A FRESH LOOK INTO VARIATIONAL ANALYSIS OF $\mathcal{C}^2$-PARTLY SMOOTH FUNCTIONS

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

Abstract. $\mathcal{C}^2$-partial smoothness of functions has been an important subject of research in optimization, on both theoretical and algorithmic aspects, since it was first introduced by Lewis in [18]. Our work aims at providing a fresh variational analysis viewpoint on the class of $\mathcal{C}^2$-partly smooth functions. Namely, we explore the relationship between $\mathcal{C}^2$-partial smoothness and strict twice epi-differentiability and demonstrate that functions from the latter class are always strictly twice epi-differentiable. On the other hand, we provide two examples to show that the opposite conclusion does not hold in general. As a consequence of our analysis, we calculate the second subderivative of $\mathcal{C}^2$-partly smooth functions. Applications to stability analysis of related generalized equations involving a general perturbation and to asymptotic analysis of the well-known sample average approximation method for stochastic programs with $\mathcal{C}^2$-partly smooth regularizers are also given.

Keywords. $\mathcal{C}^2$-partial smoothness, prox-regularity, strict proto-differentiability, differential stability, (strong) metric regularity, sample average approximation method.

Mathematics Subject Classification (2000) 49J52, 65K10, 90C15, 90C31

1 Introduction

This paper is a continuation of our recent work [11] on strict proto-differentiability of subgradient mappings and its applications in stability analysis of generalized equations, with an emphasis on $\mathcal{C}^2$-*partly smooth* functions. The notion of partial smoothness, first formalized in [18] by Lewis, expresses an underlying smooth structure of a nonsmooth function on a smooth manifold and offers a powerful framework for both stability and algorithmic analysis of optimization problems. Therefore, there has been considerable efforts to study various theoretical aspects (cf. [2, 3, 7, 12, 14, 20, 23, 34]) as well as numerical advantages (cf. [1, 2, 4, 12, 19, 21, 22, 34]) of $\mathcal{C}^2$-partly smooth functions.

Partial smoothness is widely satisfied by important classes of functions in optimization and variational analysis, including the polyhedral and pointwise maximum functions [18], $\mathcal{C}^2$-decomposable functions [32], and spectral functions [3, 18]. Calculus rules for this property, established in [18, Section 4] and [12, Chapter 3], furthermore prove the abundance of partly smooth functions in these areas.

Motivated by the aforementioned applications, we plan to study new second-order variational properties of $\mathcal{C}^2$-partly smooth functions. More precisely, we follow our recent developments in [9–11], where novel second-order variational properties, namely strict twice epi-differntiability and strict proto-differentiability, of important classes of composite functions were brought to

the surface. We also paid special attention to continuous differentiability of the proximal mappings and characterizations of regularity properties of certain generalized equations. Similar results about the proximal mappings were established for $\mathcal{C}^2$-partly smooth functions in [4, Theorem 28] and [34, Corollaries 4.4.2 and 4.4.3]. In addition, in [11, Section 5], strict twice epi-differentiability was verified for a subclass of $\mathcal{C}^2$-partly smooth functions, called therein reliably $\mathcal{C}^2$-decomposable functions; see Example 3.2 for the definition of this class of functions. These inspire us to explore further the relationships between $\mathcal{C}^2$-partial smoothness and strict twice epi-differentiability of functions. We should add here that strict twice epi-differentiability was defined by Rockafellar and Poliquin in [27] and was further studied for prox-regular functions in [28]. In particular, a characterization of this concept in term of epigraphical convergence of the second subderivative of prox-regular functions was achieved in [28, Theorem 4.1], which will be leveraged in this paper to show under a relative interior condition for the chosen subgradient that $\mathcal{C}^2$-partly smooth functions are always strictly twice epi-differentiable in a neighborhood of a given point relative to the graph of subgradient mappings. Moreover, our proof allows us to calculate the second subderivative of $\mathcal{C}^2$-partly smooth functions. We also provide two examples to show that there are strictly twice epi-differentiable functions that are not $\mathcal{C}^2$-partly smooth. Not only does this enable us to extend the scope of our developments in [11] to a larger class of functions but also provides a fresh viewpoint to investigate such important functions in optimization and variational analysis. We then concentrate on stability analysis of a generalized equation related to subgradient mapping of a $\mathcal{C}^2$-partly smooth function under a general perturbation and provide sufficient conditions for its solution mapping to have a single-valued Lipschitz continuous graphical localization. Using this, we analyze asymptotic behavior of the sequence, generated by a sample average approximation (SAA) method, for certain regularized stochastic programs.

The rest of the paper is organized as follows. Section 2 recaps definitions of important concepts used in this paper. Section 3 begins with reviewing the notion of partial smoothness. We then present our main findings about strict twice epi-differentiability of $\mathcal{C}^2$-partly smooth functions and regularity properties of the corresponding subgradient mappings. As important consequences, we establish next in this section $\mathcal{C}^1$-smoothness and active manifold identifiability of proximal mappings as well as differential stability for generalized equations involving a general perturbation. The last section, Section 4, concerns asymptotic analysis of an SAA method for stochastic programs with $\mathcal{C}^2$-partly smooth regularizers.

2   Tools of Variational Analysis

Throughout, suppose that $\mathbf{X}$ and $\mathbf{Y}$ are finite dimentional 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 $r > 0$. Given a nonempty set $C \subset \mathbf{X}$, the symbols $\mathrm{ri}\,C$, $\mathrm{aff}\,C$, C^*, and $\mathrm{par}\,C$ signify its relative interior, affine span, polar cone, and the linear subspace parallel to $\mathrm{aff}\,C$, respectively. The indicator function δ_C of the set C is defined by $\delta_C = 0$ for $x \in C$ and $\delta_C(x) = \infty$ for $x \in \mathbf{X} \setminus C$. We denote by P_C the projection mapping onto C and by $\mathrm{dist}(x, C)$ the distance between $x \in \mathbf{X}$ and the set C. The domain, range, and graph of a set-valued mapping $F : \mathbf{X} \rightrightarrows \mathbf{Y}$ are defined, respectively, by $\mathrm{dom}\,F = \{x \in \mathbf{X} \,|\, F(x) \neq \emptyset\}$, $\mathrm{rge}\,F = \{y \in \mathbf{Y} \,|\, \exists\,x \in \mathbf{X} \text{ with } y \in F(x)\}$, and $\mathrm{gph}\,F = \{(x, y) \in \mathbf{X} \times \mathbf{Y} \,|\, y \in F(x)\}$. Consider an extended-real-valued function $\varphi : \mathbf{X} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. We denote by $\mathrm{epi}\,\varphi$ its epigraph set given by $\{(x, \alpha) \in \mathbf{X} \times \mathbb{R} \,|\, \alpha \geq \varphi(x)\}$. Assuming that φ is twice differentiable at $\bar{x}$ with Hessian $\nabla^2\varphi(\bar{x})$, we also use the same notation $\nabla^2\varphi(\bar{x})$ to denote the self-adjoint linear operator of $\mathbf{X}$ associated with the symmetric bilinear form $\nabla^2\varphi(\bar{x}) : \mathbf{X} \times \mathbf{X} \to \mathbb{R}$.

Let $\{C^t\}_{t>0}$ be a parameterized family of sets in $\mathbf{X}$. Its inner and outer limit sets are defined,

respectively, by

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

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

see [31, 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 $\operatorname{epi} f^k \to \operatorname{epi} f$ in $\mathbf{X} \times \mathbb{R}$ as $k \to \infty$; see [31, Definition 7.1] for more details on the epi-convergence of a sequence of extended-real-valued functions. 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 cone to Ω at $\bar{x}$, denoted $T_\Omega(\bar{x})$, is defined by $T_\Omega(\bar{x}) = \limsup_{t \searrow 0} \frac{\Omega - \bar{x}}{t}$. The regular/Fréchet normal cone $\widehat{N}_\Omega(\bar{x})$ to Ω at $\bar{x}$ is defined by $\widehat{N}_\Omega(\bar{x}) = T_\Omega(\bar{x})^*$. For $x \notin \Omega$, we set $\widehat{N}_\Omega(x) = \emptyset$. The limiting/Mordukhovich normal cone $N_\Omega(\bar{x})$ to Ω at $\bar{x}$ is the collection of all vectors $\bar{v} \in \mathbf{X}$ for which there exist sequences $\{x^k\}_{k \in \mathbb{N}}$ 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, the subderivative function $\mathrm{d}f(\bar{x}) \colon \mathbf{X} \to \overline{\mathbb{R}}$ is defined by

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

A vector $v \in \mathbf{X}$ is called a subgradient of f at $\bar{x}$ if $(v, -1) \in N_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))$. The set of all sub-gradients of f at $\bar{x}$ is denoted by $\partial f(\bar{x})$. Replacing the limiting normal cone with $\widehat{N}_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))$ in the definition of $\partial f(\bar{x})$ gives us $\widehat{\partial} f(\bar{x})$, the regular subdifferential of f at $\bar{x}$. The function f is called subdifferentially regular at $\bar{x}$ provided that $N_{\operatorname{epi} f}(\bar{x}, f(\bar{x})) = \widehat{N}_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))$. By defi-nition, subdifferential regularity of f at $\bar{x}$ particularly ensures the coincidence $\partial f(\bar{x}) = \widehat{\partial} f(\bar{x})$. The critical cone of f at $\bar{x}$ for $\bar{v}$ with $\bar{v} \in \partial f(\bar{x})$ is defined by

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

When $f = \delta_\Omega$ for 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, one has $\mathrm{d}\delta_\Omega(\bar{x}) = \delta_{T_\Omega(\bar{x})}$, the above definition of the critical cone of a function therefore boils down to the well-known concept of critical cone to a set (see [6, page 109]), namely $K_\Omega(\bar{x}, \bar{v}) = T_\Omega(\bar{x}) \cap [\bar{v}]^\perp$, where $[v]$ denotes the one-dimensional subspace $\{tv \,|\, t \in \mathbb{R}\}$. If f is subdifferentially regular and $\partial f(\bar{x}) \neq \emptyset$, then $\mathrm{d}f(\bar{x})$ is the support function of $\partial f(\bar{x})$ (cf. [31, Theorem 8.30]) and the critical cone $K_f(\bar{x}, \bar{v})$ can be equivalently described by

$$K_f(\bar{x}, \bar{v}) = N_{\partial f(\bar{x})}(\bar{v}). \tag{2.1}$$

The second subderivative of f at $\bar{x}$ for $\bar{v}$, denoted $\mathrm{d}^2 f(\bar{x}, \bar{v})$, is an extended-real-valued function 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 $\Delta_t^2 f(\bar{x}, \bar{v})$, $t > 0$, is the parametric family of second-order difference quotients of f at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ defined by

$$\Delta_t^2 f(\bar{x}, \bar{v})(w) = \frac{f(\bar{x} + tw) - f(\bar{x}) - t\langle \bar{v}, w \rangle}{\frac{1}{2} t^2}$$

3

for any $w \in \mathbf{X}$ and $t > 0$. According to [31, Definition 13.6], f is said to be 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 $t \searrow 0$. Further, we say that f is strictly twice epi-differentiable at $\bar{x}$ for $\bar{v}$ if the functions $\Delta_t^2 f(x, v)$ epi-converge to some function as $t \searrow 0$ and $(x, v) \to (\bar{x}, \bar{v})$ with $f(x) \to f(\bar{x})$ and $(x, v) \in \mathrm{gph}\, \partial f$. If this condition holds, the limit function is then the second subderivative $\mathrm{d}^2 f(\bar{x}, \bar{v})$.

Consider a set-valued mapping $F : \mathbf{X} \rightrightarrows \mathbf{Y}$. According to [31, Definition 8.33], the graphical derivative of F at $\bar{x}$ for $\bar{y}$ with $(\bar{x}, \bar{y}) \in \mathrm{gph}\, F$ is the set-valued mapping $DF(\bar{x}, \bar{y}) : \mathbf{X} \rightrightarrows \mathbf{Y}$ defined via the tangent cone to $\mathrm{gph}\, F$ at $(\bar{x}, \bar{y})$ by $\mathrm{gph}\, DF(\bar{x}, \bar{y}) = T_{\mathrm{gph}\, F}(\bar{x}, \bar{y})$. Using the definition of the tangent cone, we can alternatively represent $\mathrm{gph}\, DF(\bar{x}, \bar{y})$ in terms of graphical limit as

$$\mathrm{gph}\, DF(\bar{x}, \bar{y}) = \limsup_{t \searrow 0} \frac{\mathrm{gph}\, F - (\bar{x}, \bar{y})}{t}. \tag{2.2}$$

The set-valued mapping F is said to be proto-differentiable at $\bar{x}$ for $\bar{y}$ if the outer graphical limit in (2.2) is actually a full limit. 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 $\nabla F(\bar{x})$. Recall from [31, Definition 9.53] that the strict graphical derivative of a set-valued mapping F at $\bar{x}$ for $\bar{y}$ with $(\bar{x}, \bar{y}) \in \mathrm{gph}\, F$, is the set-valued mapping $\widetilde{D}F(\bar{x}, \bar{y}) : \mathbf{X} \rightrightarrows \mathbf{Y}$, defined by

$$\mathrm{gph}\, \widetilde{D}F(\bar{x}, \bar{y}) = \limsup_{\substack{t \searrow 0 \\ (x,y) \xrightarrow{\mathrm{gph}\, F} (\bar{x}, \bar{y})}} \frac{\mathrm{gph}\, F - (x, y)}{t}. \tag{2.3}$$

We say that F is strictly proto-differentiable at $\bar{x}$ for $\bar{y}$ if the outer graphical limit in (2.3) is attained as a full limit, or, equivalently, if its graph $\mathrm{gph}\, F$ is strictly smooth at $(\bar{x}, \bar{y})$, cf. [30, page 173]. It is obvious from definitions that $\mathrm{gph}\, DF(\bar{x}, \bar{y}) \subset \mathrm{gph}\, \widetilde{D}F(\bar{x}, \bar{y})$ in general. Strict proto-differentiability of F at $\bar{x}$ for $\bar{y}$ particularly implies that the latter inclusion must hold as equality, that is $\widetilde{D}F(\bar{x}, \bar{y})$ and $DF(\bar{x}, \bar{y})$ coincide.

3 Strict Twice Epi-Differentiability of $\mathcal{C}^2$-Partly Smooth Functions

This section aims at establishing new second-order variational properties of $\mathcal{C}^2$-partly smooth functions as well as their applications in stability analysis of related variational systems. We begin with reviewing this notion and some of its well-known consequences established in [18,20]. Recall that a set $\mathcal{M} \subset \mathbf{X}$ is called a $\mathcal{C}^2$-smooth manifold of codimension $m \leq \dim \mathbf{X}$ around a point $\bar{x} \in \mathbf{X}$ if $\bar{x} \in \mathcal{M}$ and there is an open neighborhood $\mathcal{O} \subset \mathbf{X}$ of $\bar{x}$ and a $\mathcal{C}^2$-smooth mapping $\Phi : \mathcal{O} \to \mathbb{R}^m$ such that

$$\mathcal{M} \cap \mathcal{O} = \left\{ x \in \mathcal{O} \,\middle|\, \Phi(x) = 0 \right\} \quad \text{and} \quad \nabla \Phi(\bar{x}) : \mathbf{X} \to \mathbb{R}^m \text{ is surjective.} \tag{3.1}$$

We should stress that the surjectivity of $\nabla \Phi(\bar{x})$, assumed above, yields the validity of the same property of $\nabla \Phi(x) : \mathbf{X} \to \mathbb{R}^m$ for all x in an open neighborhood of $\bar{x}$. Thus, we can assume without loss of generality that $\nabla \Phi(x)$ is surjective for all $x \in \mathcal{M} \cap \mathcal{O}$. It is well known (cf. [31, Example 6.8]) that the tangent and normal cones to $\mathcal{M}$ at $x \in \mathcal{M} \cap \mathcal{O}$ are linear subspaces of $\mathbf{X}$, given by $T_{\mathcal{M}}(x) = \ker \nabla \Phi(x)$ and $N_{\mathcal{M}}(x) = \mathrm{rge}\, \nabla \Phi(x)^*$, respectively. We should point out that these two orthogonal complement linear subspaces are independent of the choice of Φ, and that the critical cone to $\mathcal{M}$ at enjoys the relationship

$$K_{\mathcal{M}}(x, y) = T_{\mathcal{M}}(x) \tag{3.2}$$

4

for all $x \in \mathcal{M} \cap \mathcal{O}$ and $y \in N_{\mathcal{M}}(x)$; see (2.1) for the definition of the critical cone.

Definition 3.1 ($\mathcal{C}^2$-partial smoothness, cf. [20, Definition 3.2]). A function $f : \mathbf{X} \to \overline{\mathbb{R}}$ is said to be $\mathcal{C}^2$-partly smooth at $\bar{x}$ relative to a set $\mathcal{M} \subset \mathbf{X}$ containing $\bar{x}$ if $\mathcal{M}$ is a $\mathcal{C}^2$-smooth manifold around $\bar{x}$ and the following properties hold:

   (a) (restricted smoothness) $f|_{\mathcal{M}}$ is $\mathcal{C}^2$-smooth around $\bar{x}$, that is, there exists a representative function $\widehat{f} : \mathbf{X} \to \mathbb{R}$ which is $\mathcal{C}^2$-smooth around $\bar{x}$ with $\widehat{f}|_{\mathcal{M}} = f|_{\mathcal{M}}$ locally around $\bar{x}$;
   (b) (regularity) at every point $x \in \mathcal{M}$ close to $\bar{x}$, the function f is subdifferentially regular and has nonempty subdifferential $\partial f(x)$;
   (c) (normal sharpness) $N_{\mathcal{M}}(\bar{x}) = \operatorname{par}\{\partial f(\bar{x})\}$;
   (d) (subdifferential continuity) the subgradient mapping ∂f is continuous at $\bar{x}$ relative to $\mathcal{M}$ in the sense that the limit set $\lim_{x \in \mathcal{M}, x \to \bar{x}} \partial f(x)$ exists and equals to $\partial f(\bar{x})$.

The manifold $\mathcal{M}$ is often referred to as the *active manifold* at the point $\bar{x}$, and also as the *optimal manifold* when $\bar{x}$ is actually a minimizer of the function f.

Observe that not only does a $\mathcal{C}^2$-partly smooth function behave smoothly along some manifold according to condition (a), but also satisfies regularity and normal sharpness properties in items (b)–(d). These additional requirements create a strong bond between the function and its active manifold. In fact, the active manifold associated with a $\mathcal{C}^2$-partly smooth function at a given point is locally unique in the presence of a suitable additional regularity, see the discussion prior to Proposition 3.7. Note also that conditions (a), (b), and $\mathcal{M}$ being a $\mathcal{C}^2$-smooth manifold in the above definition are properties required in a neighborhood of the point $\bar{x}$ in question. On the other hand, the normal sharpness condition (c) was proved to be stable in the sense that if the function f is $\mathcal{C}^2$-partly smooth at $\bar{x}$ relative to $\mathcal{M}$, then the normal sharpness property in Definition 3.1(c) must be satisfied at all nearby points $x \in \mathcal{M}$ (cf. [18, Proposition 2.10]). Thus, if continuity of the subgradient mapping $x \in \mathcal{M} \mapsto \partial f(x)$ in (d) is maintained around $\bar{x}$, then $\mathcal{C}^2$-partial smoothness of f at $\bar{x}$ relative to $\mathcal{M}$ yields the same property at all points $x \in \mathcal{M}$ nearby. A localized version of the condition in (d) with respect to a dual variable $\bar{v} \in \partial f(\bar{x})$, which is absent from the definition above, will be justified in Proposition 3.9 to hold around $\bar{x}$. In the sequel, we refer to the properties described in Definition 3.1 with condition (d) being replaced by

   (d') there exists a neighborhood V of $\bar{v}$ such that $\lim_{x \in \mathcal{M}, x \to \bar{x}}(V \cap \partial f(x)) = V \cap \partial f(\bar{x})$

as the $\mathcal{C}^2$-*partial smoothness of f at $\bar{x}$ for $\bar{v}$.* Note also that this modified version of the $\mathcal{C}^2$-partial smoothness is stable (cf. Proposition 3.9) and suffices for second-order variational analysis in this paper, since we mainly concern with pairs (x, v) in $\operatorname{gph} \partial f$.

The class of $\mathcal{C}^2$-partly smooth functions is known to encompass important functions that often appear in constrained and composite optimization problems, as mentioned in Section 1. We particularly refer the interested readers to [13] for examples of $\mathcal{C}^2$-partly smooth functions and relations between partial smoothness structure and other smooth substructures of nonsmooth functions. Below we briefly discuss the conditions outlined in Definition 3.1 for some important classes of functions.

Example 3.2 (reliably $\mathcal{C}^2$-decomposable functions). Following [32], a function $f : \mathbf{X} \to \overline{\mathbb{R}}$ is called $\mathcal{C}^2$-decomposable at $\bar{x} \in \mathbf{X}$ if $f(\bar{x})$ is finite and f enjoys the composite representation

$$f(x) = f(\bar{x}) + (\vartheta \circ \Phi)(x) \quad \text{for } x \in \mathcal{O}, \tag{3.3}$$

where $\mathcal{O} \subset \mathbf{X}$ is an open neighborhood of $\bar{x}$, $\vartheta : \mathbf{Y} \to \overline{\mathbb{R}}$ is proper, lsc, and sublinear, and $\Phi : \mathcal{O} \to \mathbf{Y}$ is $\mathcal{C}^2$-smooth with $\Phi(\bar{x}) = 0$. Denote by S the subspace of $\mathbf{Y}$ parallel to $\operatorname{aff}\{\partial\vartheta(0)\}$, namely

$S = \text{par}\,\{\partial\vartheta(0)\}$. It was shown in [32, pp. 683–684] that under the following nondegeneracy condition

$$S \cap \ker \nabla\Phi(\bar{x})^* = \{0\}, \tag{3.4}$$

the composite function f in (3.3) is $\mathcal{C}^2$-partly smooth at $\bar{x}$ relative to the $\mathcal{C}^2$-smooth manifold $\mathcal{M} = \{x \in \mathbf{X} \,|\, (P_S \circ \Phi)(x) = 0\}$ with $\widehat{f}(x) = \langle y, \Phi(x) \rangle$ for all $x \in \mathcal{M}$, where $y \in \partial\vartheta(0)$ can be chosen arbitrarily. Moreover, it is argued therein that subdifferential continuity in Definition 3.1(d) holds around the point $\bar{x}$ relative to $\mathcal{M}$. In [11, Section 5], a composite function, satisfying the local representation in (3.3) and the nondegeneracy condition in (3.4), is called reliably $\mathcal{C}^2$-decomposable at $\bar{x}$. Also, it was verified in [11, Example 5.1] that this class encompasses a variety of important functions including polyhedral functions, certain composite functions, and indicator functions of important cones such as the second-order cone and the cone of semidefinite matrices.

Example 3.3 (spectral functions). Let $\mathbf{X} = \mathbb{R}^{n \times m}$, the space of $n \times m$ real matrices equipped with the trace inner product $\langle X, Y \rangle := \text{tr}\,(X^\top Y)$ for all $X, Y \in \mathbb{R}^{n \times m}$. Recall that singular values of a matrix $X \in \mathbb{R}^{n \times m}$ are the square roots of the eigenvalues of $X^\top X$. Denote by $\sigma(X)$ the vector of m singular values of a given matrix $X \in \mathbb{R}^{n \times m}$ with components $\sigma_1(X) \geq \sigma_2(X) \geq \cdots \geq \sigma_m(X) \geq 0$. The correspondence $X \mapsto \sigma(X)$ defines a mapping from $\mathbb{R}^{n \times m}$ to $\mathbb{R}^m$. In this example, we consider a spectral function $f : \mathbb{R}^{n \times m} \to \overline{\mathbb{R}}$ formulated as a composition by $f(X) = (\vartheta \circ \sigma)(X)$, where the outer function $\vartheta : \mathbb{R}^m \to \overline{\mathbb{R}}$ is lsc and absolutely permutation-invariant, meaning invariant under all signed permutations of coordinates. Let $\overline{X} \in \mathbb{R}^{n \times m}$ and let $S \subset \mathbb{R}^m$ be a $\mathcal{C}^2$-smooth manifold around $\sigma(\overline{X})$ with δ_S being absolutely permutation-invariant, and $\mathcal{M}_S = \sigma^{-1}(S)$. It was shown in [3, Theorem 5.3] that f is $\mathcal{C}^2$-partly smooth at $\overline{X}$ relative to $\mathcal{M}_S$ if and only if ϑ is $\mathcal{C}^2$-partly smooth at $\sigma(\overline{X})$ relative to S.

The next lemma presents a simple observation on normal cone to a convex set that will be employed to analyze next the class convex piecewise linear-quadratic functions to see under which conditions such functions are $\mathcal{C}^2$-partly smooth.

Lemma 3.4. *Assume that $\Omega \subset \mathbf{X}$ is a convex set and $\bar{x} \in \Omega$. Then $\text{par}\,\{\Omega\}^\perp \subset N_\Omega(\bar{x})$ and equality holds if and only if $\bar{x} \in \text{ri}\,\Omega$.*

Proof. It is clear that $\text{par}\,\{\Omega\}^\perp \subset N_\Omega(x)$ for all $x \in \Omega$. Assume now that $\bar{x} \in \text{ri}\,\Omega$ and pick $\bar{v} \in N_\Omega(\bar{x})$. There must exist $\epsilon > 0$ such that $\mathbb{B}_\epsilon(\bar{x}) \cap \text{aff}\,\{\Omega\} \subset \Omega$. Observe that

$$\langle \bar{v}, u \rangle = \frac{1}{\epsilon}\langle \bar{v}, (\bar{x} + \epsilon u) - \bar{x} \rangle \leq 0$$

for all $u \in \mathbb{B} \cap \text{par}\,\{\Omega\}$, which in turn yields $\bar{v} \in \text{par}\,\{\Omega\}^\perp$. Conversely, if $N_\Omega(\bar{x}) = \text{par}\,\{\Omega\}^\perp$, a subspace, then we infer via [25, Proposition 2.51] that $\bar{x} \in \text{ri}\,\Omega$. The proof is then complete. $\square$

Example 3.5 (convex piecewise linear-quadratic functions). Let $\mathbf{X} = \mathbb{R}^n$ and consider a convex piecewise linear-quadratic (CPLQ) function f defined on $\mathbb{R}^n$. Recall that a proper function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is piecewise linear-quadratic if $\text{dom}\, f = \bigcup_{i \in \mathcal{I}} C_i$ with C_i being a polyhedral convex set for any index i from a finite set $\mathcal{I}$, relative to each of which f takes the form

$$f(x) = \tfrac{1}{2}\langle A_i x, x \rangle + \langle a_i, x \rangle + \alpha_i, \quad x \in C_i, \tag{3.5}$$

for some $n \times n$ symmetric matrix A_i, vector $a_i \in \mathbb{R}^n$, and scalar $\alpha_i \in \mathbb{R}$; cf. [31, Definition 10.20]. Since f is convex, we know from [31, Proposition 10.21] that $\text{dom}\, f$ is polyhedral convex and

$\partial f(x) \neq \emptyset$ for all $x \in \text{dom } f$. Thus, a CPLQ function always satisfies regularity condition in Definition 3.1(b). Pick $\bar{x} \in \text{dom } f$ and define the set of active indices of the domain of f at $\bar{x}$ by

$$\mathcal{I}(\bar{x}) = \big\{i \in \mathcal{I} \,\big|\, \bar{x} \in C_i\big\}.$$

We first analyze the simplest case when $\mathcal{I}(\bar{x})$ is a singleton, say $\mathcal{I}(\bar{x}) = \{i_0\}$ for some $i_0 \in \mathcal{I}$. Since $\mathcal{I}(x) \subset \mathcal{I}(\bar{x})$ for all $x \in \text{dom } f$ sufficiently close to $\bar{x}$, there is a neighborhood of $\bar{x}$ in which the function f behaves as if it is comprised of a single quadratic piece, defined on a polyhedral convex set, namely

$$f(x) = \tfrac{1}{2}\langle A_{i_0} x, x\rangle + \langle a_{i_0}, x\rangle + \alpha_{i_0} + \delta_{C_{i_0}}(x), \quad x \in \mathbb{R}^n.$$

According to [29, Theorem 18.2], there exists a unique face of C_{i_0} containing $\bar{x}$ in its relative interior. Let $\mathcal{M}_{\bar{x}}$ be the relative interior of such a face. Clearly, $\mathcal{M}_{\bar{x}}$ coincides with an affine subspace locally around $\bar{x}$, and thus forms a $\mathcal{C}^2$-smooth manifold around that point. This tells us that f restricted to $\mathcal{M}_{\bar{x}}$ is a quadratic function and thus is $\mathcal{C}^2$-smooth. Moreover, observing that

$$\partial f(x) = A_{i_0} x + a_{i_0} + N_{C_{i_0}}(x) = A_{i_0} x + a_{i_0} + N_{C_{i_0}}(\bar{x})$$

for all $x \in \mathcal{M}_{\bar{x}}$, we get the fulfillment of the subdifferential continuitity in Definition 3.1(d). It is also straightforward to see that $\text{par}\,\{\partial f(\bar{x})\} = \text{span}\,\{N_{C_{i_0}}(\bar{x})\} = N_{\mathcal{M}_{\bar{x}}}(\bar{x})$, meaning that the normal sharpness property in Definition 3.1(c) holds. Thus, if $\mathcal{I}(\bar{x})$ is a singleton, we can conclude that f is $\mathcal{C}^2$-partly smooth at $\bar{x}$ relative to $\mathcal{M}_{\bar{x}}$, where $\mathcal{M}_{\bar{x}}$ is the maximal relatively open subset of $\text{dom } f = C_{i_0}$ containing $\bar{x}$, cf. [29, Theorem 18.2].

We now turn to the general case when $\mathcal{I}(\bar{x})$ may not be a singleton. Observe from the analysis above for the case of a single linear-quadratic piece that the manifold $\mathcal{M}_{\bar{x}}$ consists of all points $x \in \mathbb{R}^n$ sharing not only a common set of active polyhedra $\{C_{i_0}\}$ but also the same normal cone $N_{C_{i_0}}(x) = N_{C_{i_0}}(\bar{x})$. We aim to construct a set $\mathcal{M}_{\bar{x}}$ enjoying this property and to examine to what extent the CPLQ function f is $\mathcal{C}^2$-partly smooth at $\bar{x}$ relative to $\mathcal{M}_{\bar{x}}$.

According to the analysis in the proof of [31, Theorem 11.14(b)], we argue that such a set $\mathcal{M}_{\bar{x}}$ can be defined as follows. Let $\mathcal{H}$ be a finite collection of closed half-spaces $H = \big\{x \in \mathbb{R}^n \,\big|\, \langle b_H, x\rangle \leq \beta_H\big\}$, for some $(b_H, \beta_H) \in (\mathbb{R}^n \setminus \{0\}) \times \mathbb{R}$, such that (i) each of the polyhedral convex sets $\text{dom } f$ and C_i, $i \in \mathcal{I}$, is the intersection of a subcollection of half-spaces in $\mathcal{H}$; and (ii) for every $H \in \mathcal{H}$, the opposite closed half-space $H' = \big\{x \in \mathbb{R}^n \,\big|\, \langle -b_H, x\rangle \leq -\beta_H\big\}$ is also in $\mathcal{H}$. Corresponding to this collection, define for each $\bar{x} \in \text{dom } f$ a subcollection $\mathcal{H}_{\bar{x}} = \big\{H \in \mathcal{H} \,\big|\, \bar{x} \in H\big\}$ and a relatively open set $\mathcal{M}_{\bar{x}} = \big\{x \in \text{dom } f \,\big|\, \mathcal{H}_x = \mathcal{H}_{\bar{x}}\big\}$. Thus, we have $\mathcal{I}(x) = \mathcal{I}(\bar{x})$ for all $x \in \mathcal{M}_{\bar{x}}$, and $N_{C_i}(x) = N_{C_i}(\bar{x})$ for any $x \in \mathcal{M}_{\bar{x}}$ and any $i \in \mathcal{I}(\bar{x})$. Note also that $\mathcal{M}_{\bar{x}}$ has the alternative representation

$$\mathcal{M}_{\bar{x}} = \text{ri} \bigcap_{H \in \mathcal{H}_{\bar{x}}} H = \Big(\bigcap_{H \in \mathcal{H}_{\bar{x}}^*} H \cap H' \Big) \bigcap \Big(\bigcap_{H \in \mathcal{H}_{\bar{x}} \setminus \mathcal{H}_{\bar{x}}^*} \text{int } H \Big), \tag{3.6}$$

where $\mathcal{H}_{\bar{x}}^* = \big\{H \in \mathcal{H}_{\bar{x}} \,\big|\, H' \in \mathcal{H}_{\bar{x}}\big\}$. This clearly shows that in some neighborhood of $\bar{x}$, $\mathcal{M}_{\bar{x}}$ coincides with the affine subspace $\bigcap_{H \in \mathcal{H}_{\bar{x}}^*}(H \cap H')$. Hence, $\mathcal{M}_{\bar{x}}$ forms a $\mathcal{C}^2$-smooth manifold around $\bar{x}$ and satisfies

$$N_{\mathcal{M}_{\bar{x}}}(x) = \text{span}\,\big\{b_H \,\big|\, H \in \mathcal{H}_{\bar{x}}^*\big\} \quad \text{for all } x \in \mathcal{M}_{\bar{x}}.$$

Fix an index $i_0 \in \mathcal{I}(\bar{x})$. Observing that $\mathcal{M}_{\bar{x}} \cap C_{i_0} \neq \emptyset$, we get $\mathcal{M}_{\bar{x}} \subset C_{i_0}$ and conclude that f restricted to $\mathcal{M}_{\bar{x}}$ coincides with $\tfrac{1}{2}\langle A_{i_0} x, x\rangle + \langle a_{i_0}, x\rangle + \alpha_{i_0}$ and thus is $\mathcal{C}^2$-smooth. Recall also from [31, page 487] that

$$\partial f(x) = \bigcap_{i \in \mathcal{I}(\bar{x})} \big(A_i x + a_i + N_{C_i}(\bar{x})\big) \subset A_{i_0} x + a_{i_0} + N_{\mathcal{M}_{\bar{x}}}(\bar{x}) \tag{3.7}$$

7

for all $x \in \mathcal{M}_{\bar{x}}$, where the last inclusion is due to [18, Proposition 2.2].

Claim. *The CPLQ function f in (3.5) is $\mathcal{C}^2$-partly smooth at $\bar{x}$ relative to the affine manifold $\mathcal{M}_{\bar{x}}$ in (3.6) if and only if the following conditions are satisfied:*

$$\begin{cases} \operatorname{span}\{N_{C_i}(\bar{x})\} = N_{\mathcal{M}_{\bar{x}}}(\bar{x}) & \text{for all } i \in \mathcal{I}(\bar{x}), \text{ and} \\ \bigcap_{i\in\mathcal{I}(\bar{x})}(A_i\bar{x} + a_i + \operatorname{ri}N_{C_i}(\bar{x})) \neq \emptyset. \end{cases} \tag{3.8}$$

To verify the "if" part, assume that the conditions in (3.8) hold. Pick $v \in \operatorname{ri}\partial f(\bar{x})$ and conclude from (3.7) and [31, Proposition 2.42] that $v \in A_i\bar{x} + a_i + \operatorname{ri}N_{C_i}(\bar{x})$ for all $i \in \mathcal{I}(\bar{x})$. Employing Lemma 3.4 and the first condition in (3.8) implies for any $i \in \mathcal{I}(\bar{x})$ that

$$N_{N_{C_i}(\bar{x})}(v - A_i\bar{x} - a_i) = \operatorname{span}\{N_{C_i}(\bar{x})\}^{\perp} = T_{\mathcal{M}_{\bar{x}}}(\bar{x}).$$

We then get from [29, Corollary 23.8.1] and (3.7) that

$$N_{\partial f(\bar{x})}(v) = \sum_{i\in\mathcal{I}(\bar{x})} N_{N_{C_i}(\bar{x})}(v - A_i\bar{x} - a_i) = \sum_{i\in\mathcal{I}(\bar{x})} T_{\mathcal{M}_{\bar{x}}}(\bar{x}) = T_{\mathcal{M}_{\bar{x}}}(\bar{x}). \tag{3.9}$$

Recalling that $v \in \operatorname{ri}\partial f(\bar{x})$, we infer from Lemma 3.4 that $N_{\partial f(\bar{x})}(v) = \operatorname{par}\{\partial f(\bar{x})\}^{\perp}$, which, together with (3.9), implies the normal sharpness condition $\operatorname{par}\{\partial f(\bar{x})\} = N_{\mathcal{M}_{\bar{x}}}(\bar{x})$. It follows from [31, Proposition 2.42] and the second condition in (3.8) that the sets $A_i\bar{x} + a_i + N_{C_i}(\bar{x})$ and $\bigcap_{k\in\mathcal{I}(\bar{x})\setminus\{i\}}(A_k\bar{x} + a_k + N_{C_k}(\bar{x}))$ have relative interior points in common for all $i \in \mathcal{I}(\bar{x})$. These amount to none of the sets $A_i\bar{x} + a_i + N_{C_i}(\bar{x})$ being separated properly from the intersection $\bigcap_{k\in\mathcal{I}(\bar{x})\setminus\{i\}}(A_k\bar{x} + a_k + N_{C_k}(\bar{x}))$ of the others, cf. [29, Theorem 11.3]. Employing [31, Exercise 4.33], we can then deduce from the representation of $\partial f(x)$ in (3.7) that ∂f is continuous at $\bar{x}$ relative to $\mathcal{M}_{\bar{x}}$. Thus, f is $\mathcal{C}^2$-partly smooth at $\bar{x}$ with respect to $\mathcal{M}_{\bar{x}}$, defined in (3.6).

We now turn to the "only if" part. First observe from (3.7) and the inclusions $\mathcal{M}_{\bar{x}} \subset C_i$, $i \in \mathcal{I}(\bar{x})$, that

$$\operatorname{par}\{\partial f(\bar{x})\} \subset \operatorname{span}\{N_{C_i}(\bar{x})\} \subset N_{\mathcal{M}_{\bar{x}}}(\bar{x}), \quad i \in \mathcal{I}(\bar{x}). \tag{3.10}$$

The normal sharpness condition $\operatorname{par}\{\partial f(\bar{x})\} = N_{\mathcal{M}_{\bar{x}}}(\bar{x})$ then yields the first condition in (3.8) for all $i \in \mathcal{I}(\bar{x})$. We prove next that $v \in A_i\bar{x} + a_i + \operatorname{ri}N_{C_i}(\bar{x})$ for all $i \in \mathcal{I}(\bar{x})$. Assume that $v - A_{i_0}\bar{x} - a_{i_0} \notin \operatorname{ri}N_{C_{i_0}}(\bar{x})$ for some $i_0 \in \mathcal{I}(\bar{x})$. We then get from $\operatorname{span}\{N_{C_{i_0}}(\bar{x})\}^{\perp} = T_{\mathcal{M}_{\bar{x}}}(\bar{x})$ and Lemma 3.4 that $T_{\mathcal{M}_{\bar{x}}}(\bar{x}) \subsetneq N_{N_{C_{i_0}}(\bar{x})}(v - A_{i_0}\bar{x} - a_{i_0})$. The first equality in (3.9), which is always valid due to (3.7), therefore leads to $T_{\mathcal{M}_{\bar{x}}}(\bar{x}) \subsetneq N_{\partial f(\bar{x})}(v) = N_{\mathcal{M}_{\bar{x}}}(\bar{x})^{\perp}$, a contradiction. Thus, we arrive at $v \in \bigcap_{i\in I(\bar{x})}\left(A_i\bar{x} + a_i + \operatorname{ri}N_{C_i}(\bar{x})\right)$ and complete verification of the claim.

Observe that conditions required in (3.8) automatically hold when $\mathcal{I}(\bar{x})$ is a singleton, as analyzed above. However, (3.8) should not be taken for granted in general. A simple counterexample is given by $f(x) = \frac{1}{2}(\max\{x, 0\})^2$, $x \in \mathbb{R}$. It can be seen that f is $\mathcal{C}^1$-smooth everywhere but not $\mathcal{C}^2$-partly smooth at $\bar{x} = 0$ relative to $\mathcal{M}_{\bar{x}} = \{0\}$, since the normal sharpness property does not hold: $\operatorname{par}\{\nabla f(\bar{x})\} = \{0\}$ while $N_{\mathcal{M}_{\bar{x}}}(\bar{x}) = \mathbb{R}$.

It is noticed, in our forthcoming analysis, that subdifferential regularity in Definition 3.1(b) should be replaced with a stronger condition called *prox-regularity*. In fact, this was well observed in the literature on partial smoothness. For example, in [18, Section 7], Lewis constructed an example of an everywhere subdifferentially regular function that is $\mathcal{C}^2$-partly smooth relative to two distinct manifolds. It was then pointed out in [20, Corollary 4.12] that such a concern is eliminated for prox-regular functions. More precisely, the additional prox-regularity ensures local uniqueness of the corresponding active manifolds, and hence, it does have appealing effects on $\mathcal{C}^2$-partly smooth functions, see, e.g., [20, Sections 4 and 5]. Prox-regularity was therefore a key ingredient in exploiting the partial smoothness structure, see, e.g., [3,4,7]. Also, prox-regularity

and subdifferential continuity of the function in question were essential assumptions throughout the analysis carried out in [11] on strict proto-differentiability of subgradient mappings. Among appealing implications of those two properties, we should mention the equivalence between (strict) twice epi-differentiability and (strict) subgradient proto-differentiability; cf. [31, Theorem 13.40] and [27, Corollary 6.2(b)]. These will be extensively exploited in our establishment of strict twice epi-differentiability of a $\mathcal{C}^2$-partly smooth function and its consequences in this section. We now recall the notion of such regularity. According to [31, Definition 13.27], a function $f: \mathbf{X} \to \overline{\mathbb{R}}$ is 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 $\epsilon > 0$ and $\rho \geq 0$ such that

$$\begin{cases} f(u) \geq f(x') + \langle v', u - x' \rangle - \frac{\rho}{2}\|u - x'\|^2 & \text{for all } u \in \mathbb{B}_\epsilon(\bar{x}) \\ \text{whenever } (x', v') \in \mathbb{B}_\epsilon(\bar{x}, \bar{v}) \cap \operatorname{gph} \partial f \text{ and } f(x') \leq f(\bar{x}) + \epsilon. \end{cases} \tag{3.11}$$

When this property holds for all $v \in \partial f(\bar{x})$, the function f is called prox-regular at $\bar{x}$. The function f is said to be 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 $k \to \infty$. The localization in function value in (3.11) can be omitted if f is subdifferentially continuous at $\bar{x}$ for $\bar{v}$.

Proposition 3.6. *Assume that $f: \mathbf{X} \to \overline{\mathbb{R}}$ is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v}$. Then, there exists $\varepsilon > 0$ such that f is prox-regular and subdifferentially continuous at x for v for any $(x, v) \in \mathbb{B}_\varepsilon(\bar{x}, \bar{v}) \cap \operatorname{gph} \partial f$.*

Proof. Pick the constants ϵ and ρ for which (3.11) holds. Since f is subdifferentially continuous at $\bar{x}$ for $\bar{v}$, we find $\varepsilon \in (0, \epsilon/2)$ such that $|f(x) - f(\bar{x})| < \epsilon$ for any $(x, v) \in \mathbb{B}_\varepsilon(\bar{x}, \bar{v}) \cap \operatorname{gph} \partial f$. Pick any $(x, v) \in \mathbb{B}_\varepsilon(\bar{x}, \bar{v}) \cap \operatorname{gph} \partial f$. It is easy to see that the prox-regularity inequality in (3.11) with the same constant ρ is satisfied for all $u \in \mathbb{B}_{\epsilon/2}(x)$ and all $(x', v') \in \mathbb{B}_{\epsilon/2}(x, v) \cap \operatorname{gph} \partial f$. This proves prox-regularity of f at x for v. Shrinking ε if necessary, we deduce from [27, Proposition 2.3] that $f(x') \to f(x)$ whenever $(x', v') \to (x, v)$ with $(x', v') \in \mathbb{B}_\varepsilon(\bar{x}, \bar{v}) \cap \operatorname{gph} \partial f$, which proves subdifferential continuity of f at x for v and completes the proof. $\square$

The following result summarizes important properties of $\mathcal{C}^2$-partly smooth functions obtained in [18, 20]. While the local representation of graphs of $\mathcal{C}^2$-partly smooth functions in (3.12) below is the cornerstone in our justification of strict twice epi-differentiability of such functions, stability of the relative interior condition plays a key role in extending the latter point-based property to all points nearby. The latter stability can be derived from [4, Lemma 20], however, the equivalence between our statement and the latter lemma is probably not apparent. We instead supply a direct proof for stability of the relative interior condition based on arguments in the proof of [20, Proposition 4.3].

Proposition 3.7. *Let $f: \mathbf{X} \to \overline{\mathbb{R}}$ be a $\mathcal{C}^2$-partly smooth function at $\bar{x}$ relative to a $\mathcal{C}^2$-smooth manifold $\mathcal{M}$. Let $\widehat{f}: \mathbf{X} \to \mathbb{R}$ be any $\mathcal{C}^2$-smooth representative of f around $\bar{x}$ relative to $\mathcal{M}$. The following assertions hold.*

(a) *We have $\nabla \widehat{f}(\bar{x}) \in \operatorname{aff} \{\partial f(\bar{x})\}$.*
(b) *Suppose, moreover, that f is prox-regular and subdifferentially continuous at $\bar{x}$ for some $\bar{v} \in \operatorname{ri} \partial f(\bar{x})$. Then there exists $\varepsilon > 0$ such that*

$$\mathbb{B}_\varepsilon(\bar{x}, \bar{v}) \cap \operatorname{gph} \partial f = \mathbb{B}_\varepsilon(\bar{x}, \bar{v}) \cap \operatorname{gph}(\nabla \widehat{f} + N_{\mathcal{M}}) \tag{3.12}$$

and $v \in \operatorname{ri} \partial f(x)$ for all $(x, v) \in \mathbb{B}_\varepsilon(\bar{x}, \bar{v}) \cap \operatorname{gph} \partial f$.

Proof. First, assertion (a) follows immediately from [18, Proposition 2.4]. We now turn to (b). Since the function f is $\mathcal{C}^2$-partly smooth at $\bar{x}$ relative to $\mathcal{M}$ and both prox-regular and

subdifferentially continuous at $\bar{x}$ for $\bar{v} \in \mathrm{ri}\,\partial f(\bar{x})$, the local representation of gph ∂f in (3.12) results from [20, Corollary 5.2].

It remains to show that the relative interior condition is stable in some neighborhood of $(\bar{x}, \bar{v})$ relative to gph ∂f. The proof of this part mimics that of [20, Proposition 4.3]. Assume the contrary that there exists a sequence $\{(x^k, v^k)\}_{k \in \mathbb{N}} \subset \mathrm{gph}\,\partial f$ with $x^k \to \bar{x}$ and $v^k \to \bar{v}$ as $k \to \infty$ and $v^k \notin \mathrm{ri}\,\partial f(x^k)$ for all k. The local representation in (3.12) then particularly tells us that eventually $x^k \in \mathcal{M}$. Recall that f is subdifferentially regular at all point $x \in \mathcal{M}$ close to $\bar{x}$. We can assume that $\partial f(x^k)$ is closed and convex for all sufficiently large $k \in \mathbb{N}$. Due to stability of the normal sharpness property as noted above, we can also assume that par $\{\partial f(x^k)\} = N_{\mathcal{M}}(x^k)$ for all such k. Employing now the separation theorem for the nonempty closed convex set $\partial f(x^k)$ and a point v^k from its relative boundary (cf. [31, Exercise 2.45(e)]), we find $\eta^k \in$ par $\{\partial f(x^k)\} = N_{\mathcal{M}}(x^k)$ with $\|\eta^k\| = 1$ such that $\langle \eta^k, v^k \rangle \geq \langle \eta^k, u \rangle$ for all $u \in \partial f(x^k)$. Let $\bar{\eta}$ be any accumulation point of $\{\eta^k\}_{k \in \mathbb{N}}$. Clearly, $\bar{\eta} \neq 0$. Moreover, we conclude via robustness of the limiting normal cone that $\bar{\eta} \in N_{\mathcal{M}}(\bar{x}) = \mathrm{par}\,\{\partial f(\bar{x})\}$. Recall that $\bar{v} \in \mathrm{ri}\,\partial f(\bar{x})$. There must exist $\epsilon > 0$ such that $\mathbb{B}_\epsilon(\bar{v}) \cap \mathrm{aff}\,\{\partial f(\bar{x})\} \subset \partial f(\bar{x})$. Pick an arbitrary $u \in \mathbb{B}_\epsilon(\bar{v}) \cap \mathrm{aff}\,\{\partial f(\bar{x})\}$. Since $u \in \partial f(\bar{x})$, we can find $u^k \to u$ with $u^k \in \partial f(x^k)$ for all k sufficiently large due to subdifferential continuity of ∂f at $\bar{x}$ relative to $\mathcal{M}$ and to the facts that $x^k \to \bar{x}$ and $x^k \in \mathcal{M}$ as $k \to \infty$. Noting that $\langle \eta^k, v^k \rangle \geq \langle \eta^k, u^k \rangle$ for all sufficiently large k, we arrive at $\langle \bar{\eta}, \bar{v} \rangle \geq \langle \bar{\eta}, u \rangle$. Since $u \in \partial f(\bar{x})$ is chosen arbitrarily from $\mathbb{B}_\epsilon(\bar{v}) \cap \mathrm{aff}\,\{\partial f(\bar{x})\}$, the latter implies that $\langle \bar{\eta}, u \rangle \leq 0$ for all $u \in \epsilon \mathbb{B} \cap \mathrm{par}\,\{\partial f(\bar{x})\}$, a contradiction with $0 \neq \bar{\eta} \in \mathrm{par}\,\{\partial f(\bar{x})\}$. The proof is then complete. $\square$

Remark 3.8. We should point out that the subdifferential continuity assumption in Proposition 3.7 can be dropped from this result with no harm. Indeed, one can conclude from [8, Proposition 10.12] that the crucial representation in (3.12) does hold without the latter condition. Having this result in our disposal, we can drop the subdifferential continuity assumption from all the results in this paper. However, we proceed with assuming it to avoid more complication in our presentation.

Proposition 3.9 (continuity of subgradient mapping). *Let* $f : \mathbf{X} \to \overline{\mathbb{R}}$ *be a* $\mathcal{C}^2$*-partly smooth function at* $\bar{x}$ *relative to a* $\mathcal{C}^2$*-smooth manifold* $\mathcal{M}$ *such that* f *is prox-regular and subdifferentially continuous at* $\bar{x}$ *for some* $\bar{v} \in \mathrm{ri}\,\partial f(\bar{x})$. *Then there exists a neighborhood* V *of* $\bar{v}$ *such that the subgradient mapping* $x \in \mathcal{M} \mapsto V \cap \partial f(x)$ *is continuous around* $\bar{x}$.

Proof. According to Proposition 3.7, the local representation of gph ∂f in (3.12) with $\mathbb{B}_\varepsilon(\bar{x}, \bar{v})$ replaced with $\mathbb{B}_\varepsilon(\bar{x}) \times \mathbb{B}_\varepsilon(\bar{v})$ is valid for some $\varepsilon > 0$ and some $\mathcal{C}^2$-smooth function $\widehat{f}$. Set $V := \mathbb{B}_{\varepsilon/2}(\bar{v})$ and pick $x \in \mathbb{B}_{\varepsilon/2}(\bar{x}) \cap \mathcal{M}$, $v \in V \cap \partial f(x)$, and $x^k \to x$ with $x^k \in \mathcal{M}$. We now construct a sequence v^k converging to v with $v^k \in \partial f(x^k)$ for all large k. It follows from (3.12) that $v - \nabla \widehat{f}(x) \in N_{\mathcal{M}}(x)$. Due to surjectivity of $\nabla \Phi(\bar{x})$ and smoothness of Φ around $\bar{x}$, the normal cone mapping $x \mapsto N_{\mathcal{M}}(x) = \mathrm{rge}\,\nabla \Phi(x)^*$ is continuous around $\bar{x}$. Choosing a smaller ε if necessary, we can assume that $N_{\mathcal{M}}$ is continuous at x and then, for the given sequence $\{x^k\}_{k \in \mathbb{N}}$, we find $y^k \to v - \nabla \widehat{f}(x)$ with $y^k \in N_{\mathcal{M}}(x^k)$ for all k. Let $v^k := \nabla \widehat{f}(x^k) + y^k$ and observe that $v^k \to v$. We then have $x^k \in \mathbb{B}_\varepsilon(\bar{x})$ and $v^k \in \mathbb{B}_\varepsilon(\bar{v})$, for all k sufficiently large, and thus arrive at $(x^k, v^k) \in (\mathbb{B}_\varepsilon(\bar{x}) \times \mathbb{B}_\varepsilon(\bar{v})) \cap \mathrm{gph}\,(\nabla \widehat{f} + N_{\mathcal{M}})$ for such k. Again, it follows from (3.12) that $v^k \in \partial f(x^k)$ for all k sufficiently large. This confirms continuity of the localization of subgradient mapping ∂f around $\bar{x}$ relative to $\mathcal{M}$ and completes the proof. $\square$

Below, we present our main result in this section in which strict twice epi-differentiability of $\mathcal{C}^2$-partly smooth functions is justified. Taking a $\mathcal{C}^2$-partly smooth function at $\bar{x}$ relative to $\mathcal{M}$, denoted f, and assuming $\widehat{f} : \mathbf{X} \to \mathbb{R}$ as a $\mathcal{C}^2$-smooth representative of f around $\bar{x}$ relative to $\mathcal{M}$, we define the Lagrangian function $L : \mathbf{X} \times \mathbb{R}^m \to \mathbb{R}$ associated with f by

$$L(u, y) := \widehat{f}(u) + \langle y, \Phi(u) \rangle, \quad (u, y) \in \mathbf{X} \times \mathbb{R}^m, \tag{3.13}$$

where the mapping Φ comes from (3.1).

Theorem 3.10 (strict twice epi-differentiability). *Let $\mathcal{M}$ be a $\mathcal{C}^2$-smooth manifold around $\bar{x}$ with the local representation (3.1) and $f : \mathbf{X} \to \overline{\mathbb{R}}$ a $\mathcal{C}^2$-partly smooth function at $\bar{x}$ relative to $\mathcal{M}$ such that f is prox-regular and subdifferentially continuous at $\bar{x}$ for some $\bar{v} \in \operatorname{ri} \partial f(\bar{x})$. Let $\widehat{f} : \mathbf{X} \to \mathbb{R}$ be any $\mathcal{C}^2$-smooth representative of f around $\bar{x}$ relative to $\mathcal{M}$. Then, for all $(x, v) \in \operatorname{gph} \partial f$ sufficiently close to $(\bar{x}, \bar{v})$, the following properties hold:*

(a) *the subgradient mapping ∂f is strictly proto-differentiable at x for v and its graphical derivative $D(\partial f)(x, v)$ can be calculated by*

$$D(\partial f)(x,v)(w) = \begin{cases} \nabla_{xx}^2 L(x, \mu)(w) + N_{\mathcal{M}}(x) & \text{if } w \in T_{\mathcal{M}}(x), \\ \emptyset & \text{otherwise,} \end{cases} \tag{3.14}$$

where $\mu \in \mathbb{R}^m$ is the unique vector satisfying $\nabla \Phi(x)^* \mu = v - \nabla \widehat{f}(x)$;

(b) *the function f is strictly twice epi-differentiable at x for v and its second subderivative can be calculated by*

$$\mathrm{d}^2 f(x, v)(w) = \nabla_{xx}^2 L(x, \mu)(w, w) + \delta_{T_{\mathcal{M}}(x)}(w), \quad w \in \mathbf{X}. \tag{3.15}$$

Proof. We first verify the claimed assertions in (a) and (b) for $(\bar{x}, \bar{v})$ with $\bar{\mu}$ being the unique vector satisfying $\nabla \Phi(\bar{x})^* \bar{\mu} = \bar{v} - \nabla \widehat{f}(\bar{x})$. To this end, we begin by showing that $\widehat{f} + \delta_{\mathcal{M}}$ is strictly twice epi-differentiable at $\bar{x}$ for $\bar{v}$. Take $(x, \eta) \in \operatorname{gph} N_{\mathcal{M}}$ with x sufficiently close to $\bar{x}$. According to [24, Theorem 6.2], $\delta_{\mathcal{M}}$ is twice epi-differentible at x for η, and its second subderivative at x for η is given by

$$\mathrm{d}^2 \delta_{\mathcal{M}}(x, \eta)(w) = \langle \nu, \nabla \Phi^2(\bar{x})(w, w) \rangle + \delta_{K_{\mathcal{M}}(x, \eta)}(w), \quad w \in \mathbf{X}, \tag{3.16}$$

where $\nu \in \mathbb{R}^m$ is the unique vector satisfying $\nabla \Phi(x)^* \nu = \eta$. Since $\widehat{f}$ is $\mathcal{C}^2$-smooth, [31, Exercise 13.18] tells us that $\widehat{f} + \delta_{\mathcal{M}}$ maintains twice epi-differentiability at such x for $\nabla \widehat{f}(x) + \eta$ and

$$\mathrm{d}^2(\widehat{f} + \delta_{\mathcal{M}})(x, \nabla \widehat{f}(x) + \eta)(w) = \nabla^2 \widehat{f}(x)(w, w) + \mathrm{d}^2 \delta_{\mathcal{M}}(x, \eta)(w), \quad w \in \mathbf{X}. \tag{3.17}$$

Note further from Proposition 3.7(a) and the normal sharpness in Definition 3.1 that $\bar{v} - \nabla \widehat{f}(\bar{x}) \in \operatorname{par} \{\partial f(\bar{x})\} = N_{\mathcal{M}}(\bar{x})$. Since $N_{\mathcal{M}}(\bar{x})$ is obviously a linear subspace, we get $\bar{v} - \nabla \widehat{f}(\bar{x}) \in \operatorname{ri} \partial \delta_{\mathcal{M}}(\bar{x})$ and conclude from [11, Theorem 5.8] that $\delta_{\mathcal{M}}$ is strictly twice epi-differentiable at $\bar{x}$ for $\bar{v} - \nabla \widehat{f}(\bar{x})$. By [11, Proposition 5.6] and the aforementioned twice epi-differentiability of $\delta_{\mathcal{M}}$, we obtain for any pair $(x, \eta) \to (\bar{x}, \bar{v} - \nabla \widehat{f}(\bar{x}))$ with $\eta \in N_{\mathcal{M}}(x)$ that

$$\mathrm{d}^2 \delta_{\mathcal{M}}(x, \eta) \xrightarrow{e} \mathrm{d}^2 \delta_{\mathcal{M}}(\bar{x}, \bar{v} - \nabla \widehat{f}(\bar{x})).$$

Employing now the sum rule for epi-convergence in [31, Theorem 7.46(b)], we deduce from the latter and (3.17) that

$$\mathrm{d}^2(\widehat{f} + \delta_{\mathcal{M}})(x, \nabla \widehat{f}(x) + \eta) \xrightarrow{e} \nabla^2 \widehat{f}(\bar{x})(\cdot, \cdot) + \mathrm{d}^2 \delta_{\mathcal{M}}(\bar{x}, \bar{v} - \nabla \widehat{f}(\bar{x}))$$

as $(x, \eta) \to (\bar{x}, \bar{v} - \nabla \widehat{f}(\bar{x}))$ with $\eta \in N_{\mathcal{M}}(x)$. We then conclude via [11, Proposition 5.6] that $\widehat{f} + \delta_{\mathcal{M}}$ is strictly twice epi-differentiable at $\bar{x}$ for $\bar{v}$ and that

$$\mathrm{d}^2(\widehat{f} + \delta_{\mathcal{M}})(\bar{x}, \bar{v})(w) = \nabla^2 \widehat{f}(\bar{x})(w, w) + \mathrm{d}^2 \delta_{\mathcal{M}}(\bar{x}, \bar{v} - \nabla \widehat{f}(\bar{x}))(w), \quad w \in \mathbf{X}.$$

11

Recall that $\bar{\mu}$ is the unique vector satisfying the equation $\nabla\Phi(\bar{x})^*\bar{\mu} = \bar{v} - \nabla\widehat{f}(\bar{x})$. In view of (3.2) and (3.16), we get from the above formula that

$$d^2(\widehat{f} + \delta_{\mathcal{M}})(\bar{x}, \bar{v})(w) = \nabla^2\widehat{f}(\bar{x})(w, w) + \langle\bar{\mu}, \nabla^2\Phi(\bar{x})(w, w)\rangle + \delta_{T_{\mathcal{M}(\bar{x})}}(w)$$
$$= \nabla^2_{xx}L(\bar{x}, \bar{\mu})(w, w) + \delta_{T_{\mathcal{M}(\bar{x})}}(w), \quad w \in \mathbf{X}.$$

By [31, Example 10.24(f)], $\widehat{f} + \delta_{\mathcal{M}}$ is strongly amenable at $\bar{x}$ in the sense of [31, Definition 10.23(b)]. Thus, by [31, Theorem 13.32], it is prox-regular and subdifferentially continuous at any $x \in \mathcal{M}$ sufficiently close to $\bar{x}$. Appealing now to [28, Corollary 4.3] and [31, Theorem 13.40], we deduce that the subgradient mapping $\partial(\widehat{f} + \delta_{\mathcal{M}}) = \nabla\widehat{f} + N_{\mathcal{M}}$ is strictly proto-differentiable at $\bar{x}$ for $\bar{v}$ and that

$$D(\nabla\widehat{f} + N_{\mathcal{M}})(\bar{x}, \bar{v})(w) = \partial\big(\tfrac{1}{2}\nabla^2_{xx}L(\bar{x}, \bar{\mu})(\cdot, \cdot) + \delta_{T_{\mathcal{M}(\bar{x})}}\big)(w)$$
$$= \nabla^2_{xx}L(\bar{x}, \bar{\mu})(w) + N_{T_{\mathcal{M}(\bar{x})}}(w), \quad w \in \mathbf{X}.$$

According to Proposition 3.7(a), gph ∂f and gph $\partial(\widehat{f} + \delta_{\mathcal{M}})$ locally coincide around $(\bar{x}, \bar{v})$. This yields strict proto-differentiability of ∂f at $\bar{x}$ for $\bar{v}$. It also confirms that

$$D(\partial f)(\bar{x}, \bar{v}) = D(\nabla\widehat{f} + N_{\mathcal{M}})(\bar{x}, \bar{v})$$

and hence proves the claimed formula in (3.14). Moreover, we infer via [31, Corollary 4.3] that f is strictly twice epi-differentiable at $\bar{x}$ for $\bar{v}$. Remembering from [31, Theorem 13.40] that

$$\partial\big(\tfrac{1}{2}d^2 f(\bar{x}, \bar{v})\big) = D(\partial f)(\bar{x}, \bar{v}),$$

and setting $\varphi(w) := \nabla^2_{xx}L(\bar{x}, \bar{\mu})(w, w) + \delta_{T_{\mathcal{M}(\bar{x})}}(w)$ for any $w \in \mathbf{X}$, we deduce from (3.14) that $\partial\big(d^2 f(\bar{x}, \bar{v})\big)(w) = \partial\varphi(w)$ for any $w \in \mathbf{X}$. By [31, Proposition 13.49], we find $r > 0$ such that the mapping $w \mapsto d^2 f(\bar{x}, \bar{v})(w) + r\|w\|^2$ is a proper lsc convex function. Choosing a bigger r if necessary, we can assume without loss of generality that $w \mapsto \varphi(w) + r\|w\|^2$ is also a proper lsc convex function. Appealing to [31, Theorem 12.25], which is an integration of the subdifferential for convex functions, tells us that there is a constant c for which we have $d^2 f(\bar{x}, \bar{v})(w) + r\|w\|^2 = \varphi(w) + r\|w\|^2 + c$ for any $w \in \mathbf{X}$. Since $d^2 f(\bar{x}, \bar{v})(0) = 0$, we get $c = 0$, which proves the formula (3.15) for $(\bar{x}, \bar{v})$.

We now argue that assertions (a) and (b) are also valid for any $(x, v) \in$ gph ∂f sufficiently close to $(\bar{x}, \bar{v})$. It follows from the discussion after Definition 3.1 and Proposition 3.9 that f is $\mathcal{C}^2$-partly smooth at x for $v \in \partial f(x)$ for all (x, v) sufficiently close to $(\bar{x}, \bar{v})$. Also, Proposition 3.6 tells us that prox-regularity and subdifferential continuity of f at x for v are preserved whenever $(x, v) \in$ gph ∂f and sufficiently close to $(\bar{x}, \bar{v})$. For any such a pair (x, v), we can deduce, without loss of generality, via Proposition 3.7(b) that $v \in$ ri $\partial f(x)$. Using an argument for any such a pair (x, v) similar to that for $(\bar{x}, \bar{v})$ proves (a) and (b) for (x, v) and hence completes the proof. $\square$

Strict twice epi-differentiability in a neighborhood relative to the graph of subgradient mapping was first characterized for polyhedral functions in [9, Theorem 4.1] via the same relative interior condition in Theorem 3.10. A similar result was then achieved for composite functions whose outer functions are polyhedral convex in [10, Theorem 3.9] under a second-order qualification condition (SOQC). We should stress that these structures, namely, polyhedral function and $\mathcal{C}^2$-smooth mapping composed with polyhedral function satisfying SOQC, are stable in the sense that if a given function has such a structure at a given point, it maintains the same structure at all points nearby. In [11, Theorem 5.8], by assuming the reliable $\mathcal{C}^2$-decomposability of the function in a neighborhood of the point in question, the characterization was extended to the

class of reliably $\mathcal{C}^2$-decomposable functions considered in Example 3.2. However, Example 3.2 and Theorem 3.10 together tell us that reliable $\mathcal{C}^2$-decomposability at the point in question solely can guarantee the aforementioned characterization. This discloses the power of robustness of $\mathcal{C}^2$-partial smoothness. We should remark, however, that in Theorem 3.10 the relative interior condition serves as a sufficient condition, not as a characterization, for strict twice epidifferentiability. In fact, the essence of $\mathcal{C}^2$-partial smoothness is rooted in the local representation of gph ∂f in (3.12) which is available only for multipliers taken from the relative interior of the subdifferential, to the best of our knowledge.

Remark 3.11. Note that it is possible to get another proof of Theorem 3.10 using [28, Theorem 4.4] and [4, Theorem 28]. Indeed, it was shown in the former that for prox-regular functions, strict twice epi-differentiability amounts to continuous differentiability of proximal mappings under an extra condition that the function has a global minimum. The latter restrictive condition was weakened recently by the authors in [11, Theorem 4.7], where it was argued that prox-boundedness suffices to achieve this equivalence. On the other hand, it was shown in [4, Theorem 28] for prox-regular and prox-bounded functions that are $\mathcal{C}^2$-partly smooth that the proximal mapping is continuously differentiable. Combining these results tells us that, providing the same relative interior condition imposed in Theorem 3.10 is satisfied, any prox-regular, prox-bounded, and $\mathcal{C}^2$-partly smooth function is strictly twice epi-differentiable. Theorem 3.10 provides new information, however. First, it shows that prox-boundedness is not required. Second, it proves strict twice epi-differentiability of a prox-regular and $\mathcal{C}^2$-partly smooth function in a neighborhood of a given point. Lastly, it presents a simple formula for the second subderivative of such functions that has been reported before and is a byproduct of our direct proof for this result.

Remark 3.12. Assume that $f : \mathbf{X} \to \overline{\mathbb{R}}$ is reliably $\mathcal{C}^2$-decomposable at $\bar{x} \in \mathbf{X}$ with representation (3.3) and that $\bar{v} \in \operatorname{ri} \partial f(\bar{x})$. Pick an arbitrary $y \in \partial \vartheta(0)$. Example 3.2 asserts that f is $\mathcal{C}^2$-partly smooth at $\bar{x}$ relative to the $\mathcal{C}^2$-smooth manifold $\mathcal{M} = \left\{ x \in \mathbf{X} \,\middle|\, \Phi(x) \in S^\perp \right\}$, where $S = \operatorname{par} \{\partial \vartheta(0)\}$, and $\widehat{f}(x) = \langle y, \Phi(x) \rangle$ is a $\mathcal{C}^2$-smooth representative of f around $\bar{x}$ relative to $\mathcal{M}$. It can be shown that the composite function (3.3) satisfying (3.4) is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v}$, since it is strongly amenable at $\bar{x}$ in the sense of [31, Definition 10.23(b)]. Let $\bar{\mu} \in \mathbf{Y}$ be any solution to $\nabla(P_S \circ \Phi)(\bar{x})^* \mu = \bar{v} - \nabla \Phi(\bar{x})^* y$. Recalling that the projection P_S is linear and self-adjoint, we get from the latter that

$$\nabla \Phi(\bar{x})^* P_S(\bar{\mu}) = \bar{v} - \nabla \widehat{f}(\bar{x}).$$

It then follows from (3.4) that $\eta := P_S(\bar{\mu})$ is the unique vector in S satisfying $\nabla \Phi(\bar{x})^* \eta = \bar{v} - \nabla \widehat{f}(\bar{x})$. We now can conclude via Theorem 3.10 that f is strictly twice epi-differentiable at $\bar{x}$ for $\bar{v}$ and that

$$\mathrm{d}^2 f(\bar{x}, \bar{v})(w) = \nabla_{xx}^2 \langle y + P_S(\bar{\mu}), \Phi \rangle(\bar{x})(w, w) + \delta_{T_{\mathcal{M}}(\bar{x})}(w), \quad w \in \mathbf{X}. \tag{3.18}$$

We next justify that $\bar{y} := y + P_S(\bar{\mu}) \in \partial \vartheta(0)$ and $\nabla \Phi(\bar{x})^* \bar{y} = \bar{v}$. Indeed, the latter follows from definition of $\bar{\mu}$. To prove the former, let $\hat{y}$ be the unique vector from $\partial \vartheta(0)$ satisfying $\nabla \Phi(\bar{x})^* \hat{y} = \bar{v}$, thanks to (3.4). Since $\bar{v} \in \operatorname{ri} \partial f(\bar{x})$, we infer via [11, Proposition 5.3(b)] that $\hat{y} \in \operatorname{ri} \partial \vartheta(0)$. For any $t \in [0, 1)$, we have $(1 - t)\hat{y} + ty \in \operatorname{ri} \partial \vartheta(0)$ which results in

$$(1 - t)\hat{y} + t\bar{y} = (1 - t)\hat{y} + ty + tP_S(\bar{\mu}) \in \partial \vartheta(0)$$

for all $t \geq 0$ sufficiently small due to $S = \operatorname{par} \{\partial \vartheta(0)\}$. Because $\nabla \Phi(\bar{x})^* ((1 - t)\hat{y} + t\bar{y}) = \bar{v}$, we then conclude from the uniqueness of $\hat{y}$ that $\bar{y} = \hat{y}$. Recalling the composite form (3.3) of f, we have $K_f(\bar{x}, \bar{v}) = \left\{ w \in \mathbf{X} \,\middle|\, \nabla \Phi(\bar{x}) w \in K_\vartheta(0, \bar{y}) \right\}$. Using sublinearity of ϑ, we have

$\vartheta(u) = \mathrm{d}\vartheta(0)(u)$ for any $u \in \mathrm{dom}\,\vartheta$, which coupled with [31, Theorem 8.24 and Corollary 8.25] leads us to

$$K_\vartheta(0, \bar{y}) = \big\{ u \in \mathbf{Y} \,\big|\, \vartheta(u) = \langle \bar{y}, u \rangle \big\} = \big\{ u \in \mathbf{Y} \,\big|\, \bar{y} \in \operatorname*{argmax}_{y \in \partial\vartheta(0)} \langle u, y \rangle \big\}.$$

Since $\bar{y} \in \mathrm{ri}\,\partial\vartheta(0)$, the linear function $\langle u, \cdot \rangle$ attains its maximum over $\partial\vartheta(0)$ at $\bar{y}$ if and only if $u \in (\mathrm{aff}\,\{\partial\vartheta(0)\})^\perp = S^\perp$. Thus, we have

$$K_f(\bar{x}, \bar{v}) = \big\{ w \in \mathbf{X} \,\big|\, \nabla\Phi(\bar{x})w \in S^\perp \big\} = T_\mathcal{M}(\bar{x}),$$

where the last equality follows from (3.4). Combining this, (3.18), and the fact that $\bar{y} = y + P_S(\bar{\mu})$ is the unique vector from $\partial\vartheta(0)$ satisfying $\nabla\Phi(\bar{x})^*\bar{y} = \bar{v}$ implies that Theorem 3.10 covers the implication (c) $\Longrightarrow$ (a), and also the implication (c) $\Longrightarrow$ (b), in [11, Theorem 5.8]. We should stress here that in order to apply Theorem 3.10, one only needs to justify the reliable $\mathcal{C}^2$-decomposability of the function f at the point $\bar{x}$. On the contrast, [11, Theorem 5.8] demands the latter property in a neighborhood of $\bar{x}$. Thus, Theorem 3.10 actually improves those implications. Particularly, we infer from the latter theorem that a proper, lsc, and sublinear function $\vartheta : \mathbf{Y} \to \overline{\mathbb{R}}$ is strictly twice epi-differentiable at $z \in \mathbf{Y}$ for $y \in \partial\vartheta(z)$ for all $(z, y) \in \mathrm{gph}\,\partial\vartheta$ close to $(0, \bar{y})$, for any $\bar{y} \in \mathrm{ri}\,\partial\vartheta(0)$, which is an improvement of [11, Theorem 5.7].

Stability of the strict twice-epidifferentiability of a $\mathcal{C}^2$-partly smooth function, established in Theorem 3.10, makes the class of such functions a proper subset of the class of strictly twice epi-differentiable functions as demonstrated in the next example.

Example 3.13 (failure of strict twice epi-diffrentiability in a neighborhood)**.** To demonstrate that the class of strictly twice epi-differentiable functions is strictly larger than that of $\mathcal{C}^2$-partly smooth functions, consider the function $f(x) = \int_0^x g(t)dt$, $x \in \mathbb{R}$, where $g : \mathbb{R} \to \mathbb{R}$ is a continuous piecewise affine function given by

$$g(x) = \begin{cases} 0 & \text{if } x \le 0, \\ 2^{2i} + 3 \cdot 2^i(x - 2^i) & \text{if } x \in [2^i, 2^{i+1}), \ i \in \mathbb{Z}. \end{cases}$$

Observe that f is $\mathcal{C}^1$-smooth with $\nabla f(x) = g(x)$ for all $x \in \mathbb{R}$. Moreover, it can be seen that g is monotonically increasing over $\mathbb{R}$, which ensures that f is convex. Note also that the affine pieces comprising the graph of g have increasing slopes of 0 and $3 \cdot 2^i$, $i \in \mathbb{Z}$. Assume that t and u are close to 0 and that $t > u$ with $t \in [2^{-i}, 2^{-i+1})$ for some $i \in \mathbb{N}$. If $u \le 0 < t$, we arrive at

$$g(t) - g(u) = g(t) = 3 \cdot 2^{-i}t - 2^{-2i+1} \le 3 \cdot 2^{-i}t \le 3 \cdot 2^{-i}(t - u).$$

If $t > u \ge 0$, then $u \in [2^{-j}, 2^{-j+1})$ for some $j \ge i$. Assume that $j > i$ and consider the lines

$$y = g_i(x) := 3 \cdot 2^{-i}x - 2^{-2i+1} \quad \text{and} \quad y = g_j(x) := 3 \cdot 2^{-j}x - 2^{-2j+1},$$

which intersect each other at the point whose x-coordinate is $x_0 = \frac{2}{3}2^{-i} + \frac{1}{3}2^{-j+1}$. Since $u \le x_0 \le t$, we have $g(u) = g_j(u) \ge g_i(u)$ and

$$g(t) - g(u) \le g_i(t) - g_i(u) = 3 \cdot 2^{-i}(t - u).$$

Clearly, the latter estimate is also valid for $j = i$. Combining these tells us that

$$0 \le \frac{g(t) - g(u)}{t - u} \le \frac{3 \cdot 2^{-i}(t - u)}{t - u} \le 3t \to 0 \quad \text{as } t, u \searrow 0.$$

14

Finally, we have $g(t) - g(u) = 0$ if $u < t \leq 0$. Thus, g is strictly differentiable at $\bar{x} = 0$ with $\nabla g(0) = 0$. Note, however, that g is not differentiable at any $x = 2^i$, $i \in \mathbb{Z}$, since those points correspond to kinks on gph g. In view of [30, Proposition 3.1(b)] and [28, Theorem 4.3], we conclude that the convex function f is strictly twice epi-differentiable at $\bar{x} = 0$ for $\nabla f(\bar{x}) = 0$ while not satisfying such a property along the sequence $x^k := 2^{-k}$, $k \in \mathbb{N}$, which converges to 0. Theorem 3.10 then tells us that f is not $\mathcal{C}^2$-partly smooth at 0.

As Theorem 3.10 shows, under the assumptions therein a $\mathcal{C}^2$-partly smooth function is always strictly twice epi-differentiable around the point under consideration. One may wonder whether strict twice epi-differentiability in a neighborhood actually amounts to $\mathcal{C}^2$-partial smoothness. The example below rules out this possibility.

Example 3.14 (failure of $\mathcal{C}^2$-partial smoothness). Define the function $f : \mathbb{R}^2 \to \mathbb{R}$ by $f(x_1, x_2) = |x_1^3 x_2^2|$, which is taken from [26, Example 4.11]. Since f can be expressed equivalently as $\max\{x_1^3 x_2^2, -x_1^3 x_2^2\}$, it is prox-regular and subdifferentially continuous on $\mathbb{R}^2$. Moreover, f is $\mathcal{C}^1$-smooth on $\mathbb{R}^2$ and $\mathcal{C}^2$-smooth around any point (x_1, x_2) with $x_1 \neq 0$. Its gradient and Hessian, when exists, can be calculated as

$$\nabla f(x_1, x_2) = \operatorname{sgn}(x_1)(3x_1^2 x_2^2, 2x_1^3 x_2)$$

and

$$\nabla^2 f(x_1, x_2) = \operatorname{sgn}(x_1) \begin{bmatrix} 6x_1 x_2^2 & 6x_1^2 x_2 \\ 6x_1^2 x_2 & 2x_1^3 \end{bmatrix} \quad \text{if } x_1 \neq 0,$$

where $\operatorname{sgn}(x_1)$ signifies the sign of x_1: $\operatorname{sgn}(x_1) = 1$ if $x_1 > 0$, $\operatorname{sgn}(x_1) = -1$ if $x_1 < 0$, and $\operatorname{sgn}(x_1) = 0$ otherwise. For any $w = (w_1, w_2) \in \mathbb{R}^2$, it is not hard to see that

$$\mathrm{d}^2 f(x_1, x_2)(w) = \begin{cases} 0 & \text{if } (x_1, x_2) \in \{0\} \times \mathbb{R}, \\ \nabla^2 f(x_1, x_2)(w, w) & \text{otherwise.} \end{cases}$$

Observe from these formulas that $\mathrm{d}^2 f(x_1', x_2') \to \mathrm{d}^2 f(x_1, x_2)$ uniformly on bounded sets whenever $(x_1', x_2') \to (x_1, x_2)$. It then follows from [31, Proposition 7.15] that the latter uniform convergence implies $\mathrm{d}^2 f(x_1', x_2') \xrightarrow{e} \mathrm{d}^2 f(x_1, x_2)$ when $(x_1', x_2') \to (x_1, x_2)$, or, equivalently, f is strictly twice epi-differentiability at any $(x_1, x_2) \in \mathbb{R}^2$. However, we see that f is not $\mathcal{C}^2$-partly smooth at $\bar{x} := (0, 0)$. Indeed, if that was the case, we would conclude from the smoothness of f at $\bar{x}$ and the normal sharpness property of f at this point that $N_{\mathcal{M}}(\bar{x}) = \{0\}$, where $\mathcal{M}$ is the associated manifold to f from Definition 3.1. By [31, Exercise 6.19], we arrive at $\bar{x} \in \operatorname{int} \mathcal{M}$, which tells us that f must be $\mathcal{C}^2$-smooth around $\bar{x}$, a contradiction to the calculation above. This confirms that f is not $\mathcal{C}^2$-partly smooth at $\bar{x}$.

Recall that for a set-valued mapping $F : \mathbf{X} \rightrightarrows \mathbf{Y}$ with $(\bar{x}, \bar{y}) \in \operatorname{gph} F$, the coderivative mapping of F at $\bar{x}$ for $\bar{y}$, denoted $D^* F(\bar{x}, \bar{v})$, is defined via the normal cone to gph F at $(\bar{x}, \bar{y})$ by

$$\eta \in D^* F(\bar{x}, \bar{y})(w) \iff (\eta, -w) \in N_{\operatorname{gph} F}(\bar{x}, \bar{y}). \tag{3.19}$$

Corollary 3.15. *Let $f : \mathbf{X} \to \overline{\mathbb{R}}$ be a $\mathcal{C}^2$-partly smooth function at $\bar{x}$ relative to a $\mathcal{C}^2$-smooth manifold $\mathcal{M}$ such that f is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v}$ with $\bar{v} \in \operatorname{ri} \partial f(\bar{x})$. Then, for any $(x, v) \in \operatorname{gph} \partial f$ sufficiently close to $(\bar{x}, \bar{v})$, we have*

$$D(\partial f)(x, v) = \widetilde{D}(\partial f)(x, v) = D^*(\partial f)(x, v). \tag{3.20}$$

Proof. According to Theorem 3.10, f is strictly proto-differentiable at x for v for any $(x, v) \in \operatorname{gph} \partial f$ sufficiently close to $(\bar{x}, \bar{v})$, which in turn justifies the first claimed equality. The second one then results from the recent observation in [11, Theorem 3.9]. $\square$

We should add here that the local representation of gph ∂f in (3.12) for a $\mathcal{C}^2$-partly smooth function f was exploited in [20, Theorem 5.3 and Corollary 5.4] to derive an exact formula for $D^*(\partial f)(\bar{x}, \bar{v})$ in terms of the covariant Hessian $\nabla^2_{\mathcal{M}} f_{\bar{v}}(\bar{x})$ of the tilted function $f_{\bar{v}} := f - \langle \bar{v}, \cdot \rangle$. Note that the bilinear form $\nabla^2_{xx} L(\bar{x}, \bar{\lambda})$ in (3.14) restricted to $T_{\mathcal{M}}(\bar{x}) \times T_{\mathcal{M}}(\bar{x})$ gives $\nabla^2_{\mathcal{M}} f_{\bar{v}}(\bar{x})$; see [20, Definition 2.11] and the discussion afterwards. Corollary 3.15 not only provides a new proof for this result but also goes one step further and calculates the strict graphical derivative of the subgradient mapping for this class of functions.

Recall that a function $f : \mathbf{X} \to \overline{\mathbb{R}}$ enjoys the quadratic growth condition at $\bar{x}$ if $f(\bar{x})$ is finite and there exists $\ell > 0$ such that

$$f(x) \geq f(\bar{x}) + \frac{\ell}{2} \|x - \bar{x}\|^2 \tag{3.21}$$

for all x close to $\bar{x}$. In the corollary below, we glean a characterization for the quadratic growth of $\mathcal{C}^2$-partly smooth functions from our calculation of the second subderivative of those functions in the proof of Theorem 3.10. The same result was previously observed in [20, Propostion 4.13] using a different approach.

Corollary 3.16. *Under the hypothesis of Theorem 3.10, there exists $\varepsilon > 0$ such that for any $(x, v) \in \mathbb{B}_\varepsilon(\bar{x}, \bar{v}) \cap \operatorname{gph} \partial f$ we have*

$$\mathrm{d}^2 f(x, v) = \mathrm{d}^2 (\widehat{f} + \delta_{\mathcal{M}})(x, v). \tag{3.22}$$

Consequently, if $\bar{v} = 0$ then f enjoys the quadratic growth condition at $\bar{x}$ if and only if $\hat{f}$ enjoys the same property at $\bar{x}$ with respect to $\mathcal{M}$, namely (3.21) holds with f replaced with $\hat{f}$ for all $x \in \mathcal{M}$ close to $\bar{x}$.

Proof. The first claim about the second subderivative of f was indeed proved in the proof of Theorem 3.10. The second claim results from (3.22) for $(x, v) = (\bar{x}, 0)$ and the fact that the validity of quadratic growth condition can be fully characterized by positiveness of the second subderivative; see [31, Theorem 13.24(c)]. $\qquad \square$

Remark 3.17. It is worth pointing out that in [20, Propostion 4.13] the subdifferential continuity was not assumed. Taking into account the discussion in Remark 3.8, it is possible to drop the latter condition from Corollary 3.16 as well.

The rest of this section is devoted to some applications of the established strict twice epi-differentiability of $\mathcal{C}^2$-partly smooth functions. We begin with regularity properties of the solution mapping to the generalized equation

$$\bar{p} \in G(x) := \overline{\psi}(x) + \partial f(x), \tag{3.23}$$

where $f : \mathbf{X} \to \overline{\mathbb{R}}$ is a $\mathcal{C}^2$-partly smooth function, $\overline{\psi} : \mathbf{X} \to \mathbf{X}$ is $\mathcal{C}^1$-smooth around the point under consideration, and $\bar{p} \in \mathbf{X}$. We are mainly interested in the solution mapping to (3.23), that is the mapping $p \mapsto G^{-1}(p)$.

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 $\operatorname{dist}(x, F^{-1}(y)) \leq \kappa \operatorname{dist}(y, F(x))$ holds for all $(x, y) \in U \times V$. The mapping F is said to be strongly metrically regular at $\bar{x}$ for $\bar{y}$ if its inverse 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 [6, 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 $\bar{x}$. These regularity properties of a set-valued mapping enjoy full characterizations via the coderivative in (3.19) (cf. [31, Theorem 9.43]), and the strict graphical derivative in (2.3) (cf. [31, Theorem 9.54(b)]), respectively. We recently demonstrated in [11, Proposition 4.1] that these regularity properties are equivalent for the mapping G in (3.23) provided that ∂f is strictly proto-differentiable. Moreover, we presented several simple characterizations for those equivalent properties of G. In the presence of strict proto-differentiability of ∂f in a neighborhood relative to $\mathrm{gph}\,\partial f$, established for $\mathcal{C}^2$-partly smooth functions in Theorem 3.10, we provide below an adaption of the aforementioned results for the mapping G from (3.23).

Proposition 3.18. *Let $\bar{x}$ be a solution to the generalized equation in (3.23) for $\bar{p} \in \mathbf{X}$. Assume that $\overline{\psi}$ is $\mathcal{C}^1$-smooth around $\bar{x}$ and f is $\mathcal{C}^2$-partly smooth at $\bar{x}$ relative to a $\mathcal{C}^2$-smooth manifold $\mathcal{M}$ with local representation (3.1). Assume in addition that $\bar{v} := \bar{p} - \overline{\psi}(\bar{x}) \in \mathrm{ri}\,\partial f(\bar{x})$, and that f is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v}$. Then the following properties are equivalent.*

  (a) *The mapping G from (3.23) is metrically regular at $\bar{x}$ for $\bar{p}$.*
  (b) *The mapping G from (3.23) is strongly metrically regular at $\bar{x}$ for $\bar{p}$.*
  (c) *The inverse mapping G^{-1} has a Lipschitz continuous single-valued localization s around $\bar{p}$ for $\bar{x}$.*
  (d) *The condition*

$$\left\{ w \in T_{\mathcal{M}}(\bar{x}) \,\middle|\, \left(\nabla\overline{\psi}(\bar{x}) + \nabla^2_{xx}L(\bar{x},\bar{\mu})\right)^*(w) \in N_{\mathcal{M}}(\bar{x}) \right\} = \{0\} \qquad (3.24)$$

  is satisfied, where L is taken from (3.13) and $\bar{\mu}$ is the unique solution to $\nabla\Phi(\bar{x})^*\bar{\mu} = \bar{p} - (\overline{\psi}(\bar{x}) + \nabla\widehat{f}(x))$ with $\widehat{f} : \mathbf{X} \to \mathbb{R}$ being any $\mathcal{C}^2$-smooth representative of f around $\bar{x}$ relative to $\mathcal{M}$.
  (e) *The composite mapping $P_{T_{\mathcal{M}}(\bar{x})} \circ (\nabla\overline{\psi}(\bar{x}) + \nabla^2_{xx}L(\bar{x},\bar{\mu}))|_{T_{\mathcal{M}}(\bar{x})}$, where $(\nabla\overline{\psi}(\bar{x}) + \nabla^2_{xx}L(\bar{x},\bar{\mu}))|_{T_{\mathcal{M}}(\bar{x})}$ stands for the restriction of the linear mapping $\nabla\overline{\psi}(\bar{x}) + \nabla^2_{xx}L(\bar{x},\bar{\mu})$ to $T_{\mathcal{M}}(\bar{x})$, is an one-to-one linear mapping from $T_{\mathcal{M}}(\bar{x})$ onto $T_{\mathcal{M}}(\bar{x})$, where L and $\bar{\mu}$ are taken from (d).*

Moreover, if any of the above equivalent properties holds, then the mapping s in (c) is $\mathcal{M}$-valued, namely $s(p) \in \mathcal{M}$ for any p in a neighborhood of $\bar{p}$, and $\mathcal{C}^1$-smooth around $\bar{p}$ with

$$\nabla s(p) = DG(x,p)^{-1} = \left(P_{T_{\mathcal{M}}(x)} \circ (\nabla\overline{\psi}(x) + \nabla^2_{xx}L(x,\mu))|_{T_{\mathcal{M}}(x)}\right)^{-1} \circ P_{T_{\mathcal{M}}(x)}. \qquad (3.25)$$

where $x = s(p)$ and $\mu \in \mathbb{R}^m$ is the unique vector satisfying $\nabla\Phi(x)^*\mu = p - (\overline{\psi}(x) + \nabla\widehat{f}(x))$.

Proof. Since $\bar{v} \in \mathrm{ri}\,\partial f(\bar{x})$, we infer from Theorem 3.10 that ∂f is strictly proto-differentiable at x for v for all $(x,v) \in \mathrm{gph}\,\partial f$ sufficiently close to $(\bar{x}, \bar{v})$ and its graphical derivative $D(\partial f)(x,v)$ is calculated by (3.14). Thus, the equivalence among assertions (a)–(d) is an immediate consequence of [11, Proposition 4.1, (a) $\Longleftrightarrow$ (b)] and [11, Theorems 4.2 and 4.3]. It remains to verify that (d) and (e) are equivalent. If (d) holds, taking the orthogonal complements from both sides of (3.24) and using [31, Corollary 11.25] tell us that (3.24) is equivalent to

$$N_{\mathcal{M}}(\bar{x}) + (\nabla\overline{\psi}(\bar{x}) + \nabla^2_{xx}L(\bar{x},\bar{\mu}))(T_{\mathcal{M}}(\bar{x})) = \mathbf{X}. \qquad (3.26)$$

This, combined with a similar argument as the proof of [11, Theorem 4.2, (a) $\Longleftrightarrow$ (b)], leads us to the desirable claim in (e). Conversely, assuming (e), we arrive at

$$\dim\left((\nabla\overline{\psi}(\bar{x}) + \nabla^2_{xx}L(\bar{x},\bar{\mu}))(T_{\mathcal{M}}(\bar{x}))\right) = \dim T_{\mathcal{M}}(\bar{x})$$

and

$$(\nabla\overline{\psi}(\bar{x}) + \nabla^2_{xx}L(\bar{x},\bar{\mu}))(T_{\mathcal{M}}(\bar{x})) \cap N_{\mathcal{M}}(\bar{x}) = \{0\},$$

17

which together give us (3.26), and hence (3.24). This completes the proof of the equivalence of (d) and (e).

Assume now that one of the equivalent properties in (a)–(e) is satisfied. By (c), we find neighborhoods U and V of $\bar{x}$ and $\bar{p}$, respectively, such that the mapping $s(p) := U \cap G^{-1}(p)$, $p \in V$, is single-valued and Lipschitz continuous on V. It then holds that $(s(p), p - \overline{\psi}(s(p))) \in (U \times \widetilde{V}) \cap \mathrm{gph}\, \partial f$, where $\widetilde{V}$ is an open set containing $(I_{\mathbf{X}} - \overline{\psi} \circ s)(V)$, where $I_{\mathbf{X}}$ stands for the identity mapping from $\mathbf{X}$ onto $\mathbf{X}$. Obviously, $\widetilde{V}$ is a neighborhood of $\bar{v} = \bar{p} - \overline{\psi}(\bar{x})$. Shrinking U and V, if necessary, we can conclude from (3.12) that $s(p) \in \mathcal{M}$ for all $p \in V$. Recall that ∂f is strictly proto-differentiable at x for v for all $(x,v) \in \mathrm{gph}\, \partial f$ near $(\bar{x}, \bar{v})$. We then conclude via [11, Theorem 4.3] that $s : V \to U \cap \mathcal{M}$ is $\mathcal{C}^1$-smooth around $\bar{p}$. We proceed by justifying (3.25). Observe that metric regularity of G holds at x for p for all $(x,p) \in \mathrm{gph}\, G$ sufficiently close to $(\bar{x}, \bar{p})$. By shrinking U and V, we can assume the latter property holds for all $(x,p) \in (U \times V) \cap \mathrm{gph}\, G$ and also differentiability of s at p. Pick now $p \in V$ and $u \in \mathbf{X}$. Since s is differentiable at p, we get that $w := \nabla s(p)(u) = Ds(p)(u) = DG^{-1}(p,x)(u)$, where $x := s(p) \in U$. The latter verifies the first equality in (3.25) and also gives us $u \in DG(x,p)(w)$. Employing the sum rule for graphical derivative from [31, Exercise 10.43(a)], we have $u \in \nabla\overline{\psi}(x)w + D(\partial f)(x, p - \nabla\overline{\psi}(x))(w)$. This, together with (3.14), yields $w \in T_{\mathcal{M}}(x)$ and $u \in \nabla\overline{\psi}(x)w + \nabla^2_{xx}L(x,\mu)w + N_{\mathcal{M}}(x)$, where μ is the unique vector satisfying $\nabla\Phi(x)^*\mu = p - (\nabla\overline{\psi}(x) + \nabla\widehat{f}(x))$. The equivalence (a) $\iff$ (e), together with G being metrically regular at x for p, tells us that the mapping $P_{T_{\mathcal{M}}(x)} \circ (\nabla\overline{\psi}(x) + \nabla^2_{xx}L(x,\mu))|_{T_{\mathcal{M}}(x)} : T_{\mathcal{M}}(x) \to T_{\mathcal{M}}(x)$ is one-to-one. Projecting u onto the subspace $T_{\mathcal{M}}(x)$, we have

$$P_{T_{\mathcal{M}}(x)}(u) = P_{T_{\mathcal{M}}(x)}((\nabla\overline{\psi}(x) + \nabla^2_{xx}L(x,\mu))(w)),$$

Since $w \in T_{\mathcal{M}}(x)$, we get from the latter that

$$w = \left(P_{T_{\mathcal{M}}(x)} \circ (\nabla\overline{\psi}(x) + \nabla^2_{xx}L(x,\mu))|_{T_{\mathcal{M}}(x)}\right)^{-1}(P_{T_{\mathcal{M}}(x)}(u)),$$

which confirms (3.25) and hence completes the proof. $\qquad\square$

Note that a characterization of continuous differentiability of the solution mapping s, defined in Proposition 3.18, which resembles the one in (3.25), was recently established in [5, Theorem 2.7] via a different approach. However, the equivalence between metric regularity and strong metric regularity was not obtained therein.

As an application of the Proposition 3.18 above, we present next $\mathcal{C}^1$-smoothness of the proximal mapping of $\mathcal{C}^2$-partly smooth functions. Moreover, we demonstrate that the latter proximal mapping is able to eventually identify the active manifold for this class of functions. The same results were established for projection mapping onto a prox-regular and $\mathcal{C}^2$-partly smooth set – a set is called $\mathcal{C}^2$-partly smooth if its indicator function is $\mathcal{C}^2$-partly smooth– in [14, Theorem 3.3], and for the proximal mapping of a prox-bounded, prox-regular, and $\mathcal{C}^2$-partly smooth function in [7, Theorem 28]. Proofs of these results rely mainly on [18, Theorem 5.7], which establishes stability of strong critical points of parametric $\mathcal{C}^2$-partly smooth functions. Below, we derive these results as an immediate consequence of Proposition 3.18. It is worth mentioning that while we will assume the subdifferential continuity in our result below, which was not assumed in [7], it is possible to drop using the discussion in Remark 3.8.

To proceed, recall that proximal mapping of a function $f : \mathbf{X} \to \overline{\mathbb{R}}$ for a parameter $r > 0$, denoted prox_{rf}, is defined by

$$\mathrm{prox}_{rf}(x) = \underset{w \in \mathbf{X}}{\mathrm{argmin}}\ \{f(w) + \tfrac{1}{2r}\|w - x\|^2\}, \quad x \in \mathbf{X}.$$

Recall also that f is said to be prox-bounded if there exists a real number α such that the function $f + \alpha\| \cdot \|^2$ is bounded from below on $\mathbf{X}$; see [31, Exercise 1.24].

Corollary 3.19. *Let $f : \mathbf{X} \to \overline{\mathbb{R}}$ be a $\mathcal{C}^2$-partly smooth function at $\bar{x}$ relative to a $\mathcal{C}^2$-smooth manifold $\mathcal{M}$ with local representation* (3.1) *and let $\bar{v} \in \operatorname{ri} \partial f(\bar{x})$. Assume that f is prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v}$ and that f is prox-bounded. Then for any $r > 0$ sufficiently small, the proximal mapping prox_{rf} is $\mathcal{M}$-valued and $\mathcal{C}^1$-smooth around $\bar{x} + r\bar{v}$ with*

$$\nabla(\operatorname{prox}_{rf})(z) = \left(P_{T_{\mathcal{M}}(x)} \circ (I_{\mathbf{X}} + r\nabla^2_{xx}L(x,\mu))|_{T_{\mathcal{M}}(x)}\right)^{-1} \circ P_{T_{\mathcal{M}}(x)}, \tag{3.27}$$

where L is taken from (3.13)*, $x = \operatorname{prox}_{rf}(z)$, and $\mu \in \mathbb{R}^m$ is the unique vector satisfying $\nabla_x L(x,\mu) = (z - x)/r$.*

Proof. Let $\epsilon > 0$ and $\rho \geq 0$ be constants for which (3.11) holds. Fixing $r \in (0, 1/\rho)$, with convention $1/0 = \infty$, we can deduce from [31, Theorem 13.37] that $\operatorname{prox}_{rf} = (I_{\mathbf{X}} + rT_\epsilon)^{-1}$, where T_ϵ is a localization of ∂f around $(\bar{x}, \bar{v})$ whose graph coincides with $\operatorname{gph} \partial f$ in $\mathbb{B}_\epsilon(\bar{x}, \bar{v})$, is moreover single-valued and Lipschitz continuous around $\bar{x} + r\bar{v}$. Considering $\overline{\psi} = I_{\mathbf{X}}$ and $\bar{p} = \bar{x} + r\bar{v}$, we then infer from Proposition 3.18 that prox_{rf}, a Lipschitz continuous single-valued localization of the inverse mapping of $G = I_{\mathbf{X}} + r\partial f$ around $\bar{x} + r\bar{v}$ for $\bar{x}$, is $\mathcal{M}$-valued and $\mathcal{C}^1$-smooth around that point. The formula for the Jacobian of prox_{rf} at z sufficiently close to $\bar{x} + r\bar{v}$ in (3.27) follows immediately from that in (3.25), which completes the proof. $\qquad\square$

We close this section with an extension of Proposition 3.18 into stability analysis of the generalized equation

$$0 \in \psi(p, x) + \partial f(x) \tag{3.28}$$

associated with the subgardient mapping of a $\mathcal{C}^2$-partly smooth function. We will deal with the mapping $\psi : \mathcal{P} \times \mathbf{X} \to \mathbf{X}$ that concerns with a general perturbation represented by parameter $p \in \mathcal{P}$, where $\mathcal{P} \subset \mathbf{P}$ is an open subset of a normed space $\mathbf{P}$. Define the solution mapping $S : \mathbf{P} \rightrightarrows \mathbf{X}$ by

$$S(p) := \left\{ x \in \mathbf{X} \,\middle|\, 0 \in \psi(p, x) + \partial f(x) \right\} \tag{3.29}$$

for $p \in \mathcal{P}$ and $S(p) = \emptyset$ otherwise. We are interested in conditions ensuring local single-valuedness and certain differential stability properties of S around a given parameter $\bar{p} \in \mathcal{P}$. With characterizations for strong metric regularity in Proposition 3.18, we are in a position to deliver the last result of this section establishing semidifferentiability and differentiability of the solution mapping in (3.29). This will be the driving force behind our investigation in the next section. Recall that for a set-valued mapping $F : \mathbf{X} \rightrightarrows \mathbf{Y}$ with $(\bar{x}, \bar{y}) \in \operatorname{gph} F$, if the limit vector

$$\lim_{\substack{t \searrow 0 \\ w' \to w}} \frac{F(\bar{x} + tw') - \bar{y}}{t}$$

exists for all $w \in \mathbf{X}$, then F is said to be semidifferentiable at $\bar{x}$ for $\bar{y}$ and the limit, which equals to $DF(\bar{x}, \bar{y})(w)$, is called the semiderivative of F at $\bar{x}$ for $\bar{y}$ and w. For a single-valued mapping F, we simply call the limit, denoted by $DF(\bar{x})(w)$, the semiderivative of F at $\bar{x}$ for w, and F is said to be semidifferentiable at $\bar{x}$. For convenience, we now recall in addition some notion from [6, Theorem 2B.7]–an extended implicit mapping theorem that will be exploited in our next proof. A mapping $\psi : \mathcal{P} \times \mathbf{X} \to \mathbf{X}$ is Lipschitz continuous with respect to p uniformly in x around $(\bar{p}, \bar{x})$ with modulus $\ell > 0$ if there exists $\epsilon > 0$ such that

$$\left\|\psi(p, x) - \psi(p', x)\right\| \leq \ell \|p - p'\| \quad \text{for all} \ \ p, p' \in \mathbb{B}_\epsilon(\bar{p}) \ \ \text{and all} \ \ x \in \mathbb{B}_\epsilon(\bar{x}). \tag{3.30}$$

A mapping $h : \mathbf{X} \to \mathbf{X}$ is a strict estimator of ψ with respect to x uniformly in p at $(\bar{p}, \bar{x})$ with constant $\tau > 0$ if $h(\bar{x}) = \psi(\bar{p}, \bar{x})$ and there exists $\epsilon >$ such that

$$\|r(p, x) - r(p, x')\| \leq \tau \|x - x'\| \ \ \text{for all} \ \ x, x' \in \mathbb{B}_\epsilon(\bar{x}) \ \ \text{and all} \ \ p \in \mathbb{B}_\epsilon(\bar{p}),$$

where $r(p, x) = \psi(p, x) - h(x)$.

Theorem 3.20. *Let $(\bar{p}, \bar{x}) \in \mathcal{P} \times \mathbf{X}$ and $\psi : \mathcal{P} \times \mathbf{X} \to \mathbf{X}$ satisfy the following conditions:*

(i) *for any $p \in \mathcal{P}$ near $\bar{p}$, the mapping $x \mapsto \psi(p, x)$ is $\mathcal{C}^1$-smooth around $\bar{x}$ and $\nabla_x \psi(\cdot, \cdot)$ is jointly continuous at $(\bar{p}, \bar{x})$; and*

(ii) *ψ is Lipschitz continuous with respect to p uniformly in x around $(\bar{p}, \bar{x})$.*

Assume that $f : \mathbf{X} \to \overline{\mathbb{R}}$ satisfies the hypotheses in Proposition 3.18, $-\psi(\bar{p}, \bar{x}) \in \operatorname{ri} \partial f(\bar{x})$ and that (3.24) holds for $\overline{\psi} = \psi(\bar{p}, \cdot)$. Then, there exist a neighborhood U of $\bar{x}$ and a positive constant ε such that $U \cap S(p)$ is a singleton for all $p \in \mathbb{B}_\varepsilon(\bar{p})$ and the mapping $s := U \cap S(\cdot)$ is $\mathcal{M}$-valued and Lipschitz continuous in $\mathbb{B}_\varepsilon(\bar{p})$, where the solution mapping S is taken from (3.29). Moreover,

(a) *if ψ has a strong partial semiderivative with respect to p at $(\bar{p}, \bar{x})$ in the sense that its partial graphical derivative $D_p \psi(\bar{p}, \bar{x}) := D(\psi(\cdot, \bar{x}))(\bar{p}) : \mathbf{P} \rightrightarrows \mathbf{X}$ is everywhere single-valued and for every $\nu > 0$ there exists $\epsilon > 0$ such that*

$$\|\psi(\bar{p} + q, x) - \psi(\bar{p}, x) - D_p \psi(\bar{p}, \bar{x})(q)\| \leq \nu \|q\| \tag{3.31}$$

for all $x \in \mathbb{B}_\epsilon(\bar{x})$ and all $q \in \epsilon \mathbb{B} \subset \mathbf{P}$, then s is semidifferentiable at $\bar{p}$ with

$$Ds(\bar{p})(q) = -DG(\bar{x}, 0)^{-1}(D_p \psi(\bar{p}, \bar{x})(q)), \quad q \in \mathbf{P}, \tag{3.32}$$

where $G := \psi(\bar{p}, \cdot) + \partial f$ and $DG(\bar{x}, 0)^{-1}$ is calculated by (3.25);

(b) *if ψ is differentiable with respect to p at $(\bar{p}, \bar{x})$ and for every $\nu > 0$ there exists $\epsilon > 0$ such that (3.31) with $\nabla_p \psi(\bar{p}, \bar{x})(\cdot)$ in the place of $D_p \psi(\bar{p}, \bar{x})(\cdot)$ is satisfied for all $x \in \mathbb{B}_\epsilon(\bar{x})$ and all $q \in \epsilon \mathbb{B} \subset \mathbf{P}$, then s is differentiable at $\bar{p}$ with*

$$\nabla s(\bar{p})(q) = -DG(\bar{x}, 0)^{-1}(\nabla_p \psi(\bar{p}, \bar{x})(q)), \quad q \in \mathbf{P},$$

where G and $DG(\bar{x}, 0)$ are defined in (a).

Proof. It is clear from $-\psi(\bar{p}, \bar{x}) \in \operatorname{ri} \partial f(\bar{x})$ that $\bar{x} \in S(\bar{p})$. We shall utilize [6, Theorem 2B.7] to establish the existence of a Lipschitz continuous single-valued localization of S around $\bar{p}$ for $\bar{x}$. To begin with, recall that $\overline{\psi} = \psi(\bar{p}, \cdot)$ is $\mathcal{C}^1$-smooth around $\bar{x}$ and $-\overline{\psi}(\bar{x}) = -\psi(\bar{p}, \bar{x}) \in \operatorname{ri} \partial f(\bar{x})$ and that the condition (3.24) is fulfilled. We then infer from Proposition 3.18 that the inverse mapping G^{-1} of $G := \overline{\psi} + \partial f$ has a Lipschitz continuous single-valued localization around 0 for $\bar{x}$ that is $\mathcal{M}$-valued and $\mathcal{C}^1$-smooth around 0. Let U be a neighborhood of $\bar{x}$ and $\epsilon' > 0$ such that

$$\gamma(v) := U \cap G^{-1}(v) \tag{3.33}$$

is a singleton for all $v \in \epsilon' \mathbb{B}$ and the mapping $\gamma : \epsilon' \mathbb{B} \to U \cap \mathcal{M}$ is $\mathcal{C}^1$-smooth. Let $\kappa > 0$ be the Lipschitzian constant of γ in $\epsilon' \mathbb{B}$. We now show that $\psi(\bar{p}, \cdot)$ is a strict estimator of ψ with respect to x uniformly in p at $(\bar{p}, \bar{x})$ with a constant $\tau \in (0, \kappa^{-1})$. Let $r(p, x) := \psi(p, x) - \psi(\bar{p}, x)$ for $(p, x) \in \mathcal{P} \times \mathbf{X}$ and $\epsilon'' > 0$ be such that $\mathbb{B}_{\epsilon''}(\bar{p}) \subset \mathcal{P}$. Pick $p \in \mathbb{B}_{\epsilon''}(\bar{p})$ and $x, x' \in \mathbb{B}_{\epsilon''}(\bar{x})$ arbitrarily and observe that

$$\begin{aligned}
\|r(p, x) - r(p, x')\| &\leq \sup_{y \in [x, x']} \|\nabla_x r(p, y)\| \cdot \|x - x'\| \\
&= \sup_{y \in [x, x']} \|\nabla_x \psi(p, y) - \nabla_x \psi(\bar{p}, y)\| \cdot \|x - x'\| \leq \tau \|x - x'\|,
\end{aligned}$$

where $[x, x'] := \{(1 - t)x + tx' \mid 0 \leq t \leq 1\}$ and $\tau := \sup_{p \in \mathbb{B}_{\epsilon''}(\bar{p}), y \in \mathbb{B}_{\epsilon''}(\bar{x})} \|\nabla_x \psi(p, y) - \nabla_x \psi(\bar{p}, y)\|$. Recalling that $\nabla_x \psi$ is jointly continuous at $(\bar{p}, \bar{x})$, we can assume without loss of generality that $\tau < \kappa^{-1}$, for otherwise we could select a smaller ϵ''. Thus, $\overline{\psi} = \psi(\bar{p}, \cdot)$ is a strict estimator of ψ with respect to x uniform in p at $(\bar{p}, \bar{x})$ with constant τ satisfying $\kappa \tau < 1$. Appealing now to [6, Theorem 2B.7], we conclude that the solution mapping S from (3.29) has a Lipschitz continuous

20

single-valued localization around $\bar{p}$ for $\bar{x}$. We then find a neighborhood of $\bar{x}$, which we denote by the same U as above for simplicity, and $\varepsilon > 0$ such that the mapping $s : \mathcal{P} \supset \mathbb{B}_\varepsilon(\bar{p}) \to U$, defined by $s(p) = U \cap S(p)$, is single-valued and Lipschitz continuous on $\mathbb{B}_\varepsilon(\bar{p})$. In view of (3.29), we get for all $p \in \mathbb{B}_\varepsilon(\bar{p})$ that

$$-(\psi(p, s(p)) - \psi(\bar{p}, s(p))) \in (\overline{\psi} + \partial f)(s(p)),$$

or, equivalently, $s(p) \in G^{-1}(r(p, s(p)))$. Shrinking neighborhoods of $\bar{x}$ and $\bar{p}$ if necessary, we can infer from (3.30) that $\|r(p, s(p))\| \leq \ell\|p - \bar{p}\| < \epsilon'$ for all $p \in \mathbb{B}_\varepsilon(\bar{p})$. Using (3.33), we arrive at $s(p) \in U \cap G^{-1}(r(p, s(p))) = \{\gamma(r(p, s(p)))\} \subset \mathcal{M}$ for all $p \in \mathbb{B}_\varepsilon(\bar{p})$. In summary, we have showed that S has a single-valued localization $s : \mathbb{B}_\varepsilon(\bar{p}) \to U$ that is $\mathcal{M}$-valued and Lipschitz continuous on $\mathbb{B}_\varepsilon(\bar{p})$.

We now turn to differential stability properties of the mapping s above. Due to Lipschitz continuity of s on $\mathbb{B}_\varepsilon(\bar{p})$, to prove that s is semidifferentiable (respectively, differentiable) at $\bar{p}$ it suffices to show that $Ds(\bar{p})$ is single-valued (respectively, $Ds(\bar{p})$ is single-valued and linear); see [31, Exercise 9.25(a)–(b)]. Pick an arbitrary $q \in \mathbf{P}$ and assume that $w \in Ds(\bar{p})(q)$. By definition, there are sequences $t_k \searrow 0$, $q^k \to q$ in $\mathbf{P}$, and $w^k \to w$ in $\mathbf{X}$ with $\bar{x} + t_k w^k = s(\bar{p} + t_k q^k)$. The latter yields $\bar{x} + t_k w^k \in U$ and

$$\begin{aligned}
0 &\in \psi(\bar{p} + t_k q^k, \bar{x} + t_k w^k) + \partial f(\bar{x} + t_k w^k) \\
&= G(\bar{x} + t_k w^k) + \psi(\bar{p} + t_k q^k, \bar{x} + t_k w^k) - \psi(\bar{p}, \bar{x} + t_k w^k)
\end{aligned}$$

Let $\nu > 0$ be arbitrarily small. By (a), ψ has a strong partial semiderivative with respect to p at $(\bar{p}, \bar{x})$. So, we get from the above inclusion that

$$-D_p\psi(\bar{p}, \bar{x})(q^k) \in G(\bar{x} + t_k w^k)/t_k + \nu\|q^k\|\mathbb{B}$$

for all k sufficiently large. Passing to the limit as $k \to \infty$ and observing that ν was arbitrary give us $-D_p\psi(\bar{p}, \bar{x})(q) \in DG(\bar{x}, 0)(w)$, or, equivalently, $w \in DG(\bar{x}, 0)^{-1}(-D_p\psi(\bar{p}, \bar{x})(q))$. It follows from (3.24) and Proposition 3.18 that $DG(\bar{x}, 0)^{-1}$ is calculated by (3.25). Thus, $DG(\bar{x}, 0)^{-1}$ is single-valued and linear. We then have

$$Ds(\bar{p})(q) = \{w\} = -DG(\bar{x}, 0)^{-1}(D_p\psi(\bar{p}, \bar{x})(q)).$$

Since $q \in \mathbf{P}$ was chosen arbitrarily, the latter verifies semidifferentiability of s at $\bar{p}$ and (3.32). For (b), we can argue similarly to drive

$$Ds(\bar{p})(q) = \{w\} = -DG(\bar{x}, 0)^{-1}(\nabla_p\psi(\bar{p}, \bar{x})(q)) \quad \text{for all } q \in \mathbf{P}.$$

Since $DG(\bar{x}, 0)^{-1}$ is single-valued and linear, the latter justifies differentiability of s at $\bar{p}$ and the claimed formula for $\nabla s(\bar{p})$ and hence completes the proof. $\qquad\square$

4 Asymptotics of Stochastic Programs with $\mathcal{C}^2$-Partly Smooth Regularizers

Consider the generic regularized stochastic problem

$$\underset{x \in \mathbf{X}}{\text{minimize}} \quad f(x) := \varphi(x) + \theta(x), \quad \text{with } \varphi(x) := \mathbb{E}[\widehat{\varphi}(x; \mathsf{Z})], \tag{P}$$

where Z is a random vector whose probability distribution $\mathbb{P}$ is supported on a measurable space $(\Xi, \mathscr{A})$ and stands for a data sample and $\widehat{\varphi} : \mathbf{X} \times \Xi \to \overline{\mathbb{R}}$ is a loss function. The regularizer

$\theta : \mathbf{X} \to \overline{\mathbb{R}}$ in (P) may either encode a geometric constraint or promote desirable low-complexity structure, such as sparsity or low-rank, for a minimizer. We assume that the function $\varphi(\cdot) = \mathbb{E}[\widehat{\varphi}(\cdot; \mathsf{Z})]$, is well-defined within some given domain, where the expectation $\mathbb{E}$ is taken with respect to $\mathsf{Z} \sim \mathbb{P}$ on $(\Xi, \mathscr{A})$. Let $\mathcal{U}$ be a bounded domain in $\mathbf{X}$, where we expect that a solution $\bar{x}$ to the true problem (P) exists. For each $x \in \mathcal{U}$, the expected value $\varphi(x)$ can be approximated through a *sample average approximation* (SAA) estimator. Namely, given k data samples $\mathsf{Z}_1, \ldots, \mathsf{Z}_k$, independent and identically distributed (iid) copies of the random vector Z, we use the averaging function of $\widehat{\varphi}(\cdot; \mathsf{Z}_i)$, $i = 1, \ldots, k$, in place of φ in (P) and then consider the so-called *SAA problem* of (P), which is formulated as

$$\underset{x \in \mathbf{X}}{\text{minimize}} \quad f_k(x) := \frac{1}{k} \sum_{i=1}^{k} \widehat{\varphi}(x; \mathsf{Z}_i) + \theta(x). \tag{P_k}$$

The emphasis of this section is on asymptotics of sequences of solutions to the SAA problem (P_k). Our analysis closely follows developments of sensitivity analysis and asymptotic theory made in [16, 17] by King and Rockafellar for (stochastic) generalized equations under proto-differentiability, *subinvertibility*, and single-valuedness of the inverse graphical derivatives; see [17, Assumptions M.2–M.4] and discussions therein. According to [16, Section 3], a set-valued mapping $F : \mathbf{X} \rightrightarrows \mathbf{Y}$ is subinvertible at $(\bar{x}, 0)$ if $0 \in F(\bar{x})$ and there exist a compact convex neighborhood $U \subset \mathbf{X}$ of $\bar{x}$, a positive constant ε, and a nonempty convex-valued mapping $G : \varepsilon \mathbb{B} \rightrightarrows U$ such that $\bar{x} \in G(0)$ and $\mathrm{gph}\, G$ is a closed subset of $(\varepsilon \mathbb{B} \times U) \cap \mathrm{gph}\, F^{-1}$. If F^{-1} admits a continuous selection $x(y)$ around 0 with $x(0) = \bar{x}$, then F is clearly subinvertible at $(\bar{x}, 0)$. Assume now that $F : \mathbf{X} \rightrightarrows \mathbf{X}$ is a maximal monotone mapping; see [31, Definitions 12.1 and 12.5] for the definition of this concept. It was shown in [16, Theorem 5.1] that the condition $DF(\bar{x}, 0)^{-1}(0) = \{0\}$ is sufficient for subinvertibility of F at $(\bar{x}, 0)$. Recalling also from [31, Theorem 12.65] that the single-valued property of $DF(\bar{x}, 0)^{-1} = DF^{-1}(0, \bar{x})$ is equivalent to semidifferentiability of F^{-1} at 0 for $\bar{x}$. Recall that Theorem 3.20 provides sufficient conditions for semidifferetiability of solution mappings to generalized equations associated with $\mathcal{C}^2$-partly smooth functions. In order to exploit our analysis in Section 3 in analyzing asymptotics of the stochastic generalized equation associated with (P), we focus on objective functions that consist of a $\mathcal{C}^2$-smooth loss function φ in conjunction with a $\mathcal{C}^2$-partial smoothness regularizer θ. Fixing these settings, we are going to study the asymptotic distribution of solutions to

$$0 \in \frac{1}{k} \sum_{i=1}^{k} \nabla_x \widehat{\varphi}(x; \mathsf{Z}_i) + \partial \theta(x), \tag{4.1}$$

which is referred to as the SAA generalized equation of the true variational system corresponding to problem (P) given by

$$0 \in \mathbb{E}[\nabla_x \widehat{\varphi}(x; \mathsf{Z})] + \partial \theta(x). \tag{4.2}$$

To begin with, we formally state standing assumptions in our analysis.

Assumption 4.1 (integrability and smoothness). (S.1) The function $\widehat{\varphi} : \mathcal{U} \times \Xi \to \overline{\mathbb{R}}$ is measurable in the second variable $\xi \in \Xi$ for each $x \in \mathcal{U}$, a bounded domain in $\mathbf{X}$.

(S.2) For almost every $\xi \in \Xi$, the function $x \mapsto \widehat{\varphi}(x; \xi)$ is $\mathcal{C}^2$-smooth on $\mathcal{U}$.

(S.3) There exists $g : \Xi \to \mathbb{R}_+$ with $\mathbb{E}[g(\mathsf{Z})^2] < \infty$ such that

$$\max \left\{ |\widehat{\varphi}(x; \xi)|, \ \|\nabla_x \widehat{\varphi}(x; \xi)\|, \ \|\nabla_{xx}^2 \widehat{\varphi}(x; \xi)\| \right\} \leq g(\xi)$$

for almost every $\xi \in \Xi$ and for all $x \in \mathcal{U}$.

(S.4) There exists $\ell : \Xi \to \mathbb{R}_{++}$ with $\mathbb{E}[\ell(\mathsf{Z})] < \infty$ satisfying

$$\max\big\{|\widehat{\varphi}(x;\xi) - \widehat{\varphi}(x';\xi)|, \|\nabla_x \widehat{\varphi}(x;\xi) - \nabla_x \widehat{\varphi}(x';\xi)\|\big\} \leq \ell(\xi)\|x - x'\|$$

for almost every $\xi \in \Xi$ and for all $x, x' \in \mathcal{U}$.

Here, in (S.3)–(S.4), the gradients $\nabla_x \widehat{\varphi}(\cdot;\xi)$ and Hessians $\nabla^2_{xx}\widehat{\varphi}(\cdot;\xi)$ of $\widehat{\varphi}(\cdot;\xi)$ exist for almost every $\xi \in \Xi$, by (S.2).

Note from assumptions (S.1)–(S.4) and [33, Theorem 7.52] that the expectation function $\varphi = \mathbb{E}[\widehat{\varphi}(\cdot;\mathsf{Z})]$ is well-defined and, moreover, twice continuously differentiable with

$$\nabla\varphi(x) = \mathbb{E}[\nabla_x \widehat{\varphi}(x;\mathsf{Z})] \quad \text{and} \quad \nabla^2\varphi(x) = \mathbb{E}[\nabla^2_{xx}\widehat{\varphi}(x;\mathsf{Z})], \quad x \in \mathcal{U}.$$

Assumption 4.2 (analytical assumptions). Assume that $\bar{x} \in \operatorname{dom} \theta$ for which the following conditions are satisifed:

(A.1) $-\nabla\varphi(\bar{x}) \in \operatorname{ri} \partial\theta(\bar{x})$.
(A.2) θ is $\mathcal{C}^2$-partly smooth at $\bar{x}$ relative to a $\mathcal{C}^2$-smooth manifold $\mathcal{M}$ with the local representation (3.1) and $\widehat{\theta} : \mathbf{X} \to \mathbb{R}$ is a $\mathcal{C}^2$-smooth representation of θ around $\bar{x}$ relative to $\mathcal{M}$; θ is prox-regular and subdifferentially continuous at $\bar{x}$ for $-\nabla\varphi(\bar{x})$.
(A.3) The second-order condition

$$\big\{w \in T_{\mathcal{M}}(\bar{x}) \,\big|\, \nabla^2_{xx}L(\bar{x}, \bar{\mu})(w) \in N_{\mathcal{M}}(\bar{x})\big\} = \{0\} \tag{4.3}$$

holds, where $L(x, \mu) := \varphi(x) + \widehat{\theta}(x) + \langle \mu, \Phi(x) \rangle$ for any $(x, \mu) \in \mathbf{X} \times \mathbb{R}^m$ and where $\bar{\mu} \in \mathbb{R}^m$ is the unique vector, satisfying $\nabla\Phi(\bar{x})^*\bar{\mu} = -(\nabla\varphi(\bar{x}) + \nabla\widehat{\theta}(\bar{x}))$.

Let $\mathbf{P}$ be a separable Banach space equipped with its Borel algebra $\mathscr{B}(\mathbf{P})$. Recall from [15, Section 3] that a sequence $\{\mathbf{p}_k\}_{k\in\mathbb{N}}$ of random vectors $\mathbf{p}_k : \Xi \to \mathbf{P}$ is said to satisfy a *generalized central limit* formula if there exist a limit $\bar{p} \in \mathbf{P}$, a sequence $\{t_k\}_{k\in\mathbb{N}}$ of positive scalars decreasing to 0, and a limit distribution q such that $(\mathbf{p}_k - \bar{p})/t_k \xrightarrow{\mathscr{D}} q$, where the notation $\xrightarrow{\mathscr{D}}$ stands for convergence in distribution. The following lemma establishes asymptotic behavior of the sequence of stochastic gradient mappings

$$\nabla\widehat{\varphi}_k := \tfrac{1}{k}\sum_{i=1}^{k} \nabla_x \widehat{\varphi}(\cdot;\mathsf{Z}_i), \quad k \in \mathbb{N}, \tag{4.4}$$

taken from (4.1). In what follows, we use $\mathcal{C}^1(\mathcal{U}, \mathbf{X})$ to denote the space of $\mathcal{C}^1$-smooth mappings $\phi : \mathcal{U} \to \mathbf{X}$ endowed with the norm

$$\|\phi\|_{1,\mathcal{U}} := \sup_{x\in\mathcal{U}} \|\phi(x)\| + \sup_{x\in\mathcal{U}} \|\nabla\phi(x)\|,$$

where $\phi(x) \in \mathbf{X}$ and $\nabla\phi(x) \in \mathcal{L}(\mathbf{X}, \mathbf{X})$, the space of linear operators from $\mathbf{X}$ into $\mathbf{X}$ itself equipped with the operator norm. And, as usual, $\mathscr{B}(\mathbf{X})$ and $\mathscr{B}(\mathcal{C}^1(\mathcal{U}, \mathbf{X}))$ stand for the Borel algebras on $\mathbf{X}$ and $\mathcal{C}^1(\mathcal{U}, \mathbf{X})$, respectively.

Lemma 4.3. *Suppose that $\widehat{\varphi} : \mathbf{X} \times \Xi :\to \overline{\mathbb{R}}$ satisfies Assumption 4.1. Then, for each k, the stochastic gradient mapping $\nabla\widehat{\varphi}_k$ in (4.4) is a random vector in $\mathcal{C}^1(\mathcal{U}, \mathbf{X})$, and $\nabla\widehat{\varphi}_k$ converges to $\nabla\varphi$ with probability one in $\mathcal{C}^1(\mathcal{U}, \mathbf{X})$. Moreover, one has*

$$\sqrt{k}\big(\nabla\widehat{\varphi}_k - \nabla\varphi\big) \xrightarrow{\mathscr{D}} \phi \tag{4.5}$$

as random vectors in $\mathcal{C}^1(\mathcal{U}, \mathbf{X})$, where ϕ is normally distributed with mean 0 and variance

$$\Sigma(x) := \mathbb{V}\mathrm{ar}\, \nabla_x \widehat{\varphi}(x;\mathsf{Z}) = \mathbb{E}[(\nabla_x \widehat{\varphi}(x;\mathsf{Z}) - \nabla\varphi(x))(\nabla_x \widehat{\varphi}(x;\mathsf{Z}) - \nabla\varphi(x))^\top], \quad x \in \mathcal{U}.$$

23

Proof. Note that $\nabla\widehat{\varphi}_k$ converging with probability one to $\nabla\varphi$ in $\mathcal{C}^1(\mathcal{U}, \mathbf{X})$ means that $\nabla\widehat{\varphi}_k(x)$ and $\nabla^2\widehat{\varphi}_k(x) := \frac{1}{k}\sum_{i=1}^{k} \nabla_{xx}^2\widehat{\varphi}(x; Z_i)$ converge with probability one to $\nabla\varphi(x)$ in $\mathbf{X}$ and to $\nabla^2\varphi(x)$ in $\mathcal{L}(\mathbf{X}, \mathbf{X})$, respectively, uniformly in $\mathcal{U}$. Both assertions are an adaptation of [33, Theorem 9.60] for $\nabla_x\widehat{\varphi}(\cdot; \xi)$ and $\nabla_{xx}^2\widehat{\varphi}(\cdot; \xi)$ that are finite-dimensional vector-valued mappings. The claimed generalized central limit formula in (4.5) follows from [17, Theorem A.3]. $\qquad\square$

We conclude this section with asymptotic distribution of the solution to the SAA generalized equation in (4.1).

Theorem 4.4 (asymptotic distribution). *Suppose that Assumptions 4.1 and 4.2 hold for $\widehat{\varphi}$ and $\varphi = \mathbb{E}[\widehat{\varphi}(\cdot; Z)]$ with respect to a bounded domain $\mathcal{U}$ in $\mathbf{X}$ containing $\bar{x}$. Then, there exists a neighborhood $U \subset \mathcal{U}$ of $\bar{x}$ such that, with probability one, for k sufficiently large, the SAA generalized equation in (4.1) admits a unique local solution $\mathbf{x}^k \in U$. Moreover, the sequence $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ converges to $\bar{x}$ with probability one and satisfies the generalized central limit formula*

$$\sqrt{k}(\mathbf{x}^k - \bar{x}) \xrightarrow{\mathscr{D}} -\left(P_{T_{\mathcal{M}}(\bar{x})} \circ \nabla_{xx}^2 L(\bar{x}, \bar{\mu})|_{T_{\mathcal{M}}(\bar{x})}\right)^{-1}(P_{T_{\mathcal{M}}(\bar{x})}(\phi(\bar{x}))), \tag{4.6}$$

where $\phi \in \mathcal{C}^1(\mathcal{U}, \mathbf{X})$ is taken from Lemma 4.3 and L and $\bar{\mu}$ are from Assumption 4.2(A.3).

Proof. Let $\mathbf{P} = \mathcal{C}^1(\mathcal{U}, \mathbf{X})$ and $\mathcal{P} \subset \mathbf{P}$ an open subset containing $\bar{p} := \nabla\varphi$. Consider the evaluation mapping $\psi : \mathcal{P} \times \mathcal{U} \to \mathbf{X}$, defined by $\psi(p, x) := p(x)$ for any $(p, x) \in \mathcal{P} \times \mathcal{U}$. This instance of ψ clearly satisfies assumptions (i) and (ii) in Theorem 3.20 at $(\bar{p}, \bar{x})$. It was also pointed out in [16, Remark 4.2] that ψ satisfies the assumption (a) in Theorem 3.20 at $(\bar{p}, \bar{x})$ with

$$D_p\psi(\bar{p}, \bar{x})(q) = q(\bar{x}) \quad \text{for all } q \in \mathcal{C}^1(\mathcal{U}, \mathbf{X}). \tag{4.7}$$

Let $U \subset \mathcal{U}$ and $\delta > 0$ be the neighborhood of $\bar{p}$ and the constant taken from Theorem 3.20. In the event of the convergence $\nabla\widehat{\varphi}_k \to \nabla\varphi$ in $\mathcal{C}^1(\mathcal{U}, \mathbf{X})$, which happens with probability one as shown in Lemma 4.3, Theorem 3.20 tells us that (4.1) eventually admits a unique local solution $\mathbf{x}^k \in U$. Compactness of U implies that the sequence $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ has at least one cluster point, says $\widehat{\mathbf{x}}$. Thus, there is a subsequence $\{\mathbf{x}^{k_i}\}_{i\in\mathbb{N}}$ of $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ that $\mathbf{x}^{k_i} \to \widehat{\mathbf{x}}$. It follows from the local Lipschitz continuity of the localization $s(\cdot) = U \cap S(\cdot)$ of the solution mapping

$$S(p) := \left\{x \in \mathbf{X} \,\middle|\, 0 \in \psi(p, x) + \partial\theta(x)\right\}, \quad p \in \mathcal{P},$$

on $\mathbb{B}_\delta(\nabla\varphi)$ that

$$\widehat{\mathbf{x}} = \lim_{i\to\infty} \mathbf{x}^{k_i} = \lim_{i\to\infty} s(\nabla\widehat{\varphi}_{k_i}) = s(\nabla\varphi) = \bar{x},$$

where the third equality holds with probability one. This tells us that all the convergent subsequences of $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ tend to $\bar{x}$ with probability one and therefore the sequence $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ converges to $\bar{x}$ with probability one. We can deduce from Theorem 3.20 together with (4.7) that the graphical localization s of S is almost surely semidifferentiable at $\bar{p} = \nabla\varphi$ for $\bar{x}$ with

$$Ds(\bar{p})(q) = -\left(P_{T_{\mathcal{M}}(\bar{x})} \circ \nabla_{xx}^2 L(\bar{x}, \bar{\mu})|_{T_{\mathcal{M}}(\bar{x})}\right)^{-1}\left(P_{T_{\mathcal{M}}(\bar{x})}(q(\bar{x}))\right), \quad q \in \mathcal{C}^1(\mathcal{U}, \mathbf{X});$$

see [15, Definition 3.1] for the definition of the latter differentiability notion. Applying [15, Theorem 3.2] to the continuous mapping $s : \mathbb{B}_\delta(\nabla\varphi) \to U$ and the sequence $\{\nabla\widehat{\varphi}_k\}_{k\in\mathbb{N}}$ that satisfies (4.5) by Lemma 4.3, we obtain the generalized central limit formula (4.6), which completes the proof. $\qquad\square$

Note that asymptotic distribution of SAA of a generalized equation was first established in [17, Theorem 2.7] under subinvertibility and the single-valuedness of the inverse of the graphical derivative of the solution mapping; see also [33, section 5.2.2] for a similar result for the

KKT system of nonlinear programming problems. These assumptions are automatically satisfied under Assumption 4.2. Furthermore, the latter result assumes the almost sure convergence of the sequence $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$ in Theorem 4.4, a fact that was not assumed in our result and indeed was proven as a consequence of Assumption 4.2. The asymptotic distribution of SAA of a generalized equation was recently achieved in [5, Theorem 3.1] when the solution mapping to the generalized equation has a $\mathcal{C}^1$ single-valued graphical localization. According to Theorem 3.20, our Assumption 4.2 ensures such a property for the solution mapping to (4.2). However, similar to [17, Theorem 2.7], [5, Theorem 3.1] assumes the almost sure convergence of the sequence $\{\mathbf{x}^k\}_{k\in\mathbb{N}}$. Note that while our framework is not as general as the one considered in [5], our approach clearly can be used for a generalized equation. Moreover, we can replace Assumption 4.2 with the mere assumption that $\partial\theta$ is strictly proto-differentiable and study the asymptotic distribution of solutions to the SAA approximation of the generalized equation

$$0 \in f(x) + \partial\theta(x),$$

where $f : \mathbf{X} \to \mathbb{R}$ is $\mathcal{C}^1$-smooth. Such a condition is weaker than assuming the solution mapping to the generalized equation has a $\mathcal{C}^1$ single-valued graphical localization, which was used in [5]. We will pursue this issue in our future works.

References

[1] G. Bareilles, F. Iutzeler, and J. Malick: *Newton acceleration on manifolds identified by proximal gradient methods*, Math. Program. 200 (2023) 37–70.

[2] J.V. Burke and A. Engle: *Strong metric (sub)regularity of KKT mappings for piecewise linear-quadratic convex composite optimization*, Math. Oper. Res. 45/3 (2020) 797–1192.

[3] A. Daniilidis, D. Drusvyatskiy, and A.S. Lewis: *Orthogonal invariance and identifiability*, SIAM J. Matrix Anal. Appl. 35/2 (2014) 580–598.

[4] A. Daniilidis, W. L. Hare, and J. Malick: *Geometrical interpretation of the predictor-corrector type algorithms in structured optimization problems*, Optimization 55/5–6 (2006), 481–503.

[5] D. Davis, D. Drusvyatskiy, and L. Jiang: *Asymptotic normality and optimality in nonsmooth stochastic approximation*, Ann. Statist. 52/4 (2024) 1485–1508.

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

[7] D. Drusvyatskiy, A. D. Ioffe, and A.S. Lewis: *Generic minimizing behavior in semialgebraic optimization*, SIAM J. Optim. 26/1 (2016) 513–534.

[8] D. Drusvyatskiy and A.S. Lewis, *Optimality, identifiability, and sensitivity*, ArXiv preprint, arXiv:1207.6628 (2012).

[9] N. T. V. Hang, W. Jung, and M. E. Sarabi: *Role of subgradients in variational analysis of polyhedral functions*, J. Optim. Theory Appl. 200 (2024) 1160–1192.

[10] N. T. V. Hang and M. E. Sarabi: *A chain rule for strict twice epi-differentiability and its applications*, SIAM J. Optim. 34/1 (2024) 918–945.

[11] N. T. V. Hang and M. E. Sarabi: *Smoothness of subgradient mappings and its applications in parametric optimization*, ArXiv preprint, `arXiv:2311.06026` (2023).

[12] W. L. Hare: *Nonsmooth Optimization with Smooth Substructure*, Ph.D. Dissertation, Simon Fraser University (2004).

[13] W. L. Hare: *Functions and sets of smooth substructure: relationships and examples*, Comput. Optim. Appl. 33 (2006) 249–270.

[14] W. L. Hare and A. S. Lewis: *Identifying active constraints via partial smoothness and prox-regularity*, J. Convex Anal. 11/2 (2004) 251–266.

[15] A. J. King: *Generalized delta theorems for multivalued mappings and measurable selections*, Math. Oper. Res. 14/4 (1989) 720–736.

[16] A. J. King and R. T. Rockafellar: *Sensitive analysis for nonsmooth generalized equations*, Math. Program., 55 (1992) 193–212.

[17] A. J. King and R. T. Rockafellar: *Asymptotic theory for solutions in statistical estimation and stochastic programming*, Math. Oper. Res. 18/1 (1993) 148–162.

Math. Program. 48 (1990) 285–301.

[18] A. S. Lewis: *Active sets, nonsmoothness, and sensitivity*, SIAM J. Optim. 13/3 (2002) 702–725.

[19] A. S. Lewis and C. Wylie: *Active-set Newton methods and partial smoothness*, Math. Oper. Res. 46/2 (2021) 712–725.

[20] A. S. Lewis and S. Zhang: *Partial smoothness, tilt stability, and generalized Hessians*, SIAM J. Optim. 23/1 (2013) 74–94.

[21] J. Liang, J. M. Fadili, and G. Peyré: *Local linear convergence of forward-backward under partial smoothness*, Adv. Neural Inf. Process. Syst. 27/2 (2014) 1970–1978.

[22] J. Liang, J. M. Fadili, and G. Peyré: *Activity identification and local liner convergence of forward–backward-type methods*, SIAM J. Optim. 27/1 (2017) 408–437.

[23] R. Mifflin and C. Sagastizábal: *Primal-dual gradient structured functions: second-order results; links to epi-derivatives and partly smooth functions*, SIAM J. Optim. 13/4 (2003) 1174–1194.

[24] A. Mohammadi, B. Mordukhovich and M. E. Sarabi: *Parabolic regularity via geometric variational analysis*, Trans. Amer. Soc. 374/3 (2021) 1711–1763.

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

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

[27] R. A. Poliquin and R. T. Rockafellar: *Prox-regular functions in variational analysis*, Trans. Amer. Math. Soc. 348/5 (1996) 1805–1838.

[28] R. A. Poliquin and R. T. Rockafellar: *Generalized Hessian properties of regularized nonsmooth functions*, SIAM J. Optim. 6/4 (1996) 1121–1137.

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

[30] R. T. Rockafellar: *Maximal monotone relations and the second derivatives of nonsmooth functions*, Ann. Inst. H. Poincaré Analyse Non Linéaire 2/3 (1985) 167–184.

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

[32] A. Shapiro: *On a class of nonsmooth composite functions*, Math. Oper. Res. 28/4 (2003) 677–692.

[33] A. Shapiro, D. Dentcheva, and A. Ruszczyński: *Lectures on Stochastic Programming: Modeling and Theory*, third edition, MOS-SIAM Series on Optimization, SIAM, Philadelphia (2021).

[34] C. Wylie: *Partly Smooth Models and Algorithms*, Ph.D. Dissertation, Cornell University (2019).