Smoothness of Subgradient Mappings and Its Applications in Parametric Optimization

NGUYEN T. V. HANG[1] and EBRAHIM SARABI[2]

Abstract. We demonstrate that the concept of strict proto-differentiability of subgradient mappings can play a similar role as smoothness of the gradient mapping of a function in the study of subgradient mappings of prox-regular functions. We then show that metric regularity and strong metric regularity are equivalent for a class of generalized equations when this condition is satisfied. For a class of composite functions, called $\mathcal{C}^2$-decomposable, we argue that strict proto-differentiability can be characterized via a simple relative interior condition. Leveraging this observation, we present a characterization of the continuous differentiability of the proximal mapping for this class of function via a certain relative interior condition. Applications to the study of strong metric regularity of the KKT system of a class of composite optimization problems are also provided.

Keywords. strictly proto-differentiable mappings, strong metric regularity, proximal mappings, $\mathcal{C}^2$-decomposable functions.

Mathematics Subject Classification (2000) 90C31, 65K99, 49J52, 49J53

1 Introduction

Smoothness of the gradient mapping of a function, or twice differentiability of the function itself, has numerous applications including those in designing efficient numerical algorithms for solving optimization problems such as Newton and Newton-like methods. That inspires us to ask the following question: Is there a counterpart of the latter concept for subgradient mapping of nonsmooth functions? If so, what should we expect from such a property when dealing with different second-order variational constructions? Borrowing the concept of *strict proto-differentiability* from [40], we denominate that this concept plays a similar role for subgradient mappings as the smoothness of the gradient mapping. Given a finite dimensional Hilbert space $\mathbf{X}$ and a proper function $f : \mathbf{X} \to \overline{\mathbb{R}} := [-\infty, \infty]$, recall from [40] that the subgradient mapping ∂f is said to be strictly proto-differentiable at $\bar{x}$ for $\bar{v}$ with $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$ if the regular (Clarke) tangent cone to $\operatorname{gph} \partial f$ at $(\bar{x}, \bar{v})$, denoted by $\widehat{T}_{\operatorname{gph} \partial f}(\bar{x}, \bar{v})$, and the paratingent cone to $\operatorname{gph} \partial f$ at $(\bar{x}, \bar{v})$, denoted by $\widetilde{T}_{\operatorname{gph} \partial f}(\bar{x}, \bar{v})$, coincide; see (2.1) for the definitions of the these cones. At first glance, it may not be clear why this concept can be considered as a proper candidate in this regard but our developments in this paper will confirm the unique potential of this concept to answer our questions. For instance, it is well-known that the Hessian matrix of a twice continuously differentiable function is symmetric. In Section 3, we are going to demonstrate that for a broad class of functions, called prox-regular, the graphical derivative and coderivative (generalized Hessian) of subgradient mappings coincide provided that strict proto-differentiability of subgradient mappings is satisfied; see Theorem 3.9. This can be viewed as a far-reaching extension of the aforementioned property of the Hessian matrix of twice continuously differentiable functions and also as a motivation to investigate further strict proto-differentiability for important classes of functions.

Poliquin and Rockafellar obtained interesting characterizations of strict proto-differentiability of subgradient mappings of prox-regular functions in [41]. In particular, it was shown there that this concept amounts to the continuous differentiability of proximal mappings under certain assumptions. The latter property of proximal mappings was studied for convex sets in Hilbert

[1]School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 639798 and Institute of Mathematics, Vietnam Academy of Science and Technology, Hanoi, Vietnam (thivanhang.nguyen@ntu.edu.sg). Research of this author is partially supported by Singapore National Academy of Science under the grant RIE2025 NRF International Partnership Funding Initiative.

[2]Department of Mathematics, Miami University, Oxford, OH 45065, USA (sarabim@miamioh.edu). Research of this author is partially supported by the U.S. National Science Foundation under the grant DMS 2108546.

spaces by Holmes in [18] in the 1970s using the implicit function theorem and imposing a certain assumption on the boundary of the convex set under consideration. The characterization in [41] indicates that concepts of second-order variational analysis are perhaps more appropriate to investigate the continuous differentiability of proximal mappings. Indeed, we recently provided in [16] a simple characterization of the continuous differentiability of the proximal mapping of a certain composite function via a relative interior condition by leveraging strict proto-differentiability of their subgradient mappings.

While strict proto-differentiability was introduced in [40], it was not studied systematically for different classes of functions commonly seen in constrained and composite optimization problems. Recently, the authors characterized it for polyhedral functions, those that their epigraphs are polyhedral convex sets, in [14] and for a class of composite functions in [15], obtained from the composition of a polyhedral function and a $\mathcal{C}^2$-smooth function. In these works, it was shown that strict proto-differentiability of subgradient mappings under consideration amounts to the subgradient taken from the relative interior of the subdifferential set. Such a characterization had two major consequences. First, it allowed the authors to justify the equivalence of metric regularity and strong metric regularity for a class of generalized equations at their nondegenerate solutions (see the paragraph after (5.25) for its definition). And second, we were able to characterize the continuous differentiability of the proximal mapping for certain composite functions.

This paper aims to lay the foundation for the systematic study of strict proto-differentiability of subgradient mappings of prox-regular functions. In doing so, we utilize a geometric approach by studying first strict smoothness of closed sets, which is different from the path taken in [41]. Then, we provide a simple characterization of this concept for an important class of functions, called $\mathcal{C}^2$-*decomposable*. This class of functions, as shown by Shapiro in [51], encompasses important classes of functions including polyhedral functions and various eigenvalue/singular value functions that often appear in applications. Moreover, we present various consequences of strict proto-differentiability in stability properties of generalized equations and use those to characterize the continuous differentiability of the proximal mapping.

The outline of the paper is as follows. We begin in Section 2 by recalling our notation and characterizing the strict smoothness of sets and using them to present a characterization of strict proto-differentiability of subgradient mappings of prox-regular functions. Section 3 presents some important consequences of strict proto-differentiability for various second-order variational constructions. In Section 4, we study the relationship between metric regualrity and strong metric regularity of a class of generalized equations under strict proto-differentiability and show that they are, indeed, equivalent. Leveraging then this equivalence, we characterize the continuous differentiability of the proximal mapping for prox-regular functions. Section 5 is devoted to establishing a chain rule for strict proto-differentiability of $\mathcal{C}^2$-decomposable functions and deriving a simple characterization of this concept via a relative interior condition. Using this, we study strong metric regularity of the KKT system of a class of composite optimization problems.

2 Strict Proto-Differentiability of Subgradient Mappings

In what follows, suppose that $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ are finite dimensional Hilbert spaces. We denote by $\mathbb{B}$ the closed unit ball in the space in question and by $\mathbb{B}_r(x) := x + r\mathbb{B}$ the closed ball centered at x with radius $r > 0$. In the product space $\mathbf{X} \times \mathbf{Y}$, we use the norm $\|(w, u)\| = \sqrt{\|w\|^2 + \|u\|^2}$ for any $(w, u) \in \mathbf{X} \times \mathbf{Y}$. Given a nonempty set $C \subset \mathbf{X}$, the symbols $\operatorname{int} C$, $\operatorname{ri} C$, C^*, and $\operatorname{par} C$ signify its interior, relative interior, polar cone, and the linear subspace parallel to the affine hull of C, respectively. For any set C in $\mathbf{X}$, its indicator function is defined by $\delta_C(x) = 0$ for $x \in C$ and $\delta_C(x) = \infty$ otherwise. We denote by P_C the projection mapping onto C and by $\operatorname{dist}(x, C)$ the distance between $x \in \mathbf{X}$ and a set C. For a vector $w \in \mathbf{X}$, the subspace $\{tw \mid t \in \mathbb{R}\}$ is denoted by $[w]$. Let $\{C^t\}_{t>0}$ be a parameterized family of sets in $\mathbf{X}$. Its inner and outer limit sets are defined, respectively, by

$$\liminf_{t\searrow 0} C^t = \big\{x \in \mathbf{X}|\ \forall\, t_k \searrow 0\ \exists\, x^{t_k} \to x \ \text{ with }\ x^{t_k} \in C^{t_k}\ \text{ for }\ k\ \text{ sufficiently large}\big\},$$
$$\limsup_{t\searrow 0} C^t = \big\{x \in \mathbf{X}|\ \exists\, t_k \searrow 0\ \exists\ x^{t_k} \to x\ \text{ with }\ x^{t_k} \in C^{t_k}\big\};$$

see [50, Definition 4.1]. The limit set of $\{C^t\}_{t>0}$ exists if $\liminf_{t\searrow 0} C^t = \limsup_{t\searrow 0} C^t =: C$, written as $C^t \to C$ when $t \searrow 0$. A sequence $\{f^k\}_{k\in\mathbb{N}}$ of functions $f^k : \mathbf{X} \to \overline{\mathbb{R}}$ is said to *epi-converge* to a function $f : \mathbf{X} \to \overline{\mathbb{R}}$ if we have $\operatorname{epi} f^k \to \operatorname{epi} f$ as $k \to \infty$, where $\operatorname{epi} f = \{(x,\alpha) \in \mathbf{X} \times \mathbb{R}|\, f(x) \le \alpha\}$ is the epigraph of f; see [50, Definition 7.1] for more details on epi-convergence. We denote by $f^k \xrightarrow{e} f$ the epi-convergence of $\{f^k\}_{k\in\mathbb{N}}$ to f.

Given a nonempty set $\Omega \subset \mathbf{X}$ with $\bar{x} \in \Omega$, the tangent and derivable (adjacent) cones to Ω at $\bar{x}$ are defined, respectively, by

$$T_\Omega(\bar{x}) = \limsup_{t\searrow 0} \frac{\Omega - \bar{x}}{t} \quad \text{and} \quad \breve{T}_\Omega(\bar{x}) = \liminf_{t\searrow 0} \frac{\Omega - \bar{x}}{t}.$$

Consider a set-valued mapping $F : \mathbf{X} \rightrightarrows \mathbf{Y}$. According to [50, Definition 8.33], the *graphical derivative* of F at $\bar{x}$ for $\bar{y}$ with $(\bar{x},\bar{y}) \in \operatorname{gph} F$ is the set-valued mapping $DF(\bar{x},\bar{y}) : \mathbf{X} \rightrightarrows \mathbf{Y}$ defined via the tangent cone to $\operatorname{gph} F$ at $(\bar{x},\bar{y})$ by

$$\eta \in DF(\bar{x},\bar{y})(w) \iff (w,\eta) \in T_{\operatorname{gph} F}(\bar{x},\bar{y}),$$

or, equivalently, $\operatorname{gph} DF(\bar{x},\bar{y}) = T_{\operatorname{gph} F}(\bar{x},\bar{y})$. When $F(\bar{x})$ is a singleton consisting of $\bar{y}$ only, the notation $DF(\bar{x},\bar{y})$ is simplified to $DF(\bar{x})$. It is easy to see that for a single-valued mapping F which is differentiable at $\bar{x}$, the graphical derivative $DF(\bar{x})$ boils down to the Jacobian of F at $\bar{x}$, denoted by $\nabla F(\bar{x})$. The set-valued mapping F is said to be *proto-differentiable* at $\bar{x}$ for $\bar{y}$ if $\breve{T}_{\operatorname{gph} F}(\bar{x},\bar{y}) = T_{\operatorname{gph} F}(\bar{x},\bar{y})$. Note that the inclusion '$\subset$' in the later equality always holds, and so proto-differentiability requires that opposite inclusion be satisfied. Proto-differentiability was introduced in [49] and has been studied extensively for subgradient mappings of different classes of functions in [27, 29, 40, 50].

A more restrictive version of proto-differentiability was introduced in [40] and is the main subject of our study in this paper. To present its definition, recall that the regular (Clarke) tangent cone and the paratingent cone to Ω at $\bar{x}$ are defined, respectively, by

$$\widehat{T}_\Omega(\bar{x}) = \liminf_{x\xrightarrow{\Omega}\bar{x},\, t\searrow 0} \frac{\Omega - x}{t} \quad \text{and} \quad \widetilde{T}_\Omega(\bar{x}) = \limsup_{x\xrightarrow{\Omega}\bar{x},\, t\searrow 0} \frac{\Omega - x}{t}, \tag{2.1}$$

where the symbol $x \xrightarrow{\Omega} \bar{x}$ means that $x \to \bar{x}$ with $x \in \Omega$. It can be immediately observed that the inclusions

$$\widehat{T}_\Omega(\bar{x}) \subset \breve{T}_\Omega(\bar{x}) \subset T_\Omega(\bar{x}) \subset \widetilde{T}_\Omega(\bar{x}) \tag{2.2}$$

always hold. Below, we record two important properties of the paratingent cone, used later in our developments.

Proposition 2.1. *Assume that Ω is a subset of $\mathbf{X}$ and $\bar{x} \in \Omega$. Then the following properties hold for the paratingent cone to Ω at $\bar{x}$.*

(a) $\widetilde{T}_\Omega(\bar{x}) = -\widetilde{T}_\Omega(\bar{x})$.

(b) $\limsup_{x\xrightarrow{\Omega}\bar{x}} T_\Omega(x) \subset \widetilde{T}_\Omega(\bar{x})$.

Proof. While the property in (a) is well-known, we provide a short proof. Take $w \in \widetilde{T}_\Omega(\bar{x})$. By definition, we find $t_k \searrow 0$, $x^k \xrightarrow{\Omega} \bar{x}$, and $w^k \to w$ such that $\tilde{x}^k := x^k + t_k w^k \in \Omega$ for all k sufficiently large. Thus, we get $\tilde{x}^k + t_k(-w^k) = x^k \in \Omega$. Since $-w^k \to -w$, we arrive at $-w \in \widetilde{T}_\Omega(\bar{x})$, which proves (a). The property in (b) results from a standard diagonalization argument and can be gleaned from [1, Proposition 4.5.6]. □

We proceed with recalling the definitions of smoothness of a set and a Lipschitzian manifold, appeared first in [47, pp. 169–173].

Definition 2.2. Assume that Ω is a subset of $\mathbf{X}$ and $\bar{x} \in \Omega$.

(a) The set Ω is called smooth at $\bar{x}$ if $\breve{T}_\Omega(\bar{x}) = T_\Omega(\bar{x})$ and $T_\Omega(\bar{x})$ is a linear subspace of $\mathbf{X}$.

(b) The set Ω is called strictly smooth at $\bar{x}$ if $\widehat{T}_\Omega(\bar{x}) = \widetilde{T}_\Omega(\bar{x})$.

(c) The set Ω is called a Lipschitzian manifold around $\bar{x}$ if there are an open neighborhoods O of $\bar{x}$, a splitting $\mathbf{X} = \mathbf{Y} \times \mathbf{Z}$ with $\mathbf{Y}$ and $\mathbf{Z}$ being two finite dimensional Hilbert spaces, and a $\mathcal{C}^1$-diffeomorphism Φ from O onto an open neighborhood of $U := U' \times U''$ of $\mathbf{Y} \times \mathbf{Z}$ such that $\Phi(\Omega \cap O) = (\operatorname{gph} f) \cap U$, where $f : U' \to U''$ is a Lipschitz continuous function.

We are now going to present a characterization of strict smoothness of sets that are Lipschitzian manifolds.

Theorem 2.3 (characterization of strictly smooth sets)**.** *Assume that Ω is a subset of $\mathbf{X}$, $\bar{x} \in \Omega$, and that Ω is locally closed around $\bar{x}$. Consider the following properties:*

(a) *Ω is strictly smooth at $\bar{x}$;*

(b) $\lim_{x \xrightarrow{\Omega} \bar{x}} T_\Omega(x) = T_\Omega(\bar{x})$.

Then, the implication (a) $\Longrightarrow$ (b) *always holds. If, in addition, Ω is a Lipschitzian manifold around $\bar{x}$, then the opposite implication holds as well. In this case,* (a) *and* (b) *are also equivalent to the following properties:*

(c) *$T_\Omega(x)$ converges to $T_\Omega(\bar{x})$ as $x \to \bar{x}$ in the set of points $x \in \Omega$ for which $\breve{T}_\Omega(x) = T_\Omega(x)$.*

(d) *$T_\Omega(x)$ converges to $T_\Omega(\bar{x})$ as $x \to \bar{x}$ in the set of points $x \in \Omega$ at which Ω is smooth.*

Proof. If (a) holds, it follows from [50, Theorem 6.26] that

$$\liminf_{x \xrightarrow{\Omega} \bar{x}} T_\Omega(x) = \widehat{T}_\Omega(\bar{x}).$$

Combining this and Proposition 2.1(b) tells us that

$$\widehat{T}_\Omega(\bar{x}) = \liminf_{x \xrightarrow{\Omega} \bar{x}} T_\Omega(x) \subset \limsup_{x \xrightarrow{\Omega} \bar{x}} T_\Omega(x) \subset \widetilde{T}_\Omega(\bar{x}),$$

which, together with (a), yields (b). The implications (b) $\Longrightarrow$ (c) $\Longrightarrow$ (d) clearly hold. Assume now that Ω is a Lipschitzian manifold at $\bar{x}$ and that $\Omega' \subset \Omega$ is the set of points $x \in \Omega$ at which Ω is smooth. We conclude from [47, Theorem 3.5(a)] that Ω' differs from Ω by only a set of measure zero. Furthermore, it follows from [47, Theorem 3.5(c)] that (d) is equivalent to (a), which completes the proof. □

Recall from [50, Definition 9.53] that the strict graphical derivative of a set-valued mapping F at $\bar{x}$ for $\bar{y}$ with $(\bar{x}, \bar{y}) \in \operatorname{gph} F$, is the set-valued mapping $\widetilde{D}F(\bar{x}, \bar{y}) : \mathbf{X} \rightrightarrows \mathbf{Y}$, defined by

$$\eta \in \widetilde{D}F(\bar{x}, \bar{y})(w) \iff (w, \eta) \in \widetilde{T}_{\operatorname{gph} F}(\bar{x}, \bar{y}).$$

Following [40], we say that the set-valued mapping F is *strictly* proto-differentiable at $\bar{x}$ for $\bar{y}$ provided that $\operatorname{gph} F$ is strictly smooth at $(\bar{x}, \bar{y})$, namely $\widehat{T}_{\operatorname{gph} F}(\bar{x}, \bar{y}) = \widetilde{T}_{\operatorname{gph} F}(\bar{x}, \bar{y})$. Indeed, since the inclusion '$\subset$' always holds due to (2.2), the strict proto-differentiability of F at $\bar{x}$ for $\bar{y}$ amounts to the validity of the opposite inclusion therein. According to (2.2), the strict proto-differentiability of F $\bar{x}$ for $\bar{y}$ implies that $\widetilde{T}_{\operatorname{gph} F}(\bar{x}, \bar{y}) = T_{\operatorname{gph} F}(\bar{x}, \bar{y})$, which in turn demonstrates that $DF(\bar{x}, \bar{y})(w) = \widetilde{D}F(\bar{x}, \bar{y})(w)$ for any $w \in \mathbf{X}$. While proto-differentiability has been studied for different classes of functions, its strict version has not received much attention. In fact, we recently characterized in [14, Theorem 3.5(c)] this property for subgradient mappings of polyhedral functions via a relative interior condition; see also [15, Theorem 3.10] for a similar result for certain composite functions. Our main objective is to extend the latter characterization for a rather large class of composite functions.

Given $\Omega \subset \mathbf{X}$ and $\bar{x} \in \Omega$, its regular normal cone $\widehat{N}_\Omega(\bar{x})$ at $\bar{x}$ is defined by $\widehat{N}_\Omega(\bar{x}) = T_\Omega(\bar{x})^*$. For $\bar{x} \notin \Omega$, we set $\widehat{N}_\Omega(\bar{x}) = \emptyset$. The (limiting/Mordukhovich) normal cone $N_\Omega(\bar{x})$ to Ω at $\bar{x}$ is the set of all vectors $\bar{v} \in \mathbf{X}$ for which there exist sequences $\{x^k\}_{k\in\mathbb{N}} \subset \Omega$ and $\{v^k\}_{k\in\mathbb{N}}$ with $v^k \in \widehat{N}_\Omega(x^k)$ such that $(x^k, v^k) \to (\bar{x}, \bar{v})$. When Ω is convex, both normal cones boil down to that of convex analysis. Given a function $f : \mathbf{X} \to \overline{\mathbb{R}}$ and a point $\bar{x} \in \mathbf{X}$ with $f(\bar{x})$ finite, a vector $v \in \mathbf{X}$ is called a subgradient of f at $\bar{x}$ if $(v, -1) \in N_{\text{epi}\, f}(\bar{x}, f(\bar{x}))$. The set of all subgradients of f at $\bar{x}$ is denoted by $\partial f(\bar{x})$. A function $f\colon \mathbf{X} \to \overline{\mathbb{R}}$ is called prox-regular at $\bar{x}$ for $\bar{v}$ if f is finite at $\bar{x}$ and locally lower semicontinuous (lsc) around $\bar{x}$ with $\bar{v} \in \partial f(\bar{x})$, and there exist constants $\varepsilon > 0$ and $r \geq 0$ such that

$$\begin{cases} f(x') \geq f(x) + \langle v, x' - x\rangle - \frac{r}{2}\|x' - x\|^2 \ \text{ for all } \ x' \in \mathbb{B}_\varepsilon(\bar{x}) \\ \text{whenever } \ (x, v) \in (\text{gph}\, \partial f) \cap \mathbb{B}_\varepsilon(\bar{x}, \bar{v}) \ \text{ with } \ f(x) < f(\bar{x}) + \varepsilon. \end{cases} \tag{2.3}$$

The function f is called subdifferentially continuous at $\bar{x}$ for $\bar{v}$ if the convergence $(x^k, v^k) \to (\bar{x}, \bar{v})$ with $v^k \in \partial f(x^k)$ yields $f(x^k) \to f(\bar{x})$ as $k \to \infty$. Important examples of prox-regular and subdifferentially continuous functions are convex functions and strongly amenable functions in the sense of [50, Definition 10.23]. Below, we provide a characterization of strict proto-differentiability of subgradient mappings of prox-regular functions, which is a slight improvement of [41, Corollary 4.3].

In what follows, we say that a sequence of set-valued mappings $F^k : \mathbf{X} \rightrightarrows \mathbf{Y}$, $k \in \mathbb{N}$, *graph*-converges to $F : \mathbf{X} \rightrightarrows \mathbf{Y}$ if the sequence of sets $\{\text{gph}\, F^k\}_{k\in\mathbb{N}}$ is convergent to $\text{gph}\, F$.

Corollary 2.4 (characterizations of strict proto-differentiability)**.** *Assume that $f : \mathbf{X} \to \overline{\mathbb{R}}$ is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$. Then the following properties are equivalent:*

(a) *∂f is strictly proto-differentiable at $\bar{x}$ for $\bar{v}$;*
(b) *$D(\partial f)(x, v)$ graph-converges to $D(\partial f)(\bar{x}, \bar{v})$ as $(x, v) \to (\bar{x}, \bar{v})$ and $(x, v) \in \text{gph}\, \partial f$;*
(c) *$D(\partial f)(x, v)$ graph-converges to $D(\partial f)(\bar{x}, \bar{v})$ as $(x, v) \to (\bar{x}, \bar{v})$ in the set of pairs $(x, v) \in \text{gph}\, \partial f$ for which ∂f is proto-differentiable;*
(d) *$D(\partial f)(x, v)$ graph-converges to $D(\partial f)(\bar{x}, \bar{v})$ as $(x, v) \to (\bar{x}, \bar{v})$ in the set of pairs $(x, v) \in \text{gph}\, \partial f$ for which ∂f is proto-differentiable and $T_{\text{gph}\, \partial f}(x, v)$ is a linear subspace.*

Proof. Recall that (a) amounts to $\text{gph}\, \partial f$ being strictly smooth at $(\bar{x}, \bar{v})$ and that for all $(x, v) \in \text{gph}\, \partial f$ we have $\text{gph}\, D(\partial f)(x, v) = T_{\text{gph}\, \partial f}(x, v)$. According to [40, Theorem 4.7], the graphical set $\text{gph}\, \partial f$ is a Lipschitzian manifold around $(\bar{x}, \bar{v})$. The equivalence of (a)-(d) is a direct consequence of Theorem 2.3. $\square$

Note that the equivalence of (a), (c), and (d) in Corollary 2.4 were observed before in [41, Corollary 4.3] using an argument via the Moreau envelope of prox-regular functions. While the properties in (c) and (d) also appeared in [41, Corollary 4.3], the latter didn't determine to which mapping the graph-convergences happen. Also, the property in (b) is new and was not observed in [41].

3 Remarkable Consequences of Strict Proto-Differentiability

This section is devoted to establishing some interesting consequences of strict proto-differentiability of subgradient mappings of prox-regular functions, which highlight the importance of this property in second-order variational analysis of various classes of optimization problems. We begin with recalling the definition of generalized quadratic form from [38, Definition 2.1].

Definition 3.1. Suppose that $\varphi : \mathbf{X} \to \overline{\mathbb{R}}$ is a proper function.

(a) We say that the subgradient mapping $\partial\varphi : \mathbf{X} \rightrightarrows \mathbf{X}$ is generalized linear if its graph is a linear subspace of $\mathbf{X} \times \mathbf{X}$.

(b) We say that φ is a generalized quadratic form on $\mathbf{X}$ if $\operatorname{dom}\varphi$ is a linear subspace of $\mathbf{X}$ and there exists a linear symmetric mapping A from $\operatorname{dom}\varphi$ to $\mathbf{X}$ (i.e. $\langle A(x), y\rangle = \langle x, A(y)\rangle$ for any $x, y \in \operatorname{dom}\varphi$) such that φ has a representation of form

$$\varphi(x) = \langle A(x), x\rangle \quad \text{for all } \ x \in \operatorname{dom}\varphi.$$

It was shown in [38, Theorem 2.5] that for a proper lsc convex function $\varphi : \mathbf{X} \to \overline{\mathbb{R}}$ with $\varphi(0) = 0$, φ is a generalized quadratic form on $\mathbf{X}$ if and only if $\partial\varphi$ is generalized linear. In this case, one can glean from the proof of the latter result that

$$\varphi(x) = \tfrac{1}{2}\langle A(x), x\rangle + \delta_S(x) \quad \text{for } \ x \in \mathbf{X}, \tag{3.1}$$

where $S := \operatorname{dom}\partial\varphi$ is a linear subspace of $\mathbf{X}$ and $A : S \to S$ is linear and symmetric mapping defined by $A(x) = P_S(y)$[‡]. Define now $\widehat{A} : \mathbf{X} \to \mathbf{X}$ by $\widehat{A}(x) = A(P_S(x))$ for $x \in \mathbf{X}$. It is not hard to see that $\widehat{A}$ is a linear symmetric mapping on X satisfying $\widehat{A}(x) = A(x)$ for all $x \in S$, and therefore can equivalently represent φ as follows

$$\varphi(x) = \tfrac{1}{2}\langle \widehat{A}(x), x\rangle + \delta_S(x) \quad \text{for all } \ x \in \mathbf{X}.$$

To present our first result in this section, we begin by recalling the concept of second subderivatives. Given a function $f : \mathbf{X} \to \overline{\mathbb{R}}$ and $\bar{x} \in \mathbf{X}$ with $f(\bar{x})$ finite, the second subderivative of f at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ is defined by

$$\mathrm{d}^2 f(\bar{x}, \bar{v})(w) = \liminf_{\substack{t \searrow 0 \\ w' \to w}} \Delta_t^2 f(\bar{x}, \bar{v})(w'), \quad w \in \mathbf{X},$$

where $\Delta_t^2 f(\bar{x}, \bar{v})(w') := (f(\bar{x} + tw') - f(\bar{x}) - t\langle \bar{v}, w'\rangle)/\frac{1}{2}t^2$ is second-order difference quotients of f at $\bar{x}$ for $\bar{v}$. According to [50, Definition 13.6], f is called *twice epi-differentiable* at $\bar{x}$ for $\bar{v}$ if the functions $\Delta_t^2 f(\bar{x}, \bar{v})$ epi-converge to $\mathrm{d}^2 f(\bar{x}, \bar{v})$ as $t \searrow 0$. Twice epi-differentiability of a prox-regular function is equivalent to proto-differentiability of its subgradient mapping, according to [50, Theorem 13.40], and has been studied extensively in [27, 29, 30, 50] for different classes of functions. Following [50, Definition 8.45], we say that $v \in \mathbf{X}$ is a proximal subgradient of f at $\bar{x}$ if there exists $r \geq 0$ and a neighborhood U of $\bar{x}$ such that for all $x \in U$, one has

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

Recall also that a set $\Omega \subset \mathbf{X}$ is called regular at $\bar{x} \in \Omega$ if it is locally closed around $\bar{x}$ and $\widehat{N}_\Omega(\bar{x}) = N_\Omega(\bar{x})$. Below, we show that strict proto-differentiability of subgradient mappings immediately implies that these subgradient mappings are generalized linear. While parts (a) and (b) of the following result were observed before in [41, Corollary 4.3], we provide a different proof for (a). The proof of (b) should be known but we could find it in any publications. We supply a proof of (b) as well for the readers' convenience.

Proposition 3.2. *Assume that $f : \mathbf{X} \to \overline{\mathbb{R}}$ is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ and that ∂f is strictly proto-differentiable at $\bar{x}$ for $\bar{v}$. Then the following properties hold.*

(a) *$D(\partial f)(\bar{x}, \bar{v})$ is generalized linear.*

(b) *The set $\overline{K} := \operatorname{dom}\mathrm{d}^2 f(\bar{x}, \bar{v})$ is a linear subspace and there is a linear symmetric mapping $A : \mathbf{X} \to \mathbf{X}$ for which the second subderivative of f at $\bar{x}$ for $\bar{v}$ has a representation of the form*

$$\mathrm{d}^2 f(\bar{x}, \bar{v})(w) = \tfrac{1}{2}\langle A(w), w\rangle + \delta_{\overline{K}}(w), \quad w \in \mathbf{X}. \tag{3.2}$$

Consequently, $\mathrm{d}^2 f(\bar{x}, \bar{v})$ is a generalized quadratic form on $\mathbf{X}$.

[‡]Interested readers can find more details about this representation in the arXiv version of this paper available at arxiv.org/abs/2311.06026

(c) $\operatorname{gph} \partial f$ *is regular at* $(\bar{x}, \bar{v})$.

Proof. Strict proto-differentiability of ∂f at $\bar{x}$ for $\bar{v}$ amounts to saying that $\widehat{T}_{\operatorname{gph} \partial f}(\bar{x}, \bar{v}) = \widetilde{T}_{\operatorname{gph} \partial f}(\bar{x}, \bar{v}) = T_{\operatorname{gph} \partial f}(\bar{x}, \bar{v})$. By Proposition 2.1(a), we have $\widetilde{T}_{\operatorname{gph} \partial f}(\bar{x}, \bar{v}) = -\widetilde{T}_{\operatorname{gph} \partial f}(\bar{x}, \bar{v})$. Since the regular tangent cone $\widehat{T}_{\operatorname{gph} \partial f}(\bar{x}, \bar{v})$ is a convex cone (cf. [50, Theorem 6.26]), we conclude that $T_{\operatorname{gph} \partial f}(\bar{x}, \bar{v})$ is a linear subspace of $\mathbf{X} \times \mathbf{X}$. This, together with $\operatorname{gph} D(\partial f)(\bar{x}, \bar{v}) = T_{\operatorname{gph} \partial f}(\bar{x}, \bar{v})$, confirms that $\operatorname{gph} D(\partial f)(\bar{x}, \bar{v})$ is generalized linear and hence proves (a).

To justify (b), observe first that strict proto-differentiability of ∂f at $\bar{x}$ for $\bar{v}$ implies proto-differentiability of ∂f at $\bar{x}$ for $\bar{v}$. So, by [50, Theorem 13.40], we have $D(\partial f)(\bar{x}, \bar{v}) = \partial(\frac{1}{2}\mathrm{d}^2 f(\bar{x}, \bar{v}))$. Prox-regularity of f at $\bar{x}$ for $\bar{v}$ implies that $\bar{v}$ is a proximal subgradient of f at $\bar{x}$. This, together with [29, Proposition 2.1], that $\mathrm{d}^2 f(\bar{x}, \bar{v})$ is a proper function. Because the second subderivative $\mathrm{d}^2 f(\bar{x}, \bar{v})$ is positive homogeneous of degree 2, we arrive at $\mathrm{d}^2 f(\bar{x}, \bar{v})(0) = 0$. By [50, Proposition 13.49], there exists $\rho \geq 0$ such that the function φ, defined by $\varphi(w) = \mathrm{d}^2 f(\bar{x}, \bar{v})(w) + \|w\|^2$ for any $w \in \mathbf{X}$, is convex. Moreover, $\varphi(0) = 0$, φ is lsc, and $\partial \varphi(w) = 2D(\partial f)(\bar{x}, \bar{v})(w) + 2\rho w$. By (a), $D(\partial f)(\bar{x}, \bar{v})$ is generalized linear, and so is $\partial \varphi$. Particularly, $\overline{K} = \operatorname{dom} \partial \varphi$ is a linear subspace. According to the discussion after (3.1), there is a linear symmetric mapping $\widehat{A} : \mathbf{X} \to \mathbf{X}$ such that $\varphi(w) = \frac{1}{2}\langle \widehat{A}(w), w \rangle + \delta_{\overline{K}}(w)$ for any $w \in \mathbf{X}$. Setting $A := \widehat{A} - 2\rho I$ with I being the identity mapping from $\mathbf{X}$ onto $\mathbf{X}$, we arrive at (3.2), which proves (b).

Turing now to (c), we deduce from strict proto-differentiability of ∂f at $\bar{x}$ that $\widehat{T}_{\operatorname{gph} \partial f}(\bar{x}, \bar{v}) = T_{\operatorname{gph} \partial f}(\bar{x}, \bar{v})$. Employing [50, Corollary 6.29(b)] illustrates that $\operatorname{gph} \partial f(\bar{x}, \bar{v})$ is regular, and hence completes the proof. □

Given a function $f : \mathbf{X} \to \overline{\mathbb{R}}$ and $\bar{x} \in \mathbf{X}$ with $f(\bar{x})$ finite, the subderivative function $\mathrm{d}f(\bar{x}) \colon \mathbf{X} \to \overline{\mathbb{R}}$ is defined by

$$\mathrm{d}f(\bar{x})(w) = \liminf_{\substack{t \searrow 0 \\ w' \to w}} \frac{f(\bar{x} + tw') - f(\bar{x})}{t}.$$

The critical cone of f at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ is defined by

$$K_f(\bar{x}, \bar{v}) = \{ w \in \mathbf{X} \,|\, \langle \bar{v}, w \rangle = \mathrm{d}f(\bar{x})(w) \}.$$

For $f = \delta_\Omega$, the indicator function of a nonempty subset $\Omega \subset \mathbf{X}$, the critical cone of δ_Ω at $\bar{x}$ for $\bar{v}$ is denoted by $K_\Omega(\bar{x}, \bar{v})$. In this case, the above definition of the critical cone of a function boils down to the well known concept of a critical cone of a set (see [9, page 109]), namely $K_\Omega(\bar{x}, \bar{v}) = T_\Omega(\bar{x}) \cap [\bar{v}]^\perp$. Recall also from [50, Definition 7.25] that an lsc function $f : \mathbf{X} \to \overline{\mathbb{R}}$ is said to be subdifferentially regular at $\bar{x} \in \mathbf{X}$ if $f(\bar{x})$ is finite and $\widehat{N}_{\operatorname{epi} f}(\bar{x}, f(\bar{x})) = N_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))$. Below, we record a simple observation about the critical cone of prox-regular function, used often in this paper.

Proposition 3.3. *Assume that* $f : \mathbf{X} \to \overline{\mathbb{R}}$ *is prox-regular and subdifferentially continuous at* $\bar{x}$ *for* $\bar{v} \in \partial f(\bar{x})$. *Then the following properties are satisfied.*

(a) *The second subderivative* $\mathrm{d}^2 f(\bar{x}, \bar{v})$ *is a proper function and*

$$\operatorname{cl}\big(\operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v})\big) \subset K_f(\bar{x}, \bar{v}) \subset N_{\widehat{\partial} f(\bar{x})}(\bar{v}).$$

(b) *If, in addition,* f *is subdifferentially regular at* $\bar{x}$, *then we have* $K_f(\bar{x}, \bar{v}) = N_{\partial f(\bar{x})}(\bar{v})$ *and consequently* $K_f(\bar{x}, \bar{v})$ *is a closed convex cone.*

Proof. Observe first that prox-regularity of f at $\bar{x}$ for $\bar{v}$ implies that $\bar{v}$ is a proximal subgradient of f at $\bar{x}$. This, together with [29, Proposition 2.1], implies that $\mathrm{d}^2 f(\bar{x}, \bar{v})$ is a proper function. Appealing now to [50, Proposition 13.5] gives us the inclusions $\operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v}) \subset K_f(\bar{x}, \bar{v}) \subset$

$N_{\widehat{\partial} f(\bar{x})}(\bar{v})$. We are now going to show that $K_f(\bar{x}, \bar{v})$ is closed. To this end, recall that $\bar{v}$ is a proximal subgradient of f at $\bar{x}$, which results in $\bar{v} \in \widehat{\partial} f(\bar{x})$. Employing [50, Exercise 8.4] then allows us to represent $K_f(\bar{x}, \bar{v})$ as the level set $\{w \in \mathbf{X} \,|\, \mathrm{d}f(\bar{x})(w) - \langle \bar{v}, w\rangle \le 0\}$ of an lsc function $w \mapsto \mathrm{d}f(\bar{x})(w) - \langle \bar{v}, w\rangle$ (cf. [50, Proposition 7.4]). Thus, we conclude that $K_f(\bar{x}, \bar{v})$ is a closed set and complete the proof of (a).

To justify (b), we infer from [50, Theorem 8.30] and subdifferential regularity of f at $\bar{x}$ that

$$K_f(\bar{x}, \bar{v}) = \{w \in \mathbf{X} \,|\, \langle \bar{v}, w\rangle = \sup_{v \in \partial f(\bar{x})} \langle v, w\rangle\} = N_{\partial f(\bar{x})}(\bar{v}),$$

where the second equality is due to convexity of $\partial f(\bar{x}) = \widehat{\partial} f(\bar{x})$. This proves (b) and completes the proof. □

We proceed with a consequence of subdifferential regularity and strict proto-differentiability, which is important for our characterization of the latter concept for $\mathcal{C}^2$-decomposable functions in Section 5.

Proposition 3.4. *Assume that $f : \mathbf{X} \to \overline{\mathbb{R}}$ is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ and that ∂f is strictly proto-differentiable at $\bar{x}$ for $\bar{v}$. Assume further that f is subdifferentially regular at $\bar{x}$ and that $\operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v}) = K_f(\bar{x}, \bar{v})$. Then, we have $\bar{v} \in \operatorname{ri} \partial f(\bar{x})$.*

Proof. It follows from Proposition 3.3 that $K_f(\bar{x}, \bar{v}) = N_{\partial f(\bar{x})}(\bar{v})$. By Proposition 3.2(b), $\operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v}) = K_f(\bar{x}, \bar{v})$ must be a linear subspace of $\mathbf{X}$. It also results from subdifferential regularity of f at $\bar{x}$ that $\partial f(\bar{x}) = \widehat{\partial} f(\bar{x})$, which tells us that $\widehat{\partial} f(\bar{x})$ is convex. Thus, it follows from a well-known fact from convex analysis (cf. [32, Proposition 2.51]) that $K_f(\bar{x}, \bar{v})$ being a linear subspace is equivalent to $\bar{v} \in \operatorname{ri} \partial f(\bar{x})$. This completes the proof. □

Note that by Proposition 3.3(b), one may expect to assume the condition $\operatorname{cl}\big(\operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v})\big) = K_f(\bar{x}, \bar{v})$ instead of $\operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v}) = K_f(\bar{x}, \bar{v})$ in the result above. They are, however, equivalent in the setting of Proposition 3.4, since $\operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v})$ must be a linear subspace.

Proposition 3.4 tells us that strict proto-differentiability of subgradient mappings necessitates that in the presence of subdifferential regularity, the subgradient under consideration must satisfy a certain relative interior condition. We recently demonstrated in [14, 15] that the latter condition is indeed equivalent to strict proto-differentiability of subgradient mappings for polyhedral functions and certain composite functions. Using Proposition 3.4, we will justify a similar characterization for $\mathcal{C}^2$-decomposable functions in Section 5.

While the assumptions in Proposition 3.4 hold for many classes of functions, important for applications to constrained and composite optimization, the assumption about the domain of the second subderivative requires more elaboration. In the framework of Proposition 3.4, it follows from Proposition 3.3 that $\operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v}) \subset K_f(\bar{x}, \bar{v})$. Equality in the latter inclusion was studied in [29, Proposition 3.4]. To present it, take $w \in \mathbf{X}$ with $\mathrm{d}f(\bar{x})(w)$ finite and define the parabolic subderivative of f at $\bar{x}$ for w with respect to z by

$$\mathrm{d}^2 f(\bar{x})(w \,|\, z) = \liminf_{\substack{t \searrow 0 \\ z' \to z}} \frac{f(\bar{x} + tw + \frac{1}{2}t^2 z') - f(\bar{x}) - t\mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t^2}.$$

It was shown in [29, Proposition 3.4] that the condition $\operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \,|\, \cdot) \neq \emptyset$ for all $w \in K_f(\bar{x}, \bar{v})$ yields $\operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v}) = K_f(\bar{x}, \bar{v})$. It is worth mentioning that the latter condition imposed on $\operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \,|\, \cdot)$ is satisfied whenever f is parabolically epi-differentiable at $\bar{x}$ for any $w \in K_f(\bar{x}, \bar{v})$ in the sense of [50, Definition 13.59]. Note that parabolic epi-differentiability has been studied extensively in [27, 29, 30] and holds for many important classes of functions including convex piecewise linear-quadratic functions (cf. [50, Theorem 13.67]), $\mathcal{C}^2$-decomposable functions (cf. [16, Theorem 6.2]), and spectral functions (cf. [30, Theorem 4.7]).

Corollary 3.5. *Assume that $f : \mathbf{X} \to \overline{\mathbb{R}}$ is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ and that ∂f is strictly proto-differentiable at $\bar{x}$ for $\bar{v}$. Assume further that f is subdifferentially regular at $\bar{x}$ and that $\operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \mid \cdot) \neq \emptyset$ for any $w \in K_f(\bar{x}, \bar{v})$, then one has $\bar{v} \in \operatorname{ri} \partial f(\bar{x})$.*

Proof. This results from the discussion above, Proposition 3.4, and [29, Proposition 3.4]. $\square$

Proposition 3.6. *Assume that $f : \mathbf{X} \to \overline{\mathbb{R}}$ is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ and that ∂f is proto-differentiable at $\bar{x}$ for $\bar{v}$. Assume further that f is subdifferentially regular at $\bar{x}$ and that $\operatorname{cl}\big(\operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v})\big) = K_f(\bar{x}, \bar{v})$. Then, we have*

$$D(\partial f)(\bar{x}, \bar{v})(0) = N_{K_f(\bar{x},\bar{v})}(0) = K_f(\bar{x}, \bar{v})^*.$$

Proof. Since ∂f is proto-differentiable at $\bar{x}$ for $\bar{v}$, it follows from [50, Theorem 13.40] that $D(\partial f)(\bar{x}, \bar{v})(0) = \partial(\frac{1}{2}\mathrm{d}^2 f(\bar{x}, \bar{v}))(0)$. Using a similar argument as the one in the proof of Proposition 3.2(b), we find $\rho \geq 0$ such that the function φ, defined by $\varphi(w) = \mathrm{d}^2 f(\bar{x}, \bar{v})(w) + \rho\|w\|^2$ for $w \in \mathbf{X}$, is a proper convex function with $\varphi(0) = 0$. Moreover, by [29, Proposition 2.1(iii)], we know that $\mathrm{d}^2 f(\bar{x}, \bar{v})(w) \geq -r\|w\|^2$ for any $w \in \mathbf{X}$, where r is taken from (2.3). Choosing ρ sufficiently large, we can then assume without loss of generality that $\varphi(w) \geq 0$ for any $w \in \mathbf{X}$. It follows from the definition of φ that $\operatorname{dom} \varphi = \operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v})$. Take now $\eta \in D(\partial f)(\bar{x}, \bar{v})(0)$ and conclude from $\partial(\frac{1}{2}\varphi)(0) = \partial(\frac{1}{2}\mathrm{d}^2 f(\bar{x}, \bar{v}))(0)$ that $\eta \in \partial(\frac{1}{2}\varphi)(0)$. Since φ is convex, we obtain for any $w \in \operatorname{dom} \varphi$ and any $t > 0$ that

$$\langle \eta, tw - 0\rangle \leq \tfrac{1}{2}\varphi(tw) - \tfrac{1}{2}\varphi(0) = \tfrac{1}{2}\mathrm{d}^2 f(\bar{x}, \bar{v})(tw) + \tfrac{1}{2}\rho\|tw\|^2 = t^2\big(\tfrac{1}{2}\mathrm{d}^2 f(\bar{x}, \bar{v})(w) + \tfrac{1}{2}\rho\|w\|^2\big),$$

where the last equality results from the fact the second subderivative is positive homogeneous of degree 2. Dividing both sides by t and letting then $t \to 0$ bring us to $\langle \eta, w\rangle \leq 0$ for any $w \in \operatorname{dom} \varphi$. Since $\operatorname{cl}(\operatorname{dom} \varphi) = K_f(\bar{x}, \bar{v})$, we arrive at $\eta \in K_f(\bar{x}, \bar{v})^*$.

To justify the opposite inclusion, take $\eta \in K_f(\bar{x}, \bar{v})^*$ and observe that

$$\langle \eta, w - 0\rangle \leq 0 \leq \tfrac{1}{2}\varphi(w) = \tfrac{1}{2}\varphi(w) - \tfrac{1}{2}\varphi(0), \quad \text{for all } \ w \in K_f(\bar{x}, \bar{v}).$$

This confirms that $\eta \in \partial(\frac{1}{2}\varphi)(0) = \partial(\frac{1}{2}\mathrm{d}^2 f(\bar{x}, \bar{v}))(0)$ and hence $\eta \in D(\partial f)(\bar{x}, \bar{v})(0)$, which completes the proof. $\square$

Recall that a set-valued mapping $F : \mathbf{X} \rightrightarrows \mathbf{Y}$ is said to be metrically subregular at x for $y \in F(x)$ if there are a constant $\ell \geq 0$ and a neighborhood U of x such that the estimate $\operatorname{dist}(x', F^{-1}(y)) \leq \ell \operatorname{dist}(y, F(x'))$ holds for any $x' \in U$. Below, we show that strict proto-differentiability of subgradient mappings can ensure metric subregularity of those mappings under certain assumptions.

Theorem 3.7. *Assume that $f : \mathbf{X} \to \overline{\mathbb{R}}$ is a proper lsc convex and $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$ and that ∂f is strictly proto-differentiable at $\bar{x}$ for $\bar{v}$. If $\operatorname{dom} \mathrm{d}^2 f^*(\bar{v}, \bar{x}) = K_{f^*}(\bar{v}, \bar{x})$, then the subgradient mapping ∂f is metrically subregular at $\bar{x}$ for $\bar{v}$.*

Proof. Suppose on the contrary that ∂f is not metrically subregular at $\bar{x}$ for $\bar{v}$. Thus, we find a sequence $x^k \to \bar{x}$ as $k \to \infty$ such that

$$k\operatorname{dist}(\bar{v}, \partial f(x^k)) < \operatorname{dist}(x^k, (\partial f)^{-1}(\bar{v})) = \operatorname{dist}(x^k, \partial f^*(\bar{v})),$$

where the last equality results from the fact that $(\partial f)^{-1} = \partial f^*$ (cf. [50, Proposition 11.3]). Since both sets $\partial f(x^k)$ and $\partial f^*(\bar{v})$ are closed and convex, we find $v^k = P_{\partial f(x^k)}(\bar{v})$ and $\tilde{x}^k = P_{\partial f^*(\bar{v})}(x^k)$ such that $\operatorname{dist}(\bar{v}, \partial f(x^k)) = \|v^k - \bar{v}\|$ and $\operatorname{dist}(x^k, \partial f^*(\bar{v})) = \|x^k - \tilde{x}^k\|$, respectively. Set $t_k := \|x^k - \tilde{x}^k\|$ and observe from the inequality above that $v^k - \bar{v} = o(t_k)$. Passing to a subsequence, if necessary, we can assume that $\{(x^k - \tilde{x}^k)/t_k\}_{k\in\mathbb{N}}$ converges to some $u \in \mathbf{X}$ with $\|u\| = 1$. Since

$$(\tilde{x}^k + t_k\big((x^k - \tilde{x}^k)/t_k\big), \bar{v} + t_k\big((v^k - \bar{v})/t_k\big) = (x^k, v^k) \in \operatorname{gph}\partial f,$$

we obtain $0 \in \widetilde{D}(\partial f)(\bar{x}, \bar{v})(u)$ via the definition of the strict graphical derivative. We have $\widetilde{D}(\partial f)(\bar{x}, \bar{v})(u) = D(\partial f)(\bar{x}, \bar{v})(u)$ by strict proto-differentiability of ∂f at $\bar{x}$ for $\bar{v}$, and therefore $0 \in D(\partial f)(\bar{x}, \bar{v})(u)$. The latter amounts to

$$u \in D(\partial f)^{-1}(\bar{v}, \bar{x})(0) = D(\partial f^*)(\bar{v}, \bar{x})(0).$$

On the other hand, strict proto-differentiability of ∂f at $\bar{x}$ for $\bar{v}$ is equivalent to that of ∂f^* at $\bar{v}$ for $\bar{x}$. Since f^* is convex, it is subdifferentially regular at $\bar{v}$. Combining these and Proposition 3.6 leads us to

$$u \in D(\partial f^*)(\bar{v}, \bar{x})(0) = K_{f^*}(\bar{v}, \bar{x})^*. \tag{3.3}$$

Moreover, it follows from $\tilde{x}^k = P_{\partial f^*(\bar{v})}(x^k)$ that $(x^k - \tilde{x}^k)/t_k \in N_{\partial f^*(\bar{v})}(\tilde{x}^k)$, which implies via Proposition 3.3(b) that

$$u \in N_{\partial f^*(\bar{v})}(\bar{x}) = K_{f^*}(\bar{v}, \bar{x}).$$

This, coupled with (3.3), results in $u = 0$, a contradiction, and completes the proof. □

Strict proto-differentiability, assumed in Theorem 3.7, will be characterized for $\mathcal{C}^2$-decomposable functions in Section 5. Below, we provide an example of a class of functions for which the assumption on the domain of the second subderivative in Theorem 3.7 is always satisfied.

Example 3.8. Let $\mathbf{S^n}$ stand for the space of all real $n \times n$ symmetric matrices equipped with the inner product

$$\langle X, Y \rangle = \operatorname{tr}(XY), \quad X, Y \in \mathbf{S^n}.$$

The induced Frobenius norm of $X \in \mathbf{S^n}$ is defined by $\|X\| = \sqrt{\operatorname{tr}(X^2)}$. Recall that $f : \mathbf{S^n} \to \overline{\mathbb{R}}$ is a spectral function if it is orthogonally invariant, meaning that for any $X \in \mathbf{S^n}$ and any $n \times n$ orthogonal matrix U, we have $f(X) = f(U^\top X U)$. It follows from [22, Proposition 4] that any spectral function f can be expressed in the composite form

$$f(X) := (\theta \circ \lambda)(X), \quad X \in \mathbf{S^n},$$

where $\theta : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a permutation-invariant function on $\mathbb{R}^n$, called symmetric, and λ is the mapping assigning to each matrix $X \in \mathbf{S^n}$ the vector $\big(\lambda_1(X), \ldots, \lambda_n(X)\big)$ of its eigenvalues arranged in nonincreasing order. Suppose that the symmetric function $\theta : \mathbb{R}^n \to \overline{\mathbb{R}}$ is polyhedral, meaning $\operatorname{epi}\theta$ is a polyhedral convex set. Consider the spectral function $f = \theta \circ \lambda$. This selection of θ allows to covers important examples of eigenvalue functions such as the maximum eigenvalue function and the sum of the first k ($1 \le k \le n$) largest eigenvalues of a matrix. It follows from [4, Theorem 5.2.2] that $f^* = \theta^* \circ \lambda$. According to [50, Theorem 11.14], θ^* is polyhedral. Appealing now to [30, Corollary 5.8] tells us that if $(X, Y) \in \operatorname{gph}\partial f$, then we always have $\operatorname{dom} \mathrm{d}^2 f^*(Y, X) = K_{f^*}(Y, X)$.

We continue with a relationship between graphical derivative and coderivative of subgradient mappings of strictly proto-differentiable functions. Recall that for a function $f : \mathbf{X} \to \overline{\mathbb{R}}$ with $(\bar{x}, \bar{v}) \in \operatorname{gph}\partial f$, the coderivative mapping of ∂f at $\bar{x}$ for $\bar{v}$, denoted $D^*(\partial f)(\bar{x}, \bar{v})$, is defined by

$$\eta \in D^*(\partial f)(\bar{x}, \bar{v})(w) \iff (\eta, -w) \in N_{\operatorname{gph}\partial f}(\bar{x}, \bar{v}). \tag{3.4}$$

Theorem 3.9. *Assume that $f : \mathbf{X} \to \overline{\mathbb{R}}$ is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ and that ∂f is strictly proto-differentiable at $\bar{x}$ for $\bar{v}$. Then, one has*

$$D(\partial f)(\bar{x}, \bar{v})(w) = D^*(\partial f)(\bar{x}, \bar{v})(w) = A(w) + N_{\overline{K}}(w) \quad \text{for all } w \in \mathbf{X}, \tag{3.5}$$

where $\overline{K}$ and A are taken from Proposition 3.2(b).

Proof. Strict proto-differentiability of ∂f at $\bar{x}$ for $\bar{v}$ implies proto-differentiability of ∂f at $\bar{x}$ for $\bar{v}$, which together with [50, Theorem 13.40] illustrates that f is twice epi-differentiable at $\bar{x}$ for $\bar{v}$. By [50, Theorem 13.57], we get the inclusion $D(\partial f)(\bar{x},\bar{v})(w) \subset D^*(\partial f)(\bar{x},\bar{v})(w)$. To obtain the opposite inclusion, take $\eta \in D^*(\partial f)(\bar{x},\bar{v})(w)$, which means that $(\eta,-w) \in N_{\mathrm{gph}\,\partial f}(\bar{x},\bar{v})$. By Proposition 3.2(c), we have $(\eta,-w) \in \widehat{N}_{\mathrm{gph}\,\partial f}(\bar{x},\bar{v})$. It also follows from Proposition 3.2(b) that $\mathrm{d}^2 f(\bar{x},\bar{v})$ is a generalized quadratic form on $\mathbf{X}$ given by (3.2). This representation brings us to

$$D(\partial f)(\bar{x},\bar{v})(w) = \partial\big(\tfrac{1}{2}\mathrm{d}^2 f(\bar{x},\bar{v})\big)(w) = A(w) + N_{\overline{K}}(w) \quad \text{for all } \ w \in \mathbf{X}, \tag{3.6}$$

where the first equality results from [50, Theorem 13.40]. This clearly confirms that

$$T_{\mathrm{gph}\,\partial f}(\bar{x},\bar{v}) = \mathrm{gph}\, D(\partial f)(\bar{x},\bar{v}) = \big\{(w,\eta)\,\big|\, w \in \overline{K},\ \eta - A(w) \in \overline{K}^{\perp}\big\},$$

which leads us via [50, Corollary 11.25(d)] to

$$\widehat{N}_{\mathrm{gph}\,\partial f}(\bar{x},\bar{v}) = \big(T_{\mathrm{gph}\,\partial f}(\bar{x},\bar{v})\big)^* = \big\{(\xi - A(\nu),\nu)\,\big|\,(\xi,\nu) \in \overline{K}^{\perp} \times \overline{K}\big\}.$$

Since $(\eta,-w) \in \widehat{N}_{\mathrm{gph}\,\partial f}(\bar{x},\bar{v})$, there exists $(\xi,\nu) \in \overline{K}^{\perp} \times \overline{K}$ such that

$$\xi - A(\nu) = \eta \quad \text{and} \quad \nu = -w.$$

These equations imply that $\eta - A(w) = \xi \in \overline{K}^{\perp}$ and $w \in \overline{K}$, which, together with the fact that $\overline{K}$ is a linear subspace of $\mathbf{X}$ and (3.6), shows that $\eta \in D(\partial f)(\bar{x},\bar{v})(w)$, as desired, and completes the proof. $\square$

As mentioned in the proof of Theorem 3.9, the inclusion $D(\partial f)(\bar{x},\bar{v})(w) \subset D^*(\partial f)(\bar{x},\bar{v})(w)$ always holds under proto-differentiability of ∂f; see [50, Theorem 13.57]. It is also not hard to see that the latter inclusion can be strict; consider the function $f(x) = x^2/2$ for $x \geq 0$ and $f(x) = -x^2/2$ for $x \leq 0$. Thus, equality requires to impose more assumptions.

We recently characterized the relationship in (3.5) via a relative interior condition for polyhedral functions in [14, Corollary 3.7] and for certain composite functions in [15, Theorem 3.12]. Moreover, it was shown therein that such a condition is equivalent to strict proto-differentiability of subgradient mappings under consideration. Whether or not a similar result can be obtained in the general setting of Theorem 3.9 remains as an open question. We should also mention that it was shown in [24, Theorem 4] that the same equality in (3.6) holds for $\mathcal{C}^2$-partly smooth functions in the sense of [23, Definition 2.7]. We should add that $\mathcal{C}^2$-partly smooth functions are always strictly proto-differentiable and so [24, Theorem 4] can be covered by Theorem 3.9. The proof of the latter result is beyond the scope of this paper and will be appeared in our forthcoming paper [17].

We close this section by showing that our observation in (3.5) has an interesting consequence about local minimizers of a function. To present it, we have to recall the concept of tilt-stable local minimizers of a function. Given a function $f : \mathbf{X} \to \overline{\mathbb{R}}$ and $\bar{x} \in \mathbf{X}$ with $f(\bar{x})$ finite, we say that $\bar{x}$ is a tilt-stable local minimizer of f if there exist neighborhoods U of $\bar{x}$ and V of $\bar{v} = 0$ such that the optimal solution mapping

$$v \mapsto \underset{x \in U}{\mathrm{argmin}}\{f(x) - \langle v, x - \bar{x}\rangle\}$$

is single-valued and Lipschitz continuous on V and its value at $\bar{v}$ is $\{\bar{x}\}$. Tilt-stability was introduced in [42] and has been studied extensively for various classes of constrained and composite optimization problems; see [33, 42]. According to [33, Theorem 3.2], tilt-stability of a local minimizer $\bar{x}$ of f yields the quadratic growth condition of f at $\bar{x}$: there exist a constant $\ell \geq 0$ and a neighborhood U of $\bar{x}$ such that

$$f(x) \geq f(\bar{x}) + \frac{\ell}{2}\|x - \bar{x}\|^2 \quad \text{for all } \ x \in U.$$

Below, we are going to show that under strict proto-differentiability of subgradient mappings the opposite implication is also valid.

Corollary 3.10. *Assume that $f : \mathbf{X} \to \overline{\mathbb{R}}$ is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v} := 0 \in \partial f(\bar{x})$ and that ∂f is strictly proto-differentiable at $\bar{x}$ for $\bar{v}$. Then, the following properties are equivalent:*

(a) *$\bar{x}$ is a tilt-stable local minimizer of f;*
(b) *f enjoys the quadratic growth condition at $\bar{x}$;*
(c) *for any nonzero $w \in \operatorname{dom} D(\partial f)(\bar{x}, \bar{v})$ and any $\eta \in D(\partial f)(\bar{x}, \bar{v})(w)$, one has $\langle w, \eta\rangle > 0$.*

Proof. Strict proto-differentiability of ∂f at $\bar{x}$ for $\bar{v}$ clearly implies proto-differentiability of ∂f at $\bar{x}$ for $\bar{v}$, which together with [50, Theorem 13.40] illustrates that f is twice epi-differentiable at $\bar{x}$ for $\bar{v}$. The equivalence of (b) and (c) was established in [5, Theorem 3.7]. It follows from [42, Theorem 1.3] that (a) is equivalent to the condition $\langle w, \eta\rangle > 0$ for any $w \in \big(\operatorname{dom} D^*(\partial f)(\bar{x}, \bar{v})\big) \setminus \{0\}$ and any $\eta \in D^*(\partial f)(\bar{x}, \bar{v})(w)$. Combining this with Theorem 3.9 tells us that (a) and (c) are equivalent, which completes the proof. □

Using a different approach, Lewis and Zhang demonstrated in [25, Theorem 6.3] the equivalence of the properties in Corollary 3.10(a) and (b) for $\mathcal{C}^2$-partly smooth functions. As pointed out earlier, the latter functions are strictly proto-differentiable, which allows us to obtain their result using our established theory in this section.

4 Regularity Properties of Generalized Equations

In this section, we aim to study important stability properties of solution mappings to a class of generalized equations under the strict proto-differentiability assumption. As Section 3 may suggest, we should expect stronger properties to be satisfied under the latter assumption. We begin with an important question about the relationship between two important stability properties of solution mappings of generalized equations, namely metric regularity and strong metric regularity. Given a differentiable mapping $\psi : \mathbf{X} \to \mathbf{X}$ and a proper function $f : \mathbf{X} \to \overline{\mathbb{R}}$, we mainly focus on the generalized equation

$$0 \in \psi(x) + \partial f(x), \tag{4.1}$$

to which we associate the set-valued mapping $G : \mathbf{X} \rightrightarrows \mathbf{X}$, defined by

$$G(x) = \psi(x) + \partial f(x), \quad x \in \mathbf{X}, \tag{4.2}$$

and the solution mapping $S : \mathbf{X} \rightrightarrows \mathbf{X}$, defined by

$$S(y) := G^{-1}(y) = \big\{x \in \mathbf{X} \,\big|\, y \in \psi(x) + \partial f(x)\big\}, \quad y \in \mathbf{X}. \tag{4.3}$$

We will assume further that the function f in (4.1) is prox-regular at the point under consideration. Recall that a set-valued mapping $F : \mathbf{X} \rightrightarrows \mathbf{Y}$ is said to be *metrically regular* at $\bar{x}$ for $\bar{y} \in F(\bar{x})$ if there exist a positive constant κ and neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that the estimate

$$\operatorname{dist}\big(x, F^{-1}(y)\big) \le \kappa \operatorname{dist}\big(y, F(x)\big) \tag{4.4}$$

holds for all $(x, y) \in U \times V$. When the estimate (4.4) holds for any $(x, y) \in \mathbf{X} \times \mathbf{Y}$, we say that F is *globally* metrically regular. The mapping is said to be *strongly metrically regular* at $\bar{x}$ for $\bar{y}$ if F^{-1} admits a Lipschitz continuous single-valued localization around $\bar{y}$ for $\bar{x}$, which means that there exist neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that the mapping $y \mapsto F^{-1}(y) \cap U$ is single-valued and Lipschitz continuous on V. According to [9, Proposition 3G.1], strong metric regularity of F at $\bar{x}$ for $\bar{y}$ amounts to F being metrically regular at $\bar{x}$ for $\bar{y}$ and the inverse mapping F^{-1} admitting a single-valued localization around $\bar{y}$ for $\bar{x}$.

When $f = \delta_C$ in (4.1) with C being a polyhedral convex set, the generalized equation (4.1) presents an example of variational inequalities. In this case, the seminal paper [8] revealed for the first time that strong metric regularity and metric regularity of the solution mapping S in (4.3) are equivalent. In this section, we are going to justify the same equivalence for the generalized equation in (4.1) when ∂f is strictly proto-differentiable.

While metric regularity of the mapping G from (4.2) can be studied using the coderivative criterion (cf. [31, Theorem 3.3(ii)]) in general, we show below that in the presence of strict proto-differetiability, such a characterization has simpler forms.

Proposition 4.1 (point-based criteria for metric regularity). *Assume that $\bar{x}$ is a solution to the generalized equation in (4.1) in which ψ is strictly differentiable at $\bar{x}$ and f is both prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v} := -\psi(\bar{x})$ and that ∂f is strictly proto-differentiable at the point $\bar{x}$ for $\bar{v}$. Take the linear subspace $\overline{K}$ and the linear operator A from (3.5). Then the following properties are equivalent:*

(a) *the mapping G from (4.2) is metrically regular at $\bar{x}$ for 0;*

(b) $\{w \in \mathbf{X} \,|\, (\nabla\psi(\bar{x}) + A)^*(w) \in \overline{K}^{\perp}\} \cap \overline{K} = \{0\}$*;*

(c) $(\nabla\psi(\bar{x}) + A)(\overline{K}) + \overline{K}^{\perp} = \mathbf{X}$*;*

(d) $|DG(\bar{x}, 0)^{-1}|^{-}$ *is finite, where the inner norm* $|DG(\bar{x}, 0)^{-1}|^{-}$ *is defined by*

$$|DG(\bar{x}, 0)^{-1}|^{-} := \sup_{\|u\|\le 1} \inf_{w\in DG(\bar{x},0)^{-1}(u)} \|w\|$$

with convention $\inf_{w\in\emptyset} \|w\| = \infty$*;*

(e) $DG(\bar{x}, 0)$ *is surjective, meaning that*

$$\operatorname{rge} DG(\bar{x}, 0) := \{u \in \mathbf{X} |\, \exists\, w \in \mathbf{X} \ \textit{with} \ u \in DG(\bar{x}, 0)(w)\} = \mathbf{X};$$

(f) $DG(\bar{x}, 0)$ *is globally metrically regular.*

Proof. To prove the equivalence of (a) and (b), we begin with calculating $D^*G(\bar{x}, 0)$. Applying the sum rule for the coderivative from [50, Exercise 10.43(b)] to $G = \psi + \partial f$ and using the formula for $D^*(\partial f)(\bar{x}, \bar{v})$ in (3.5), we obtain for all $w \in \mathbf{X}$ that

$$D^*G(\bar{x}, 0)(w) = \nabla\psi(\bar{x})^*(w) + D^*(\partial f)(\bar{x}, \bar{v})(w) = (\nabla\psi(\bar{x}) + A)^*(w) + N_{\overline{K}}(w). \tag{4.5}$$

Since $\overline{K}$ is a linear subspace of $\mathbf{X}$, we then deduce from the above calculation that

$$D^*G(\bar{x}, 0)^{-1}(0) := \{w \in \mathbf{X} \,|\, 0 \in D^*G(\bar{x}, 0)(w)\} = \{w \in \mathbf{X} \,|\, (\nabla\psi(\bar{x}) + A)^*(w) \in \overline{K}^{\perp}\} \cap \overline{K}.$$

By [31, Theorem 3.3(ii)], metric regularity of G at $\bar{x}$ for 0 amounts to $D^*G(\bar{x}, 0)^{-1}(0) = \{0\}$. Combining it with the discussion above yields the equivalence of (a) and (b).

Turing to the equivalence of (b) and (c), observe first via [50, Corollary 11.25(c)] that

$$\{w \in \mathbf{X} \,|\, (\nabla\psi(\bar{x}) + A)^*(w) \in \overline{K}^{\perp}\}^{\perp} = (\nabla\psi(\bar{x}) + A)(\overline{K}).$$

Taking the orthogonal complements of both sides of either (b) or (c) and using the above equality tell us that (b) and (c) are equivalent.

To prove the equivalence of (c) and (e), we proceed with calculating $DG(\bar{x}, 0)$. Since ψ is differentiable at $\bar{x}$, the sum rule for the graphical derivative from [50, Exercise 10.43(a)], together with (3.5), gives us

$$DG(\bar{x}, 0)(w) = \nabla\psi(\bar{x})(w) + D(\partial f)(\bar{x}, \bar{v})(w) = (\nabla\psi(\bar{x}) + A)(w) + N_{\overline{K}}(w), \quad w \in \mathbf{X}. \tag{4.6}$$

Thus, we get

$$\operatorname{rge} DG(\bar{x},0) = (\nabla\psi(\bar{x}) + A)(\overline{K}) + \overline{K}^{\perp}, \tag{4.7}$$

which clearly implies that $\operatorname{rge} DG(\bar{x},0) = \mathbf{X}$ if and only if (c) holds. This verifies the equivalence of (c) and (e).

According to (4.6), $\operatorname{gph} DG(\bar{x},0)$ is closed and convex due to $\overline{K}$ being a linear subspace of $\mathbf{X}$. It follows then from [9, Proposition 4A.6] that $|DG(\bar{x},0)^{-1}|^{-} < \infty$ amounts to $DG(\bar{x},0)$ being surjective, which proves the equivalence of (d) and (e). Finally, observe that (e) amounts to the condition $0 \in \operatorname{int}\big(\operatorname{rge} DG(\bar{x},0)\big)$ due to the fact that $\operatorname{rge} DG(\bar{x},0)$ is indeed a linear subspace of $\mathbf{X}$; see (4.7). This, combined with $0 \in DG(\bar{x},0)(0)$ and [9, Theorem 5B.4], implies that (e) is equivalent to the mapping $DG(\bar{x},0)$ being metrically regular at $0 \in \mathbf{X}$ for $0 \in \mathbf{X}$. According to [19, Theorem 5.9(a)], the latter amounts to (f), confirming that (e) and (f) are equivalent. This completes the proof. $\square$

To proceed, we recall the following sufficient conditions for strong metric regularity of a set-valued mapping, taken from [9, Theorem 4D.1]: a set-valued mapping $F : \mathbf{X} \rightrightarrows \mathbf{Y}$ is strongly metrically regular at $\bar{x}$ for $\bar{y} \in F(\bar{x})$ provided that its graph is locally closed at $(\bar{x},\bar{y})$ and the conditions

$$0 \in \widetilde{D}F(\bar{x},\bar{y})(w) \implies w = 0 \tag{4.8}$$

and

$$\bar{x} \in \liminf_{y \to \bar{y}} F^{-1}(y) \tag{4.9}$$

hold.

Theorem 4.2. *Assume that $\bar{x}$ is a solution to the generalized equation in (4.1) in which ψ is strictly differentiable at $\bar{x}$ and f is both prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v} := -\psi(\bar{x})$. Then the following properties are equivalent:*

(a) *the mapping G, taken from (4.2), is metrically regular at $\bar{x}$ for 0 and ∂f is strictly proto-differentiable at $\bar{x}$ for $\bar{v}$;*

(b) *the solution mapping S from (4.3) has a Lipschitz continuous single-valued localization s around $0 \in \mathbf{X}$ for $\bar{x}$, which is strictly differentiable at 0.*

Moreover, if (a) holds, then the derivative of s at 0 can be calculated by

$$\nabla s(0) = \big((\nabla\psi(\bar{x}) + A)|_{\overline{K}}\big)^{-1} \circ P_{\overline{K}}, \tag{4.10}$$

where the linear mapping A and the linear subspace $\overline{K}$ are taken from (3.5), where $P_{\overline{K}} : \mathbf{X} \to \mathbf{X}$ is the projection mapping onto $\overline{K}$, and where $(\nabla\psi(\bar{x}) + A)|_{\overline{K}}$ stands for the restriction of the linear mapping $\nabla\psi(\bar{x}) + A$ to $\overline{K}$.

Proof. Suppose that (a) holds. We begin by proving that G is strongly metrically regular at $\bar{x}$ for 0 by validating (4.8) and (4.9). To this end, it can be seen via the definition of G in (4.2) that the graph of G is closed. Moreover, the distance estimate in (4.4), adopted for metric regularity of G at $\bar{x}$ for 0, clearly yields (4.9). It remains to justify (4.8). Making use of strict proto-differentiability of ∂f at $\bar{x}$ for $\bar{v}$, we conclude via [14, Proposition 5.3] that G is strictly proto-differetiable at $\bar{x}$ for 0, which results in $\widetilde{D}G(\bar{x},0)(w) = DG(\bar{x},0)(w)$ for any $w \in \mathbf{X}$. Also, we know that $DG(\bar{x},0)$ enjoys the representation in (4.6). Moreover, we deduce from metric regularity of G at $\bar{x}$ for 0 and Proposition 4.1(c) that the condition

$$(\nabla\psi(\bar{x}) + A)(\overline{K}) + \overline{K}^{\perp} = \mathbf{X} \tag{4.11}$$

is satisfied. On the other hand, it is not hard to see that $\dim\big((\nabla\psi(\bar{x}) + A)(\overline{K})\big) \le \dim \overline{K}$. We claim that $\big((\nabla\psi(\bar{x})+A)(\overline{K})\big)\cap\overline{K}^{\perp} = \{0\}$. If not, we would get $\dim\Big(\big((\nabla\psi(\bar{x})+A)(\overline{K})\big)\cap\overline{K}^{\perp}\Big) >$

0. This, together with (4.11), would lead us to

$$\dim \mathbf{X} = \dim\big((\nabla\psi(\bar{x})+A)(\overline{K})\big) + \dim \overline{K} - \dim\big(((\nabla\psi(\bar{x})+A)(\overline{K})) \cap \overline{K}^{\perp}\big)$$
$$< \dim \overline{K} + \dim \overline{K}^{\perp} = \dim \mathbf{X},$$

a contradiction. The above claim implies that

$$(\nabla\psi(\bar{x})+A)(\overline{K}) \subset \overline{K}, \tag{4.12}$$

which in turn demonstrates that $H := (\nabla\psi(\bar{x})+A)|_{\overline{K}}$, the restriction of the linear mapping $\nabla\psi(\bar{x})+A$ to $\overline{K}$, is a linear mapping from $\overline{K}$ into $\overline{K}$. We claim now that H is a bijection from $\overline{K}$ onto $\overline{K}$. To this end, it suffices to show that H is surjective due to the rank–nullity theorem from linear algebra. If H were not surjective, we would get $\dim\big(H(\overline{K})\big) < \dim \overline{K}$. This would yield

$$\dim\big(H(\overline{K}) + \overline{K}^{\perp}\big) \le \dim\big(H(\overline{K})\big) + \dim \overline{K}^{\perp} < \dim \overline{K} + \dim \overline{K}^{\perp} = \dim \mathbf{X},$$

which contradicts (4.11) and hence confirms that H is a bijection from $\overline{K}$ onto $\overline{K}$. Turning to the proof the implication in (4.8) for G, suppose that $0 \in \widetilde{D}G(\bar{x},0)(w) = DG(\bar{x},0)(w)$. We then get from (4.6) that $w \in \overline{K}$ and $-(\nabla\psi(\bar{x})+A)(w) \in \overline{K}^{\perp}$. These indicate that

$$(\nabla\psi(\bar{x})+A)(w) \in \big((\nabla\psi(\bar{x})+A)(\overline{K})\big) \cap \overline{K}^{\perp} = \overline{K} \cap \overline{K}^{\perp} = \{0\},$$

which in turn leads us to $H(w) = 0$. Since H is a bijection, we get $w = 0$, proving (4.8) for G, which confirms that G is strongly metrically regular at $\bar{x}$ for 0. This amounts to saying that $S = G^{-1}$ has a Lipschitz continuous single-valued localization around 0 for $\bar{x}$. So, we find neighborhoods U of $\bar{x}$ and V of 0 such that the mapping $y \mapsto S(y) \cap U$ is single-valued and Lipschitz continuous on V. Define the mapping $s : V \to U$ by $s(y) = S(y) \cap U$ for $y \in V$. We then conclude that

$$\operatorname{gph} s = \operatorname{gph} S \cap (V \times U), \tag{4.13}$$

which clearly yields $T_{\operatorname{gph} s}(0,\bar{x}) = T_{\operatorname{gph} S}(0,\bar{x})$. Recall that G is strictly proto-differentiable at $\bar{x}$ for 0. It is worth mentioning that strict proto-differentiability is a geometric property of the graph of the mapping and that $\operatorname{gph} G^{-1}$ and $\operatorname{gph} s$ are the same around $(0,\bar{x})$. Thus, strict proto-differentiability of G^{-1} at 0 for $\bar{x}$ yields that of s at 0 for $\bar{x}$. By Definition 2.2(b), $\operatorname{gph} s$ is strictly smooth at $(0,\bar{x})$. Since s is Lipschitz continuous on V, we deduce from [47, Proposition 3.1] that s is strictly differentiable at 0. To justify the formula for $\nabla s(0)$ in (4.10), take $u \in \mathbf{X}$. We conclude from differentibility of s at 0 that $DG^{-1}(0,\bar{x})(u) = Ds(0)(u) = \nabla s(0)(u)$, which in turn yields $u \in DG(\bar{x},0)(\nabla s(0)(u))$. Appealing now to (4.6), we obtain

$$\nabla s(0)(u) \in \overline{K} \quad \text{and} \quad u - (\nabla\psi(\bar{x})+A)(\nabla s(0)(u)) \in \overline{K}^{\perp}. \tag{4.14}$$

By (4.12), we have $(\nabla\psi(\bar{x})+A)(\nabla s(0)(u)) \in \overline{K}$. Since $P_{\overline{K}}$ is a linear mapping, employing the latter and (4.14) yields

$$P_{\overline{K}}(u) = P_{\overline{K}}\big((\nabla\psi(\bar{x})+A)(\nabla s(0)(u))\big) = (\nabla\psi(\bar{x})+A)(\nabla s(0)(u)).$$

Recalling that $(\nabla\psi(\bar{x})+A)|_{\overline{K}}$ is a bijection and using the latter relationships, we arrive at

$$\nabla s(0)(u) = \big((\nabla\psi(\bar{x})+A)|_{\overline{K}}\big)^{-1}(P_{\overline{K}}(u)).$$

This confirms (4.10) and hence completes the proof of the implication (a) $\Longrightarrow$ (b). To prove the opposite implication, assume that (b) is satisfied. Thus, G is strongly metrically regular at $\bar{x}$ for 0, which clearly shows that it is metrically regular at $\bar{x}$ for 0. Moreover, we find neighborhoods U of $\bar{x}$ and V of 0 for which (4.13) holds. Since s is strictly differentiable and Lipschitz continuous around 0, we deduce from [47, Proposition 3.1] that s is strictly proto-differentiable at 0 for $\bar{x}$. It follows from (4.13) that the solution mapping S is strictly proto-differentiable at 0 for $\bar{x}$. Since $S = G^{-1}$, we conclude that G is strictly proto-differentiable at $\bar{x}$ for 0. Using the sum rule for strict proto-differentiability in [14, Proposition 5.3] and the fact that $G = \psi + \partial f$ tells us that ∂f is strictly proto-differentiable at $\bar{x}$ for $-\psi(\bar{x})$, which proves (a) and completes the proof. ☐

One question that may arise is how one can ensure continuous differentiability ($\mathcal{C}^1$) of the localization s in Theorem 4.2(b), a property that is useful in dealing with a number of applications including the one discussed at the end of this section. A close look into the proof of Theorem 4.2 tells us that if we assume strict proto-differentiability of ∂f at x for v for any pair $(x, v) \in \operatorname{gph} \partial f$ close to $(\bar{x}, -\psi(\bar{x}))$ and if ψ is $\mathcal{C}^1$ in a neighborhood of $\bar{x}$, we can ensure strict proto-differentiability of the mapping G from (4.2) on a neighborhood of $(\bar{x}, 0)$. This implies that the localization s is strictly differentiable in a neighborhood of 0, a property equivalent to $\mathcal{C}^1$-smoothness of s around 0 according to [9, Exercise 1D.8]. This brings us to the following observation.

Theorem 4.3. *Assume that $\bar{x}$ is a solution to the generalized equation in (4.1) in which ψ is $\mathcal{C}^1$ around $\bar{x}$ and f is prox-regular and subdifferentially continuous at $\bar{x}$ for $-\psi(\bar{x})$. Then the following properties are equivalent:*

- (a) *the mapping G, taken from (4.2), is metrically regular at $\bar{x}$ for 0 and ∂f is strictly proto-differentiable at x for v for all $(x, v) \in \operatorname{gph} \partial f$ close to $(\bar{x}, -\psi(\bar{x}))$;*
- (b) *the solution mapping S from (4.3) has a Lipschitz continuous single-valued localization s around $0 \in \mathbf{X}$ for $\bar{x}$, which is $\mathcal{C}^1$ around 0.*

As an immediate consequence of Theorem 4.2, we arrive at the following equivalence of metric regularity and strong metric regularity of G in the presence of a strict proto-differetiability assumption.

Corollary 4.4 (equivalence between metric regularity and strong metric regularity)**.** *Assume that $\bar{x}$ is a solution to the generalized equation in (4.1) in which ψ is strictly differentiable at $\bar{x}$ and f is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v} := -\psi(\bar{x})$ and that ∂f is strictly proto-differentiable at the point $\bar{x}$ for $\bar{v}$. Then, the mapping G, taken from (4.2), is metrically regular at $\bar{x}$ for 0 if and only if it is strongly metrically regular at $\bar{x}$ for 0.*

Proof. This directly falls out of the established equivalence in Theorem 4.2. ☐

As mentioned above, Dontchev and Rockafellar in [8, Theorem 1] derived the equivalence of metric regularity and strong metric regularity of the solution mapping S from (4.3) when $f = \delta_C$, where C is a polyhedral convex subset of $\mathbf{X}$. In such a framework, Corollary 4.4 covers the aforementioned seminal result under the extra assumption of strict proto-differentiability of N_C at $\bar{x}$ for $-\psi(\bar{x})$. Note that it was shown in [14, Theorem 3.5] that the latter condition on N_C amounts to the relative interior condition $-\psi(\bar{x}) \in \operatorname{ri} N_C(\bar{x})$. While not knowing whether or not a similar conclusion can be achieved for prox-regular functions in general, we will demonstrate in the next section that such an observation holds true for $\mathcal{C}^2$-decomposable functions. We should add to this discussion that our proof is fundamentally different from the approach exploited in [8] which relied heavily on Robinson's results in [45] and did not utilize the theory of second-order variational analysis, used in this paper. Note also that when f in (4.1) enjoys a certain composite representation, Corollary 4.4 boils down to our recent result in [16, Theorem 4.3]. Finally, the recent results in [13, Corollary 7.3] presents a characterization of strong metric regularity of set-valued mappings that are graphically Lipschitzian manifolds. It is not clear, however, whether the latter Lipschitzian manifold assumption does hold for the mapping G from (4.2) and hence whether the latter result can be exploited in our setting.

It was conjectured in [10, Conjecture 4.7] that if the proper function $f : \mathbf{X} \to \overline{\mathbb{R}}$ is prox-regular and subdifferentially continuous and $\bar{x}$ is a local minimum of f, then metric regularity and strong metric regularity of ∂f at $\bar{x}$ for 0 are equivalent. First, observe that the condition of $\bar{x}$ being a local minimizer is not essential, since one can choose a $\mathcal{C}^2$ mapping $g : \mathbf{X} \to \mathbb{R}$ and a polyhedral convex set $C \subset \mathbf{X}$ and set $f = g + \delta_C$. Employing [8, Theorem 1] tells us that the desired equivalence of metric regularity and strong metric regularity of ∂f at $\bar{x}$ for 0 holds without demanding that $\bar{x}$ be a local minimum of f. Dropping the latter condition from [10, Conjecture 4.7], one can find in [20, Example BE.4] an example of a $\mathcal{C}^1$ function with Lipschitz continuous derivative (thus prox-regular and subdifferentially continuous by [50, Proposition 13.34]) that is metrically regular but is not strongly metrically regular. This example suggests that the latter equivalence does not hold for prox-regular in general. Corollary 4.4 provides an answer to this question by demonstrating that if, in addition, ∂f is strictly proto-differentiable at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$, we can ensure that metric regularity and strong metric regularity of ∂f at $\bar{x}$ for $\bar{v}$ are equivalent, which confirms the conjecture in [10, Conjecture 4.7] under this extra assumption. While Dontchev and Rockafellar's result in [8, Theorem 1] indicates that strict proto-differentiability is not required for such an equivalence in general, it remains as an open question to proceed if the strictly proto-differentiability condition fails.

When $\nabla\psi$ in the generalized equation in (4.1) enjoys a certain symmetry property, we can improve Corollary 4.4. To do this, recall that a set-valued mapping $F : \mathbf{X} \rightrightarrows \mathbf{Y}$ is called strongly metrically subregular at x for $y \in F(x)$ if there exist a constant $\ell \ge 0$ and a neighborhood U of x such that the estimate $\|x' - x\| \le \ell \, \mathrm{dist}(y, F(x'))$ holds for any $x' \in U$.

Corollary 4.5. *Assume that $\bar{x}$ is a solution to the generalized equation in* (4.1) *in which $\nabla\psi(\bar{x}) = \nabla\psi(\bar{x})^*$ and f is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v} := -\nabla\psi(\bar{x})$ and that ∂f is strictly proto-differentiable at the point $\bar{x}$ for $\bar{v}$. Then the following properties are equivalent:*

(a) *the mapping G, taken from* (4.2)*, is strongly metrically regular at $\bar{x}$ for* 0*;*
(b) *the mapping G is metrically regular at $\bar{x}$ for* 0*;*
(c) *the mapping G is strongly metrically subregular at $\bar{x}$ for* 0.

Proof. The equivalence of (a) and (b) was already established in Corollary 4.4. To prove (b) and (c) are also equivalent, we deduce from [9, Theorem 4E.1] that G is strongly metrically subregular at $\bar{x}$ for 0 if and only if the implication

$$0 \in DG(\bar{x}, 0)(w) = \nabla\psi(\bar{x})w + D(\partial f)(\bar{x}, \bar{v})(w) \implies w = 0$$

holds, where the equality comes from (4.6). By [31, Theorem 3.3(ii)], G is metrically regular at $\bar{x}$ for 0 if and only if $0 \in D^*G(\bar{x}, 0)(w)$ yields $w = 0$. By the first equality in (4.5) and the assumption $\nabla\psi(\bar{x}) = \nabla\psi(\bar{x})^*$, we conclude from (3.5) that

$$D^*G(\bar{x}, 0)(w) = \nabla\psi(\bar{x})^* w + D^*(\partial f)(\bar{x}, \bar{v})(w) = \nabla\psi(\bar{x})w + D(\partial f)(\bar{x}, \bar{v})(w),$$

and hence we arrive at $DG(\bar{x}, 0)(w) = D^*G(\bar{x}, 0)(w)$ for any $w \in \mathbf{X}$. This, coupled with the above characterizations of strong metric subregularity and metric regularity, demonstrates that (b) and (c) are also equivalent and so completes the proof. □

Note that the condition $\nabla\psi(\bar{x}) = \nabla\psi(\bar{x})^*$ in the generalized equation in (4.1) is satisfied for an important instance of generalized equations, namely the KKT system of optimization problems; see Theorem 5.14 for more detail.

We proceed with a characterization of continuous differentiability of the proximal mapping of prox-regular functions. Recall that proximal mapping of a function $f : \mathbf{X} \to \overline{\mathbb{R}}$ for a parameter value $\gamma > 0$, denoted by $\mathrm{prox}_{\gamma f}$, is defined by

$$\mathrm{prox}_{\gamma f}(x) = \underset{w \in \mathbf{X}}{\mathrm{argmin}} \Big\{ f(w) + \frac{1}{2\gamma}\|w - x\|^2 \Big\}.$$

Recall from [50, Exercise 1.24] that a function $f : \mathbf{X} \to \overline{\mathbb{R}}$ is called prox-bounded if the function $f+\alpha\|\cdot\|^2$ is bounded from below on $\mathbf{X}$ for some $\alpha \in \mathbb{R}$. We begin with recording some properties of the proximal mapping of prox-regular functions from [50, Proposition 13.37], which will be used in our characterization of continuous differentiability of their proximal mappings.

Proposition 4.6. *Assume that $f : \mathbf{X} \to \overline{\mathbb{R}}$ is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ and that f is prox-bounded. Then there exist positive constants ε and r such that for any $\gamma \in (0, 1/r)$, there is a neighborhood U_γ of $\bar{x} + \gamma\bar{v}$ on which $\operatorname{prox}_{\gamma\varphi}$ is nonempty, single-valued, Lipschitz continuous, and can be calculated by*

$$\operatorname{prox}_{\gamma f} = (I + \gamma T_\varepsilon)^{-1}, \tag{4.15}$$

where the set-valued mapping $T_\varepsilon : \mathbf{X} \rightrightarrows \mathbf{X}$ is defined by

$$T_\varepsilon(x) = \begin{cases} \partial f(x) \cap \mathbb{B}_\varepsilon(\bar{v}) & \text{if } x \in \mathbb{B}_\varepsilon(\bar{x}), \\ \emptyset & \text{otherwise}, \end{cases} \tag{4.16}$$

and where I stands for the identity operator on $\mathbf{X}$. Moreover, we have

$$\nabla e_\gamma f(x) = \frac{1}{\gamma}(x - \operatorname{prox}_{\gamma f}(x)), \quad x \in U_\gamma, \tag{4.17}$$

where the Moreau envelope function $e_\gamma f$ is defined by

$$e_\gamma f(x) = \inf_{w \in \in \mathbf{X}} \Big\{ f(w) + \frac{1}{2\gamma}\|w - x\|^2 \Big\}, \quad x \in \mathbf{X}.$$

In addition, if f is convex, then T_ε in (4.15) can be replaced with ∂f and the constant r can be taken as 0 with convention $1/0 = \infty$.

Theorem 4.7. *Assume that $f : \mathbf{X} \to \overline{\mathbb{R}}$ is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ and that f is prox-bounded. Then the following properties are equivalent:*

(a) *there exists a positive constant r such that for any $\gamma \in (0, 1/r)$, the proximal mapping $\operatorname{prox}_{\gamma f}$ is $\mathcal{C}^1$ around $\bar{x} + \gamma\bar{v}$;*
(b) *there exists a positive constant r such that for any $\gamma \in (0, 1/r)$, the envelope function $e_\gamma f$ is $\mathcal{C}^2$ around $\bar{x} + \gamma\bar{v}$;*
(c) *the subgradient mapping ∂f is strictly proto-differentiable at x for v for all $(x, v) \in \operatorname{gph} \partial f$ close to $(\bar{x}, \bar{v})$.*

If, in addition, f is convex, then the constant r in (a) and (b) can be taken as 0 with convention $1/0 = \infty$.

Proof. The equivalence of (a) and (b) results directly from (4.17). We now show that (a) is also equivalent to (c). To do so, take the positive constants ε and r from Proposition 4.6, pick $\gamma \in (0, 1/r)$, and take also the neighborhood U_γ of $\bar{x} + \gamma\bar{v}$ from Proposition 4.6. Define the mapping $\psi : \mathbf{X} \to \mathbf{X}$ by $\psi(x) = x - (\bar{x} + \gamma\bar{v})$ for any $x \in \mathbf{X}$ and consider the generalized equation

$$0 \in \psi(x) + \gamma T_\varepsilon(x), \tag{4.18}$$

where T_ε comes from (4.16). It is not hard to see that $\bar{x}$ is a solution to (4.18). Define now the solution mapping $S : \mathbf{X} \rightrightarrows \mathbf{X}$ by

$$S(u) := \big\{ x \in \mathbf{X} \,|\, u \in \psi(x) + \gamma T_\varepsilon(x) \big\}, \quad u \in \mathbf{X}.$$

Since U_γ is a neighborhood of $\bar{x} + \gamma\bar{v}$, we find a $\delta > 0$ such that $\delta\mathbb{B} \subset U_\gamma - (\bar{x} + \gamma\bar{v})$. Take $u \in \delta\mathbb{B}$ and observe that $x \in S(u)$ amounts to $u + \bar{x} + \gamma\bar{v} \in (I + \gamma T_\varepsilon)(x)$, which is equivalent via (4.15) to

$$x = (I + \gamma T_\varepsilon)^{-1}(u + \bar{x} + \gamma\bar{v}) = \operatorname{prox}_{\gamma f}(u + \bar{x} + \gamma\bar{v}).$$

These allow us to conclude that $S(u) = \operatorname{prox}_{\gamma f}(u + \bar{x} + r\bar{v})$ whenever $u \in \delta\mathbb{B}$. Suppose that (a) holds. We deduce from the latter that S is $\mathcal{C}^1$ around 0. By Theorem 4.3, the mapping γT_ε is strictly proto-differentiable at x for γv for all $(x, v) \in \operatorname{gph} \partial f$ close to $(\bar{x}, \bar{v})$. Since $\operatorname{gph} T_\varepsilon = (\operatorname{gph} \partial f) \cap \big(\mathbb{B}_\varepsilon(\bar{x}) \times \mathbb{B}_\varepsilon(\bar{v})\big)$, the latter property of γT_ε is equivalent to (c) and hence we are done with (a) $\Longrightarrow$ (c).

Assume now that (c) is satisfied. Thus, γT_ε is strictly proto-differentiable at x for γv for all $(x, v) \in \operatorname{gph} \partial f$ close to $(\bar{x}, \bar{v})$. By [14, Proposition 5.3], $\psi + \gamma T_\varepsilon$ is is strictly proto-differentiable at x for $x + \gamma v - (\bar{x} + \gamma\bar{v})$ for all $(x, v) \in \operatorname{gph} \partial f$ close to $(\bar{x}, \bar{v})$. Using again $S(u) = \operatorname{prox}_{\gamma f}(u + \bar{x} + r\bar{v})$ for all $u \in \delta\mathbb{B}$ and employing Proposition 4.6, we get that S is single-valued and Lipschitz continuous on $\delta\mathbb{B}$. Appealing now to [9, Propostion 3G.1] and using the facts that $S(0) = \bar{x}$ and $S = (I + \gamma T_\varepsilon)^{-1}$ particularly tell us that $\psi + \gamma T_\varepsilon$ is metrically regular at $\bar{x}$ for 0. Note that we can apply Theorem 4.2, and therefore also Theorem 4.3, to the generalized equation in (4.18) since $\operatorname{gph} T_\varepsilon = (\operatorname{gph} \partial f) \cap \big(\mathbb{B}_\varepsilon(\bar{x}) \times \mathbb{B}_\varepsilon(\bar{v})\big)$ due to (4.16) and since Theorem 4.3 exploits only local points of $\operatorname{gph} \partial f$ close to $(\bar{x}, -\psi(\bar{x}))$ in the generalized equation in (4.1). It follows from Theorem 4.3 that the solution mapping S has a Lipschitz continuous localization around 0 for $\bar{x}$, which is $\mathcal{C}^1$ in a neighborhood of 0. Since $S(u) = \operatorname{prox}_{\gamma f}(u+\bar{x}+\gamma\bar{v})$ for $u \in \delta\mathbb{B}$, we conclude that the proximal mapping $\operatorname{prox}_{\gamma f}$ is $\mathcal{C}^1$ in a neighborhood of $\bar{x} + \gamma\bar{v}$, which proves the implication (c) $\Longrightarrow$ (a).

If f is convex, the same argument can be utilized to justify the equivalence (a)-(c) using the fact that in this case, T_ε can be replaced with ∂f; see the final part of Proposition 4.6. This completes the proof. $\square$

The characterization of continuous differentiability of proximal mappings using a strict proto-differentiability assumption on the subgradient mapping as part (c) in Theorem 4.7 was first developed by Poliquin and Rockafellar in [41, Theorem 4.4]. The latter result, however, requires in the setting of Theorem 4.7 that $\bar{x}$ be a global minimizer of f, a condition that was replaced in our result by prox-boundedness of f, which is more realistic. We should emphasis that it is very likely that the latter requirement on $\bar{x}$ in [41, Theorem 4.4] can be dropped by inspecting carefully its proof. We, however, didn't proceed in that way and justify this result as an immediate consequence of the equivalence of metric regularity and strong metric regularity for generalized equations under strict proto-differentiability in Theorem 4.2. Note that while Theorem 4.7 presents not only a sufficient condition but a characterization of continuous differentiability of proximal mappings of prox-regular functions, it still requires more effort to disentangle the strict proto-differentiability assumption in part (c) therein. This was accomplished for polyhedral functions in [14, Theorem 3.5] and a certain composite functions in [16, Theorem 3.10] by showing that the latter strict proto-differentiability assumption amounts to a relative interior condition. While we will be pursuing a similar characterization for $\mathcal{C}^2$-decomposable functions in the next section, it remains as an open question whether a similar result can be justified for a prox-regular function in general.

Note that continuous differentiability of the projection mapping to convex sets was studied by Holmes in [18] in Hilbert spaces. His main result, [18, Theorem 2], states that if $\Omega \subset \mathbb{R}^d$ is a closed convex set, $x \in \mathbb{R}^d$, the boundary of Ω is a $\mathcal{C}^2$ smooth manifold around $y = P_\Omega(x)$, then the projection mapping P_Ω is $\mathcal{C}^1$ in a neighborhood of the open normal ray $\{y + t(x - y) | \ t > 0\}$; see also [6, Theorem 2.4] for an extension of Holmes' result for prox-regular sets.

5 Chain Rule for Strict Proto-Differentiability

The final section of this paper is devoted to provide a simple and verifiable characterization of strict proto-differentiability for a class of composite functions, which encompasses important examples of functions that we often encounter in different classes of constrained and composite optimization problems. To achieve this goal, recall from [51] that $g : \mathbf{Y} \to \overline{\mathbb{R}}$ is said to be

$\mathcal{C}^2$-decomposable at $u \in \mathbf{Y}$ if $g(u)$ is finite and g can be locally represented in the composite form

$$g(u') = g(u) + \vartheta(\Xi(u')) \quad \text{for} \quad u' \in \mathcal{O}, \tag{5.1}$$

where $\mathcal{O} \subset \mathbf{Y}$ is an open neighborhood of the given point u, $\vartheta : \mathbf{Z} \to \overline{\mathbb{R}}$ is a proper, lsc, sublinear function, and $\Xi : \mathcal{O} \to \mathbf{Z}$ is a $\mathcal{C}^2$-smooth mapping with $\Xi(u) = 0$ and $\mathbf{Z}$ being a finite dimensional Hilbert space. One can immediately conclude from [50, Definition 3.18 and Exercise 3.19] that $\vartheta(\Xi(u)) = 0$. Variational analysis of $\mathcal{C}^2$-decomposable functions often requires a constraint qualification. The most common condition, used for this purpose, is called the *nondegeneracy* condition. In what follows, we say that the nondegeneracy condition is satisfied for a $\mathcal{C}^2$-decomposable function g with representation in (5.1) at $u \in \mathbf{Y}$ if

$$\operatorname{par}\{\partial\vartheta(\Xi(u))\} \cap \ker \nabla\Xi(u)^* = \{0\} \tag{5.2}$$

holds, where $\operatorname{par}\{\partial\vartheta(\Xi(u))\}$ stands for the linear subspace of $\mathbf{Z}$ parallel to the affine hull of $\partial\vartheta(\Xi(u))$. It is important to note that the nondegeneracy condition in (5.2) holds automatically for many important examples of $\mathcal{C}^2$-decomposable functions; see [51, Examples 2.1 and 2.3]. Since our analysis in this section heavily relies on the nondegeneracy condition in (5.2), we call the function g *reliably* $\mathcal{C}^2$-decomposable at $u \in \mathbf{Y}$ if both conditions (5.1) and (5.2) are satisfied concurrently.

As shown in [51, Example 2.4], the class of $\mathcal{C}^2$-decomposable functions is a generalization of $\mathcal{C}^2$-*cone reducible* sets in the sense of [3, Definition 3.135], which is defined as follows: A closed convex set $C \subset \mathbf{Y}$ is $\mathcal{C}^2$-cone reducible at $u \in C$ to a closed convex cone $\Theta \subset \mathbf{Z}$ if there exist a neighborhood $\mathcal{O} \subset \mathbf{Y}$ of u and a $\mathcal{C}^2$-smooth mapping $\Xi : \mathbf{Y} \to \mathbf{Z}$ such that

$$C \cap \mathcal{O} = \big\{u' \in \mathcal{O} \,\big|\, \Xi(u') \in \Theta\big\}, \quad \Xi(u) = 0, \quad \text{and} \quad \nabla\Xi(u) : \mathbf{Y} \to \mathbf{Z} \;\; \text{is surjective.} \tag{5.3}$$

Below, we provide some important examples of reliably $\mathcal{C}^2$-decomposable functions that often appear in constrained and composite optimization problems.

Example 5.1. Assume that $g : \mathbf{Y} \to \overline{\mathbb{R}}$ and $\bar{u} \in \mathbf{Y}$ with $g(\bar{u})$ finite.

(a) If C is $\mathcal{C}^2$-cone reducible at $\bar{u} \in C$, then $g = \delta_C$ is reliably $\mathcal{C}^2$-decomposable at $\bar{u}$. It is known that polyhedral convex sets (cf. [3, Example 3.139]), the second-order cone, and the cone of $n \times n$ positive semidefinite matrices, denoted by $\mathbf{S}^{\mathbf{n}}_+$, (cf. [3, Example 3.140]) are $\mathcal{C}^2$-cone reducible at all of their points.

(b) If g is a polyhedral function, it was shown in [51, Example 2.1] that g is reliably $\mathcal{C}^2$-decomposable at any points of its domain. Furthermore, if $\Phi : \mathbf{X} \to \mathbf{Y}$ is a $\mathcal{C}^2$-smooth mapping, $\bar{u} := \Phi(\bar{x}) \in \operatorname{dom} g$ for some $\bar{x} \in \mathbf{X}$, and the condition

$$\operatorname{par}\{\partial g(\Phi(\bar{x}))\} \cap \ker \nabla\Phi(\bar{x})^* = \{0\} \tag{5.4}$$

is satisfied, the composite function $f = g \circ \Phi$ is reliably $\mathcal{C}^2$-decomposable at $\bar{x}$. To justify it, it follows from [14, Proposition 2.1(b)] that there is a neighborhood $\mathcal{O}$ of $\bar{u}$ for which we have

$$g(u) = g(\bar{u}) + \vartheta(u - \bar{u}) \quad \text{with} \;\; \vartheta := \mathrm{d}g(\bar{u}), \tag{5.5}$$

which leads us to the representation

$$f(x) = f(\bar{x}) + \vartheta(\Phi(x) - \bar{u})$$

for all x close to $\bar{x}$. Moreover, we deduce from [50, Exercise 8.44] that $\partial\vartheta(0) \subset \partial g(\bar{u})$. It is not hard to see via (5.5) that the latter inclusion becomes equality, and thus we obtain $\operatorname{par}\{\partial g(\bar{u})\} = \operatorname{par}\{\partial\vartheta(0)\}$. Combining this with (5.4) confirms that the nondegeneracy condition in (5.2) is satisfied at $u = \bar{u}$, and hence proves that f is reliably $\mathcal{C}^2$-decomposable at $\bar{x}$.

(c) Given $i \in \{1, \ldots, n\}$ and $X \in \mathbf{S^n}$, denote by $\ell_i(X)$ the number of eigenvalues that are equal to $\lambda_i(X)$ but are ranked before i including $\lambda_i(X)$. In what follows, we often drop X from $\ell_i(X)$ when the dependence of l_i on X can be seen clearly from the context. This integer allows us to locate λ_i in the group of the eigenvalues of X as follows:

$$\lambda_1(X) \geq \cdots \geq \lambda_{i-\ell_i(X)} > \lambda_{i-\ell_i+1}(X) = \cdots = \lambda_i(X) \geq \cdots \geq \lambda_n(X).$$

The eigenvalue $\lambda_{i-\ell_i(X)+1}(X)$, ranking first in the group of eigenvalues equal to $\lambda_i(X)$, is called a *leading* eigenvalue of X. For any $i \in \{1, \ldots, n\}$, define now the function $\alpha_i : \mathbf{S}^n \to \mathbb{R}$ by

$$\alpha_i(X) = \lambda_{i-\ell_i+1}(X) + \cdots + \lambda_i(X), \quad X \in \mathbf{S^n}.$$

According to [51, Example 2.3], α_i is reliably $\mathcal{C}^2$-decomposable at any $X \in \mathbf{S^n}$. In particular, when $\lambda_i(X)$ ranks first in a group of equal eigenvalues, meaning either $i = 1$ or $\lambda_{i-1}(X) > \lambda_i(X)$ if $i > 1$, the function α_i reduces to λ_i. This tells us that all the leading eigenvalue functions are always reliably $\mathcal{C}^2$-decomposable. Note that, except the first leading eigenvalue function, the other leading eigenvalue functions are nonconvex functions, hence they provide examples of nonconvex functions satisfying the reliable $\mathcal{C}^2$-decomposability property. Despite the latter fact, it was shown in [53, Theorem 2.3] that all the leading eigenvalue functions are subdifferentially regular.
Given $i \in \{1, \ldots, n\}$, define the sum of the first i largest eigenvalues of X by

$$g_i(X) = \lambda_1(X) + \cdots + \lambda_i(X).$$

It follows from [50, Exercise 2.54] that g_i is convex. Observe also that $g_i(X) = \alpha_i(X) + g_{i-\ell_i}(X)$. By [53, Proposition 1.3], $g_{i-\ell_i}$ is $\mathcal{C}^2$-smooth on $\mathbf{S}^n$. This, coupled with [51, Remark 2.2], demonstrates that g_i is reliably $\mathcal{C}^2$-decomposable at any $X \in \mathbf{S^n}$. The later can be extended for singular values of a matrix; see [26, Example 5.3.18] for more details. The readers can find more examples of $\mathcal{C}^2$-decomposable functions in [26, Section 5.3.3].

Below, we record some properties of sublinear functions, which is often used in this section.

Proposition 5.2. *Assume that $\vartheta : \mathbf{Z} \to \overline{\mathbb{R}}$ is a proper, lsc, and sublinear function and $\bar{z} = 0$. Then the following properties are fulfilled.*

(a) *For any $z \in \mathbf{Z}$, one has*

$$\vartheta(z) = \sup\big\{\langle \eta, z\rangle \,|\, \eta \in \vartheta(\bar{z})\big\} =: \sigma_{\partial\vartheta(\bar{z})}(z) \quad \textit{and} \quad \partial\vartheta(z) = \operatorname{argmax}\big\{\langle \eta, z\rangle \,|\, \eta \in \vartheta(\bar{z})\big\}.$$

(b) *For any $z \in \mathbf{Z}$, the inclusion $\emptyset \neq \partial\vartheta(z) \subset \partial\vartheta(\bar{z})$ holds.*

(c) *For any $w \in \operatorname{dom}\vartheta$, we have $\mathrm{d}\vartheta(\bar{z})(w) = \vartheta(w)$.*

Proof. Part (a) follows from [50, Theorem 8.24] and [50, Corollary 8.25]. The claimed inclusion in (b) results from the the second equality in (a). Finally, (c) follows from the fact that $\bar{z} = 0$ and ϑ is positive homogenous. □

Suppose that $g : \mathbf{Y} \to \overline{\mathbb{R}}$ is $\mathcal{C}^2$-decomposable at $u \in \mathbf{Y}$ with representation in (5.1). Given $(u, y) \in \operatorname{gph}\partial g$, we define the set of Lagrange multipliers associated with (u, y) by

$$M(u, y) := \big\{\mu \in \mathbf{Z} \;\big|\; \nabla\Xi(u)^*\mu = y,\ \mu \in \partial\vartheta(\Xi(u))\big\}. \tag{5.6}$$

The next result collects some simple consequences of the nondegeneracy condition in (5.2), important for our analysis of reliably $\mathcal{C}^2$-decomposable functions in this section.

Proposition 5.3. *Assume that $g : \mathbf{Y} \to \overline{\mathbb{R}}$ is reliably $\mathcal{C}^2$-decomposable at $\bar{u} \in \mathbf{Y}$ satisfying the representation in (5.1) for $u = \bar{u}$. Then, the following properties hold.*

(a) *There is a neighborhood $U \subset \mathcal{O}$ of $\bar{u}$ such that for any $u \in U \cap \operatorname{dom} g$, the nondegeneracy condition in* (5.2) *and the following basic constraint qualification* (*BCQ*)

$$N_{\operatorname{dom}\vartheta}(\Xi(u)) \cap \ker \nabla\Xi(u)^* = \{0\} \tag{5.7}$$

hold. Moreover, g is prox-regular and subdifferentially continuous at any $u \in U \cap \operatorname{dom} g$ for any $y \in \partial g(u)$ and is subdifferentially regular at any such u.

(b) *The Lagrange multiplier set $M(u, y)$ from* (5.6) *is a singleton for any $u \in U \cap \operatorname{dom} g$ and any $y \in \partial g(u)$, where U is taken from (a). Moreover, the dual condition*

$$K_\vartheta(\Xi(u), \mu)^* \cap \ker \nabla\Xi(u)^* = \{0\} \tag{5.8}$$

and the equivalence

$$y \in \operatorname{ri} \partial g(u) \iff \mu \in \operatorname{ri} \partial\vartheta(\Xi(u)) \tag{5.9}$$

are satisfied for any $u \in U \cap \operatorname{dom} g$, where μ is a unique element in $M(u, y)$ for any $y \in \partial g(u)$.

(c) *The set-valued mapping $\widetilde{\Xi} : \mathbf{Y} \rightrightarrows \mathbf{Z}$, defined by $\widetilde{\Xi}(u) := \Xi(u) + L^\perp$ with $L := \operatorname{par}\{\partial\vartheta(\Xi(\bar{u}))\}$, is metrically regular at $\bar{u}$ for $\Xi(\bar{u})$.*

Proof. To prove (a), we begin by justifying the nondegeneracy condition in (5.2) for any $u \in \operatorname{dom} g$ sufficiently close to $\bar{u}$. Observe that the condition (5.2) at $u = \bar{u}$ yields

$$\operatorname{par}\{\partial\vartheta(\Xi(\bar{u}))\} \cap \ker \nabla\Xi(u)^* = \{0\} \tag{5.10}$$

for all u sufficiently close to $\bar{u}$. Since $\Xi(\bar{u}) = 0$, it follows from Proposition 5.2(b) that $\emptyset \neq \partial\vartheta(\Xi(u)) \subset \partial\vartheta(\Xi(\bar{u}))$ for all u with $\Xi(u) \in \operatorname{dom}\vartheta$, which then tells us that the nondegeneracy condition (5.2) also holds for $u \in \operatorname{dom} g$ sufficiently close to $\bar{u}$ such that (5.10) is fulfilled. Thus, we find a neighborhood U of $\bar{u}$ such that for any $u \in U \cap \operatorname{dom} g$, the nondegeneracy condition in (5.7) is satisfied. We are now going to show that the BCQ condition (5.7) holds at $u = \bar{u}$. To do so, we claim that $N_{\operatorname{dom}\vartheta}(\Xi(\bar{u})) \subset \operatorname{par}\{\partial\vartheta(\Xi(\bar{u}))\}$. Since $\Xi(\bar{u}) = 0$, it follows from Proposition 5.2(c) that $\mathrm{d}\vartheta(\Xi(\bar{u}))(w) = \vartheta(w)$ for all $w \in \mathbf{Z}$, which in turn implies $K_\vartheta(\Xi(\bar{u}), \mu) \subset \operatorname{dom}\vartheta$, where μ is taken arbitrarily from $\partial\vartheta(\Xi(\bar{u}))$. Since ϑ is sublinear, $\operatorname{dom}\vartheta$ is a convex cone. Thus, we arrive at

$$N_{\operatorname{dom}\vartheta}(\Xi(\bar{u})) = (\operatorname{dom}\vartheta)^* \subset K_\vartheta(\Xi(\bar{u}), \mu)^* = T_{\partial\vartheta(\Xi(\bar{u}))}(\mu) \subset \operatorname{par}\{\partial\vartheta(\Xi(\bar{u}))\}, \tag{5.11}$$

which confirms our claim. We then conclude from (5.2) for $u = \bar{u}$ that the BCQ condition (5.7) holds at that point. Shrinking the neighborhood U of $\bar{u}$, if necessary, and employing robustness of the normal cone mapping $N_{\operatorname{dom}\vartheta}$ and $\mathcal{C}^2$-smoothness of Ξ, we can argue further that the BCQ condition (5.7) is valid for all $u \in U \cap \operatorname{dom} g$. Finally, the composite function g in (5.1), satisfying the nondegeneracy condition in (5.2), is strongly amenable at any $u \in U \cap \operatorname{dom} g$ in the sense of [50, Definition 10.23(a)] and therefore is prox-regular and subdifferentially continuous at all points $u \in U \cap \operatorname{dom} g$ for any $y \in \partial g(u)$ according to [50, Theorem 13.32]. The subdifferential regularity of g at any such u falls out of [50, Exercise 10.25(a)], which completes the proof of (a).

Turning now to (b), take the neighborhood U from (a) and pick $u \in U \cap \operatorname{dom} g$ and $y \in \partial g(u)$. It follows from [50, Example 10.8], the composite representation (5.1), and (5.7) that

$$\partial g(u) = \nabla\Xi(u)^* \partial\vartheta(\Xi(u)). \tag{5.12}$$

Since $y \in \partial g(u)$, the latter implies that $M(u, y)$ is nonempty. We claim that $M(u, y)$ is a singleton. Indeed, assuming that $\mu^1, \mu^2 \in M(u, y)$, we get from (5.6) that $\mu^1 - \mu^2 \in \ker \nabla\Xi(u)^*$ and $\mu^1 - \mu^2 \in \operatorname{par}\{\partial\vartheta(\Xi(u))\}$. Employing the nondegeneracy condition in (5.2), we arrive at $\mu^1 = \mu^2$, which proves our claim. The dual condition in (5.8) then follows from the inclusion in

(5.11) together with the validity of the nondegeneracy condition at any $u \in U \cap \operatorname{dom} g$, which was established in (a). Regarding the equivalence in (5.9), by [50, Proposition 2.44], it is a consequence of (5.12).

Finally, to justify (c), define the set-valued mapping $F : \mathbf{Y} \rightrightarrows \mathbf{Z}$ by $F(u) = L^{\perp}$. Observe that $\operatorname{gph} F = \mathbf{Y} \times L^{\perp}$. Using the definition of coderivative from (3.4), we obtain $D^*F(\bar{u}, 0)(w) = \{0\}$ for $w \in L$ and $D^*F(\bar{u}, 0)(w) = \emptyset$ otherwise. Since $\widetilde{\Xi}(u) = \Xi(u) + F(u)$ for all $u \in \mathbf{Y}$, using the sum rule for the coderivative from [50, Exercise 10.43(b)] yields

$$D^*\widetilde{\Xi}(\bar{u}, \Xi(\bar{u}))(w) = \nabla\Xi(\bar{u})^*w + D^*F(\bar{u}, 0)(w) = \nabla\Xi(\bar{u})^*w \quad \text{for all } \ w \in L. \tag{5.13}$$

By [50, Theorem 9.43], the mapping $\widetilde{\Xi}$ is metrically regular at $\bar{u}$ for $\Xi(\bar{u})$ if and only if $0 \in D^*\widetilde{\Xi}(\bar{u}, \Xi(\bar{u}))(w)$ yields $w = 0$. According to (5.13), the condition $0 \in D^*\widetilde{\Xi}(\bar{u}, \Xi(\bar{u}))(w)$ amounts to $w \in \operatorname{par}\{\partial\vartheta(\Xi(\bar{u}))\} \cap \ker \nabla\Xi(\bar{u})^*$. Since the nondegeneracy condition in (5.2) holds at $u = \bar{u}$, we obtain $w = 0$, which proves (c) and hence completes the proof. □

We proceed with a chain rule for the second subderivative of reliably $\mathcal{C}^2$-decomposable functions.

Theorem 5.4. *Assume that $g : \mathbf{Y} \to \overline{\mathbb{R}}$ is reliably $\mathcal{C}^2$-decomposable at $\bar{u} \in \mathbf{Y}$ satisfying the representation in (5.1) for $u = \bar{u}$. Then, there is a neighborhood U of $\bar{u}$ such that for any $u \in U \cap \operatorname{dom} g$ and any $y \in \partial g(u)$, the second subderivative of g at u for y can be calculated by*

$$\mathrm{d}^2 g(u, y)(w) = \langle \mu, \nabla^2\Xi(u)(w, w)\rangle + \mathrm{d}^2\vartheta(\Xi(u), \mu)(\nabla\Xi(u)w), \qquad \textit{for } \ w \in \mathbf{Y}, \tag{5.14}$$

where $\mu \in \mathbf{Z}$ is the unique element in the multiplier set $M(u, y)$ from (5.6).

Proof. Take the neighborhood U from Proposition 5.3(a) and pick $u \in U \cap \operatorname{dom} g$ and $y \in \partial g(u)$. By Proposition 5.3(b), we know that the multiplier set $M(u, y)$ is a singleton, say $M(u, y) = \{\mu\}$. It follows from Proposition 5.3(a) that the BCQ in (5.7) holds, which, coupled with [50, Theorem 13.14], gives us the inequality

$$\mathrm{d}^2 g(u, y)(w) \geq \langle \mu, \nabla^2\Xi(u)(w, w)\rangle + \mathrm{d}^2\vartheta(\Xi(u), \mu)(\nabla\Xi(u)w) \quad \text{for all } \ w \in \mathbf{Y}. \tag{5.15}$$

To obtain the opposite inequality, fix $w \in \mathbf{Y}$ and pick sequences $t_k \searrow 0$ and $\xi^k \to \nabla\Xi(u)w$ as $k \to \infty$, satisfying

$$\mathrm{d}^2\vartheta(\Xi(u), \mu)(\nabla\Xi(u)w) = \lim_{k\to\infty} \frac{\vartheta(\Xi(u) + t_k\xi^k) - \vartheta(\Xi(u)) - t_k\langle \mu, \xi^k\rangle}{\frac{1}{2}t_k^2}. \tag{5.16}$$

We know from the nondegeneracy condition in (5.2) that the set-valued mapping $\widetilde{\Xi} : \mathbf{Y} \rightrightarrows \mathbf{Z}$, defined in Proposition 5.3(c), is metrically regular at $\bar{u}$ for $\Xi(\bar{u})$, meaning that there exist $\ell \geq 0$ and neighborhoods V of $\bar{u}$ and W of $\Xi(\bar{u})$ for which the estimate

$$\operatorname{dist}(p, \widetilde{\Xi}^{-1}(q)) \leq \ell \operatorname{dist}(q, \widetilde{\Xi}(p)) = \ell \operatorname{dist}(q, \Xi(p) + L^{\perp}) \quad \text{for all } \ (p, q) \in V \times W$$

is satisfied, where $L = \operatorname{par}\{\partial\vartheta(\Xi(\bar{u}))\}$. Since Ξ is $\mathcal{C}^2$-smooth, we can assume by shrinking the neighborhood U if necessary that $(p^k, q^k) := (u + t_k w, \Xi(u) + t_k\xi^k) \in V \times W$ for any k sufficiently large. Thus, for any k sufficiently large, we can find $u^k \in \widetilde{\Xi}^{-1}(q^k)$ such that

$$\|u + t_k w - u^k\| = \|p^k - u^k\| \leq \ell \operatorname{dist}\big(q^k, \widetilde{\Xi}(p^k)\big) \leq \ell\, \|\Xi(u) + t_k\xi^k - \Xi(u + t_k w)\|.$$

Let $w^k := (u^k - u)/t_k$ and observe that

$$\|w^k - w\| = \Big\|\frac{u^k - u - t_k w}{t_k}\Big\| \leq \ell\, \Big\|\frac{\Xi(u + t_k w) - \Xi(u)}{t_k} - \xi^k\Big\| \to 0 \quad \text{ as } \ k \to \infty,$$

which yields $w^k \to w$ as $k \to \infty$. Since $u^k \in \widetilde{\Xi}^{-1}(q^k)$, we find by the definition of $\widetilde{\Xi}$ a vector $\nu^k \in L^\perp$ such that

$$\Xi(u) + t_k \xi^k = q^k = \Xi(u^k) + \nu^k. \tag{5.17}$$

Recall that $\mu \in M(u, y)$ and hence $\mu \in \partial\vartheta(\Xi(u)) \subset \partial\vartheta(\Xi(\bar{u}))$. Since $\nu^k \in L^\perp = \text{par}\,\{\partial\vartheta(\Xi(\bar{u}))\}^\perp$, we get $\langle \nu^k, \eta \rangle = \langle \nu^k, \mu \rangle$ for any $\eta \in \partial\vartheta(\Xi(\bar{u}))$. This, coupled with Proposition 5.2(a), allows us to conclude for all k sufficiently large that

$$\begin{aligned} \vartheta(\Xi(u) + t_k \xi^k) = \vartheta(\Xi(u^k) + \nu^k) = \sigma_{\partial\vartheta(\Xi(\bar{u}))}\big(\Xi(u^k) + \nu^k\big) &= \langle \mu, \nu^k \rangle + \sigma_{\partial\vartheta(\Xi(\bar{u}))}\big(\Xi(u^k)\big) \\ &= \langle \mu, \nu^k \rangle + \vartheta(\Xi(u^k)). \end{aligned} \tag{5.18}$$

We therefore obtain from (5.1), (5.16), and (5.17) as well as $w^k \to w$ that

$$\begin{aligned} &\mathrm{d}^2\vartheta(\Xi(u), \mu)(\nabla\Xi(u)w) = \lim_{k\to\infty} \frac{\vartheta(\Xi(u^k)) + \langle \mu, \nu^k \rangle - \vartheta(\Xi(u)) - \langle \mu, \Xi(u^k) + \nu^k - \Xi(u) \rangle}{\frac{1}{2}t_k^2} \\ &= \lim_{k\to\infty} \Big(\frac{g(u + t_k w^k) - g(u) - t_k \langle \nabla\Xi(u)^*\mu, w^k \rangle}{\frac{1}{2}t_k^2} - \frac{\langle \mu, \Xi(u^k) - \Xi(u) - t_k \nabla\Xi(u) w^k \rangle}{\frac{1}{2}t_k^2} \Big) \\ &\ge \mathrm{d}^2 g(u, y)(w) - \langle \mu, \nabla^2\Xi(u)(w, w) \rangle, \end{aligned}$$

which verifies the opposite inequality in (5.15) and hence completes the proof. □

Note that the chain rule for the second subderivative of reliably $\mathcal{C}^2$-decomposable functions, established above, cannot be obtained from previous efforts in this direction. In fact, it was shown in [29, Theorem 5.4] that such a chain rule can be achieved for a broader class of functions, which includes $\mathcal{C}^2$-decomposable functions, under the metric subregularity constraint qualification, which is strictly weaker than the nondegeneracy condition, imposed in Theorem 5.4. However, [29, Theorem 5.4] gives us this chain rule just at the point under consideration and cannot ensure a similar result for all the points nearby. Indeed, had we assumed in Theorem 5.4 the reliable $\mathcal{C}^2$-decomposability of g for all u close to $\bar{u}$, we could have applied [29, Theorem 5.4] to obtain (5.14). Even in such a case, we should have dealt with the fact that the mapping Ξ in (5.14) would depend on u, an unpleasant issue which causes a major obstacle in the proof of Theorem 5.8. The nondegeneracy condition allows us to bypass this hurdle by exerting metric regularity into the mapping $\widetilde{\Xi}$ in Propsoition 5.3(c). Note also that a similar chain rule was recently obtained in [2, Corollary 4.3] for a general composite function under the surjectivity assumption on the jacobian of the inner mapping. Imposing the latter surjectivity condition in Theorem 5.4, one can obtain (5.14) without facing the above-mentioned obstacle. We, however, require to proceed with the nondegeneracy condition, which is in general strictly weaker than the surjectivity condition.

To establish our characterization of strict proto-differentiability of subgradient mappings of $\mathcal{C}^2$-decomposable functions, we begin to analyze some variational properties of the sublinear function ϑ from (5.1) at the origin.

Lemma 5.5. *Let $\vartheta : \mathbf{Z} \to \overline{\mathbb{R}}$ be a proper, lsc, and sublinear function, $\bar{z} = 0$, and $\bar{\mu} \in \text{ri}\,\partial\vartheta(\bar{z})$. Then there exists $\varepsilon > 0$ such that for all $(z, \mu) \in \big(\text{gph}\,\partial\vartheta\big) \cap \mathbb{B}_\varepsilon(\bar{z}, \bar{\mu})$, we have $\mu \in \text{ri}\,\partial\vartheta(\bar{z})$ and $\mu \in \text{ri}\,\partial\vartheta(z)$.*

Proof. Since $\bar{\mu} \in \text{ri}\,\partial\vartheta(\bar{z})$, we find $\varepsilon > 0$ such that $\big(\text{aff}\,\partial\vartheta(\bar{z})\big) \cap \mathbb{B}_{2\varepsilon}(\bar{\mu}) \subset \partial\vartheta(\bar{z})$. Take any $(z, \mu) \in \big(\text{gph}\,\partial\vartheta\big) \cap \mathbb{B}_\varepsilon(\bar{z}, \bar{\mu})$. It is not hard to see that

$$\big(\text{aff}\,\partial\vartheta(\bar{z})\big) \cap \mathbb{B}_\varepsilon(\mu) \subset \big(\text{aff}\,\partial\vartheta(\bar{z})\big) \cap \mathbb{B}_{2\varepsilon}(\bar{\mu}) \subset \partial\vartheta(\bar{z}),$$

which in turn proves that $\mu \in \text{ri}\,\partial\vartheta(\bar{z})$. To prove the second claim, take $\widehat{q} \in \big(\text{aff}\,\partial\vartheta(z)\big) \cap \mathbb{B}_\varepsilon(\mu)$. We then deduce from $\mu \in \mathbb{B}_\varepsilon(\bar{\mu})$ and Proposition 5.2(b) that $\widehat{q} \in \big(\text{aff}\,\partial\vartheta(\bar{z})\big) \cap \mathbb{B}_{2\varepsilon}(\bar{\mu}) \subset \partial\vartheta(\bar{z})$.

Since $\widehat{q} \in \operatorname{aff} \partial\vartheta(z)$, we find $m \in \mathbb{N}$, $\alpha_i \in \mathbb{R}$ and $q^i \in \partial\vartheta(z)$ for $i = 1, \ldots, m$ with $\sum_{i=1}^m \alpha_i = 1$ such that $\widehat{q} = \sum_{i=1}^m \alpha_i q^i$. We conclude from Proposition 5.2(a) and $q^i \in \partial\vartheta(z)$ that $\langle q^i, z\rangle = \beta$ for any $i = 1, \ldots, m$, where $\beta := \max_{q\in\partial\vartheta(\bar{z})} \langle q, z\rangle$. Thus, we obtain

$$\langle \widehat{q}, z\rangle = \sum_{i=1}^m \alpha_i \langle q^i, z\rangle = \beta = \max_{q\in\partial\vartheta(\bar{z})} \langle q, z\rangle.$$

Employing again Proposition 5.2(a) and using the fact that $\widehat{q} \in \partial\vartheta(\bar{z})$ imply that $\widehat{q} \in \partial\vartheta(z)$. This proves the inclusion $\big(\operatorname{aff} \partial\vartheta(z)\big) \cap \mathbb{B}_\varepsilon(\mu) \subset \partial\vartheta(z)$, which yields $\mu \in \operatorname{ri} \partial\vartheta(z)$ and hence completes the proof. □

We continue with recording a characterization of strict proto-differentiability of subgradient mappings of prox-regular functions via the concept of strict twice epi-differentiability. Following [40, page 1830], we say that a function $f : \mathbf{X} \to \overline{\mathbb{R}}$ is *strictly* twice epi-differentiable at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ if the functions $\Delta_t^2 f(x, v)$ epi-converge to a function as $t \searrow 0$, $(x, v) \to (\bar{x}, \bar{v})$ with $f(x) \to f(\bar{x})$ and $(x, v) \in \operatorname{gph} \partial f$. If this condition holds, the limit function must be the second subderivative $\mathrm{d}^2 f(\bar{x}, \bar{v})$.

Proposition 5.6. *Under the hypothesis of Corollary 2.4, the following properties can be added to the list of equivalences:*

- (a) *f is strictly twice epi-differentiable at $\bar{x}$ for $\bar{v}$;*
- (b) *$\mathrm{d}^2 f(x, v)$ epi-converges to $\mathrm{d}^2 f(\bar{x}, \bar{v})$ as $(x, v) \to (\bar{x}, \bar{v})$ in the set of pairs $(x, v) \in \operatorname{gph} \partial f$ for which f is twice epi-differentiable;*
- (c) *$\mathrm{d}^2 f(x, v)$ epi-converges to $\mathrm{d}^2 f(\bar{x}, \bar{v})$ as $(x, v) \to (\bar{x}, \bar{v})$ in the set of pairs $(x, v) \in \operatorname{gph} \partial f$ for which $\mathrm{d}^2 f(x, v)$ is generalized quadratic.*

Proof. The equivalence of the properties in (a)-(c) to those in Corollary 2.4 were established in [41, Corollary 4.3] using the Attouch' theorem (cf. [50, Theorem 12.35]). Note that [41, Corollary 4.3] does not determine to which function $\mathrm{d}^2 f(x, v)$ epi-converges in (b) and (c). The properties in Theorem 2.4 allow us to demonstrate that the latter function is indeed the second subderivative $\mathrm{d}^2 f(\bar{x}, \bar{v})$. One can use a similar argument as [41, Corollary 4.3] to justify the claimed equivalences. □

Below, we present our first major result in this section, a characterization of strict twice epi-differentiability of sublinear functions.

Theorem 5.7. *Let $\vartheta : \mathbf{Z} \to \overline{\mathbb{R}}$ be a proper, lsc, and sublinear function. Then, ϑ is strictly twice epi-differentiable at $\bar{z} = 0$ for $\bar{\mu}$ if and only if $\bar{\mu} \in \operatorname{ri} \partial\vartheta(\bar{z})$.*

Proof. Observe first from [26, page 138] that

$$\mathrm{d}^2\vartheta(\bar{z}, \bar{\mu})(w) = \delta_{K_\vartheta(\bar{z},\bar{\mu})}(w), \quad \text{for } w \in \mathbf{Z}. \tag{5.19}$$

If ϑ is strictly twice epi-differentiable at $\bar{z} = 0$ for $\bar{\mu} \in \partial\vartheta(\bar{z})$, it follows from Proposition 5.6 that $\partial\vartheta$ is strictly proto-differentiable at $\bar{z}$ for $\bar{\mu}$. Moreover, by (5.19), we have $\operatorname{dom} \mathrm{d}^2\vartheta(\bar{z}, \bar{\mu}) = K_\vartheta(\bar{z}, \bar{\mu})$. Employing Proposition 3.4 yields $\bar{\mu} \in \operatorname{ri} \partial\vartheta(\bar{z})$.

Assume now that $\bar{\mu} \in \operatorname{ri} \partial\vartheta(\bar{z})$. To justify strict twice epi-differentiability of ϑ at $\bar{z}$ for $\bar{\mu}$, we claim that the functions $\Delta_t^2\vartheta(z, \mu)$ epi-converge to $\delta_{K_\vartheta(\bar{z},\bar{\mu})}$ as $t \searrow 0$ and $(z, \mu) \to (\bar{z}, \bar{\mu})$ with $(z, \mu) \in \operatorname{gph} \partial\vartheta$. According to [50, Proposition 7.2], the latter epi-convergence claim holds if and only if for any $w \in \mathbf{Z}$, any sequence $t_k \searrow 0$, and any sequence $(z^k, \mu^k) \to (\bar{z}, \bar{\mu})$ with $\mu^k \in \partial\vartheta(z^k)$, the second-order difference quotients $\Delta_{t_k}^2\vartheta(z^k, \mu^k)$ satisfy the conditions

$$\liminf_{k\to\infty} \Delta_{t_k}^2\vartheta(z^k, \mu^k)(w^k) \geq \delta_{K_\vartheta(\bar{z},\bar{\mu})}(w) \quad \text{for every sequence } w^k \to w, \tag{5.20a}$$

$$\limsup_{k\to\infty} \Delta_{t_k}^2\vartheta(z^k, \mu^k)(w^k) \leq \delta_{K_\vartheta(\bar{z},\bar{\mu})}(w) \quad \text{for some sequence } w^k \to w. \tag{5.20b}$$

We split the verification of (5.20a) and (5.20b) into the two possible cases: $w \in K_\vartheta(\bar z,\bar\mu)$ and $w \notin K_\vartheta(\bar z,\bar\mu)$. Assume first that $w \in K_\vartheta(\bar z,\bar\mu)$ and choose arbitrary sequences $t_k \searrow 0$, and $(z^k,\mu^k) \to (\bar z,\bar\mu)$ with $\mu^k \in \partial\vartheta(z^k)$. We deduce from the convexity of ϑ and $\mu^k \in \partial\vartheta(z^k)$ that

$$\Delta^2_{t_k}\vartheta(z^k,\mu^k)(w^k) = \frac{\vartheta(z^k+t_kw^k)-\vartheta(z^k)-t_k\langle\mu^k,w^k\rangle}{\frac{1}{2}t_k^2} \ge 0,$$

which clearly justifies (5.20a) for any sequence $w^k \to w$, since $w \in K_\vartheta(\bar z,\bar\mu)$. To prove (5.20b), set $w^k = w$ for all $k \in \mathbb{N}$. Since $\bar\mu \in \operatorname{ri}\partial\vartheta(\bar z)$, the critical cone $K_\vartheta(\bar z,\bar\mu) = N_{\partial\vartheta(\bar z)}(\bar\mu)$ is a linear subspace, where the last equality comes from Proposition 3.3(b). By Proposition 5.2(b), the inclusion $\partial\vartheta(z^k) \subset \partial\vartheta(\bar z)$ holds for any k, which, coupled with $w \in N_{\partial\vartheta(\bar z)}(\bar\mu)$, yields $\langle\mu^k-\bar\mu,w\rangle = 0$ for all $k \in \mathbb{N}$. This, together with Proposition 5.2(c), leads us to $\langle\mu^k,w\rangle = \langle\bar\mu,w\rangle = \mathrm{d}\vartheta(\bar z)(w) = \vartheta(w)$ for all k, and so we arrive at

$$\Delta^2_{t_k}\vartheta(z^k,\mu^k)(w) = \frac{\vartheta(z^k+t_kw)-\vartheta(z^k)-t_k\vartheta(w)}{\frac{1}{2}t_k^2} \le 0$$

due to the sublinearity of ϑ. Passing to the limit clearly proves (5.20b) in this case.

Suppose now that $w \notin K_\vartheta(\bar z,\bar\mu)$. Clearly, (5.20b) is fulfilled for any sequence $w^k \to w$ with $w \notin K_\vartheta(\bar z,\bar\mu)$. Pick arbitrary sequences $w^k \to w$ and $(z^k,\mu^k) \to (\bar z,\bar\mu)$ with $\mu^k \in \partial\vartheta(z^k)$. Recalling that $\bar\mu \in \operatorname{ri}\partial\vartheta(\bar z)$, we deduce from Lemma 5.5 that $\mu^k \in \operatorname{ri}\partial\vartheta(\bar z)$ for all k sufficiently large, and therefore arrive at $N_{\partial\vartheta(\bar z)}(\mu^k) = N_{\partial\vartheta(\bar z)}(\bar\mu) = K_\vartheta(\bar z,\bar\mu)$ for such k. It follows from Proposition 5.2(a) that $z^k \in N_{\partial\vartheta(\bar z)}(\mu^k)$. Combining these tells us that $z^k \in K_\vartheta(\bar z,\bar\mu)$, and also $-z^k \in K_\vartheta(\bar z,\bar\mu)$ for k sufficiently large, since $K_\vartheta(\bar z,\bar\mu)$ is a linear subspace of $\mathbf{Z}$. Thus, we conclude from Proposition 5.2(c) that

$$\mathrm{d}\vartheta(\bar z)(z^k) = \vartheta(z^k) = \langle\bar\mu,z^k\rangle = -\langle\bar\mu,-z^k\rangle = -\mathrm{d}\vartheta(\bar z)(-z^k) = -\vartheta(-z^k),$$

which, together with the sublinearity of ϑ, yields

$$\Delta^2_{t_k}\vartheta(z^k,\mu^k)(w^k) = \frac{\vartheta(z^k+t_kw^k)+\vartheta(-z^k)-t_k\langle\mu^k,w^k\rangle}{\frac{1}{2}t_k^2} \ge \frac{\vartheta(w^k)-\langle\mu^k,w^k\rangle}{\frac{1}{2}t_k}$$

for all k sufficiently large. Observe that

$$\liminf_{k\to\infty}\vartheta(w^k)-\langle\mu^k,w^k\rangle \ge \vartheta(w)-\langle\bar\mu,w\rangle > 0,$$

where the first inequality results from lower semicontinuiuty of ϑ and the second one comes from $w \notin K_\vartheta(\bar z,\bar\mu)$. Combining these estimates and passing to the limit clearly justify (5.20a) for any $w \notin K_\vartheta(\bar z,\bar\mu)$ and hence completes the proof. □

We are now in a position to establish our main result in the section, a simple characterization of strict proto-differentiability of subgradient mappings of $\mathcal{C}^2$-decomposable functions.

Theorem 5.8 (strict proto-differentiability of $\mathcal{C}^2$-decomposable functions). *Assume that $g : \mathbf{Y} \to \overline{\mathbb{R}}$, $(\bar u,\bar y) \in \operatorname{gph}\partial g$, and that there is a neighborhood U of $\bar u$ such that g is reliably $\mathcal{C}^2$-decomposable at any $u \in U$. Then the following properties are equivalent:*

(a) *g is strictly twice epi-differentiable at u for y for any pair $(u,y) \in \operatorname{gph}\partial g$ close to $(\bar u,\bar y)$;*
(b) *∂g is strictly proto-differentiable at u for y for any pair $(u,y) \in \operatorname{gph}\partial g$ close to $(\bar u,\bar y)$;*
(c) *$\bar y \in \operatorname{ri}\partial g(\bar u)$.*

Proof. The equivalence of (a) and (b) results from Proposition 5.6. We proceed by concluding from Theorem 5.4 and (5.19) that

$$\operatorname{dom}\mathrm{d}^2g(\bar u,\bar y) = \big\{w \in \mathbf{Y}\,\big|\,\nabla\Xi(\bar u)w \in K_\vartheta(\Xi(\bar u),\bar\mu)\big\},$$

where $\bar{\mu}$ is the unique element of the multiplier set $M(\bar{u},\bar{y})$; see Proposition 5.3(b). By Proposition 5.3(a), the BCQ condition in (5.7) is satisfied at $u=\bar{u}$. Thus, it follows from the chain rule for the subderivative in [50, Theorem 10.6] that

$$\begin{aligned}K_g(\bar{u},\bar{y}) &= \big\{w\in\mathbf{Y}\,\big|\,\mathrm{d}g(\bar{u})(w)=\mathrm{d}\vartheta(\Xi(\bar{u}))(\nabla\Xi(\bar{u})w)=\langle\bar{y},w\rangle=\langle\bar{\mu},\nabla\Xi(\bar{u})w\rangle\big\}\\ &= \big\{w\in\mathbf{Y}\,\big|\,\nabla\Xi(\bar{u})w\in K_\vartheta(\Xi(\bar{u}),\bar{\mu})\big\}.\end{aligned}$$

Combining these tells us that $\operatorname{dom}\mathrm{d}^2g(\bar{u},\bar{y})=K_g(\bar{u},\bar{y})$. Moreover, g is subdifferentially regular at $\bar{u}$ due to Proposition 5.3(a). Thus, Proposition 3.4 confirms the implication (b) $\Longrightarrow$ (c).

We now turn to the implication (c) $\Longrightarrow$ (a). Assume first that $\bar{y}\in\operatorname{ri}\partial g(\bar{u})$, g has the representation in (5.1) with $u=\bar{u}$, and the nondegeneracy condition (5.2) holds at $u=\bar{u}$. We are going to verify strict twice epi-differentiability of g at $\bar{u}$ for $\bar{y}$ via justifying its characterization in Proposition 5.6(b). To this end, suppose that $(u,y)\to(\bar{u},\bar{y})$ with $y\in\partial g(u)$ and that g is twice epi-differentiable at u for y. The latter is equivalent via [50, Proposition 7.2] to the fact that for any $w\in\mathbf{Y}$ and any sequence of $t_k\searrow 0$, there exists a sequence $w^k\to w$ with

$$\lim_{k\to\infty}\frac{g(u+t_kw^k)-g(u)-t_k\langle y,w^k\rangle}{\frac{1}{2}t_k^2}=\mathrm{d}^2g(u,y)(w). \tag{5.21}$$

Pick $(u,y)\in\operatorname{gph}\partial g$ sufficiently close to $(\bar{u},\bar{y})$ such that $u+t_kw^k\in\mathcal{O}$ for any k sufficiently large with $\mathcal{O}$ taken from (5.1). It follows from (5.1) that

$$g(u+t_kw^k)-g(u)=\vartheta(\Xi(u+t_kw^k))-\vartheta(\Xi(u)).$$

Set $\xi^k:=(\Xi(u+t_kw^k)-\Xi(u))/t_k$. We can rewrite the left-hand side of (5.21) to obtain

$$\mathrm{d}^2g(u,y)(w)=\lim_{k\to\infty}\Big(\Delta_{t_k}^2\vartheta(\Xi(u),\mu)(\xi^k)+\frac{\langle\mu,\Xi(u+t_kw^k)-\Xi(u)-\nabla\Xi(u)w^k\rangle}{\frac{1}{2}t_k^2}\Big), \tag{5.22}$$

where μ is the unique element in the multiplier set $M(u,y)$; see Proposition 5.3(b). According to Theorem 5.4, (5.14) holds for any $(u,y)\in\operatorname{gph}\partial g$ with u sufficiently close to $\bar{u}$. This, coupled with (5.22), leads us to

$$\begin{aligned}\lim_{k\to\infty}\Delta_{t_k}^2\vartheta(\Xi(u),\mu)(\xi^k) &= \mathrm{d}^2g(u,y)(w)-\lim_{k\to\infty}\frac{\langle\mu,\Xi(u+t_kw^k)-\Xi(u)-\nabla\Xi(u)w^k\rangle}{\frac{1}{2}t_k^2}\\ &= \mathrm{d}^2\vartheta(\Xi(u),\mu)(\nabla\Xi(u)w).\end{aligned} \tag{5.23}$$

Our goal is to show that ϑ is twice epi-differentiable at $\Xi(u)$ for μ. In to order to achieve it, pick $\xi\in\mathbf{Z}$. It is not hard to see that the nondegeneracy condition in (5.2) for $u=\bar{u}$ can be equivalently described as $\mathbf{Z}=\operatorname{rge}\nabla\Xi(u)+\operatorname{par}\{\partial\vartheta(\Xi(\bar{u}))\}^\perp$. Thus, we can find $w\in\mathbf{Y}$ and $\nu\in\operatorname{par}\{\partial\vartheta(\Xi(\bar{u}))\}^\perp$ such that $\xi=\nabla\Xi(u)w+\nu$. A similar argument as that for (5.18) brings us to

$$\vartheta(\Xi(u)+t_k(\xi^k+\nu))=\vartheta(\Xi(u)+t_k\xi^k)+\langle\mu,\nu\rangle,$$

which in turn yields

$$\Delta_{t_k}^2\vartheta(\Xi(u),\mu)(\xi^k)=\Delta_{t_k}^2\vartheta(\Xi(u),\mu)(\xi^k+\nu). \tag{5.24}$$

Observing that this equality is, indeed, valid for any $\xi^k\in\mathbf{Y}$, we arrive at a similar equality for the second subderivative of ϑ, namely

$$\mathrm{d}^2\vartheta(\Xi(u),\mu)(\nabla\Xi(u)w)=\mathrm{d}^2\vartheta(\Xi(u),\mu)(\nabla\Xi(u)w+\nu)=\mathrm{d}^2\vartheta(\Xi(u),\mu)(\xi).$$

Using this, we infer from (5.23) and (5.24) that

$$\lim_{k\to\infty}\Delta_{t_k}^2\vartheta(\Xi(u),\mu)(\xi^k+\nu)=\mathrm{d}^2\vartheta(\Xi(u),\mu)(\xi).$$

Since $\xi^k + \nu \to \nabla\Xi(u)w + \nu = \xi$ and since $\xi \in \mathbf{Z}$ was taken arbitrary, the above equality confirms that ϑ is twice epi-differentiable at $\Xi(u)$ for μ. Recalling the condition $\bar{y} \in \operatorname{ri} \partial g(\bar{u})$, we get from (5.9) that $\bar{\mu} \in \operatorname{ri} \partial\vartheta(\Xi(\bar{u}))$. Thus, Proposition 5.7 demonstrates that ϑ is strictly twice epi-differentiable at $\Xi(\bar{u}) = 0$ for $\bar{\mu}$. We know from Proposition 5.3(b) that the dual condition in (5.8) is satisfied. Appealing now to [16, Proposition 7.1] tells us that $\mu \to \bar{\mu}$ as $(u, y) \to (\bar{u}, \bar{y})$. So, we conclude from Proposition 5.6 that $\mathrm{d}^2\vartheta(\Xi(u), \mu) \xrightarrow{e} \mathrm{d}^2\vartheta(\Xi(\bar{u}), \bar{\mu})$. It then follows from [50, Exercise 7.47(a)] that $\mathrm{d}^2\vartheta(\Xi(u), \mu) \circ \nabla\Xi(u) \xrightarrow{e} \mathrm{d}^2\vartheta(\Xi(\bar{u}), \bar{\mu}) \circ \nabla\Xi(\bar{u})$ if the condition

$$0 \in \operatorname{int}\big(K_\vartheta(\Xi(\bar{u}), \bar{\mu}) - \operatorname{rge}\nabla\Xi(\bar{u})\big)$$

is satisfied. We know from [3, Proposition 2.97] that this condition is equivalent to the dual condition in (5.8) with $(u, \mu) = (\bar{u}, \bar{\mu})$, which clearly holds under the nondegeneracy condition in (5.2). Finally, the sum rule for epi-convergence in [50, Theorem 7.46(b)] yields

$$\langle \mu, \nabla^2\Xi(u)(\cdot,\cdot)\rangle + \mathrm{d}^2\vartheta(\Xi(u), \mu)(\nabla\Xi(u)\cdot) \xrightarrow{e} \langle \bar{\mu}, \nabla^2\Xi(\bar{u})(\cdot,\cdot)\rangle + \mathrm{d}^2\vartheta(\Xi(\bar{u}), \bar{\mu})(\nabla\Xi(\bar{u})\cdot),$$

which is equivalent by (5.14) to $\mathrm{d}^2 g(u, y)$ epi-converging to $\mathrm{d}^2 g(\bar{u}, \bar{y})$ as $(u, y) \to (\bar{u}, \bar{y})$ in the set of pairs $(u, y) \in \operatorname{gph} \partial g$ for which g is twice epi-differentiable. Appealing to Proposition 5.6 confirms strict twice epi-differentiability of g at $\bar{u}$ for $\bar{y}$.

We now proceed with verifying similar conclusion for any pair $(u, y) \in \operatorname{gph} \partial g$ in some neighborhood of $(\bar{u}, \bar{y})$. Observe from Lemma 5.5 that there exists $\varepsilon > 0$ such that $\mu \in \operatorname{ri} \partial\vartheta(z)$ for all $(z, \mu) \in \big(\operatorname{gph} \partial\vartheta\big) \cap \mathbb{B}_\varepsilon(\Xi(\bar{u}), \bar{\mu})$. Take $(u, y) \in \operatorname{gph} \partial g$ sufficiently close to $(\bar{u}, \bar{y})$ so that g is reliably $\mathcal{C}^2$-decomposable at u, and that $(\Xi(u), \mu) \in \big(\operatorname{gph} \partial\vartheta\big) \cap \mathbb{B}_\varepsilon(\Xi(\bar{u}), \bar{\mu})$, where μ is the unique multiplier in $M(u, y)$ – the multiplier set defined in (5.6) associated with the composite representation (5.1) of g around u. The latter inclusion, which can be ensured via [16, Proposition 7.1], tells us that $\mu \in \operatorname{ri} \partial\vartheta(\Xi(u))$. Employing the equivalence from (5.9) again, shrinking the neighborhood around $\bar{u}$ if necessary, we have $y \in \operatorname{ri} \partial g(u)$. Repeating arguments carried out in the above proof for $(\bar{u}, \bar{y})$ demonstrates that g is strict twice epi-differentiable at u for y and hence proves (a). This completes the proof. □

Remark 5.9. Observe from the proof of Theorem 5.8 that had we assumed the reliable $\mathcal{C}^2$-decomposability of g only at $\bar{u}$, we would have ensured strict twice epi-differentiability of g in (a) and strict proto-differentiability of ∂g in (b) only at $\bar{u}$ for $\bar{y}$. We should also point out that the reliable $\mathcal{C}^2$-decomposability of g in a neighborhood of $\bar{u}$, used in Theorem 5.8, is not restrictive, since most functions, satisfying the reliable $\mathcal{C}^2$-decomposability, enjoy this property at all points in their domains; see, e.g., Corollary 5.10.

Theorem 5.8 is a far-reaching extension of our recent results in [14, Theorem 4.3] in which a similar characterization of strict twice epi-differentiability was achieved for polyhedral functions. We should point out that [15, Theorem 3.9] presented a similar result for a composite functions with the outer function being a polyhedral function and the nondegeneracy condition being satisfied. The latter result can be distilled from Theorem 5.8 using the observation in Example 5.1(b) for such composite functions.

Corollary 5.10. *Assume that* $(\overline{X}, \overline{Y}) \in \operatorname{gph} N_{\mathbf{S}^n_+}$. *Then the following properties are equivalent:*

(a) $\delta_{\mathbf{S}^n_+}$ *is strictly twice epi-differentiable at* X *for* Y *for any pair* $(X, Y) \in \operatorname{gph} N_{\mathbf{S}^n_+}$ *in a neighborhood of* $(\overline{X}, \overline{Y})$;

(b) $N_{\mathbf{S}^n_+}$ *is strictly proto-differentiable at* X *for* Y *for any pair* $(X, Y) \in \operatorname{gph} N_{\mathbf{S}^n_+}$ *in a neighborhood of* $(\overline{X}, \overline{Y})$;

(c) $\overline{Y} \in \operatorname{ri} N_{\mathbf{S}^n_+}(\overline{X})$;

(d) $\operatorname{rank} \overline{X} + \operatorname{rank} \overline{Y} = n$.

Proof. It is known that the cone of positive semidefinite matrics $\mathbf{S}^{\mathbf{n}}_+$ is $\mathcal{C}^2$-cone reducible – and thus its indicator function is reliably $\mathcal{C}^2$-decomposable – at any of its points; see [3, Example 3.140]. Employing now Theorem 5.8 justifies the equivalences of (a)-(c). The equivalence of (c) and (d) is well-known; see [3, page 320]. □

Several important consequences can be distilled from Theorem 5.8. We begin with a useful characterization of continuous differentiability of the proximal mapping of $\mathcal{C}^2$-decomposable functions.

Theorem 5.11. *Assume that $g : \mathbf{Y} \to \overline{\mathbb{R}}$, $(\bar u, \bar y) \in \operatorname{gph} \partial g$, and that there is a neighborhood U of $\bar u$ such that g is reliably $\mathcal{C}^2$-decomposable at any $u \in U$. If g is prox-bounded, then the following properties are equivalent:*

- (a) $\bar y \in \operatorname{ri} \partial g(\bar u)$;
- (b) *there exists a positive constant r such that for any $\gamma \in (0, 1/r)$, the proximal mapping $\operatorname{prox}_{\gamma g}$ is $\mathcal{C}^1$ around $\bar u + \gamma \bar y$;*
- (c) *there exists a positive constant r such that for any $\gamma \in (0, 1/r)$, the envelope function $e_\gamma g$ is $\mathcal{C}^2$ around $\bar u + \gamma \bar y$.*

In addition, if g is convex, the constant r in (b) and (c) can be taken as 0 with convention $1/0 = \infty$.

Proof. According to Proposition 5.3(a), g is both prox-regular and subdifferentially continuous at $\bar u$ for $\bar y$. Appealing to Theorem 5.8 tells us that (a) is equivalent to strict proto-differentiability of ∂g at u for y for any $(u, y) \in \operatorname{gph} \partial g$ sufficiently close to $(\bar u, \bar y)$. It follows from Theorem 4.7 that (a)-(c) are equivalent. □

For the indicator function of a $\mathcal{C}^2$-cone reducible set, the above result is simplified as recorded below.

Corollary 5.12. *Assume that $C \subset \mathbf{Y}$ is a nonempty closed convex subset, $\bar u \in C$, and that C is $\mathcal{C}^2$-cone reducible set at any u in a neighborhood of $\bar u$. Then the following properties are equivalent:*

- (a) $\bar y \in \operatorname{ri} N_C(\bar u)$;
- (b) *for any $\gamma > 0$, the proximal mapping $\operatorname{prox}_{\gamma g}$ is $\mathcal{C}^1$ around $\bar u + \gamma \bar y$;*
- (c) *for any $\gamma > 0$, the envelope function $e_\gamma g$ is $\mathcal{C}^2$ around $\bar u + \gamma \bar y$.*

In particular, the properties in (a)-(c) are equivalent when $C = \mathbf{S}^{\mathbf{n}}_+$.

Proof. The claimed equivalence of (a)-(c) results from Theorems 5.11 and 4.7. The last claim about $\mathbf{S}^{\mathbf{n}}_+$ falls out of Corollary 5.10. □

Note that Shapiro in [52, Proposition 3.1] characterized the differentiability of projection onto $\mathcal{C}^2$-cone reducible convex sets using a similar relative interior condition in Corollary 5.12(a). Shapiro's result was generalized in [26, Lemma 5.3.32] for $\mathcal{C}^2$-decomposable convex functions. Corollary 5.12 and Theorem 5.11, respectively, improve both latter results by showing that indeed differentiability can be strengthened to continuous differentiability. We should, however, point out that while the authors in [52] and [26] assume $\mathcal{C}^2$-cone reducibility and reliable $\mathcal{C}^2$-decomposability at the point under consideration, Corollary 5.12 and Theorem 5.11 requires that these properties be satisfied in a neighborhood of such a point. Also, it was shown in [7, Theorem 28] that the proximal mapping of any prox-bounded and prox-regular $\mathcal{C}^2$-partly smooth function (cf. [7, Definition 14]) satisfying the relative interior condition in Theorem 5.11(c) enjoys local $\mathcal{C}^1$-smoothness similar to that in Theorem 5.11(b). It was shown by Shapiro in [51] that reliably $\mathcal{C}^2$-decomposable functions are $\mathcal{C}^2$-partly smooth. This suggests that the implication (a) $\Longrightarrow$ (b) in Theorem 5.11 can be extracted from [7, Theorem 28]. Note that it is possible to show that $\mathcal{C}^2$-partly smooth functions are strictly twice epi-differentiable. This makes it possible to completely recover [7, Theorem 28] using our approach and even shows that the relative interior condition indeed characterizes the continuous differentiability of proximal mappings for this class of functions. We will leave this issue for our forthcoming paper [17].

Next, we provide a characterization of the existence of a continuously differentiable single-valued graphical localization of the solution mapping to the generalized equation

$$0 \in \psi(u) + \partial g(u), \tag{5.25}$$

where $\psi : \mathbf{Y} \to \mathbf{Y}$ is a continuously differentiable function and $g : \mathbf{Y} \to \overline{\mathbb{R}}$ is a proper function. When g enjoys the reliable $\mathcal{C}^2$-decomposability, we can apply Theorem 4.3 to obtain the result below. In what follows, we call a solution $\bar{u}$ to the generalized equation in (5.25) *nondegenerate* if $-\psi(\bar{u}) \in \operatorname{ri} \partial g(\bar{u})$.

Corollary 5.13. *Assume that $\bar{u}$ is a solution to the generalized equation in (5.25), that there is a neighborhood U of $\bar{u}$ such that g in (5.25) is reliably $\mathcal{C}^2$-decomposable at any $u \in U$. Then, the following properties are equivalent:*

(a) *the mapping $\psi + \partial g$ is metrically regular at $\bar{u}$ for 0 and $\bar{u}$ is a nondegenerate solution to (5.25);*

(b) *the solution mapping S, defined by*

$$S(y) := \big\{ u \in \mathbf{Y} \,\big|\, y \in \psi(u) + \partial g(u) \big\}, \quad y \in \mathbf{Y},$$

has a Lipschitz continuous single-valued localization s around $0 \in \mathbf{Y}$ for $\bar{u}$, which is $\mathcal{C}^1$ around 0.

In particular, the properties in (a) and (b) are equivalent when $g = \delta_{\mathbf{S}^{\mathbf{n}}_{+}}$.

Proof. The claimed equivalence results from Theorems 5.8 and 4.3. The last claim about $\mathbf{S}^{\mathbf{n}}_{+}$ falls out of Corollary 5.10. □

We close this section with an application of our main result in studying stability properties of the KKT system of the composite optimization problem

$$\text{minimize} \;\; \varphi(x) + (g \circ \Phi)(x) \quad \text{subject to} \quad x \in \mathbf{X}, \tag{5.26}$$

where $\varphi : \mathbf{X} \to \mathbb{R}$ and $\Phi : \mathbf{X} \to \mathbf{Y}$ are $\mathcal{C}^2$-smooth mappings and $g : \mathbf{Y} \to \overline{\mathbb{R}}$ is a convex function. We are going to impose the reliable $\mathcal{C}^2$-decomposability assumption on g to make use of our developments of strict proto-differentiability of subgradient mappings in this section. With that in mind, the composite program in (5.26) encompasses important classes of constrained and composite optimization problems including second-order cone programming problems, when $g = \delta_{\mathcal{Q}}$, where $\mathcal{Q}$ stands for the second-order cone, and semidefinite programming problems, when $g = \delta_{\mathbf{S}^{\mathbf{m}}_{+}}$. Given $(x, y) \in \mathbf{X} \times \mathbf{Y}$, define the Lagrangian of (5.26) by $L(x, y) := \varphi(x) + \langle y, \Phi(x) \rangle$. The KKT system associated with the composite problem (5.26) is given by

$$0 = \nabla_x L(x, y), \quad y \in \partial g(\Phi(x)). \tag{5.27}$$

Define the mapping $\Psi : \mathbf{X} \times \mathbf{Y} \rightrightarrows \mathbf{X} \times \mathbf{Y}$ by

$$\Psi(x, y) := \begin{bmatrix} \nabla_x L(x, y) \\ -\Phi(x) \end{bmatrix} + \begin{bmatrix} 0 \\ \partial g^*(y) \end{bmatrix}. \tag{5.28}$$

It can be easily seen that a pair $(\bar{x}, \bar{y}) \in \mathbf{X} \times \mathbf{Y}$ is a solution to the KKT system in (5.27) if and only if $(0, 0) \in \Psi(\bar{x}, \bar{y})$. Define also the solution mapping $S : \mathbf{X} \times \mathbf{Y} \rightrightarrows \mathbf{X} \times \mathbf{Y}$ to the canonical perturbed of the KKT system in (5.27) by

$$S_{KKT}(p, q) := \Psi^{-1}(p, q) = \big\{ (x, y) \in \mathbf{X} \times \mathbf{Y} \,\big|\, (p, q) \in \Psi(x, y) \big\}, \quad \text{for} \;\; (p, q) \in \mathbf{X} \times \mathbf{Y}.$$

The next result presents a characterization of strong metric regularity of the solution mapping S_{KKT} under a relative interior condition imposed for the Lagrange multiplier under consideration.

Theorem 5.14. *Assume that* $(\bar{x}, \bar{y})$ *is a solution to the KKT system in* (5.27) *and* $\bar{y} \in \operatorname{ri} \partial g(\bar{u})$ *with* $\bar{u} := \Phi(\bar{x})$, *that the convex function* g *in* (5.26) *is reliably* $\mathcal{C}^2$*-decomposable in a neighborhood of* $\bar{u}$. *Then the following properties are equivalent:*

(a) *the mapping* Ψ *is strongly metrically regular at* $(\bar{x}, \bar{y})$ *for* $(0, 0)$;

(b) *the mapping* Ψ *is metrically regular at* $(\bar{x}, \bar{y})$ *for* $(0, 0)$;

(c) *the mapping* Ψ *is strongly metrically subregular at* $(\bar{x}, \bar{y})$ *for* $(0, 0)$;

(d) *the solution mapping* S_{KKT} *has a Lipschitz continuous single-valued localization around* $(0, 0)$ *for* $(\bar{x}, \bar{y})$, *which is* $\mathcal{C}^1$ *in a neighborhood of* $(0, 0)$;

(e) *the implication*

$$\begin{cases} \nabla^2_{xx} L(\bar{x}, \bar{y})w - \nabla \Phi(\bar{x})^* w' = 0, \\ \nabla \Phi(\bar{x}) w \in K_g(\bar{u}, \bar{y}), \\ w' + \nabla^2 \langle \bar{\mu}, \Xi \rangle(\bar{u})(\nabla \Phi(\bar{x}) w) \in K_g(\bar{u}, \bar{y})^\perp \end{cases} \implies (w, w') = (0, 0)$$

holds, where $\bar{\mu} \in \mathbf{Z}$ *is the unique element in the Lagrange multiplier set* $M(\bar{u}, \bar{y})$ *from* (5.6).

Proof. Setting

$$\psi(x, y) := \begin{bmatrix} \nabla_x L(x, y) \\ -\Phi(x) \end{bmatrix} \quad \text{and} \quad f(x, y) := g^*(y), \quad (x, y) \in \mathbf{X} \times \mathbf{Y},$$

and observing via [50, Proposition 10.5] that

$$\partial f(x, y) = \{0\} \times \partial g^*(y), \tag{5.29}$$

we can equivalently write the KKT system in (5.27) as the generalized equation

$$0 \in \psi(x, y) + \partial f(x, y).$$

Since $(\bar{x}, \bar{y})$ is a solution to the KKT system in (5.27), it is a solution to the above generalized equation. It follows from Theorem 5.8 and $\bar{y} \in \operatorname{ri} \partial g(\bar{u})$ that there is $\varepsilon > 0$ such that ∂g is strictly proto-differentiable at u for y for any $(u, y) \in (\operatorname{gph} \partial g) \cap \mathbb{B}_\varepsilon(\bar{u}, \bar{y})$. Since $\partial g^* = (\partial g)^{-1}$ due to the convexity of g, we deduce that ∂g^* is strictly proto-differentiable at y for u for any $(y, u) \in (\operatorname{gph} \partial g^*) \cap \mathbb{B}_\varepsilon(\bar{y}, \bar{u})$. By (5.29), we have

$$(x, y, v, u) \in \operatorname{gph} \partial f \iff (x, v, y, u) \in \mathbf{X} \times \{0\} \times \operatorname{gph} \partial g^*.$$

Pick $(x, y, v, u) \in \operatorname{gph} \partial f$ so that $(y, u) \in (\operatorname{gph} \partial g^*) \cap \mathbb{B}_\varepsilon(\bar{y}, \bar{u})$. Using the definition of the regular tangent cone, we obtain

$$(\xi_1, \xi_2, \eta_1, \eta_2) \in \widehat{T}_{\operatorname{gph} \partial f}(x, y, v, u) \iff (\xi_1, \eta_1, \xi_2, \eta_2) \in \mathbf{X} \times \{0\} \times \widehat{T}_{\operatorname{gph} \partial g^*}(y, u).$$

Similarly, it is not hard to see via the definition of the paratingent cone that

$$(\xi_1, \xi_2, \eta_1, \eta_2) \in \widetilde{T}_{\operatorname{gph} \partial f}(x, y, w, u) \iff (\xi_1, \eta_1, \xi_2, \eta_2) \in \mathbf{X} \times \{0\} \times \widetilde{T}_{\operatorname{gph} \partial g^*}(y, u).$$

Since ∂g^* is strictly proto-differentiable at y for u, we have $\widehat{T}_{\operatorname{gph} \partial g^*}(y, u) = \widetilde{T}_{\operatorname{gph} \partial g^*}(y, u)$. Thus, we arrive at $\widehat{T}_{\operatorname{gph} \partial f}(x, y, v, u) = \widetilde{T}_{\operatorname{gph} \partial f}(x, y, v, u)$, which is equivalent to saying that ∂f is strictly proto-differentiable at (x, y) for (v, u) whenever $(y, u) \in (\operatorname{gph} \partial g^*) \cap \mathbb{B}_\varepsilon(\bar{y}, \bar{u})$. Observe also that f is clearly prox-regular and subdifferentially continuous at $(\bar{x}, \bar{y})$ for $(0, \bar{u})$, since g is convex. Appealing now to Theorem 4.3 demonstrates that (a), (b), and (d) are equivalent. On the other hand, it is easy to see that $\nabla \psi(\bar{x}, \bar{y}) = \nabla \psi(\bar{x}, \bar{y})^*$. Thus, it results from Corollary 4.5 that (b) and (c) are also equivalent.

Turning to (e), recall that ∂f is strictly proto-differentiable at $(\bar{x}, \bar{y})$ for $(0, \bar{u})$. By Theorem 3.9 and (5.29), we conclude for any $(w, w') \in \mathbf{X} \times \mathbf{Y}$ that

$$D^*(\partial f)\big((\bar{x}, \bar{y}), (0, \bar{u})\big)(w, w') = D(\partial f)\big((\bar{x}, \bar{y}), (0, \bar{u})\big)(w, w') = \{0\} \times D(\partial g^*)(\bar{y}, \bar{u})(w').$$

Using this, the sum rule for coderivatives from [50, Exercise 10.43(b)], we get for any $(w, w') \in \mathbf{X} \times \mathbf{Y}$ that

$$\begin{aligned} D^*(\psi + \partial f)\big((\bar{x}, \bar{y}), (0, 0)\big)(w, w') \\ &= \nabla \psi(\bar{x}, \bar{y})^*(w, w') + D^*(\partial f)\big((\bar{x}, \bar{y}), (0, \Phi(\bar{x}))\big)(w, w') \\ &= \nabla \psi(\bar{x}, \bar{y})^*(w, w') + D(\partial f)\big((\bar{x}, \bar{y}), (0, \Phi(\bar{x}))\big)(w, w') \\ &= \big(\nabla^2_{xx} L(\bar{x}, \bar{y}) w - \nabla \Phi(\bar{x})^* w', \nabla \Phi(\bar{x}) w\big) + \{0\} \times D(\partial g^*)(\bar{y}, \bar{u})(w') \\ &= \big(\nabla^2_{xx} L(\bar{x}, \bar{y}) w - \nabla \Phi(\bar{x})^* w', \nabla \Phi(\bar{x}) w\big) + \{0\} \times D(\partial g)(\bar{u}, \bar{y})^{-1}(w'), \end{aligned}$$

where the last equality comes from $\partial g^* = (\partial g)^{-1}$. By [31, Theorem 3.3(ii)], the mapping $\Psi = \psi + \partial f$ is metrically regular at $(\bar{x}, \bar{y})$ for $(0, 0)$ if and only if the implication

$$(0, 0) \in D^*\Psi\big((\bar{x}, \bar{y}), (0, 0)\big)(w, w') = D^*(\psi + \partial f)\big((\bar{x}, \bar{y}), (0, 0)\big)(w, w') \implies w = 0,\ w' = 0$$

is satisfied. To show that this implication is the same as the one in (e), assume $(0, 0) \in D^*\Psi\big((\bar{x}, \bar{y}), (0, 0)\big)(w, w')$. By the calculation above, we get $\nabla^2_{xx} L(\bar{x}, \bar{y}) w - \nabla \Phi(\bar{x})^* w' = 0$ and $-\nabla \Phi(\bar{x}) w \in D(\partial g)(\bar{u}, \bar{y})^{-1}(w')$. The latter amounts to the inclusion $w' \in D(\partial g)(\bar{u}, \bar{y})(-\nabla \Phi(\bar{x}) w)$. Employing [16, Theorem 6.2(b)] tells us that

$$D(\partial g)(\bar{u}, \bar{y})(-\nabla \Phi(\bar{x}) w) = -\nabla^2 \langle \bar{\mu}, \Xi \rangle(\bar{u})(\nabla \Phi(\bar{x}) w) + N_{K_g(\bar{u}, \bar{y})}(-\nabla \Phi(\bar{x}) w).$$

Since $K_g(\bar{u}, \bar{y})$ is a linear subspace of $\mathbf{Y}$ due to $\bar{y} \in \operatorname{ri} \partial g(\bar{u})$, we have

$$N_{K_g(\bar{u}, \bar{y})}(-\nabla \Phi(\bar{x}) w) = \begin{cases} K_g(\bar{u}, \bar{y})^\perp & \text{if } \nabla \Phi(\bar{x}) w \in K_g(\bar{u}, \bar{y}), \\ \emptyset & \text{otherwise.} \end{cases}$$

Combining these shows that (e) is equivalent to (b) and hence completes the proof. □

For classical nonlinear programming problems (NLPs), it is well-known that metric regularity and strong metric regularity of KKT systems are equivalent; see [9, Theorem 4I.2] and [20, Section 7.5]. By using a new approach, Theorem 5.14 extends this result for the composite problem (5.26) under a relative interior condition. This extra condition allows us to demonstrate further that the Lipschitz continuous single-valued localization of the solution mapping to the KKT system of (5.26) is continuously differentiable. This can be viewed as an extension of Fiacco and McCormick's result in [12] for NLPs, which was achieved under the classical second-order sufficient condition, strict complementarity condition, and linear independence constraint qualification.

References

[1] Aubin, J.-P., Frankowska, H.: Set-Valued Analysis. Birkhäuser, Boston (2009)

[2] Benko, M., Mehlitz, P.: Why second-order sufficient conditions are, in a way, easy – or – revisiting calculus for second subderivatives. J. Convex Anal. 30, 541–589 (2023)

[3] Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer New York, New York (2000)

[4] Borwein, J.M., Lewis A.S.: Convex Analysis and Nonlinear Optimization: Theory and Examples. 2nd edition, Springer-Verlag, New York (2006)

[5] Chieu, N.H., Hien, L.V., Nghia, T.T.A., Tuan, H.A.: Quadratic growth and strong metric subregularity of the subdifferential via subgradient graphical derivative. SIAM J. Optim. 31, 545–568 (2021)

[6] Correa, R., Salas, D., Thibault, L.: Smoothness of the metric projection onto nonconvex bodies in Hilbert spaces. J. Math. Anal. Appl. 457, 1307–1332 (2018)

[7] Daniilidis, A., Hare, W., Malick, J.: Geometrical interpretation of the predictor-corrector type algorithms in structured optimization problems. Optimization 55, 481–503 (2006)

[8] Dontchev, A.L., Rockafellar, R.T.: Characterizations of strong regularity for variational inequalities over polyhedral convex sets. SIAM J. Optim. 6, 1087–1105 (1996)

[9] Dontchev, A.L., Rockafellar, R.T.: Implicit Functions and Solution Mappings: A View from Variational Analysis. Springer New York, New York (2014)

[10] Drusvyatskiy, D., Mordukhovich, B.S., Nghia, T.T.A.: Second-order growth, tilt stability, and metric regularity of the subdifferential. J. Convex Anal. 21, 1165-1192 (2014)

[11] Facchinei, F., Pang, J.-S.: Finite-Dimesional Variational Inequalities and Complementarity Problems, Springer New York, New York (2003)

[12] Fiacco, A.V., McCormick, G.P.: Nonlinear Programming: Sequential Unconstrained Minimization Techniques. John Wiley & Sons, New York (1968)

[13] Gfrerer, H., Outrata, J. V.: On (local) analysis of multifunctions via subspaces contained in graphs of generalized derivatives. J. Math. Anal. Appl. 508 (2022), doi.org/10.1016/j.jmaa.2021.125895

[14] Hang, N.T.V., Jung, W., Sarabi, M.E.: Role of subgradients in variational analysis of polyhedral functions. ArXiv. https://arxiv.org/abs/2207.07470 (2022).

[15] Hang, N.T.V., Sarabi, M.E.: A chain rule for strict twice epi-differentiability and its applications. To appear in SIAM J. Optim., ArXiv. https://arxiv.org/abs/2209.01489 (2022).

[16] Hang, N.T.V., Sarabi, M.E.: Convergence of augmented Lagrangian methods for composite optimization problems. ArXiv. https://arxiv.org/abs/2307.15627 (2023).

[17] Hang, N.T.V., Sarabi, M.E.: A fresh look into variational analysis of $\mathcal{C}^2$-partly smooth functions. preprint, (2023).

[18] Holmes, R.B.: Smoothness of certain metric projections on Hilbert space. Trans. Amer. Math. Soc. 183, 87–100 (1973)

[19] Ioffe, A.D.: Variational Analysis of Regular Mappings: Theory and Applications. Springer, Cham, Switzerland (2017)

[20] Klatte D., Kummer B.: Nonsmooth Equations in Optimization: Regularity, Calculus, Methods and Applications. Springer New York, New York (2002)

[21] Lemaréchal, Sagastizábal, C.: More than first-order developments of convex functions: primal-dual relations. J. Convex Anal. 3, 255–268 (1996)

[22] Lewis, A.S.: Nonsmooth analysis of eigenvalues. Math. Program. 84, 1–24 (1999)

[23] Lewis, A.S.: Active sets, nonsmoothness, and sensitivity. SIAM J. Optim. 13, 702–725 (2002)

[24] Lewis A. S., Wylie C.: Active-set Newton methods and partial smoothness, Math. Oper. Res., 46 (2021), pp. 712–725.

[25] Lewis, A.S, Zhang, S.: Partial smoothness, tilt stability, and generalized Hessians. SIAM J. Optim. 23, 74–94 (2013)

[26] Milzarek, A.: Numerical Methods and Second Order Theory for Nonsmooth Problems. Ph.D. Dissertation. University of Munich (2016)

[27] Mohammadi, A., Mordukhovich, B.S., Sarabi, M.E.: Parabolic regularity via geometric variational analysis. Trans. Amer. Soc. 374, 1711–1763 (2021)

[28] Mohammadi, A., Mordukhovich, B.S., Sarabi, M.E.: Variational analysis of composite models with applications to continuous optimization. Math. Oper. Res. 47, 397–426 (2022)

[29] Mohammadi, A., Sarabi, M.E.: Twice epi-differentiability of extended-real-valued functions with applications in composite optimization. SIAM J. Optim. 30, 2379–2409 (2020)

[30] Mohammadi, A., Sarabi, M.E.: Parabolic regularity of spectral functions. Part I: Theory. Arxiv. arXiv:2301.04240 (2023).

[31] Mordukhovich, B.S.: Variational Analysis and Applications. Springer, Cham, Switzerland (2018)

[32] Mordukhovich, B.S., Nam, N.M.: An Easy Path To Convex Analysis and Applications. Springer Cham (2023)

[33] Mordukhovich, B.S., Nghia, T.T.A.: Second-order characterizations of tilt stability with applications to nonlinear programming. Math. Program. 149, 83–104 (2015)

[34] Mordukhovich, B.S., Sarabi, M.E.: Generalized differentiation of piecewise linear functions in second-order variational analysis. Nonlinear Anal. 132, 240–273 (2016)

[35] Mordukhovich, B.S., Sarabi, M.E.: Second-order analysis of piecewise linear functions with applications to optimization and stability. J. Optim. Theory Appl. 171, 504–526 (2016)

[36] Mordukhovich, B.S., Sarabi, M.E.: Critical multipliers in variational systems via second-order generalized differentiation. Math. Program. 169, 605–648 (2018)

[37] Mordukhovich, B.S., Sarabi, M.E.: Criticality of Lagrange multipliers in variational systems. SIAM J. Optim. 29, 1524–1557 (2019)

[38] Ndoutoume, J.L., Théra, M.: Generalised second-order derivatives of convex functions in reflexive Banach spaces. Bull. Austral. Math. Soc. 51, 55–72 (1995)

[39] Poliquin, R.A., Rockafellar, R.T.: Second-order nonsmooth analysis in nonlinear programming. In: Du, D., Qi L., Womersley, R.S. (eds) Recent Advances in Optimization, pp. 322–350. World Scientific (1995)

[40] Poliquin, R.A., Rockafellar, R.T.: Prox-regular functions in variational analysis. Trans. Amer. Math. Soc. 348, 1805–1838 (1996)

[41] Poliquin, R.A., Rockafellar, R.T.: Generalized Hessian properties of regularized nonsmooth functions. SIAM J. Optim. 6, 1121–1137 (1996)

[42] Poliquin, R.A., Rockafellar, R.T.: Tilt stability of a local minimum. SIAM J. Optim. 8, 287–299 (1998)

[43] Robinson, S.M.: Local structure of feasible sets in nonlinear programming. Part II: Nondegeneracy. Math. Program. Studies 22, 217–230 (1984)

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

[45] Robinson, S.M.: Normal maps induced by linear transformations. Math. Oper. Res. 17, 691–714 (1992)

[46] Rockafellar, R.T.: Convex Ananlysis. Priceton University Press, Princeton, New Jersey (1970)

[47] Rockafellar R. T.: Maximal monotone relations and the second derivatives of nonsmooth functions, Ann. Inst. H. Poincaré Anal. Non Linéaire 2, 167–184 (1985)

[48] Rockafellar, R.T.: First- and second-order epi-differentiability in nonlinear programming. Trans. Amer. Math. Soc. 307, 75–108 (1988)

[49] Rockafellar, R.T.: Proto-differentiability of set-valued mappings and its applications in optimization. Ann. Inst. H. Poincaré Anal. Non Linnéaire 6, 449–482 (1989)

[50] Rockafellar, R.T., Wets, R.J-B.: Variational Analysis. Springer, Berlin (1998)

[51] Shapiro, A.: On a class of nonsmooth composite functions. Math. Oper. Res. 28, 677–692 (2003)

[52] Shapiro, A.: Differentiability properties of metric projections onto convex sets. J. Optim. Theory Appl. 169, 953–964 (2016)

[53] Torki, M.: First- and second-order epi-differentiability in eigenvalue optimization. J. Math. Anal. Appl. 234, 391–416 (1998)