM J. Optim.* **6** (1996), 1087–1105.
6. A. L. Dontchev and R. T. Rockafellar, Primal-dual solution perturbations in convex optimization, preprint.
7. A. D. Ioffe, Metric regularity and subdifferential calculus, preprint.
8. A. J. King and R. T. Rockafellar, Sensitivity analysis for nonsmooth generalized equations, *Math. Programming* **55** (1992) Ser. A, 193–212.
9. D. Klatte, Upper Lipschitz behavior of solutions to perturbed $C^{1,1}$ programs, preprint.

10. D. Klatte and B. Kummer, Strong stability in nonlinear programming revisited, *J. Austral. Math. Soc.* **40** (1999) Ser. B, 336–352.

11. B. Kummer, Lipschitzian and pseudo-Lipschitzian inverse functions and applications to nonlinear optimization, in *Mathematical programming with data perturbations*, pp. 201–222, Lecture Notes in Pure and Appl. Math. 195, Dekker, New York, 1998.

12. A. B. Levy, Calm minima in parameterized finite-dimensional optimization, preprint.

13. A. B. Levy and R. T. Rockafellar, Sensitivity analysis of solutions to generalized equations, *Trans. Amer. Math. Soc.* **345** (1994), 661–671.

14. A. B. Levy and R. T. Rockafellar, Proto-derivatives and the geometry of solution mappings in nonlinear programming, in *Nonlinear Optimization and Applications* (G. Di Pillo and F. Giannessi, eds.), pp. 249–260, Plenum, New York, 1996.

15. A. B. Levy, R. A. Poliquin and R. T. Rockafellar, Stability of locally optimal solutions, *SIAM J. Optim.* **10** (2000), 580–604.

16. R. A. Poliquin and R. T. Rockafellar, Prox-regular functions in variational analysis, *Trans. Amer. Math. Soc.* **348** (1996), 1805–1838.

17. R. A. Poliquin and R. T. Rockafellar, Tilt stability of local minima, *SIAM J. Optim.* **8** (1998), 287–299.

18. S. M. Robinson, An implicit function theorem for generalized variational inequalities, Technical Summary Report 1672, University of Wisconsin-Madison, 1976.

19. S. M. Robinson, Strongly regular generalized equations, *Math. Oper. Res.* **5** (1980), 43–62.

20. S. M. Robinson, Some continuity properties of polyhedral multifunctions, *Math. Programming Study* **14** (1981), 206–214.

21. S. M. Robinson, An implicit-function theorem for a class of nonsmooth functions, *Math. Oper. Res.* **16** (1991), 292–309.

22. R. T. Rockafellar, Proto-differentiability of set-valued mappings and its applications in optimization, *Analyse non linéaire (Perpignan, 1987); Ann. Inst. H. Poincaré Anal. Non Linéaire* **6** (1989), suppl., 449–482.

23. R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer Verlag, Berlin, 1998.