Vietnam Journal of Mathematics 2019
http://dx.doi.org/10.1007/s10013-019-00339-5

VARIATIONAL CONVEXITY AND THE LOCAL MONOTONICITY OF SUBGRADIENT MAPPINGS

R. Tyrrell Rockafellar[1]

Abstract

It is well known that the subgradient mapping associated with a lower semicontinuous function is maximal monotone if and only if the function is convex, but what characterization can be given for the case in which a subgradient mapping is only maximal monotone locally instead of globally? That question is answered here in terms of a condition more subtle than local convexity. Applications are made to the tilt stability of a local minimum and to the local execution of the proximal point algorithm in optimization.

Keywords: *second-order variational analysis, local maximal monotonicity, variational convexity, local optimality, tilt stability, proximal point algorithm.*

[1]University of Washington, Department of Mathematics, Box 354350, Seattle, WA 98195-4350;
E-mail: *rtr@uw.edu*, URL: www.math.washington.edu/∼rtr/mypage.html

1 Introduction

For a lower semicontinuous (lsc) function $f : I\!R^n \to (-\infty, \infty]$, $f \not\equiv \infty$, a *regular* (or "Fréchet") subgradient at a point $x \in \operatorname{dom} f$ in the terminology of the book [16] is a vector v such that

$$f(x') \geq f(x) + v{\cdot}(x' - x) + o(|x' - x|), \qquad \text{notation:} \quad v \in \widehat{\partial} f(x), \tag{1.1}$$

while a *general* (or "limiting") subgradient at x is a vector v for which

$$\exists v^\nu \in \widehat{\partial} f(x^\nu) \text{ such that } v^\nu \to v,\ x^\nu \to x,\ f(x^\nu) \to f(x), \qquad \text{notation:} \quad v \in \partial f(x). \tag{1.2}$$

The set-valued *subgradient mapping* $\partial f : x \to \partial f(x)$ is the vehicle for many properties of importance in variational analysis. One of those properties is monotonicity in the sense of Minty, according to which a set-valued mapping $T : I\!R^n \rightrightarrows I\!R^n$ is *monotone* if

$$(v' - v){\cdot}(x' - x) \geq 0 \text{ for all } (x, v),\ (x', v') \in \operatorname{gph} T, \tag{1.3}$$

and *maximally* monotone if, in addition,

$$\nexists \text{ monotone } T' : I\!R^n \rightrightarrows I\!R^n \text{ with } \operatorname{gph} T' \supset \operatorname{gph} T \text{ but } \operatorname{gph} T' \neq \operatorname{gph} T. \tag{1.4}$$

For the case of $T = \partial f$, it was observed very early by Moreau [9] that maximal monotonicity holds when f is convex. Much later, Poliquin [11] established the converse in terms of subgradients beyond convex analysis, namely that maximal monotonicity of ∂f requires the convexity of f. In terms of subgradients, of course, this is characterized by ∂f coinciding with $\hat{\partial} f$ and being given by the subgradient inequality of convex analysis, which is (1.1) without the $o(|x' - x|)$ term.

All this carries over also to *strong* monotonicity with modulus $\sigma > 0$ is defined as in (1.3) but with the condition $(v' - v){\cdot}(x' - x) \geq 0$ being strengthened to $(v' - v){\cdot}(x' - x) \geq \sigma ||x' - x||^2$. Maximal strong monotonicity again refers to the graphical condition in (1.4) with strong monotonicity replacing monotonicity. This is equivalent to $T - \mu I$ being monotone or maximal monotone. Strong monotonicity of ∂f with modulus σ corresponds to strong convexity of f with the same modulus, cf. [16, page 565], which is equivalent to $f(x) - \frac{\sigma}{2}||x||^2$ being convex, or the subgradients $v \in \partial f(x)$ satisfying the strengthened inequality

$$f(x') \geq f(x) + v{\cdot}(x' - x) + \frac{\sigma}{2}||x' - x||^2. \tag{1.5}$$

However, monotonicity can also be articulated just *locally* for a mapping T around a pair $(\bar{x}, \bar{v})$ in $\operatorname{gph} T$, namely with respect to the existence of a neighborhood of $(\bar{x}, \bar{v})$ such that (1.3) holds when the pairs (x, v), (x', v') are restricted to that neighborhood, and likewise with maximality. This idea was pursued by Pennanen [10] in extensions of the proximal point algorithm of [14]. Local strong monotonicity of T can be defined similarly with $(v' - v){\cdot}(x' - x) \geq \sigma ||x' - x||^2$.

What are the consequences of such local monotonicity in the case of a subgradient mapping $T = \partial f$? Up to now they haven't been pinned down, and that's the challenge we intend to make progress on here. The issue is of basic theoretical interest in variational analysis but also comes up in trying to understand the extent to which Pennanen's localized extensions of the proximal point algorithm can be executed by minimization steps when applied to $T = \partial f$.

Local convexity in the sense of f being convex relative to some neighborhood X of $\bar{x}$ would of course imply the maximal monotonicity of ∂f with respect to $X \times I\!R^n$, but such local convexity can't be the characterization of local maximal monotonicity of ∂f in general. This is clear from simple one-dimensional examples like

$$f(x) = \max\{1 - e^x,\, 1 - e^{-x}\}, \quad \partial f(x) = \begin{cases} \{-e^x\} & \text{when } x < 0, \\ [-1,1] & \text{when } x = 0, \\ \{e^{-x}\} & \text{when } x > 0. \end{cases} \tag{1.6}$$

which is concave on $(-\infty, 0)$ and on $(0, \infty)$ but has $0 \in \partial f(0)$ and a neighborhood of $(0,0)$ where $\operatorname{gph} \partial f$ is reduced to a vertical segment. That segment exhibits maximal strong monotonicity for arbitrarily high σ.

We have been speaking in simple terms about taking $T = \partial f$ and then localizing by intersecting $\operatorname{gph} T$ with a neighborhood of one of its elements $(\bar{x}, \bar{v})$, but we really need to proceed more delicately in dealing with subgradients because of the role of function values in the definition of $\partial f(x)$ in (1.2). This is illustrated by another elementary one-dimensional example:

$$\begin{aligned} &f(x) = 0 \ \text{ when } \ x = 0 \ \text{ but } \ f(x) = 1 \ \text{ when } \ x \neq 0, \ \text{ for which} \\ &\operatorname{gph} \partial f \ \text{ is the union of the horizontal and vertical axes of } I\!R^2, \end{aligned} \tag{1.7}$$

and in particular $0 \in \partial f(0)$. Even though f has its global mininum attained uniquely at 0, there's no local monotonicity of ∂f around $(0,0)$. However, this trouble is artificial in a picture of minimization. For that, we shouldn't be taking all of $\operatorname{gph} \partial f$ into account in localizing arount $(0,0)$ but only the portion where $f(x) < f(0) + \varepsilon$ for some $\varepsilon > 0$. With this restriction the set of pairs (x, v) reduces to just the vertical axis of $I\!R^2$, which is the graph of a globally maximal monotone mapping. These considerations suggest the following definition.

Definition 1 (f-local monotonicity of a subgradient mapping). *For lsc $f : I\!R^n \to (-\infty, \infty]$ the mapping $\partial f : I\!R^n \rightrightarrows I\!R^n$ will be called f-locally monotone around $(\bar{x}, \bar{v})$ if there is a neighborhood $X \times V$ of $(\bar{x}, \bar{v})$ such that*

$$[X_\varepsilon \times V] \cap \operatorname{gph} \partial f \ \textit{ is monotone in } \ X \times V \tag{1.8}$$

for some $\varepsilon > 0$, where

$$X_\varepsilon = \{\, x \in X \mid f(x) < f(\bar{x}) + \varepsilon \}. \tag{1.9}$$

The same pattern defines f-local maximal monotonicity, f-local strong monotonicity, and f-local maximal strong monotonicity.

The distinction between asserting monotonicity in $X \times V$ instead of just in $X_\varepsilon \times V$ comes out only when maximality is demanded, of course. Maximality with respect to $X \times V$ opens up more potential conflicts than with respect to $X_\varepsilon \times V$. Anyway, the latter might not even be a neighborhood of $(\bar{x}, \bar{v})$, as seen in example (1.7), where $X_\varepsilon \times V$ for small ε would be contained in the vertical axis of $I\!R^2$. The distinction falls away, obviously, if f is *subdifferentially continuous* at $\bar{x}$ in the sense that

$$(x^\nu, v^\nu) \in \operatorname{gph} \partial f, \ (x^\nu, v^\nu) \to (\bar{x}, \bar{v}) \quad \Longrightarrow \quad f(x^\nu) \to f(\bar{x}). \tag{1.10}$$

That property, introduced in [12], has appeared widely in circumstances of second-order variational analysis, but we wish to avoid imposing it here.

The concept we feature in our effort to capture the impact of f-local monotonicity of subgradient mappings is the following, where the notation in (1.9) for an *f-truncated neighborhood* of $\bar{x}$ will again be used — and subsequently many times as well.

Definition 2 (variational convexity). *The lsc function f will be called variationally convex at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ if for some open convex neighborhood $X \times V$ of $(\bar{x}, \bar{v})$ there is a convex lsc function $\widehat{f} \leq f$ on X such that, for some $\varepsilon > 0$,*

$$[X_\varepsilon \times V] \cap \operatorname{gph} \partial f = [X \times V] \cap \operatorname{gph} \partial \widehat{f} \;\text{ and }\; f(x) = \widehat{f}(x) \;\text{ at the common elements }\; (x, v). \quad (1.11)$$

It will be called variationally strongly convex at $\bar{x}$ for $\bar{v}$ with modulus $\sigma > 0$ if this holds with $\widehat{f}$ strongly convex on X with that modulus.

Variational convexity is new in its general form, but the strong version was utilized, without a name, in [13] together with the subdifferential continuity in (1.10) (which obviated putting X_ε in place of X). More will be said about that later in this introduction.

Observe that variational convexity reduces to local convexity over X when $V = I\!R^n$, but not necessarily otherwise. In example (1.6), for instance, condition (1.11) holds at $\bar{x} = 0$ for $\bar{v} = 0$ with $\widehat{f}(x) = \lambda|x|$ and any $\lambda \in (0, 1)$, the neighborhoods X and V being intervals adjusted for the size of λ. The same thing works also in example (1.7).

The following theorems present our main results about local monotonicity of subgradient mappings. The proofs will presented in Section 2. In these results we are obliged to assume, for a technical reason having to do with the argument to be given, that $\bar{v} \in \hat{\partial} f(\bar{x})$ instead of just $\bar{v} \in \partial f(\bar{x})$. However, we have no counterexample showing why the results might not apply to the more general situation.

Certainly there's no difference between $\bar{v} \in \partial f(\bar{x})$ and $\bar{v} \in \hat{\partial} f(\bar{x})$ if $\partial f(\bar{x}) = \hat{\partial} f(\bar{x})$, which is very common in applications and holds in particular when f is *subdifferentially regular* at $\bar{x}$ [16, 7.26]. Subdifferential regularity can conveniently be confirmed by various rules of subgradient calculus in [16] in the presence of constraint qualifications applicable to the specific structure of f.

Theorem 1 (characterization of f-local monotonicity). *The following are equivalent for the lsc function f when $\bar{v} \in \hat{\partial} f(\bar{x})$:*

(a) *∂f is monotone f-locally around $(\bar{x}, \bar{v})$,*
(a′) *∂f is maximal monotone f-locally around $(\bar{x}, \bar{v})$,*
(b) *f is variationally convex at $\bar{x}$ for $\bar{v}$,*
(c) *there is a convex neighborhood $X \times V$ of $(\bar{x}, \bar{v})$ along with $\varepsilon > 0$ such that*

$$(x, v) \in [X_\varepsilon \times V] \cap \operatorname{gph} \partial f \implies f(x') \geq f(x) + v \cdot (x' - x) \;\text{ for all }\; x' \in X. \quad (1.12)$$

Theorem 2 (characterization of f-local strong monotonicity). *The following are equivalent for the lsc function f when $\bar{v} \in \hat{\partial} f(\bar{x})$:*

(a) *∂f is strongly monotone f-locally around $(\bar{x}, \bar{v})$ with modulus $\sigma > 0$,*
(a′) *∂f is maximal strongly monotone f-locally around $(\bar{x}, \bar{v})$ with modulus $\sigma > 0$,*
(b) *f is variationally strongly convex at $\bar{x}$ for $\bar{v}$ with modulus $\sigma > 0$,*
(c) *there is a convex neighborhood $X \times V$ of $(\bar{x}, \bar{v})$ along with $\varepsilon > 0$ such that*

$$(x, v) \in [X_\varepsilon \times V] \cap \operatorname{gph} \partial f \implies f(x') \geq f(x) + v \cdot (x' - x) + \frac{\sigma}{2} ||x' - x||^2 \;\text{ for all }\; x' \in X. \quad (1.13)$$

These results in the case of $\bar{v} = 0$ reveal a deep connection between f having a local minimum at $\bar{x}$ and a number of other valuable properties. Having $0 \in \partial f(\bar{x})$ is a first-order necessary condition for such a local minimum, but the generally stronger property of having $0 \in \hat{\partial} f(\bar{x})$ is actually necessary as well. These coincide when f is convex and then are sufficient for global optimality. In the absence of global convexity, which can be seen as a *global second-order* property, some kind of *local second-order* property of f can be contemplated as furnishing, in tandem with $0 \in \partial f(\bar{x})$ or $0 \in \hat{\partial} f(\bar{x})$, a sufficient condition for f to have a local minimum at $\bar{x}$. In fact variational convexity of f at $\bar{x}$ for $\bar{v}=0$ is just such a condition. It identifies the first-order necessary conditions on f with $0 \in \partial \widehat{f}(\bar{x})$ for a convex function $\widehat{f} \leq f$ on X that agrees with f at $\bar{x}$, thus entailing a global minimum of $\widehat{f}$ at $\bar{x}$ which ensures a local minimum of f at $\bar{x}$.

An interesting aspect is that this approach to a second-order sufficient condition for local optimality doesn't demand an appeal to the theory of generalized second-derivatives of nonsmooth functions, such as has been laid out in [16, Chapter 13]. Instead it promotes, through equivalence also with (a) of Theorem 1 under the necessary condition $0 \in \hat{\partial} f(\bar{x})$, the analogy that local monotonicity of ∂f around $(\bar{x}, 0)$ is the natural replacement for the global monotonicity of ∂f in assuring local optimality.

The implications of Theorem 1(c) when $\bar{v}=0$ extend to more than just having a local minimum at $\bar{x}$. They also concern pairs $(x, v) \in \operatorname{gph} \partial f$ near to $(\bar{x}, 0)$ with $f(x) < f(\bar{x}) + \varepsilon$, saying that

$$f(x') - v \cdot x' \geq f(x) - v \cdot x \;\text{ for }\; x' \in X. \tag{1.14}$$

This relation can also be expressed as

$$x \in \operatorname*{argmin}_{x' \in X} \{\, f(x') - v \cdot (x' - \bar{x}) \} \tag{1.15}$$

in which the subtraction from f of the affine term (which is set up to vanish at $\bar{x}$) is a so-called *tilt perturbation* of f with respect to it having a local minimum at $\bar{x}$. From the equivalence of conditions (c) and (b) in Theorem 1 it's evident that this kind of perturbed local minimization of f coincides with the same thing for the convex function $\widehat{f}$:

$$\operatorname*{argmin}_{x' \in X} \{\, f(x') - v \cdot (x' - \bar{x}) \} = \operatorname*{argmin}_{x' \in X} \{\, \widehat{f}(x') - v \cdot (x' - \bar{x}) \} \tag{1.16}$$

for v in the range of the truncated mapping that has $[X_\varepsilon \times V] \cap \operatorname{gph} \partial f$ as its graph. The argmin is therefore a convex set (under the convexity of X) and behaves in the manner of convex minimization, indeed, through Theorem 1(a′), with the dependence of the argmin on v being maximal monotone.[2]

Tilt perturbations of a local minimum of f at $\bar{x}$ are even more important in appreciating the content of Theorem 2, which touches on the following concept, already much studied.

Definition 3 (tilt stability) [13]**.** *The lsc function f has a tilt-stable local minimum at $\bar{x}$ if there are neighborhoods X of $\bar{x}$ and V of $\bar{v}=0$ such that the mapping M defined by*

$$M(v) = \operatorname*{argmin}_{x' \in X} \{\, f(x') - v \cdot (x' - \bar{x}) \} \;\text{ for }\; v \in V \tag{1.17}$$

is single-valued and Lipschitz continuous with $M(0) = \bar{x}$.

Theorem 3 (criteria for tilt stability)**.** *The properties in Theorem 2 guarantee, in the case of $\bar{v}=0$, that f has a tilt-stable local minimum at $\bar{x}$ with σ^{-1} as the Lipschitz modulus for the mapping M, which is locally maximal monotone as well.*

In summary, just as variational convexity at $\bar{x}$ for $\bar{v}=0$ constitutes what we can call the *variational second-order sufficient condition* for a local minimum of f at $\bar{x}$, associated with the properties in Theorem 1, variational *strong* convexity at $\bar{x}$ for $\bar{v}=0$ furnishes the corresponding variational *strong* second-order sufficient condition associated with the properties in Theorems 2 and 3.

[2]In fact it is maximal *cyclically* monotone, since such monotonicity characterises the subgradient mappings of lsc convex functions [16, 12.25].

Theorem 3 doesn't claim that tilt stability is actually equivalent to the properties in Theorem 2, but just that it is implied by them. The original results about tilt stability in [13] identified it with properties (a′) and (b) of Theorem 2, but they required f not only to be subdifferentially continuous, but also to be *prox-regular* at $\bar{x}$ for $\bar{v} = 0$. Prox-regularity, introduced in [12], is equivalent to local *hypomonotonicity*, namely a property defined like strong maximal monotonicity but with $\sigma < 0$ instead of > 0; it is thus weaker than local maximal monotonicity of ∂f and thus is entailed by the conditions in Theorem 1. See [16, 13F] for a discussion of both subdifferential continuity and prox-regularity, which later proved important also in going beyond tilt stability to *full stability*, where tilt perturbations are augmented by perturbations built into f through the introduction of other parameters [5, 6, 7].

Connections between tilt stability and the quadratic growth condition in Theorem 2(c) have also attracted much interest, as in seen in [2, 3, 8], although not with the generality of our function f. The central theme of such work, some of it infinite-dimensional or avoiding the assumption of subdifferential continuity, was determining whether a localization of $\operatorname{gph}\partial f$ had, as its inverse, a single-valued Lipschitz continuous mapping — a property called *strong metric regularity.*

That topic has more recently been addressed also in the 2017 book of Ioffe [4, Section 9.4.3] in terms of a condition resembling (c) of Theorem 2, but with a key difference. He refers to $\bar{x}$ as a *stable strong local minimizer* of f if (in our notation) there exist neighborhoods X of $\bar{x}$ and V of $\bar{v} = 0$ along with $\sigma > 0$ such that

$$\forall v \in V,\ \exists x_v \in X \ \text{ such that } \ f(x') \geq f(x) + v{\cdot}(x' - x_v) + \frac{\sigma}{2}||x' - x_v||^2 \ \text{ for all } \ x' \in X. \tag{1.18}$$

Without assuming any more than the lower semicontinuity of f, he shows that this property is equivalent to the tilt stability of f at $\bar{x}$. Although (1.18) entails having $v \in \partial f(x)$, there is no way of knowing from it whether *every* pair $(x, v) \in [X_\varepsilon \times V] \cap \partial f$ is of the form (x_v, v). Thus, (1.18) may be weaker in general than (c) of Theorem 2. Ioffe's equivalence therefore doesn't allow us to conclude that tilt stability is actually equivalent to the properties in Theorem 2, even though that has been known to be true since [13] under the additional assumption of subdifferential continuity plus prox-regularity.

Also inadequately understood at present, and for that reason absent from this paper, are equivalences with conditions on the second-order subdifferential $\partial^2 f(\bar{x})$, which were in [13] and have since been utilized in other work on stability.

Anyway, none of this previous research has focused on characterizing local monotonicity or local strong monotonicity of ∂f at the present level of generality, although we wish our assumption that $\bar{v} \in \hat{\partial} f(\bar{x})$ might be weakened to $\bar{v} \in \partial f(\bar{x})$. Nor has anyone up to now related such monotonicity to the local implementation of the proximal point algorithm, which we will undertake in Section 3.

2 The proofs of Theorems 1, 2 and 3.

Let's start by confirming Theorem 3 on the basis of Theorem 2. Condition (c) of Theorem 2 implies the weaker condition (c) of Theorem 1, which we have seen to mean (1.14) holding for $(x, v) \in \operatorname{gph}\partial f$ in the neighborhood $X \times V$ of $(\bar{x}, \bar{v})$, which we can take to be open, when $f(x) < f(\bar{x}) + \varepsilon$. In the notation of (1.16), this says that $(x, v) \in [X_\varepsilon \times V] \cap \operatorname{gph}\partial f$ implies $x \in M(v)$. On the other hand, with X open it's clear that $x \in M(v)$ implies $v \in \partial f(x)$. Furthermore, because $x' = \bar{x}$ is a candidate in the minimization in (1.17), $x = M(v)$ entails $f(x) - v{\cdot}(x - \bar{x}) \leq f(\bar{x})$ and consequently through the convergence of $M(v)$ to $\bar{x}$ as $v \to 0$ that $f(x) < f(\bar{x}) + \varepsilon$ for v near enough to 0. Hence, with a possible adjustment to a smaller open neighborhood V if necessary, we have

$$(x, v) \in [X_\varepsilon \times V] \cap \operatorname{gph}\partial f \quad \Longleftrightarrow \quad (v, x) \in [V \times X_\varepsilon] \cap \operatorname{gph} M,$$

so that these two localizations are the inverses of each other. In (a) of Theorem 2, the truncated mapping with its graph on the left is maximal strongly monotone with modulus $\sigma > 0$. That is known to mean that its inverse is a monotone mapping that is Lipschitz continuous with modulus σ^{-1}.

Next we portray Theorem 2 as hardly more than a corollary of Theorem 1. We simply need to apply Theorem 1 to the function

$$f_\sigma(x) = f(x) - \frac{\sigma}{2}||x - \bar{x}||^2 \text{ with } \partial f_\sigma(x) = \partial f(x) - \sigma(x - \bar{x}), \;\; f_\sigma(\bar{x}) = f(\bar{x}), \;\; \partial f_\sigma(\bar{x}) = \partial f(\bar{x}).$$

Conditions (b) and (c) of Theorem 2 for f can be identified in this way with conditions (b) and (c) for f_σ. The same holds for conditions (a) and (a′) of the two theorems through the following fact, as invoked for the mapping T with $\operatorname{gph} T = \operatorname{gph} \partial f - (\bar{x}, \bar{v})$, which has $(0, 0)$ in its graph: T is σ-strongly monotone in the neighborhood W of $(0, 0)$ if and only if $T_\sigma = T - \sigma I$ is monotone in the neighborhood $L(W)$ of $(0, 0)$, where L is the linear transformation that takes (x, v) to $(x, v - \sigma x)$. This linear transformation changes the graph of T into the graph of T_σ and, because it's invertible, converts a neighborhood of $(0, 0)$ into another neighborhood of $(0, 0)$. The strong monotonicity inequality for pairs in the graph of T thereby turns into the monotonicity inequality for pairs in the graph of T_σ, as is well known already from global analysis of strong monotonicity.

Turning now to the proof of Theorem 1, we suppose first that (c) holds and aim at getting (b). Define $\widehat{f}$ by

$$\widehat{f}(x') = \sup \{\, f(x) + v{\cdot}(x' - x) \mid (x, v) \in [X_\varepsilon \times V] \cap \operatorname{gph} \partial f \}.$$

As the supremum of a collection of affine functions of x', $\widehat{f}$ is lsc convex. On the basis of (c) it's clear also that $\widehat{f} \leq f$ on X and that having $(x, v) \in [X_\varepsilon \times V] \cap \operatorname{gph} \partial f$ is equivalent to having $(x, v) \in [X \times V] \cap \operatorname{gph} \partial \widehat{f}$ and $\widehat{f}(x) = f(x)$. To reach (b), we only have to demonstrate the impossiblity of having $(x, v) \in [X \times V] \cap \operatorname{gph} \partial f$ but $\widehat{f}(x) < f(x)$, at least when (x, v) is near enough to $(\bar{x}, \bar{v})$, because the definition of variational convexity will then be satisfied, although perhaps for a neighborhood $X' \times V'$ of $(\bar{x}, \bar{v})$ that's smaller than $X \times V$.

Let C be a compact convex neighborhood of $\bar{x}$ within X and consider for $x \in \operatorname{int} C$ and $(x, v) \in [X \times V] \cap \operatorname{gph} \partial \widehat{f}$ the function $h_{x,v}$ defined by

$$h_{x,v}(x') = f(x') - \widehat{f}(x) - v{\cdot}(x' - x) + \frac{1}{2}||x' - x||^2 + \delta_C(x').$$

As the sum of the lsc function $f + \delta_C$ and a quadratic function which depends continuously on (x, v) (through the fact that convexity of $\widehat{f}$ makes $\widehat{f}(x)$ depend continuously on $(x, v) \in \operatorname{gph} \partial \widehat{f}$, cf. [16, 13.30]), the function $h_{x,v}$ depends epicontinuously on (x, v) as seen from the characterization of this property in [16, 7.2]. That will have crucial impact shortly in understanding how the nonempty set $\operatorname{argmin} h_{x,v} \subset C$ behaves with respect to (x, v).

From the inequalities $f(x') \geq \widehat{f}(x') \geq \widehat{f}(x) + v{\cdot}(x' - x)$ we have $h_{x,v}(x') > 0$ unless $x' = x$ and $f(x) = \widehat{f}(x)$. Thus,

$$\operatorname{argmin} h_{x,v} = \{x\} \text{ and } \min h_{x,v} = 0 \text{ when } (x, v) \in [X_\varepsilon \times V] \cap \operatorname{gph} \partial f.$$

This applies in particular to $(x, v) = (\bar{x}, \bar{v})$, and the epicontinuous dependence of $h_{x,v}$ on (x, v) therefore yields by [16, 7.33] that

$$\operatorname{argmin} h_{x,v} \to \{\bar{x}\} \text{ and } \min h_{x,v} \to 0 \text{ as } (x, v) \to (\bar{x}, \bar{v}) \text{ in } \operatorname{gph} \partial \widehat{f}.$$

Hence by taking (x,v) close enough to $(\bar{x},\bar{v})$ we can guarantee that an element $\tilde{x}\in\operatorname{argmin} h_{x,v}$ is arbitrarily close to $\bar{x}$ (and within $\operatorname{int} C$) and that $h_{x,v}(\tilde{x})$ is arbitrarily close to 0, which entails $f(\tilde{x})$ being arbitrarily close to $f(\bar{x})$ (ensuring $\tilde{x}\in X_\varepsilon$). Such $\tilde{x}$ in $\operatorname{argmin} h_{x,v}$ must also satisfy the necessary condition $0\in\partial h_{x,v}(\tilde{x})=\partial f(\tilde{x})-v+(\tilde{x}-x)$, so that $\tilde{v}\in\partial f(\bar{x})$ for $\tilde{v}=v-(\tilde{x}-x)$. Again, by taking (x,v) close enough to $(\bar{x},\bar{v})$ we can get $\tilde{v}$ arbitrarily close to $\bar{v}$. In summary, for (x,v) in a neighborhood of $(\bar{x},\bar{v})$ that's small enough, we will have $(\tilde{x},\tilde{v})\in[X_\varepsilon\times V]\cap\operatorname{gph}\partial f$ and consequently $(\tilde{x},\tilde{v})\in\operatorname{gph}\partial\widehat{f}$ and $\widehat{f}(\tilde{x})=f(\tilde{x})$. We can finish up now by showing this can't hold unless $\tilde{x}=x$. Indeed, having both (x,v) and $(\tilde{x},\tilde{v})$ belong to $\operatorname{gph}\partial\widehat{f}$ entails through the convexity of $\widehat{f}$ that

$$0\le(\tilde{x}-x){\cdot}(\tilde{v}-v)=(\tilde{x}-x){\cdot}(v-(\tilde{x}-x)-v)=-||\tilde{x}-x||^2,$$

which requires $\tilde{x}=x$. This confirms that (c) implies (b).

Next we suppose in Theorem 1 that (b) holds and aim at getting (a′). This is easy. From the definition of variational convexity we have a function $\widehat{f}$ that's lsc convex, and its subgradient mapping $\partial\widehat{f}$ is therefore maximal monotone globally, the graphical agreement in (1.11) ensures that ∂f is maximal monotone in the neighborhood $X\times V$ of $(\bar{x},\bar{v})$, not just with respect to $X_\varepsilon\times V$. Trivially, (a′) in turn implies (a).

Finally, we suppose in Theorem 1 that (a) holds and aim at getting (c). This is the seriously difficult part for which diverse tools of variational analysis will have to be brought in.

Because all the analysis will be local, we can suppose without loss of generality that $\operatorname{dom} f$ is bounded. In the notation that

$$\mathbb{B}\text{ and }\mathbb{B}^\circ\text{ are the closed and open unit balls in }\mathbb{R}^n,$$

we can furthermore assume, through rescaling if necessary, that

$$\partial f\ \text{ is monotone with respect to }\ (x,v)\in(\bar{x}+3\mathbb{B}^\circ)\times(\bar{v}+3\mathbb{B}^\circ)\ \text{ with }\ f(x)<f(\bar{x})+\varepsilon. \tag{2.2}$$

A complicating feature of the situation we face is the need to determine a common set X over which the minimization properties in (c) hold, instead of a separate set for each (x,v) in some neighborhood of $(\bar{x},\bar{v})$ in $\operatorname{gph}\partial f$. Our tactic for that is to introduce with respect to the lsc functions

$$\begin{array}{l} g_{x,v}(u)=\lambda^{-1}[f(x+\lambda u)-f(x)]-v{\cdot}u\ \text{ for }\ (x,v)\in\operatorname{gph}\partial f\ \text{ in choosing}\\ \lambda=\min\{1,\varepsilon/6\},\ \text{ so that }\ \partial g_{x,v}(u)=\partial f(x+\lambda u)-v\ \text{ and furthermore}\\ f(x+u)<f(\bar{x})+\varepsilon\ \text{ if }\ g_{x,v}(u)<\lambda,\ f(x)<f(\bar{x})+\lambda,\ ||u||<2,\ ||v||<2. \end{array} \tag{2.3}$$

In particular these functions all have $\operatorname{dom} g_{x,v}$ bounded and

$$g_{x,v}(0)=0\ \text{ and }\ 0\in\partial g_{x,v}(0). \tag{2.4}$$

From (2.2) and (2.3) we are assured that

$$\begin{array}{c} \partial g_{x,v}\ \text{ is monotone in }\ \{\,(u,w)\in 2\mathbb{B}^\circ\times 2\mathbb{B}^\circ\,|\,g_{x,v}(u)<\lambda\}\\ \text{as long as }\ (x,v)\in(\bar{x},\bar{v})+[\mathbb{B}^\circ\times\mathbb{B}^\circ],\ f(x)<f(\bar{x})+\lambda, \end{array} \tag{2.5}$$

and therefore

$$(x,v)\text{ is henceforth restricted to }\ (\bar{x},\bar{v})+[\mathbb{B}^\circ\times\mathbb{B}^\circ],\ f(x)<f(\bar{x})+\lambda, \tag{2.6}$$

We will be occupied with understanding whether (2.6) ensures that $g_{x,v}$ has a local minimum at 0 and whether there is neighborhood in localization that is the same for all (x,v) near enough to $(\bar{x},\bar{v})$ with $f(x) < f(\bar{x}) + \lambda$. For this it is be useful to consider the nonincreasing function

$$\gamma_{x,v}(r) = \min_{||u||\leq r} g_{x,v}(u) \leq 0, \quad \text{with} \quad \gamma_{x,v}(0) = 0, \tag{2.7}$$

which likewise is lsc because of the level boundedness of $g_{x,v}(u)+\delta_{r\boldsymbol{B}}(u)$ with respect to (u,r). Clearly

$$g_{x,v} \ \text{ has a local minimum at } 0 \qquad \Longleftrightarrow \qquad \gamma_{x,v}(r) = 0 \ \text{ for some } \ r > 0. \tag{2.8}$$

Our goal of arriving at (c) will correspond through this to *determining the existence of* $r_0 > 0$ *such that* $\gamma_{x,v}(r_0) = 0$ *for all* (x,v) *near enough to* $(\bar{x},\bar{v})$ *in* $\operatorname{gph}\partial f$ *with* $f(x) < f(\bar{x}) + \lambda$.

A direct approach to that could be to work with the optimization problem behind (2.7), viewed as the minimization of $g_{x,v}(u) + \delta_{r\boldsymbol{B}}(u)$ in $u \in I\!R^n$ with parameter $r \geq 0$, as a means of getting a formula for the subgradients of $\gamma_{x,v}$. However, there is a pitfall to that because the subgradients of $g_{x,v} + \delta_{r\boldsymbol{B}}$ might not be accessible without an appeal to some constraint qualification. That could lead to a major restriction on f, which we wish to avoid.

For that reason, instead of working with $\delta_{r\boldsymbol{B}}(u)$ we'll work an expression $r\theta(r^{-1}||u||)$ in which θ is a convex function on $I\!R$ that is finite and differentiable on $(-\infty,3)$:

$$\begin{array}{l} \theta(t) \equiv 0 \ \text{ on } (-\infty,1], \qquad \theta(t) \equiv \infty \ \text{ on } [3,\infty), \\ \theta'(t) = 2(t-1) \ \text{ on } [1,2], \text{ then increases to } \infty \text{ on } (2,3), \\ \ \text{so that } \ \theta(t) = (t-1)^2 \ \text{ on } [1,2], \ \theta'(2) = 2, \ \theta(2) = 1. \end{array} \tag{2.9}$$

Define the lsc convex function $h \geq 0$ on $I\!R^n \times I\!R$ by

$$h(u,r) = \begin{cases} r\theta(r^{-1}||u||) \ \text{ for } \ r > 0, \\ \delta_{\{0\}}(u) \ \text{ for } \ r = 0, \\ \infty \ \text{ for } \ r < 0, \end{cases} \tag{2.10}$$

observing that

$$\operatorname{dom} h = \{(0,0)\} \cup \{\, (u,r) \,|\, ||u|| < 3r\} \tag{2.11}$$

and that h is differentiable on $\operatorname{dom} h$ away from $(0,0)$ with

$$\nabla h(u,r) = \begin{cases} \{(0,0)\} \ \text{ when } \ ||u|| \leq r, \ (u,r) \neq (0,0), \\ \Big(\theta'(r^{-1}||u||)\frac{u}{||u||}, \ \theta(r^{-1}||u||) - \theta'(r^{-1}||u||)(r^{-1}||u||)\Big) \ \text{ when } \ r < ||u|| < 3r. \end{cases} \tag{2.12}$$

Consider now the function

$$b_{x,v}(u,r) = g_{x,v}(u) + h(u,r) \tag{2.13}$$

and define

$$\beta_{x,v}(r) = \inf\nolimits_u b_{x,v}(u,r), \qquad U_{x,v}(r) = \operatorname{argmin}_u b_{x,v}(u,r), \tag{2.14}$$

noting that the minimum is attained when $r \geq 0$ because $b(x,v)$ is lsc and level-bounded, due to $\operatorname{dom} g_{x,v}$ being bounded, and because $u = 0$ is always a candidate when $r \geq 0$, and $b_{x,v}(0,r) = 0$. Indeed, the lower semicontinuity and level-boundedness further ensures that $\beta_{x,v}$ is lsc [16, 1.17] as well as nonincreasing on $I\!R$, with $\beta_{x,v}(r) = \infty$ when $r < 0$ but $\beta_{x,v}(0) = 0$. Moreover, since $\beta_{x,v}(r) \leq 0$ when $r \geq 0$ we actually get in (2.14), through having $b_{x,v}(u,r) \leq b_{x,v}(u,\bar{r})$ when $r \geq \bar{r}$, that

$$\left.\begin{array}{l} \beta_{x,v}(r) = \inf\nolimits_u \{\, b_{x,v}(u,r) \,|\, b_{x,y}(u,\bar{r}) \leq 0\} \leq 0, \\ U_{x,v}(r) = \operatorname{argmin}_u \{\, b_{x,v}(u,r) \,|\, b_{x,y}(u,\bar{r}) \leq 0\} \subset \{\, u \,|\, g_{x,v}(u) \leq 0\} \end{array}\right. \quad \text{when } \ r \geq \bar{r} \geq 0. \tag{2.15}$$

This will be helpful shortly as a simplification. Another important fact is that

$$\gamma_{x,v}(r) \geq \beta_{x,v}(r) \geq \gamma_{x,v}(3r) \ \text{ when } \ r \geq 0, \tag{2.16}$$

as follows from (2.9). Hence being able to show that $\beta_{x,v}(r) \equiv 0$ on an interval $[0, 3\rho]$ for some $\rho > 0$ would lead to the conclusion that $\gamma_{x,v}(r) \equiv 0$ on $[0, \rho]$ and thus that $g_{x,v}$ has a minimum at 0 relative to $\rho\mathbb{B}$. Therefore, instead of analyzing the subgradients of $\gamma_{x,v}$, as initially considered, we will work at analyzing those of $\beta_{x,v}$.

Properties coming from the definition of $\beta_{x,v}$ in (2.14) as the pointwise minimum over $(0, \infty)$ of the function collection $\{b_{x,v}(u, \cdot)\}_{u \in \boldsymbol{R}^n}$ will provide crucial support. To better apply the results available in such a situation, let's exploit it in the form (2.15) and express it in maximization mode:

$$\begin{gathered} -\beta_{x,v}(r) = \max_{(u,\alpha) \in A(\bar{r})} \{ -\alpha - h(u, r) \} \ \text{ for } \ r \in (0, \bar{r}], \\ \text{where } \ A(\bar{r}) = \{ (u, \alpha) \mid \alpha \geq g_{x,v}(u), \ \alpha + h(u, \bar{r}) \leq 0 \}. \end{gathered} \tag{2.17}$$

The advantage of the switch is that the set $A(\bar{r})$ is compact and works uniformly for $r \in (0, \bar{r}]$. Since $\bar{r}$ can be arbitrarily close to 0, that tells us that $-\beta_{x,v}$ is a lower-$\mathcal{C}^1$ function on $(0, \infty)$ [16, 10.29]. In consequence, it has left and right derivatives at every r which satisfy the limit relations familiar for convex functions [16, 10.31]. Specifically, in translating from $-\beta_{x,v}$ back to $\beta_{x,v}$, there exist nonincreasing left and right derivatives when $r > 0$,

$$(\beta_{x,v})'_-(r) \geq (\beta_{x,v})'_+(r) \ \text{ such that } \ \begin{cases} \lim\limits_{s \nearrow r} (\beta_{x,v})'_+(s) = \lim\limits_{s \nearrow r} (\beta_{x,v})'_-(s) = (\beta_{x,v})'_-(r), \\ \lim\limits_{s \searrow r} (\beta_{x,v})'_-(s) = \lim\limits_{s \searrow r} (\beta_{x,v})'_+(s) = (\beta_{x,v})'_+(r), \end{cases} \tag{2.18}$$

where having $(\beta_{x,v})'_-(r) = (\beta_{x,v})'_+(r)$ corresponds to having $\beta_{x,v}$ differentiable at r with this common value as the derivative $\beta'_{x,v}(r)$.

Moving on now to determining the elements of the subgradient set $\partial\beta_{v,x}(r)$, we have at our disposal an important implication from the first line of (2.18), namely that

$$\partial\beta_{x,v}(r) = \begin{cases} \{\beta'_{v,x}(r)\} \ \text{ if } \ (\beta_{x,v})'_-(r) = (\beta_{x,v})'_+(r), \\ \{ (\beta_{x,v})'_-(r), \ (\beta_{x,v})'_+(r)\} \ \text{ if } \ (\beta_{x,v})'_-(r) > (\beta_{x,v})'_+(r). \end{cases} \tag{2.19}$$

Let's next apply to $\beta_{x,v}$, as defined by parametric minimization in (2.15), the rule in [16, 10.13] for subgradients in parametic optimization:

$$\partial\beta_{x,v}(r) \subset \{ s \mid \exists\, u \in U_{x,v}(r) \ \text{ with } \ (0, s) \in \partial b_{x,v}(u, r)\}. \tag{2.20}$$

Because $h(u, r)$ is differentiable when $r > 0$, $||u|| < 3r$, we know also that

$$\partial b_{x,v}(u, r) \ \subset \ (\partial g_{x,v}(u), 0) + \nabla h(u, r) \ \text{ when } \ r > 0, \ ||u|| < 3r. \tag{2.21}$$

Hence

$$\partial\beta_{x,v}(r) \subset \{ s \mid \exists\, u \in U_{x,v}(r) \ \text{ with } \ (0, s) \in \partial g_{x,v}(u) + \nabla h(u, r)\} \ \text{ when } \ r > 0, \ ||u|| < 3r, \tag{2.22}$$

from which we see through (2.12) that

$$\partial\beta_{x,v}(r) \subset \bigcup \begin{cases} \{0\} \ \text{ if } \ \exists\, u \in U_{x,v}(r) \ \text{ with } \ ||u|| \leq r, \\ \{ \theta(r^{-1}||u||) - \theta'(r^{-1}||u||)(r^{-1}||u||)\} \ \text{ if } \ \exists\, u \in U_{x,v}(r) \ \text{ with} \\ \qquad r < ||u|| < 3r \ \text{ and also } \ -\theta'(r^{-1}||u||)\frac{u}{||u||} \in \partial g_{x,v}(u). \end{cases} \tag{2.23}$$

Lemma 1. *As long as $0 < r < 1$, the second case in (2.23) is impossible unless $||u|| \geq 2r$.*

Proof. The second case first of all entails $g_{x,v}(u) \leq 0$ according to (2.15). When $r < ||u|| < 2r$ with $0 < r < 1$, we then have from (2.5) and (2.6) that the pair $(u, -\theta'(r^{-1}||u||)\frac{u}{||u||}) \in \operatorname{gph} \partial g_{x,v}$ also has $||u|| < 2$ and $\theta'(r^{-1}||u||) < \theta'(2) = 2$. It thus lies in the range where $\partial g_{x,v}$ is monotone. Since $(0,0)$ is also in this range, the monotonicty implies that

$$0 \leq \Big(-\theta'(r^{-1}||u||)\frac{u}{||u||} - 0\Big)\cdot(u - 0) = -\theta'(r^{-1}||u|||)|u||.$$

But $\theta'(r^{-1}||u||) > 0$ and $||u|| > 0$, so this is contradictory. $\square$

Lemma 2. *When $||u|| \geq 2r$, the second case of (2.23) has $\theta(r^{-1}||u||) - \theta'(r^{-1}||u||)(r^{-1}||u||) \leq -3$.*

Proof. The convex function conjugate to θ on $I\!R$ has $\theta^*(s) = st - \theta(t)$ when $s = \theta'(t)$, so we are looking at this for $t = r^{-1}||u||$. We are claiming that $\theta^*(s) > 3$ when $t \in [2,3)$, as corresponds to $s \in [2,\infty)$ according to the formula for θ in (2.9). That formula indicates further that θ^* is an increasing function on $(0,\infty)$ with $\theta^*(s) = s + \frac{1}{4}s^2$ on [0,2]. Hence the value of $\theta^*(s)$ for $s \in [2,\infty)$ is bounded below by $\theta^*(2)$, which is 3. $\square$

The combination of Lemma 1 and Lemma 2 reveals that the possibilities for $\partial\beta_{x,v}(r)$ that are compatible with (2.23) are extremely limited when $0 < r < 1$.

Lemma 3. *There must exist $r_{x,y} \in [0,1]$ such that $\beta_{x,v}(r) = 0$ when $0 \leq r < r_{x,v}$ but $\beta_{x,v}(r) \leq 3(r - r_{x,v})$ when $r_{x,v} \leq r < 1$, in which case*

$$\gamma_{x,v}(r) = 0 \text{ when } 0 \leq r < r_{x,v} \text{ but } \gamma_{x,v}(r) \leq r - r_{x,v} \text{ when } r_{x,v} \leq r < 1. \tag{2.24}$$

Moreover $r_{x,v} = 0$ is impossible in the case of a regular subgradient, $v \in \hat{\partial} f(x)$.

Proof. Lemma 1 and Lemma 2 have determined that $\partial\beta_{x,v}(r)$ can only contain 0 or values ≤ -3 when $r \in (0,1)$. The description of $\partial\beta_{x,v}(r)$ in (2.19) and the limit relations in (2.18) make clear that $\beta_{x,v}$ must be continuously differentiable on $(0,1)$ except perhaps at one point $r_{x,v}$ where $\beta'_{x,v}(r) = 0$ to the left and $\beta'_{x,v}(r) < -3$ to the right. Otherwise $\beta'_{x,v}(r) = 0$ for all $r \in (0,1)$ or $\beta'_{x,v}(r) < -3$ for all $r \in (0,1)$. This supports the claim about the structure of $\beta_{x,v}$ on $[0,1)$, and then (2.24) follows from (2.16). For the second part, if $v \in \hat{\partial} f(x)$ we have $0 \in \hat{\partial} g_{x,v}(0)$ from (2.3) and therefore $g_{x,v}(u) \geq g_{x,v}(0) + o(||u||)$ by the definition of a regular subgradient in (1.1), where $g_{x,v}(0) = 0$. Then likewise $\gamma_{x,v}(r) \geq o(r)$ by (2.7). This precludes (2.24) from holding with $r_{x,y} = 0$. $\square$

Lemma 3 provides important information about whether $g_{x,v}$ has a local minimum at 0:

$$g_{x,v}(u) \geq g_{x,v}(0) = 0 \text{ when } ||u|| < r_{x,v}, \tag{2.25}$$

which translates through (2.3) to

$$x \in \operatorname{argmin}\Big\{ f(x') - v\cdot(x' - x) \,\Big|\, ||x' - \bar{x}|| < \lambda r_{x,v} \Big\}. \tag{2.26}$$

This comes close to our goal of getting condition (c) of Theorem 1, but for that we'll need to know more about the behaviour of the threshold $r_{x,v}$ with respect to (x,v).

Lemma 4. *If $(x^\nu, v^\nu) \to (\bar{x}, \bar{v})$ and $f(x^\nu) \to f(\bar{x})$ within the confines of (2.6), then $r_{x^\nu, v^\nu} \to r_{\bar{x}, \bar{v}}$.*

Proof. The argument will rely on properties of epiconvergence explained in [16, Chapter 7]. Consider a sequence of pairs $(x^\nu, v^\nu) \to (\bar{x}, \bar{v})$ in the set (2.6) with $f(x^\nu) \to f(\bar{x})$. We'll show first that this implies for the functions $b_{x,v}$ in (2.14), expressible directly in terms of f itself by

$$b_{x,v}(u, r) = \lambda^{-1} f(x + \lambda u) - f(x)] - v{\cdot}u + h(u, r), \tag{2.27}$$

that b_{x^ν, v^ν} epiconverges to $b_{\bar{x}, \bar{v}}$. A characterization of this in [16, 7.3] is that for all (u, r),

$$\begin{aligned} &\forall\, (u^\nu, r^\nu) \to (u, r),\ \liminf_\nu b_{x^\nu, v^\nu}(u^\nu, r^\nu) \geq b_{\bar{x}, \bar{v}}(u, r), \\ &\exists\, (u^\nu, r^\nu) \to (u, r),\ \limsup_\nu b_{x^\nu, v^\nu}(u^\nu, r^\nu) \leq b_{\bar{x}, \bar{v}}(u, r). \end{aligned} \tag{2.28}$$

The first condition in (2.28) is evident from (2.27) and the lower semicontinuity of f and h. The second in the case of $r > 0$ holds through the continuity of h when $r > 0$ by taking $r^\nu = r$ and $u^\nu = u + \bar{x} - x^\nu$. The only other case neeeding to be considered for the second condition in (2.28) is that of $(u, r) = (0, 0)$, for which one can simply take $(u^\nu, r^\nu) = (0, 0)$ as well.

This epiconvergence of b_{x^ν, v^ν} to $b_{\bar{x}, \bar{v}}$ implies the epiconvergence of β_{x^ν, v^ν} to $\beta_{\bar{x}, \bar{v}}$, because $\operatorname{epi} \beta_{x,v}$ is the image of $\operatorname{epi} b_{x,v}$ under $(x, v, u, r, \alpha) \to (x, v, r, \alpha)$. That follows from [16, 4.27 via 4.24].

Next we appeal to the local Lipschitz continuity of $\beta_{x,v}$ on $(0, \infty)$, which is a consequence of our having ascertained that $-\beta_{x,v}$ is lower $\mathcal{C}^1$, causes the epiconvergence of β_{x^ν, v^ν} to $\beta_{\bar{x}, \bar{v}}$ to entail not only pointwise but also uniform convergence of β_{x^ν, v^ν} to $\beta_{\bar{x}, \bar{v}}$ on all closed, bounded subintervals of $(0, \infty)$. That in turn guarantees for the compact sets

$$C_{\delta,\rho}(x, v) = \{\, (r, \mu) \,|\, r \in [\delta, \rho],\ \bar{\mu} \leq \mu \leq \beta_{x,v}(r) \} \ \text{ for } \ [\delta, \rho] \subset (0, 1), \tag{2.29}$$

where $\bar{\mu}$ is a lower bound to $\beta_{x,v}$ on [0,1] for (x, v) satisfying (2.6), that $C_{\delta,\rho}(x^\nu, v^\nu) \to C_{\delta,\rho}(\bar{x}, \bar{v})$. with respect to set convergence [16, Chapter 4]. Then too, for any continuous function φ on $I\!R^2$,

$$\limsup_{\nu \to \infty} \operatorname{argmax} \{\, \varphi(r, \mu) \,|\, (r, \mu) \in C_{\delta,\rho}(x^\nu, v^\nu) \} \subset \operatorname{argmax} \{\, \varphi(r, \mu) \,|\, (r, \mu) \in C_{\delta,\rho}(\bar{x}, \bar{v}) \} \tag{2.30}$$

by [16, 7.33]. In specializing this to $\varphi(r, \mu) = r + \mu$ we get from the very particular structure of $\beta_{x,v}$ in Lemma 3 that

$$\operatorname{argmax} \{\, r + \mu \,|\, (r, \mu) \in C_{\delta,\rho}(\bar{x}, \bar{v}) \} = \operatorname*{argmax}_{\delta \leq r \leq \rho} \{\, r + \beta_{x,v}(r) \} = m_{\delta,\rho}(x, v),$$

where

$$m_{\delta,\rho}(x, v) = \begin{cases} \beta_{x,v}(\delta) \ \text{ if } \ r_{x,v} < \delta, \\ r_{x.v} \ \text{ if } \ \delta \leq r_{x,v} \leq \rho, \\ 1 \ \text{ if } \ r_{x,v} > \rho. \end{cases} \tag{2.31}$$

Therefore, from (2.30),

$$\lim_{v \to \infty} m_{\delta,\rho}(x^\nu, v^\nu) = m_{\delta,\rho}(\bar{x}, \bar{v}). \tag{2.32}$$

There's no loss of generality in taking $\delta < r_{\bar{x}, \bar{v}}$, inasmuch as $r_{\bar{x}, \bar{v}} > 0$ by Lemma 3 under our assumption that $\bar{v} \in \hat{\partial} f(\bar{x})$, in which case $m_{\delta,\rho}(\bar{x}, \bar{v}) = 0$ by (2.31). Then, through the uniform convergence of β_{x^ν, v^ν} to $\beta_{\bar{x}, \bar{v}}$ on $[\delta, \rho]$, the combination of (2.32) and definition (2.31) tells us that $r_{x^\nu, v^\nu} \to r_{\bar{x}, \bar{v}}$ if $r\bar{x}, \bar{v} \leq \rho$. Since ρ can be arbitrarily near to 1, this confirms the claim of the lemma. □

This brings us nearly to our goal of getting part (c) of Theorem 1 out of part (a). From Lemma 4 and the fact that $r_{\bar{x},\bar{v}} > 0$ by Lemma 3 and the regularity of $\bar{v}$ as a subgradient of f at $\bar{x}$, there must exist $\bar{r} > 0$ along with $\bar{\varepsilon} > 0$, which can be taken $\leq$ the λ in (2.6), such that

$$r_{x,v} \geq \bar{r} \;\text{ when }\; (x,v) \in (\bar{x},\bar{v}) + 2\bar{\varepsilon}[I\!B^{\circ} \times I\!B^{\circ}] \;\text{ and }\; f(x) \leq f(\bar{x}) + \bar{\varepsilon}. \tag{2.33}$$

Then by taking $\bar{X} \times \bar{V} = (\bar{x},\bar{v}) + \bar{\varepsilon}[I\!B^{\circ} \times I\!B^{\circ}]$ we get from $(\bar{X}_{\bar{\varepsilon}}, \bar{V})$ an f-localization of ∂f for which the desired property in (c) holds. This ends the proof of Theorem 1.

3 Application to the proximal point algorithm

The proximal point algorithm in its original development in [14] was aimed at solving the following problem for a maximal monotone mapping $T : I\!R^n \rightrightarrows I\!R^n$:

$$\text{determine } \bar{x} \text{ such that } 0 \in T(\bar{x}), \;\text{ or equivalently, }\; \bar{x} \in T^{-1}(0). \tag{3.1}$$

It did this through iterations of the form

$$x_{k+1} = (I + c_k T)^{-1}(x_k) \;\text{ with }\; c_k > 0, \tag{3.2}$$

or some appoximation thereof. These were based on the fact that maximal monotonicity of T makes the mapping $(I + c_k T)^{-1}$ be single-valued and nonexpansive (Lipschitz continuous with modulus 1) from $I\!R^n$ into $I\!R^n$, moreover with the (convex) set of solutions to (3.1) as its set of fixed points. That was relaxed by Eckstein and Bertsekas in [1] to iterations of the form

$$x_{k+1} = [(1-\lambda_k)I + \lambda_k (I + c_k T)^{-1}](x_k) \;\text{ with }\; \lambda_k \in (0,2), \tag{3.3}$$

in which (3.2) fits as the case where $\lambda_k = 1$.

When $T = \partial f$ for a proper lsc function $f : I\!R^n \to (-\infty, \infty]$, the maximal monotonicity of T corresponds to the convexity of f, as already recalled at the start of this paper. The iterations in (3.2) or (3.3) can be executed then through observing that

$$(I + c_k \partial f)^{-1}(x_k) = \{\, x \,|\, 0 \in \partial f_k(x) \} \;\text{ where }\; f_k(x) = f(x) + \frac{1}{2c_k}||x - x_k||^2. \tag{3.4}$$

When f is convex, this is the same as

$$(I + c_k \partial f)^{-1}(x_k) = \operatorname*{argmin}_{x \in \boldsymbol{R}^n} \Big\{ f(x) + \frac{1}{2c_k}||x - x_k||^2 \Big\}, \tag{3.5}$$

where the quadratic term ensures that a strongly convex function is being minimized and therefore that the global minimum is attained at a unique point.

Our concern here is the possible extension of such facts to *local* monotonicity and *local* minimization. Pennanen in [10] showed that the same properties of convergence of the proximal point algorithm hold locally for either (3.2) or (3.3), as long as the iterations are initiated near enough to a solution and the graph of T is maximally monotone locally around there. Specifically, in our notation, his result [10, Proposition 6] postulates that[3]

$$\begin{array}{l} \exists \text{ open } X \times V \text{ with } 0 \in V \text{ and } X \cap T^{-1}(0) \text{ closed} \neq \emptyset \\ \text{such that also } [X \cap T^{-1}(0)] + \delta I\!B \subset X \text{ for some } \delta > 0, \\ \text{and furthermore } T \text{ is maximal monotone in } X \times V. \end{array} \tag{3.6}$$

[3]In a more recent article than here we have shown that this condition can be weakened; see [15, Theorem 2].

Then, supposing $\inf c_k > 0$, $\inf \lambda_k \geq 1$ and $\sup \lambda_k < 2$, if x_0 is near enough to $X \cap T^{-1}(0)$, there will exist $\varepsilon > 0$ and a particular solution $\bar{x} \in X \cap T^{-1}(0)$ such that in each iteration there is a *unique*

$$x_{k+1} \in [\bar{x} + \varepsilon I\!B] \cap [(1 - \lambda_k)I + \lambda_k(I + c_k T)^{-1}](x_k), \tag{3.7}$$

and the sequence so generated will converge to $\bar{x}$ in the manner expected from the theory for globally maximal monotone T.

Here we're not focused on the details of that convergence but on how the execution step in (3.7) specializes when $T = \partial f$. The difficulty is that local maximal monotonicity in $X \times V$, in contrast to global maximal monotonicity of ∂f, doesn't ensure convexity of f even locally. This is where our results in Theorems 1 and 2 can supply insights.

For general f, the implementation of (3.7) might be carried out by the rule in (3.4), of course, but what would that mean in practice? The presence of λ_k makes the coordination with the neighborhood $\bar{x} + \varepsilon I\!B$ murky, so in what follows we'll restrict attention to $\lambda_k = 1$. Then the iteration step reduces to taking

$$x_{k+1} \in [\bar{x} + \varepsilon I\!B] \cap \{\, x \mid 0 \in \partial f_k(x) \} \;\text{ where }\; f_k(x) = f(x) + \frac{1}{2c_k}||x - x_k||^2. \tag{3.8}$$

This can be interpreted as determining x_k as a sort of "stationary point" of f_k that's ε-near to $\bar{x}$. Many so-called optimization algorithms are really oriented around finding a solution for first-order optimality conditions and thus fit into such a picture. But with our new knowledge about variational convexity, we can get more.

Theorem 4 (proximal point iterations as local minimization)**.** *Let the variational second-order sufficient condition for local optimality hold at $\bar{x}$, i.e., f is variationally convex at $\bar{x}$ for $0 \in \partial f(\bar{x})$, and suppose $0 \in \hat{\partial} f(\bar{x})$. Then Pennanen's assumptions (3.6) lead, when x_k is near enough to $\bar{x}$, to this step taking the form*

$$x_{k+1} = \operatorname*{argmin}_{x \in \bar{x} + \varepsilon \boldsymbol{B}} \Big\{ f(x) + \frac{1}{2c_k}||x - x_k||^2 \Big\}, \tag{3.9}$$

with the minimizer being uniquely determined in the interior of the ball $\bar{x} + \varepsilon I\!B$. Convergence to $\bar{x}$ is then moreover convergence to a local minimizer of f instead of just to a stationary point. Under the strong variational second-order sufficient condition, providing strong variational convexity, $\bar{x}$ is sure even to be a tilt-stable local minimizer.

Proof. This is obvious from Theorems 1, 2 and 3. □

Of course like most algorithmic prescriptions that only kick in when "close enough" to a solution, the big question is how to get close enough. In Pennanen's local version of the proximal point algorithm, the mystery is how to carry out the local minimization in (3.9) without already having identified $\bar{x}$. Presumably, the answer is to invoke a minimization scheme with respect to the successive functions $f_k(x) = f(x) + \frac{1}{2c_k}||x - x_k||^2$ which may at first only produce x_{k+1} a local stationary point, but eventually, with luck, will approach a local minimizer and then be attracted to it.

References

[1] Eckstein, J., and Bertsekas, D. P., "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematicl Programming Series A* **55** (1992), 293–318.

[2] Drusvyatskiy, D., and Lewis, A. S., "Tilt stablity, uniform quadratic growth, and strong metric subregularity of the subdifferential," *SIAM J. Optimization* **23** (2013), 256–267.

[3] Drusvyatskiy, D., Mordukhovich, B. S., and Nhgia, T. T. A., "Second-order growth, tilt stability and metric regularity of the subdifferential," *Journal of Convex Analysis* **21** (2014), 1165–1192.

[4] Ioffe, A. D., *Variational Analysis of Regular Mappings: Theory and Applications*, Springer Monographs in Mathematics, 2017.

[5] Levy, A. B., Poliquin, R. A., and Rockafellar, R. T., "Stability of locally optimal solutions," *SIAM J. Optimization* **10** (2000), 580–604.

[6] Mordukhovich, B. M., Rockafellar, R. T., and Sarabi, M. E., "Characterizations of full stability in constrained optimization," *SIAM J. Optimization* **23** (2013), 1810–1849.

[7] Mordukhovich, B. M., Nghia, T. T. A., and Rockafellar, R. T., "Full stability in finite-dimensional optimization," *Mathematics of Operations Research* **40** (2015), 226–252

[8] Mordukhovich, B. M., and Nghia, T. T. A., "Second-order variational analysis and characterizations of tilt-stable optimal solutions in finite and infinite dimensions," *Nonlinear Analysis* **86** (2013), 159–180.

[9] Moreau, J.-J., "Proximité et dualité dans un espace hilbertien," *Bulletin de la Societé Mathématique de France* **93** (1965), 273-299.

[10] Pennanen, T., "Local convergence of the proximal point algorithm and multiplier methods without monotonicity," *Mathematics of Operations Research* **27** (2002), 170–191.

[11] Poliquin, R. A., "Subgradient monotonicity and convex functions," *Nonlinear Analysis* **14** (1990), 385–398.

[12] Poliquin, R. A., and Rockafellar, R. T., "Prox-regularity in variational analysis," *Transactions of the American Mathematical Society* **348** (1996), 1805–1838.

[13] Poliquin, R. A., and Rockafellar, R. T., "Tilt stability of a local minimum," *SIAM J. Optimization* **8** (1998), 287–299.

[14] Rockafellar, R. T., "Monotone operators and the proximal point algorithm," *SIAM J. Control and Optimization* **14** (1976), 877–898.

[15] Rockafellar, R. T., "Progressive decoupling of linkages in optimization and variational inequalities with elicitable convexity or monotonicity," *Set-valued and Variational Analysis*, 2018.

[16] Rockafellar, R. T., and Wets, R. J-B, *Variational Analysis*, No. 317 in the series *Grundlehren der Mathematischen Wissenschaften*, Springer-Verlag, 1998.