Derivative Tests for Prox-Regularity and the Modulus of Convexity

R. Tyrrell Rockafellar[1]

Abstract

Prox-regularity is a fundamental property of functions in second-order variational analysis that has come to be understood as marking the boundary for convexity-like behavior. Broad classes of examples have long been familiar, but there has not been any pointwise test based on generalized second derivatives to check for its presence. Such tests are developed here in terms of newly defined strict second subderivatives and strict second-order subdifferentials. They also identify the exact associated level of variational s-convexity, whether positive or negative. A recent criterion of Gfrerer for the level of variational s-convexity of a function already identified as prox-regular is shown to follow also from the theory of generalized twice differentiability of convex functions and their conjugates. Tests of the strong variational sufficient condition for local optimality are obtained as an application.

Keywords: *second-order variational analysis, generalized second derivatives, modulus of convexity, variational convexity, local subdifferential monotonicity, uniform quadratic growth, prox-regularity, tilt stability, variational sufficiency.*

Version of 16 November 2024

[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

Prox-regularity of a function is a localized property of its subgradient mapping that demarcates the territory around variational convexity and local subdifferential monotonicity in second-order variational analysis. Here we characterize its presence, and the exact level of that presence, by new sorts of generalized second derivatives, including strict second-order subdifferentials based on strict graphical derivatives of subgradient mappings. In contrast, the familiar second subdifferentials based on graphical coderivatives of subgradient mappings are unable to provide a test for the presence of prox-regularity. In recent efforts parallel to ours, however, Gfrerer has shown in [2] that, with modification, they can at least serve in the known presence of prox-regularity to identify its level by a parallel formula, which moreover can be quantified by properties of special bundles of subspaces associated with graphs of subgradient mappings. We add to that by demonstrating how Gfrerer's bundle criterion can also be viewed as a consequence of extending to all prox-regular functions the theory of generalized twice differentiability previously articulated only for convex functions.

Prox-regularity arises from questions about proximal subgradients of an extended-real-valued function on f on $\mathbb{R}^n$, taken here always to be closed (lower semicontinuous) and proper in having nonempty $\operatorname{dom} f = \{\, x \mid f(x) < \infty\}$ and nowhere taking on $-\infty$. To begin with, a vector $\bar{v}$ is a *regular subgradient* of f at $\bar{x}$ if

$$f(x) \geq f(\bar{x}) + \bar{v}\cdot(x-\bar{x}) + o(|x-\bar{x}|), \;\; \text{notation} \;\; \bar{v} \in \widehat{\partial} f(\bar{x}), \tag{1.1}$$

and a (*general*) *subgradient* $\bar{v} \in \partial f(\bar{x})$ if there is a sequence $(x^\nu, v^\nu) \to (\bar{x}, \bar{v})$ with $f(x^\nu) \to f(\bar{x}) < \infty$ such that $v^\nu \in \widehat{\partial} f(x^\nu)$ [16, 8B]. It is a *proximal subgradient* if, more specifically, there is a neighborhood $\mathcal{X}$ of $\bar{x}$ and some r such that

$$f(x) \geq f(\bar{x}) + \bar{v}\cdot(x-\bar{x}) - rj(x-\bar{x}) \;\text{ for }\; x \in \mathcal{X}, \;\; \text{where} \;\; j(x) := \tfrac{1}{2}|x|^2, \tag{1.2}$$

which is equivalent to having in (1.1) an error of type $o(|x-\bar{x}|^2)$ instead.[2] Having f be *prox-regular* at $\bar{x}$ for $\bar{v}$ means having this hold uniformly in an f-attentively-local primal-dual manner. Such localization refers to restricting (x, v) to $\mathcal{X}^\alpha_f \times \mathcal{V}$ for an ordinary neighborhood $\mathcal{V}$ of $\bar{v}$ and an f*-attentive* neighborhood $\mathcal{X}^\alpha_f$ of $\bar{x}$, given by

$$\mathcal{X}^\alpha_f = \{\, x \in \mathcal{X} \mid f(x) < \alpha\} \;\text{ for an ordinary neighborhood } \mathcal{X} \text{ of } \bar{x} \text{ and } \alpha > f(\bar{x}), \tag{1.3}$$

with the uniformity condition that defines prox-regularity being

$$f(x') \geq f(x) + v\cdot(x'-x) - rj(x'-x) \;\text{ for }\; x' \in \mathcal{X}, \; (x,v) \in [\mathcal{X}^\alpha_f \times \mathcal{V}] \cap \operatorname{gph} \partial f. \tag{1.4}$$

The adjusted kind of $\mathcal{X}^\alpha_f$ "neighborhood"[3] in (1.3) enters (1.4) because of the role of function values in defining ∂f from $\widehat{\partial} f$. The truncated mapping having as its graph the intersection in (1.4) is an f*-attentive localization* of ∂f; with just $\mathcal{X}$ instead of $\mathcal{X}^\alpha_f$ it would be an *ordinary* localization.

Proximal subgradients and prox-regularity are usually described with r taken to be positive, but that's a superfluous restriction because, if the property holds for any r at all, it holds for all higher r. In fact, negative values of r are really important. In rewriting (1.4) with s in place of $-r$ as

$$\begin{gathered} f(x') \geq f(x) + v\cdot(x'-x) + sj(x'-x) \\ \text{for } \; x' \in \mathcal{X}, \; (x,v) \in [\mathcal{X}^\alpha_f \times \mathcal{V}] \cap \operatorname{gph} \partial f, \end{gathered} \tag{1.5}$$

[2]Here we are denoting the canonical norm on $\mathbb{R}^n$ by $|\cdot|$ and employing the classical notation that $o(t)$ refers to an expression such that $o(t)/t \to 0$ as $t \to 0$.

[3]This isn't an ordinary neighborhood in $\mathbb{R}^n$, because it doesn't necessarily have an ordinary interior that contains $\bar{x}$. Instead, it's a neighborhood of $\bar{x}$ in the f-attentive topology on $\mathbb{R}^n$, this being the weakest refinement of the ordinary topology of $\mathbb{R}^n$ with respect to which f is continuous.

we get what is called the *uniform quadratic growth property* at level s, at least when s is positive. Here we embrace the full spectrum for both properties by admitting all levels of r and s in $(-\infty,\infty)$. That way we achieve a wider perspective in which we see in both cases the same concept, just opposite ways of calibrating it, and moreover can see how prox-regularity connects with variational convexity.

As defined in [11], f is *variationally convex* at $\bar{x}$ for $\bar{v}\in\partial f(\bar{x})$ if, for a neighborhood $\mathcal{X}\times\mathcal{V}$ of $(\bar{x},\bar{v})$ and $\mathcal{X}_f^\alpha$ as in (1.3), there is a closed proper *convex* function $\widehat{f}$ such that

$$\begin{aligned}&\widehat{f}\leq f \text{ on } \mathcal{X} \text{ with } [\mathcal{X}\times\mathcal{V}]\cap\operatorname{gph}\partial\widehat{f} = [\mathcal{X}_f^\alpha\times\mathcal{V}]\cap\operatorname{gph}\partial f \text{ and}\\ &\widehat{f}(x)=f(x) \text{ at all } (x,v) \text{ in the joint intersection, such as } (\bar{x},\bar{v}).\end{aligned} \tag{1.6}$$

This property has the interpretation that, in a primal-dual local sense, the subgradients and function values of f behave exactly as if f is convex. More broadly in concept,

$$f \text{ is called } \textit{variationally } s\textit{-convex} \text{ if (1.8) holds with } \widehat{f} \text{ being } s\textit{-convex}, \tag{1.7}$$

where s-convexity of $\widehat{f}$ has several equivalent descriptions, such as obeying

$$\widehat{f}(\lambda x+(1-\lambda)x')\leq\lambda\widehat{f}(x)+(1-\lambda)\widehat{f}(x')-\lambda(1-\lambda)sj(x-x') \text{ for } \lambda\in(0,1), \tag{1.8}$$

but for our purposes it can be most usefully be identified with $\widehat{f}-sj$ being convex. Strong convexity corresponds to a positive s, while hypoconvexity allows s to be negative.

Variational convexity arose from efforts at determining when a subdifferential mapping might have a local property of monotonicity. Recall that a mapping $T:\mathbb{R}^n\rightrightarrows\mathbb{R}^n$ is *locally s-monotone* at $\bar{x}$ for $\bar{v}\in T(\bar{x})$ if there is a neighborhood $\mathcal{X}\times\mathcal{V}$ of $(\bar{x},\bar{v})$ such that

$$(x'-x)\cdot(v'-v)\geq s|x'-x|^2 \text{ for } (x,v),\,(x',v'), \text{ in } [\mathcal{X}\times\mathcal{V}]\cap\operatorname{gph}T, \tag{1.9}$$

with this property being *maximal* ("max" for short) if the graph of T can't be enlarged in $\mathcal{X}\times\mathcal{V}$ without upsetting it. For $s=0$, s-monotonicity is plain *monotonicity*, while for $s>0$ it is strong monotonicity; $s<0$ signals hypomonotonicity. Local s monotonicity of T corresponds, with an adjustment of location, to local s-monotonicity of $T+sI$. Subdifferential mappings of closed proper s-convex functions are globally max s monotone in particular, cf. [16, 12.17]. In applying monotonicity concepts to ∂f for general f, however, it's best to pass to *f-local* s-monotonicity by changing (1.9) to

$$(x'-x)\cdot(v'-v)\geq s|x'-x|^2 \text{ for } (x,v),\,(x',v'), \text{ in } [\mathcal{X}_f^\alpha\times\mathcal{V}]\cap\operatorname{gph}\partial f. \tag{1.10}$$

The tight relationship between all these concepts is featured in the following statement, which provides the platform on which we will construct much more. Note that the cases for general s can be obtained from the case for a single $\bar{s}$, say $\bar{s}=0$, by applying that to $g(x)=f(x)+(\bar{s}-s)j(x-\bar{x})$, which still has $\bar{v}\in\partial g(\bar{x})$. This kind of shift will assist us over and over in simplifying proofs.

Theorem 1.1 (enhanced restatement of [15, Theorem 1]). *The following properties at $\bar{x}$ for $\bar{v}\in\partial f(\bar{x})$ are equivalent for any $s\in(-\infty,\infty)$ and correspond to f being prox-regular at level $r=-s$:*

(a) *f has uniform quadratic growth at level s,*
(b) *f is variationally s-convex,*
(c) *∂f is f-locally s-monotone and $\bar{v}\in\widehat{\partial}f(\bar{x})$.*
(d) *∂f is f-locally max s-monotone,*
(e) *∂f has an $\mathcal{X}_f^\alpha\times\mathcal{V}$ localization that is max s-monotone relative to $\mathcal{X}\times\mathcal{V}$.*

This theorem statement differs from the original in [15] in replacing prox-regularity by its equivalent description as uniform quadratic growth in (a), and by the addition of (e), which is immediate from (b) in its reduction of properties of ∂f to those of $\partial\widehat{f}$ for a closed proper convex function $\widehat{f}$.

Prox-regularity has typically been combined in applications with subdifferential continuity in the sense of $f(x)$ depending continuously on $(x,v) \in [\mathcal{X} \times \mathcal{V}] \cap \operatorname{gph} \partial f$. But Gfrerer has shown in [2, Lemma 3.4] that prox-regularity automatically entails the existence of $\mathcal{X}_f^\alpha \times \mathcal{V}$ such that

$$\begin{array}{l} \text{the closure of } [\mathcal{X}_f^\alpha \times \mathcal{V}] \cap \operatorname{gph} \partial f \text{ lies in } \operatorname{gph} \partial f, \text{ and} \\ f(x) \text{ depends continuously on } (x,v) \text{ in that closure.} \end{array} \tag{1.11}$$

This property, which we'll call *f-attentive subdifferential continuity* at $\bar{x}$ for $\bar{v}$, builds on a much earlier observation in [8, Lemma 2.3] that prox-regularity entails having $f(x)$ depend continuously on $(x,v) \in [\mathcal{X}_f^\alpha \times \mathcal{V}] \cap \operatorname{gph} \partial f$. In fact, f-attentive subdifferential continuity can be seen from Theorem 1.1 to be a consequence of (c), like (e), in simply reflecting behavior known for convex functions. That substitute for the original concept of subdifferential continuity should work just as well for most theoretical purposes. As pointed out in [2], for example, where in the past it was shown that in the presence of both prox-regularity and subdifferential continuity, variational strong convexity is necessary and sufficient for tilt stability [9], the assumption of subdifferential continuity is redundant.

The watershed question in finite-dimensional second-order variational analysis is therefore whether prox-regularity is present or not, and if it is, then at what level. When the equivalent properties of Theorem 1.1 hold for some s, they also hold for every lower s, so the s range, if nonempty, is an interval unbounded from below. We can aim at understanding the upper bound to that interval as a special value associated with f at $\bar{x}$ for $\bar{v}$.

Definition 1.2 (convexity modulus, adapted from Gfrerer [2]). *The modulus of convexity of f at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ is the common value $\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) \in [-\infty,\infty]$ that, according to Theorem 1.1, can be identified equivalently as the modulus of variational convexity,*[4]

$$\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) := \begin{cases} \textit{limsup of the possible variational } s\textit{-convexity levels} \\ \textit{in (1.6) as } \mathcal{X} \times \mathcal{V} \textit{ shrinks down to } (\bar{x},\bar{v}),\ \alpha \searrow f(\bar{x}), \end{cases} \tag{1.12}$$

the modulus of local monotonicity,

$$\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) = \begin{cases} \textit{limsup of the possible } f\textit{-local } s\textit{-monotonicity levels} \\ \textit{in (1.10) as } \mathcal{X} \times \mathcal{V} \textit{ shrinks down to } (\bar{x},\bar{v}),\ \alpha \searrow f(\bar{x}), \end{cases} \tag{1.13}$$

the modulus of uniform quadratic growth,

$$\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) = \begin{cases} \textit{limsup of the possible } s\textit{-uniform quadratic growth levels} \\ \textit{in (1.5) as } \mathcal{X} \times \mathcal{V} \textit{ shrinks down to } (\bar{x},\bar{v}) \textit{ and } \alpha \searrow f(\bar{x}), \end{cases} \tag{1.14}$$

and in reversed sign as the modulus of prox-regularity,

$$-\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) = \begin{cases} \textit{liminf of the possible } r\textit{-prox-regularity levels} \\ \textit{in (1.4) as } \mathcal{X} \times \mathcal{V} \textit{ shrinks to } (\bar{x},\bar{v}),\ \alpha \searrow f(\bar{x}). \end{cases} \tag{1.15}$$

Note that, by the conventions for a supremum, the value coming from the first three formulas is $-\infty$ when the set of s levels in question is empty.

[4] Gfrerer in [2] defined that modulus by this formula, but the fact that several very different formulas produce the same value is, in our view, vitally important — hence the need for a common modulus that calibrates the hidden "quantum" of these common properties aligned with convex analysis at this location.

Corollary 1.3 (modulus test for the presence or absence of prox-regularity). *Prox-regularity of f at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ fails if and only if* $\operatorname{cnv} f(\bar{x} \,|\, \bar{v}) = -\infty$.

This is valuable in telling us that any formula capable of determining $\operatorname{cnv} f(\bar{x} \,|\, \bar{v})$ without knowing in advance whether or not it might be $-\infty$ will furnish a test for prox-regularity.

The central goal in this framework, then, is developing formulas for determining $\operatorname{cnv} f(\bar{x} \,|\, \bar{v})$. Such formulas have recently been obtained by Gfrerer in [2] for situations in which f is already known to be prox-regular at $\bar{x}$ for $\bar{v}$. They reveal important information but can't serve as tests for prox-regularity in the manner indicated by Corollary 1.3. Here we'll contribute formulas different from Gfrerer's which, along with other capabilities, do provide tests. In Section 2, $\operatorname{cnv} f(\bar{x} \,|\, \bar{v})$ will be assessed by using a new *strict second-subderivative* $d^2_* f(\bar{x} \,|\, \bar{v})$. In Section 3, the engine will instead be a new *strict second-order subdifferential* $\partial^2_* f(\bar{x} \,|\, \bar{v})$ and its levels of "s-definiteness," which extrapolate up and down from positive-semidefiniteness ($s = 0$) and positive-definiteness ($s > 0$). The new subdifferential concept employs strict graphical derivatives of set-valued mappings, but with the twist that, in application to ∂f, f-attentive convergence is brought in. The modulus formula we obtain closely resembles the formula obtained by Gfrerer [2] by using an f-attentive modification of the coderivative-based second-order subdifferential $\partial^2 f(\bar{x} \,|\, \bar{v})$, but unlike that, it can operate without assuming prox-regularity, just that $\bar{v}$ itself is a proximal subgradient at $\bar{x}$. In Section 4, we work instead at presenting an alternative statement and derivation of a different formula of Gfrerer [2], which came out of his research with Outrata in [3] on subspaces contained in the graphs of graphical derivatives. There, prox-regularity has to be assumed, but the formula has special attraction because of parallels with classical second-order ideas and a better potential for application in some cases. Section 5 applies the tests to verifying the strong variational sufficient condition for local optimality in [12] and [13].

All of the described formulas for $\operatorname{cnv} f(\bar{x} \,|\, \bar{v})$ offer criteria, in particular, for tilt stability and its modulus, since that's known to be reciprocal to the value expressed by (1.12), cf. Gfrerer [2, Theorem 5.1]. They also relate to properties of Moreau envelopes through the connections in [4].

Might these formulations and results of second-order variational analysis in finite dimensions carry over to infinite dimensions? To some extent, even though many of the arguments are compactness-oriented, but infinite-dimensional results on variational convexity and its connections with local monotonicity of subgradient mappings, such as in [5] and the comprehensive new book [6], have so far always relied on first assuming not only prox-regularity but also subdifferential continuity. Whether there is a full-scale counterpart to Theorem 1.1, our touchstone here, remains to be seen.

2 Tests utilizing strict second subderivatives

Second-order variational analyis as laid out in [16, Chapter 13] revolves around second-order difference quotient functions in the form

$$\Delta^2_\tau f(x \,|\, v)(\xi) := \Big[f(x + \tau\xi) - f(x) - \tau\xi{\cdot}v \Big] \Big/ \tfrac{1}{2}\tau^2 \quad \text{for } \tau > 0 \text{ and } v \in \partial f(x). \tag{2.1}$$

It associates with f at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ the basic *second subderivative* function $d^2 f(\bar{x} \,|\, \bar{v}) : I\!R^n \to \overline{I\!R} = [-\infty, \infty]$ defined by

$$d^2 f(\bar{x} \,|\, \bar{v})(\xi) = \liminf_{\xi' \to \xi,\ \tau \searrow 0} \Delta^2_\tau f(\bar{x} \,|\, \bar{v})(\xi'), \tag{2.2}$$

which for instance in the case of f being $\mathcal{C}^2$ and $\bar{v} = \nabla f(\bar{x})$ comes out as equal to $\xi{\cdot}\nabla^2 f(\bar{x})\xi$. We consider now the alternative function obtained by taking the limit in (2.2) with $(\bar{x}, \bar{v})$ on the right replaced by pairs $(x, v) \in \operatorname{gph} \partial f$ that approach it f-attentively.

Definition 2.1 (strict second subderivative function). *The function $d^2_*f(\bar{x}\,|\,\bar{v}) : I\!R^n \to \overline{I\!R}$ defined by*

$$d^2_*f(\bar{x}\,|\,\bar{v})(\xi) = \liminf_{\substack{\xi'\to\xi,\ \tau\searrow 0\\ (x,v)\to(\bar{x},\bar{v}),\ f(x)\to f(\bar{x})}} \Delta^2_\tau f(x\,|\,v)(\xi'). \tag{2.3}$$

will be called the strict second subderivative f at $\bar{x}$ for $\bar{v}\in\partial f(\bar{x})$.

The form of the limit in (2.3) guarantees that the function $d^2_*f(\bar{x}\,|\,\bar{v})$ is closed (lower semicontinuous) and positively homogeneous of degree 2, satisfying

$$d^2_*f(\bar{x}\,|\,\bar{v})(\lambda\xi) = \lambda^2 d^2_*f(\bar{x}\,|\,\bar{v})(\xi) \ \text{ for } \lambda > 0, \tag{2.4}$$

although not necessarily proper. Whether or not it is proper, or takes on $-\infty$, will turn out to signal whether or not f is prox-regular at $\bar{x}$ for $\bar{v}$. For $f\in\mathcal{C}^2$, the extended liminf in (2.3) yields the same result as the liminf in (2.2), namely $\xi{\cdot}\nabla^2 f(\bar{x})\xi$, but that's essentially because $\nabla^2 f(x)$ depends continuously on x. The formula is being applied here, though, to functions f without making any assumptions of second-order or even first-order differentiability. Perhaps surprisingly, the result of the construction nonetheless captures something fundamental about the local behavior of f.

Theorem 2.2 (modulus formula from strict second subderivatives). *In general when $\bar{v}\in\partial f(\bar{x})$,*

$$\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) = \inf_{|\xi|=1} d^2_*f(\bar{x}\,|\,\bar{v})(\xi), \quad \textit{where the infimum is attained when finite.} \tag{2.5}$$

Proof. The description of $\operatorname{cnv} f(\bar{x}\,|\,\bar{v})$ in (1.14) will be the key, but to tie it to strict second subderivatives we must begin by showing that the property in (1.5) on which it depends can be recast as the existence of $\mathcal{X}^\alpha_f\times\mathcal{V}$ some $\beta>0$ such that

$$\begin{aligned} &f(x') \geq f(x) + v{\cdot}(x'-x) + sj(x'-x)\\ &\quad\text{when } \ |x'-x|\leq\beta,\ (x,v)\in[\mathcal{X}^\alpha_f\times\mathcal{V}]\cap\operatorname{gph}\partial f.\end{aligned} \tag{2.6}$$

For this, it's enough to compare two ways of specifying an open set, on the one hand in terms of $\varepsilon>0$, on the other hand with $\delta>0$ and $\beta>0$:

$$\begin{aligned} U_\varepsilon &= \{\,(x',x,v)\,|\,|x'-\bar{x}|<\varepsilon,\ |x-\bar{x}|<\varepsilon,\ |v-\bar{v}|<\varepsilon\},\\ V_{\delta,\beta} &= \{\,(x',x,v)\,|\,|x'-x|<\beta,\ |x-\bar{x}|<\delta,\ |v-\bar{v}|<\delta\}.\end{aligned}$$

Obviously, given δ and β we can get ε such that $U_\varepsilon\subset V_{\delta,\beta}$ by taking $\varepsilon=\min\{\delta,\beta\}$. But also, given ε we can get δ and β such that $V_{\delta,\beta}\subset U_\varepsilon$ but taking $\delta+\beta\leq\varepsilon$. The formula in (1.14) can thus be written equivalently as

$$\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) = \begin{cases}\text{limsup of the possible } s\text{-uniform quadratic growth levels}\\ \text{in (2.6) as } \mathcal{X}\times\mathcal{V} \text{ shrinks down to } (\bar{x},\bar{v}),\ \alpha\searrow f(\bar{x}),\ \beta\searrow 0.\end{cases} \tag{2.7}$$

A change of variables will take this further. Vectors x' can be expressed in the form $x+\tau\xi$ with $\tau\geq 0$ and $|\xi|=1$ to rewrite (2.6) as $f(x+\tau\xi)\geq f(x)+\tau v{\cdot}\xi + s\tau^2 j(\xi)$ for $|\xi|=1$ when $\tau\in[0,\beta]$ and $(x,v)\in[\mathcal{X}^\alpha_f\times\mathcal{V}]\cap\operatorname{gph}\partial f$, where the function inequality is trivial when $\tau=0$ and otherwise means $\Delta^2_\tau f(x\,|\,v)(\xi)\geq s$. That lets us turn (2.7) into

$$\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) = \limsup_{\substack{\beta\searrow 0\\ (x,v)\to(\bar{x},\bar{v}),\ f(x)\to f(\bar{x})}} \ \min_{|\xi|=1,\,\tau\in(0,\beta]} \Delta^2_\tau f(x\,|\,v)(\xi),$$

which is easy to reconcile with the claimed formula in (2.5) based on (2.3). The attainment of the infimum in (2.5) when finite is guaranteed by the lower semicontinuity of $d^2_*f(\bar{x}\,|\,\bar{v})(\xi)$. □

Applying the formula in (2.5) by assessing second subderivatives directly from the defining formula looks like no easy matter. There are, though, some rules of calculus that might help.

Proposition 2.3 (elementary addition rule for strict second subderivatives)**.** *Suppose $f = f_1 + f_2$ with $f_2 \in \mathcal{C}^2$, so that $\partial f(x) = \partial f_1(x) + \nabla f_2(x)$, and let $\bar v \in \partial f(\bar x)$, setting $\bar v_1 = \bar v - \nabla f_2(\bar x)$. Then*

$$d^2_* f(\bar x \,|\, \bar v)(\xi) = d^2_* f_1(\bar x \,|\, \bar v_1)(\xi) + \xi{\cdot}\nabla^2 f_2(\bar x)\xi, \tag{2.8}$$

and consequently in Theorem 2.2

$$\operatorname{cnv} f(\bar x \,|\, \bar v) \;\geq\; \operatorname{cnv} f_1(\bar x \,|\, \bar v_1) + \text{lowest eigenvalue of } \nabla^2 f_2(\bar x). \tag{2.9}$$

Proof. The subgradient formula is known from [16, 8.8]. As second-order difference quotients these circumstances give us $\Delta^2_\tau f(x \,|\, v)(\xi) = \Delta^2_\tau f_1(x \,|\, v - \nabla f_2(x))(\xi) + \Delta^2_\tau f_2(x \,|\, \nabla f_2(x))(\xi)$. The last term converges uniformly on bounded sets to $\xi{\cdot}\nabla^2 f_2(\bar x)\xi$ on the basis of f_2 being $\mathcal{C}^2$, and that makes the conclusion obvious. □

To go beyond the elementary rule we need *horizon subgradients* $v \in \partial^\infty f(x)$, which are defined for f as the possible limits of sequences $\lambda^\nu v^\nu$ with $v^\nu \in \widehat\partial f(x^\nu)$ when $\lambda^\nu \searrow 0$ and $(x^\nu, v^\nu) \to (\bar x, \bar v)$ with $f(x^\nu) \to f(x)$ [16, 8B].

Proposition 2.4 (general addition rule for strict second subderivatives)**.** *Suppose that $f = f_1 + f_2$ and $\bar x$ satisfies the constraint qualification that no nonzero $v \in \partial^\infty f_1(\bar x)$ has $-v \in \partial^\infty f_2(\bar x)$, in which case $\partial f(x) \subset \partial f_1(x) + \partial f_2(x)$ for all x in an f-attentive neighborhood of $\bar x$. Then for any $\bar v \in \partial f(\bar x)$,*

$$d^2_* f(\bar x \,|\, \bar v)(\xi) \geq \inf_{v_i \in \partial f_i(\bar x),\, v_1 + v_2 = v} \Big\{ d^2_* f_1(\bar x \,|\, v_1)(\xi) + d^2_* f_2(\bar x \,|\, v_2)(\xi) \Big\}, \tag{2.10}$$

and consequently in Theorem 2.2

$$\operatorname{cnv} f(\bar x \,|\, \bar v) \geq \inf_{v_i \in \partial f_i(\bar x),\, v_1 + v_2 = v} \{ \operatorname{cnv} f_1(\bar x \,|\, v_1) + \operatorname{cnv} f_2(\bar x \,|\, v_2) \}. \tag{2.11}$$

For efficiency, instead of proving Proposition 2.4 directly, we postpone that until after establishing the following, broader result from which it can be derived as a special case.

Proposition 2.5 (chain rule for strict second subderivatives)**.** *Suppose that $f(x) = g(F(x))$ for a $\mathcal{C}^2$ mapping $F : I\!R^n \to I\!R^m$ and a closed proper function g on $I\!R^m$. Let $\bar x$ satisfy the constraint qualification that no nonzero $y \in \partial^\infty g(F(\bar x))$ has $\nabla[yF](\bar x) = 0$, where $[yF](x) = y{\cdot}F(x)$, in which case $\partial f(x) \subset \{\, v = \nabla[yF](x) \,|\, y \in \partial g(F(x)) \}$ for all x in some f-attentive neighborhood of $\bar x$. Then for any $\bar v \in \partial f(\bar x)$,*

$$d^2_* f(\bar x \,|\, \bar v)(\xi) \geq \inf_{y \in \partial g(F(\bar x)),\, \nabla[yF](\bar x) = \bar v} \Big\{ d^2_* g(F(\bar x) \,|\, y)(\nabla F(\bar x)\xi) + \xi{\cdot}\nabla^2 [yF](\bar x)\xi \Big\}, \tag{2.12}$$

and consequently in Theorem 2.2

$$\operatorname{cnv} f(\bar x \,|\, \bar v) \geq \inf_{y \in \partial g(F(\bar x)),\, \nabla[yF](\bar x) = \bar v} \{ \operatorname{cnv} g(F(\bar x) \,|\, y) + \text{lowest eigenvalue of } \nabla^2 [yF](\bar x) \}. \tag{2.13}$$

Proof. The subgradient formula specializes [16, 10.6] in combination with the observation that the constraint qualification persists for all x in some f-attentive neighborhood X^α_f of $\bar x$. In fact the constraint qualification guarantees, f-attentively around $(\bar x, \bar v)$, that the set-valued mapping

$$Y : \operatorname{gph} \partial f \rightrightarrows I\!R^m \;\text{ given by }\; Y(x, v) = \{\, y \in \partial g(F(x)) \,|\, \nabla[yF](x) = v \} \tag{2.14}$$

is nonempty-compact-valued and locally bounded with closed graph. With this in mind, consider any ξ with $|\xi| = 1$ and any sequences $\xi^\nu \to \xi$, $\tau \searrow 0$ and $(x^\nu, v^\nu) \to (\bar{x}, \bar{v})$ with $f(x^\nu) \to f(\bar{v})$ such that

$$d_*^2 f(\bar{x} \,|\, \bar{v}) = \lim_{\nu \to \infty} \Delta_{\tau^\nu}^2 f(x^\nu \,|\, v^\nu)(\xi^\nu), \tag{2.15}$$

as exist by Definition 2.1. Through the constraint qualification and passing to subsequences if necessary, we can choose $y^\nu \in Y(x^\nu, v^\nu)$ with $y^\nu \to y$ for some $y \in Y(\bar{x}, \bar{v})$. For such y^k we have

$$\Delta_{\tau^\nu}^2 f(x^\nu \,|\, v^\nu)(\xi^\nu) = \Big[g(F(x^\nu + \tau^\nu \xi^\nu)) - g(F(x^\nu)) - \tau^\nu \nabla [y^\nu F](x^\nu) \cdot \xi^\nu \Big] \Big/ \tfrac{1}{2} {\tau^\nu}^2. \tag{2.16}$$

In terms of $\Delta_\tau F(x)(\xi) := \tau^{-1}[F(x + \tau\xi) - F(x)]$, for which $F(x + \tau\xi) = F(x) + \tau \Delta_\tau F(x)(\xi)$ and

$$\Delta_\tau^2 [yF](x \,|\, v)(\xi) = \Big[[yF](x + \tau\xi) - [yF](x) - \tau v \cdot \xi \Big] \Big/ \tfrac{1}{2}\tau^2 = \Big[\tau y \cdot \Delta_\tau F(x)(\xi) - \tau v \cdot \xi \Big] \Big/ \tfrac{1}{2}\tau^2,$$

we can rewrite (2.16) as

$$\begin{gathered} \Delta_{\tau^\nu}^2 f(x^\nu \,|\, v^\nu)(\xi^\nu) = \Delta_{\tau^\nu}^2 g(F(x^\nu) \,|\, y^\nu)(\omega^\nu) + \Delta_{\tau^\nu}^2 [y^\nu F](x^\nu \,|\, v^\nu)(\xi^\nu), \\ \text{where } \ \omega^\nu = \Delta_{\tau^\nu} F(x^\nu)(\xi^\nu). \end{gathered} \tag{2.17}$$

This combines with (2.15) to give us

$$d_*^2 f(\bar{x} \,|\, \bar{v})(\xi) \ \geq \ d_*^2 g(F(\bar{x}) \,|\, y)(\nabla F(\bar{x})\xi) + \xi \cdot \nabla^2 [yF](\bar{x})\xi, \tag{2.18}$$

inasmuch as the $\mathcal{C}^2$ assumption on F makes $\Delta_\tau F(x)(\xi)$ and $\Delta_\tau^2 [yF](x \,|\, v)(\xi)$ converge uniformly on bounded ξ-sets to $\nabla F(\bar{x}) \cdot \xi$ and $\xi \cdot \nabla^2 [yF](\bar{x})\xi$; the inequality comes from the fact that the liminf for $d_*^2 g(F(\bar{x}) \,|\, y)(\nabla F(\bar{x})\xi)$ may involve more possibilities for $\omega^\nu \to \omega = \nabla F(\bar{x})(\xi)$ than those of form $\omega^\nu = \Delta_{\tau^\nu} F(x^\nu)(\xi^\nu)$ dictated by (2.17). The particular $y \in Y(\bar{x}, \bar{v})$ in (2.18) was obtained in dependence on ξ, so to get a lower bound that works simultaneously for all ξ we need to pass to the infimum in (2.11). □

Proof of Proposition 2.4. Apply Proposition 2.5 to the case of $g(x_1, x_2) = f_1(x_1) + f_2(x_2)$ on $I\!R^n \times I\!R^n$ and the linear mapping $F : x \mapsto (x, x) \in I\!R^n \times I\!R^n$. Then the constraint qualification comes out as indicated in the statement of Proposition 2.4, and likewise the subgradient formula based on having $\partial g(x_1, x_2)) = (\partial f_1(x_1), \partial f_2(x_2))$, so $y = (v_1, v_2)$ and $[yF](x) = [v_1 + v_2] \cdot x$ with $\nabla [yF](x) = v_1 + v_2$ and $\nabla^2 [yF](x) = 0$-matrix. □

3 Tests utilizing strict second-order subdifferentials

Strict graphical derivatives will need adaptation to work best with subdifferential mappings. In preparation for that, we recall the picture they provide for any set-valued mapping $T : I\!R^n \rightrightarrows I\!R^n$, without making assumptions other than $\bar{v} \in T(\bar{x})$.

For background, the (plain) *graphical derivative* of T at $\bar{x}$ for $\bar{v}$ is the mapping $DT(\bar{x} \,|\, \bar{v}) : I\!R^n \rightrightarrows I\!R^n$ having as its graph the tangent cone to $\operatorname{gph} T$ at $(\bar{x}, \bar{v})$, or equivalently to $\operatorname{cl} \operatorname{gph} T$ there. In terms of difference quotient mappings defined by

$$\Delta_\tau T(x \,|\, v) : \xi \mapsto [T(x + \tau\xi) - v]/\tau \ \text{ for } \ \tau > 0, \tag{3.1}$$

this means

$$\operatorname{gph} DT(\bar{x} \,|\, \bar{v}) = \limsup_{\tau \searrow 0} \operatorname{gph} \Delta_\tau T(x \,|\, v). \tag{3.2}$$

The strict graphical derivative of T at $\bar{x}$ for $\bar{v}$ is the mapping $D_*T(\bar{x}\,|\,\bar{v}) : I\!R^n \rightrightarrows I\!R^n$ obtained by taking instead a grander outer limit of difference quotient mappings,

$$\operatorname{gph} D_*T(\bar{x}\,|\,\bar{v}) = \limsup_{\substack{(x,v)\to(\bar{x},\bar{v}) \text{ in } \operatorname{gph} T \\ \tau \searrow 0}} \operatorname{gph} \Delta_\tau T(x\,|\,v). \tag{3.3}$$

Thus, in terms of sequences indexed by ν,

$$\begin{array}{l} (\xi,\psi) \in \operatorname{gph} D_*T(\bar{x}\,|\,\bar{v}) \quad\Longleftrightarrow\quad \exists\,(x^\nu,v^\nu) \to (\bar{x},\bar{v}) \text{ in } \operatorname{gph} T \text{ and} \\ (\xi^\nu,\psi^\nu) \to (\xi,\psi),\; \tau^\nu \searrow 0, \text{ having } (x^\nu + \tau^\nu\xi^\nu, v^\nu + \tau^\nu\psi^\nu) \in \operatorname{gph} T. \end{array} \tag{3.4}$$

This says geometrically that $\operatorname{gph} D_*T(\bar{x}\,|\,\bar{v})$ is the paratingent cone to $\operatorname{gph} T$ at $(\bar{x},\bar{v})$ [16, 9H]. That kind of cone is symmetric (equal to its own reflection), hence it consists of a union of lines rather than just a union of rays. It is a closed subset of $I\!R^n \times I\!R^n$ regardless of the closedness of $\operatorname{gph} T$.

In general, the mappings $H : I\!R^n \rightrightarrows I\!R^n$ such that $\operatorname{gph} H$ is a cone are the ones called *positively homogeneous* because

$$0 \in H(0) \text{ and } \tau\psi \in H(\tau\xi) \text{ for all } \tau > 0 \text{ when } \psi \in H(\xi). \tag{3.5}$$

Examples are $H = DT(\bar{x}\,|\,\bar{v})$ and $H = D_*T(\bar{x}\,|\,\bar{v})$, but for the latter there is more. By the symmetry of its graph,

$$\psi \in D_*T(\bar{x}\,|\,\bar{v})(\xi) \quad\Longrightarrow\quad \tau\psi \in D_*T(\bar{x}\,|\,\bar{v})(\tau\xi) \text{ for every } \tau \in (-\infty,\infty). \tag{3.6}$$

Strict graphical derivatives have a fundamental role in assessing possibilities for single-valued Lipschitz continuity. They have been seen recently in [1] to yield valuable information that way, even though, in general, they might be very hard to compute for specific mappings T. The key fact known about them is coming next. Its statement utilizes the concept of the outer norm

$$|H|^+ := \inf\{\lambda \geq 0 \,|\, |\psi| \leq \lambda|\xi| \text{ when } \psi \in H(\xi)\} \text{ for positively homogeneous H.} \tag{3.7}$$

Also involved is the concept of the *Lipschitz modulus* $\operatorname{lip} T(\bar{x}\,|\,\bar{v})$ of a localization of T at $(\bar{x},\bar{v}) \in \operatorname{gph} T$, this being the lower limit, as the neighborhood $\mathcal{X} \times \mathcal{V}$ shrinks down to just $(\bar{x},\bar{v})$ of the constants $\kappa > 0$ such that $|v'-v| \leq \kappa|x'-x|$ for (x,v) and (x',v') in $[\mathcal{X} \times \mathcal{V}] \cap \operatorname{gph} T$.

Proposition 3.1 (test for single-valued Lipschitz continuity, from [16, proof of 9.54])**.** *A necessary and sufficient condition for T to have a graphical localization at $\bar{x}$ for $\bar{v}$ that is single-valued and Lipschitz continuous relative to its domain is that*

$$D_*T(\bar{x}\,|\,\bar{v})(0) = \{0\}. \tag{3.8}$$

The Lipschitz modulus of that localization is then the outer norm $|D_*T(\bar{x}\,|\,\bar{v})|^+$, with this equaling the (attained) lowest λ such that $|\psi| \leq \lambda|\xi|$ when $\psi \in D_*T(\bar{x}\,|\,\bar{v})(\xi)$.

Likewise, a necessary and sufficient condition for T^{-1} to have a graphical localization at $\bar{v}$ for $\bar{x}$ that is single-valued and Lipschitz continuous relative to its domain is that

$$D_*T(\bar{x}\,|\,\bar{v})^{-1}(0) = \{0\}. \tag{3.9}$$

The Lipschitz modulus of that localization is then the outer norm $|D_*T(\bar{x}\,|\,\bar{v})^{-1}|^+$, with this equaling the (attained) lowest κ such that $|\xi| \leq \kappa|\psi|$ when $\psi \in D_*T(\bar{x}\,|\,\bar{v})(\xi)$.

Proof. This is really elementary, but we provide the full argument, extracting it from the proof of [16, 9.54], so that the result will easily be seen to carry over when we later pass to modified circumstances. The second part of the theorem mirrors the first because

$$\xi \in D_*[T^{-1}](\bar{v}\,|\,\bar{x})(\psi) \quad\Longleftrightarrow\quad \psi \in D_*T(\bar{x}\,|\,\bar{v})(\xi), \tag{3.10}$$

so we only need to deal with the first part. The claim of attainment of the value of the outer norm is new, not touched in [16, 9.54], so we quickly get that out of the way before dealing with the rest. By definition in (3.7), this value is the infimum of $\lambda \geq 0$ such that $|\psi| \leq \lambda|\xi|$ when $\psi \in D_*T(\bar{x}\,|\,\bar{v})(\xi)$, but because (3.8) precludes having $\xi = 0$ when $\psi \neq 0$, this is the infimum of λ such that $|\xi|^{-1}| \leq \lambda$ when $\psi \in D_*T(\bar{x}\,|\,\bar{v})(\xi)$ with $|\psi| = 1$. Because the graph of $D_*T(\bar{x}\,|\,\bar{v})$ is a closed set, the maximum of such $|\xi|^{-1}$ has an attained maximum value, $\bar{\lambda}$, which is then the attained minimum value of λ.

Necessity. Suppose there is a neighborhood $\mathcal{X} \times \mathcal{V}$ such that the localization $x \in \mathcal{X} \cap \operatorname{dom} T \mapsto \mathcal{V} \cap T(x)$ is a single-valued Lipschitz continuous function $t(x)$ with Lipschitz constant κ. Then in (3.4) we have for large enough ν that $v^\nu = t(x^\nu)$ and $v^\nu + \tau\psi^\nu = t(x^\nu + \tau^\nu\xi^\nu)$, so $|(v^\nu + \tau\psi^\nu) - v^\nu| \leq \kappa|(x^\nu + \tau^\nu\xi^\nu)x^\nu|$. Then $|\psi^\nu| \leq \kappa|\xi^\nu|$, and in the limit $|\psi| \leq \kappa|\xi|$. Therefore, having $\xi = 0$ entails having $\psi = 0$. The validity of the claim about the Lipschitz modulus is apparent from this context.

Suffiency. Because $D_*T(\bar{x}\,|\,\bar{v})$ has closed graph and is positively homogeneous in particular, according to (3.5), having $D_*T(\bar{x}\,|\,\bar{v})(0) = \{0\}$ ensures by [16, 9.23] the existence of $\bar{\kappa}$ such that $|\psi| \leq \bar{\kappa}|\xi|$ when $\psi \in D_*T(\bar{x}\,|\,\bar{v})(\xi)$. Then, for any choice of $\kappa > \bar{\kappa}$, there has to be a neighborhood $\mathcal{X} \times \mathcal{V}$ of $(\bar{x}, \bar{v})$ such that the sequences in (3.4), once they reach it, have $|\psi^\nu| \leq \kappa|\xi^\nu|$. Since any sequence $(x'^\nu, v'^\nu) \to (\bar{x}, \bar{v})$ can be expressed as $(x^\nu + \tau\xi^\nu, v^\nu + \tau\psi^\nu)$ with $\tau^\nu \searrow 0$ and the sequence of pairs (ξ^ν, ψ^ν) bounded (hence reducible to subsequences converging to pairs $(\xi, \psi) \in \operatorname{gph} D_*T(\bar{x}\,|\,\bar{v})$, we see that eventually $|v'^\nu - v^\nu| \leq \kappa|(x'^\nu - x^\nu|$. Then having $x'^\nu = x^\nu$ entails having $v'^\nu = v^\nu$, hence the localization of $\operatorname{gph} T$ to $\mathcal{X} \times \mathcal{V}$ is the graph of a single-valued, Lipschitz continuous mapping. □

The weakness of Proposition 3.1 for some purposes is that domain of the localization might not include a neighborhood of $\bar{x}$. That can be remedied with a coderivative or inner semicontinuity assumption as in [16, 9.54], or by taking T to be graphically Lipschitzian or merely crypto-continuous as in [1, Theorem 2.3]. But we'll have recourse below to other tactics and won't need this.

For our closed proper convex function f and its subdifferential mapping $\partial f : I\!R^n \rightrightarrows I\!R^n$, the strict graphical derivative $D_*(\partial f)(\bar{x}\,|\,\bar{v}) : I\!R^n \rightrightarrows I\!R^n$ itself has the shortcoming that its definition isn't tuned to the f-attentive convergence in the definition of general subgradients. Unless f is subdifferentially continuous, $D_*(\partial f)(\bar{x}\,|\,\bar{v})$ may reflect features of epi f that aren't localizable to its geometry at $(\bar{x}, f(\bar{x}))$. A substitute with finer tuning is required.

Definition 3.2 (strict second-order subdifferential)**.** *The mapping $\partial^2_* f(\bar{x}\,|\,\bar{v}) : I\!R^n \rightrightarrows I\!R^n$, called the strict second-order subdifferential of f at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$, is constructed by adding to the limit in (3.3) in the case of $T = \partial f$ in (3.1) the requirement that $f(x) \to f(\bar{x})$ and $f(x + \tau\xi) \to f(\bar{x})$. Thus,*

$$\begin{aligned} &(\xi, \psi) \in \operatorname{gph} \partial^2_* f(\bar{x}\,|\,\bar{v}) \iff \exists\, (x^\nu, v^\nu) \to (\bar{x}, \bar{v}) \text{ in } \operatorname{gph} T \text{ and} \\ &(\xi^\nu, \psi^\nu) \to (\xi, \psi),\ \tau^\nu \searrow 0, \text{ with } (x^\nu + \tau^\nu\xi^\nu, v^\nu + \tau^\nu\psi^\nu) \in \operatorname{gph} T, \\ &\text{and furthermore both } f(x^\nu) \to f(\bar{x}) \text{ and } f(x^\nu + \tau^\nu\xi^\nu) \to f(\bar{x}), \end{aligned} \tag{3.11}$$

The graph of $\partial^2_* f(\bar{x}\,|\,\bar{v})$ is again a symmetric cone, a union of lines through the origin of $I\!R^n \times I\!R^n$ instead merely rays. The rule in (3.5) therefore carries over,

$$\psi \in \partial^2_* f(\bar{x}\,|\,\bar{v})(\xi) \implies \tau\psi \in \partial^2_* f(\bar{x}\,|\,\bar{v})(\tau\xi) \text{ for every } \tau \in (-\infty, \infty). \tag{3.12}$$

Moreover, the graph of $\partial^2_* f(\bar{x}\,|\,\bar{v})$ is closed as a subset of $I\!R^n \times I\!R^n$, even though the graph of ∂f might not be.[5] And the facts in Proposition 3.1 likewise remain valid:

Proposition 3.3 (test of single-valued Lipschitz continuity for subgradient mappings)**.** *A necessary and sufficient condition for ∂f to have an f-attentive graphical localization at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ that is*

[5]The usual diagonalization argument for such closedness works just as well in the framework of (3.11).

single-valued and Lipschitz continuous relative to its domain is

$$\psi \in \partial_*^2 f(\bar{x}\,|\,\bar{v})(0) \quad\Longrightarrow\quad \psi = 0. \tag{3.13}$$

The Lipschitz modulus of that localization is then the outer norm $|\partial_*^2 f(\bar{x}\,|\,\bar{v})|^+$*, with this equaling the (attained) lowest* λ *such that* $|\psi| \le \lambda|\xi|$ *when* $\psi \in \partial_*^2 f(\bar{x}\,|\,\bar{v})(\xi)$*.*

Likewise, a necessary and sufficient condition for $(\partial f)^{-1}$ *to have a graphical localization at* $\bar{v}$ *for* $\bar{x}$ *that is single-valued and Lipschitz continuous relative to its domain is that*

$$0 \in \partial_*^2 f(\bar{x}\,|\,\bar{v})(\xi) \quad\Longrightarrow\quad \xi = 0. \tag{3.14}$$

The Lipschitz modulus of that localization is then the outer norm $|\partial_*^2 f(\bar{x}\,|\,\bar{v})^{-1}|^+$*, with this equaling the (attained) lowest* κ *such that* $|\xi| \le \kappa|\psi|$ *when* $\psi \in \partial_*^2 f(\bar{x}\,|\,\bar{v})(\xi)$*.*

Proof. The argument for Proposition 3.1 works just as well with f-attentive convergence. □

The last line of (3.11) provides the contrast between $\partial_*^2 f(\bar{x}\,|\,\bar{v})$ and $D_*(\partial f)(\bar{x}\,|\,\bar{v})$. In modifying strict graphical derivatives in this way we follow in the footsteps of Gfrerer [2] in modifying the Mordukhovich second-order subdifferential

$$\partial^2 f(\bar{x}\,|\,\bar{v}) := D^*(\partial f)(\bar{x}\,|\,\bar{v}), \tag{3.15}$$

cf. [6], as defined directly from the coderivative mapping $D^*(\partial f)(\bar{x}\,|\,\bar{v})$. That coderivative mapping is itself defined in terms of limits of normal cones to $\operatorname{gph}\partial f$ at elements (x, v) that approach $(\bar{x}, \bar{v})$, but Gfrerer adds the requirement that also $f(x) \to f(\bar{x})$ to replace (3.12) by, in his notation, $D_f^*(\partial f)(\bar{x}\,|\,\bar{v})$. For prox-regular f, he was able to characterize monotonicity properties of ∂f to "semidefiniteness" properties of $D_f^*(\partial f)(\bar{x}\,|\,\bar{v})$. We will proceed toward that with our $\partial_*^2 f(\bar{x}\,|\,\bar{v})$ as well, and the results of the two approaches will then be compared.

An explanation of what we mean by "semidefiniteness" properties is needed as a start. For any positively homogeneous mapping $H : I\!R^n \rightrightarrows I\!R^n$ and any $s \in (-\infty, \infty)$, we say that H is

$$\begin{array}{l} s\text{-}\textit{semidefinite} \text{ (positive-semidefinite when } s = 0) \text{ if } \ \psi{\cdot}\xi \ge s|\xi|^2 \text{ when } \psi \in H(\xi), \\ s\text{-}\textit{definite} \text{ (positive-definite when } s = 0) \text{ if } \ \psi{\cdot}\xi > s|\xi|^2 \text{ when } \psi \in H(\xi),\ \xi \ne 0. \end{array} \tag{3.16}$$

We introduce the *modulus of definiteness* of H to be

$$\operatorname{def} H := \sup\{\, s \mid H \text{ is } s\text{-semidefinite} \,\} = \inf\{\, \psi{\cdot}\xi \mid \psi \in H(\xi),\ |\xi| = 1 \} \tag{3.17}$$

(where by convention the sup is $-\infty$ if no such s exists and the inf is ∞ of no such ξ exists).

Obviously, in the case of a linear mapping, $H(x) = Ax$, s-definiteness for $s = 0$ is monotonicity of A, and for $s > 0$ is strong monotonicity of A, with the first corresponding to having $\operatorname{def} H \ge 0$ and the second to surely having $\operatorname{def} H > 0$. Hypo-monotonicity of H allows for negative $\operatorname{def} H$. It's easy to see that

$$\begin{array}{l} T \text{ locally s-monotone at } \bar{x} \text{ for } \bar{v} \in T(\bar{x}) \quad\Longrightarrow\quad \operatorname{def} T(\bar{x}\,|\,\bar{v}) \ge s, \\ \partial f\ f\text{-locally s-monotone at } \bar{x} \text{ for } \bar{v} \in \partial f(\bar{x}) \quad\Longrightarrow\quad \operatorname{def} \partial_*^2 f(\bar{x}\,|\,\bar{v}) \ge s, \end{array} \tag{3.18}$$

but what about the opposite direction. If $\operatorname{def} T(\bar{x}\,|\,\bar{v}) > s$, say, is T sure to be locally s-monotone there? That would be beautiful, but it's false.[6]

[6]The mapping in the following counterexample will also serve us in Counterexample 3.9, but the falsity can likewise be seen from the mapping utilized earlier in [7, Example 3.5] for the analogous conjecture for $D^*T(\bar{x}\,|\,\bar{y})$.

Counterexample 3.4 (failure of monotonicity test for set-valued mappings in general)**.** *A mapping* $T : I\!R^n \rightrightarrows I\!R^n$ *with* $\bar v \in T(\bar x)$ *can have* $\operatorname{def} D_*T(\bar x\,|\,\bar v) > 0$ *and yet not be strongly monotone or even hypomonotone in any localization there.*

Details. Define $T : I\!R^2 \rightrightarrows I\!R^2$ by $T(x_1, x_2) = (x_1, -x_2^{1/3})$ and take $\bar x = (0,0)$, $\bar v = (0,0)$. Direct calculation indicates that

$$\begin{aligned} &D_*T(\bar x\,|\,\bar v)(\xi_1,\xi_2) = \emptyset \ \text{ unless } \ \xi_2 = 0,\\ &(\psi_1,\psi_2) \in D_*T(\bar x\,|\,\bar v)(\xi_1, 0) \quad\Longleftrightarrow\quad \psi_1 = \xi_1 \ \ (\psi_2 \text{ anything}). \end{aligned} \tag{3.19}$$

Then $\operatorname{def} D_*T(\bar x\,|\,\bar v) = 1$, yet T doesn't have an s-monotone localization at $\bar x$ for $\bar v$ for any $s < 1$. □

Nonetheless, despite this counterexample to the conjectured reversal of the first implication in (3.18), that reversal turns out to be valid for the second implication in (3.18). In preparation for the proof of that, we confirm in precise *local* form a fact about convex functions that has long been known in a *global* form, namely that g is strongly convex if and only if g^* is differentiable with ∇g^* Lipschitz continuous, cf. [16, 12.60]. For a localization of this relationship, we appropriately replace global strong convexity of g by variational strong convexity of g at $\bar x$ for $\bar v$.

Proposition 3.5 (local dualization of strong convexity)**.** *For a closed proper convex function* g *on* $I\!R^n$ *and* $\bar v \in \partial g(\bar x)$*, variational strong convexity holds at* $\bar x$ *for* $\bar v$ *if and only if* g^* *is differentiable on a neighborhood of* $\bar v$ *with its gradient mapping* ∇g^* *Lipschitz continuous there. Moreover in that case*

$$\operatorname{lip} \nabla g^*(\bar v) = 1/\operatorname{cnv} g(\bar x\,|\,\bar v). \tag{3.20}$$

Proof. Variational strong convexity of g at $\bar x$ for $\bar v \in \partial g(\bar x)$ is equivalent by Theorem 1.1 to strong monotonicity of a localization of ∂g to a neighborhood $\mathcal{X}\times\mathcal{V}$ of $(\bar x, \bar v)$, which can be taken to be convex and open. Such monotonicity as a level $s > 0$ implies the single-valuedness and Lipschitz continuity of the inverse mapping $(\partial g)^{-1}$ in localization to $\mathcal{V}\times\mathcal{X}$ with Lipschitz constant s^{-1}: for (x, v) and (x', v') in the localization,

$$\begin{aligned} (x'-x)\cdot(v'-v) \geq s|x'-x|^2 \quad&\Longrightarrow\quad |x'-x|^2 \leq (1/s)|x'-x|\cdot|v'-v|\\ &\Longrightarrow\quad |x'-x| \leq s^{-1}|v'-v|. \end{aligned}$$

Since $(\partial g)^{-1} = \partial g^*$ with $\bar x \in \partial g^*(\bar x)$ in particular, this property of $(\partial g)^{-1}$ is equivalent to g^* being differentiable on $\mathcal{V}$ with that Lipschitz constant for ∇g^*. With v' expressed as $v+\psi$, we have then from $g^*(v+\psi) - g^*(v) = \int_0^1 \nabla g^*(v+t\psi)\cdot\psi dt$ with $|\nabla g^*(v+t\psi) - \nabla g^*(v)| \leq s^{-1}t|\psi|$ that

$$g^*(v+\psi) - g^*(v) - \nabla g^*(v)\cdot\psi \ \leq\ s^{-1}j(\psi) \ \text{ when } v \in \mathcal{V},\ v+\psi \in \mathcal{V}.$$

Denote the left side of this by $\gamma_{x,v}(\psi)$ and choose $\lambda > 0$ small enough that $\bar v + 2\lambda I\!B \subset \mathcal{V}$, so that

$$\gamma_{x,v}(\psi) \leq k(\psi) \ \text{ for } \ k(\psi) := s^{-1}j(\psi) + \delta_{\lambda B}(\psi) \ \text{ when } x \in \mathcal{X},\ |v-\bar v| \leq \lambda.$$

Taking conjugates, we get

$$\gamma^*_{x,v}(\xi) \ \geq\ k^*(\xi) \ \text{ for all } \xi \text{ when } x \in \mathcal{X},\ |v-\bar v| \leq \lambda, \tag{3.21}$$

where $\gamma^*_{x,v}(\xi) = g(x+\xi) - g(x) - v\cdot\xi$ for $x = \nabla g^*(v)$ and k^* is given by infimal convolution of the conjugates of $s^{-1}j$ and $\delta_{\lambda B}$. Those being sj and $\lambda|\cdot|$, that comes out as $k^*(\xi) = sj(\xi)$ when $|\xi| \leq \lambda s^{-1}$ and $k^*(\xi) = \lambda|\xi| - \frac{1}{2}\lambda^2 s^{-1}$ when $|\xi| \geq \lambda s^{-1}$. Hence from (3.21),

$$g(x+\xi) - g(x) - v\cdot\xi \ \geq\ sj(\xi) \ \text{ for } \ \xi \in \lambda s^{-1} I\!B \ \text{ when } x \in \mathcal{X},\ |v-\bar v| \leq \lambda,$$

which by definition indicates prox-regularity of g at $\bar x$ for $\bar v$ at level s. That's equivalent to variational s-convexity of g by Theorem 1.1, so we have come full circle with the same $s > 0$, which been of any magnitude such that $s < \operatorname{cnv} g(\bar x\,|\,\bar v)$ or $s^{-1} > \operatorname{lip} \nabla g^*(\bar v)$. This verifies (3.20). □

Theorem 3.6 (modulus formula from strict second-order subdifferential)**.** *When the subgradient* $\bar{v} \in \partial f(\bar{x})$ *is a proximal subgradient,*

$$\operatorname{cnv} f(\bar{x} \,|\, \bar{v}) = \operatorname{def} \partial_*^2 f(\bar{x} \,|\, \bar{v}). \tag{3.22}$$

Proof. From the validity of the second implication in (3.18) we already know by way of the formula for $\operatorname{cnv} f(\bar{x} \,|\, \bar{v})$ in (1.13), expressing it as the modulus of f-local monotonicity of ∂f, that if $\operatorname{cnv} f(\bar{x} \,|\, \bar{v}) \geq s$, then $\operatorname{def} \partial_*^2 f(\bar{x} \,|\, \bar{v}) \geq s$. All that remains is showing that

$$\operatorname{def} \partial_*^2 f(\bar{x} \,|\, \bar{v}) > s \quad \Longrightarrow \quad \operatorname{cnv} f(\bar{x} \,|\, \bar{v}) \geq s. \tag{3.23}$$

We can harmlessly normalize by supposing that $(\bar{x}, \bar{v}) = (0, 0)$ and $f(0) = 0$. Then in passing from f to $f + rj$ for any r both inequalities would simply be shifted by adding r, so by confirming (3.23) any convenient s, assuming for instance that $s > 0$, we'll be able to conclude that it holds for all s.

The same principle of shifts leads to further simplification based on our assumption that the subgradient we're dealing with is proximal. That gives us the existence of $r > 0$ such that $f(x) \geq -r|x|^2$ on a neighborhood of 0. We are only concerned with local properties connected with these circumstances, so there's no loss of generality is taking $\operatorname{dom} f$ to be bounded, inasmuch as that can be arranged by adding to f the indicator of a closed neighborhood of the origin. This allows us to pass, through an increase in r if needed to having $f(x) > -r|x|^2$ for all $x \neq 0$. Since a shift from f to $f + rj$ is at our disposal, we see we can just as well normalize to supposing

$$\operatorname{argmin} \bar{f} = \{0\} \;\text{ and }\; \min \bar{f} = 0. \tag{3.24}$$

Our task in these circumstances comes down to proving that

$$\operatorname{def} D_*[\partial f](0 \,|\, 0) > s > 0 \quad \Longrightarrow \quad \operatorname{cnv} f(0 \,|\, 0) \geq s, \tag{3.25}$$

where on the left we have

$$\psi{\cdot}\xi > s|\xi|^2 \;\text{ when }\; \psi \in \partial_*^2 f(0 \,|\, 0)(\xi),\; \xi \neq 0. \tag{3.26}$$

In particular, $0 \in \partial_*^2 \bar{f}(0 \,|\, 0)(\xi)$ only for $\xi = 0$. We know then from Proposition 3.3 that $\partial \bar{f}^{-1}$ has a graphical localization to a convex neighborhood $\mathcal{V} \times \mathcal{X}$ of $(0, 0)$ that's single-valued and locally Lipschitz continuous on $\mathcal{V}$, the Lipschitz modulus being the lowest κ such that $|\xi| \leq \kappa|\psi|$ when $\psi \in \partial_*^2 \bar{f}(0 \,|\, 0)(\xi)$. As seen from (3.26), s^{-1} is greater than that modulus and thus can serve as a local Lipschitz constant.

Let $g = f^{**}$. Because of $\operatorname{dom} f$ being bounded, g is closed proper convex. We have $g^* = f^*$ with $-g^*(v) = \min \{ f(x) - v{\cdot}x \}$, the minimum being surely attained, that being uniquely at $x = 0$ when $v = 0$ by (3.24). In general, its attainment at x, which corresponds to having $x \in \partial g^*(v)$ and $f(x) - v{\cdot}x = -g^*(v)$, requires $v \in \partial f(x)$, the same as $x \in \partial f^{-1}$. Also, though, having $x \in \partial g^*(v)$ entails $x{\cdot}v - g^*(v) = -g^{**}(x) = g(x)$. Thus,

$$x \in \partial g^*(v) \;\text{ with }\; (v, x) \in \mathcal{V} \times \mathcal{X} \quad \Longrightarrow \quad x \in \partial f^{-1}(v) \;\text{ and }\; f(x) = g(x).$$

From what we know about ∂f^{-1}, we see that g^* must be differentiable on $\mathcal{V}$ with ∇g^* Lipschitz continuous there with constant $\bar{s}^{-1}$. Proposition 3.5 says that in this case g is variationally s-convex with respect to localizing ∂g to $\mathcal{X} \times \mathcal{V}$, moreover with g coinciding with f in that localization. Then, by definition, f is variationally s-convex at 0 for $0 \in \partial f(0)$. That confirms (3.25). ◻

Counterexample 3.7 (showing the need for a proximal subgradient in Theorem 3.6)**.** *The formula in Theorem 3.6 can fail when* $\bar{v}$ *is a regular subgradient but not a proximal subgradient. Moreover, it can fail with* $\operatorname{def} f(\bar{x}\,|\,\bar{v}) > 0$ *and* $\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) = -\infty$*, indicating even a lack of prox-regularity.*

Details. The mapping T in Counterexample 3.4 is in fact the gradient mapping for $f(x_1, x_2) = \frac{1}{2}x_1^2 - \frac{3}{4}x^{4/3}$, which is continuously differentiable. The subgradients of f are its gradients and are all regular, most actually proximal, but not the ones where $x_2 = 0$. In particular, in taking $\bar{x} = (0, 0)$ and $\bar{v} = \nabla f(\bar{x}) = (0, 0)$, we have $\bar{v}$ regular but not proximal. The continuous differentiability of f encompasses subdifferential continity, so we have $D_*T(\bar{x}\,|\,\bar{v}) = \partial_*^2 f(\bar{x}\,|\,\bar{v})$ and $\operatorname{def} \partial_*^2 f(\bar{x}\,|\,\bar{v}) = 1$. Thus, for any $s \in (0, 1)$ we have $\psi{\cdot}\xi \geq s|\xi|^2$ when $\psi \in \partial_*^2 f(\bar{x}\,|\,\bar{v})(\xi)$. This might be expected to guarantee variational strong convexity, but f isn't even variationally hypoconvex at this location. □

The modulus formula in Theorem 3.6 using $\partial_*^2 f(\bar{x}\,|\,\bar{v})$ can be compared, in the notation we've introduced here, to a formula recently proved by Gfrerer in terms of his f-attentive modification $D_f^*(\partial f)(\bar{x}\,|\,\bar{v})$ of the Mordukhovich second-order subdifferential in (3.15).

Theorem 3.8 (coderivative-based modulus formula from Gfrerer [2, Theorem 5.1])**.** *As long as* f *is prox-regular at* $\bar{x}$ *for* $\bar{v} \in \partial f(\bar{x})$*,*

$$\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) = \operatorname{def} D_f^*(\partial f)(\bar{x}\,|\,\bar{v}). \tag{3.27}$$

Here it's not just the subgradient $\bar{v}$ that's assumed to be proximal, but all subgradients nearby, and the proximality has to be uniform in a sense. But is that stronger assumption truly needed? Yes. as shown by an example contributed to us by M. Benko.[7]

Counterexample 3.9 (showing why prox-regularity needs to be assumed in Theorem 3.8; Benko)**.** *The formula in Theorem 3.8 can fail when prox-regularity is weakened to the proximal subgradient assumption in Theorem 3.6. It can even fail with* $\operatorname{def} D_f^*(\partial f(\bar{x}\,|\,\bar{v}) > 0$ *and* $\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) = -\infty$.

Details. For $(x_1, x_2) \in I\!R^2$, let A_1 and A_2 denote the x_1-axis and x_2 axis, and let $A = A_1 \cup A_2$. It will be demonstrated for $\bar{x} = (0, 0)$ and $\bar{v} = (0, 0)$ that δ_A is subdifferentially continuous there with $\bar{v}$ a proximal subgradient, but not prox-regular although

$$\xi \in \partial^2 \delta_A(\bar{x}\,|\,\bar{v})(\psi) \quad \Longrightarrow \quad \psi{\cdot}\xi = 0. \tag{3.28}$$

Then for $f = \delta_A + j$ we will have these subgradient properties with $\partial^2 f(\bar{x}\,|\,\bar{v})$ positive-definite.

The subdifferential continuity of δ_A is assured its values on its domain being constantly 0. The graph of $\partial\delta_A$ is then simply the closure of the graph of $\widehat{\partial}\delta_A$, which is the regular normal cone mapping $\widehat{N}_A$. Obviously, $\widehat{N}_A(0, 0) = \{(0, 0)\}$, while for $x \neq (0, 0)$ we have $\widehat{N}_A(x) = A_2$ when $x \in A_1$ and $\widehat{N}_A(x) = A_1$ when $x \in A_2$. Hence $\operatorname{gph} \partial\delta_A = S \cup S^\perp$ for $S = (I\!R, 0; 0, I\!R)$ and $S^\perp = (0, I\!R; I\!R, 0)$, this graphical configuration being incompatible with prox-regularity at the origin. The pairs (ψ, ξ) in (3.28) are, by the definition of the coderivative in (3.14), the ones such that $(\xi, -\psi) \in N_{S \times S^\perp}(\bar{x}, \bar{v})$. But that normal cone is just $S \times S^\perp$ again, so the implication in (3.28) is correct. □

We finish the section with an observation that adds to Theorem 3.6.

Proposition 3.10 (formula enhancement under prox-regularity)**.** *When* f *is prox-regular at* $\bar{x}$ *for* $\bar{v} \in \partial f(\bar{x})$*, the function*

$$h(\xi) := \inf \{\, \psi{\cdot}\xi \mid \psi \in \partial_*^2 f(\bar{x}\,|\,\bar{v})(\xi) \,\} \tag{3.29}$$

[7] Private communication.

is not only proper and positively homogeneous of degree 2, but also lower semicontinuous with the infimum in its definition being attained when finite. In these terms, then,

$$\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) = \min\{\, h(\xi) \mid |\xi| = 1 \,\}. \tag{3.30}$$

Proof. The assumption of prox-regularity, in corresponding to $\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) > -\infty$, ensures that h is proper. The positive homogeneity of degree 2 for h follows from the positive homogeneity of degree 1 for the mapping $\partial^2_* f(\bar{x}\,|\,\bar{v})$, and through properness it entails $h(0) = 0$. Prox-regularity at level r corresponds in fact to $h \geq -2rj$. Verifying the lower semicontinuity of h comes down, through the positive homogeneity, to verifying it relative to the unit sphere $|\xi| = 1$, and that will also confirm the claim of attainment in (3.29) and the re-expression of (3.22) as (3.30).

Recall that, in replacing f by $f + sj$, $\partial^2_* f(\bar{x}\,|\,\bar{v})$ would be replaced by $\partial^2_* f(\bar{x}\,|\,\bar{v}) + sI$ and h would be replaced by $h + sj$. This kind of shift has no effect on the presence or absence of lower semicontinuity, so we can take advantage of it, starting from having $h \geq 2rj$, to reduce to having $h \geq tj$ for some $t > 0$. This yields $\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) > 0$ and the variational strong convexity of f at $\bar{x}$ for $\bar{v}$. Through the definition of that, we can just as well suppose that f is itself convex and fits the framework of localized strong convexity in Proposition 3.5, with ∂f^* being single-valued and locally Lipschitz continous around $\bar{v}$. Proposition 3.3 informs us that then (3.13) holds, and this will soon be important.

The lower semicontinuity to be verified corresponds to the closedness of the level sets $C_\alpha = \{\, \xi \mid h(\xi) \leq \alpha,\, |\xi| = 1\}$. Suppose $\xi^\nu \in C_\alpha$ and $\xi^\nu \to \bar{\xi}$. Then $|\bar{\xi}| = 1$, and from the definition of h in (3.29) we have the existence of $\psi^\nu \in \partial^2_* f(\bar{x}\,|\,\bar{v})(\xi^\nu)$ and $\alpha^\nu \searrow \alpha$ such that $\psi^\nu \cdot \xi^\nu \leq \alpha^\nu$. If the sequence of vectors ψ^ν is bounded, we can pass to a subsequence converging to some $\bar{\psi}$, which must belong to $\partial^2_* f(\bar{x}\,|\,\bar{v})(\bar{\xi})$, because the graph of $\partial^2_* f(\bar{x}\,|\,\bar{v})$ is closed. Then in the limit also $\bar{\psi} \cdot \bar{\xi} \leq \alpha$, so $\bar{\xi} \in C_\alpha$ and we are done. All that remains is excluding the possibility that the sequence of vectors ψ^ν is unbounded. If that were true, we could reduce to having $\psi^\nu \neq 0$ with $\lambda^\nu = 1/|\psi^\nu| \searrow 0$ and $\lambda^\nu \psi^\nu$ converging to some $\psi \neq 0$. In that case from $\psi^\nu \in \partial^2_* f(\bar{x}\,|\,\bar{v})(\xi^\nu)$ we also have $\lambda^\nu \psi^\nu \in \partial^2_* f(\bar{x}\,|\,\bar{v})(\lambda^\nu \xi^\nu)$ and therefore $\psi \in \partial^2_* f(\bar{x}\,|\,\bar{v})(0)$. But that contradicts our knowledge that (3.13) holds. □

4 Tests utilizing generalized Hessian constructs

A classical theorem of Alexandrov asserts that a finite convex function on an open convex set has a quadratic "expansion" almost everywhere in the sense of its second-order difference quotients converging uniformly on bounded sets to something quadratic. Back in 1985, this was extended in [10] by generalizing "quadratic" and replacing the uniform pointwise convergence by epiconvergence, which reduces to it in the finiteness context. More recently in [13], "quadratic bundles" were created by taking limits and employed in characterizing strong convexity of augmented Lagrangians by an appeal to positive-definiteness. That line of theory, essentially primal-dual local in relying on properties of convex functions based on localizations of the graphs of their subdifferential mappings, immediately carries over to variationally convex functions, even to variationally s-convex functions for any s. But by Theorem 1.1, that means it's available any time prox-regularity is available — which has not previously been pointed out and guides our thinking here.

Generalized second-order differentiability will take some explaining, but with that out of the way, a definitive connection with the modulus $\operatorname{cnv} f(\bar{x}\,|\,\bar{v})$ will easily be made. A formula will emerge that's essentially a repackaging of a formula already offered by Gfrerer in [2], but seen this way as coming from second-order convex analysis instead of his broader theory with Outrata [3] of subspaces associated with graphical derivatives and coderivatives, with its algorithmic inspirations.

The subsets of $I\!R^n \times I\!R^m$ are, by definition, the graphs of the generally set-valued mappings from $I\!R^n$ to $I\!R^m$. When its graph is a subspace, a mapping is *generalized linear* in the terminology of [16, p. 329]. The adjoint of a generalized linear mapping $L : I\!R^n \rightrightarrows I\!R^m$ is the generalized linear mapping $L^* : I\!R^m \rightrightarrows I\!R^n$ given by

$$\operatorname{gph} L^* = \{\, (\eta,\zeta) \,|\, \zeta{\cdot}\xi = \eta{\cdot}\psi \text{ for all } (\xi,\psi) \in \operatorname{gph} L\} = \{\, (\eta,\zeta) \,|\, (\zeta,-\eta) \in (\operatorname{gph} L)^\perp\}, \tag{4.1}$$

which yields $L^{**} = L$ and works out to the right thing when L is a linear mapping in the original sense. The inverse of a generalized linear mapping is another such, and so forth. On the other hand, a function $q : I\!R^n \to \overline{I\!R}$ is a *generalized (purely) quadratic form* when

$$q(\xi) = \left\{ \begin{array}{ll} \frac{1}{2}\xi{\cdot}Q\xi & \text{if } \xi \in S \\ \infty & \text{if } \xi \notin S \end{array} \right\} \text{ for a subspace } S \text{ and linear self-adjoint } Q : S \to S. \tag{4.2}$$

The generalized linear mappings $L : I\!R^n \rightrightarrows I\!R^n$ that are self-adjoint, $L^* = L$, are the subdifferentials of such functions q, with

$$L(\xi) = Q\xi + S^\perp \text{ when } \xi \in S, \text{ but otherwise } L(\xi) = \emptyset, \tag{4.3}$$

in relation to (4.2), and consequently

$$\operatorname{dom} q = \operatorname{dom} L \text{ and } q(\xi) = \tfrac{1}{2}\xi{\cdot}L(\xi) := \tfrac{1}{2}\xi{\cdot}\psi \text{ for any } \psi \in L(\xi). \tag{4.4}$$

Because generalized linear mappings L are positively homogeneous, the "definiteness" terminology in Section 3 applies to them and carries over through (4.4) to generalized quadratic forms q:

$$\begin{array}{lll} L \text{ is } s\text{-semidefinite ("positive-semidefinite" when } s=0) & \Longleftrightarrow & q(\xi) \geq \frac{s}{2}|\xi|^2, \\ L \text{ is } s\text{-definite ("positive-definite" when } s=0) & \Longleftrightarrow & q(\xi) > \frac{s}{2}|\xi|^2 \text{ for } \xi \neq 0, \end{array} \tag{4.5}$$

or from another angle,[8]

$$\operatorname{def} L = \min\{\, 2q(\xi) \,|\, |\xi| = 1\} = \text{ lowest eigenvalue of } Q \text{ on } S \text{ in (4.2)–(4.3)}. \tag{4.6}$$

For instance we can speak this way of q as being positive-definite or positive-semidefinite. Note that s-semidefiniteness of L is the same as s-monotonicity of L and therefore corresponds in (4.5) to $2q$ being s-convex — because having $(\xi_1-\xi_0){\cdot}(\psi_1-\psi_0) \geq s|\xi_1-\xi_0|^2$ when $(\xi_i,\psi_i) \in \operatorname{gph} L$ is the same as having $\xi{\cdot}\psi \geq s|\xi|^2$ when $(\xi,\psi) \in \operatorname{gph} L$, due to that graph being a subspace. Furthermore,

$$\begin{array}{l} \text{for convex } q \text{ having } \partial q = L, \text{ the conjugate function} \\ q^* \text{ is the generalized quadratic form with } \partial q^* = L^{-1}. \end{array} \tag{4.7}$$

The early developments in [10] on second-order convex analysis depended on these extensions of "linear" and "quadratic" and their interrelationship in articulating a sort of neoclassical approach to differentiability. Both of the following concepts were essential in that, although only the second was given the name then that we now attach to it.

[8] Here $\operatorname{def} L := \infty$ when $\operatorname{dom} L = \{0\}$, which corresponds to $q =$ indicator of $\{0\}$.

Definition 4.1 (generalized differentiability, first-order and second-order)**.**

(a) $T : I\!R^n \rightrightarrows I\!R^m$ *is generalized differentiable at* $\bar{x}$ *for* $\bar{v} \in T(\bar{x})$ *if the graphical derivative mapping* $DT(\bar{x}\,|\,\bar{v})$ *is generalized linear and the difference quotient mappings* $\Delta_\tau T(\bar{x}\,|\,\bar{v})$ *in (3.1) converge to it in graph as* $\tau \searrow 0$*, instead of only having it as the outer limit in (3.2).*

(b) f *is generalized twice differentiable at* $\bar{x}$ *for* $\bar{v} \in \partial f(\bar{x})$ *if the second-order subderivative* $d^2 f(\bar{x}\,|\,\bar{v})$ *is a generalized quadratic form and the second-order difference quotient functions* $\Delta_\tau^2 f(\bar{x}\,|\,\bar{v})$ *in (2.1) epi-converge to it, instead of only having it as the lower epi-limit in (2.2).*

The property in (a) means that the tangent cone to $\operatorname{gph} T$ at $(\bar{x}, \bar{v})$ is a subspace, and $\operatorname{gph} T$ is "derivable" there [16, p. 198]. In [10], where T was ∂f with f convex, this was called the *smoothness* of $\operatorname{gph} T$ at $(\bar{x}, \bar{v})$. In that special case, of course, the graph is a Lipschitzian manifold of dimension n through its Minty parameterization [16, 12.15] as the graph of a maximal monotone mapping, so by Rademacher's theorem about the almost everywhere differentiability of Lipschitz continuous mappings [16, 9.60], smoothness is present almost everwhere (with respect to the Minty parameterization). But through Theorem 1.1, this holds locally not just for convex functions, but whenever prox-regularity is at hand. The neoclassical picture in [10] is that way function grandly enlarged.

Theorem 4.2 (generalized twice differentiability under prox-regularity)**.** *Let* f *be prox-regular at* $\bar{x}$ *for* $\bar{v} \in \partial f(\bar{x})$*, and let* $\mathcal{X}_f^\alpha \times \mathcal{V}$ *be an* f*-attentive neighborhood of* $(\bar{x}, \bar{v})$ *for localizing* ∂f *as in Theorem 1.1(e), with* $\mathcal{X} \times \mathcal{V}$ *taken to be open.*

(a) *For* $(x, v) \in \mathcal{X}_f^\alpha \times \mathcal{V}$*,* ∂f *is generalized differentiable (in the localization) if and only if* f *is generalized twice differentiable. If so, then* $L = \partial q$ *for* $L = D[\partial f](\bar{x}\,|\,\bar{v})$ *and* $q = \frac{1}{2} d^2 f(\bar{x}\,|\,\bar{v})$*.*

(b) *Furthermore, these properties are sure to hold for almost every* $(x, v) \in (\mathcal{X}_f^\alpha \times \mathcal{V}) \cap \operatorname{gph} \partial f$*, if not necessarily for* $(\bar{x}, \bar{v})$ *itself.*

Proof. This just translates results in [10, Section 4] and [8, Theorem 6.1] ([16, 13.40]) in the obvious way, using shifts between f and $f + sj$, to the broader setting here. □

Theorem 4.2 reinterprets and makes applicable to all prox-regular functions, despite their ∞ values and discontinuities, the theorem of Alexandrov about a finite convex function having almost everywhere a second-order expansion. That theorem is the corollary based on the fact that when finite convex functions epiconverge to a finite convex function, the convergence is uniform on bounded sets [16, 7.17] — as applied to second-order difference quotients. There is no longer a Hessian *matrix* to point to in the extended framework, but we can refer just as well to L in Theorem 4.2(a) as the generalized *Hessian mapping* at x for $v \in \partial f(x)$.

The next step down the path followed in [10] is taking advantage of having generalized differentiability available almost everywhere to generate "information repositories" at points where there isn't such generalized differentiability. This follows Clarke in his long-ago introduction of a generalized Jacobian for a Lipschitz continuous mapping. However, unlike him, we don't end up with convex hulls, and the limits we take are graphical and epigraphical.

Proposition 4.3 (compactness properties)**.**

(a) *The collection of all* s*-semidefinite generalized quadratic forms* q *on* $I\!R^n$ *is compact in the topology of epi-convergence.*

(b) *The collection of all self-adjoint* s*-semidefinite generalized linear mappings* $L : I\!R^n \rightrightarrows I\!R^n$ *is compact in the topology of graphical convergence.*

(c) *The passage from* q *to* $L = \partial q$ *is continuous with respect to these topologies, and the passage from* L *back to* q *is continuous that way as well.*

Proof. Reduction can be made to the case of $s = 0$. Then (a) and (b) follow from the compactness properties of set-convergence in [16, 4E] as applied to epigraphs and graphs. On the other hand, (c) specializes Attouch's theorem about the relationship between epi-convergence of sequences of convex functions and graphical convergence of their subdifferential mappings [16, 12.35]. □

Definition 4.4 (generalized quadratic and Hessian bundles)**.** *For f prox-regular at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$, the generalized quadratic bundle there is*

$$\operatorname{quad} f(\bar{x}\,|\,\bar{v}) = \begin{cases} \text{the collection of all generalized quadratic forms obtainable} \\ \text{as epi-limits of such forms } q^\nu = \tfrac{1}{2}d^2 f(x^\nu\,|\,v^\nu) \text{ in generalized} \\ \text{twice differentiability of } f \text{ as } (x^\nu, v^\nu) \to (\bar{x}, \bar{v}),\ f(x^\nu) \to f(\bar{x}). \end{cases} \tag{4.8}$$

The generalized Hessian bundle there is

$$\nabla^2_* f(\bar{x}\,|\,\bar{v}) = \{\, L = \partial q \mid q \in \operatorname{quad} f(\bar{x}\,|\,\bar{v})\}, \tag{4.9}$$

thus consisting, from the perspective of Proposition 4.3(c), of all associated graphical limits of generalized Hessian mappings L^ν at nearby (x^ν, v^ν).

The generalized quadratic bundle is nonempty and compact with respect to epi-convergence, and the generalized Hessian bundle is nonempty and compact with respect to graphical convergence. The purpose of introducing $\operatorname{quad} f(\bar{x}\,|\,\bar{v})$ in [13, Section 4] was to test for local strong convexity, and testing for variational strong convexity is now a continuation of that. The generalized Hessian bundle $\nabla^2_* f(\bar{x}\,|\,\bar{v})$ was implicit in [13], but here we bring it forth explicitly.

With this framework in place, we can state a result that in essense is a partial reformulation in such terms of one of Gfrerer, as will be discussed afterward.

Theorem 4.5 (modulus via generalized Hessian bundles, reformulating Gfrerer [2, Theorem 4.3])**.** *If f is prox-regular at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$, then*

$$\operatorname{cnv} f(\bar{x}\,|\,\bar{v}) = \min\{\operatorname{def} L \mid L \in \nabla^2_* f(\bar{x}\,|\,\bar{v})\} =: \operatorname{def} \nabla^2_* f(\bar{x}\,|\,\bar{v}). \tag{4.10}$$

Equivalently in quadratic bundle terms,

$$\begin{aligned} \tfrac{1}{2} \operatorname{cnv} f(\bar{x}\,|\,\bar{v}) &= \min\{\ \operatorname{def} q \mid q \in \operatorname{quad} f(\bar{x}\,|\,\bar{v})\,\} \\ &= \min_{q \in \operatorname{quad} f(\bar{x}\,|\,\bar{v})} \left\{ \min_{|\xi|=1} q(\xi) \right\} \ = \ \min_{|\xi|=1} \left\{ \min_{q \in \operatorname{quad} f(\bar{x}\,|\,\bar{v})} q(\xi) \right\}. \end{aligned} \tag{4.11}$$

Proof. Since the claimed formula is unaffected by adding a term sj to f, we can restrict to f being variationally s-convex for $s > 0$, so that the function $\widehat{f}$ in the definition of that property is variationally strongly convex at $\bar{x}$ for $\bar{v}$. But then, through the agreements between f and $\widehat{f}$ in the definition, we might as well suppose that f itself is variationally strongly convex at $\bar{x}$ for $\bar{v}$. This puts us in the setting of Proposition 3.5 with $g = f$ and $\operatorname{cnv} f(\bar{x}\,|\,\bar{v})$ being reciprocal to $\operatorname{lip} \nabla f^*(\bar{v})$ in connection with ∂f^* reducing locally to a single-valued Lipschitz continuous mapping $F = \nabla f^*$. For that kind of mapping F, it's known that $\operatorname{lip} F(\bar{v})$ is the maximum of the norms $\|J\|$ over the compact set $\overline{\nabla} F(\bar{v})$ comprised of the linear mappings J obtained by taking limits of the Jacobian mappings J^ν existing at points $v^\nu \to \bar{v}$. But here it's more specific through application of generalized twice differentiability to f^*: the limit set in question is $\nabla^2_* f^*(\bar{v}\,|\,\bar{x})$. Thus, $\operatorname{lip} \nabla f^*(\bar{v})$ is the max of the norms of the linear mappings $J \in \nabla^2_* f^*(\bar{v}\,|\,\bar{x})$, hence the max of $\operatorname{def} J$ for such J, since these mappings are self-adjoint and positive semi-definite. But from (4.7) we have $\nabla^2_* f^*(\bar{v}\,|\,\bar{x}) = \{\, L^{-1} \mid L \in \nabla^2_* f(\bar{x}\,|\,\bar{v})\}$, so this max is reciprocal to the minimum on the right of (4.10), which was our goal to confirm. The alternative statement in (4.11) is the translation of (4.10) by way of the q-L connection in (4.6). □

The formula of Gfrerer in [2, Theorem 4.3] that corresponds to (4.10) is expressed in other language. Instead of the generalized linear mappings $L \in \nabla^2_* f(\bar{x}\,|\,\bar{v})$ it centers on the subspaces of $I\!R^n \times I\!R^n$ comprising their graphs and introduces in them bases and coordinate systems to obtain matrices that can be put to tests of definiteness. This relates to Gfrerer's different route to the formula's derivation, which passes through his efforts with Outrata in [3] at understanding graphical derivatives and coderivatives of mappings T more general than ∂f, especially those having Lipschitzian manifolds as their graphs. In that work, $\operatorname{gph} T$ was deemed to be "smooth" at $(\bar{x}, \bar{v})$ whenever the tangent cone there, the graph of $DT(\bar{x}\,|\,\bar{v})$, was a subspace — a weaker property in general than ours in Definition 4.1(a), which reflects the original view of smoothness in [10]. However, this discrepancy in concept has no impact on the modulus formula in the spotlight here, with $T = \partial f$ being graphically Lipshitzian locally by Theorem 1.1(e). That can be seen through Lipschitzian local parameterization of $\operatorname{gph} T$ and the fact (as applied to difference quotient functions) that continuous convergence to a continuous function implies uniform convergence [16, 7.14].

Conjecture 4.6 (equivalent expressions for strict second-order subderivatives)**.** *When f is prox-regular at $\bar{x}$ for $\bar{v}$, the function*

$$h(\xi) := \min \{\, q(\xi) \mid q \in \operatorname{quad} f(\bar{x}\,|\,\bar{v}) \,\} \tag{4.12}$$

in (4.11) may coincide with $h(\xi) = d^2_* f(\bar{x}\,|\,\bar{v})(\xi)$ in Definition 2.1 and furthermore with the function h in Proposition 3.10.

Proving or disproving this conjecture could lead to deeper understanding of the different facets of prox-regularity in second-order variational analysis.

5 Application to variational sufficiency

To illustrate how the tests developed here can be put to use, we adapt them into criteria for local optimality in the fundamental problem

$$\text{minimize } \varphi(x, u) \text{ subject to } u = 0, \tag{5.1}$$

for a closed proper function φ on $I\!R^n \times I\!R^m$. The first-order condition

$$\exists\, \bar{y} \;\text{ such that }\; (0, \bar{y}) \in \partial\varphi(\bar{x}, 0) \tag{5.2}$$

is known to be necessary under the basic constraint qualification that only $y = 0$ satisfies $(0, y) \in \partial^\infty \varphi(\bar{x}, 0)$. Our concern here, though, is with what can be combined with this first-order condition on the second-order level to get a sufficient condition for local optimality with attractive consequences. For that, we focus on the *strong variational sufficient condition* introduced in [12] with elicitation parameter e, which combines (5.2) with

$$\begin{array}{l}\exists\, e > 0 \;\text{ such that the function }\; \varphi_e(x, u) := \varphi(x, u) + e j(u), \;\text{ still with} \\ (0, \bar{y}) \in \partial\varphi_e(\bar{x}, 0), \;\text{ is variationally strongly convex at } (\bar{x}, 0) \text{ for } (0, \bar{y}).\end{array} \tag{5.3}$$

Strong variational sufficiency as the combination of (5.2) and (5.3) is sufficient for local optimality regardless of any constraint qualification and brings with it an array of properties which support the local convergence of the progressive decoupling algorithm in [12] and the augmented Lagrangian method for generalized nonlinear programming in [14], in particular. The open door to something new is our ability now to check for variational strong convexity by means of formulas for the convexity modulus $\operatorname{cnv}\varphi(\bar{x},0\,|\,0,\bar{y})$.

In [13], strong variational sufficiency was shown to be equivalent in classical nonlinear programming to the standard strong second-order sufficient condition, and ties to new and old sufficient conditions in generalized nonlinear programming were worked out at well. On the other hand, it was pointed out in [12] that in the elementary case of problem (5.1) where φ is a $\mathcal{C}^2$ function on $\mathbb{R}^n \times \mathbb{R}^m$, strong variational sufficiency comes down to the sufficient condition in multivariate calculus that

$$\nabla\varphi(\bar{x},0) = (0,\bar{y}), \quad \nabla^2_{xx}\varphi(\bar{x},0) \text{ positive-definite.} \tag{5.4}$$

It can be imagined therefore that our tests might therefore produce substitutes for (5.3) in the form of *positive-definiteness of generalized partial second derivatives of $\varphi(x,u)$ in x*, and indeed this will come into view.

Our claims will depend on first knowing that φ is prox-regular at $(\bar{x},0)$ for the subgradient $(0,\bar{y})$. But of course that corresponds by Corollary 1.3 to testing whether $\operatorname{cnv}\varphi(\bar{x},0\,|\,0,\bar{y}) > -\infty$, which can be carried out by the formula in Theorem 2.2 or the alternative in Theorem 3.6.

Theorem 5.1 (derivative tests for strong variational sufficiency). *Under the first-order condition (5.2) and the assumption that $\operatorname{cnv}\varphi(\bar{x},0\,|\,0,\bar{y}) > -\infty$, the following are equivalent substitutes for the condition (5.3) in the combination that constitutes strong variational sufficiency at $\bar{x}$ for $\bar{y}$:*

(a) $d^2_*\varphi(\bar{x},0\,|\,0,\bar{y})(\xi,0) > 0$ *when* $\xi \neq 0$.

(b) $(0,\bar{y})$ *is a proximal subgradient, and* $(\psi,\eta) \in \partial^2_*\varphi(\bar{x},0\,|\,0,\bar{y})(\xi,0),\ \xi \neq 0 \Longrightarrow \psi{\cdot}\xi > 0$.

(c) $\operatorname{cnv}\varphi(\bar{x},0\,|\,0,\bar{y}) > -\infty$ *and also* $L \in \nabla^2_*\varphi(\bar{x},0\,|\,0,\bar{y}) \Longrightarrow (\xi,0){\cdot}L(\xi,0) > 0$ *for* $\xi \neq 0$ *in (4.4), or equivalently, the partial forms* $q_{xx}(\xi) = q(\xi,0)$ *on* $\mathbb{R}^n$ *coming from the generalized quadratic forms* $q \in \operatorname{quad}\varphi(\bar{x},0\,|\,0,\bar{y})$ *on* $\mathbb{R}^n \times \mathbb{R}^n$ *are all positive-definite.*

Proof. In (a) we are dealing with a closed function $h = d^2_*\varphi(\bar{x},0\,|\,0,\bar{y})$, positively homogeneous of degree 2, that is proper by Theorem 2.2 and our assumption that $\operatorname{cnv}\varphi(\bar{x},0\,|\,0,\bar{y}) > -\infty$. On the unit sphere $S = \{\,(\xi,\omega) \,|\, |(\xi,\omega)| = 1\,\}$, which dictates the values of h everywhere else, h is bounded below by some b. We are assuming that $h(\xi,0) > 0$ when $\xi \neq 0$, and are claiming that

$$\exists\, e \geq 0,\ \varepsilon > 0, \text{ such that } h(\xi,\omega) + ej(\omega) > \varepsilon \text{ on the sphere } S. \tag{5.5}$$

If not, there would be $e^\nu \nearrow \infty$ and $\varepsilon^\nu \searrow 0$ along with $(\xi^\nu,\omega^\nu) \in S$ such that $\varepsilon^\nu \geq h(\xi^\nu,\omega^\nu) + e^\nu j(\omega^\nu)$, and it could be supposed that $(\xi^\nu,\omega^\nu) \to (\xi,\omega) \in S$. Then $\varepsilon^\nu \geq b + e^\nu j(\omega^\nu)$, so necessarily $j(\omega^\nu) \to 0$, and by lower semicontinuity $h(\xi,\omega) \leq 0$ with $\omega = 0$. But that violates our assumption.

In (b) the argument is the same, except for being applied instead to the function h associated with φ by Proposition 3.10. In (c), we turn to the function

$$h(\xi,\omega) = 2\inf\{\,q(\xi,\omega) \,|\, q \in \operatorname{quad} f(\bar{x}\,|\,\bar{v})\,\}. \tag{5.6}$$

Because the collection of functions q in (5.6) is compact with respect to epiconvergence, the union of their epigraphs is closed, and that union is then the epigraph of $\frac{1}{2}h$, hence h is lower semicontinuous. It's also proper and positively homogenous of degree 2, so we are back in the same picture and the argument for (a) goes through once more. ☐

Note that if the answer to Conjecture 4.6 were known to be positive, the fact that the three different prescriptions in Theorem 5.1 lead to the same conclusion would be obvious.

References

[1] Benko, M., and Rockafellar, R.T., "Primal-dual stability in local optimality," *J. Optimization Theory and Applications*, forthcoming.

[2] Gfrerer, H., "On second-order variational analysis of variational convexity of prox-regular functions," submitted for publication in August 2024.

[3] Gfrerer, H., and Outrata, J., "On (local) analysis of multifunctions via subspaces contained in graphs of generalized derivatives," *J. Math. Analysis and Applications* **508** (2022), 125895.

[4] Khanh, P.D., Mordukhovich, B.S., and Phat, V.T., "Variational convexity of functions and variational sufficiency in optimization," *SIAM J. Optimization* **33** (2023), 1121–1158.

[5] Khanh, P.D., Khoa, V.V.H., Mordukhovich, B.S., and Phat, V.T., "Variational and strong variational convexity in infinite-dimensional variational analysis," *SIAM J. Optimization* **34** (2024), 2756–2787.

[6] Mordukhovich, B.S., *Second-Order Variational Analysis in Optimization, Variational Stability and Control*, Springer 2024.

[7] Nghia, T.T.A., Pham, D.T., and Tran, T.T.T., "On the positive definiteness of the limiting coderivative for set-valued mappings," *Set-Valued Analysis* **29** (2021), 283–300.

[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., "Maximal monotone relations and the second derivatives of nonsmooth functions," *Annales Inst. Henri Poincaré* **2** (1985), 167–184.

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

[12] 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.

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

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

[15] Rockafellar, R.T., "Variational convexity and prox-regularity," *J. Convex Analysis*, forthcoming.

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