Variational Convexity and Prox-Regularity

R. Tyrrell Rockafellar[1]

Abstract

Variationally convex functions and prox-regular functions are placed in a single framework that coordinates and enhances the many connections between them. Their characterization in terms of monotonicity properties of subgradient mappings is confirmed by a more informative argument than has previously been offered, which yields more a definitive result.

Keywords: *variational convexity, prox-regularity, local subdifferential monotonicity, uniform quadratic growth, second-order variational analysis*

Version of 18 March 2024

Dedicated to the memory of Hedy Attouch

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

1 Introduction and statement of results

Prox-regular functions, since their introduction in [8] in 1996, have been prominent in studies of stability of solutions in problems of optimization as in [9], [3], [7], [4], [5] and [1], and also in second-order calculus in variational analysis more broadly, such as in [6] and [14, 13F]. Variationally convex functions, having subgradient mappings that are primal-dual-locally indistinguishable from those of convex functions, emerged much more recently in [10] in being motivated by algorithmic needs in [11], [13], and the study of local optimization duality in [12]. The two concepts, although developed independently for different reasons, have very much in common. The aim here is to place them in a single framework where their interactions are revealed in greater detail and can be more clearly appreciated.

Prox-regularity builds on the notion of a proximal subgradient. Let $f : I\!R^n \to (-\infty, \infty]$ be closed (lower semicontinuous) and proper, and let $x \in \operatorname{dom} f$. A vector v is a *proximal* subgradient of f at x, written $v \in \partial_p f(x)$, if for some r the inequality

$$f(x') \geq f(x) + \langle v, x' - x\rangle - \frac{r}{2}|x' - x|^2 \tag{1.1}$$

holds when x' is near enough to x and $f(x')$ is near enough to $f(x)$. In particular then, v is a *regular* subgradient, written $v \in \widehat{\partial} f(x)$, meaning that

$$f(x') \geq f(x) + \langle v, x' - x\rangle + o(|x' - x|). \tag{1.2}$$

Subgradients $v \in \partial f(x)$ in general are defined by taking limits of regular subgradients: there is a sequence $x^\nu \to x$ with $f(x^\nu) \to f(x)$ and $v^\nu \in \widehat{\partial} f(x^\nu)$ such that $v^\nu \to v$ [14, 8B]. But regular subgradients $v^\nu \in \widehat{\partial} f(x^\nu)$ can be restricted to being proximal subgradients $v^\nu \in \partial_p f(x^\nu)$ [14, 8.5] without affecting the limits obtained. Prox-regularity of f refers in this setting to a local reduction of the general subgradient mapping $\partial f : I\!R^n \rightrightarrows I\!R^n$ to the proximal subgradient mapping $\partial_p f$ in which the proximal parameter is kept at a uniform level. The localization is *primal-dual* in restricting both x and v, as well as *f-attentive* in also controlling $f(x)$ in the absence of f being continuous.

Although prox-regularity is usually described just in terms of the existence of a proximal parameter r fitting the requirements, with r treated as positive, there are good reasons to look more closely at the level of r and also allow it to be negative. That way, the perspective of "growth properties" can be brought in. Rewriting the proximal subgradient inequality (1.1) in the form

$$f(x') \geq f(x) + \langle v, x' - x\rangle + \frac{s}{2}|x' - x|^2, \tag{1.3}$$

for instance, produces in the case of $s = -r > 0$ a condition of quadratic growth that has long been recognized as essential to the analysis of optimality conditions and associated numerical methodology.

With this in mind, let's say that f has the s-level *uniform quadratic growth* property at the point $\bar{x} \in \operatorname{dom} f$ for a subgradient $\bar{v} \in \partial f(\bar{x})$ if there are neighborhoods $\mathcal{X}$ of $\bar{x}$ and $\mathcal{V}$ of $\bar{v}$ (taken for our purposes to be convex) along with $\rho > f(\bar{x})$ such that

$$\text{(1.3) holds for all } x' \in \mathcal{X} \text{ when } (x, v) \in (\mathcal{X} \times \mathcal{V}) \cap \operatorname{gph}_\rho \partial f, \tag{1.4}$$

where, in terms of $\operatorname{gph} \partial f$ being the graph of the subdifferential mapping ∂f, the set

$$\operatorname{gph}_\rho \partial f := \{ (x, v) \in \operatorname{gph} \partial f \mid f(x) < \rho \} \tag{1.5}$$

specifies an f-*truncation* of ∂f. But the parameter s is not restricted here to being positive. It may be 0 or negative. We can refine, and at the same time broaden, the meaning of prox-regularity by saying then that

$$\begin{array}{l} f \text{ is } r\text{-}level \text{ prox-regular at the point } \bar{x} \text{ for } \bar{v} \in \partial f(\bar{x}) \quad \Longleftrightarrow \\ f \text{ has uniform quadratic growth at } \bar{x} \text{ for } \bar{v} \text{ at the level } s = -r. \end{array} \tag{1.6}$$

The restriction in (1.4) to an f-truncation of ∂f via (1.5) bases the property in question on an f-*attentive* localization that focuses not only on (x, v) being close to $(\bar{x}, \bar{v})$, but also having $f(x)$ close to $f(\bar{x})$ That kind of localization is important for many reasons in variational analysis, because it ensures that a concept depends only on a neighborhood of $(\bar{x}, f(\bar{x}))$ in the epigraph of f, cf. [14, Chapter 8].[2] It's superfluous when f is *subdifferentially continuous* at $(\bar{x}, \bar{v})$ in having the mapping $(x, v) \mapsto f(x)$ be continuous relative to $\operatorname{gph} \partial f$ at $(\bar{x}, \bar{v})$. Subdifferential continuity covers a lot of territory, but we want to proceed here without presupposing it.

A closely related property of $\operatorname{gph} \partial f$ that doesn't concern the growth of function values involves f-truncation as well. Specifically, ∂f has at the point $\bar{x}$ for the subgradient $\bar{v} \in \partial f(\bar{x})$ an f-*attentive localization that is s-monotone* if, in the same context of $\mathcal{X}$, $\mathcal{V}$ and ρ,

$$\langle v_1 - v_0, x_1 - x_0 \rangle \geq s|x_1 - x_0|^2 \;\text{ when }\; (x_i, v_i) \in (\mathcal{X} \times \mathcal{V}) \cap \operatorname{gph}_\rho \partial f. \tag{1.7}$$

For $s = 0$, this is plain *monotonicity*, whereas for $s > 0$ it is *strong* monotonicity. For $s < 0$ it is *hypo-monotonicity.*

Level s of the uniform quadratic growth condition on f at $\bar{x}$ for $\bar{v}$ implies the existence of an f-attentive localization of ∂f that is s-monotone. That's elementary, but what about the converse? In the early paper [8], where the focus was on $s = -r < 0$, it was demonstrated such a localization guaranteed at least prox-regularity at *some* (inexactly determined) level $r' > r$, corresponding to $s' < s$. But why not at the same level, without a shift?

This is where the variational convexity can come in. As defined in [10], f is *variationally convex* at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ if, once more in the same context of $\mathcal{X}$, $\mathcal{V}$ and ρ, there is a closed proper function $\widehat{f}$ that is *convex* on $\mathcal{X}$,[3] for which

$$\begin{array}{l} (\mathcal{X} \times \mathcal{V}) \cap \operatorname{gph} \partial \widehat{f} = (\mathcal{X} \times \mathcal{V}) \cap \operatorname{gph}_\rho \partial f, \;\text{ with }\; \widehat{f} \leq f \;\text{ on } \mathcal{X} \text{ and} \\ \widehat{f}(x) = f(x) \;\text{ for all the pairs } (x, v) \text{ in the common intersection.} \end{array} \tag{1.8}$$

It is *variationally strongly convex* if $\widehat{f}$ is not just convex, but strongly convex on $\mathcal{X}$. But it will be good now to extend this to f being *variationally s-convex* if $\widehat{f}$ is s-convex on $\mathcal{X}$, meaning that

$$\begin{array}{c} \widehat{f}((1-\lambda)x_0 + \lambda x_1) \leq (1-\lambda)\widehat{f}(x_0) + \lambda \widehat{f}(x_1) - \frac{s}{2}\lambda(1-\lambda)|x_1 - x_0|^2 \\ \text{for all }\; x_0, x_1 \in \mathcal{X} \;\text{ and }\; \lambda \in (0,1). \end{array} \tag{1.9}$$

This is plain convexity when $s = 0$, strong convexity when $s > 0$, and *hypo-convexity* when $s < 0$.

Our featured theorem precisely coordinates these extended versions of variational convexity, monotone subdifferential localization, and quadratic growth with its mirror image of prox-regularity. Local maximality properties in the monotonicity enter in, and a distinction needs to be made with respect to what kind of neighborhood is taken. The s-monotonicity property in (1.7) is *locally max* if no subset of $\mathcal{X} \times \mathcal{V}$ strictly larger than the one in (1.7) has it. It is f-*attentive-locally max* when $\mathcal{X} \times \mathcal{V}$ is replaced in this by its maybe smaller subset where also $f(x) < \rho$.

[2] A illustration of the need for this is the function f on $\mathbb{R}$ having $f(x) = 0$ for $x \leq 0$ and $f(x) = 1$ for $x > 0$. It has $\operatorname{gph} \partial f$ equal to $(\mathbb{R}, 0) \cup (0, \mathbb{R}_+)$, whereas $\operatorname{gph}_\rho \partial f$ for $\rho = \frac{1}{2}$ is equal to $(\mathbb{R}_-, 0) \cup (0, \mathbb{R}_+)$. In localizing around $(\bar{x}, \bar{v}) = (0, 0)$, this makes a big difference.

[3] By adding to $\widehat{f}$ the indicator of the closure of the convex set $\mathcal{X}$, it's possible to pass to $\widehat{f}$ being convex on $\mathbb{R}^n$

Theorem 1 (fundamental equivalence of properties)**.** *The following properties are equivalent to each other at the same s value.*

(a) *f is variationally s-convex at $\bar{x}$ for $\bar{v}$.*

(b) *f has the s-level uniform quadratic growth property at $\bar{x}$ for $\bar{v}$.*

(c$'$) *∂f has an f-truncation that is locally max monotone around $(\bar{x}, \bar{v})$.*

(c) *∂f has an f-truncation that is f-attentive-locally max monotone around $(\bar{x}, \bar{v})$.*

(d) *∂f has an f-attentive localization at $\bar{x}$ for $\bar{v}$ that is s-monotone (no maximality assumed), and the subgradient $\bar{v}$ is regular, $\bar{v} \in \widehat{\partial} f(\bar{x})$.*

Note that, although the equivalence in Theorem 1 keeps s as the same level throughout, it's not claimed that the ρ and neighborhoods $\mathcal{X}$ and $\mathcal{V}$ in the definitions of the properties are kept the same.

In particular, Theorem 1 puts new light on prox-regularity by tying it to the seemingly much stronger property of variational hypo-convexity.

Corollary (top consequence for prox-regularity)**.** *The function f is prox-regular at $\bar{x}$ for $\bar{v}$ at some level $r > 0$ if and only if it is variationally hypo-convex there at the level $s = -r$.*

The vector $\bar{v}$ in Theorem 1 ends up necessarily being a proximal subgradient, $\bar{v} \in \partial_p f(\bar{x})$, without that having to be brought in as an assumption. Yet being a regular subgradient enters as an assumption in part (d). Why? This is made clear by our supplementary theorem.

Theorem 2 (counterexample)**.** *The equivalence of all the properties in Theorem 1 fails without the assumption of regularity in part (d). Specifically, for $s = 0$ there is an example of a function f such that ∂f has an s-monotone f-attentive localization at $\bar{x}$ for $\bar{v}$, yet f is not variationally s-convex at $\bar{x}$ for $\bar{v}$. In this example, $\bar{v} \in \partial f(\bar{x})$ but $\bar{v} \notin \widehat{\partial} f(\bar{x})$,*

Theorem 1 can be compared to Theorems 1 and 2 of the original paper on variational convexity [10], with the first theorem directed at $s = 0$ and the second at $s > 0$. The key difference in statement is that the previous results omitted (c$'$) and assumed the regularity of $\bar{v}$ in both (c) and (d) (it's automatic in (a) and (b)), while speculating that it might be in fact be superfluous in both. Here we resolve that issue by managing in Theorem 1 to drop that assumption in (c) and leave it out of (c$'$), while ending further speculation by determining in Theorem 2 that it can't be dropped from (d).

We reach this definitive improvement by taking a different route than in [10]. The proof is more informative and at the same time avoids a recently identified troublespot in the earlier proof, where at one place an uncertain property of the one-sided derivatives of lower-$\mathcal{C}^1$ functions [14, 10G] on $\mathbb{R}$ is taken for granted.[4] The old argument put all the hardest effort into going directly from (d) to (a), but the new argument goes from (d) to (c$'$) to (a). Novel to this subject, its success depends on appealing at a critical point to the fact that a nonempty subset of a given open set can't be both open and closed relative to that set without being all of that set.

Although (c$'$) didn't appear alongside of (c) originally in [10], it was explicitly brought out in [11, Theorem 7] as implied by the variational convexity in (a). Obviously (c) is in principle an intermediate property between (c$'$) and (d).

The full extent to which the equivalences in Theorem 1 might carry over to an infinite-dimensional Hilbert space is unclear, because compactness techniques are central to our approach. The infinite-dimensional results of Khanh, Khoa, Mordokhovich and Phat in [2] require assuming in the background that f is already prox-regular, at some level at least. With that they were able in [2, Theorem 6.6] to show the equivalence of (c) and (a), in particular, and therefore with (c$'$) as well. Whether their prox-regularity assumption might be removed is an open question.

[4] Is the right derivative really continuous from the right, and the left derivative from the left, as it is with lower-$\mathcal{C}^2$ functions? The author thanks Khoa Vu at Wayne State University, Detroit, for bringing this to his attention.

2 Proving the results

The counterexample behind Theorem 2, the initial item on our agenda, is one-dimensional. On $I\!R$, let $f(x) = \min\{x, 0\}$. The mapping $\partial f : I\!R \rightrightarrows I\!R$ is given then by

$$\partial f(x) = \begin{cases} \{0\} \text{ for } x > 0, \\ \{0, 1\} \text{ for } x = 0, \\ \{1\} \text{ for } x < 0. \end{cases}$$

Take $\bar{x} = 0$ and $\bar{v} = 0$, this being a general subgradient at $\bar{x}$ that is not a regular subgradient. As neighborhoods of $\bar{x}$ and $\bar{v}$ take $\mathcal{X} = (-1, 1)$ and $\mathcal{V} = (-1, 1)$ along with any $\rho > 0$. The corresponding f-attentive localization of ∂f, obtained by intersecting $\mathrm{gph}_\rho \partial f$ with $\mathcal{X} \times \mathcal{V}$, is then the mapping

$$x \mapsto \begin{cases} 0 \text{ if } x \in [0, 1) \\ \emptyset \text{ if } x \in (-1, 0). \end{cases}$$

That mapping is monotone, satisfying (1.7) for $s = 0$, yet f is not variationally convex at $\bar{x}$. It does not satisfy (1.8) for any convex function $\widehat{f}$, regardless of any possible adjustment of the localization.

With Theorem 2 out of the way, we turn to Theorem 1. All the concerns there are local, so to simplify notation and other aspects of the task ahead we can take

$$\bar{x} = 0, \qquad \bar{v} = 0, \qquad f(0) = 0, \qquad \mathrm{dom}\, f \text{ bounded.} \tag{2.1}$$

We can further then reduce to the case where $s = 0$, because in terms of $j(x) = \frac{1}{2}|x|^2$, the s-convexity of f relative to some convex set is equivalent to the convexity of $f - sj$ on that set, as is well known. That translates into the variational s-convexity in (a) corresponding to the variational convexity of $f - sj$. (Neighborhoods need adjustment in the translation, but that's easy.) In the same vein, the s-level uniform quadratic growth condition on f in (b) is equivalent to the 0-level version for $f - sj$, which simple examination confirms. Because

$$\partial[f - sj](x) = \partial f(x) - s\nabla j(x) = \partial f(x) - sx,$$

the s-monotone localizations of ∂f in (c) and (d) are equivalent to the existence of 0-monotone localizations of $f - sj$. Thus, by establishing the equivalence in Theorem 1 in the case of $s = 0$, we will establish it for every s.

The proof will proceed from here in the cycle of (a)$\Rightarrow$(b)$\Rightarrow$(d)$\Rightarrow$(c$'$)$\Rightarrow$(a), thereby covering (c) because of it being a property stronger than (d) but weaker than (c$'$). Understanding first of all that (a)$\Rightarrow$(b) is elementary. It's seen from the definition of variational convexity at level $s = 0$ in its identification of the subgradient properties of f locally with those of a convex function $\widehat{f}$, since (b) at level $s = 0$ is just the subgradient inequality of convex analysis. Understanding that (b)$\Rightarrow$(d) for $s = 0$ is also elementary. The combination of having $f(x_1) \geq f(x_0) + \langle v_0, x_1 - x_0 \rangle$ and $f(x_0) \geq f(x_1) + \langle v_1, x_0 - x_1 \rangle$ yields $\langle v_1 - v_0, x_1 - x_0 \rangle \geq 0$.

In working next toward confirming that (d)$\Rightarrow$(c$'$), we can rely on a strengthened interpetation of what it means for $\bar{v}$ to be a regular subgradient of f at $\bar{x}$ in the circumstances in (2.1). While by definition it just means $f(x) \geq o(|x|)$, it can be replaced by the more powerful condition that

$$\begin{array}{l} f(x) \geq -h(x) \text{ for all } x, \text{ equal only when } x = 0, \text{ where} \\ h \text{ is strongly convex and } \mathcal{C}^1 \text{ with } h(0) = 0,\ \nabla h(0) = 0. \end{array} \tag{2.2}$$

This depiction of regularity goes back to [14, 8.5], which in turn was derived from a normal cone result in [14, 6.11]. Although not made explicit in the statements of those results, the argument for [14, 6.11] produced such an h as $\theta(|x|)$ for a function θ that happens also to be strongly convex.[5]

Now let $g = f + h$, a closed proper function having the same (bounded) effective domain as f. By (2.2), g satisfies

$$g(x) > 0 \;\text{ for all } x \neq 0,\;\; g(0) = 0, \tag{2.3}$$

and further, by the calculus rule for obtaining the subgradients of the sum of a function with a smooth function [14, 8.8(c)], has

$$\partial g(x) = \partial f(x) + \nabla h(x) \;\text{ for all } x. \tag{2.4}$$

Then the conjugate function g^*, having[6] $-g^*(y) = \min_x \{\, g(x) - \langle y, x\rangle \}$, is finite convex and globally Lipschitz continuous (that being equivalent to the boundedness of $\operatorname{dom} g$) with

$$g^*(y) \geq 0 \;\text{ for all } y, \text{ while } g^*(0) = 0 \text{ and } \nabla g^*(0) = 0. \tag{2.5}$$

In convex analysis, the biconjugate function g^{**} is the convex hull of g (no closure necessary because of $\operatorname{dom} g$ being bounded), and

$$\begin{array}{c} y \in \partial g^{**}(x) \quad\Longleftrightarrow\quad x \in \partial g^*(y) \quad\Longleftrightarrow\quad x \in \operatorname{conv} M(y) \\ \text{for the mapping} \quad M:\; y \mapsto \operatorname{argmin}_x \{\, g(x) - \langle y, x\rangle \}, \end{array} \tag{2.6}$$

which is is nonempty-valued with closed graph,

$$\operatorname{gph} M = \{\, (y,x) \,|\, g(x) + g^*(y) - \langle y, x\rangle \leq 0, \;\text{ or equivalently, } \; = 0 \}. \tag{2.7}$$

From the subgradient relationship in (2.4) and the formula for the graph of M in (2.7) it's apparent that

$$M(y) \subset \partial g^*(y) \;\text{ and }\; x \in M(y) \Longrightarrow \begin{cases} f(x) = \langle y, x\rangle - g^*(y) - h(x), \\ y - \nabla h(x) \in \partial f(x). \end{cases} \tag{2.8}$$

The properties of g^* in (2.5), along with the boundedness and graphical closedness of its subgradient mapping ∂g^*, tell us that, as $y \to 0$, both $g^*(y) \to 0$ and $\partial g(y) \to \{0\}$. Hence by (2.8), there is a neighborhood $\mathcal{Y} \subset \mathcal{V}$ of 0 such that

$$x \in M(y),\; y \in \mathcal{Y} \quad\Longrightarrow\quad (x, y - \nabla h(x)) \in \mathcal{X} \times \mathcal{V} \;\text{ with }\; f(x) < \rho. \tag{2.9}$$

This has the consequence that M must be single-valued on $\mathcal{Y}$, because having $x_0, x_1 \in M(y)$ and therefore $y - \nabla h(x_i) \in \partial f(x_i)$ by (2.8) leads through our monotonicity assumption in (d) at level $s = 0$ to having

$$0 \leq \langle x_1 - x_0, [y - \nabla h(x_1)] - [y - \nabla h(x_0)] \rangle = -\langle x_1 - x_0, \nabla h(x_1) - \nabla h(x_0) \rangle.$$

But $\langle x_1 - x_0, \nabla h(x_1) - \nabla h(x_0)\rangle > 0$ unless $x_1 = x_0$ by the strong convexity of h in (2.2). Then ∂g^* is single-valued on $\mathcal{Y}$ as well by (2.6), reducing in that case to a gradient mapping:

$$M(y) = \nabla g^*(y) \;\text{ for }\; y \in \mathcal{Y}. \tag{2.10}$$

[5]The argument, recast from normals to subgradients, starts from defining $\theta_0(t) = -\inf \{ f(x) \,\big|\, |x - \bar{x}| \leq t \}$ on $[0, \infty)$, noting that this function is nondecreasing, lower semicontinuous and finite (as follows from f being lower semicontinuous). It takes $\theta_1(t) = \frac{1}{t}\int_t^{2t} \theta_0(s)ds$ and then defines $\theta_2(t) = \frac{1}{t}\int_t^{2t} \theta_1(s)ds$, showing that θ_2 is convex, differentiable and $\geq \theta_0$. Finally, it sets $\theta(t) = \theta_2(t) + t^2$.

[6]The notation for the vector dual to x switches here from v to y in order to retain v for later in an appeal to (2.4) as representing vectors $v \in \partial f(x)$ as differences $y - \nabla h(x)$.

This reveals importantly that the graphical localization of M to the neighborhood $\mathcal{Y}\times\mathcal{X}$ of $(0,0)$ is maximal monotone. So too then is the graphical localization of M^{-1} to $\mathcal{X}\times\mathcal{Y}$, with

$$\begin{aligned}(\mathcal{X}\times\mathcal{Y})\cap\operatorname{gph}M^{-1} &= (\mathcal{X}\times\mathcal{Y})\cap\operatorname{gph}\partial g^{**}\\ &= (\mathcal{X}\times\mathcal{Y})\cap\operatorname{gph}\partial g = (\mathcal{X}\times\mathcal{Y})\cap\operatorname{gph}[\partial f+\nabla h].\end{aligned} \tag{2.11}$$

Although we have proceeded only from assuming monotonicity in the localization $(\mathcal{X}\times\mathcal{V})\cap\operatorname{gph}_\rho\partial f$, we see from this that ∂f must be maximally monotone in a neighorhood $X_0\times\mathcal{V}_0$ of $(0,0)$ within $\mathcal{X}\times\mathcal{V}$, because otherwise there would be a contradiction to the maximal monotonicity in (2.11). Thus, (d)$\Rightarrow$(c$'$) is correct.

We turn now to verifying the final implication (c$'$)$\Rightarrow$(a), this task being the hardest. A similar pattern to that of (d)$\Rightarrow$(c$'$) will be observed, but new and different ideas will also be brought in to cope with a more challenging environment. We start in this case from having an open convex neighborhood $\mathcal{X}\times\mathcal{V}$ of $(0,0)$ and a $\rho>0$ such that the set

$$\mathcal{G} := (\mathcal{X}\times\mathcal{V})\cap\operatorname{gph}_\rho\partial f \tag{2.12}$$

is a maximally monotone graph in $\mathcal{X}\times\mathcal{V}$. By [14, 2.6] there then exists a maximal monotone mapping $T: I\!R^n \rightrightarrows I\!R^n$ such that

$$\mathcal{G} = (\mathcal{X}\times\mathcal{V})\cap\operatorname{gph}T. \tag{2.13}$$

Through Minty parameterization [14, 12.14+12.15], the mapping $(T+rI)^{-1}$ for any $r>0$ is single-valued and Lipschitz continuous from $I\!R^n$ onto $I\!R^n$ with Lipschitz constant r^{-1}. Relative to the open convex set

$$\mathcal{Y}_r = \{\, y = x+rv) \,|\, (x,v)\in\mathcal{G}\},\ \text{ with }\ 0\in\mathcal{Y}_r, \tag{2.14}$$

it furnishes the representation

$$(x,v)\in\mathcal{G}\ \longleftrightarrow\ y\in\mathcal{Y}_r\ \text{ one-to-one, with }\ x=(T+rI)^{-1}(y),\ v=y-rx. \tag{2.15}$$

By [14, 8.47], every pair $(x,v)\in\mathcal{G}$ can be approximated arbitrarily closely by a pair $(x,v)\in\mathcal{G}$ in which v is a proximal subgradient of f at x, thereby satisfying

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

locally around x for some $r>0$. Because $\operatorname{dom}f$ is bounded, this inequality can be made global, moreover strict away from x, merely by increasing r if necessary. Then, in terms of the function

$$g_r = f+rj\ \text{ for }\ j(x)=\tfrac{1}{2}|x|^2,\ \text{ having }\ \partial g_r=\partial f+rI,\ g_r(0)=0,\ 0\in\partial g_r(0) \tag{2.16}$$

by (2.1), the strict global inequality can be rewritten as

$$g_r(x')-\langle y,x'\rangle > g_r(x)-\langle y,x\rangle\ \text{ for all } x'\neq x, \text{ where }\ y=v+rx. \tag{2.17}$$

That way, in describing th "globally proximal" subgradient relationship for pairs (x,v) by the nested families of sets

$$\mathcal{G}^r_{\text{prox}} := \{\,(x,v)\,|\ (2.17)\text{ holds}\,\},\qquad \mathcal{Y}^r_{\text{prox}} = \{\,y=v+rx\,|\,(x,v)\in\mathcal{G}^r_{\text{prox}}\}, \tag{2.18}$$

we have $\bigcup_{r>0}\mathcal{G}^r_{\text{prox}}\cap\mathcal{G}$ dense in $\mathcal{G}$ and $\bigcup_{r>0}\mathcal{Y}^r_{\text{prox}}\cap\mathcal{Y}_r$ dense in $\mathcal{Y}_r$. Furthermore, in terms of the mapping $M_r: I\!R^n\rightrightarrows I\!R^n$ defined by

$$M_r(y) := \operatorname{argmin}_x\Big\{\,g_r(x)-\langle y,x\rangle\Big\} \tag{2.19}$$

we have that M_r is single-valued on $\mathcal{Y}^r_{\rm prox}$, with the unique $x = M_r(y)$ having $(x, y-rx) \in \mathcal{G}^r_{\rm prox}$. However, then by (2.5),

$$M_r(y) = \{(T+rI)^{-1}(y)\} \;\text{ when }\; y \in \mathcal{Y}^r_{\rm prox} \cap \mathcal{Y}_r, \tag{2.20}$$

while also more generally,

$$M_r(y) = \{(T+rI)^{-1}(y)\} \;\text{ on }\; \mathcal{O}_r := \{\, y \,|\, M_r(y) \subset \mathcal{X},\; y - rM_r(y) \subset \mathcal{V}\} \subset \mathcal{Y}_r. \tag{2.21}$$

Although potentially set-valued in general, apart from (2.20) and(2.21), the mapping M_r has closed graph and is everywhere nonempty-valued with $M_r(y) \subset \operatorname{dom} f$ (bounded). Moreover

$$\operatorname{conv} M_r(y) = \partial g_r^*(y) \;\text{ for }\; g_r^*(y) = \max_x \{\, \langle y, x\rangle - g_r(x)\}, \tag{2.22}$$

as will be recalled from the pattern in the proof of (d)$\Rightarrow$(c$'$), but here with $g_r = f + rj$ instead of $g = f + h$. From the graphical closedness and local boundedness, the set $\mathcal{O}_r$ in (2.12) is open [14, 5.19]. On the other hand, the set

$$\mathcal{C}_r := \{\, y \in \mathcal{Y}_r \,|\, M_r(y) = \{(T+rI)^{-1}(y)\}\}$$

is closed relative to $\mathcal{Y}_r$. We have from (2.20) and (2.21) that

$$\mathcal{Y}^r_{\rm prox} \cap \mathcal{Y}_r \;\subset\; \mathcal{C}_r \;\subset\; \mathcal{O}_r \;\subset\; \mathcal{C}_r,$$

where the first set on the left is nonempty, at least when r is large enough, through the density of the union of such sets in $\mathcal{Y}$. Thus $\mathcal{C}_r$ and $\mathcal{O}_r$ are actually the same nonempty subset of $\mathcal{Y}_r$, simultaneously open and closed relative to $\mathcal{Y}_r$. However, the only such set, lacking a boundary point, is $\mathcal{Y}_r$ itself. We must have

$$M_r(y) = \{(T+rI)^{-1}(y)\} = \nabla g_r^*(y) \;\text{ for all }\; y \in \mathcal{Y}_r, \tag{2.23}$$

where the gradient comes in through (2.22) and the fact that a convex function is differentiable at a given point if and only if its subgradient set is a singleton. Note from the equation in (2.23) that, because $\operatorname{gph} M_r = \{\, (x,y) \,|\, g_r(x) + g_r^*(y) = \langle x, y\rangle\}$ by (2.19) and (2.22), we have

$$y \in \mathcal{Y}_r,\; x = \nabla g_r^*(y) \quad\Longrightarrow\quad g_r^{**}(x) = g_r(x) = \langle y, x\rangle - g_r^*(y) \;\text{ with }\; y \in \partial g_r^{**}(x) \tag{2.24}$$

and in particular out of (2.16)

$$g_r^*(0) = 0, \quad \nabla g_r(0) = 0, \quad g_r^{**}(0) = 0, \quad 0 \in \partial g_r^{**}(0). \tag{2.25}$$

Recall now that $(T+rI)^{-1}$ is Lipschitz continuous with Lipschitz constant r^{-1}. By (2.23), this property is inherited by ∇g_r^* on $\mathcal{Y}^r$, so $|\nabla g_r^*(y_1) - \nabla g_r^*(y_0)| \le r^{-1}|y_1 - y_0|$ for $y_0, y_1 \in \mathcal{Y}_r$. Then, since $\mathcal{Y}_r$ is a convex set, we have for any $y, y' \in \mathcal{Y}_r$

$$g_r^*(y') - g_r^*(y) - \langle \nabla g_r^*(y), y' - y\rangle = \int_0^1 \langle \nabla g_r^*(y + t(y'-y)) - \nabla g_r^*(y)\rangle, y' - y\rangle dt,$$

where $\langle \nabla g_r^*(y + t(y'-y)) - \nabla g_r^*(y)\rangle, y' - y\rangle \le r^{-1} t |y' - y|^2$. In consequence,

$$g_r^*(y') \le g_r^*(y) + \langle \nabla g_r^*(y), y' - y\rangle + r^{-1} j(y' - y) \;\text{ for all }\; y, y' \in \mathcal{Y}_r^*.$$

Putting this another way, we have for any $y \in \mathcal{Y}_r$ that

$$g_r^* \leq q_{r,y} + \delta_{\mathrm{cl}\,\mathcal{Y}_r} \text{ on } I\!R^n \text{ for } q_{r,y}(y') = g_r^*(y) + \langle \nabla g_r^*(y), y' - y\rangle + r^{-1} j(y' - y).$$

This dualizes through conjugacy and the sum rule in [14, 11.23] to

$$g_r^{**} \geq (q_{r,y} + \delta_{\mathrm{cl}\,\mathcal{Y}_r})^* \text{ on } I\!R^n \text{ with } (q_{r,y} + \delta_{\mathrm{cl}\,\mathcal{Y}_r})^* = q_{r,y}^* \# \sigma_{\mathrm{cl}\,\mathcal{Y}_r}, \tag{2.26}$$

where $\sigma_{\mathrm{cl}\,\mathcal{Y}_r}$ is the support function of $\mathrm{cl}\,\mathcal{Y}_r$ [14, 11.4] and $\#$ stands for epi-addition [14, 1H] (infimal convolution). With respect to the vector $x = \nabla g_r^*(y)$ we can take advantage of (2.24) to re-express

$$q_{r,y}(y') = -g_r^{**}(x) + \langle x, y'\rangle + r^{-1} j(y' - y),$$

and then calculate

$$\begin{aligned} q_{r,y}^*(x') &= \sup\nolimits_{y'} \{ \langle y', x'\rangle - q_{r,y}(y')\} = \sup\nolimits_u \{ \langle y + u, x'\rangle - q_{r,y}(y + u)\} \\ &= \sup\nolimits_u \{ \langle y + u, x'\rangle + g_r^{**}(x) - \langle x, y + u\rangle - r^{-1} j(u), \} \\ &= g_r^{**}(x) + \langle y, x' - x\rangle + \sup\nolimits_u \{ \langle u, x' - x\rangle - r^{-1} j(u)\} \\ &= g_r^{**}(x) + \langle y, x' - x\rangle + r j(x' - x), \end{aligned} \tag{2.27}$$

inasmuch as $(r^{-1}j)^* = rj$, getting as a byproduct of that

$$\nabla q_{r,y}^*(x') = y + r(x' - x). \tag{2.28}$$

By definition, $[q_{r,y}^* \# \sigma_{\mathrm{cl}\,\mathcal{Y}_r}](x')$ is the minimum of $q_{r,y}^*(x' - x'') + \sigma_{\mathrm{cl}\,\mathcal{Y}_r}](x'')$ in x'', which is sure to be attained uniquely because of the quadratic function $q_{r,y}^*$ is strongly convex. According to the elementary optimality condition for this minimum, it is attained at x'' if and only if $\nabla q_{r,y}^*(x')$ belongs to the set of subgradients of $\sigma_{\mathrm{cl}\,\mathcal{Y}_r}$ at x''. That set is in the boundary of $\mathrm{cl}\,\mathcal{Y}_r$ if $x'' \neq 0$, whereas for $x'' = 0$ it is all of $\mathrm{cl}\,\mathcal{Y}_r$. Therefore, as long as $\nabla q_{r,y}^*(x')$ lies in the open set $\mathcal{Y}_r$, the minimum will be attained at $x'' = 0$, and $[q_{r,y}^* \# \sigma_{\mathrm{cl}\,\mathcal{Y}_r}](x')$ will agree with $q_{r,y}^*(x')$. Through (2.27) and (2.28) we can translate (2.26) now to

$$g_r^{**}(x') \geq g_r^{**}(x) + \langle y, x' - x\rangle + r j(x' - x) \text{ when } y + r(x' - x) \in \mathcal{Y}_r, \tag{2.29}$$

where, since x is $\nabla g_r^*(y)$,

$$y + r(x' - x) \in \mathcal{Y}_r \iff x' \in r^{-1}(\mathcal{Y}_r - y) + \nabla g_r^*(y). \tag{2.30}$$

We are looking at this for $y \in \mathcal{Y}_r$ and can appeal to having $\nabla g_r^*(y) \to 0$ as $y \to 0$ by (2.25) and to the notation

$$I\!B = \text{ the closed unit ball in } I\!R^n$$

to see in (2.29) and (2.30) that there exist $\lambda > 0$ and $\kappa > 0$ such that $\kappa I\!B \subset \mathcal{Y}_r$ and when $y \in \kappa I\!B$ also $\lambda I\!B \subset r^{-1}(\mathcal{Y}_r - y) + \nabla g_r^*(y)$. Then (2.29) and (2.30) give us

$$g_r^{**}(x') \geq g_r^{**}(x) + \langle y, x' - x\rangle + r j(x' - x) \text{ for all } x' \in \lambda I\!B \text{ when } y \in \kappa I\!B.$$

Using $j(x' - x) = j(x') + j(x) - \langle x, x'\rangle$, we can write this as in terms of the vector $v = y - rx \in \mathcal{V}$ as

$$g_r^{**}(x') - r j(x') \geq g_r^{**}(x) - r j(x) + \langle v, x' - x\rangle \text{ for all } x' \in \lambda I\!B \text{ when } y \in \kappa I\!B. \tag{2.31}$$

Here from (2.16) and (2.24) we have $g_r^{**}(x) - rj(x) = g_r(x) - rj(x) = f(x)$ while on the other hand $g_r^{**}(x') - rj(x') \leq g_r(x') - rj(x') = f(x')$. The inequality in (2.31) comes out that way as

$f(x') \geq f(x) + \langle v, x' - x\rangle$ for all $x' \in \lambda I\!B$, or $[f + \delta_{\lambda B}](x') \geq f(x) + \langle v, x' - x\rangle$ for all $x' \in I\!R^n$. Therefore, in taking

$$\widehat{f} = \text{conv}[f + \delta_{\lambda B}] = (f + \delta_{\lambda B})^{**},$$

we get a closed proper convex function that satisfies

$$\begin{array}{l} \widehat{f}(x') \geq \widehat{f}(x) + \langle v, x' - x\rangle, \quad f(x') \geq \widehat{f}(x'), \quad f(x) = \widehat{f}(x), \\ \text{for } x' \in \lambda I\!B \text{ when } x = \nabla g_r^*(y),\ v = y - rx, \text{ with } y \in \kappa I\!B. \end{array} \tag{2.32}$$

Remember next that the conditions on x and v in (2.32) entail (x, v) belonging to the localization $\mathcal{G}$ of ∂f from which we began. There is then a neighborhood $\mathcal{X}' \times \mathcal{V}'$ small enough that, on the basis of (2.32), we have

$$\begin{array}{l} \widehat{f}(x') \geq \widehat{f}(x) + \langle v, x' - x\rangle, \quad f(x') \geq \widehat{f}(x'), \quad f(x) = \widehat{f}(x) \\ \text{for } x' \in \mathcal{X}' \text{ whenever } (x, v) \in \mathcal{G}' = (\mathcal{X}' \times \mathcal{V}') \cap \text{gph}_\rho \partial f, \end{array} \tag{2.33}$$

with the $\mathcal{G}'$ localization inheriting maximal monotonicity in $\mathcal{X}' \times \mathcal{V}'$ from that of $\mathcal{G}$ in $\mathcal{X} \times \mathcal{V}$. In particular by (2.33), the pairs $(x, v) \in \mathcal{G}'$ belong to $\text{gph}\, \partial\widehat{f}$. But $\partial\widehat{f}$ is a monotone mapping, so there can't be any pairs $(x, v) \in (\mathcal{X}' \times \mathcal{V}') \cap \text{gph}\, \partial\widehat{f}$ other than these, because that would contradict the maximality of $\mathcal{G}'$. The requirements of variational convexity in (a) at level $s = 0$ are therefore satisfied, and the proof of Theorem 1 is finished. $\square$

References

[1] Benko, M., and Rockafellar, R.T., "Primal-dual stability in local optimality," submitted for publication 2023.

[2] Khanh,P.D., Khao, V.V.H., Mordukhovich, B.S., and Phat, V.T., "Variational and strong variational convexity in infinite-dimensional variational analysis," arXiv:2309.13779 (2023).

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

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

[5] Mordukhovich, B.S., and Sarabi, E., "Variational analysis and full stability of optimal solutions to constrained and minimax problems," *Nonlinear Analysis* **121** (2015), 36-53.

[6] Mordukhovich, B.S., and Rockafellar, R.T., "Second-order subdifferential calculus with applications to tilt stability in optimization," *SIAM J. Optimization* **22** (2012), 953–986.

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

[8] Poliquin, R.A., and Rockafellar, R.T., "Prox-regular functions in variational analysis," *Transactions Amer. Math. Soc.* **348** (1996), 1805–1838.

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

[10] Rockafellar, R.T., "Variational convexity and the local monotonicity of subgradient mappings," *Vietnam Journal of Mathematics* **47** (2019), 547–561.

[11] Rockafellar, R.T., "Progressive decoupling of linkages in optimization and variational inequalities with elicitable convexity or monotonicity," *Set-Valued and Variational Analysis* **27** (2019), 863–893.

[12] Rockafellar, R.T., "Augmented Lagrangians and hidden convexity in sufficient conditions for local optimality," *Math. Programming* **198** (2023), 159–194.

[13] Rockafellar, R.T., "Convergence of augmented Lagrangian methods in extensions beyond nonlinear programming," *Mathematical Programming* **199** (2023), 375–420.

[14] Rockafellar, R. T., and Wets, R.J-B, *Variational Analysis*, No. 317 in the series *Grundlehren der Mathematischen Wissenschaften*, Springer-Verlag, 1997.