

STABILITY OF LOCALLY OPTIMAL SOLUTIONS

A. B. Levy,¹ R. A. Poliquin,² R. T. Rockafellar³

Abstract: Necessary and sufficient conditions are obtained for the Lipschitzian stability of local solutions to finite-dimensional parameterized optimization problems in a very general setting. Properties of prox-regularity of the essential objective function and positive definiteness of its coderivative Hessian are the key to these results. A previous characterization of tilt stability comes out as a special case.

Keywords: Parameterized optimization, Lipschitzian stability, tilt stability, coderivative Hessians, prox-regular functions, amenable functions.

¹ Department of Mathematics, Bowdoin College, Brunswick, ME 04011.

² Department of Mathematics, University of Alberta, Edmonton, Alberta T6G 2G1, Canada. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under grant OGP41983.

³ Department of Mathematics, Box 354350, University of Washington, Seattle, WA 98195-4350. This work was supported in part by the National Science Foundation under grant DMS-9803089.

1. Introduction

In concept, any problem of optimization in n real variables can be represented as a problem of minimizing, over the entire space \mathbb{R}^n , a function f with values in $\overline{\mathbb{R}} = [-\infty, \infty]$. Points x that should not be candidates in the minimization can effectively be excluded by setting $f = \infty$ there. Such a representation is especially useful in getting to the heart of theoretical issues in parametric optimization, because it allows problem parameters to be viewed just as additional variables on which f depends.

Our aim is to try to understand in this abstract setting, on the most fundamental level of variational analysis, the circumstances in which locally optimal solutions behave in a “stable” manner with respect to shifts in parameter values. The model we adopt is that of a family of minimization problems in $x \in \mathbb{R}^n$ parameterized by $u \in \mathbb{R}^d$, as specified by a function $f : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$. Within the family we single out a problem

$$\overline{\mathcal{P}} \quad \text{minimize } f(x, \bar{u}) \text{ over } x \in \mathbb{R}^n,$$

and compare it with perturbed versions that come from shifting the associated parameter vector \bar{u} to some nearby vector u . For technical reasons, we further consider, along with such *basic* perturbations, *tilt* perturbations that correspond to adding a small linear term to the objective. Thus, we regard $\overline{\mathcal{P}}$ as imbedded in the larger family of problems

$$\mathcal{P}(u, v) \quad \text{minimize } f(x, u) - \langle v, x \rangle \text{ over } x \in \mathbb{R}^n,$$

with both $u \in \mathbb{R}^d$ and $v \in \mathbb{R}^n$ parameters, so that $\overline{\mathcal{P}} = \mathcal{P}(\bar{u}, \bar{v})$ for $\bar{v} = 0$. In the developments that follow, however, \bar{v} might just as well be any vector, so we refer to the unperturbed problem around which we work as $\mathcal{P}(\bar{u}, \bar{v})$ rather than $\overline{\mathcal{P}}$.

Throughout, we assume that f is lower semicontinuous (lsc) and proper, i.e., not identically ∞ and nowhere taking on $-\infty$. The set of *feasible solutions* to $\mathcal{P}(u, v)$ consists then, by definition, of the points x such that $f(u, x)$ is finite. We denote by \bar{x} a feasible solution to $\mathcal{P}(\bar{u}, \bar{v})$ and investigate it in terms of the functions $m_\delta : \mathbb{R}^d \times \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ and mappings $M_\delta : \mathbb{R}^d \times \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ (set-valued) that are defined for $\delta > 0$ by

$$\begin{aligned} m_\delta(u, v) &= \inf_{|x - \bar{x}| \leq \delta} \{ f(x, u) - \langle v, x \rangle \}, \\ M_\delta(u, v) &= \operatorname{argmin}_{|x - \bar{x}| \leq \delta} \{ f(x, u) - \langle v, x \rangle \}. \end{aligned} \tag{1.1}$$

Here $M_\delta(u, v)$ could consist of a single-point x , in which case M_δ is *single-valued* at (u, v) , but it might contain many points or be empty. By convention, $\operatorname{argmin} = \emptyset$ when the

expression being minimized can only be ∞ ; that ensures having $M_\delta(u, v)$ be empty when $\mathcal{P}(u, v)$ has no feasible solutions x satisfying $|x - \bar{x}| \leq \delta$, i.e., when $m_\delta(u, v) = \infty$. Aside from that case, $M_\delta(u, v)$ is nonempty and $m_\delta(u, v)$ is finite.

In such notation, to say that \bar{x} is a *locally optimal solution* to $\mathcal{P}(\bar{u}, \bar{v})$ is to say that $\bar{x} \in M_\delta(\bar{u}, \bar{v})$ for some $\delta > 0$ (sufficiently small). The stability properties of locally optimal solutions that we target for study revolve around \bar{x} being the *only* point of $M_\delta(\bar{u}, \bar{v})$ and having this single-valuedness of the mapping M_δ at (\bar{u}, \bar{v}) persist in a Lipschitzian manner with respect to certain parameter shifts away from (\bar{u}, \bar{v}) .

Definition 1.1 (solution stability). *A point \bar{x} is a stable locally optimal solution to $\mathcal{P}(\bar{u}, \bar{v})$ (in the basic sense, i.e., relative to the specified parameterization in u only) if there is a $\delta > 0$ such that, on some neighborhood U of \bar{u} , the mapping $u \mapsto M_\delta(u, \bar{v})$ is single-valued and Lipschitz continuous with $M_\delta(\bar{u}, \bar{v}) = \bar{x}$, and the function $u \mapsto m_\delta(u, \bar{v})$ is likewise Lipschitz continuous on U .*

It is a tilt stable locally optimal solution if these properties hold with respect to v instead of u , i.e., for the mapping $v \mapsto M_\delta(\bar{u}, v)$ and the function $v \mapsto m_\delta(\bar{u}, v)$ on some neighborhood V of \bar{v} . It is a fully stable locally optimal solution if these properties hold with respect to (u, v) for the full mapping $(u, v) \mapsto M_\delta(u, v)$ and function $(u, v) \mapsto m_\delta(u, v)$ on some neighborhood $U \times V$ of (\bar{u}, \bar{v}) .

Full stability implies both (basic) stability and tilt stability but in general may differ from those properties. With x and u in \mathbb{R} and $(\bar{x}, \bar{u}) = (0, 0)$, for instance, the case of $f(x, u) = (x - u)^4$ exhibits stability without full stability, whereas $f(x, u) = (x - u^{1/3})^2$ has tilt stability without full stability.

Note that in the definition of tilt stability it would not really be necessary to say anything about m_δ , since the formula for this function in (1.1) implies that $m_\delta(\bar{u}, v)$ is finite and concave in v (as long as $f(\bar{x}, \bar{u})$ is finite). In other situations the Lipschitz continuity of m_δ is not automatic, however, even in the face of Lipschitz continuity of M_δ . For example, the lsc, proper function $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ defined by $f(x, u) = x^2$ when $u = 0$ but $f(x, u) = 1 + x^2$ when $u \neq 0$ has, for $(\bar{x}, \bar{u}) = (0, 0)$ and $\bar{v} = 0$, that $M_\delta(u, \bar{v}) = 0$ for all u , yet $m_\delta(u, \bar{v})$ is discontinuous at $u = \bar{u}$.

Stability properties of one kind or another have extensively been investigated for optimal solutions to conventional nonlinear programming problems as well as for Karush-Kuhn-Tucker pairs in such problems or, more broadly, solutions to “generalized equations” and variational inequalities. The pioneering contribution of Robinson [1] put the focus on single-valued Lipschitzian behavior of optimal solutions. The literature on the subject is vast; the articles of Klatte and Kummer [2] and Dontchev and Rockafellar [3] provide

an overview with many references to Lipschitzian behavior and also to calmness (“upper Lipschitzian” behavior) under perturbations.

The approach we take to stability differs from most of that literature, not merely in adopting the format of extended-real-valued functions, but in the tools we use. Crucial among them is the form of localized Lipschitz continuity for set-valued mappings that was defined by Aubin [4] and the criterion for it that was derived by Mordukhovich [5] in terms of his coderivative mappings. These tools of variational analysis have already been applied to stability issues by those authors in some general ways and also by Rockafellar and Wets in their recent book [6], which offers a thorough exposition of the concepts and their history (in finite dimensions). In other work, Dontchev and Rockafellar [7] have applied such methodology in finer detail to nonlinear programming and variational inequalities over polyhedral sets. Closest to our present effort, however, is the paper of Poliquin and Rockafellar [8], where tilt stability was first explored—in the simpler framework of a minimization problem perturbed by tilt vectors only.

The chief contribution in [8] was a characterization of tilt stability of locally optimal solutions in terms of positive definiteness of the generalized Hessian for f in the sense of Mordukhovich [5]. Here we build on the results in [8] by adding a parameterization in u alongside of the tilt perturbations in v . As in [8], a function property called prox-regularity turns out to be essential. That property, which was introduced by Poliquin and Rockafellar [9] for the sake of fundamental developments in second-order nonsmooth analysis, must be adapted however to the additional parameterization. Likewise, the generalized Hessian in x is no longer enough and must be extended as part of the effort to make sure that the functions $f(\cdot, u)$ depend reasonably on u .

We concentrate on characterizing full stability, being content with the fact that necessary and sufficient conditions for full stability immediately yield sufficient conditions for basic stability. The task of characterizing basic stability on its own appears much more difficult and perhaps not even appropriate. After all, tilt perturbations are a special case of other perturbations (one could have $f(x, u) = f_0(x) - \langle u, x \rangle$, say), so a universal result about basic stability could not escape having to account for them somehow. Indeed, it might well be that such a result would require a sort of extra “constraint qualification” that is tantamount to insisting on good tilt behavior. Anyway, from a practical point of view, as in connection with numerical methodology for instance, there is likely to be little interest in situations where tilt stability is absent.

The assumptions behind our characterization of full stability, stated in Theorem 2.3, cover a very broad range of parameterized optimization problems expressible in the pattern

of $\mathcal{P}(u, v)$. That includes not only nonlinear programming models in standard formats but also extended nonlinear programming models in which the objective function can be represented as the composition of a \mathcal{C}^2 mapping with a proper, lsc, convex function. We establish this in Proposition 2.2.

In order to apply our results to such special cases, one has to invoke a calculus of generalized Hessian mappings to see what one gets for the particular forms of $f(x, u)$ that come up. We have not undertaken to do that because it is a major project in itself and is better reserved for other papers in which the calculus rules suited for the job can systematically be laid out. Here, as a critical first step, we identify the underpinnings to stability at a depth not previously plumbed.

2. Main Results

In dealing with subgradients, we follow the notation and terminology of the book [6]. For a function $g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ and a point $x \in \mathbb{R}^n$, a vector $v \in \mathbb{R}^n$ is a *regular subgradient* of g at x if $g(x)$ is finite and $g(x+w) \geq g(x) + \langle v, w \rangle + o(|w|)$. It is a (*general*) *subgradient* at x if $g(x)$ is finite and there exist sequences $\{x^\nu\}_{\nu=1}^\infty$ and $\{v^\nu\}_{\nu=1}^\infty$ with v^ν a regular subgradient of g at x^ν , such that $v^\nu \rightarrow v$, $x^\nu \rightarrow x$, and $g(x^\nu) \rightarrow g(x)$. The set of all such (general) subgradients of g at x includes the regular subgradients at x and is denoted by $\partial g(x)$. A set-valued *subgradient mapping* $\partial g : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is thereby defined, which is empty-valued outside of $\text{dom } g = \{x \mid g(x) < \infty\}$. The *graph* of ∂g is the set $\text{gph } \partial g \subset \mathbb{R}^n \times \mathbb{R}^n$ consisting of the pairs (x, v) such that $v \in \partial g(x)$.

Also of use to us will be the concept of v being a *horizon subgradient* of g at x . This refers to the existence of sequences $\{x^\nu\}_{\nu=1}^\infty$ and $\{v^\nu\}_{\nu=1}^\infty$ with v^ν a regular subgradient of g at x^ν , such that $x^\nu \rightarrow x$, $g(x^\nu) \rightarrow g(x)$, and $\lambda^\nu v^\nu \rightarrow v$ for some scalar sequence $\{\lambda^\nu\}_{\nu=1}^\infty$ with $\lambda^\nu \searrow 0$. The set of horizon subgradients v of g at x is denoted by $\partial^\infty g(x)$.

Prox-regularity arises from consideration of regular subgradients with a second-order aspect. A *proximal subgradient* of g at x is a regular subgradient v for which the error term $o(|w|)$ can be specialized to $(r/2)|w|^2$. Prox-regularity refers to a situation in which proximal subgradients prevail locally and with the same r . Specifically, g is *prox-regular* at \bar{x} for \bar{v} if it is locally lsc at \bar{x} (cf. [6; 1.33, 1.34]), has $\bar{v} \in \partial g(\bar{x})$, and there are neighborhoods X of \bar{x} and V of \bar{v} along with $\varepsilon > 0$ and $r \geq 0$ such that

$$\begin{aligned} g(x') &\geq g(x) + \langle v, x' - x \rangle - \frac{r}{2}|x' - x|^2 \quad \text{for all } x' \in X \\ &\text{when } v \in \partial g(x), v \in V, x \in X, g(x) \leq g(\bar{x}) + \varepsilon. \end{aligned} \tag{2.1}$$

It is *continuously prox-regular* at \bar{x} for \bar{v} if, in addition, $g(x)$ is continuous as a function of $(x, v) \in \text{gph } \partial g$ at (\bar{x}, \bar{v}) . (The latter property, by itself, is known as the *subdifferential*

continuity of g at \bar{x} for \bar{v} .) In that case one can arrange, by a shrinking of the neighborhoods X and V if necessary, that

$$\begin{aligned} g(x') &\geq g(x) + \langle v, x' - x \rangle - \frac{r}{2} |x' - x|^2 && \text{for all } x' \in X \\ &&& \text{when } v \in \partial g(x), v \in V, x \in X. \end{aligned} \quad (2.2)$$

The class of continuously prox-regular functions is very wide and includes not only convex functions, \mathcal{C}^2 functions and lower- \mathcal{C}^2 functions, but also any such function plus the indicator of a set defined by finitely many \mathcal{C}^2 constraints under a constraint qualification. Many, if not most, of the essential objective functions in finite-dimensional optimization are covered. An overview is provided in [6; Chap. 13]. An elaboration for the parametric situation at hand will be given below in Proposition 2.2.

For the indicator δ_D of a set $D \subset \mathbb{R}^n$, the subgradient set $\partial\delta_D(x)$ is denoted by $N_D(x)$ and its elements are called the *normal vectors* to D at x .

Generalized Hessians are derived from normal vectors to the graphs of subgradient mappings. For any mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^p$, we denote by $\text{gph } S$ the set of all pairs $(z, w) \in \mathbb{R}^m \times \mathbb{R}^n$ such that $w \in S(z)$. For any such pair (z, w) , the *coderivative* of S at z for w is the mapping $D^*S(z|w) : \mathbb{R}^p \rightrightarrows \mathbb{R}^m$ defined by

$$D^*S(z|w)(w') = \{z' \mid (z', -w') \in N_{\text{gph } S}(z, w)\}. \quad (2.3)$$

When S is single-valued and \mathcal{C}^1 around z with Jacobian matrix $\nabla S(z)$, the coderivative for $w = S(z)$ reduces to the adjoint linear mapping $w' \mapsto \nabla S(z)^* w'$.

For a subgradient mapping $\partial g : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and a pair $(x, v) \in \text{gph } \partial g$, the mapping $D^*(\partial g)(x|v)$ is the *coderivative Hessian* associated with g at x for v in the sense of Mordukhovich [5] and is denoted by $\partial^2 g(x|v)$. If g is \mathcal{C}^2 around x with Hessian matrix $\nabla^2 g(x)$, then $\partial^2 g(x|v)$ for $v = \nabla g(x)$ reduces to the linear mapping $v' \mapsto \nabla^2 g(x)v'$.

In the context of our parametric model, as specified by the function $f : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \overline{\mathbb{R}}$, these concepts need some adaptation. The spotlight there is on the *partial* subgradient mapping $\partial_x f : \mathbb{R}^n \times \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ defined by

$$\partial_x f(x, u) = \{ \text{set of subgradients } v \text{ of } f_u := f(\cdot, u) \text{ at } x \} = \partial f_u(x). \quad (2.4)$$

The importance of $\partial_x f$ comes from the elementary rule that wherever a function on \mathbb{R}^n has a local minimum, its subgradient set must contain 0. Application of that rule to $f(\cdot, u) - \langle v, \cdot \rangle$ yields the first-order necessary condition with which we must work:

$$x \text{ locally optimal in } \mathcal{P}(u, v) \implies v \in \partial_x f(x, u). \quad (2.5)$$

In particular, any local optimal solution \bar{x} to $\mathcal{P}(\bar{u}, \bar{v})$ must have $\bar{v} \in \partial_x f(\bar{x}, \bar{u})$.

Although the constraints in $\mathcal{P}(u, v)$ are only implicit in our general framework, as signaled by ∞ values of f , a notion of “constraint qualification” comes in anyway. The *basic constraint qualification* at a feasible solution x to $\mathcal{P}(u, v)$ is the condition

$$\mathcal{Q}(x, u) \quad (0, y) \in \partial^\infty f(x, u) \implies y = 0.$$

In our reference problem $\mathcal{P}(\bar{u}, \bar{v})$, we will be concerned primarily with \bar{x} and $\mathcal{Q}(\bar{x}, \bar{u})$.

Note that $\partial^\infty f(x, u)$ refers to horizon subgradients of f as a function of both arguments, not just in x . As demonstrated in [6; 10.12], the constraint qualification $\mathcal{Q}(x, u)$ guarantees in connection with the optimality condition in (2.5) the existence of y such that $(v, y) \in \partial f(x, u)$. In other words, it implies that

$$\partial_x f(x, u) \subset \{v \mid \exists y \text{ with } (v, y) \in \partial f(x, u)\}.$$

In the circumstances we ultimately will be working with (in Theorem 2.3), this inclusion will turn out actually to be an equation (cf. Proposition 3.4). Nonetheless, the mapping $\partial_x f : \mathbb{R}^n \times \mathbb{R}^d \rightrightarrows \mathbb{R}^n$ rather than the mapping $\partial f : \mathbb{R}^n \times \mathbb{R}^d \rightrightarrows \mathbb{R}^n \times \mathbb{R}^d$ will be the vehicle for stating our results.

In analyzing the parametric behavior of locally optimal solutions on the platform of the optimality condition in (2.5), we will inevitably be concerned not only with $\partial_x f$ but also with its partial inverse

$$M : (u, v) \mapsto \{x \mid v \in \partial_x f(x, u)\}. \quad (2.6)$$

Because the first-order condition $v \in \partial_x f(x, u)$ is also necessary for optimality in the minimization problem that defines $M_\delta(u, v)$ in (1.1) when $|x - \bar{x}| < \delta$, we know that

$$x \in M_\delta(u, v), |x - \bar{x}| < \delta \implies x \in M(u, v). \quad (2.7)$$

Much will hinge on ascertaining when the graphs of M_δ and M actually coincide around $(\bar{u}, \bar{v}, \bar{x})$ for small δ , with M single-valued and Lipschitz continuous in such localization. The analysis will center on the coderivative mappings $D^*(\partial_x f)(x, u|v) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n \times \mathbb{R}^d$ at points $(x, u, v) \in \text{gph } \partial_x f$ near $(\bar{x}, \bar{u}, \bar{v})$.

It should be observed that the mapping $D^*(\partial_x f)(x, u|v)$ is not the same as the coderivative Hessian mapping

$$\partial_x^2 f(x, u|v) := D^*(\partial f_u)(x|v) = \partial^2 f_u(x|v) \text{ for } f_u = f(\cdot, u). \quad (2.8)$$

With $f \in \mathcal{C}^2$ and $v = \nabla_x f(x, u)$ for instance, $\partial_x^2 f(x, u|v)$ comes out as $v' \mapsto \nabla_{xx}^2 f(x, u)v'$, while $D^*(\partial_x f)(x, u|v)$ comes out as $v' \mapsto (\nabla_{xx}^2 f(x, u)v', \nabla_{ux}^2 f(x, u)v')$. But the mapping $\partial_x^2 f(x, u|v)(v')$ cannot even be identified in general with the mapping

$$v' \mapsto \{x' \mid \exists u', (x', u') \in D^*(\partial_x f)(x, u|v)(v')\}. \quad (2.9)$$

The former has u fixed in its definition, whereas the latter, which for comparison might be denoted by $\tilde{\partial}_x^2 f(x, u|v)$, depends on limits being taken in the u argument as well, and its graph may therefore be larger. Limits in u are a source of strength, however. The positive definiteness that we eventually require will be imposed on $\tilde{\partial}_x^2 f(\bar{x}, \bar{u}|\bar{v})$ instead of $\partial_x^2 f(\bar{x}, \bar{u}|\bar{v})$, although the notation $\tilde{\partial}_x^2 f(\bar{x}, \bar{u}|\bar{v})$ will not be employed in expressing it.

The notion of prox-regularity must now be expanded in order for it to be able to account for parametric effects in u .

Definition 2.1 (parametric prox-regularity). *The lsc expression $f(x, u)$ is prox-regular in x at \bar{x} for \bar{v} with compatible parameterization by u at \bar{u} if $\bar{v} \in \partial_x f(\bar{x}, \bar{u})$ and there exist neighborhoods U of \bar{u} , X of \bar{x} , and V of \bar{v} , along with $\varepsilon > 0$ and $r \geq 0$ such that*

$$\begin{aligned} f(x', u) &\geq f(x, u) + \langle v, x' - x \rangle - \frac{r}{2}|x' - x|^2 \quad \text{for all } x' \in X \\ \text{when } v &\in \partial_x f(x, u), v \in V, x \in X, u \in U, f(x, u) \leq f(\bar{x}, \bar{u}) + \varepsilon. \end{aligned} \quad (2.10)$$

It is continuously prox-regular in x at \bar{x} for \bar{v} with compatible parameterization by u at \bar{u} if, in addition, $f(x, u)$ is continuous as a function of $(x, u, v) \in \text{gph } \partial_x f$ at $(\bar{x}, \bar{u}, \bar{v})$.

Our attention will be focused on the parametric version here of continuous prox-regularity, which obviously entails continuous prox-regularity of $f(\cdot, \bar{u})$ at \bar{x} for \bar{v} , in particular, but spreads some of it uniformly to subgradients of neighboring functions $f(\cdot, u)$. According to its definition, it provides the existence of a neighborhood $X \times U \times V$ of $(\bar{x}, \bar{u}, \bar{v}) \in \text{gph } \partial_x f$ such that, for a certain $r \geq 0$, one has

$$\begin{aligned} f(x', u) &\geq f(x, u) + \langle v, x' - x \rangle - \frac{r}{2}|x' - x|^2 \quad \text{for all } x' \in X \\ \text{when } (x, u, v) &\in [X \times U \times V] \cap \text{gph } \partial_x f. \end{aligned} \quad (2.11)$$

Strongly amenable functions furnish a prime source of examples for parametric continuous prox-regularity, as we show next. Amenable functions were first studied as a class in [10]. Parametric amenability, as defined in the next proposition, was introduced in [11].

Proposition 2.2 (prox-regularity from amenability). *Suppose that $f(x, u)$ is strongly amenable in x at \bar{x} with compatible parameterization by u at \bar{u} , in the sense that on some*

neighborhood of (\bar{x}, \bar{u}) there is a composite representation $f(x, u) = g(F(x, u))$ in which $F : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}^m$ is a \mathcal{C}^2 mapping and $g : \mathbb{R}^m \rightarrow \overline{\mathbb{R}}$ is a convex, proper, lsc function for which $F(\bar{x}, \bar{u}) \in D := \text{dom } g$ and

$$z \in N_D(F(\bar{x}, \bar{u})), \nabla_x F(\bar{x}, \bar{u})^* z = 0 \implies z = 0. \quad (2.12)$$

Then, as long as $\bar{v} \in \partial_x f(\bar{x}, \bar{u})$, one has $f(x, u)$ continuously prox-regular in x at \bar{x} for \bar{v} with compatible parameterization by u at \bar{u} . Moreover $\mathcal{Q}(\bar{x}, \bar{u})$ holds.

Proof. From (2.12) we have in particular that f is strongly amenable at (\bar{x}, \bar{u}) as a function of (x, u) , since that property by definition (cf. [6; 10.23]) concerns a representation $f = g \circ F$ of the same kind but which need only satisfy

$$z \in N_D(F(\bar{x}, \bar{u})), \nabla_x F(\bar{x}, \bar{u})^* z = 0, \nabla_u F(\bar{x}, \bar{u})^* z = 0 \implies z = 0,$$

where $N_D = \partial^\infty g$ (because g is convex; cf. [6; 8.12]). This condition implies by the subgradient chain rule in [6; 10.6] that $\partial^\infty f(\bar{x}, \bar{u})$ consists of all (v, y) such that there exists $z \in N_D(F(\bar{x}, \bar{u}))$ with $\nabla_x F(\bar{x}, \bar{u})^* z = v$ and $\nabla_u F(\bar{x}, \bar{u})^* z = y$. Clearly, then, it is impossible to have $(0, y) \in \partial^\infty f(\bar{x}, \bar{u})$ unless $y = 0$. Thus, $\mathcal{Q}(\bar{x}, \bar{u})$ is satisfied.

The condition in (2.12) carries over from (\bar{x}, \bar{u}) to all nearby (x, u) with $F(x, u) \in D$, for if not there would be a contradiction based on a simple argument of taking limits. This condition ensures by the same subgradient chain rule that for such (x, u) one has

$$\partial_x f(x, u) = \nabla_x F(x, u)^* \partial g(F(x, u)) = \{v = \nabla_x F(x, u)^* z \mid z \in \partial g(F(x, u))\}. \quad (2.13)$$

Assuming $\bar{v} \in \partial f(\bar{x}, \bar{u})$, let S be the mapping that associates with (x, u, v) the set of vectors z on the right of (2.13). We argue that S is locally bounded at $(\bar{x}, \bar{u}, \bar{v})$, i.e., that there exist $\varepsilon > 0$ and $\zeta > 0$ such that

$$|(x, u, v) - (\bar{x}, \bar{u}, \bar{v})| \leq \varepsilon, z \in S(x, u, v) \implies |z| \leq \zeta, \quad (2.14)$$

moreover with (2.13) holding under these circumstances. The reason is that if we had sequences $(x^\nu, u^\nu, v^\nu) \rightarrow (\bar{x}, \bar{u}, \bar{v})$ and $z^\nu \in S(x^\nu, u^\nu, v^\nu)$ with $0 < |z^\nu| \rightarrow \infty$, the vectors $\lambda^\nu z^\nu$ for $\lambda^\nu = 1/|z^\nu| \searrow 0$ would cluster at some $\bar{z} \neq 0$. Then from having $\nabla_x F(x^\nu, u^\nu)^* [\lambda^\nu z^\nu] = \lambda^\nu v^\nu$ and $z^\nu \in \partial g(F(x^\nu, u^\nu))$ we would get $\nabla_x F(\bar{x}, \bar{u})^* \bar{z} = 0$ and $\bar{z} \in \partial^\infty g(F(\bar{x}, \bar{u}))$. Here we have $\partial^\infty g(F(\bar{x}, \bar{u})) = N_D(F(\bar{x}, \bar{u}))$, so this would contradict (2.12).

Now let $X \times U \times V$ be a neighborhood of $(\bar{x}, \bar{u}, \bar{v})$ small enough that $f = g \circ F$ on $X \times U$ and $|(x, u, v) - (\bar{x}, \bar{u}, \bar{v})| \leq \varepsilon$ when $(x, u, v) \in X \times U \times V$. Suppose $(x^\nu, u^\nu, v^\nu) \rightarrow (\bar{x}, \bar{u}, \bar{v})$

in $X \times U \times V$ with $v^\nu \in \partial_x f(x^\nu, u^\nu)$. Is it true that $f(x^\nu, u^\nu) \rightarrow f(\bar{x}, \bar{u})$? Taking advantage of the formula in (2.13) at (x^ν, u^ν, v^ν) , select $z^\nu \in \partial g(F(x^\nu, u^\nu))$ such that $\nabla_x F(x^\nu, u^\nu)^* z^\nu = v^\nu$. We have $|z^\nu| \leq \zeta$ through (2.14), so by passing to subsequences we can reduce to the case where z^ν converges to some \bar{z} . The pairs $(F(x^\nu, u^\nu), z^\nu) \in \text{gph } \partial g$ converge then to $(F(\bar{x}, \bar{u}), \bar{z})$, and since g is convex (hence subdifferentially continuous) this implies that $g(F(x^\nu, u^\nu)) \rightarrow g(F(\bar{x}, \bar{u}))$. Thus, $f(x^\nu, u^\nu) \rightarrow f(\bar{x}, \bar{u})$ as required.

Observe next that because F is of class \mathcal{C}^2 and the neighborhood U is bounded, there exists $r > 0$ such that, for all z with $|z| \leq \zeta$ and $u \in U$, the function $h_{zu} : x \mapsto \langle z, F(x, u) \rangle$ has $h_{zu}(x') \geq h_{zu}(x) + \langle \nabla_x h_{zu}(x), x' - x \rangle - \frac{r}{2}|x' - x|^2$ for all $x, x' \in X$. This tells us that

$$\begin{aligned} \langle z, F(x', u) - F(x, u) \rangle &\geq \langle \nabla_x F(x, u)^* z, x' - x \rangle - \frac{r}{2}|x' - x|^2 \\ &\text{when } x, x' \in X, \quad u \in U, \quad |z| \leq \zeta. \end{aligned} \quad (2.15)$$

For any $x, x' \in X$, $u \in U$, and $v \in V$ with $v \in \partial_x f(x, u)$, we have $v = \nabla_x F(x, u)^* z$ for some $z \in \partial g(F(x, u))$, necessarily satisfying $|z| \leq \zeta$ by the local boundedness of S in (2.14). The convexity of g yields $g(F(x', u)) \geq g(F(x, u)) + \langle z, F(x', u) - F(x, u) \rangle$, and in combination with (2.15) we therefore have $f(x', u) - f(x, u) = g(F(x', u)) - g(F(x, u)) \geq \langle v, x' - x \rangle - \frac{r}{2}|x' - x|^2$. In other words we have (2.11), as required. \square

As obvious very special cases of Proposition 2.2, f could be any \mathcal{C}^2 function (take $F = f$ and let $g(t) = t$ on \mathbb{R}) or any lsc, proper convex function (take $g = f$ and $F = I$). For a broader discussion of the rich possibilities, see [11] and [6; 10.24].

Theorem 2.3 (full stability). *Let \bar{x} be a feasible solution to $\mathcal{P}(\bar{u}, \bar{v})$ at which the first-order condition $\bar{v} \in \partial_x f(\bar{x}, \bar{u})$ is satisfied along with the constraint qualification $\mathcal{Q}(\bar{x}, \bar{u})$. Suppose $f(x, u)$ is continuously prox-regular in x at \bar{x} for \bar{v} with compatible parameterization by u at \bar{u} . Then for \bar{x} to be a locally optimal solution to $\mathcal{P}(\bar{u}, \bar{v})$ that is fully stable, it is necessary and sufficient that the following second-order conditions be fulfilled:*

- (a) $(x', u') \in D^*(\partial_x f)(\bar{x}, \bar{u} | \bar{v})(v')$, $v' \neq 0 \Rightarrow \langle v', x' \rangle > 0$,
- (b) $(0, u') \in D^*(\partial_x f)(\bar{x}, \bar{u} | \bar{v})(0) \Rightarrow u' = 0$.

Moreover in that case it follows, when $\delta > 0$ is sufficiently small, that for all (u, v) in some neighborhood of (\bar{u}, \bar{v}) one has $M_\delta(u, v) = M(u, v) \cap \{x \mid |x - \bar{x}| < \delta\}$. In addition, the Lipschitz modulus of M_δ at (\bar{u}, \bar{v}) is given then by

$$(\text{lip } M_\delta)(\bar{u}, \bar{v}) = \max \left\{ \frac{|(u', v')|}{|x'|} \mid (x', u') \in D^*(\partial_x f)(\bar{x}, \bar{u} | \bar{v})(v'), \quad x' \neq 0 \right\}, \quad (2.16)$$

where $(\text{lip } M_\delta)(\bar{u}, \bar{v})$ is the upper limit of $|M_\delta(u_1, v_1) - M_\delta(u_2, v_2)| / |(u_1, v_1) - (u_2, v_2)|$ as $(u_1, v_1) \rightarrow (\bar{u}, \bar{v})$ and $(u_2, v_2) \rightarrow (\bar{u}, \bar{v})$ with $(u_1, v_1) \neq (u_2, v_2)$.

This is our main result. It will be proved in §5. The proof of equivalence really centers just on the single-valuedness and Lipschitz continuity of M_δ . The local Lipschitz continuity of m_δ that has been incorporated into the definition of full stability is already a consequence merely of assuming $\mathcal{Q}(\bar{x}, \bar{u})$ (cf. Proposition 3.5).

Theorem 2.3 covers the chief characterization of tilt stability in [8] as the case where the parameterization in u drops out and only the tilt vectors v remain. It adds to that characterization the corresponding specialization of the modulus formula in (2.16), i.e., $(\text{lip } M_\delta)(\bar{v}) = \max\{|v'|/|x'| \mid x' \in \partial^2 f(\bar{x}|\bar{v})(v'), x' \neq 0\}$. Of course it also provides a criterion for the basic form of stability in Definition 1.1.

Corollary 2.4 (basic stability). *The properties in Theorem 2.3 suffice for \bar{x} to be a locally optimal solution to $\mathcal{P}(\bar{u}, \bar{v})$ that is stable (in the basic sense).*

Corollary 2.5 (amenable case). *Suppose $f(x, u)$ is strongly amenable in x at \bar{x} with compatible parameterization by u at \bar{u} . Then for \bar{x} to be a locally optimal solution to $\mathcal{P}(\bar{u}, \bar{v})$ that is fully stable, it is necessary and sufficient that the second-order conditions (a) and (b) of Theorem 2.3 be fulfilled along with the first-order condition $\bar{v} \in \partial_x f(\bar{x}, \bar{u})$.*

Proof. This is immediate from Theorem 2.3 and Proposition 2.2. □

Corollary 2.6 (smooth case). *Let f be of class \mathcal{C}^2 around (\bar{x}, \bar{u}) . In order for \bar{x} to be a locally optimal solution to $\mathcal{P}(\bar{u}, \bar{v})$ that is fully stable, it is necessary and sufficient that $\nabla_x f(\bar{x}, \bar{u}) = \bar{v}$ with $\nabla_{xx}^2 f(\bar{x}, \bar{u})$ positive definite.*

Proof. For f of this type we have the amenability in Corollary 2.5. The coderivative mapping $D^*(\partial_x f)(\bar{x}, \bar{u})$ reduces to the linear mapping $v' \mapsto (\nabla_{xx}^2 f(\bar{x}, \bar{u})v', \nabla_{ux}^2 f(\bar{x}, \bar{u})v')$, as noted earlier. Condition (a) of Theorem 2.3 turns into the positive definiteness of $\nabla_{xx}^2 f(\bar{x}, \bar{u})$, while condition (b) trivializes. □

It would be possible to derive the fact in Corollary 2.6 by classical methods, but we present it this way to show how it fits into the broader scene. The direct argument is not as easy as might be imagined, however; cf. the corresponding case of tilt stability in [8].

Corollary 2.6 brings attention to the “positive definiteness” in (a) of Theorem 2.3 as expressing a *second-order sufficient condition for optimality*, at least in combination with (b). This role was observed previously by Poliquin and Rockafellar in their tilt stability setting in [8]. Although second-order conditions in terms of coderivative Hessians can, in general, be far from the sharpest conditions for confirming local optimality, if only that were the issue, our results show that they are sharp for confirming local optimality together with stability. In the unconstrained optimization in Corollary 2.6, especially the

tilt case with u suppressed, such a gap between stable and unstable second-order sufficient conditions is absent, but it appears to prevail almost everywhere else.

Theorem 2.3 requires f to belong to a class of prox-regular functions. Proposition 2.2 underscores the breadth of this class. Still, one can ask whether the stability conclusions might hold for an even larger class. The answer is essentially negative, however.

Theorem 2.7 (effective need for prox-regularity). *Let \bar{x} be a locally optimal solution to $\mathcal{P}(\bar{u}, \bar{v})$ that is fully stable and satisfies $\mathcal{Q}(\bar{x}, \bar{u})$. Then there is a proper, lsc function \hat{f} that has the prox-regularity ascribed to f in Theorem 2.3 and is locally equivalent to f for purposes of optimization, in the following sense: For the problems $\hat{\mathcal{P}}(u, v)$ obtained with \hat{f} in place of f , the associated \hat{m}_δ and \hat{M}_δ for δ sufficiently small agree with m_δ and M_δ on a neighborhood of (\bar{u}, \bar{v}) . Indeed, one can take $\hat{f}(x, u)$ convex in x and such that, for (u, v) near to (\bar{u}, \bar{v}) , if $v \in \partial_x \hat{f}(x, u)$ then $v \in \partial_x f(x, u)$ and $\hat{f}(x, u) = f(x, u)$.*

This theorem will be proved in §5 as well. The need for replacing f by a “locally equivalent” function \hat{f} to get a converse result can be seen already from examples focused on tilt stability. On \mathbb{R}^2 , let $f(x, u) = |x| \sin(1/x) + 2|x|$ with $f(0, u) = 0$. The increasingly wild oscillations prevent f from having the prox-regularity demanded in Theorem 2.3 relative to $(\bar{x}, \bar{u}) = (0, 0)$ and $\bar{v} = 0$. The function $\hat{f}(x, u) = |x|$ does have all the properties though. (It is convex and therefore covered by Proposition 2.2.) For any $\delta > 0$ and $(u, v) \in W = \mathbb{R} \times (-1, 1)$ we have $\hat{m}_\delta(u, v) = m_\delta(u, v) = 0$ and $\hat{M}_\delta(u, v) = M_\delta(u, v) = 0$. Thus, f and \hat{f} are equivalent in the sense described in Theorem 2.7.

3. Prox-Regularity Under the Constraint Qualification

Laying the groundwork for the proof of Theorem 2.3, we show that the combination of parametric prox-regularity with the constraint qualification $\mathcal{Q}(\bar{x}, \bar{u})$ produces even more uniformity than has been explicitly built into Definition 2.1. The analysis revolves around a form of “graphically localized Lipschitz continuity” of set-valued mappings which will also be important later in the study of the mappings $\partial_x f$ and M but for now is utilized in an epigraphical context.

A mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^p$ has the *Aubin property* at \bar{z} for \bar{w} , an element of $S(\bar{z})$, if there are neighborhoods Z of \bar{z} and W of \bar{w} along with $\kappa \geq 0$ such that

$$S(z') \cap W \subset S(z) + \kappa|z' - z|\mathcal{B} \text{ for all } z, z' \in Z. \quad (3.1)$$

Here \mathcal{B} is the closed unit ball in \mathbb{R}^p . This property, which Aubin called “pseudo-Lipschitz continuity” in [4], reduces for single-valued S to Lipschitz continuity around \bar{z} . A powerful

criterion has been found by Mordukhovich [5], [12], [13]: As long as $\text{gph } S$ is closed relative to a neighborhood of (\bar{z}, \bar{w}) , the Aubin property holds if and only if

$$z' \in D^*S(\bar{z}|\bar{w})(0) \implies z' = 0, \quad (3.2)$$

where moreover the lowest limiting value at (\bar{z}, \bar{w}) of the moduli κ that work in (3.1) has been characterized as the “norm” of the coderivative mapping $D^*S(\bar{z}|\bar{w})$. (That characterization will ultimately be the source of formula (2.16) in Theorem 2.3.) The great advantage of the Mordukhovich criterion is that, because coderivatives of S arise from normal vectors to $\text{gph } S$, it can be invoked in tandem with the calculus of coderivatives that comes out of the calculus of normal vectors. See [6; Chap. 9] as well as [14].

The constraint qualification $\mathcal{Q}(\bar{x}, \bar{u})$ has an interpretation in this context in terms of the epigraphs

$$\text{epi } f_u = \text{epi } f(\cdot, u) := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x, u) \leq \alpha\}.$$

As shown in [6; 10.16], it amounts to the Mordukhovich criterion for the epigraphical mapping $E : u \mapsto \text{epi } f_u$ at \bar{u} for $(\bar{x}, f(\bar{x}, \bar{u}))$ and therefore to the Aubin property holding there. (The graph of this mapping is closed because f is lsc.)

Proposition 3.1 (consequences of the basic constraint qualification). *Under the constraint qualification $\mathcal{Q}(\bar{x}, \bar{u})$, there exist neighborhoods X_1 of \bar{x} and U_1 of \bar{u} along with $\varepsilon > 0$ and $\kappa \geq 0$ such that*

$$\left. \begin{array}{l} x \in X_1, \quad u, u' \in U_1 \\ f(x, u) \leq f(\bar{x}, \bar{u}) + \varepsilon \end{array} \right\} \implies \exists x' \text{ with } \begin{cases} |x' - x| \leq \kappa|u' - u|, \\ f(x', u') \leq f(x, u) + \kappa|u' - u|. \end{cases} \quad (3.3)$$

Proof. We have just observed that $\mathcal{Q}(\bar{x}, \bar{u})$ corresponds to having the Aubin property of the set-valued mapping $E : u \mapsto \text{epi } f_u$ hold at \bar{u} for $(\bar{x}, \bar{\alpha})$, where $\bar{\alpha} := f(\bar{x}, \bar{u})$, so the task is to show that this yields (3.3).

With convenient adjustments of notation to fit the epigraphical setting, the Aubin property in question can be identified with the existence of neighborhoods X_1 of \bar{x} and U_1 of \bar{u} along with $\varepsilon > 0$ and $\kappa \geq 0$ such that, for all $u, u' \in U_1$, one has

$$[\text{epi } f_u] \cap ([X_1 \times [\bar{\alpha} - \varepsilon, \bar{\alpha} + \varepsilon]]) \subset [\text{epi } f_{u'}] + \kappa|u' - u|(\mathbb{B} \times [-1, 1]),$$

or in other words the implication

$$\left. \begin{array}{l} x \in X_1 \\ \alpha \geq f(x, u) \\ |\alpha - \bar{\alpha}| \leq \varepsilon \end{array} \right\} \implies \exists(x', \alpha') \text{ with } \begin{cases} f(x', u') \leq \alpha', \\ |x' - x| \leq \kappa|u' - u|, \\ |\alpha' - \alpha| \leq \kappa|u' - u|. \end{cases} \quad (3.4)$$

Because f is lsc in this implication, we can arrange (by shrinking X_1 and U_1 if necessary) that $f(x, u) \geq \bar{\alpha} - \varepsilon$ when $(x, u) \in X_1 \times U_1$. Then only the inequality $\alpha \leq \bar{\alpha} + \varepsilon$ has force on the left. On the other hand, only the upper bound provided by the inequality $|\alpha' - \alpha| \leq \kappa|u' - u|$ has force on the right. Thus, we can enhance (3.4) to

$$\left. \begin{array}{l} x \in X_1 \\ \alpha \geq f(x, u) \\ \alpha \leq \bar{\alpha} + \varepsilon \end{array} \right\} \implies \exists(x', \alpha') \text{ with } \begin{cases} f(x', u') \leq \alpha', \\ |x' - x| \leq \kappa|u' - u|, \\ \alpha' \leq \alpha + \kappa|u' - u|. \end{cases} \quad (3.5)$$

When (3.5) is invoked in the case of $\alpha = f(x, u)$, the x' it produces has $f(x', u') \leq f(x, u) + \kappa|u' - u|$. Since (3.5) holds for arbitrary $u, u' \in U_1$, we have (3.3). \square

We use this now to bring out some important consequences of parametric prox-regularity.

Proposition 3.2 (persistence of prox-regularity). *Let the constraint qualification $\mathcal{Q}(\bar{x}, \bar{u})$ hold with $\bar{v} \in \partial_x f(\bar{x}, \bar{u})$, and suppose that $f(x, u)$ is continuously prox-regular in x at \bar{x} for \bar{v} with compatible parameterization by u at \bar{u} . Then an open neighborhood $X \times U \times V$ of $(\bar{x}, \bar{u}, \bar{v})$ can be found for which the uniform proximal subgradient property in (2.11) holds and, in addition,*

- (a) $f(x, u)$ is continuous as a function of $(x, u, v) \in [X \times U \times V] \cap \text{gph } \partial_x f$,
- (b) $\text{gph } \partial_x f$ is closed relative to $X \times U \times V$.

In particular, then, one has for all $(\tilde{x}, \tilde{u}, \tilde{v}) \in [X \times U \times V] \cap \text{gph } \partial_x f$ that $f(x, u)$ is continuously prox-regular in x at \tilde{x} for \tilde{v} with compatible parameterization by u at \tilde{u} .

Proof. Let X_0, U_0 and V_0 be neighborhoods as in the definition of continuous prox-regularity, so that (2.11) holds for them and a certain r . Let X_1, U_1, λ and κ have the property in Proposition 3.1. Choose an open neighborhood $X \times U \times V$ of $(\bar{x}, \bar{u}, \bar{v})$ such that $X \times U \times V \subset X_0 \times U_0 \times V_0$, $X \times U \subset X_1 \times U_1$, and

$$(x, u, v) \in [X \times U \times V] \cap \text{gph } \partial_x f \implies f(x, u) < f(\bar{x}, \bar{u}) + \lambda,$$

the latter being possible because $f(x, u)$ is continuous at (\bar{x}, \bar{u}) as a function of $(x, u, v) \in [X \times U \times V] \cap \text{gph } \partial_x f$. Then (2.11) holds for the neighborhoods X, U and V , and (3.3) can be invoked in the simplified form:

$$\left. \begin{array}{l} (x, u) \in X \times U, u' \in U \\ (x, u, v) \in \text{gph } \partial_x f, v \in V \end{array} \right\} \implies \exists x' \text{ with } \begin{cases} |x' - x| \leq \kappa|u' - u|, \\ f(x', u') \leq f(x, u) + \kappa|u' - u|. \end{cases} \quad (3.6)$$

Consider any sequence of points $(x^\nu, u^\nu, v^\nu) \in [X \times U \times V] \cap \text{gph } \partial_x f$ that converges to a point $(\tilde{x}, \tilde{u}, \tilde{v}) \in [X \times U \times V]$. We have to demonstrate that $f(x^\nu, u^\nu) \rightarrow f(\tilde{x}, \tilde{u})$ and $(\tilde{x}, \tilde{u}, \tilde{v}) \in \text{gph } \partial_x f$.

We first apply (3.6) to $x = \tilde{x}$, $u = \tilde{u}$, and $u' = u^\nu$ to obtain for each ν the existence of \tilde{x}^ν such that $|\tilde{x}^\nu - \tilde{x}| \leq \kappa|u^\nu - \tilde{u}|$ and $f(\tilde{x}^\nu, u^\nu) \leq f(\tilde{x}, \tilde{u}) + \kappa|u^\nu - \tilde{u}|$. Then $\tilde{x}^\nu \rightarrow \tilde{x}$ and $f(\tilde{x}^\nu, u^\nu) \rightarrow f(\tilde{x}, \tilde{u})$ (because f is lsc). Eventually $\tilde{x}^\nu \in X$, so that we have

$$f(\tilde{x}^\nu, u^\nu) \geq f(x^\nu, u^\nu) + \langle v^\nu, \tilde{x}^\nu - x^\nu \rangle - \frac{r}{2}|\tilde{x}^\nu - x^\nu|^2.$$

The second and third terms on the right tend to 0 as $(x^\nu, u^\nu) \rightarrow (\tilde{x}, \tilde{u})$, so from knowing that $f(\tilde{x}^\nu, u^\nu) \rightarrow f(\tilde{x}, \tilde{u})$ we may conclude that $f(x^\nu, u^\nu) \rightarrow f(\tilde{x}, \tilde{u})$ (because f is lsc). This establishes (a).

Next we consider any point $\hat{x} \in X$ and apply (3.6) to $x = \hat{x}$, $u = \tilde{u}$ and $u' = u^\nu$ to get for each ν the existence of \hat{x}^ν such that $|\hat{x}^\nu - \hat{x}| \leq \kappa|u^\nu - \tilde{u}|$ and $f(\hat{x}^\nu, u^\nu) \leq f(\hat{x}, \tilde{u}) + \kappa|u^\nu - \tilde{u}|$. We have $\hat{x}^\nu \rightarrow \hat{x}$ and $f(\hat{x}^\nu, u^\nu) \rightarrow f(\hat{x}, \tilde{u})$ (again because f is lsc). Furthermore, we have from (2.11) that

$$f(\hat{x}^\nu, u^\nu) \geq f(x^\nu, u^\nu) + \langle v^\nu, \hat{x}^\nu - x^\nu \rangle - \frac{r}{2}|\hat{x}^\nu - x^\nu|^2.$$

Limits are known for all the terms in this inequality, and in passing to them we obtain

$$f(\hat{x}, \tilde{u}) \geq f(\tilde{x}, \tilde{u}) + \langle \tilde{v}, \hat{x} - \tilde{x} \rangle - \frac{r}{2}|\hat{x} - \tilde{x}|^2.$$

This has been shown to hold for arbitrary \hat{x} in X , which is a neighborhood of \tilde{x} , so it follows that \tilde{v} is a regular subgradient of $f(\cdot, \tilde{u})$ at \tilde{x} and hence in particular that $\tilde{v} \in \partial_x f(\tilde{x}, \tilde{u})$. This establishes (b). \square

Corollary 3.3 (nonparametric case). *Suppose that a function $g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ is continuously prox-regular at \bar{x} for \bar{v} . Then an open neighborhood $X \times V$ of (\bar{x}, \bar{v}) can be found for which the uniform proximal subgradient property in (2.2) holds and, in addition,*

- (a) $g(x)$ is continuous as a function of $(x, v) \in [X \times V] \cap \text{gph } \partial g$,
- (b) $\text{gph } \partial g$ is closed relative to $X \times V$.

In particular, then, one has for all $(\tilde{x}, \tilde{v}) \in [X \times V] \cap \text{gph } \partial g$ that g is continuously prox-regular at \tilde{x} for \tilde{v} .

Proof. Here we take $f(x, u) \equiv g(x)$. \square

Proposition 3.4 (subgradients under parametric prox-regularity). *Under the hypothesis of Proposition 3.2, there is a neighborhood of $(\bar{x}, \bar{u}, \bar{v})$ such that, as long as (x, u, v) lies in this neighborhood, one has*

$$v \in \partial_x f(x, u) \iff \exists y \text{ with } (v, y) \in \partial f(x, u). \quad (3.7)$$

Proof. Because $\mathcal{Q}(\bar{x}, \bar{u})$ holds, the constraint qualification $\mathcal{Q}(x, u)$ holds too when (x, u) is close enough to (\bar{x}, \bar{u}) with $f(x, u)$ close enough to $f(\bar{x}, \bar{u})$. (Otherwise a contradiction can be reached by a simple argument based on the definition of $\partial^\infty f(\bar{x}, \bar{u})$.) As part of the continuous prox-regularity that is assumed, we know that when (x, u, v) approaches $(\bar{x}, \bar{u}, \bar{v})$ within $\text{gph } \partial_x f$, $f(x, u)$ automatically approaches $f(\bar{x}, \bar{u})$, so the proviso about $f(x, u)$ being close enough to $f(\bar{x}, \bar{u})$ is superfluous.

The constraint qualification $\mathcal{Q}(x, u)$ guarantees that “ \Rightarrow ” holds in (3.7); see [6; 10.11]. For the converse, suppose that $(v, y) \in \partial f(x, u)$ with (x, u, v) in an open neighborhood $X \times U \times V$ of $(\bar{x}, \bar{u}, \bar{v})$ of the kind in Proposition 3.2. Then by definition there is a sequence of points $(x^\nu, u^\nu, v^\nu, y^\nu) \rightarrow (x, u, v, y)$ with $f(x^\nu, u^\nu) \rightarrow f(x, u)$ and (v^ν, y^ν) a regular subgradient of f at (x^ν, u^ν) . Then v^ν is a regular subgradient of $f(\cdot, u^\nu)$ at x^ν and in particular $v^\nu \in \partial_x f(x^\nu, u^\nu)$. Eventually (x^ν, u^ν, v^ν) belongs to the neighborhood $X \times U \times V$, and by appealing to (b) of Proposition 3.2 we see that the limit (x, u, v) still lies in $\text{gph } \partial_x f$. Thus, “ \Leftarrow ” holds in (3.7) when $(x, u, v) \in X \times U \times V$. \square

We finish off with a result about the behavior of the functions m_δ and mappings M_δ in (1.1), which will be needed later in the proof of Theorem 2.3.

Proposition 3.5 (convergence in local optimality). *Suppose $M_\delta(\bar{u}, \bar{v}) = \{\bar{x}\}$ for some $\delta > 0$ and the constraint qualification $\mathcal{Q}(\bar{x}, \bar{u})$ is satisfied. Then m_δ is Lipschitz continuous around (\bar{u}, \bar{v}) , and for every $\varepsilon > 0$ there is a neighborhood W of (\bar{u}, \bar{v}) such that*

$$(u, v) \in W \implies \emptyset \neq M_\delta(u, v) \subset \{x \mid |x - \bar{x}| < \varepsilon\}.$$

Proof. In terms of the function

$$g_\delta(u, v, x) := \begin{cases} f(x, u) - \langle v, x \rangle & \text{if } |x - \bar{x}| \leq \delta, \\ \infty & \text{if } |x - \bar{x}| > \delta, \end{cases}$$

we have $m_\delta(u, v) = \inf_x g_\delta(u, v, x)$ and $M_\delta(u, v) = \text{argmin}_x g_\delta(u, v, x)$. Here g_δ is lsc and proper on $\mathbb{R}^d \times \mathbb{R}^n \times \mathbb{R}^n$, and for each (u, v) the level sets of the form $\{x \mid g_\delta(u, v, x) \leq \alpha\}$, $\alpha \in \mathbb{R}$, are of course all contained in the ball $\{x \mid |x - \bar{x}| \leq \delta\}$. Further, we have

$$\partial^\infty g_\delta(\bar{u}, \bar{v}, \bar{x}) = \{(y, 0, w) \mid (w, y) \in \partial^\infty f(\bar{x}, \bar{u})\}$$

by the calculus rule in [6; 8.8(c)], so that, from $\mathcal{Q}(\bar{x}, \bar{u})$, g_δ has $(y, z, 0) \in \partial^\infty g_\delta(\bar{u}, \bar{v}, \bar{x})$ only for $(y, z) = (0, 0)$. On the basis of this constraint qualification we know that m_δ is Lipschitz continuous on some neighborhood of (\bar{u}, \bar{v}) ; cf. [6; 10.13]. The rest then follows from the fundamental theorem on parametric optimization in [6; 1.17]. \square

4. Coderivative Analysis of Subgradient Mappings

Our investigation shifts now to coderivatives of the mapping $\partial_x f$ and its partial inverse M introduced in (2.6).

Proposition 4.1 (partial inverse mapping). *The mapping M has its coderivatives related to those of $\partial_x f$ by*

$$(u', -v') \in D^*M(u, v|x)(-x') \iff (x', u') \in D^*(\partial_x f)(x, u|v)(v'). \quad (4.1)$$

When $\text{gph } \partial_x f$ is closed locally around (x, u, v) , the condition

$$(0, u') \in D^*(\partial_x f)(x, u|v)(v') \implies (u', v') = (0, 0) \quad (4.2)$$

is necessary and sufficient for M to have the Aubin property at (u, v) for x .

Proof. By definition, $(u', -v') \in D^*M(u, v|x)(-x')$ means that $(u', -v', x')$ belongs to $N_{\text{gph } M}(u, v, x)$. Since the elements (u, v, x) of $\text{gph } M$ correspond simply to the elements (x, u, v) of $\text{gph } \partial_x f$, this is the same as having $(x', u', -v') \in N_{\text{gph } \partial_x f}(x, u, v)$. But that means $(x', u') \in D^*(\partial_x f)(x, u|v)(v')$.

Local closedness of $\text{gph } \partial_x f$ around (x, u, v) corresponds to local closedness of $\text{gph } M$ around (u, v, x) and allows the Aubin property of M at (u, v) for x to be captured by the Mordukhovich criterion: $(u', -v') \in D^*M(u, v|x)(0)$ only for $(u', -v') = (0, 0)$. When the latter is translated through (4.1), it comes out as (4.2). \square

Corollary 4.2 (Aubin property of the partial inverse). *Under the hypothesis of Theorem 2.3, conditions (a) and (b) in the statement of that theorem guarantee that M has the Aubin property at (\bar{u}, \bar{v}) for \bar{x} .*

Proof. The hypothesis in question guarantees through Proposition 3.2 that $\text{gph } \partial_x f$ is closed locally around $(\bar{x}, \bar{u}, \bar{v})$. The issue then is whether (4.2) holds there. Let $(0, u') \in D^*(\partial_x f)(\bar{x}, \bar{u}|\bar{v})(v')$. From condition (a) of Theorem 2.3, we must have $v' = 0$. But then by condition (b) of Theorem 2.3, we must have $u' = 0$. Thus, (4.2) is correct. \square

Proposition 4.3 (partial coderivatives). *Consider in terms of $f_u = f(\cdot, u)$ the set-valued mapping $G : \mathbb{R}^d \rightrightarrows \mathbb{R}^n \times \mathbb{R}^n$ defined by*

$$G(u) = \text{gph } \partial f_u = \{(x, v) \mid (x, u, v) \in \text{gph } \partial_x f\}. \quad (4.3)$$

When $\text{gph } \partial_x f$ is closed locally around $(\bar{x}, \bar{u}, \bar{v})$, condition (b) of Theorem 2.3 is equivalent to G having the Aubin property at \bar{u} for (\bar{x}, \bar{v}) . Furthermore, (b) ensures that for all $(x, u, v) \in \text{gph } \partial_x f$ in some neighborhood of $(\bar{x}, \bar{u}, \bar{v})$, one has

$$\partial^2 f_u(x|v)(v') \subset \{x' \mid \exists u' \text{ with } (x', u') \in D^*(\partial_x f)(x, u|v)(v')\}. \quad (4.4)$$

Proof. The elements (u, x, v) of $\text{gph } G$ correspond under permutation to the elements (x, u, v) of $\text{gph } \partial_x f$. From this we get

$$\begin{aligned} (x', u') \in D^*(\partial_x f)(x, u|v)(v') &\iff (x', u', -v') \in N_{\text{gph } \partial_x f}(x, u, v) \\ &\iff (u', x', -v') \in N_{\text{gph } G}(u, x, v) \\ &\iff u' \in D^*G(u|x, v)(-x', v'). \end{aligned} \quad (4.5)$$

The local closedness of $\text{gph } \partial_x f$ around $(\bar{x}, \bar{u}, \bar{v})$ corresponds to the local closedness of $\text{gph } G$ around $(\bar{u}, \bar{x}, \bar{v})$. With such closedness, G has the Aubin property at \bar{u} for (\bar{x}, \bar{v}) if and only if the Mordukhovich criterion is satisfied, namely that $u' \in D^*G(\bar{u}|\bar{x}, \bar{v})(0, 0)$ only for $u' = 0$. This is identical under (4.5) to condition (b) of Theorem 2.3.

The Aubin property of G at \bar{u} for (\bar{x}, \bar{v}) entails the Aubin property at u for (x, v) whenever (u, x, v) is near enough to $(\bar{u}, \bar{x}, \bar{v})$ in $\text{gph } G$. Thus, for all such (u, x, v) in $\text{gph } G$, also within the neighborhood of $(\bar{u}, \bar{x}, \bar{v})$ where $\text{gph } G$ is locally closed, the Mordukhovich criterion is satisfied; we can write this as

$$(u', 0, 0) \in N_{\text{gph } G}(u, x, v) \implies u' = 0. \quad (4.6)$$

Fix any such element of $\text{gph } G$, say $(\tilde{u}, \tilde{x}, \tilde{v})$. By determining the normal vectors to the set $G(\tilde{u}) = \text{gph } \partial f_{\tilde{u}}$ at (\tilde{x}, \tilde{v}) , we can determine the coderivative mapping $D^*(\partial f_{\tilde{u}})(\tilde{x}|\tilde{v}) = \partial^2 f_{\tilde{u}}(\tilde{x}|\tilde{v})$. Observing that

$$G(\tilde{u}) = \{(x, v) \mid F(x, v) \in \text{gph } G\} \text{ for } F : (x, v) \mapsto (\tilde{u}, x, v), \quad (4.7)$$

we apply the chain rule for normal vectors in [6; 6.14]. Because $\text{gph } G$ is locally closed around $(\tilde{u}, \tilde{x}, \tilde{v})$, this chain rule is valid as long as the constraint qualification holds that

$$(u', x', v') \in N_{\text{gph } G}(\tilde{u}, \tilde{x}, \tilde{v}), \quad \nabla F(\tilde{x}, \tilde{v})^*(u', x', v') = (0, 0) \implies (u', x', v') = (0, 0, 0).$$

Trivially, though, $\nabla F(\tilde{x}, \tilde{v})^*(u', x', v') = (0, 0)$ only when $(x', v') = (0, 0)$, so this constraint qualification comes out as (4.6) in the case of $(u, x, v) = (\tilde{u}, \tilde{x}, \tilde{v})$ and thus is indeed satisfied.

The chain rule allows us to deduce from (4.7) that

$$\begin{aligned} N_{G(\tilde{u})}(\tilde{x}, \tilde{v}) &\subset \{(x'', v'') \mid \exists (u', x', v') \in N_{\text{gph } G}(\tilde{u}, \tilde{x}, \tilde{v}) \\ &\quad \text{with } \nabla F(\tilde{x}, \tilde{v})^*(u', x', v') = (x'', v'')\} \\ &= \{(x', v') \mid \exists u' \text{ with } (u', x', v') \in N_{\text{gph } G}(\tilde{u}, \tilde{x}, \tilde{v})\}. \end{aligned} \quad (4.8)$$

Noting that $\text{gph } D^*(\partial f_{\tilde{u}})(\tilde{x}|\tilde{v})$ consists of the pairs (v', x') with $(x', -v') \in N_{G(\tilde{u})}(\tilde{x}, \tilde{v})$, whereas $\text{gph } D^*(\partial_x f)(\tilde{x}, \tilde{u}|\tilde{v})$ consists by (4.5) of all (v', x', u') such that $(u', x', -v') \in N_{\text{gph } G}(\tilde{u}, \tilde{x}, \tilde{v})$, we obtain from (4.8) that (4.4) holds. \square

In support of the final proposition in this section, the following lemma will be crucial.

Lemma 4.4 (positive definiteness estimate). *Let $g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be continuously prox-regular at \tilde{x} for \tilde{v} and let $\varepsilon > 0$. If the inequality $\langle x', v' \rangle \geq \varepsilon |v'|^2$ holds for all $(v', x') \in \text{gph } \partial^2 g(\tilde{x}, \tilde{v})$ such that $x' = \lambda v'$ for some $\lambda \in \mathbb{R}$, then it also holds without that restriction.*

Proof. Consider any $\mu \in (0, \varepsilon)$. Let $G = \text{gph } \partial g$. Under our inequality assumption there must be an open neighborhood $X_0 \times V_0$ of (\tilde{x}, \tilde{v}) such that

$$(x, v) \in [X_0 \times V_0] \cap \text{gph } \partial g, \quad (\lambda v', v') \in \text{gph } \partial^2 g(x|v), \quad |v'| = 1 \implies \lambda \geq \mu, \quad (4.9)$$

inasmuch as $\text{gph } \partial^2 g(x|v)$ consists of the vectors (v', x') with $(x', -v') \in N_{\text{gph } \partial g}(x, v)$, and graph of the mapping $N_{\text{gph } \partial g}$ is closed (by the general definition of normal cones).

We can suppose (by shrinking X_0 and V_0 if necessary) that $X_0 \times V_0$ lies within a neighborhood $X \times V$ for which the continuous prox-regularity property in (2.2) is operational and moreover, through Corollary 3.3, makes g continuously prox-regular at x for v when $(x, v) \in [X \times V] \cap \text{gph } \partial g$. Consider now within $[X_0 \times V_0] \cap \text{gph } \partial g$ any point (x, v) with the special property that the ∂g is proto-differentiable at x for v and the corresponding derivative mapping $D(\partial g)(x|v) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is generalized linear. (This property is known actually to hold in an almost everywhere sense because continuous prox-regularity makes $\text{gph } \partial g$ be a graphically Lipschitzian manifold of dimension n in its localization relative to $X \times V$; cf. [9; Prop. 4.8]. The points (x, v) in question are the ‘‘Rademacher points’’ of $\text{gph } \partial g$ near (\tilde{x}, \tilde{v}) . Proto-differentiability is the graphical counterpart to function differentiability; see [6]. A mapping is generalized linear when its graph is a subspace.)

In this situation, three facts are at our disposal. First, according to a theorem of Rockafellar and Zagrodny [15], the graph of $D(\partial g)(x|v)$ is included in the graph of $D^*(\partial g)(x|v) = \partial^2 g(x|v)$, so that by (4.9) we have

$$(\lambda v', v') \in \text{gph } D(\partial g)(x|v), \quad |v'| = 1 \implies \lambda \geq \mu, \quad (4.10)$$

Second, because of the proto-differentiability, $D(\partial g)(x|v)$ is the subgradient mapping ∂h for $h = d^2 g(x|v)$, the second subderivative function associated with g at x for v ; this holds through prox-regularity as shown in [9; Cor. 6.2]. Third, the generalized linearity of ∂h corresponds to h being a generalized (purely) quadratic function: the sum of a purely quadratic function on \mathbb{R}^n and the indicator of a subspace. Thus, there is a subspace L of \mathbb{R}^n along with a symmetric, positive semidefinite matrix $Q \in \mathbb{R}^{n \times n}$ such that

$$D(\partial g)(x|v)(v') = \begin{cases} Qv' + L^\perp & \text{when } v' \in L, \\ \emptyset & \text{when } v' \notin L. \end{cases} \quad (4.11)$$

In combining (4.11) with (4.10), we see that the eigenvalues λ of Q relative to L must all satisfy $\lambda \geq \mu$. This tells us that the generalized linear mapping $D(\partial g)(x|v)$ is μ -strongly

monotone. We invoke next the criterion in [9; Prop. 5.7]: because the mappings $D(\partial g)(x|v)$ of the special type just investigated are all μ -strongly monotone, the localization of ∂g that we are working with is itself μ -strongly monotone.

A monotone mapping T has $\langle x', v' \rangle \geq 0$ whenever $x' \in D^*T(x|v)(v')$, as shown by Poliquin and Rockafellar [8; Thm. 2.1]; therefore, a μ -monotone mapping T (for which $T - \mu I$ is monotone) has $\langle x', v' \rangle \geq \mu|v'|^2$ whenever $x' \in D^*T(x|v)(v')$. In particular, then, in taking T to be our localization of ∂g , we see that

$$(x, v) \in [X_0 \times V_0] \cap \text{gph } \partial g, \quad x' \in \partial^2 g(x|v)(v') \quad \Longrightarrow \quad \langle x', v' \rangle \geq \mu|v'|^2.$$

Applying this at $(x, v) = (\tilde{x}, \tilde{v})$ and recalling that μ was an arbitrary value in $(0, \varepsilon)$, we reach the desired conclusion that $\langle x', v' \rangle \geq \varepsilon|v'|^2$ whenever $x' \in \partial^2 g(\tilde{x}|\tilde{v})(v')$. \square

Proposition 4.5 (uniform positive definiteness). *Let the constraint qualification $\mathcal{Q}(\bar{x}, \bar{u})$ hold with $\bar{v} \in \partial_x f(\bar{x}, \bar{u})$, and suppose that $f(x, u)$ is continuously prox-regular in x at \bar{x} for \bar{v} with compatible parameterization by u at \bar{u} . If conditions (a) and (b) of Theorem 2.3 hold as well, there must actually exist a constant $\varepsilon > 0$ and a neighborhood $X \times U \times V$ of $(\bar{x}, \bar{u}, \bar{v})$ for which, in terms of $f_u = f(\cdot, u)$, one has*

$$\left. \begin{array}{l} x' \in \partial^2 f_u(x|v)(v') \\ (x, u, v) \in [X \times U \times V] \cap \text{gph } \partial_x f \end{array} \right\} \Longrightarrow \langle x', v' \rangle \geq \varepsilon|v'|^2. \quad (4.12)$$

Conversely, if this property holds, then condition (a) of Theorem 2.3 must hold with

$$(x', u') \in D^*(\partial_x f)(\bar{x}, \bar{u}|\bar{v})(v') \quad \Longrightarrow \quad \langle x', v' \rangle \geq \varepsilon|v'|^2. \quad (4.13)$$

Proof. Our hypothesis ensures through Proposition 3.2 that for all (x, u, v) near enough to $(\bar{x}, \bar{u}, \bar{v})$ with $v \in \partial f_u(x)$ the function f_u is continuously prox-regular at x for v . In combining it with condition (b) of Theorem 2.3 and invoking Proposition 4.3, we get the coderivative inclusion in (4.4) to hold locally. Suppose now that condition (a) of Theorem 2.3 is satisfied along with condition (b). To justify the locally uniform positive definiteness property claimed in that case, we will rely on Lemma 4.4, according to which we can obtain (4.12) by demonstrating that

$$\left. \begin{array}{l} \lambda z \in \partial^2 f_u(x|v)(z) \\ (x, u, v) \in [X \times U \times V] \cap \text{gph } \partial_x f \end{array} \right\} \Longrightarrow \lambda \geq \varepsilon.$$

Through the inclusion in (4.4), it suffices to verify the existence of $\varepsilon > 0$ such that

$$\left. \begin{array}{l} (\lambda z, w) \in D^*(\partial_x f)(x, u|v)(z) \\ (x, u, v) \in [X \times U \times V] \cap \text{gph } \partial_x f \end{array} \right\} \Longrightarrow \lambda \geq \varepsilon. \quad (4.14)$$

Suppose there is no such ε . Then there must exist sequences $(x^\nu, u^\nu, v^\nu) \rightarrow (\bar{x}, \bar{u}, \bar{v})$ in $\text{gph } \partial_x f$ along with scalars $\lambda^\nu \searrow 0$ and vectors z^ν and w^ν with $z^\nu \neq 0$, such that $(\lambda^\nu z^\nu, w^\nu) \in D^*(\partial_x f)(x^\nu, u^\nu | v^\nu)(z^\nu)$. The latter means by definition that $(\lambda^\nu z^\nu, w^\nu, -z^\nu)$ is a normal vector to $\text{gph } \partial_x f$ at (x^ν, u^ν, v^ν) . Rescaling, we can make $|z^\nu| = 1$.

By passing to subsequences, we can suppose z^ν converges to some z with $|z| = 1$ and, as for w^ν , reduce to two cases: either w^ν converges to some w or $0 < |w^\nu| \rightarrow \infty$. In the first case we have in the limit that $(0, w, -z)$ is normal to $\text{gph } \partial_x f$ at $(\bar{x}, \bar{u}, \bar{v})$, so $(0, w) \in D^*(\partial_x f)(\bar{x}, \bar{u} | \bar{v})(z)$. But that is excluded by condition (a) of Theorem 2.3. In the second case, let $\hat{w}^\nu = w^\nu/|w^\nu|$ and $\hat{z}^\nu = z^\nu/|z^\nu|$. Then $\hat{z}^\nu \rightarrow 0$, whereas, by passing once more to subsequences if necessary, we can suppose \hat{w}^ν converges to some \hat{w} with $|\hat{w}| = 1$. We have $(\lambda^\nu \hat{z}^\nu, \hat{w}^\nu, -\hat{z}^\nu)$ normal to $\text{gph } \partial_x f$ at (x^ν, u^ν, v^ν) , and hence in the limit that $(0, \hat{w}, 0)$ is normal to $\text{gph } \partial_x f$ at $(\bar{x}, \bar{u}, \bar{v})$. Then $(0, \hat{w}) \in D^*(\partial_x f)(\bar{x}, \bar{u} | \bar{v})(0)$, but that is impossible under condition (b) of Theorem 2.3.

Turning now to the converse claim at the end of the proposition, we drop the assumption that (a) and (b) of Theorem 2.3 hold and suppose instead that (4.12) is satisfied by ε and a neighborhood $X \times U \times V$. Let $(x', u', v') \in D^*(\partial_x f)(\bar{x}, \bar{u} | \bar{v})(v')$, so that $(x', u', -v')$ is a normal vector to $\text{gph } \partial_x f$ at $(\bar{x}, \bar{u}, \bar{v})$. By definition, then, there exist sequences $(\bar{x}^\nu, \bar{u}^\nu, \bar{v}^\nu) \rightarrow (\bar{x}, \bar{u}, \bar{v})$ in $X \times U \times V$ and $(\tilde{x}^\nu, \tilde{u}^\nu, \tilde{v}^\nu) \rightarrow (x', u', v')$ in which $(\tilde{x}^\nu, \tilde{u}^\nu, -\tilde{v}^\nu)$ is a *regular* normal vector to $\text{gph } \partial_x f$ at $(\bar{x}^\nu, \bar{u}^\nu, \bar{v}^\nu)$. Since $\text{gph } \partial f_u$ is merely the cross section of $\text{gph } \partial_x f$ obtained by fixing the u argument, $(\tilde{x}^\nu, -\tilde{v}^\nu)$ is then a regular normal vector to $\text{gph } \partial f_{\bar{u}^\nu}$ at $(\bar{x}^\nu, \bar{v}^\nu)$. This implies $\tilde{x}^\nu \in D^*(\partial f_{\bar{u}^\nu})(\bar{x}^\nu | \bar{v}^\nu)(\tilde{v}^\nu) = \partial^2 f_{\bar{u}^\nu}(\bar{x}^\nu | \bar{v}^\nu)(\tilde{v}^\nu)$, so $\langle \tilde{x}^\nu, \tilde{v}^\nu \rangle \geq \varepsilon |\tilde{v}^\nu|^2$ by (4.12). Taking the limit we get the inequality in (4.13), as desired. \square

5. Proof of the Main Result

Two auxiliary facts still have to be established in order to set the stage completely for the proof of necessity and sufficiency in Theorem 2.3. We first deal with one needed in the sufficiency argument. We denote by $\mathcal{B}(v, \lambda)$ the closed ball of radius λ around v .

Lemma 5.1 (subgradient inversion estimate). *Let $g : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be convex and let O be an open convex set on which g is finite and strongly convex with modulus μ . Suppose $v_0 \in O$ and $w_0 \in \partial g(v_0)$, and let $\lambda > 0$ be small enough that the $\mathcal{B}(v_0, \lambda)$ lies in O . Then for every $w \in \mathcal{B}(w_0, \lambda\mu)$ there is a unique $v \in \mathcal{B}(v_0, \lambda)$ with $w \in \partial g(v)$. Furthermore, the single-valued mapping $w \mapsto v$ defined in this way is Lipschitz continuous on $\mathcal{B}(w_0, \lambda\mu)$ with constant $1/\mu$.*

Proof. Fix any $\lambda_0 > \lambda$ small enough that $\mathcal{B}(v_0, \lambda_0)$ still lies in O . Define $g_0(v)$ to be $g(v_0 + v) - \langle w_0, v \rangle$ when $v \in \lambda_0 \mathcal{B}$ but ∞ otherwise. Then $\partial g_0(v) = \partial g(v_0 + v) - w_0$ for

$v \in \lambda B$ and in particular $0 \in \partial g_0(0)$. It will suffice to prove that for every $w \in \mu\lambda B$ there is a unique $v \in \lambda B$ with $w \in \partial g_0(v)$, and that the associated mapping $w \mapsto v$ has the Lipschitz property claimed.

Because g is continuous on O by virtue of its convexity, g_0 is lsc on \mathbb{R}^n as well as μ -strongly convex on its effective domain $\lambda_0 B$. Then the subgradient mapping ∂g_0 is μ -strongly monotone, and the conjugate convex function g_0^* is differentiable on \mathbb{R}^n , its gradient mapping being globally Lipschitz continuous with constant $1/\mu$; see [6; 11.13, 12.60]. This makes ∂g_0^* reduce to ∇g_0^* , and since $\partial g_0^* = (\partial g_0)^{-1}$ in general, it follows that we have $v = \nabla g_0^*(w)$ if and only if $w \in \partial g_0(v)$.

Our task reduces to demonstrating that in these circumstances we have $v \in \lambda B$ when $w \in \lambda\mu B$. We know in general from the theory of conjugate functions that

$$\partial g_0^*(w) = \operatorname{argmin}_v \{g_0(v) - \langle v, w \rangle\} = \operatorname{argmin}_v \{g(v_0 + v) - \langle v, w_0 + w \rangle + \delta_{\lambda_0 B}(v)\}.$$

The rules of subgradient calculus tell us that the minimum is attained at v if and only if $0 \in \partial g(v_0 + v) - [w_0 + w] + N_{\lambda_0 B}(v)$. Therefore, $v = \nabla g_0^*(w)$ if and only if $v \in \lambda_0 B$ and there exists $\theta \geq 0$ such that $w_0 + w - \theta v \in \partial g(v_0 + v)$, in which case $w - \theta v \in \partial g_0(v)$. Here necessarily $\theta = 0$ unless $|v| = \lambda_0$. Thus we can finish off by showing that if $w - \theta v \in \partial g_0(v)$ and $|v| = \lambda_0$, then $|w| > \mu\lambda$.

We accomplish this by appealing to the fact that ∂g_0 is μ -strongly monotone with $0 \in \partial g_0(0)$. In combination with the relation $w - \theta v \in \partial g_0(v)$ this yields $\langle w - \theta v, v \rangle \geq \mu|v|^2$, hence $\langle w, v \rangle \geq (\mu + \theta)|v|^2$. That implies $|w| \geq (\mu + \theta)|v| = (\mu + \theta)\lambda_0 > \mu\lambda$. \square

Proof of sufficiency in Theorem 2.3. Assume the hypothesis of Theorem 2.3 along with conditions (a) and (b). Full stability will be demonstrated, and the assertion about $M_\delta(u, v)$ equaling $M(u, v)$ will be obtained as a by-product.

Our assumptions yield the uniform positive definiteness property in Proposition 4.5. In particular, in order to get started, we observe that this implies for the function $f_{\bar{u}}$ that

$$x' \in \partial^2 f_{\bar{u}}(\bar{x} | \bar{v})(v'), v' \neq 0 \implies \langle v', x' \rangle > 0.$$

Since $f_{\bar{u}}$ is continuously prox-regular at \bar{x} for \bar{v} in consequence of the parametric continuous prox-regularity of f , we have everything in place to apply the main result of Poliquin and Rockafellar in [8] and conclude that we at least have tilt stability. All that we really need from this, however, is the fact that, for $\delta > 0$ sufficiently small, we have $M_\delta(\bar{u}, \bar{v}) = \{\bar{x}\}$. Then we can invoke Proposition 3.5 in tandem with (2.7) to see that, for some neighborhood W of (\bar{u}, \bar{v}) , we have m_δ Lipschitz continuous on W and

$$\begin{aligned} \emptyset \neq M_\delta(u, v) &= M_\delta(u, v) \cap \{x \mid |x - \bar{x}| < \delta\} \\ &\subset M(u, v) \cap \{x \mid |x - \bar{x}| < \delta\} \quad \text{for all } (u, v) \in W. \end{aligned} \tag{5.1}$$

We know from Corollary 4.2, on the other hand, that M has the Aubin property at (\bar{u}, \bar{v}) for \bar{x} . The Lipschitz modulus of M at (\bar{u}, \bar{v}) for \bar{x} (in the set-valued sense of the Aubin property—see [6; 9.36]) is given then by the “norm” $|D^*M(\bar{u}, \bar{v})|^+$ in [6; 9.40]. By virtue of the equivalence in (4.1), and this “norm” value can be expressed as the max on the right side of (2.16).

Thus, if we can prove that the mapping $(u, v) \mapsto M(u, v) \cap \{x \mid |x - \bar{x}| < \delta\}$ is single-valued around (\bar{u}, \bar{v}) , it will follow that, on some neighborhood of (\bar{u}, \bar{v}) , this single-valued mapping is Lipschitz continuous and agrees with M_δ , as claimed. Furthermore, we will have the formula in (2.16) for the Lipschitz modulus of M_δ at (\bar{u}, \bar{v}) , and be done.

Everything therefore hinges on establishing this single-valuedness. From [8], as already noted, we already have it for $M(\bar{u}, v)$ as a function of v around \bar{v} . It might seem an easy step to go from that to the local single-valuedness of $M(u, v)$ in v for parameter vectors u near \bar{u} , using the fact the functions f_u , like $f_{\bar{u}}$, exhibit prox-regularity locally by Proposition 3.2, together with the fact that the coderivative Hessians associated with these functions are positive definite by Proposition 4.5. At best, though, we could only get from such an argument a separate domain of single-valuedness of $M(u, v)$ in v for each u , whereas we require that these domains come together as a neighborhood of (\bar{u}, \bar{v}) in (u, v) jointly. That makes everything much more complicated.

Let $X \times U \times V$ be a bounded open neighborhood of $(\bar{x}, \bar{u}, \bar{v})$ small enough to ensure the properties in Proposition 3.2 (for a certain prox-regularity parameter $r \geq 0$) and also the uniform positive definiteness in Proposition 4.5. Suppose further that $U \times V$ is small enough that it lies in the neighborhood W where (5.1) holds. Fix any $s > r$ and let

$$\begin{aligned} \bar{f}(x, u) &:= \begin{cases} f(x, u) & \text{if } |x - \bar{x}| \leq \delta, \\ \infty & \text{if } |x - \bar{x}| > \delta, \end{cases} \\ k(x, u, v) &:= \bar{f}(x, u) - \langle v, x \rangle + (s/2)|x - \bar{x}|^2. \end{aligned} \tag{5.2}$$

Further, in terms of this define

$$\varphi(u, v) := \inf_x k(x, u, v), \quad \Phi(u, v) := \operatorname{argmin}_x k(x, u, v). \tag{5.3}$$

Our first objective is to show by techniques of variational analysis that φ is Lipschitz continuous on a neighborhood of (\bar{u}, \bar{v}) .

To this end we note first that when $(u, v) \in U \times V$ there exists x with $k(x, u, v) < \infty$; indeed, any $x \in M_\delta(u, v)$ has this property, since (5.1) holds and $U \times V \subset W$. Therefore $\varphi < \infty$ on $U \times V$. On the other hand, k is lsc and we have for each $\alpha \in \mathbb{R}$ that the set $\{(v, u, x) \mid (v, u) \in V \times U, k(v, u, x) \leq \alpha\}$ is bounded. This guarantees by the basic

theorem on parametric optimization in [6; 1.17] that φ is lsc on $U \times V$ with $\varphi > -\infty$ and

$$\Phi(u, v) \neq \emptyset \text{ when } (u, v) \in U \times V, \text{ where } \Phi(\bar{u}, \bar{v}) = \{\bar{x}\}. \quad (5.4)$$

Moreover we have then from [6; 10.13] that

$$\begin{aligned} \partial\varphi(u, v) &\subset \{(y, w) \mid \exists x \in \Phi(u, v) \text{ with } (0, y, w) \in \partial k(x, u, v)\}, \\ \partial^\infty\varphi(u, v) &\subset \{(y, w) \mid \exists x \in \Phi(v, u) \text{ with } (0, y, w) \in \partial^\infty k(x, u, v)\}, \end{aligned} \quad (5.5)$$

where we calculate via [6; 8.8(c)] that

$$\begin{aligned} (0, y, w) \in \partial k(x, u, v) &\iff (0, y, w) \in (\partial \bar{f}(x, u), 0) + (s[x - \bar{x}] - v, 0, -x) \\ &\iff (v - s[x - \bar{x}], y) \in \partial \bar{f}(x, u) \text{ and } w = -x, \\ (0, y, w) \in \partial^\infty k(x, u, v) &\iff (0, y) \in \partial^\infty \bar{f}(x, u) \text{ and } w = 0. \end{aligned} \quad (5.6)$$

Applying the last formula to (\bar{u}, \bar{v}) and observing that $\partial^\infty \bar{f}(\bar{x}, \bar{u}) = \partial^\infty f(\bar{x}, \bar{u})$ because $\Phi(\bar{u}, \bar{v}) = \{\bar{x}\}$, we see through the constraint qualification $\mathcal{Q}(\bar{x}, \bar{u})$ that the only choice of (y, w) satisfying $(0, y, w) \in \partial^\infty k(\bar{x}, \bar{u}, \bar{v})$ is $(y, w) = (0, 0)$. The second formula in (5.5) then yields $\partial^\infty\varphi(\bar{u}, \bar{v}) = (0, 0)$. A function is Lipschitz continuous on a neighborhood of any point where it is finite, lsc, has no nonzero horizon subgradient [6; 9.13], so we conclude that φ is Lipschitz continuous around (\bar{u}, \bar{v}) .

Continuity of φ at (\bar{u}, \bar{v}) implies continuity of the set-valued mapping Φ at (\bar{u}, \bar{v}) , where it is single-valued; see [6; 1.17(b)]. Thus, for some open neighborhood $U_0 \times V_0$ of (\bar{v}, \bar{u}) within $U \times V$, which can be taken to be convex, we have

$$\Phi(u, v) \subset \{x \mid |x - \bar{x}| < \delta\} \text{ when } (u, v) \in U_0 \times V_0. \quad (5.7)$$

By choosing $U_0 \times V_0$ even smaller, we can arrange to have the additional property, needed below, that

$$x \in \Phi(u, v) \implies x \in X, \quad v - s[x - \bar{x}] \in V. \quad (5.8)$$

Under (5.7), $\partial \bar{f}(x, u)$ reduces to $\partial f(x, u)$ in (5.6), and we obtain then from (5.5) that

$$\begin{aligned} \partial\varphi(u, v) &\subset \{(y, -x) \mid x \in \Phi(v, u), (v, y) \in \partial f(x, u) + (s[x - \bar{x}], 0)\} \\ &\text{when } (u, v) \in U_0 \times V_0. \end{aligned} \quad (5.9)$$

The Lipschitz continuity of φ on $U_0 \times V_0$ implies that $\partial^\infty\varphi(u, v) = \{(0, 0)\}$ for $(u, v) \in U_0 \times V_0$ [6; 9.13] and allows us to apply the partial subgradient rule in [6; 10.11] to see that $\emptyset \neq \partial_v\varphi(u, v) \subset \{w \mid \exists y \text{ with } (y, w) \in \partial\varphi(u, v)\}$ and get then from (5.9) that

$$\partial_v\varphi(u, v) \subset \{-x \mid x \in \Phi(u, v)\} \text{ when } (u, v) \in U_0 \times V_0. \quad (5.10)$$

Next we determine what it means for x to belong to $\Phi(u, v)$ when $(u, v) \in U_0 \times V_0$. Because of (5.7), the subgradient optimality condition for x to furnish the minimum in (5.3) takes the form of requiring $0 \in \partial_x f(x, u) - v + s[x - \bar{x}]$. Hence

$$\Phi(u, v) \subset \{x \mid v - s[x - \bar{x}] \in \partial_x f(x, u)\} \quad \text{when } (u, v) \in U_0 \times V_0. \quad (5.11)$$

It will be demonstrated that this makes Φ be single-valued. Fix any $(u, v) \in U_0 \times V_0$ and suppose that $x, x' \in \Phi(u, v)$. In particular we have $(x, u, v - s[x - \bar{x}])$ and $(x', u, v - s[x' - \bar{x}])$ in $X \times U \times V$ by (5.8) and therefore by prox-regularity

$$\begin{aligned} f(x', u) &\geq f(x, u) + \langle v - s[x - \bar{x}], x' - x \rangle - \frac{r}{2}|x' - x|^2, \\ f(x, u) &\geq f(x', u) + \langle v - s[x' - \bar{x}], x - x' \rangle - \frac{r}{2}|x' - x|^2, \end{aligned}$$

from which it follows (by adding the two inequalities) that $0 \geq (s - r)|x' - x|^2$. Thus $x' = x$ (inasmuch as $s > r$), and the single-valuedness of Φ is confirmed.

The single-valuedness of Φ on $U_0 \times V_0$ produces the single-valuedness of the mapping $\partial_v \varphi$ on that set by (5.10) and reveals that for each $u \in U_0$ the function $\varphi_u = \varphi(\cdot, u)$ is strictly differentiable with respect to $v \in V_0$ [6; 9.18], in fact with gradient $\nabla \varphi_u(u, v) = -x$ for the unique $x \in \Phi(u, v)$. Strict differentiability at every point of an open set is equivalent to continuous differentiability on that set [6; 9.19].

The achievement so far can be summarized as follows in terms of φ and its ‘‘slices’’ φ_u . We have an open neighborhood $U_0 \times V_0$ of (\bar{v}, \bar{u}) on which φ is finite and Lipschitz continuous and such that, for each $u \in U_0$, φ_u is continuously differentiable on V_0 with

$$\begin{aligned} -\nabla \varphi_u(v) &= \text{unique } x \in \Phi(u, v) \\ &= \text{unique } x \text{ with } |x - \bar{x}| < \delta, \quad v - s[x - \bar{x}] \in \partial f_u(x). \end{aligned} \quad (5.12)$$

In particular, $-\nabla \varphi_{\bar{u}}(\bar{v}) = \bar{x}$.

Keeping u as an arbitrary element of U_0 , let $F_u(v) = -\nabla \varphi_u(v)$ on V_0 for simplicity. Then $F_{\bar{u}}(\bar{v}) = \bar{x}$ and F_u is a continuous, single-valued mapping from V_0 to \mathbb{R}^n with its graph related to that of ∂f_u through (5.12) by

$$\begin{aligned} (v, x) \in \text{gph } F_u &\iff (v, x) \in \Omega, \quad L(v, x) \in \text{gph } f_u, \quad \text{where} \\ \Omega &= V_0 \times \{x \mid |x - \bar{x}| < \delta\}, \quad L(v, x) = (x, v - s[x - \bar{x}]). \end{aligned} \quad (5.13)$$

The affine mapping L is invertible and gives a ‘‘change of coordinates’’ through which normal cones to $\text{gph } F_u$ can be identified with normal cones to $\text{gph } \partial f_u$; by way of the rule in [6; 6.7] we obtain

$$(v', -x') \in N_{\text{gph } F_u}(v, x) \iff (sv' - x', v') \in N_{\text{gph } \partial f_u}(x, v - s[x - \bar{x}])$$

and can write this in coderivative form as

$$v' \in D^*F_u(v|x)(x') \iff sv' - x' \in D^*(\partial f_u)(x|w)(-v') \text{ for } w = v - s[x - \bar{x}]. \quad (5.14)$$

Appealing now to the fact that the pairs (v, x) in this situation have $x \in X$ and $w \in V$ by (5.7), we make use of the uniform positive definiteness of $D^*(\partial f_u)(x|w)$ for such (x, w) (as we arranged by making our neighborhoods be such that (4.12) holds) to see from (5.14) that

$$\begin{aligned} v' \in D^*F_u(v|x)(x') &\implies \langle sv' - x', -v' \rangle \geq \varepsilon|v'|^2 \\ &\implies \langle -x', -v' \rangle \geq s|v'|^2 + \varepsilon|v'|^2 \\ &\implies |x'| |v'| \geq (s + \varepsilon)|v'|^2 \implies |v'| \leq (s + \varepsilon)^{-1}|x'|. \end{aligned}$$

This inequality on the coderivatives of F_u guarantees, in the face of the stipulated convexity of V_0 , that F_u itself is Lipschitz continuous on V_0 with constant $(s + \varepsilon)^{-1}$. That is an immediate outcome of the calculus of the Lipschitz modulus in [6; 9.31, 9.38, 9.40] as specialized to the case of a single-valued mapping like F_u .

We now introduce on V_0 the mapping $G_u : v \mapsto v - s[F_u(v) - \bar{x}]$, noting that $G_u(v) = \nabla_v \psi(u, v)$ for the function $\psi : (u, v) \mapsto \frac{1}{2}|v|^2 + s\varphi(u, v) + s\langle v, \bar{x} \rangle$. The choice of this mapping is motivated by the fact that $w = G_u(v)$ if and only if $w = v - s[x - \bar{x}]$ for the unique x such that $|x - \bar{x}| < \delta$ and $v - s[x - \bar{x}] \in \partial f_u(x)$. Then obviously $w \in \partial f_u(x)$, so that $x \in M(u, w)$. In particular we have $G_{\bar{u}}(\bar{v}) = \bar{v}$. If we can determine a neighborhood V_1 of \bar{v} along with a neighborhood U_1 of \bar{u} such that for each $(u, w) \in U_1 \times V_1$ there is a unique $v \in V_0$ with $G_u(v) = w$, we will be able to conclude that for such (u, w) there is a unique $x \in M(u, w)$ with $|x - \bar{x}| < \delta$. That will confirm that the mapping $(u, w) \mapsto M(u, w) \cap \{x \mid |s - \bar{x}| < \delta\}$ is single-valued on $U_1 \times V_1$, and we will be finished.

Our key to this final stage will be Lemma 5.1. As preparation for using it, we demonstrate that the gradient mapping G_u is strongly monotone: for $v, v' \in V_0$ we have

$$\begin{aligned} \langle G_u(v') - G_u(v), v' - v \rangle &= \langle v' - sF_u(v') + s\bar{x} - v + sF_u(v) - s\bar{x}, v' - v \rangle \\ &= |v' - v|^2 - s\langle F_u(v') - F_u(v), v' - v \rangle \\ &\geq |v' - v|^2 - s|F_u(v') - F_u(v)||v' - v| \\ &\geq |v' - v|^2 - s(s + \varepsilon)^{-1}|v' - v|^2 = \varepsilon(s + \varepsilon)^{-1}|v' - v|^2. \end{aligned}$$

This monotonicity implies that $\psi(u, v)$ is μ -strongly convex in $v \in V_0$ with modulus $\mu = \varepsilon(s + \varepsilon)^{-1}$. Since $\psi(u, v)$ is continuous in $(u, v) \in U_0 \times V_0$ (it inherits this from φ), the vector $G_u(v) = \nabla_v \psi(u, v)$ depends continuously on $(u, v) \in U_0 \times V_0$ as well [16; 25.7].

Take $\lambda > 0$ small enough that $\mathcal{B}(\bar{v}, 2\lambda) \subset V_0$. Let $g_u(v) = \psi(u, v)$ if $v \in \mathcal{B}(\bar{v}, 2\lambda)$ but $g_u(v) = \infty$ otherwise. Then g_u is convex as a function on \mathbb{R}^n and agrees with $\psi(u, \cdot)$ on the open set $O_u = \{v \mid |v - \bar{v}| < 2\lambda\}$. There, g_u is strongly convex with modulus μ , and its gradient mapping is G_u ; the unique subgradient $w \in \partial g_u(v)$ is $w = G_u(v)$ when $v \in O_u$. By virtue of Lemma 5.1, there exists then for each $w \in \mathcal{B}(G_u(\bar{v}), \lambda\mu)$ a unique $v \in \mathcal{B}(\bar{v}, \lambda)$ with $w = G_u(v)$.

All that remains is to observe that by choosing U_1 to be a small enough neighborhood of \bar{u} within U_0 we can obtain (through the continuous dependence of $G_u(\bar{v})$ on u) the existence of a neighborhood V_1 of \bar{v} within V_0 such that, for all $u \in U_1$ we have $\mathcal{B}(G_u(\bar{v}), \lambda\mu) \supset V_1$. \square

In moving on to the necessity in Theorem 2.3, we will have to have help from a different auxiliary result.

Lemma 5.2 (dual criterion for localized strong convexity). *Let $h : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ be a proper, lsc, convex function whose conjugate h^* is differentiable on a certain open convex set $O \subset \mathbb{R}^n$, moreover with its gradient mapping $\nabla h^* : O \rightarrow \mathbb{R}^n$ Lipschitz continuous on O with constant $1/\sigma$ (for some $\sigma > 0$). Let $\lambda > 0$ and $O_\lambda = \{v \mid \mathcal{B}(v, \lambda) \subset O\}$. Then*

$$h(x') \geq h(x) + \langle v, x' - x \rangle + \frac{\sigma}{2}|x' - x|^2 \quad \text{if } v \in \partial h(x) \cap O_\lambda, \quad |x' - x| \leq \frac{\lambda}{\sigma}, \quad (5.15)$$

and therefore also

$$\langle x' - x, v' - v \rangle \geq \sigma|x' - x|^2 \quad \text{whenever } \begin{cases} v \in \partial h(x), & v' \in \partial h(x'), \\ v, v' \in O_\lambda, & |x' - x| \leq \lambda/\sigma. \end{cases} \quad (5.16)$$

Proof. For any $v, v' \in O$ we have $h^*(v') - h^*(v) = \int_0^1 \langle \nabla h^*(v + t[v' - v]), v' - v \rangle dt$. The estimate $\langle \nabla h^*(v + t[v' - v]), v' - v \rangle \leq \langle \nabla h^*(v), v' - v \rangle + (t/\sigma)|v' - v|^2$ holds under our assumptions, so the integral gives us

$$h^*(v') - h^*(v) \leq \langle \nabla h^*(v), v' - v \rangle + \frac{1}{2\sigma}|v' - v|^2.$$

Therefore, in terms of the indicator function $\delta_{\lambda\mathcal{B}}$ of the closed λ -ball around 0 and the function $j(w) = \frac{1}{2}|w|^2$, we have for any choice of $v \in O_\lambda$ that

$$h^*(v') \leq k(v' - v) \quad \text{for all } v' \in \mathbb{R}^n, \quad \text{where} \quad (5.17)$$

$$k(w) := h^*(v) + \langle \nabla h^*(v), w \rangle + \sigma^{-1}j(w) + \delta_{\lambda\mathcal{B}}(w).$$

Fix $v \in O_\lambda$ and take conjugates of both sides of (5.17) as convex functions of v' , using x' as the variable to describe the conjugate functions. That produces the inequality

$$h^{**}(x') \geq k^*(x') + \langle v, x' \rangle \quad \text{for all } x' \in \mathbb{R}^n. \quad (5.18)$$

Here $h^{**} = h$ because h is lsc, proper and convex, and k^* calculates to

$$k^*(x') = -h^*(v) + (\sigma^{-1}j + \delta_{\lambda B})^*(x' - \nabla h^*(v)).$$

The function conjugate to $\sigma^{-1}j$ is σj and the function conjugate to $\delta_{\lambda B}$ is $\lambda|\cdot|$, and consequently $(\sigma^{-1}j + \delta_{\lambda B})^* = \sigma j \# \lambda|\cdot|$, with “ $\#$ ” denoting the operation of epi-addition (inf-convolution):

$$(\sigma j \# \lambda|\cdot|)(u) = \inf_{u'} \{ \sigma j(u') + \lambda|u - u'| \} = \begin{cases} \sigma j(u) & \text{when } |u| \leq \lambda/\sigma, \\ \lambda(\sigma^{-1} + |u|) & \text{when } |u| \geq \lambda/\sigma. \end{cases} \quad (5.19)$$

Let $x = \nabla h^*(v)$; this relation is the same as $x \in \partial h^*(v)$ when h^* is differentiable at v and hence is equivalent also to $v \in \partial h(x)$ as well as to $h(x) + h^*(v) = \langle x, v \rangle$ (by convex analysis; cf. [6; 11.3]). We obtain from (5.18) and our calculations that

$$h(x') \geq h(x) + \langle v, x' - x \rangle + (\sigma j \# \lambda|\cdot|)(x' - x) \text{ for all } x' \in \mathbb{R}^n.$$

This yields (5.15) through (5.19). By symmetry, of course, we also have

$$h(x) \geq h(x') + \langle v', x - x' \rangle + \frac{\sigma}{2}|x - x'|^2 \text{ if } v' \in \partial h(x') \cap O_\lambda, \quad |x - x'| \leq \frac{\lambda}{\sigma}.$$

In combining this inequality with the one in (5.15) we obtain (5.16). □

Proof of necessity in Theorem 2.3. The hypothesis furnishes for us a neighborhood $X \times U \times V$ of $(\bar{x}, \bar{u}, \bar{v})$ for which the properties in Proposition 3.2 hold. An additional assumption now is that, for some $\delta > 0$ sufficiently small, the mapping M_δ is single-valued and Lipschitz continuous around (\bar{u}, \bar{v}) , its value at (\bar{u}, \bar{v}) being \bar{x} . Without loss of generality we can suppose these properties hold for M_δ on $U \times V$, and that

$$M_\delta(u, v) \in \{x \mid |x - \bar{x}| < \delta\} \subset X \text{ for } (u, v) \in U \times V. \quad (5.20)$$

We can also arrange that (5.1) holds for $W = U \times V$, through Proposition 3.5 and (2.7).

Define \bar{f} , k , φ and Φ as in (5.2) and (5.3) but with $s = 0$, so $\varphi = m_\delta$ and $\Phi = M_\delta$. The subgradient calculus used in the proof of sufficiency after those definitions remains valid and reveals that φ , which is Lipschitz continuous on an open convex neighborhood $U_0 \times V_0$, say, of (\bar{u}, \bar{v}) in $U \times V$, exhibits as instances of (5.10) and (5.11) the relations

$$\begin{aligned} \partial_v \varphi(u, v) &= -M_\delta(u, v), \\ M_\delta(u, v) &\in \{x \mid v \in \partial_x f(x, u)\} = M(u, v). \end{aligned} \quad (5.21)$$

The first of these implies moreover that for each $u \in U_0$ the function $\varphi_u = \varphi(u, \cdot)$ is continuously differentiable on V_0 with gradient $\nabla\varphi_u(v) = -M_\delta(u, v)$. In fact our Lipschitz assumption on M_δ gives us a constant $\kappa > 0$ such that for each $u \in U_0$ the mapping $\nabla\varphi_u$ is Lipschitz continuous on V_0 with constant κ .

Let $g_u = -\varphi_u$, so that $g_u(v) = \sup_x \{\langle v, x \rangle - \bar{f}(x, u)\}$, or in other words, g_u is conjugate to \bar{f}_u under the Legendre-Fenchel transform. In particular, g_u is a proper, lsc, convex function on \mathbb{R}^n that is differentiable on V_0 with $\nabla g_u(v) = M_\delta(u, v)$. Let h_u be conjugate in turn to g_u . Then $h_u = g_u^* = \bar{f}_u^{**}$ and $g_u = h_u^* = \bar{f}_u^*$, and we have by the usual relation between subgradients of conjugate convex functions that $v \in \partial h_u(x)$ if and only if $x \in \partial g_u(v)$, so that

$$v \in \partial h_u(x) \iff x = \nabla g_u(v) = M_\delta(u, v), \quad \text{as long as } u \in U_0, v \in V_0. \quad (5.22)$$

We apply Lemma 5.2 now to h_u and its conjugate function g_u on the set $O = V_0$ with $1/\sigma = \kappa$. Let $\lambda > 0$ be small enough that $\mathcal{B}(\bar{v}, \lambda) \subset V_0$, so the set $O_\lambda = \{v \mid \mathcal{B}(\bar{v}, \lambda) \subset V_0\}$ is an open neighborhood of \bar{v} . Then (5.16) holds for h_u , where by (5.24) the relations $v \in \partial h(x)$ and $v' \in \partial h(x')$ can be written as $x = M_\delta(u, v)$ and $x' = M_\delta(u, v')$.

Choose X_1 to be a neighborhood of \bar{x} within X so small that $|x' - x| \leq \lambda/\sigma$ when $x, x' \in X_1$. Let $U_1 \times V_1$ be a neighborhood of (\bar{u}, \bar{v}) within $U_0 \times O_\lambda$ small enough that $(u, v) \in U_1 \times V_1$ implies $M_\delta(u, v) \in X_1$. Then (5.16) yields the inequality

$$\langle x' - x, v' - v \rangle \geq \sigma |x' - x|^2 \quad \text{when} \quad \begin{cases} x = M_\delta(u, v), & x' = M_\delta(u, v'), \\ u \in U_1, & v, v' \in V_1. \end{cases} \quad (5.23)$$

In terms of the mapping T_u obtained by restricting $M_\delta(u, \cdot)$ to V_1 , (5.23) says that T_u^{-1} is strongly monotone with constant σ . Let S_u be the mapping whose graph is the intersection of $\text{gph } M(u, \cdot)$ with $V_1 \times \{x \mid |x - \bar{x}| < \delta\}$, so that S_u^{-1} is the mapping whose graph is the intersection of $\text{gph } \partial f_u$ with $\{x \mid |x - \bar{x}| < \delta\} \times V_1$. We have $\text{gph } T_u \subset \text{gph } S_u$ by (5.20) and the second relation in (5.21), hence also $T_u^{-1}(x) \subset S_u^{-1}(x) \subset \partial f_u(x)$ for all x .

For the constant r in the prox-regularity of f , we know that the mappings ∂f_u are monotone when $u \in U_1$. Let $s > r$ and consider the mappings $T_u^{-1} + sI$ and $S_u^{-1} + sI$. As long as $u \in U_1$, both of these are strongly monotone, the first with constant $\sigma + s$ and the second surely with constant $s - r$. Hence the inverses $(T_u^{-1} + sI)^{-1}$ and $(S_u^{-1} + sI)^{-1}$ are single-valued on their domains. Because $\text{gph } T_u^{-1} \subset \text{gph } S_u^{-1}$, we have $\text{gph}(T_u^{-1} + sI)^{-1} \subset \text{gph}(S_u^{-1} + sI)^{-1}$, so it follows that

$$x \in (T_u^{-1} + sI)^{-1}(z) \implies (T_u^{-1} + sI)^{-1}(z) = (S_u^{-1} + sI)^{-1}(z) = \{x\}. \quad (5.24)$$

Expressing z in the form $v + sx$, we find that this means

$$x \in (T_u^{-1} + sI)^{-1}(v + sx) \iff v + sx \in (T_u^{-1} + sI)(x) \iff x \in T_u(v),$$

and similarly with S_u substituted for T_u . Thus, (5.24) asserts that whenever $x \in T_u(v)$ we have $T_u(v) = S_u(v) = \{x\}$. This has been established for arbitrary $u \in U_1$, so in recalling the definitions of T_u and S_u we are able to conclude that

$$M(u, v) \cap \{x \mid |x - \bar{x}| < \delta\} = M_\delta(u, v) \text{ for all } (u, v) \in U_1 \times V_1. \quad (5.25)$$

This localization of M therefore inherits the Lipschitz continuity of M_δ . Hence in particular the Aubin property holds for M at (\bar{u}, \bar{v}) for \bar{x} . That implies by Proposition 4.1 that condition (b) of Theorem 2.3 must hold. Furthermore, in terms of the inverse mappings, (5.25) states that $\text{gph } T_u = [\{x \mid |x - \bar{x}| < \delta\} \times V_1] \cap \text{gph } \partial f_u$ when $u \in U_1$. This reveals that the coderivatives of these truncated mappings must coincide: $DT_u(x|v) = D^*(\partial f_u)(x|v) = \partial^2 f_u(x|v)$ at the common graph points (x, v) . Because T_u is strongly monotone with constant σ we have $\langle x', v' \rangle \geq \sigma|v'|^2$ for $x' \in DT_u(x|v)(v')$, hence likewise

$$\langle x', v' \rangle \geq \sigma|v'|^2 \text{ for } x' \in \partial^2 f_u(x|v)(v') \text{ when } v \in \partial f_u(x) = \partial_x f(x, u),$$

provided that $(u, v) \in U_1 \times V_1$. That guarantees through the converse part of Proposition 4.5 that the positive definiteness condition (a) holds in Theorem 2.3. \square

Proof of Theorem 2.7. This is really just an extension of the proof of necessity in Theorem 2.3. That proof utilized the function \bar{f} in (5.2) and, in terms of $\bar{f}_u = \bar{f}(\cdot, u)$, introduced the conjugate functions $g_u = \bar{f}^* = -m_\delta(u, \cdot)$ and $h_u = g_u^* = \bar{f}_u^{**}$. The conjugacy relations imply in turn that $h_u^* = g_u$ and also that $h_u(x) = \infty$ when $|x - \bar{x}| > \delta$, since $\bar{f}_u(x)$ has this property by definition. Hence

$$g_u(v) = \sup_{x \in \mathbf{R}^n} \{\langle v, x \rangle - h_u(x)\} = \sup_{|x - \bar{x}| \leq \delta} \{\langle v, x \rangle - h_u(x)\}, \quad (5.26)$$

where the maximum is attained at x if and only if $v \in \partial h_u(x)$.

Take $\hat{f}(x, u) = h_u(x)$. Then $\partial_x \hat{f}(x, u) = \partial h_u(x)$, and since $g_u(v) = -m_\delta(u, v)$ the conjugacy formula $h_u(x) = \sup_v \{\langle v, x \rangle - g_u(x)\}$ converts to

$$\hat{f}(x, u) = \sup_v \{\langle v, x \rangle + m_\delta(u, v)\}, \quad (5.27)$$

while from (5.26) we get

$$\begin{aligned} \inf_{|x - \bar{x}| \leq \delta} \{\hat{f}(x, u) - \langle v, x \rangle\} &= m_\delta(u, v), \\ \operatorname{argmin}_{|x - \bar{x}| \leq \delta} \{\hat{f}(x, u) - \langle v, x \rangle\} &= \{x \mid v \in \partial_x \hat{f}(x, u)\}. \end{aligned} \quad (5.28)$$

For the problems $\widehat{\mathcal{P}}(u, v)$, the expressions on the left of (5.28) are $\widehat{m}_\delta(u, v)$ and $\widehat{M}_\delta(u, v)$. On the other hand, according to (5.22), the right side of the second equation in (5.28) gives $M_\delta(u, v)$ when (u, v) lies in a certain neighborhood $U_0 \times V_0$ of (\bar{u}, \bar{v}) . Therefore, $\widehat{m}_\delta(u, v)$ and $\widehat{M}_\delta(u, v)$ agree with $m_\delta(u, v)$ and $M_\delta(u, v)$ around (\bar{u}, \bar{v}) ; thus, \widehat{f} is equivalent to f in the sense described in Theorem 2.7.

Furthermore, \widehat{f} is lsc on $\mathbb{R}^n \times \mathbb{R}^d$ by (5.27) because m_δ is lsc on $\mathbb{R}^d \times \mathbb{R}^n$, that being true since m_δ is a special case of the function φ defined in (5.3) through (5.2) (namely for $s = 0$), and φ was shown to be lsc in the argument leading up to (5.4). In addition we have, for $(x, u, v) \in \text{gph } \partial_x \widehat{f}$ with $(u, v) \in U_0 \times V_0$, that

$$\widehat{f}(x, u) - \langle v, x \rangle = m_\delta(u, v) = f(x, u) - \langle v, x \rangle$$

and consequently $\widehat{f}(x, u) = f(x, u) = m_\delta(u, v) + \langle v, x \rangle$, an expression that is continuous with respect to the elements (x, u, v) in question. The convexity of $\widehat{f}(x, u)$ in x combined with that continuity makes \widehat{f} be continuously prox-regular at (\bar{x}, \bar{u}) for \bar{v} . (Convexity allows the constant r in the definition of prox-regularity to be taken to be 0.) \square

References

1. S. M. Robinson, “Strongly regular generalized equations,” *Math. Oper. Research* 5 (1980), 43–62.
2. D. Klatte and B. Kummer, “Strong stability in nonlinear programming revisited,” *J. Austral. Math. Soc. Ser. B* 40 (1999), 336–352.
3. A. Dontchev and R. T. Rockafellar, “Characterizations of Lipschitzian stability in nonlinear programming,” in *Mathematical Programming With Data Perturbations* (A. V. Fiacco, ed.) Marcel Dekker, 1997, 65–82.
4. J.-P. Aubin, “Lipschitz behavior of solutions to convex minimization problems,” *Math. Oper. Research* 9 (1984), 87–111.
5. B. S. Mordukhovich, “Sensitivity analysis in nonsmooth optimization,” in *Theoretical Aspects of Industrial Design* (D. A. Field and V. Komkov, eds.), SIAM Volumes in Applied Mathematics No. 58, SIAM Publications, 1992, 32–46.
6. R. T. Rockafellar and R. J-B Wets, *Variational Analysis*, Grundlehren der Mathematik No. 317, Springer-Verlag, 1997.
7. A. Dontchev and R. T. Rockafellar, “Characterizations of strong regularity for variational inequalities over polyhedral convex sets,” *SIAM J. Optim.* 6 (1996), 1121–1137.

8. R. A. Poliquin and R. T. Rockafellar, “Tilt stability of a local minimum,” *SIAM J. Optim.* 8 (1998), 287–299.
9. R. A. Poliquin and R. T. Rockafellar, “Prox-regular functions in variational analysis,” *Trans. Amer. Math. Soc.* 348 (1996), 1805–1838.
10. R. A. Poliquin and R. T. Rockafellar, “Amenable functions in optimization,” in *Non-smooth Optimization Methods and Applications* (F. Giannessi, ed.), Gordon and Breach, Philadelphia, 1992, 338–353.
11. A. B. Levy and R. T. Rockafellar, “Variational conditions and the proto-differentiation of partial subgradient mappings,” *Nonlin. Analysis: Th. Meth. & Appl.* 26 (1996), 1951–1964
12. B. S. Mordukhovich, “Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions,” *Trans. Amer. Math. Soc.*, **340** (1993), 1–36.
13. B. S. Mordukhovich, “Lipschitzian stability of constraint systems and generalized equations,” *Nonlinear Analysis Th. Meth. Appl.* **22** (1994), 173–206.
14. B. S. Mordukhovich, “Generalized differential calculus for nonsmooth and set-valued mappings,” *J. Math. Anal. Appl.* 183 (1994), 250–288.
15. R. T. Rockafellar and D. Zagrodny, “A derivative-coderivative inclusion in second-order nonsmooth analysis,” *Set-Valued Analysis* 5 (1997), 89–105.
16. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.