ROLE OF SUBGRADIENTS IN VARIATIONAL ANALYSIS OF POLYHEDRAL FUNCTIONS

NGUYEN T. V. HANG[1], WOOSUK JUNG [2], and M. EBRAHIM SARABI[3]

Abstract. Understanding the role that subgradients play in various second-order variational analysis constructions can help us uncover new properties of important classes of functions in variational analysis. Focusing mainly on the behavior of the second subderivative and subgradient proto-derivative of polyhedral functions, functions with polyhedral epigraphs, we demonstrate that choosing the underlying subgradient, utilized in the definitions of these concepts, from the relative interior of the subdifferential of polyhedral functions ensures stronger second-order variational properties such as strict twice epi-differentiability and strict subgradient proto-differentiability. This allows us to characterize continuous differentiability of the proximal mapping and twice continuous differentiability of the Moreau envelope of polyhedral functions. We close the paper with proving the equivalence of metric regularity and strong metric regularity of a class of generalized equations at their nondegenerate solutions.

Keywords Polyhedral functions, reduction lemma, nondegenerate solutions, strict proto-differentiability, strict twice epi-differentiability, proximal mappings, strong metric regularity.

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

1 Introduction

Second-order variational constructions such as the second subderivative and subgradient proto-derivative play an important role in parametric optimization and convergence analysis of important numerical algorithms such as the augmented Lagrangian method [7, 32]. Given an extended-real-valued function $f : \mathbb{R}^n \to \overline{\mathbb{R}} := [-\infty, \infty]$, these second-order variational constructions are defined at a point (x, v) in the graph of the subgradient mapping of f. One may wonder what impacts the selection of the subgradient v can have in these constructions. Our main goal in this paper is to study the underlying role that the selection of a subgradient can play for such constructions. To achieve this goal, we focus mainly on a particular class of convex functions called polyhedral functions. Recall that a proper function $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is called polyhedral if its epigraph is a polyhedral convex set. Given $\bar{z} \in \mathbb{R}^m$ with $g(\bar{z})$ finite, consider a subgradient $\bar{\lambda} \in \partial g(\bar{z})$. It is well-known (cf. [33, Proposition 13.9]) that the polyhedral function g is twice epi-differentiable at $\bar{z}$ for $\bar{\lambda}$, that the subgradient mapping ∂g is proto-differentiable at $\bar{z}$ for $\bar{\lambda}$, and that its proximal mapping is directionally differentiable at $\bar{z} + \bar{\lambda}$; see Sections 3 and 4 for the definitions of these concepts.

Should we expect further properties if, in addition, we assume that $\bar{\lambda} \in \operatorname{ri} \partial g(\bar{z})$? This is the main question that we are going to investigate for this class of convex functions. Note that such a relative interior condition has been utilized in several studies related to different numerical methods, including the partial smoothness [14] and the $\mathcal{U}$-Lagrangian of convex functions [12]. Therefore understanding the role that is played by this condition in second-order variational

[1]N. T. V. Hang: Institute of Mathematics, Vietnam Academy of Science and Technology, Hanoi, Vietnam (ntvhang@math.ac.vn).

[2]W. Jung: Department of Mathematics, Miami University, Oxford, OH 45065, USA (jungw2@miamioh.edu).

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

analysis can lead to stronger stability properties for important classes of functions in variational analysis as demonstrated in Sections 4 and 5.

Following this question, we uncover new second-order variational properties of polyhedral functions, including strict twice epi-differentiability and strict subgradient proto-differentiability, under this extra assumption. These findings allow us to achieve a useful characterization of continuous differentiability of the proximal mapping and twice continuous differentiability of the Moreau envelope of polyhedral functions. As an important application, we turn to study stability properties of the solution mapping to the generalized equation

$$0 \in \psi(x) + \partial g(x), \tag{1.1}$$

where $\psi : \mathbb{R}^m \to \mathbb{R}^m$ is a $\mathcal{C}^1$ mapping and $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a polyhedral function. Our interest is mainly in examining the relationship between metric regularity and strong metric regularity for (1.1). The seminal work of Donchev and Rockafellar in [3] demonstrated that these properties are equivalent for (1.1) when the polyhedral function g is the indicator function of a polyhedral convex set. Employing our new developments under the relative interior condition, we are going to show that if $\bar{x}$ is a *nondegenerate* solution to the generalized equation (1.1), meaning that it satisfies the condition

$$-\psi(\bar{x}) \in \operatorname{ri} \partial g(\bar{x}), \tag{1.2}$$

then metric regularity and strong metric regularity of (1.1) are equivalent. Furthermore, we show that the solution mapping to the canonical perturbation of (1.1) has a Lipschitz continuous single-valued localization, which is continuously differentiable. The latter smoothness of a localization of solution mappings of generalized equations resembles a similar conclusion from the classical inverse mapping theorem.

The rest of the paper is organized as follows. Section 2 contains definitions of important concepts, used in this paper. We also establish some properties of polyhedral functions. Section 3 begins with a new proof of the reduction lemma for polyhedral functions and then we present its important consequences in various second-order variational constructions. In particular, we show that under the relative interior condition, the subgradient mappings of polyhedral functions are strictly proto-differentiable. Section 4 is devoted to study strict twice epi-differentiablity of polyhedral functions. As an important consequence, we characterize continuous differentiability of the proximal mapping and twice continuous differentiability of the Moreau envelope of polyhedral functions. The final section, Section 5, concerns the equivalence of metric regularity and strong metric regularity for the generalized equation (1.1) under (1.2). Using this equivalence, we present sufficient conditions for a smooth single-valued localization of the solution mapping to the canonical perturbation of (1.1) as well as KKT systems of a subclass of composite optimization problems.

2 Notation and Preliminary Results

In what follows, we denote by $\mathbb{B}$ the closed unit ball in the space in question and by $\mathbb{B}_r(x) := x + r\mathbb{B}$ the closed ball centered at x with radius $r > 0$. In the product space $\mathbb{R}^n \times \mathbb{R}^m$, we use the norm $\|(w, u)\| = \sqrt{\|w\|^2 + \|u\|^2}$ for any $(w, u) \in \mathbb{R}^n \times \mathbb{R}^m$. Given a nonempty set $C \subset \mathbb{R}^n$, the symbols $\operatorname{int} C$, $\operatorname{ri} C$, $\operatorname{cone} C$, and $\operatorname{co} C$ signify its interior, relative interior, conic hull, and convex hull, respectively. For any set C in $\mathbb{R}^n$, its indicator function is defined by $\delta_C(x) = 0$ for $x \in C$ and $\delta_C(x) = \infty$ otherwise. We denote by P_C the projection mapping onto C and by $\operatorname{dist}(x, C)$ the distance between $x \in \mathbb{R}^n$ and a set C. For a vector $w \in \mathbb{R}^n$, the subspace $\{tw \mid t \in \mathbb{R}\}$ is denoted by $[w]$. The domain and range of a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ are defined, respectively, by $\operatorname{dom} F := \{x \in \mathbb{R}^n \mid F(x) \neq \emptyset\}$ and $\operatorname{rge} F = \{u \in \mathbb{R}^m \mid \exists\, w \in \mathbb{R}^n \text{ with } u \in F(x)\}$.

In this paper, the convergence of a family of sets is always understood in the sense of Painlevé-Kuratowski (cf. [33, Definition 4.1]). This means that the inner limit set of a parameterized family of sets $\{C^t\}_{t>0}$ in $\mathbb{R}^d$, denoted $\liminf_{t\searrow 0} C^t$, is the set of points x such that for every sequence $t_k \searrow 0$, x is the limit of a sequence of points $x^{t_k} \in C^{t_k}$. The outer limit set of this family of sets, denoted $\limsup_{t\searrow 0} C^t$, is the set of points x such that there exist sequences $t_k \searrow 0$ and $x^{t_k} \in C^{t_k}$ such that $x^{t_k} \to x$ as $k \to \infty$. A sequence $\{f^k\}_{k\in\mathbb{N}}$ of functions $f^k : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to *epi-converge* to a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ if we have $\operatorname{epi} f^k \to \operatorname{epi} f$ as $k \to \infty$; see [33, Definition 7.1] for more details on the epi-convergence of a sequence of extended-real-valued functions.

Given a nonempty set $\Omega \subset \mathbb{R}^n$ with $\bar{x} \in \Omega$, the tangent cone to Ω at $\bar{x}$, denoted $T_\Omega(\bar{x})$, is defined by

$$T_\Omega(\bar{x}) = \limsup_{t\searrow 0} \frac{\Omega - \bar{x}}{t}. \tag{2.1}$$

The regular/Fréchet normal cone $\widehat{N}_\Omega(\bar{x})$ to Ω at $\bar{x}$ is defined by $\widehat{N}_\Omega(\bar{x}) = T_\Omega(\bar{x})^*$, the polar of the tangent cone (2.1). The (limiting/Mordukhovich) normal cone $N_\Omega(\bar{x})$ to Ω at $\bar{x}$ is the set of all vectors $\bar{v} \in \mathbb{R}^n$ 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 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $\bar{x} \in \mathbb{R}^n$ with $f(\bar{x})$ finite, the subderivative function $\mathrm{d}f(\bar{x})\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is defined by

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

A vector $v \in \mathbb{R}^n$ is called a subgradient of f at $\bar{x}$ if $(v, -1) \in N_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))$ with $\operatorname{epi} f = \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \,|\, f(x) \le \alpha\}$ being the epigraph of f. The set of all subgradients of f at $\bar{x}$ is denoted by $\partial f(\bar{x})$. If f is a convex function, the latter set reduces to the well-known subdifferential of convex functions.

As pointed earlier, a proper function $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is called *polyhedral* if $\operatorname{epi} g$ is a polyhedral convex set. According to [33, Theorem 2.49], this class of convex functions enjoys the representation

$$g(z) = \begin{cases} \max\limits_{j\in J} \big\{\langle a^j, z\rangle - \alpha_j\big\} & \text{if } z \in \operatorname{dom} g, \\ \infty & \text{otherwise,} \end{cases}$$

where $J = \{1, \ldots, l\}$ for some $l \in \mathbb{N}$, $a^j \in \mathbb{R}^m$ and $\alpha_j \in \mathbb{R}$ for all $j \in J$, and where $\operatorname{dom} g = \{z \in \mathbb{R}^m \,|\, g(z) < \infty\}$ is a polyhedral convex set with the representation

$$\operatorname{dom} g = \big\{z \in \mathbb{R}^m \,\big|\, \langle b^i, z\rangle \le \beta_i, i \in I = \{1, \ldots, s\}\big\}, \tag{2.2}$$

where $s \in \mathbb{N}$, and $b^i \in \mathbb{R}^m$ and $\beta_i \in \mathbb{R}$ for all $i \in I$. Thus, we can equivalently express a polyhedral function g as

$$g(z) = \max_{j\in J} \big\{\langle a^j, z\rangle - \alpha_j\big\} + \delta_{\operatorname{dom} g}(z), \quad z \in \mathbb{R}^m. \tag{2.3}$$

It was observed in [18, Proposition 3.2] that $\operatorname{dom} g$ can be expressed as the finite union of the polyhedral convex sets $C_j, j \in J$, defined by

$$C_j = \big\{z \in \operatorname{dom} g \,\big|\, g(z) = \langle a^j, z\rangle - \alpha_j\big\} = \big\{z \in \operatorname{dom} g \,\big|\, \langle a^i - a^j, z\rangle \le \alpha_i - \alpha_j, i \in J\big\}.$$

Pick $\bar{z} \in \operatorname{dom} g$ and define the sets of active indices at $\bar{z}$ corresponding to the representation (2.2) and to the partition of $\operatorname{dom} g$ via the sets C_j by

$$I(\bar{z}) = \big\{i \in I \,\big|\, \langle b^i, \bar{z}\rangle = \beta_i\big\} \quad \text{and} \quad J(\bar{z}) = \big\{j \in J \,\big|\, \bar{z} \in C_j\big\}. \tag{2.4}$$

These sets allow us to conclude from [18, Proposition 3.3] that the subdifferential of g at $\bar z$ can be calculated as

$$\partial g(\bar z) = \text{co}\,\big\{a^j \,\big|\, j \in J(\bar z)\big\} + \text{cone}\,\big\{b^i \,\big|\, i \in I(\bar z)\big\}, \tag{2.5}$$

which tells us that any $\bar\lambda \in \partial g(\bar z)$ can be written as $\bar\lambda = \bar\lambda_1 + \bar\lambda_2$, where

$$\bar\lambda_1 = \sum_{j\in J(\bar z)} \bar\sigma_j a^j \quad\text{and}\quad \bar\lambda_2 = \sum_{i\in I(\bar z)} \bar\tau_i b^i \quad\text{with}\quad \bar\sigma_j, \bar\tau_i \ge 0 \quad\text{and}\quad \sum_{j\in J(\bar z)} \bar\sigma_j = 1. \tag{2.6}$$

Taking into account these representations, define the sets of positive coefficients at $\bar z$ for $\bar\lambda_1$ and $\bar\lambda_2$, respectively, by

$$J_+(\bar z, \bar\lambda_1) = \big\{j \in J(\bar z) \,\big|\, \bar\sigma_j > 0\big\} \quad\text{and}\quad I_+(\bar z, \bar\lambda_2) = \big\{i \in I(\bar z) \,\big|\, \bar\tau_i > 0\big\}. \tag{2.7}$$

The next result presents an important observation about the graph of subgradient mappings of polyhedral functions. A similar result was established in [18, Theorem 3.4] but the neighborhood obtained therein depends on a chosen representation of the subgradient $\bar\lambda$ as $\bar\lambda = \bar\lambda_1 + \bar\lambda_2$. Next, we show that such a neighborhood can be chosen to be independent of a given decomposition of $\bar\lambda$. We also simplify the proof presented in [18] significantly.

Lemma 2.1. *Assume that $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a polyhedral function and that $(\bar z, \bar\lambda) \in \text{gph}\,\partial g$. Then there exists $r > 0$ such that for any decomposition of $\bar\lambda$ as $\bar\lambda = \bar\lambda_1 + \bar\lambda_2$ with $\bar\lambda_1$ and $\bar\lambda_2$ taken from (2.6) and every $(z, \lambda) \in \big(\text{gph}\,\partial g\big) \cap \mathbb{B}_r(\bar z, \bar\lambda)$, we have*

$$J_+(\bar z, \bar\lambda_1) \subset J(z) \quad\textit{and}\quad I_+(\bar z, \bar\lambda_2) \subset I(z). \tag{2.8}$$

Proof. Suppose by contradiction that for each $k \in \mathbb{N}$, there is a decomposition $\bar\lambda = \bar\lambda_1^k + \bar\lambda_2^k$ with

$$\bar\lambda_1^k = \sum_{j\in J(\bar z)} \bar\sigma_j^k a^j \quad\text{and}\quad \bar\lambda_2^k = \sum_{i\in I(\bar z)} \bar\tau_i^k b^i \quad\text{with}\quad \bar\sigma_j^k, \bar\tau_i^k \ge 0 \quad\text{and}\quad \sum_{j\in J(\bar z)} \bar\sigma_j^k = 1, \tag{2.9}$$

and $(z^k, \lambda^k) \in \text{gph}\,\partial g$ such that $(z^k, \lambda^k) \to (\bar z, \bar\lambda)$ as $k \to \infty$ with $J_+^k \not\subset J(z^k)$ or $I_+^k \not\subset I(z^k)$ where $J_+^k := J_+(\bar z, \bar\lambda_1^k)$ and $I_+^k := I_+(\bar z, \bar\lambda_2^k)$ are defined via (2.7). Since $z^k \to \bar z$ as $k \to \infty$, the inclusions $J(z^k) \subset J(\bar z)$ and $I(z^k) \subset I(\bar z)$ hold for all k sufficiently large. Passing to a subsequence if necessary, we can assume that there exist subsets $\bar J \subset J(\bar z)$ and $\bar I \subset I(\bar z)$ such that

$$J(z^k) = \bar J \quad\text{and}\quad I(z^k) = \bar I \quad\text{for all } k \in \mathbb{N},$$

which, together with (2.5), lead us to

$$\partial g(z^k) = \text{co}\,\big\{a^j | \, j \in \bar J\big\} + \text{cone}\,\big\{b^i | \, i \in \bar I\big\} =: \Omega \quad\text{for all } k \in \mathbb{N}.$$

Since $\lambda^k \in \partial g(z^k) = \Omega$ and $\lambda^k \to \bar\lambda$ as $k \to \infty$, we arrive at $\bar\lambda \in \Omega = \partial g(z^k)$ for all k sufficiently large. Fix such a $k \in \mathbb{N}$ and deduce from $\bar\lambda \in \partial g(z^k)$ that

$$\langle \bar\lambda, \bar z - z^k \rangle \le g(\bar z) - g(z^k). \tag{2.10}$$

By the decomposition $\bar\lambda = \bar\lambda_1^k + \bar\lambda_2^k$ and (2.9), we obtain

$$\begin{aligned}
\langle \bar\lambda, z^k - \bar z \rangle &= \sum_{j\in J(\bar z)} \bar\sigma_j^k \langle a^j, z^k - \bar z\rangle + \sum_{i\in I(\bar z)} \bar\tau_i^k \langle b^i, z^k - \bar z\rangle \\
&= \sum_{j\in J(\bar z)} \bar\sigma_j^k \big(\langle a^j, z^k\rangle - \alpha_j - g(\bar z)\big) + \sum_{i\in I(\bar z)} \bar\tau_i^k \big(\langle b^i, z^k\rangle - \beta_i\big), \\
&\le \sum_{j\in J(\bar z)} \bar\sigma_j^k \big(g(z^k) - g(\bar z)\big),
\end{aligned} \tag{2.11}$$

where the second equality results from (2.4) and the last inequality comes from the fact that $z^k \in \operatorname{dom} g$, combined with (2.2)–(2.3). If $J_+^k \not\subset J(z^k)$, there exists $j_0 \in J_+^k$ such that $z^k \notin C_{j_0}$, meaning that

$$\bar{\sigma}_{j_0}^k\big(\langle a^{j_0}, z^k\rangle - \alpha_{j_0} - g(\bar{z})\big) < \bar{\sigma}_{j_0}^k\big(g(z^k) - g(\bar{z})\big).$$

If $I_+^k \not\subset I(z^k)$, we find $i_0 \in I_+^k$ such that $\langle b^{i_0}, z^k\rangle < \beta_{i_0}$, which implies that

$$\bar{\tau}_{i_0}^k\big(\langle b^{i_0}, z^k\rangle - \beta_{i_0}\big) < 0,$$

which tells us that in both cases the last inequality in (2.11) is strict. Thus, our assumption that either $J_+^k \not\subset J(z^k)$ or $I_+^k \not\subset I(z^k)$, together with the last condition in (2.9), yields

$$\langle \bar{\lambda}, z^k - \bar{z}\rangle < \sum_{j \in J(\bar{z})} \bar{\sigma}_j^k\big(g(z^k) - g(\bar{z})\big) = g(z^k) - g(\bar{z}),$$

which contradicts (2.10) and therefore completes the proof. □

We finish this section by recording some of first- and second-order variational properties of polyhedral functions, important for our developments in this paper. Recall from [33, Exercise 6.47] that for any polyhedral convex set $C \subset \mathbb{R}^n$ and $\bar{x} \in C$, we can find a neighborhood $\mathcal{O}$ of $\bar{x}$ for which we have

$$T_C(\bar{x}) \cap \mathcal{O} = (C - \bar{x}) \cap \mathcal{O}. \tag{2.12}$$

Proposition 2.2 (first-order variational properties)**.** *Assume that $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a polyhedral function and $\bar{z} \in \operatorname{dom} g$. Then the following properties hold.*

(a) *The domain of the subderivative function $\mathrm{d}g(\bar{z})$ can be calculated by*

$$\operatorname{dom} \mathrm{d}g(\bar{z}) = T_{\operatorname{dom} g}(\bar{z}) = \bigcup_{j \in J(\bar{z})} T_{C_j}(\bar{z}).$$

(b) *If $w \in T_{C_j}(\bar{z})$ for some $j \in J(\bar{z})$, then we have $\mathrm{d}g(\bar{z})(w) = \langle a^j, w\rangle$. Moreover, there exists $r > 0$ such that any $w \in T_{\operatorname{dom} g}(\bar{z}) \cap \mathbb{B}_r(0)$, the representation*

$$g(\bar{z} + w) = g(\bar{z}) + \mathrm{d}g(\bar{z})(w) \tag{2.13}$$

holds.

Proof. The first equality in (a) was established in [33, Proposition 10.21]. The second equality results immediately from the fact that $\operatorname{dom} g = \bigcup_{j \in J} C_j$. The first claim in (b) can be found again in [33, Proposition 10.21]. To prove (2.13), recall that for any $j \in J(\bar{z})$, C_j is a polyhedral convex set. Employing (2.12) for these sets, we can find $r > 0$ such that

$$T_{C_j}(\bar{z}) \cap \mathbb{B}_r(0) = (C_j - \bar{z}) \cap \mathbb{B}_r(0) \quad \text{for all } \ j \in J(\bar{z}).$$

Pick any $w \in T_{\operatorname{dom} g}(\bar{z}) \cap \mathbb{B}_r(0)$, and conclude from (a) that $w \in T_{C_{j_0}}(\bar{z}) \cap \mathbb{B}_r(0)$ for some $j_0 \in J(\bar{z})$ and from the first claim in (b) that $\mathrm{d}g(\bar{z})(w) = \langle a^{j_0}, w\rangle$. Thus, we get $\bar{z} + w \in C_{j_0}$, which, together with the definition of C_{j_0} and (2.4), leads us to

$$g(\bar{z} + w) = \langle a^{j_0}, \bar{z} + w\rangle - \alpha_{j_0} = g(\bar{z}) + \langle a^{j_0}, w\rangle = g(\bar{z}) + \mathrm{d}g(\bar{z})(w),$$

and hence completes the proof. □

Given a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $\bar{x} \in \mathbb{R}^n$ with $f(\bar{x})$ finite, the critical cone of f at $\bar{x}$ for $\bar{v}$ with $\bar{v} \in \partial f(\bar{x})$ is defined by

$$K_f(\bar{x}, \bar{v}) = \big\{ w \in \mathbb{R}^n \,\big|\, \langle \bar{v}, w \rangle = \mathrm{d}f(\bar{x})(w) \big\}.$$

If, in addition, f is convex and $\partial f(\bar{x}) \neq \emptyset$, it follows from [33, Theorem 8.30] that its subderivative function is the support function of $\partial f(\bar{x})$, that is

$$\mathrm{d}f(\bar{x})(w) = \sup \big\{ \langle v, w \rangle \,\big|\, v \in \partial f(\bar{x}) \big\},$$

which in turn allows us to equivalently describe the critical cone $K_f(\bar{x}, \bar{v})$ as

$$K_f(\bar{x}, \bar{v}) = N_{\partial f(\bar{x})}(\bar{v}). \tag{2.14}$$

The following equivalent description of the critical cone of a polyhedral function was established in [20, Proposition 3.2].

Proposition 2.3 (critical cone of polyhedral functions). *Assume that $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a polyhedral function and $(\bar{z}, \bar{\lambda}) \in \mathrm{gph}\, \partial g$. If $\bar{\lambda} = \bar{\lambda}_1 + \bar{\lambda}_2$ with $\bar{\lambda}_1$ and $\bar{\lambda}_2$ taken from (2.6), then $w \in K_g(\bar{z}, \bar{\lambda})$ if and only if w satisfies the conditions*

$$\begin{cases} \langle a^i - a^j, w \rangle = 0, & \text{for } \; i, j \in J_+(\bar{z}, \bar{\lambda}_1), \\ \langle a^i - a^j, w \rangle \leq 0, & \text{for } \; i \in J(\bar{z}) \setminus J_+(\bar{z}, \bar{\lambda}_1) \; \text{ and } \; j \in J_+(\bar{z}, \bar{\lambda}_1), \\ \langle b^i, w \rangle = 0, & \text{for } \; i \in I_+(\bar{z}, \bar{\lambda}_2), \\ \langle b^i, w \rangle \leq 0, & \text{for } \; i \in I(\bar{z}) \setminus I_+(\bar{z}, \bar{\lambda}_2). \end{cases}$$

3 Reduction Lemma for Polyhedral Functions and its Applications

We begin this section by providing an extension of the reduction lemma for polyhedral functions. The reduction lemma, established first by Robinson in [26, Proposition 4.4] for polyhedral convex sets, shows that the graph of the normal cone to a polyhedral convex set coincides locally with that of the normal cone to its critical cone. Robinson's proof of this result relies upon his sticky face lemma, which was established in [25, Lemma 3.5] with a rather involved proof. A simpler proof of this result was presented by Dontchev and Rockafellar in [2, Lemma 2E.4]. Recently, the reduction lemma was extended for an important class of convex functions, called piecewise linear-quadratic (cf. [33, Definition 10.20]), by the third author in [34, Theorem 2.3]. This class of convex functions clearly encompasses polyhedral functions and thus the result below can be derived from the recent result in [34]. However, the presented proof in [34] relies upon two major results: 1) Robinson's reduction lemma for polyhedral convex sets and 2) the fact that the graph of subgradient mappings of convex piecewise linear-quadratic functions can be expressed as a finite union of polyhedral convex sets (see the proof of [33, Theorem 11.14(b)]). Below, we present a direct proof of the reduction lemma for polyhedral functions, which is solely based on Lemma 2.1.

Theorem 3.1 (reduction lemma). *Assume that $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a polyhedral function and $(\bar{z}, \bar{\lambda}) \in \mathrm{gph}\, \partial g$. Then there exists $r > 0$ such that*

$$\big((\mathrm{gph}\, \partial g) - (\bar{z}, \bar{\lambda})\big) \cap \mathbb{B}_r(0, 0) = \big(\mathrm{gph}\, N_{K_g(\bar{z}, \bar{\lambda})}\big) \cap \mathbb{B}_r(0, 0). \tag{3.1}$$

Proof. Suppose that the polyhedral function g has the representation (2.3). We first show that there is $r>0$ for which the inclusion '$\supset$' in (3.1) holds. To do so, using (2.12) for the polyhedral convex set $\partial g(\bar z)$, we find $r_0>0$ such that

$$T_{\partial g(\bar z)}(\bar\lambda)\cap \mathbb{B}_{r_0}(0)=\big(\partial g(\bar z)-\bar\lambda\big)\cap \mathbb{B}_{r_0}(0). \tag{3.2}$$

Pick now a pair (w,u) from the right-hand side of (3.1) with $r=r_0$ therein, which implies that the conditions $w\in K_g(\bar z,\bar\lambda)\cap \mathbb{B}_{r_0}(0)$, $u\in K_g(\bar z,\bar\lambda)^*\cap \mathbb{B}_{r_0}(0)$, and $\langle u,w\rangle=0$ are satisfied. Moreover, using (2.14) for the polyhedral function g, we get $K_g(\bar z,\bar\lambda)^*=T_{\partial g(\bar z)}(\bar\lambda)$. Therefore u belongs to the left-hand side of (3.2), and thus $\bar\lambda+u\in\partial g(\bar z)$. We now show that $(\bar z+w,\bar\lambda+u)\in \operatorname{gph}\partial g$. To this end, take $z\in \mathbb{R}^m$ and observe that

$$\langle\bar\lambda+u,z-(\bar z+w)\rangle=\langle\bar\lambda+u,z-\bar z\rangle-\langle\bar\lambda,w\rangle-\langle u,w\rangle\le g(z)-g(\bar z)-\mathrm{d}g(\bar z)(w), \tag{3.3}$$

where the last inequality is deduced from the facts that $\bar\lambda+u\in\partial g(\bar z)$, $w\in K_g(\bar z,\bar\lambda)$, and $\langle u,w\rangle=0$. Choosing a smaller radius r_0 if necessary, we can assume that (2.13) is valid on $T_{\operatorname{dom} g}(\bar z)\cap\mathbb{B}_{r_0}(0)$. It follows from Proposition 2.2(a) and $w\in K_g(\bar z,\bar\lambda)\cap\mathbb{B}_{r_0}(0)$ that $w\in T_{\operatorname{dom} g}(\bar z)\cap\mathbb{B}_{r_0}(0)$, which in turn allows us to conclude via (2.13) that $g(\bar z)+\mathrm{d}g(\bar z)(w)=g(\bar z+w)$. Combining this and (3.3) leads us to

$$\langle\bar\lambda+u,z-(\bar z+w)\rangle\le g(z)-g(\bar z+w)$$

for any arbitrary $z\in\mathbb{R}^m$, and thus to $\bar\lambda+u\in\partial g(\bar z+w)$. This proves that (w,u) belongs to the left-hand side of (3.1) with $r=r_0$.

We now proceed with proving the inclusion '$\subset$' in (3.1) for some $r>0$. Decompose $\bar\lambda$ into $\bar\lambda=\bar\lambda_1+\bar\lambda_2$ with $\bar\lambda_1,\bar\lambda_2$ taken from (2.6). Pick r_0 from (3.2) and choose $r\in(0,r_0]$ such that the inclusions in (2.8) hold for all $(z,v)\in\big(\operatorname{gph}\partial g\big)\cap\mathbb{B}_r(\bar z,\bar\lambda)$. Pick now a pair (w,v) from the left-hand side of (3.1), and conclude that $(w,u)\in\mathbb{B}_r(0,0)$ with $\bar\lambda+u\in\partial g(\bar z+w)$. We are going to show that $u\in N_{K_g(\bar z,\bar\lambda)}(w)$, which amounts via [2, Proposition 2A.3] to the conditions

$$w\in K_g(\bar z,\bar\lambda),\quad u\in K_g(\bar z,\bar\lambda)^*,\quad\text{and}\quad \langle u,w\rangle=0. \tag{3.4}$$

Taking a smaller radius r if necessary, we can assume that $\partial g(\bar z+w)\subset\partial g(\bar z)$ (cf. [18, Proposition 3.3(a)]). Thus, $u\in\big(\partial g(\bar z)-\bar\lambda\big)\cap\mathbb{B}_r(0)$. Since $r\le r_0$, we deduce from (3.2) that $u\in T_{\partial g(\bar z)}(\bar\lambda)=K_g(\bar z,\bar\lambda)^*$, which justifies the second inclusion in (3.4). By shrinking r again if necessary, we can derive from $(\bar z+w,\bar\lambda+u)\in\big(\operatorname{gph}\partial g\big)\cap\mathbb{B}_r(\bar z,\bar\lambda)$ and (2.8) that

$$J_+(\bar z,\bar\lambda_1)\subset J(\bar z+w)\subset J(\bar z)\quad\text{and}\quad I_+(\bar z,\bar\lambda_2)\subset I(\bar z+w)\subset I(\bar z).$$

To prove that $w\in K_g(\bar z,\bar\lambda)$, we use Proposition 2.3 in which an equivalent description of $K_g(\bar z,\bar\lambda)$ was given. We break this task into four cases as follows:

(i) $i,j\in J_+(\bar z,\bar\lambda_1)$. In this case, we obtain from the inclusions above that $i,j\in J(\bar z+w)\subset J(\bar z)$. These inclusions bring us via (2.4) to

$$\langle a^i,\bar z\rangle-\alpha_i=\langle a^j,\bar z\rangle-\alpha_j\quad\text{and}\quad\langle a^i,\bar z+w\rangle-\alpha_i=\langle a^j,\bar z+w\rangle-\alpha_j.$$

Combining these confirms that $\langle a^i-a^j,w\rangle=0$ for all $i,j\in J_+(\bar z,\bar\lambda_1)$.

(ii) $i\in J(\bar z)\setminus J_+(\bar z,\bar\lambda_1)$ and $j\in J_+(\bar z,\bar\lambda_1)$. In this case, we arrive at $i,j\in J(\bar z)$ and $j\in J(\bar z+v)$, which imply via (2.4) that

$$\langle a^i,\bar z\rangle-\alpha_i=\langle a^j,\bar z\rangle-\alpha_j\quad\text{and}\quad\langle a^i,\bar z+w\rangle-\alpha_i\le\langle a^j,\bar z+w\rangle-\alpha_j.$$

Combining these confirms that $\langle a^i-a^j,w\rangle\le 0$ for all $i\in J(\bar z)\setminus J_+(\bar z,\bar\lambda_1)$ and $j\in J_+(\bar z,\bar\lambda_1)$.

(iii) $i \in I_+(\bar{z}, \bar{\lambda}_2)$. In this case, we have $i \in I(\bar{z})$ and $i \in I(\bar{z}+w)$, which result in $\langle b^i, \bar{z}\rangle = \langle b^i, \bar{z}+w\rangle = \beta_i$. Combining these confirms that $\langle b^i, w\rangle = 0$ for all $i \in I_+(\bar{z}, \bar{\lambda}_2)$.

(iv) $i \in I(\bar{z}) \setminus I_+(\bar{z}, \bar{\lambda}_2)$. In this case, we deduce from $\bar{z}+w \in \operatorname{dom} g$ and $i \in I(\bar{z})$ that $\langle b^i, \bar{z}+w\rangle \le \beta_i = \langle b^i, \bar{z}\rangle$, which in turn yields $\langle b^i, w\rangle \le 0$ for all $i \in I(\bar{z}) \setminus I_+(\bar{z}, \bar{\lambda}_2)$.

In summary, we showed that w satisfies the equivalent description of $K_g(\bar{z}, \bar{\lambda})$ from Proposition 2.3, and so $w \in K_g(\bar{z}, \bar{\lambda})$. It remains to demonstrate that $\langle w, u\rangle = 0$. Recall that $\bar{\lambda} + u \in \partial g(\bar{z}+w)$. Using (2.5), we can express $\bar{\lambda}+w$ as

$$\bar{\lambda} + u = \sum_{j\in J(\bar{z}+w)} \sigma_j a^j + \sum_{i\in I(\bar{z}+w)} \tau_i b^i \quad \text{with} \quad \sigma_j, \tau_i \ge 0 \;\text{ and } \sum_{j\in J(\bar{z}+w)} \sigma_j = 1.$$

This and the decomposition $\bar{\lambda} = \bar{\lambda}_1 + \bar{\lambda}_2$ with $\bar{\lambda}_1, \bar{\lambda}_2$ taken from (2.6) allow us to conclude that

$$\begin{aligned}
\langle u, w\rangle &= \langle \bar{\lambda}+u, w\rangle - \langle \bar{\lambda}, w\rangle \\
&= \sum_{j\in J(\bar{z}+w)} (\sigma_j - \bar{\sigma}_j)\langle a^j, w\rangle + \sum_{i\in I(\bar{z}+w)} (\tau_i - \bar{\tau}_i)\langle b^i, w\rangle \\
&= \sum_{j\in J(\bar{z}+w)} (\sigma_j - \bar{\sigma}_j)\big[\langle a^j, \bar{z}+w\rangle - \alpha_j - (\langle a^j, \bar{z}\rangle - \alpha_j)\big] + \sum_{i\in I(\bar{z}+w)} (\tau_i - \bar{\tau}_i)\big[\langle b^i, \bar{z}+w\rangle - \langle b^i, \bar{z}\rangle\big] \\
&= \big(g(\bar{z}+w) - g(\bar{z})\big) \sum_{j\in J(\bar{z}+w)} (\sigma_j - \bar{\sigma}_j) + \sum_{i\in I(\bar{z}+w)} (\tau_i - \bar{\tau}_i)(\beta_i - \beta_i) \\
&= 0,
\end{aligned}$$

where the second equality results from the inclusions $J_+(\bar{z}, \bar{\lambda}_1) \subset J(\bar{z}+w)$ and $I_+(\bar{z}, \bar{\lambda}_2) \subset I(\bar{z}+w)$, where the fourth equality comes from the definitions of the active index sets $J(\bar{z}+w)$ and $I(\bar{z}+w)$ as well as the inclusions $J(\bar{z}+w) \subset J(\bar{z})$ and $I(\bar{z}+w) \subset I(\bar{z})$, and where the last equality follows from the assumption on σ_j and $\bar{\sigma}_j$. This proves the inclusion '$\subset$' in (3.1) for some $r > 0$ and thus completes the proof. $\square$

The rest of this section will be devoted to presenting several important consequences of Theorem 3.1 for different second-order variational constructions of polyhedral functions. We begin with a duality relation between critical cones of a polyhedral function and its (Fenchel) conjugate in the sense of convex analysis. Both results in the following corollary were observed by Rockafellar in [32, equations (3.23) & (3.29)]. Since an explicit proof was not presented for (3.5) in [32] and since this result plays an important role in strict twice epi-differentiability of polyhedral functions in the next section, we supply a short proof for readers' convenience. Let us recall two important results, used in the proof of next result. The first is that the conjugate function g^* of a polyhedral function g is again a polyhedral function due to [33, Theorem 11.14(a)]. The second fact is that if C is a convex subset of $\mathbb{R}^n$ and $\bar{x} \in C$, then $N_C(\bar{x})$ is a linear subspace if and only if $\bar{x} \in \operatorname{ri} C$; see [13, Proposition 2.2].

Corollary 3.2 (polar relation of critical cones)**.** *Assume that $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a polyhedral function and $(\bar{z}, \bar{\lambda}) \in \operatorname{gph} \partial g$. Then $K_g(\bar{z}, \bar{\lambda})$ enjoys the duality relationship*

$$K_g(\bar{z}, \bar{\lambda}) = K_{g^*}(\bar{\lambda}, \bar{z})^*. \tag{3.5}$$

Consequently, we have

$$\bar{\lambda} \in \operatorname{ri} \partial g(\bar{z}) \;\; \textit{if and only if} \;\; \bar{z} \in \operatorname{ri} \partial g^*(\bar{z}). \tag{3.6}$$

Proof. Let $r > 0$ be such that (3.1) holds. Then, we conclude from (2.14) that

$$\begin{aligned} w \in K_g(\bar z, \bar\lambda) &\iff (w, 0) \in \operatorname{gph} N_{K_g(\bar z, \bar\lambda)} \\ &\iff \exists\, r' > 0 : t(w, 0) \in \big(\operatorname{gph} N_{K_g(\bar z, \bar\lambda)}\big) \cap \mathbb{B}_r(0, 0) \quad \text{for all} \quad t \in [0, r') \\ &\iff \exists\, r' > 0 : (\bar z + tw, \bar\lambda) \in \big(\operatorname{gph} \partial g\big) \cap \mathbb{B}_r(\bar z, \bar\lambda) \quad \text{for all} \quad t \in [0, r') \\ &\iff \exists\, r' > 0 : \bar\lambda \in \partial g(\bar z + tw) \quad \text{for all} \quad t \in [0, r') \\ &\iff \exists\, r' > 0 : \bar z + tw \in \partial g^*(\bar\lambda) \quad \text{for all} \quad t \in [0, r') \\ &\iff w \in T_{\partial g^*(\bar\lambda)}(\bar z) = \big(N_{\partial g^*(\bar\lambda)}(\bar z)\big)^* = K_{g^*}(\bar\lambda, \bar z)^*, \end{aligned}$$

where the third equivalence follows from (3.1) and the last one relies on the polyhedrality of $\partial g^*(\bar\lambda)$. To justify (3.6), it follows from $K_g(\bar z, \bar\lambda) = N_{\partial g(\bar z)}(\bar\lambda)$ that $K_g(\bar z, \bar\lambda)$ is a linear subspace if and only if $\bar\lambda \in \operatorname{ri} \partial g(\bar z)$. A similar observation can be made for $K_{g^*}(\bar\lambda, \bar z)$, which together with (3.5) completes the proof of (3.6). □

We proceed with presenting two important results for critical cones of polyhedral functions, which allow us to draw an important conclusion about the points in the graph of their subgradient mappings.

Theorem 3.3. *For a polyhedral function $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ and $(\bar z, \bar\lambda) \in \operatorname{gph} \partial g$, the following properties hold.*

(a) *There exists $r > 0$ such that for all $(z, \lambda) \in \big(\operatorname{gph} \partial g\big) \cap \mathbb{B}_r(\bar z, \bar\lambda)$ the following inclusions hold:*

$$K_g(z, \lambda) \subset K_g(\bar z, \bar\lambda) - K_g(\bar z, \bar\lambda), \tag{3.7a}$$

$$K_g(\bar z, \bar\lambda) \cap -K_g(\bar z, \bar\lambda) \subset K_g(z, \lambda). \tag{3.7b}$$

(b) *If, in addition, $\bar\lambda \in \operatorname{ri} \partial g(\bar z)$, then there exists $r > 0$ such that for all $(z, \lambda) \in \big(\operatorname{gph} \partial g\big) \cap \mathbb{B}_r(\bar z, \bar\lambda)$, we have $\lambda \in \operatorname{ri} \partial g(z)$ and*

$$K_g(z, \lambda) = K_g(\bar z, \bar\lambda). \tag{3.8}$$

Proof. We begin with proving (3.7a) for any $(z, \lambda) \in \operatorname{gph} \partial g$ sufficiently close to $(\bar z, \bar\lambda)$. Since the critical cone $K_g(\bar z, \bar\lambda)$ is a polyhedral convex cone, the latter inclusion is equivalent to the inclusion

$$K_g(z, \lambda)^* \supset \big(K_g(\bar z, \bar\lambda) - K_g(\bar z, \bar\lambda)\big)^* = K_g(\bar z, \bar\lambda)^* \cap -K_g(\bar z, \bar\lambda)^*,$$

which, by virtue of the relationships $K_g(z, \lambda)^* = (N_{\partial g(z)}(\lambda))^* = T_{\partial g(z)}(\lambda)$, amounts to

$$T_{\partial g(z)}(\lambda) \supset T_{\partial g(\bar z)}(\bar\lambda) \cap -T_{\partial g(\bar z)}(\bar\lambda).$$

Pick an arbitrary vector w from the right-hand side of this inclusion. For $(z, \lambda) \in \operatorname{gph} \partial g$ sufficiently close to $(\bar z, \bar\lambda)$, we can assume that

$$\partial g(z) \subset \partial g(\bar z) \quad \text{and} \quad T_{\partial g(\bar z)}(\bar\lambda) \subset T_{\partial g(\bar z)}(\lambda),$$

which imply that $\lambda \in \partial g(\bar z)$ and $w \in T_{\partial g(\bar z)}(\lambda)$, respectively. Employing (2.12) for the polyhedral convex set $\partial g(\bar z)$, we find a neighborhood $\mathcal{O}_\lambda$ of $0 \in \mathbb{R}^m$ such that

$$T_{\partial g(\bar z)}(\lambda) \cap \mathcal{O}_\lambda = \big(\partial g(\bar z) - \lambda\big) \cap \mathcal{O}_\lambda,$$

which, together with $w \in T_{\partial g(\bar z)}(\lambda)$, ensures the existence of $t_{\bar z, \lambda} > 0$ such that $\lambda + tw \in \partial g(\bar z)$ for any $t \in [0, t_{\bar z, \lambda}]$. Choosing $(z, \lambda) \in \operatorname{gph} \partial g$ sufficiently close to $(\bar z, \bar\lambda)$, we conclude from

Theorem 3.1 that $(z-\bar{z}, \lambda-\bar{\lambda}) \in \operatorname{gph} N_{K_g(\bar{z},\bar{\lambda})}$. This clearly yields $z-\bar{z} \in K_g(\bar{z},\bar{\lambda})$, which together with $w \in K_g(\bar{z},\bar{\lambda})^* \cap -K_g(\bar{z},\bar{\lambda})^*$ leads us to $\langle w, z-\bar{z}\rangle = 0$. We now claim that $\lambda+tw \in \partial g(z)$ for any $t \in [0, t_{\bar{z},\lambda}]$. To justify it, for any $z' \in \mathbb{R}^m$, we deduce from $\lambda+tw \in \partial g(\bar{z})$ and $\lambda \in \partial g(z)$ that

$$\begin{aligned}\langle \lambda+tw, z'-z\rangle &= \langle \lambda+tw, z'-\bar{z}\rangle + \langle \lambda, \bar{z}-z\rangle + t\langle w, \bar{z}-z\rangle \\ &\le g(z') - g(\bar{z}) + g(\bar{z}) - g(z) + 0 \\ &= g(z') - g(z),\end{aligned}$$

which proves our claim. It clearly follows from the latter claim that $w \in T_{\partial g(z)}(\lambda) = K_g(z,\lambda)^*$. This completes the proof of (3.7a).

We now proceed to justify the inclusion (3.7b) for all $(z,\lambda) \in \operatorname{gph} \partial g$ sufficiently close to $(\bar{z},\bar{\lambda})$. As mentioned before, the conjugate function g^* is a polyhedral function. Applying (3.7a) for g^* gives us the inclusion

$$K_{g^*}(\lambda,z) \subset K_{g^*}(\bar{\lambda},\bar{z}) - K_{g^*}(\bar{\lambda},\bar{z})$$

for all $(\lambda,z) \in \operatorname{gph} \partial g^*$ sufficiently close to $(\bar{\lambda},\bar{z})$. This inclusion, the polyhedrality of the critical cone of a polyhedral function, and (3.5) bring us to

$$\begin{aligned} K_g(z,\lambda) = K_{g^*}(\lambda,z)^* \quad &\supset \quad \big(K_{g^*}(\bar{\lambda},\bar{z}) - K_{g^*}(\bar{\lambda},\bar{z})\big)^* \\ &= \quad K_{g^*}(\bar{\lambda},\bar{z})^* \cap -K_{g^*}(\bar{\lambda},\bar{z})^* = K_g(\bar{z},\bar{\lambda}) \cap -K_g(\bar{z},\bar{\lambda}),\end{aligned}$$

which proves (3.7b).

To justify (b), it follows from $\bar{\lambda} \in \operatorname{ri} \partial g(\bar{z})$ and $K_g(\bar{z},\bar{\lambda}) = N_{\partial g(\bar{z})}(\bar{\lambda})$ that $K_g(\bar{z},\bar{\lambda})$ is a linear subspace. This and the inclusions (3.7a) and (3.7b) tell us that

$$K_g(\bar{z},\bar{\lambda}) = K_g(\bar{z},\bar{\lambda}) - K_g(\bar{z},\bar{\lambda}) \supset K_g(z,\lambda) \supset K_g(\bar{z},\bar{\lambda}) \cap -K_g(\bar{z},\bar{\lambda}) = K_g(\bar{z},\bar{\lambda})$$

for all $(z,\lambda) \in \operatorname{gph} \partial g$ sufficiently close to $(\bar{z},\bar{\lambda})$, which in turn confirms that $K_g(z,\lambda)$ is a linear subspace. This, together with $K_g(z,\lambda) = N_{\partial g(z)}(\lambda)$ taken from (2.14), results in $\lambda \in \operatorname{ri} \partial g(z)$ and hence completes the proof of (b). $\square$

Both inclusions in Theorem 3.3(a) were perviously established in [2, Proposition 2E.10] for polyhedral convex sets without appealing to the reduction lemma, an important tool used in our proof. Note that while the observation in Theorem 3.3(b) is a simple and direct consequence of the inclusions (3.7a)-(3.7b), it plays an indispensable role in the next two sections in which we are going to study strict twice epi-differentiability of polyhedral functions. In fact, this result reveals that when we are converging in the graph of subgradient mappings of polyhedral functions to a given point therein under the extra relative interior condition, all those points, used in this convergence, enjoy this relative interior condition. This has a major implication for second-order variational constructions such as limiting coderivatives and strict subgradient graphical derivatives; see Theorem 3.6(c) and Theorem 3.7(c).

One can proceed further to characterize critical cones of polyhedral functions for points nearby a given point in the graph of their subgradient mappings. Such a result for polyhedral convex sets can be found in [2, Lemma 4H.2]. Using a similar proof, we extend the latter observation for polyhedral functions. Recall that a closed face F of a polyhedral convex cone $C \subset \mathbb{R}^d$ is defined by

$$F = C \cap [v]^\perp \quad \text{for some } \; v \in C^*.$$

Proposition 3.4. *Assume that $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a polyhedral function and $(\bar z, \bar\lambda) \in \operatorname{gph} \partial g$. Then there exists $r > 0$ such that for all $(z, \lambda) \in \big(\operatorname{gph} \partial g\big) \cap \mathbb{B}_r(\bar z, \bar\lambda)$ the corresponding critical cone $K_g(z, \lambda)$ enjoys the following representation*

$$K_g(z, \lambda) = F_1 - F_2,$$

for some faces F_1, F_2 of $K_g(\bar z, \bar\lambda)$ with $F_2 \subset F_1$. Conversely, for any pair of faces F_1, F_2 of $K_g(\bar z, \bar\lambda)$ with $F_2 \subset F_1$ and any sufficiently small real number $r > 0$, there exists $(z, \lambda) \in \big(\operatorname{gph} \partial g\big) \cap \mathbb{B}_r(\bar z, \bar\lambda)$ with $K_g(z, \lambda) = F_1 - F_2$.

Proof. Let $r > 0$ be such that (3.1) holds. Choosing a smaller radius r if necessary, we can assume without loss of generality that the inclusions $\partial g(z) \subset \partial g(\bar z)$ and $N_{\partial g(\bar z)}(\lambda) \subset N_{\partial g(\bar z)}(\bar\lambda)$ hold for any $(z, \lambda) \in \big(\operatorname{gph} \partial g\big) \cap \mathbb{B}_r(\bar z, \bar\lambda)$, which can be guaranteed by g being a polyhedral function. We first show that for any $(z, \lambda) \in \big(\operatorname{gph} \partial g\big) \cap \mathbb{B}_r(\bar z, \bar\lambda)$, we have

$$K_g(z, \lambda) = K_g(\bar z, \bar\lambda) \cap [\lambda - \bar\lambda]^\perp + [z - \bar z]. \tag{3.9}$$

To this end, pick $(z, \lambda) \in \big(\operatorname{gph} \partial g\big) \cap \mathbb{B}_r(\bar z, \bar\lambda)$ and observe that the inclusion

$$N_{\partial g(z)}(\lambda) \subset N_{\partial g(z)}(\bar\lambda) \tag{3.10}$$

holds. To justify it, it is not hard to see $\partial g(z) = \partial g(\bar z) \cap D$ with $D := \{v \in \mathbb{R}^m |\, \langle v, z - \bar z\rangle = g(z) - g(\bar z)\}$. It follows from $\lambda \in \partial g(z) \subset D$ and $\lambda - \bar\lambda \in N_{K_g(\bar z, \bar\lambda)}(z - \bar z)$ that $\bar\lambda \in D$. This, together with the definition of D, leads us to $N_D(\lambda) = N_D(\bar\lambda) = [z - \bar z]$. Since both $\partial g(\bar z)$ and D are polyhedral, we conclude from the intersection rule for normal cones, holding without any constraint qualification for polyhedral convex sets, that

$$\begin{aligned} N_{\partial g(z)}(\lambda) &= N_{\partial g(\bar z) \cap D}(\lambda) = N_{\partial g(\bar z)}(\lambda) + N_D(\lambda) \\ &\subset N_{\partial g(\bar z)}(\bar\lambda) + N_D(\bar\lambda) = N_{\partial g(\bar z) \cap D}(\bar\lambda) = N_{\partial g(z)}(\bar\lambda), \end{aligned}$$

proving our claimed inclusion. Since g^* is a polyhedral function as well, using Lemma 2.1 for g^* at $(\bar\lambda, \bar z) \in \operatorname{gph} \partial g^*$ and shrinking r, if necessary, ensure that $\bar z \in \partial g^*(\lambda)$. Moreover, it results from (3.10) that $N_{\partial g^*(\lambda)}(z) \subset N_{\partial g^*(\lambda)}(\bar z)$, which in combination with (2.14) leads us to

$$K_{g^*}(\lambda, z) = N_{\partial g^*(\lambda)}(z) = N_{\partial g^*(\lambda)}(\bar z) \cap [z - \bar z]^\perp = K_{g^*}(\lambda, \bar z) \cap [z - \bar z]^\perp.$$

Using this and (3.5) brings us to

$$K_g(z, \lambda) = K_{g^*}(\lambda, z)^* = \big(K_{g^*}(\lambda, \bar z) \cap [z - \bar z]^\perp\big)^* = K_g(\bar z, \lambda) + [z - \bar z]. \tag{3.11}$$

Similarly, by (2.14) and the definition of the normal cone to the polyhedral convex set $\partial g(\bar z)$, we obtain

$$K_g(\bar z, \lambda) = N_{\partial g(\bar z)}(\lambda) = N_{\partial g(\bar z)}(\bar\lambda) \cap [\lambda - \bar\lambda]^\perp = K_g(\bar z, \bar\lambda) \cap [\lambda - \bar\lambda]^\perp,$$

which, together with (3.11), proves (3.9).

After these preparations, we are in a position to justify the claimed descriptions of critical cones of g. Pick $(z, \lambda) \in \big(\operatorname{gph} \partial g\big) \cap \mathbb{B}_r(\bar z, \bar\lambda)$ and set $F_1 := K_g(\bar z, \bar\lambda) \cap [\lambda - \bar\lambda]^\perp$, which is clearly a face of $K_g(\bar z, \bar\lambda)$. Because $\lambda - \bar\lambda \in N_{K_g(\bar z, \bar\lambda)}(z - \bar z)$, resulting from (3.1), we conclude that $z - \bar z \in F_1$. Since the relative interiors of nonempty faces of F_1 form a partition of this set (cf. [28, Theorem 18.2]), we find a face of F_1, denoted by F_2, that $z - \bar z \in \operatorname{ri} F_2$. This tells us that $F_2 \subset F_1$. Moreover, F_2 is a face of $K_g(\bar z, \bar\lambda)$ as well. By (3.9), the inclusion $K_g(z, \lambda) \subset F_1 - F_2$

clearly holds. To get the opposite inclusion, pick $x_i \in F_i$ for $i = 1, 2$. It follows from $z - \bar{z} \in \operatorname{ri} F_2$ and [28, Theorem 6.4] that there is $t > 0$ such that $(1 + t)(z - \bar{z}) - tx_2 \in F_2 \subset F_1$, which yields

$$t(x_1 - x_2) = tx_1 + (1 + t)(z - \bar{z}) - tx_2 - (1 + t)(z - \bar{z}) \in F_1 + [z - \bar{z}], \tag{3.12}$$

which confirms via (3.9) that the inclusion $F_1 - F_2 \subset K_g(z, \lambda)$ holds. This shows that $K_g(z, \lambda) = F_1 - F_2$.

Assume now that F_1 and F_2 are faces of $K_g(\bar{z}, \bar{\lambda})$ with $F_2 \subset F_1$. Pick $r > 0$ from Theorem 3.1 and choose $u \in K_g(\bar{z}, \bar{\lambda})^*$ with $\|u\| < r/2$ such that $F_1 = K_g(\bar{z}, \bar{\lambda}) \cap [u]^\perp$. Since $u \in K_g(\bar{z}, \bar{\lambda})^* = N_{K_g(\bar{z},\bar{\lambda})}(0)$, it follows from Theorem 3.1 that $\bar{\lambda} + u \in \partial g(\bar{z})$. Pick now $w \in \operatorname{ri} F_2$ with $\|w\| < r/2$ and observe from $w \in F_1$ that $(w, u) \in \operatorname{gph} N_{K_g(\bar{z},\bar{\lambda})}$. By Theorem 3.1, we arrive at $(\bar{z} + w, \bar{\lambda} + u) \in \operatorname{gph} \partial g$. Since $w \in \operatorname{ri} F_2$, as in (3.12), we can show that $F_1 - F_2 = F_1 + [w]$. Employing this and (3.9) leads us to

$$F_1 - F_2 = F_1 + [w] = K_g(\bar{z}, \bar{\lambda}) \cap [u]^\perp + [w] = K_g(\bar{z} + w, \bar{\lambda} + u),$$

which completes the proof. □

We proceed with two other important applications of the established reduction lemma for polyhedral functions in calculating the proto-derivative and coderivative of subgradient mappings of this class of functions. To this end, consider a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$. According to [33, Definition 8.33], the *graphical derivative* of F at $\bar{x}$ for $\bar{y}$ with $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ is the set-valued mapping $DF(\bar{x}, \bar{y}) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ defined via the tangent cone to $\operatorname{gph} F$ at $(\bar{x}, \bar{y})$ by

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

or, equivalently, $\operatorname{gph} DF(\bar{x}, \bar{y}) = T_{\operatorname{gph} F}(\bar{x}, \bar{y})$. Using the definition of the tangent cone, we can present an alternative definition of $DF(\bar{x}, \bar{y})$ in terms of graphical limits as

$$\operatorname{gph} DF(\bar{x}, \bar{y}) = \limsup_{t \searrow 0} \frac{\operatorname{gph} F - (\bar{x}, \bar{y})}{t}. \tag{3.13}$$

The set-valued mapping F is said to be *proto-differentiable* at $\bar{x}$ for $\bar{y}$ if the outer graphical limit in (3.13) is actually a full limit. If F is proto-differentiable at $\bar{x}$ for $\bar{y}$, its graphical derivative $DF(\bar{x}, \bar{y})$ is called the *proto-derivative* of F at $\bar{x}$ for $\bar{y}$. When $F(\bar{x})$ is a singleton consisting of $\bar{y}$ only, the notation $DF(\bar{x}, \bar{y})$ is simplified to $DF(\bar{x})$. It is easy to see that for a single-valued function F, which is differentiable at $\bar{x}$, the graphical derivative $DF(\bar{x})$ boils down to the Jacobian matrix of F at $\bar{x}$, denoted by $\nabla F(\bar{x})$. Recall from [33, Definition 9.53] that the strict graphical derivative of a set-valued mapping F at $\bar{x}$ for $\bar{y}$ with $(\bar{x}, \bar{y}) \in \operatorname{gph} F$, is the set-valued mapping $D_*F(\bar{x}, \bar{y}) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, defined by

$$\operatorname{gph} D_*F(\bar{x}, \bar{y}) = \limsup_{\substack{t \searrow 0 \\ (x,y) \xrightarrow{\operatorname{gph} F} (\bar{x},\bar{y})}} \frac{\operatorname{gph} F - (x, y)}{t}. \tag{3.14}$$

The set-valued mapping F is said to be *strictly* proto-differentiable at $\bar{x}$ for $\bar{y}$ if the outer graphical limit in (3.14) is a full limit. When F is strictly proto-differentiable at $\bar{x}$ for $\bar{y}$, its strict graphical derivative $D_*F(\bar{x}, \bar{y})$ is called the *strict* proto-derivative of F at $\bar{x}$ for $\bar{y}$.

Remark 3.5 (comparison of proto-derivative and strict proto-derivative)**.** It is important to remind our readers that if a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is strictly proto-differentiable at x for y with $(x, y) \in \operatorname{gph} F$, then it is proto-differentiable at x for y and its proto-derivative and strict proto-derivative coincide, namely we have

$$D_* F(x, y) = DF(x, y). \tag{3.15}$$

The latter claim results directly from the definitions of the outer and inner limits of a sequence of sets, which subsequently bring us to the inclusions

$$\liminf_{\substack{t \searrow 0 \\ (x',y') \xrightarrow{\operatorname{gph} F} (x,y)}} \frac{\operatorname{gph} F - (x', y')}{t} \subset \liminf_{t \searrow 0} \frac{\operatorname{gph} F - (x, y)}{t},$$

$$\limsup_{t \searrow 0} \frac{\operatorname{gph} F - (x, y)}{t} \subset \limsup_{\substack{t \searrow 0 \\ (x',y') \xrightarrow{\operatorname{gph} F} (x,y)}} \frac{\operatorname{gph} F - (x', y')}{t}.$$

These, combined with the definitions of proto-differentiability and strict proto-differentiability, justifies (3.15).

As Remark 3.5 demonstrates, strict proto-differentiability of a set-valued mapping has far-reaching consequences. In fact, it allows us to evaluate the strict proto-derivative of a set-valued mapping, which can be utilized to characterize its strong metric regularity; see [2, Theorem 4D.1] and Section 5 for more details on this application. The intriguing question is whether strict proto-differentiability holds for any class of set-valued mappings, important for constrained and composite optimization problems. In the next theorem, we are going to show that subgradient mappings of polyhedral functions enjoy this property when a relative interior condition is satisfied for subgradients under consideration. Furthermore, we will prove that such a relative interior condition, indeed, characterizes strict proto-differentiability of these set-valued mappings. While it is not easy to study strict proto-differentiability of set-valued mappings, its weaker version, namely proto-differentiability, has been well understood for many important classes of functions that are important for various applications; see [15,16,31,33] for more details and examples. In particular, it is well-known (cf. [33, Corollary 13.41]) that subgradient mappings of polyhedral functions are always proto-differentiable. Below, we also give another proof of this result via the reduction lemma, established in Theorem 3.1.

In what follows, we say that a sequence of set-valued mappings $F^k : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, $k \in \mathbb{N}$, *graph*-converges to $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ if the sequence $\{\operatorname{gph} F^k\}_{k \in \mathbb{N}}$ is convergent to $\operatorname{gph} F$ in the sense of Painlevé-Kuratowski.

Theorem 3.6. *For a polyhedral function $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ and $(\bar{z}, \bar{\lambda}) \in \operatorname{gph} \partial g$, the following properties hold.*

(a) *The tangent cone to $\operatorname{gph} \partial g$ at $(\bar{z}, \bar{\lambda})$ can be obtained by*

$$T_{\operatorname{gph} \partial g}(\bar{z}, \bar{\lambda}) = \operatorname{gph} N_{K_g(\bar{z}, \bar{\lambda})}. \tag{3.16}$$

Consequently, there exists $\varepsilon > 0$ such that the equality

$$T_{\operatorname{gph} \partial g}(\bar{z}, \bar{\lambda}) \cap \mathbb{B}_\varepsilon(0, 0) = \big(\operatorname{gph} \partial g - (\bar{z}, \bar{\lambda})\big) \cap \mathbb{B}_\varepsilon(0, 0) \tag{3.17}$$

holds.

(b) *∂g is proto-differentiable at $\bar{z}$ for $\bar{\lambda}$.*

(c) *If, in addition, $\bar{\lambda} \in \operatorname{ri} \partial g(\bar{z})$, then there is a neighborhood O of $(\bar{z}, \bar{\lambda})$ such that for any $(z, \lambda) \in O \cap \operatorname{gph} \partial g$, ∂g is strictly proto-differentiable at z for λ and its proto-derivative and strict proto-derivative coincide, namely*

$$D_*(\partial g)(z, \lambda) = D(\partial g)(z, \lambda). \tag{3.18}$$

Proof. Using the equality (3.1) and the definition of the tangent cone to $\operatorname{gph}\partial g$ at $(\bar z,\bar\lambda)$, we immediately arrive at (3.16). The claimed identity (3.17) results from (3.1) together with (3.16).

To justify (b), it suffices to prove the inclusion

$$T_{\operatorname{gph}\partial g}(\bar z,\bar\lambda)=\operatorname{gph} D(\partial g)(\bar z,\bar\lambda)\subset\liminf_{t\searrow 0}\frac{\operatorname{gph}\partial g-(\bar z,\bar\lambda)}{t}.$$

To justify it, take $(w,u)\in T_{\operatorname{gph}\partial g}(\bar z,\bar\lambda)$ and assume $t_k\searrow 0$. We can assume without loss of generality that $t_k\|(w,u)\|\le\varepsilon$, where ε is taken from part (a). Appealing now to (3.17) tells us that

$$(w,u)\in\frac{\operatorname{gph}\partial g-(\bar z,\bar\lambda)}{t_k}\quad\text{for all}\ \ k\in\mathbb{N}.$$

This, combined with the definition of the inner limit set, proves that

$$(w,u)\in\liminf_{t\searrow 0}\frac{\operatorname{gph}\partial g-(\bar z,\bar\lambda)}{t},$$

and thus completes the proof of (b). Finally, we proceed with the proof of (c). By Theorem 3.3(b), we find $r>0$ such that for any $(z,\lambda)\in\mathbb{B}_r(\bar z,\bar\lambda)\cap\operatorname{gph}\partial g$, the property $K_g(z,\lambda)=K_g(\bar z,\bar\lambda)$ holds. Take any such a pair (z,λ) and conclude from part (b) that ∂g is proto-differentiable at z for λ and from part (a) that $\operatorname{gph} D(\partial g)(z,\lambda)=\operatorname{gph} N_{K_g(z,\lambda)}=\operatorname{gph} N_{K_g(\bar z,\bar\lambda)}$. This clearly tells us that $D(\partial g)(z,\lambda)$ graph-converges to $N_{K_g(\bar z,\bar\lambda)}$ as $(z,\lambda)\to(\bar z,\bar\lambda)$ with $(z,\lambda)\in\operatorname{gph}\partial g$. Appealing now to [23, Corollary 4.3] (see also Proposition 4.2) and the fact that g is convex ensures that the latter conclusion amounts to strict proto-differentiability of g at $\bar z$ for $\bar\lambda$. To achieve a similar conclusion for any pair $(z,\lambda)\in\operatorname{gph}\partial g$ sufficiently close to $(\bar z,\bar\lambda)$, observe from Theorem 3.3(b) that for any such a pair, we have $\lambda\in\operatorname{ri}\partial g(z)$. A similar argument as the one, presented above for $(\bar z,\bar\lambda)$, shows that g is strictly proto-differentiable at z for λ whenever $(z,\lambda)\in\operatorname{gph}\partial g$ is sufficiently close to $(\bar z,\bar\lambda)$. Finally, (3.18) results from the discussion in Remark 3.5, which finishes the proof of (c). □

Note that the observation in Theorem 3.6(c) has several major consequences in sensitivity analysis of generalized equations, which will be pursued in the final section of this paper. It is tempting to ask whether strict proto-differentiability can be justified for subgradient mappings of polyhedral functions without the relative interior assumption in Theorem 3.6(c). We will demonstrate in the next section that this relative interior condition is, indeed, equivalent to strict proto-differentiability of ∂g, which implies that strict proto-differentiability for subgradient mappings of polyhedral functions does not hold in the absence of the relative interior condition.

We continue with one more direct application of the reduction lemma for polyhedral functions in finding the regular and limiting normal cones to $\operatorname{gph}\partial g$.

Theorem 3.7 (regularity of subgradient mappings of polyhedral functions)**.** *For a polyhedral function $g:\mathbb{R}^m\to\overline{\mathbb{R}}$ and $(\bar z,\bar\lambda)\in\operatorname{gph}\partial g$, the following properties hold.*

(a) *The regular normal cone to $\operatorname{gph}\partial g$ at $(\bar z,\bar\lambda)$ can be calculated by*

$$\widehat N_{\operatorname{gph}\partial g}(\bar z,\bar\lambda)=K_g(\bar z,\bar\lambda)^*\times K_g(\bar z,\bar\lambda).\tag{3.19}$$

(b) *The (limiting) normal cone to $\operatorname{gph}\partial g$ at $(\bar z,\bar\lambda)$ can be calculated by*

$$N_{\operatorname{gph}\partial g}(\bar z,\bar\lambda)=\bigcup_{\substack{F_1,F_2\subset\mathcal{F}_{K_g(\bar z,\bar\lambda)}\\ F_2\subset F_1}}(F_1-F_2)^*\times(F_1-F_2),\tag{3.20}$$

where $\mathcal{F}_{K_g(\bar z,\bar\lambda)}$ is the collection of all faces of $K_g(\bar z,\bar\lambda)$.

(c) *The condition* $\bar\lambda \in \operatorname{ri} \partial g(\bar z)$ *is equivalent to*

$$N_{\operatorname{gph} \partial g}(\bar z, \bar\lambda) = \widehat{N}_{\operatorname{gph} \partial g}(\bar z, \bar\lambda). \tag{3.21}$$

Proof. We begin with the proof of (a). Using the polar representation of the regular normal cone via the tangent cone and the relation in (3.16), we can rewrite (3.19) equivalently as

$$\big(\operatorname{gph} N_{K_g(\bar z,\bar\lambda)}\big)^* = \big(T_{\operatorname{gph} \partial g}(\bar z, \bar\lambda)\big)^* = K_g(\bar z, \bar\lambda)^* \times K_g(\bar z, \bar\lambda). \tag{3.22}$$

To prove it, assume that (w, u) is taken from the left-hand side of (3.22). We next show that $w \in K_g(\bar z, \bar\lambda)^*$ and $u \in K_g(\bar z, \bar\lambda)$. To this end, take any $\tilde w \in K_g(\bar z, \bar\lambda)$ and $\tilde u \in K_g(\bar z, \bar\lambda)^*$ and observe that $(\tilde w, 0) \in \operatorname{gph} N_{K_g(\bar z,\bar\lambda)}$ and $(0, \tilde u) \in \operatorname{gph} N_{K_g(\bar z,\bar\lambda)}$, since $K_g(\bar z, \bar\lambda)^* = N_{K_g(\bar z,\bar\lambda)}(0)$. The latter conditions, together with $(w, u) \in \big(\operatorname{gph} N_{K_g(\bar z,\bar\lambda)}\big)^*$, give us

$$\langle w, \tilde w\rangle = \langle (w, u), (\tilde w, 0)\rangle \le 0, \quad \text{and} \quad \langle u, \tilde u\rangle = \langle (w, u), (0, \tilde u)\rangle \le 0,$$

which in turn justify that $(w, u) \in K_g(\bar z, \bar\lambda)^* \times K_g(\bar z, \bar\lambda)$.

Pick now arbitrary pairs (w, u) from the right-hand side of (3.22) and $(\tilde w, \tilde u)$ from $\operatorname{gph} N_{K_g(\bar z,\bar\lambda)}$. We then obtain in particular that $\tilde w \in K_g(\bar z, \bar\lambda)$ and $\tilde u \in K_g(\bar z, \bar\lambda)^*$. Combining these yields

$$\langle (w, u), (\tilde w, \tilde u)\rangle = \langle w, \tilde w\rangle + \langle u, \tilde u\rangle \le 0,$$

which verifies that $(w, u) \in \big(\operatorname{gph} N_{K_g(\bar z,\bar\lambda)}\big)^*$, and thus completes the proof of (3.19).

To justify (b), observe that $(u, w) \in N_{\operatorname{gph} \partial g}(\bar z, \bar\lambda)$ if and only if there are sequences $(z^k, \lambda^k) \to (\bar z, \bar\lambda)$ with $(z^k, \lambda^k) \in \operatorname{gph} \partial g$ and $(u^k, w^k) \to (u, w)$ with $(u^k, w^k) \in \widehat{N}_{\operatorname{gph} \partial g}(z^k, \lambda^k)$. Using the definition of the regular normal cone and (3.1), we conclude for any k sufficiently large that

$$\begin{aligned}(u^k, w^k) \in \widehat{N}_{\operatorname{gph} \partial g}(z^k, \lambda^k) \quad &\Longleftrightarrow \quad (u^k, w^k) \in \widehat{N}_{\operatorname{gph} \partial g - (\bar z, \bar\lambda)}(z^k - \bar z, \lambda^k - \bar\lambda)\\ &\Longleftrightarrow \quad (u^k, w^k) \in \widehat{N}_{\operatorname{gph} N_{K_g(\bar z,\bar\lambda)}}(z^k - \bar z, \lambda^k - \bar\lambda).\end{aligned}$$

The last inclusion yields $(u, w) \in N_{\operatorname{gph} N_{K_g(\bar z,\bar\lambda)}}(0, 0)$ and thus justifies the inclusion $N_{\operatorname{gph} \partial g}(\bar z, \bar\lambda) \subset N_{\operatorname{gph} N_{K_g(\bar z,\bar\lambda)}}(0, 0)$. A similar argument via (3.1) can be used to justify the opposite inclusion and obtain

$$N_{\operatorname{gph} \partial g}(\bar z, \bar\lambda) = N_{\operatorname{gph} N_{K_g(\bar z,\bar\lambda)}}(0, 0).$$

Recall that $K_g(\bar z, \bar\lambda)$ is a polyhedral convex cone. Employing the representation of the limiting normal cone to the normal cone to the graph of a polyhedral convex set, obtained in the proof of [3, Theorem 2], we arrive at the representation on the right-hand side of (3.20) for $N_{\operatorname{gph} N_{K_g(\bar z,\bar\lambda)}}(0, 0)$, which completes the proof of (b).

To prove (c), observe that it follows from $\bar\lambda \in \operatorname{ri} \partial g(\bar z)$ and Theorem 3.3(b) that there exists a neighborhood O of $(\bar z, \bar\lambda)$ such that for any $(z, \lambda) \in O \cap \operatorname{gph} \partial g$, we have $K_g(z, \lambda) = K_g(\bar z, \bar\lambda)$. This and (3.19) bring us to

$$\widehat{N}_{\operatorname{gph} \partial g}(z, \lambda) = \widehat{N}_{\operatorname{gph} \partial g}(\bar z, \bar\lambda) \quad \text{for all } \ (z, \lambda) \in O \cap \operatorname{gph} \partial g, \tag{3.23}$$

which, together with the definition of the limiting normal cone, justifies the claimed equality in (3.21). Conversely, suppose that (3.21) holds. We are going to show that $K_g(\bar z, \bar\lambda)$ is a linear subspace. To this end, it suffices to show that if $w \in K_g(\bar z, \bar\lambda)$, then $-w \in K_g(\bar z, \bar\lambda)$. Pick $w \in K_g(\bar z, \bar\lambda)$ and conclude from (3.16) that there is $u \in \mathbb{R}^m$ such that $(w, u) \in T_{\operatorname{gph} \partial g}(\bar z, \bar\lambda)$. Appealing now to [33, Theorem 13.57] implies that $(u, -w) \in N_{\operatorname{gph} \partial g}(\bar z, \bar\lambda)$. This, together with (3.21) and (3.19), tells us that $-w \in K_g(\bar z, \bar\lambda)$. Remember that $K_g(\bar z, \bar\lambda) = N_{\partial g(\bar z)}(\bar\lambda)$. Since $K_g(\bar z, \bar\lambda)$ is a linear subspace, we arrive at $\bar\lambda \in \operatorname{ri} \partial g(\bar z)$, which completes the proof of (c). $\square$

The description (3.19) of the regular normal cone to $\operatorname{gph} \partial g$ in terms of the critical cone of g was established in [18, Theorem 4.3(i)] using a different approach. Our current proof relies upon the reduction lemma, which allows us to simplify the proof of this result. We should mention that a similar result was established for polyhedral convex sets using Robinson's reduction lemma for polyhedral convex sets in the proof of [3, Theorem 2]. A similar expression of the limiting normal cone to $\operatorname{gph} \partial g$ was achieved in [18, Theorem 5.1] via a lengthy direct argument. Our proof, which heavily uses Theorem 3.1, reduces the calculation to the case of a polyhedral convex cone and then utilizes the available result for this setting. Thus, Theorem 3.7 can be considered as a generalization of Dontchev and Rockafellar's result, obtained in [3, Theorem 2], for polyhedral convex sets. Note that part (c) of Theorem 3.7 offers a new piece of information about $\operatorname{gph} \partial g$, which has not been observed before to the best of our knowledge. Indeed, it tells us that $\operatorname{gph} \partial g$ is *regular* at $(\bar{z}, \bar{\lambda}) \in \operatorname{gph} \partial g$ in the sense of [33, Definition 6.4] if and only if the subgradient $\bar{\lambda}$ is taken from the relative interior of $\partial g(\bar{z})$.

Recall that for a polyhedral function g with $(\bar{z}, \bar{\lambda}) \in \operatorname{gph} \partial g$, the coderivative mapping of ∂g at $\bar{z}$ for $\bar{\lambda}$, denoted $D^*(\partial g)(\bar{z}, \bar{\lambda})$, is defined by

$$u \in D^*(\partial g)(\bar{z}, \bar{\lambda})(w) \iff (u, -w) \in N_{\operatorname{gph} \partial g}(\bar{z}, \bar{\lambda}).$$

It is well-known (cf. [33, Theorem 13.57]) that the inclusion $D(\partial g)(\bar{z}, \bar{\lambda})(w) \subset D^*(\partial g)(\bar{z}, \bar{\lambda})(w)$, for any $w \in \mathbb{R}^m$, always holds for a polyhedral function g. Below, we show that this inclusion becomes equality provided that the subgradient $\bar{\lambda}$ is taken from the relative interior of $\partial g(\bar{z})$.

Corollary 3.8. *Assume that $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a polyhedral function and that $(\bar{z}, \bar{\lambda}) \in \operatorname{gph} \partial g$. Then $\bar{\lambda} \in \operatorname{ri} \partial g(\bar{z})$ if and only if*

$$D(\partial g)(\bar{z}, \bar{\lambda})(w) = D^*(\partial g)(\bar{z}, \bar{\lambda})(w) \quad \textit{for all } w \in \mathbb{R}^m. \tag{3.24}$$

Proof. Suppose that $\bar{\lambda} \in \operatorname{ri} \partial g(\bar{z})$. Thus the critical cone $K_g(\bar{z}, \bar{\lambda})$ is a linear subspace. By Theorem 3.6(a), we have $T_{\operatorname{gph} \partial g}(\bar{z}, \bar{\lambda}) = \operatorname{gph} N_{K_g(\bar{z}, \bar{\lambda})} = K_g(\bar{z}, \bar{\lambda}) \times K_g(\bar{z}, \bar{\lambda})^{\perp}$. Moreover, we conclude from Theorem 3.7(c) that $N_{\operatorname{gph} \partial g}(\bar{z}, \bar{\lambda}) = K_g(\bar{z}, \bar{\lambda})^{\perp} \times K_g(\bar{z}, \bar{\lambda})$. Combining these with the definitions of coderivative and proto-derivative justifies (3.24).

Assume now that (3.24) holds. We are going to show that $K_g(\bar{z}, \bar{\lambda})$ is a linear subspace. To this end, it suffices to show that if $w \in K_g(\bar{z}, \bar{\lambda})$, then $-w \in K_g(\bar{z}, \bar{\lambda})$. Pick $w \in K_g(\bar{z}, \bar{\lambda})$ and conclude from (3.16) that there is $u \in \mathbb{R}^m$ such that $u \in D(\partial g)(\bar{z}, \bar{\lambda})(w) = N_{K_g(\bar{z}, \bar{\lambda})}(w)$, which yields that $(w, u) \in K_g(\bar{z}, \bar{\lambda}) \times K_g(\bar{z}, \bar{\lambda})^*$. This, combined with (3.20), tells us that $(u, w) \in K_g(\bar{z}, \bar{\lambda})^* \times K_g(\bar{z}, \bar{\lambda}) \subset N_{\operatorname{gph} \partial g}(\bar{z}, \bar{\lambda})$. Employing the definition of coderivative and (3.24), we obtain $u \in D^*(\partial g)(\bar{z}, \bar{\lambda})(-w) = D(\partial g)(\bar{z}, \bar{\lambda})(-w) = N_{K_g(\bar{z}, \bar{\lambda})}(-w)$, which clearly confirms that $-w \in K_g(\bar{z}, \bar{\lambda})$. Since $K_g(\bar{z}, \bar{\lambda}) = N_{\partial g(\bar{z})}(\bar{\lambda})$ and $K_g(\bar{z}, \bar{\lambda})$ is a linear subspace, we obtain $\bar{\lambda} \in \operatorname{ri} \partial g(\bar{z})$, which completes the proof. $\square$

We close this section with an application of (3.20) to calculate $D^*(\partial g)(\bar{z}, \bar{\lambda})(0)$ for a polyhedral function g with $(\bar{z}, \bar{\lambda}) \in \operatorname{gph} \partial g$, important for stability analysis of composite optimization problems. By definition, we know that $u \in D^*(\partial g)(\bar{z}, \bar{\lambda})(0)$ if and only if $(u, 0) \in N_{\operatorname{gph} \partial g}(\bar{z}, \bar{\lambda})$. Now, we claim that

$$(u, 0) \in N_{\operatorname{gph} \partial g}(\bar{z}, \bar{\lambda}) \iff u \in \big(K_g(\bar{z}, \bar{\lambda}) \cap -K_g(\bar{z}, \bar{\lambda})\big)^{\perp}. \tag{3.25}$$

To justify this claim, pick $u \in \big(K_g(\bar{z}, \bar{\lambda}) \cap -K_g(\bar{z}, \bar{\lambda})\big)^{\perp}$. Observe that $K_g(\bar{z}, \bar{\lambda}) \cap -K_g(\bar{z}, \bar{\lambda})$ is a closed face of $K_g(\bar{z}, \bar{\lambda})$. Letting $F_1 = F_2 = K_g(\bar{z}, \bar{\lambda}) \cap -K_g(\bar{z}, \bar{\lambda})$ in (3.20) shows that $(u, 0) \in N_{\operatorname{gph} \partial g}(\bar{z}, \bar{\lambda})$. Conversely, suppose $(u, 0) \in N_{\operatorname{gph} \partial g}(\bar{z}, \bar{\lambda})$. By (3.20), we find closed faces

F_1 and F_2 of $K_g(\bar z,\bar\lambda)$ with $F_2\subset F_1$ that $u\in(F_1-F_2)^*$. Since $K_g(\bar z,\bar\lambda)\cap -K_g(\bar z,\bar\lambda)$ is the smallest closed face of $K_g(\bar z,\bar\lambda)$, we get $K_g(\bar z,\bar\lambda)\cap -K_g(\bar z,\bar\lambda)\subset F_1$ and $K_g(\bar z,\bar\lambda)\cap -K_g(\bar z,\bar\lambda)\subset F_2$, which bring us to

$$u\in(F_1-F_2)^*\subset\big(K_g(\bar z,\bar\lambda)\cap -K_g(\bar z,\bar\lambda)\big)^*=\big(K_g(\bar z,\bar\lambda)\cap -K_g(\bar z,\bar\lambda)\big)^\perp$$

and thus finish the proof of (3.25). If we assume further that g has the representation (2.3), it is not hard to see that (3.25) amounts to

$$u\in\operatorname{span}\big\{a^i-a^j\,|\,i,j\in J(\bar z)\big\}+\operatorname{span}\big\{b^i\,|\,i\in I(\bar z)\big\}. \tag{3.26}$$

Indeed, employing the description of $K_g(\bar z,\bar\lambda)$ from Proposition 2.3, we obtain

$$\big(K_g(\bar z,\bar\lambda)\cap -K_g(\bar z,\bar\lambda)\big)^\perp=\operatorname{span}\big\{a^i-a^j\,|\,i\in J(\bar z),\,j\in J_+(\bar z,\bar\lambda_1)\big\}+\operatorname{span}\big\{b^i\,|\,i\in I(\bar z)\big\}, \tag{3.27}$$

for any decomposition $\bar\lambda=\bar\lambda_1+\bar\lambda_2$ as in (2.6). Pick an arbitrary $j_0\in J_+(\bar z,\bar\lambda_1)$ and note that for any $i,j\in J(\bar z)$ we have $a^i-a^j=(a^i-a^{j_0})-(a^j-a^{j_0})$. Thus, we can replace $j\in J_+(\bar z,\bar\lambda_1)$ in (3.27) with $j\in J(\bar z)$, which proves (3.26).

4 Strict Twice Epi-Differentiability of Polyhedral Functions

In this section, we study another important second-order variational property, called strict twice epi-differentiability, for polyhedral functions and its applications in continuous differentiability of proximal mappings for this class of functions. Once again, our results rely heavily on our extension of the reduction lemma for polyhedral functions. To achieve our goal, consider a function $f:\mathbb{R}^n\to\overline{\mathbb{R}}$ with $\bar x\in\mathbb{R}^n$ with $f(\bar x)$ finite and define the parametric family of second-order difference quotients of f at $\bar x$ for $\bar v\in\partial f(\bar x)$ by

$$\Delta_t^2 f(\bar x,\bar v)(w)=\frac{f(\bar x+tw)-f(\bar x)-t\langle\bar v,\,w\rangle}{\frac12 t^2}$$

for any $w\in\mathbb{R}^n$ and $t>0$. The second subderivative of f at $\bar x$ for $\bar v$, denoted $\mathrm{d}^2f(\bar x,\bar v)$, is an extended-real-valued function defined by

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

Following [33, 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 $\mathrm{d}^2f(\bar x,\bar v)$ 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 a function as $t\searrow 0$, $(x,v)\to(\bar x,\bar v)$ with $f(x)\to f(\bar x)$ and $(x,v)\in\operatorname{gph}\partial f$. If this condition holds, the limit function is then the second subderivative $\mathrm{d}^2f(\bar x,\bar v)$. Twice epi-differentiability of extended-real-valued functions, introduced by Rockafellar in [29], has been investigated for important classes of functions appearing in constrained and composite optimization problems in [15, 16, 30]. Its strict version, introduced in [22], was only studied in [21] for nonlinear programming and minimax problems and so it is tempting to ask when this property holds and to explore its applications in parametric optimization. To this end, we define the *strict second subderivative* of f at $\bar x$ for $\bar v$ with $\bar v\in\partial f(\bar x)$ at $w\in\mathbb{R}^n$ by

$$\mathrm{d}_s^2f(\bar x,\bar v)(w)=\liminf_{\substack{t\searrow 0,\,w'\to w\\ (x,v)\xrightarrow{\operatorname{gph}\partial f}(\bar x,\bar v)\\ f(x)\to f(\bar x)}}\Delta_t^2 f(x,v)(w').$$

When f is subdifferentially continuous at $\bar{x}$ for $\bar{v}$ in the sense of [33, Definition 13.28], we can drop the requirement $f(x) \to f(\bar{x})$ in the definition of the strict second subderivative. According to [33, Example 13.30], convex functions are always subdifferentialy continuous. Clearly, we always have $\mathrm{d}_s^2 f(\bar{x}, \bar{v})(w) \le \mathrm{d}^2 f(\bar{x}, \bar{v})(w)$ for any $w \in \mathbb{R}^n$. If f is strictly twice epi-differentiable at $\bar{x}$ for $\bar{v}$, the latter inequality becomes equality. Being able to calculate the strict second subderivative of a function and comparing with its second subderivative can tell us when strict twice epi-differentiability should be expected for such a function. Note also that the second subderivative was exploited to characterize the quadratic growth condition for extended-real-valued functions in [33, Theorem 13.24(c)]. Similarly, we can use the strict second subderivative to achieve a characterization of the uniform quadratic growth condition (cf. [1, Definition 5.16]), which plays an important role in parametric optimization. This is beyond the scope of this paper and so we postpone it to our future research. Below, we use the characterization of the faces of the critical cone of a polyhedral function in Proposition 3.4 to find its strict second subderivative.

Proposition 4.1. *For a polyhedral function $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ and $(\bar{z}, \bar{\lambda}) \in \operatorname{gph} \partial g$, its strict second subderivative can be calculated by*

$$\mathrm{d}_s^2 g(\bar{z}, \bar{\lambda})(w) = \delta_{K_g(\bar{z},\bar{\lambda}) - K_g(\bar{z},\bar{\lambda})}(w) \quad \text{for all } w \in \mathbb{R}^m. \tag{4.1}$$

Proof. We first prove that $\mathrm{d}_s^2 g(\bar{z}, \bar{\lambda})(w) = 0$ for all $w \in K_g(\bar{z}, \bar{\lambda}) - K_g(\bar{z}, \bar{\lambda})$. First, observe that the convexity of g yields $\mathrm{d}_s^2 g(\bar{z}, \bar{\lambda})(w) \ge 0$ for all $w \in \mathbb{R}^m$. To obtain the opposite inequality, pick $w \in K_g(\bar{z}, \bar{\lambda}) - K_g(\bar{z}, \bar{\lambda})$. By Proposition 3.4, we find a sequence $(z^k, \lambda^k) \to (\bar{z}, \bar{\lambda})$ such that $(z^k, \lambda^k) \in \operatorname{gph} \partial g$ and $K_g(z^k, \lambda^k) = K_g(\bar{z}, \bar{\lambda}) - K_g(\bar{z}, \bar{\lambda})$. This implies that $w \in K_g(z^k, \lambda^k)$, which is equivalent to saying that $\mathrm{d}g(z^k)(w) = \langle \lambda^k, w \rangle$. Employing Proposition 2.2(b), we find a sequence $t_k \searrow 0$ such that $g(z^k + t_k w) - g(z^k) = \mathrm{d}g(z^k)(t_k w) = t_k \mathrm{d}g(z^k)(w)$. By definition, we get $\Delta_{t_k}^2 g(z^k, \lambda^k)(w) = 0$, which results in

$$\mathrm{d}_s^2 g(\bar{z}, \bar{\lambda})(w) \le \lim_{k \to \infty} \Delta_{t_k}^2 g(z^k, \lambda^k)(w) = 0.$$

This confirms that $\mathrm{d}_s^2 g(\bar{z}, \bar{\lambda})(w) = 0$ for any $w \in K_g(\bar{z}, \bar{\lambda}) - K_g(\bar{z}, \bar{\lambda})$. To obtain (4.1), it suffices to to justify that $\operatorname{dom} \mathrm{d}_s^2 g(\bar{z}, \bar{\lambda}) \subset K_g(\bar{z}, \bar{\lambda}) - K_g(\bar{z}, \bar{\lambda})$. Taking any $w \in \mathbb{R}^m$ with $\mathrm{d}_s^2 g(\bar{z}, \bar{\lambda})(w) < \infty$, we can find sequences $t_k \searrow 0$, $w^k \to w$, and $(z^k, \lambda^k) \to (\bar{z}, \bar{\lambda})$ with $(z^k, \lambda^k) \in \operatorname{gph} \partial g$ such that

$$\lim_{k \to \infty} \frac{g(z^k + t_k w^k) - g(z^k) - t_k \langle \lambda^k, w^k \rangle}{\frac{1}{2} t_k^2} < \infty.$$

This, together with the convexity of g, allows us to find a constant $M > 0$ such that

$$0 \le \frac{g(z^k + t_k w^k) - g(z^k) - t_k \langle \lambda^k, w^k \rangle}{t_k} \le M t_k$$

for all k sufficiently large and therefore to obtain

$$\lim_{k \to \infty} \frac{g(z^k + t_k w^k) - g(z^k) - t_k \langle \lambda^k, w^k \rangle}{t_k} = 0. \tag{4.2}$$

Take an arbitrary

$$u \in \big(K_g(\bar{z}, \bar{\lambda}) - K_g(\bar{z}, \bar{\lambda})\big)^{\perp} = K_g(\bar{z}, \bar{\lambda})^* \cap -K_g(\bar{z}, \bar{\lambda})^* = T_{\partial g(\bar{z})}(\bar{\lambda}) \cap -T_{\partial g(\bar{z})}(\bar{\lambda}),$$

where the last equality results from (2.14). We now show that $\langle u, w\rangle = 0$, which in turn yields $w \in K_g(\bar z, \bar\lambda) - K_g(\bar z, \bar\lambda)$. To this end, by (2.12), we find $\alpha > 0$ such that $\bar\lambda \pm \alpha u \in \partial g(\bar z)$. This clearly implies that

$$\langle \bar\lambda \pm \alpha u, z^k + t_k w^k - \bar z\rangle \le g(z^k + t_k w^k) - g(\bar z). \tag{4.3}$$

Employing Theorem 3.1, we can conclude that $(z^k - \bar z, \lambda^k - \bar\lambda) \in \operatorname{gph} N_{K_g(\bar z,\bar\lambda)}$ for all k sufficiently large. This tells us that $z^k - \bar z \in K_g(\bar z, \bar\lambda) = N_{\partial g(\bar z)}(\bar\lambda)$ for all k sufficiently large. Moerover, Lemma 2.1 confirms that $\bar\lambda \in \partial g(z^k)$ for all k sufficiently large. Combining these, we can conclude that

$$\langle u, z^k - \bar z\rangle = 0 \quad \text{and} \quad \langle \bar\lambda, z^k - \bar z\rangle = g(z^k) - g(\bar z).$$

These, together with (4.3), yield

$$\langle \bar\lambda \pm \alpha u, t_k w^k\rangle \le g(z^k + t_k w^k) - g(z^k),$$

and therefore we get

$$\langle \bar\lambda - \lambda^k \pm \alpha u, w^k\rangle \le \frac{g(z^k + t_k w^k) - g(z^k) - t_k\langle \lambda^k, w^k\rangle}{t_k}.$$

Passing to the limit as $k \to \infty$ and employing (4.2), we arrive at $\pm\alpha\langle u, w\rangle \le 0$, meaning $\langle u, w\rangle = 0$. This confirms that $w \in K_g(\bar z, \bar\lambda) - K_g(\bar z, \bar\lambda)$ and hence completes the proof. □

The established formula for the strict second subderivative of a polyhedral function in (4.1) suggests a path forward in the study of strict twice epi-differentiability of this class of functions. As pointed out earlier, strict twice epi-differentiability requires that the second subderivative and strict second subderivative coincide. For polyhedral functions, Proposition 4.1 immediately suggests that the given subgradient $\bar\lambda$ must belong to $\operatorname{ri} \partial g(\bar z)$; see the proof of the implication (a) $\Longrightarrow$ (c) in Theorem 4.3 for a detailed proof. One may wonder whether the opposite holds as well, namely the relative interior condition $\bar\lambda \in \operatorname{ri} \partial g(\bar z)$ implies strict twice epi-differentiability of polyhedral functions. Our next goal is to indeed demonstrate that this is true. In doing so, we rely heavily upon the reduction lemma, obtained in Theorem 3.1, as well as a characterization of strict twice epi-differentiablity of prox-regular functions from [23]. The next result is a simplified version of [23, Corollary 4.3], and presents a useful characterization of strict twice epi-differentiability of convex functions, which comprise an important subclass of prox-regular functions according to [33, Example 3.30].

Proposition 4.2 (characterization of strict twice epi-differentiability). *Assume that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $\bar x \in \mathbb{R}^n$ with $f(\bar x)$ finite, and $\bar v \in \partial f(\bar x)$ and that f is a convex function. Then there is a neighborhood O of $(\bar x, \bar v)$ such that for any $(x, v) \in O \cap \operatorname{gph} f$, the following properties are equivalent:*

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

The next result is an immediate consequence of Theorem 3.3(b) and reveals that polyhedral functions are always strictly twice epi-differentiable under a relative interior condition.

Theorem 4.3 (strict twice epi-differentiability of polyhedral functions). *Assume that $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a polyhedral function and that $(\bar z, \bar\lambda) \in \operatorname{gph} \partial g$. Then the following properties are equivalent:*

(a) *there is a neighborhood O of $(\bar{z}, \bar{\lambda})$ such that for any $(z, \lambda) \in O \cap \operatorname{gph} \partial g$, g is strictly twice epi-differentiable at z for λ;*

(b) *there is a neighborhood O of $(\bar{z}, \bar{\lambda})$ such that for any $(z, \lambda) \in O \cap \operatorname{gph} \partial g$, ∂g is strictly proto-differentiable at z for λ;*

(c) $\bar{\lambda} \in \operatorname{ri} \partial g(\bar{z})$.

Proof. Since g is convex, the equivalence of (a) and (b) falls out of Proposition 4.2. It follows from Theorem 3.6(c) that (c) yields (b). Using a similar argument presented in Theorem 3.6(c), one can show directly the implication (c) $\implies$ (a). Indeed, for any $(z, \lambda) \in \operatorname{gph} \partial g$, we infer from [33, Proposition 13.9] that g is twice epi-differentiable at z for λ and

$$\mathrm{d}^2 g(z, \lambda) = \delta_{K_g(z,\lambda)}. \tag{4.4}$$

It follows from Theorem 3.3(b) and $\bar{\lambda} \in \operatorname{ri} \partial g(\bar{z})$ that for any $(z, \lambda) \in \operatorname{gph} \partial g$ sufficiently close to $(\bar{z}, \bar{\lambda})$, the equality $K_g(z, \lambda) = K_g(\bar{z}, \bar{\lambda})$ holds, which together with (4.4) brings us to $\mathrm{d}^2 g(z, \lambda) = \delta_{K_g(\bar{z},\bar{\lambda})}$. Using [33, Proposition 7.4(f)], we conclude that $\mathrm{d}^2 g(z, \lambda)$ epi-coverges to $\mathrm{d}^2 g(\bar{z}, \bar{\lambda})$ as $(z, \lambda) \to (\bar{z}, \bar{\lambda})$ with $(z, \lambda) \in \operatorname{gph} \partial g$. This, combined with Proposition 4.2, demonstrates that g is strictly twice epi-differentiable at $\bar{z}$ for $\bar{\lambda}$. To achieve a similar conclusion for any pair $(z, \lambda) \in \operatorname{gph} \partial g$ sufficiently close to $(\bar{z}, \bar{\lambda})$, observe from Theorem 3.3(b) that for any such a pair, we have $\lambda \in \operatorname{ri} \partial g(z)$. A similar argument as the one, presented above for $(\bar{z}, \bar{\lambda})$, shows that g is strictly twice epi-differentiable at z for λ whenever $(z, \lambda) \in \operatorname{gph} \partial g$ is sufficiently close to $(\bar{z}, \bar{\lambda})$, which proves (a).

Finally, we show the implication (a) $\implies$ (c). Suppose that (a) holds. As pointed out earlier, when strict twice epidifferentiablity holds for a function, its second subderivative and strict second subderivative coincide. Strict twice epi-differentiability of g at $\bar{z}$ for $\bar{\lambda}$, together with (4.1) and (4.4), tells us that

$$\delta_{K_g(\bar{z},\bar{\lambda}) - K_g(\bar{z},\bar{\lambda})} = \mathrm{d}_s^2 g(\bar{z}, \bar{\lambda}) = \mathrm{d}^2 g(\bar{z}, \bar{\lambda}) = \delta_{K_g(\bar{z},\bar{\lambda})}.$$

This implies that $K_g(\bar{z}, \bar{\lambda}) - K_g(\bar{z}, \bar{\lambda}) = K_g(\bar{z}, \bar{\lambda})$, meaning that $K_g(\bar{z}, \bar{\lambda})$ is a linear subspace. By (2.14), we get $\bar{\lambda} \in \operatorname{ri} \partial g(\bar{z})$, which completes the proof. $\square$

Combining the obtained characterization of strict twice epi-differentiability of polyhedral functions with (3.6) allows us to conclude that strict twice epi-differentiablity is preserved under the Fenchel conjugate for polyhedral functions, as shown below.

Corollary 4.4. *Assume that $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a polyhedral function and that $(\bar{z}, \bar{\lambda}) \in \operatorname{gph} \partial g$. Then the following properties are equivalent:*

(a) *there is a neighborhood O of $(\bar{z}, \bar{\lambda})$ such that for any $(z, \lambda) \in O \cap \operatorname{gph} \partial g$, g is strictly twice epi-differentiable at z for λ;*

(b) *there is a neighborhood U of $(\bar{\lambda}, \bar{z})$ such that for any $(\lambda, z) \in U \cap \operatorname{gph} \partial g^*$, g^* is strictly twice epi-differentiable at λ for z;*

(c) $\bar{\lambda} \in \operatorname{ri} \partial g(\bar{z})$;

(d) $\bar{z} \in \operatorname{ri} \partial g^*(\bar{\lambda})$.

Proof. We know from [33, Theorem 11.14(a)] that g^* is a polyhedral function. We obtain the equivalence of (a) and (c) and of (b) and (d) from Theorem 4.3. Corollary 3.2 also tells us that (c) and (d) are equivalent, which completes the proof. $\square$

We close this section with a major consequence of the established characterization in Theorem 4.3 about smoothness of the Moreau envelope and proximal mapping for polyhedral functions. To this end, recall that for a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and parameter value $r > 0$, the Moreau envelope function $e_r f$ and the proximal mapping prox_{rf} are defined, respectively, by

$$e_r f(x) = \inf_{w \in \mathbb{R}^n} \Big\{ f(w) + \frac{1}{2r}\|w - x\|^2 \Big\},$$

and

$$\text{prox}_{rf}(x) = \underset{w \in \mathbb{R}^n}{\text{argmin}} \Big\{ f(w) + \frac{1}{2r}\|w - x\|^2 \Big\}.$$

When f is convex, the subdifferential sum rule from convex analysis implies that

$$\text{prox}_{rf}(x) = \big(I + r\partial f\big)^{-1}(x), \quad x \in \mathbb{R}^n, \tag{4.5}$$

where I stands for the $n \times n$ identity matrix. Furthermore, it is known that the envelope function $e_r f$ is continuously differentiable (cf. [33, Theorem 2.26]) for any convex function f. If, in addition, f is a polyhedral function, we deduce from [33, Proposition 13.9] and [33, Exercise 13.45] that the proximal mapping prox_{rf} is semidifferentiable. According to [2, Proposition 2D.1], the latter is equivalent to directional differentiability of prox_{rf}. Below, we present a simple but useful characterization of continuous differentiability of the proximal mapping of polyhedral functions.

Theorem 4.5. *Assume that $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a polyhedral function and that $(\bar{z}, \bar{\lambda}) \in \text{gph}\,\partial g$. Then the following properties are equivalent:*

(a) $\bar{\lambda} \in \text{ri}\,\partial g(\bar{z})$;

(b) *for any $r > 0$, the envelope function $e_r g$ is $\mathcal{C}^2$ in a neighborhood of $\bar{z} + r\bar{\lambda}$;*

(c) *for any $r > 0$, the proximal mapping prox_{rg} is $\mathcal{C}^1$ in a neighborhood of $\bar{z} + r\bar{\lambda}$.*

Furthermore, if $\bar{\lambda} \in \text{ri}\,\partial g(\bar{z})$ and $r > 0$, then for any x sufficiently close to $\bar{z} + r\bar{\lambda}$, the Jacobian matrix $\nabla(\text{prox}_{rg})(x)$ and the Hessian matrix $\nabla^2(e_r g)(x)$ can be calculated, respectively, by

$$\nabla(\text{prox}_{rg})(x) = P_{K_g(\bar{z},\bar{\lambda})} \quad \textit{and} \quad \nabla^2(e_r g)(x) = \frac{1}{r}\big(I - P_{K_g(\bar{z},\bar{\lambda})}\big) = \frac{1}{r} P_{K_{g^*}(\bar{\lambda},\bar{z})}.$$

Proof. Set $\varphi(z) = g(z) - \langle \bar{\lambda}, z\rangle$ for any $z \in \mathbb{R}^m$. Since $\bar{\lambda} \in \partial g(\bar{z})$, we get $0 \in \partial\varphi(\bar{z})$. This, together with the convexity of φ, yields $\bar{z} \in \text{argmin}\,\varphi$, where $\text{argmin}\,\varphi$ stands for the set of global minimizers of φ over $\mathbb{R}^m$. Observe also that the condition $\bar{\lambda} \in \text{ri}\,\partial g(\bar{z})$ is equivalent to $0 \in \text{ri}\,\partial\varphi(\bar{z})$ and that

$$\text{prox}_{r\varphi}(z) = \text{prox}_{rg}(z + r\bar{\lambda}) \quad \text{for all} \;\; z \in \mathbb{R}^m. \tag{4.6}$$

Picking $r > 0$ and employing [23, Theorem 4.4] imply that $\text{prox}_{r\varphi}$ is $\mathcal{C}^1$ in a neighborhood of $\bar{z}$ if and only if φ is strictly twice epi-differentiable at z for v for all $(z, v) \in \text{gph}\,\partial\varphi$ sufficiently close to $(\bar{z}, 0)$ (the parameter r in [23, Theorem 4.4] should be chosen sufficiently small since the function under consideration in [23] is prox-regular. It is, however, well-known that such a restriction on r for convex functions is not necessary). By Theorem 4.3, the latter property of φ amounts to the condition $0 \in \text{ri}\,\partial\varphi(\bar{z})$. It is not hard to see that φ is strictly twice epi-differentiable at z for v for any $(z, v) \in \text{gph}\,\partial\varphi$ sufficiently close to $(\bar{z}, 0)$ if and only if g enjoys the same property at z for λ for any $(z, \lambda) \in \partial g$ sufficiently close to $(\bar{z}, \bar{\lambda})$. Combining these with (4.6) and Theorem 4.3, we conclude the equivalence of (a) and (c). To obtain the equivalence of (a) and (b), one can see that

$$e_r\varphi(z) = e_r g(z + r\bar{\lambda}) - \langle \bar{\lambda}, z\rangle - \frac{r}{2}\|\bar{\lambda}\|^2 \quad \text{for all} \;\; z \in \mathbb{R}^m.$$

This, combined with a similar argument via [23, Theorem 4.4], confirms that (a) and (b) are equivalent.

Finally, pick $r > 0$ and suppose that $\bar{\lambda} \in \operatorname{ri} \partial g(\bar{z})$. By (b) and (c), there exists a neighborhood U of $\bar{z} + r\bar{\lambda}$ such that prox_{rg} is $\mathcal{C}^1$ and $e_r g$ is $\mathcal{C}^2$ on U. Take $x \in U$ and set $y = \operatorname{prox}_{rg}(x)$. By (4.5), we get $v := r^{-1}(x - y) \in \partial g(y)$. It also follows from (4.5) and the definition of graphical derivative that

$$D(\operatorname{prox}_{rg})(x)(w) = \big(I + rD(\partial g)(y, v)\big)^{-1}(w) \quad \text{for all } \ w \in \mathbb{R}^m,$$

which, together with (3.16), brings us to

$$D(\operatorname{prox}_{rg})(x)(w) = \big(I + rN_{K_g(y,v)}\big)^{-1}(w) = \big(I + N_{K_g(y,v)}\big)^{-1}(w) = P_{K_g(y,v)}(w)$$

for any $w \in \mathbb{R}^m$. By (c), the proximal mapping prox_{rg} is differentiable at x and thus we get $D(\operatorname{prox}_{rg})(x) = \nabla(\operatorname{prox}_{rg})(x)$. Combining these confirms the claimed formula for the Jacobian matrix $\nabla(\operatorname{prox}_{rg})(x)$. Recall also from [33, Theorem 2.26] that $\nabla(e_r g)(x) = r^{-1}(x - \operatorname{prox}_{rg}(x))$ for any $x \in \mathbb{R}^m$. By (b), the envelope function $e_r g$ is twice differentiable at x and thus we have

$$\nabla^2(e_r g)(x) = \frac{1}{r}\big(I - \nabla(\operatorname{prox}_{rg})(x)\big) = \frac{1}{r}\big(I - P_{K_g(y,v)}\big) = \frac{1}{r}P_{K_{g^*}(v,y)},$$

where the last equality comes from the identity $P_{K_g(y,v)} + P_{K_g(y,v)^*} = I$ together with (3.5). Now, we claim that $K_g(y, v) = K_g(\bar{z}, \bar{\lambda})$ whenever $x \in U$. This can be accomplished via Theorem 3.3(b) provided that we show $(y, v) \in \operatorname{gph} \partial g$ is sufficiently close to $(\bar{z}, \bar{\lambda})$. Since the proximal mapping is nonexpansive and since $\operatorname{prox}_{rg}(\bar{z} + r\bar{\lambda}) = \bar{z}$, we get

$$\|y - \bar{z}\| = \|\operatorname{prox}_{rg}(x) - \operatorname{prox}_{rg}(\bar{z} + r\bar{\lambda})\| \le \|x - \bar{z} - r\bar{\lambda}\|.$$

Moreover, we have

$$\|v - \bar{\lambda}\| = \|r^{-1}(x - y) - \bar{\lambda}\| \le r^{-1}\|x - \bar{z} - r\bar{\lambda}\| + r^{-1}\|y - \bar{z}\| \le 2r^{-1}\|x - \bar{z} - r\bar{\lambda}\|.$$

Using these estimates and shrinking U, if necessary, confirm our claim and hence prove the claimed formulas for the Jacobian matrix $\nabla(\operatorname{prox}_{rg})(x)$ and the Hessian matrix $\nabla^2(e_r g)(x)$. $\square$

We should mention that smoothness of projection mapping onto a closed convex set was first studied by Holmes in [8] in Hilbert spaces. His main result, [8, Theorem 2], states that if $C \subset \mathbb{R}^d$ is a closed convex set, $x \in \mathbb{R}^d$, the boundary of C is a $\mathcal{C}^2$ smooth manifold (cf. [33, Example 6.8]) around $y = P_C(x)$, then the projection mapping P_C is $\mathcal{C}^1$ in a neighborhood of the open normal ray $\{y + t(x - y)|\ t > 0\}$. As pointed out by Hiriart-Urruty in [9], when the projection point y is a *corner point*, Holmes's result can not be utilized to study smoothness of the projection mapping because the boundary of C fails to be a $\mathcal{C}^2$ smooth manifold around y. In contrast, Theorem 4.5 goes beyond the projection mapping and provides a characterization of smoothness of the proximal mapping of a polyhedral function via a verifiable condition. While our result is limited to polyhedral functions, our approach via second-order variational analysis opens a new door to study smoothness of projection mappings of convex sets. It is important to emphasize that our approach to characterize smoothness of proximal mappings demonstrates that instead of expecting smoothness of the boundary of the convex set under consideration, we should look for a second-order regularity condition, which seems to be the driving force for such a result.

Corollary 4.6. *Assume that $C \subset \mathbb{R}^m$ is a polyhedral convex set and $x \in \mathbb{R}^m$. Then P_C is $\mathcal{C}^1$ in a neighborhood of x if and only if $x - z \in \operatorname{ri} N_C(z)$, where $z = P_C(x)$.*

Proof. Applying Theorem 4.5 to the polyhedral function $g = \delta_C$ proves the claimed equivalence. $\square$

Note that a characterization of differentiability at a point, but not continuous differentiability *around* a point, of the projection mapping P_C, C being a polyhedral convex set, via the same relative interior condition as in Corollary 4.6 can be found in [4, Corollary 4.1.2]. Not only is our proof different from the one in [4], also Corollary 4.6 improves the latter result by showing that differentiability of the projection mapping at a given point can be strengthened to the $\mathcal{C}^1$ property of this mapping in a neighborhood of that point.

5 Regularity Properties of Variational Systems

In this section, we aim to explore the relationship between important regularity properties of the solution mapping to the generalized equation (1.1). When $g = \delta_C$ with C being a polyhedral convex set, the generalized equation (1.1) will be an example of variational inequalities. In this case, the seminal paper [3] revealed for the first time that strong metric regularity and metric regularity of the solution mapping to the canonical perturbation of (1.1) are equivalent; see below for the definitions of both concepts. The given proof in [3] for the latter equivalence relied heavily on Robinson's results in [27] and did not utilize second-order generalized differentiation tools that we are going to exploit in this section. We aim to present a proof of a similar result for the generalized equation (1.1), which is based upon our development of strict proto-differentiability of polyhedral functions. This forces us to restrict our analysis to the so-called nondegenerate solutions to (1.1), meaning solutions that satisfy the condition (1.2). Appealing to Theorem 4.5, we immediately arrive at the following equivalent descriptions of nondegenerate solutions to (1.1). Note that prox_g stands for the proximal mapping of the polyhedral function g with the parameter value $r = 1$.

Proposition 5.1. *Assume that $\bar{x}$ is a solution to the generalized equation (1.1). Then the following properties are equivalent:*

(a) *$\bar{x}$ is a nondegenerate solution to (1.1);*
(b) *the critical cone $K_g(\bar{x}, -\psi(\bar{x}))$ is a linear subspace;*
(c) *the proximal mapping prox_g is $\mathcal{C}^1$ in a neighborhood of $\bar{x} - \psi(\bar{x})$.*

To explore regularity properties of nondegenerate solutions to (1.1), define the set-valued mapping $G : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ by

$$G(x) := \psi(x) + \partial g(x), \quad x \in \mathbb{R}^m, \tag{5.1}$$

and then consider the solution mapping $S : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ to the canonical perturbation of the generalized equation (1.1) by

$$S(y) := G^{-1}(y) = \big\{x \in \mathbb{R}^m \,\big|\, y \in \psi(x) + \partial g(x)\big\}, \quad y \in \mathbb{R}^m. \tag{5.2}$$

Recall that a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is called *metrically regular* at $\bar{x}$ for $\bar{y} \in F(\bar{x})$ if there exist $\kappa \geq 0$ and neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that the distance estimate

$$\text{dist}\big(x, F^{-1}(y)\big) \leq \kappa \, \text{dist}\big(y, F(x)\big) \tag{5.3}$$

holds for all $(x, y) \in U \times V$. When the estimate (5.3) holds for any $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$, we call F globally metrically regular. It is known (cf. [10, Theorem 5.9(a)]) that when F is positively homogeneous and metrically regular at $\bar{x}$ for $\bar{y}$, it is globally metrically regular. The mapping is called *strongly metrically regular* at $\bar{x}$ for $\bar{y}$ if F^{-1} admits a Lipschitz continuous single-valued localization around $\bar{y}$ for $\bar{x}$, which means that there exist neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that the mapping $y \mapsto F^{-1}(y) \cap U$ is single-valued and Lipschitz continuous on V. According to [2, 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 its inverse F^{-1} admitting a single-valued localization around $\bar{y}$ for $\bar{x}$.

Our main goal in the rest of this section is to show that if $\bar{x}$ is a nondegenerate solution to the generalized equation (1.1), then metric regularity and strong metric regularity of the mapping G from (5.1) at $\bar{x}$ for 0 are equivalent. Since strong metric regularity of G translates as its inverse mapping $G^{-1} = S$ having a Lipschitz continuous single-valued localization and since S is the solution mapping to the canonical perturbation of the generalized equation (1.1), we will be able to find verifiable conditions for single-valuedness and Lipschitz continuity of the solution mapping S. To this end, we begin with understanding metric regularity of G from (5.1) using a characterization of this property via the graphical derivative, established in [2, Theorem 4B.1]: a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ whose graph is locally closed around $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ is metrically regular at $\bar{x}$ for $\bar{y}$ if and only if the condition

$$\limsup_{(x,y)\xrightarrow{\operatorname{gph} F}(\bar{x},\bar{y})} |DF(x,y)^{-1}|^{-} < \infty \tag{5.4}$$

is satisfied, where the inner norm $|DF(x,y)^{-1}|^{-}$ is calculated by

$$|DF(x,y)^{-1}|^{-} = \sup_{\|u\|\le 1} \inf_{w\in DF(x,y)^{-1}(u)} \|w\| \tag{5.5}$$

with the convention $\inf_{w\in\emptyset}\|w\| = \infty$. For any $(x,y) \in \operatorname{gph} F$, we deduce from (5.5) that the condition $|DF(x,y)^{-1}|^{-} < \infty$ yields $DF(x,y)^{-1}(u) \neq \emptyset$ for all $u \in \mathbb{R}^m$. The latter is equivalent to saying that for any $u \in \mathbb{R}^m$, there exists $w \in \mathbb{R}^m$ such that $u \in DF(x,y)(w)$, meaning that

$$\operatorname{rge} DF(x,y) = \big\{u \in \mathbb{R}^m \big|\ \exists\, w \in \mathbb{R}^n \text{ with } u \in DF(x,y)(w)\big\} = \mathbb{R}^m.$$

This tells us that the condition $|DF(x,y)^{-1}|^{-} < \infty$ implies that the graphical derivative mapping $DF(x,y) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is surjective. To get the opposite conclusion, we conclude from [2, Proposition 4A.6] that if $\operatorname{gph} DF(x,y) = T_{\operatorname{gph} F}(x,y)$ is convex, the surjectivity of $DF(x,y)$ is necessary and sufficient for the condition $|DF(x,y)^{-1}|^{-} < \infty$.

Next, we demonstrate that a point-based counterpart of (5.4) can be used to characterize metric regularity of the mapping G from (5.1) for nondegenerate solutions to (1.1).

Theorem 5.2 (point-based criteria for metric regularity)**.** *Assume that $\bar{x}$ is a nondegenerate solution to the generalized equation (1.1) and set $\overline{K} = K_g(\bar{x}, -\psi(\bar{x}))$. Then the following properties are equivalent:*

(a) *the mapping G from (5.1) is metrically regular at $\bar{x}$ for 0;*
(b) $|DG(\bar{x},0)^{-1}|^{-} < \infty$*;*
(c) $DG(\bar{x},0)$ *is surjective;*
(d) $DG(\bar{x},0)$ *is globally metrically regular;*
(e) $\big(\nabla\psi(\bar{x})\overline{K}\big) + \overline{K}^{\perp} = \mathbb{R}^m$*;*
(f) $\{w \in \mathbb{R}^m \,|\, \nabla\psi(\bar{x})^* w \in \overline{K}^{\perp}\} \cap \overline{K} = \{0\}$*.*

Proof. By the definition of G from (5.1), the sum rule for the graphical derivative from [33, Exercise 10.43(b)], and Theorem 3.6(a), we obtain for any $(x,y) \in \operatorname{gph} G$ and $w \in \mathbb{R}^m$ that

$$\begin{aligned} DG(x,y)(w) &= \nabla\psi(x)w + D(\partial g)(x, y-\psi(x))(w) \\ &= \nabla\psi(x)w + N_{K_g(x,y-\psi(x))}(w). \end{aligned}$$

Since $\bar{x}$ is a nondegenerate solution to (1.1), we deduce from Theorem 3.3(b) that there exists $r > 0$ such that for any $(x, y) \in \mathbb{B}_r(\bar{x}, 0) \cap \operatorname{gph} G$, we have $K_g(x, y - \psi(x)) = K_g(\bar{x}, -\psi(\bar{x})) = \overline{K}$. Combing these with $\overline{K}$ being a linear subspace brings us to

$$u \in DG(x, y)(w) \iff u \in \nabla\psi(x)w + \overline{K}^{\perp}, \;\; w \in \overline{K}. \tag{5.6}$$

This clearly tells us that $\operatorname{gph} DG(\bar{x}, 0)$ is a convex set and thus proves the equivalence of (b) and (c) using our discussion prior to this theorem. Also, we can conclude from (5.6) that

$$\operatorname{rge} DG(\bar{x}, 0) = (\nabla\psi(\bar{x})\overline{K}) + \overline{K}^{\perp}, \tag{5.7}$$

which implies that $DG(\bar{x}, 0)$ is surjective if and only if $\mathbb{R}^m = \operatorname{rge} DG(\bar{x}, 0) = \nabla\psi(\bar{x})\overline{K} + \overline{K}^{\perp}$. This justifies that (c) and (e) are equivalent. Furthermore, the surjectivity of $DG(\bar{x}, 0)$ means $\operatorname{rge} DG(\bar{x}, 0) = \mathbb{R}^m$, which amounts to the condition $0 \in \operatorname{int} \operatorname{rge} DG(\bar{x}, 0)$ due to the fact that the graphical derivative mapping $DG(\bar{x}, 0)$ is positively homogeneous (cf. [2, page 216]). This, combined with $0 \in DG(\bar{x}, 0)(0)$ and [2, Theorem 5B.4], implies that (c) is equivalent to the fact that the mapping $DG(\bar{x}, 0)$ is metrically regular at 0 for 0. According to [10, Theorem 5.9(a)], the latter amounts to (d), confirming that (c) and (d) are equivalent. The equivalence of (e) and (f) results from taking the orthogonal complement from both sides of either (e) or (f) and observing by [33, Corollary 11.25(c)] that

$$\big(\nabla\psi(\bar{x})\overline{K}\big)^{\perp} = \big\{w \in \mathbb{R}^m \,|\; \nabla\psi(\bar{x})^* w \in \overline{K}^{\perp}\big\}.$$

Suppose now that (a) holds. By (5.4), we obtain (b). To justify the implication (c) $\Longrightarrow$ (a), suppose that (a) fails. By (5.4), there exists a sequence $\{(x^k, y^k)\}_{k\in\mathbb{N}} \subset \operatorname{gph} G$, converging to $(\bar{x}, 0)$, such that $|DG(x^k, y^k)^{-1}|^{-} > k$ for any $k \in \mathbb{N}$. It follows from (5.5) that for any k sufficiently large, we find u^k with $\|u^k\| \le 1$ and $w^k \in DG(x^k, y^k)^{-1}(u^k)$ such that $\|w^k\| > k$. Thus, by (5.6), we arrive at the conditions

$$\frac{u^k}{\|w^k\|} - \nabla\psi(x^k)\frac{w^k}{\|w^k\|} \in \overline{K}^{\perp} \quad \text{and} \quad \frac{w^k}{\|w^k\|} \in \overline{K}.$$

Passing to a subsequence, if neccessary, we get $-\nabla\psi(\bar{x})\bar{w} \in \overline{K}^{\perp}$ for some $0 \ne \bar{w} \in \overline{K}$. We are going to show that $|DG(\bar{x}, 0)^{-1}|^{-} = \infty$. To do so, let $s = \dim \overline{K}$ and observe that $\dim \overline{K}^{\perp} = m - s$ and $\dim\big(\nabla\psi(\bar{x})\overline{K}\big) \le s$. If $\nabla\psi(\bar{x})\bar{w} = 0$, the classical rank-nullity theorem from linear algebra leads us to $\dim\big(\nabla\psi(\bar{x})\overline{K}\big) < s$. Otherwise, we have $0 \ne \nabla\psi(\bar{x})\bar{w} \in (\nabla\psi(\bar{x})\overline{K}) \cap \overline{K}^{\perp}$, which confirms that $\dim\big((\nabla\psi(\bar{x})\overline{K}) \cap \overline{K}^{\perp}\big) > 0$. Combining these with (5.7) results in

$$\dim\big(\operatorname{rge} DG(\bar{x}, 0)\big) = \dim\big((\nabla\psi(\bar{x})\overline{K}) + \overline{K}^{\perp}\big) < s + (m - s) = m,$$

which means that $DG(\bar{x}, 0)$ is not surjective, a contradiction of (c). This proves the implication (c) $\Longrightarrow$ (a) and hence completes the proof. □

When $g = 0$ in the generalized equation (1.1), Theorem 5.2 boils down to the classical result for strictly differentiable functions (cf. [17, Theorem 1.57]). The latter result states that metric regularity for such a function amounts to surjectivity of its Jacobian matrix.

We have one more step to take before proving the main result of this section, which is to show that strict proto-differentiability is preserved for the sum of two functions.

Proposition 5.3 (sum rule for strict proto-derivative)**.** *Let $f:\mathbb{R}^n \to \mathbb{R}^m$ be $\mathcal{C}^1$ around $\bar{x}$ and let $F:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$ with $\bar{y}\in F(\bar{x})$. Then we have*

$$D_*(f+F)(\bar{x},f(\bar{x})+\bar{y})(w)=\nabla f(\bar{x})w+D_*F(\bar{x},\bar{y})(w) \quad \textit{for all } w\in \mathbb{R}^n.$$

Moreover, F is strictly proto-differentiable at $\bar{x}$ for $\bar{y}$ if and only if $f+F$ is strictly proto-differentiable at $\bar{x}$ for $f(\bar{x})+\bar{y}$.

Proof. The given formula for $D_*(f+F)(\bar{x},f(\bar{x})+\bar{y})$ was already established in [33, Exercise 10.43(b)]. To justify the second claim, suppose that F is strictly proto-differentiable at $\bar{x}$ for $\bar{y}$. Let $(w,u)\in \operatorname{gph} D_*(f+F)(\bar{x},\bar{y}+f(\bar{x}))$ and take arbitrary sequences $t_k\searrow 0$, and $(x^k,z^k)\to(\bar{x},\bar{y}+f(\bar{x}))$ with $\{(x^k,z^k)\}_{k\in\mathbb{N}}\subset \operatorname{gph}(f+F)$. The latter tells us that $(x^k,y^k)\to(\bar{x},\bar{y})$ with $y^k:=z^k-f(x^k)$. By the sum rule for the strict graphical derivative, we get $(w,u-\nabla f(\bar{x})w)\in \operatorname{gph} D_*F(\bar{x},\bar{y})$. Since F is strictly proto-differentiable at $\bar{x}$ for $\bar{y}$, we find a sequence $(w^k,v^k)\to(w,u-\nabla f(\bar{x})w)$ such that $y^k+t_kv^k\in F(x^k+t_kw^k)$ for all $k\in\mathbb{N}$, which in turn implies for any $k\in\mathbb{N}$ that

$$z^k+t_ku^k\in(f+F)(x^k+t_kw^k) \quad \text{with } \ u^k:=\frac{f(x^k+t_kw^k)-f(x^k)}{t_k}+v^k.$$

Since $\{(w^k,u^k)\}_{k\in\mathbb{N}}$ converges to (w,u), we conclude that $f+F$ is strictly proto-differentiable at $\bar{x}$ for $f(\bar{x})+\bar{y}$.

Assume now that $f+F$ is strictly proto-differentiable at $\bar{x}$ for $f(\bar{x})+\bar{y}$. By the argument above, we can conclude that $F=f+F+(-f)$ is strictly proto-differentiable at $\bar{x}$ for $\bar{y}$, which completes the proof. $\square$

To present the main result of this section, let us recall a result, established in [2, Theorem 4D.1] in which sufficient conditions for strong metric regularity of a set-valued mapping were obtained: a set-valued mapping $F:\mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is strongly metrically regular at $\bar{x}$ for $\bar{y}\in F(\bar{x})$ provided that it has a closed graph and the conditions

$$0\in D_*F(\bar{x},\bar{y})(w) \implies w=0 \tag{5.8}$$

and

$$\bar{x}\in \liminf_{y\to\bar{y}} F^{-1}(y) \tag{5.9}$$

hold.

Theorem 5.4 (equivalence between metric regularity and strong metric regularity)**.** *Assume that $\bar{x}$ is a nondegenerate solution to the generalized equation (1.1). Then the mapping G, taken from (5.1), is metrically regular at $\bar{x}$ for 0 if and only if it is strongly metrically regular at $\bar{x}$ for 0.*

Proof. As mentioned in the paragraph after (5.3), strong metric regularity always implies metric regularity. To justify the opposite conclusion for the mapping G, suppose that G is metrically regular at $\bar{x}$ for 0. By the definition of G, $\operatorname{gph} G$ is closed. Moreover, the estimate (5.3), adopted for metric regularity of G at $\bar{x}$ for 0, clearly confirms the validity of (5.9). To prove (5.8), suppose that $0\in D_*G(\bar{x},0)(w)$. It follows from (1.2) and Theorem 3.6(c) that ∂g is strictly proto-differetiable at $\bar{x}$ for $-\psi(\bar{x})$, which, together with Proposition 5.3, tells us that G is strictly proto-differetiable at $\bar{x}$ for 0. By (3.15), we get $D_*G(\bar{x},0)(w)=DG(\bar{x},0)(w)$. Combining this with (5.6) indicates that the condition $0\in D_*G(\bar{x},0)(w)$ amounts to $-\nabla\psi(\bar{x})w\in \overline{K}^{\perp}$ and $w\in\overline{K}$, where $\overline{K}=K_g(\bar{x},-\psi(\bar{x}))$. According to Proposition 5.1, the critical cone $\overline{K}$ is a linear subspace of $\mathbb{R}^m$ of dimension $s \le m$. Let $B \in \mathbb{R}^{m\times s}$ be a matrix whose columns form a basis for $\overline{K}$. Thus, we obtain $\overline{K} = \text{rge}\, B$, which amounts to the condition $\overline{K}^\perp = \ker B^*$. So, the conditions $-\nabla\psi(\bar{x})w \in \overline{K}^\perp$ and $w \in \overline{K}$ tell us that there exists $\mu \in \mathbb{R}^s$ such that $w = B\mu$ and

$$0 = B^*\nabla\psi(\bar{x})w = B^*\nabla\psi(\bar{x})B\mu.$$

We claim that $B^*\nabla\psi(\bar{x})B$ is an $s \times s$ nonsingular matrix. This immediately yields $w = B\mu = 0$ and hence proves (5.8) for G. To justify the claim, we conclude from metric regularity of G at $\bar{x}$ for 0 and Theorem 5.2 that $DG(\bar{x}, 0)$ is surjective, meaning that for any $u \in \mathbb{R}^m$ there exists $w \in \mathbb{R}^m$ such that $u \in DG(\bar{x}, 0)(w)$. By (5.6), the latter condition is equivalent to $u - \nabla\psi(\bar{x})w \in \overline{K}^\perp$ and $w \in \overline{K}$. Since $\overline{K} = \text{rge}\, B$ and $\overline{K}^\perp = \ker B^*$, we find $q \in \mathbb{R}^s$ such that $w = Bq$ and that $u - \nabla\psi(\bar{x})Bq \in \ker B^*$, or equivalently, $B^*u = B^*\nabla\psi(\bar{x})Bq$. Since $u \in \mathbb{R}^m$ was taken arbitrary, the latter equality leads us

$$\text{rge}\,\big(B^*\nabla\psi(\bar{x})B\big) = B^*\mathbb{R}^m = \text{rge}\, B^* = \mathbb{R}^s,$$

where the last equality comes from B having full column rank. This confirms that $B^*\nabla\psi(\bar{x})B$ is an $s \times s$ nonsingular matrix and thus completes the proof. □

As pointed earlier in this section, the equivalence of metric regularity and strong metric regularity for the generalized equation (1.1) with $g = \delta_C$, where C is a polyhedral convex set, was established by Donchev and Rockafellar in [3, Theorem 3] without the extra assumption (1.2) using a different approach; see also [10, Corollary 9.7] for another proof of this result. Theorem 5.4 presents an extension of Donchev and Rockafellar's seminal result in [3] under the condition (1.2) with a proof relying heavily on second-order generalized differentiation. As we will show in our subsequent paper, our approach can be applied to any prox-regular functions that is strictly proto-differentiable. This allows us to obtain a similar result for a broader class of generalized equations.

Remark 5.5. Note that one can argue via [2, Corollary 3F.5] that (strong) metric regularity of the mapping G from (5.1) at $\bar{x}$ for $\bar{y} \in G(\bar{x})$ is equivalent to that of the mapping $x \mapsto \psi(\bar{x}) + \nabla\psi(\bar{x})(x-\bar{x}) + \partial g(x)$ at $\bar{x}$ for $\bar{y}$. Employing then Theorem 3.1 tells us that (strong) metric regularity of the latter amounts to the same property of the mapping $\Phi(w) := \nabla\psi(\bar{x})w + N_{\overline{K}}(w)$ with $\overline{K} = K_g(\bar{x}, -\psi(\bar{x}))$ and $w \in \mathbb{R}^m$ at 0 for 0. Since Φ fits into the framework of [3], one can conclude via [3, Theorem 3] (or [10, Corollary 9.7]) that metric regularity and strong metric regularity are equivalent for Φ, which implies that these properties are equivalent for G as well without the assumption (1.2). It is, however, important to emphasis that our approach is different from both [3] and [10] and relies upon the characterization of strong metric regularity via the strict graphical derivative. We will show in our subsequent paper that the approach in this section can be extended to any prox-regular function g in (1.1) that is strictly proto-differentiable under the same relative interior condition (1.2). In contrast, both approaches in [3] and [10] depend on the particular properties of polyhedral convex sets and do not seem to be applicable to other classes of functions. While finishing this paper, we came across the recent results in [6, Corollary 7.3] in which a characterization of strong metric regularity of set-valued mappings that are graphically Lipschitzian manifold in the sense of [33, Definition 9.66] was obtained. It is not clear, however, whether the latter Lipschitzian assumption does hold for the mapping G from (5.1).

Next, we present conditions under which the solution mapping S from (5.2) admits a Lipschitz continuous single-valued localization, which is continuously differentiable. To do so, recall from [29, page 173] that a set $C \subset \mathbb{R}^d$ is called *smooth* at $\bar{x} \in C$ if the tangent cone $T_C(\bar{x})$ is a linear subspace of $\mathbb{R}^d$ and the "lim sup" in (2.1) is the "lim." It is called *strictly smooth* at $\bar{x}$ if $T_C(\bar{x})$ is a linear subspace of $\mathbb{R}^d$ and

$$T_C(\bar{x}) = \lim_{x \xrightarrow{C} \bar{x}, t \searrow 0} \frac{C - x}{t}.$$

It follows from [29, Proposition 3.1] that if $f : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous around $\bar{x} \in \mathbb{R}^n$, then $\operatorname{gph} f$ is (strictly) smooth at $(\bar{x}, f(\bar{x}))$ if and only if f is (strictly) differentiable at $\bar{x}$.

Theorem 5.6. *Assume that $\bar{x}$ is a nondegenerate solution to the generalized equation (1.1). Then the solution mapping S to (1.1) has a Lipschitz continuous single-valued localization σ around $0 \in \mathbb{R}^m$ for $\bar{x}$ if and only if one of the equivalent properties (a)-(f) in Theorems 5.2 holds. In this case, the function σ is $\mathcal{C}^1$ in a neighborhood of 0 and*

$$\nabla\sigma(y) = B\big(B^*\nabla\psi(\sigma(y))B\big)^{-1}B^*$$

for all y sufficiently close to 0, where $B \in \mathbb{R}^{m\times s}$ is a matrix whose columns form a basis for the linear subspace $\overline{K} = K_g(\bar{x}, -\psi(\bar{x}))$.

Proof. The first part of this corollary results immediately from Theorems 5.2 and 5.4, the definition of strong metric regularity of G at $\bar{x}$ for 0, and the fact that $S = G^{-1}$ with G is taken from (5.1). Assume now that one of the equivalent properties (a)-(f) holds, which implies that G is strongly metrically regular at $\bar{x}$ for 0. This confirms that S has a Lipschitz continuous single-valued localization around 0 for $\bar{x}$, so we find neighborhoods U of 0 and V of $\bar{x}$ such that the mapping $y \mapsto S(y) \cap U$ is single-valued and Lipschitz continuous on V. Define the function $\sigma : V \to U$ by $\sigma(y) = S(y)\cap U$ for any $y \in V$. Shrinking the neighborhoods U and V, if necessary, we conclude from (5.6) that for any $(x, y) \in (U \times V) \cap \operatorname{gph} G$, the tangent cone $\operatorname{gph} DG(x, y) = T_{\operatorname{gph} G}(x, y)$ is a linear subspace. Since $\operatorname{gph}\sigma = \operatorname{gph} S$ locally around $(0, \bar{x})$ and since $S = G^{-1}$, we conclude that $T_{\operatorname{gph}\sigma}(y, x)$ is a linear subspace as well. It follows from the nondegeneracy condition (1.2) and Theorem 3.6(c) that ∂g is strictly proto-differentiable at x for z whenever $(x, z) \in \operatorname{gph}\partial g$ is sufficiently close to $(\bar{x}, -\psi(\bar{x}))$, which, together with Proposition 5.3, tells us that G is strictly proto-differentiable at x for y whenever $(x, y) \in \operatorname{gph} G$ is sufficiently close to $(\bar{x}, 0)$. Suppose without loss of generality that G is strictly proto-differentiable at x for y whenever $(x, y) \in (U \times V)\cap\operatorname{gph} G$. Choose a pair $(x, y) \in (U \times V)\cap\operatorname{gph} G$ and observe that σ is strictly proto-differentiable at y for x. This, together with $T_{\operatorname{gph}\sigma}(y, x)$ being a linear subspace, implies that $\operatorname{gph}\sigma$ is strictly smooth at (y, x). Since σ is Lipschitz continuous on V, it follows from [29, Proposition 3.1] that σ is strictly differentiable at y. This means that σ is strictly differentiable on V, a property equivalent to saying that σ is $\mathcal{C}^1$ on V (cf. [2, Exercise 1D.8]).

Finally, to justify the claimed formula for the Jacobian matrix of σ, take $y \in V$. Thus, for any $u \in \mathbb{R}^n$ we have $w = \nabla\sigma(y)u = D\sigma(y)(u)$, which is equivalent to $u \in DG(x, y)(w)$, where $x = \sigma(y)$. Appealing now to (5.6) brings us to the conditions $u - \nabla\psi(x)w \in \overline{K}^{\perp}$ and $w \in \overline{K}$. By the definition of B, we have $\overline{K} = \operatorname{rge} B$, or equivalently, $\overline{K}^{\perp} = \ker B^*$. Thus, we find $q \in \mathbb{R}^s$ such that $w = Bq$ and that $u - \nabla\psi(x)Bq \in \ker B^*$, or equivalently, $B^*u = B^*\nabla\psi(x)Bq$. Similar to the proof of Theorem 5.4, we can show that the matrix $B^*\nabla\psi(x)B$ is nonsingular. This leads us to

$$\nabla\sigma(y)u = w = Bq = B\big(B^*\nabla\psi(x)B\big)^{-1}B^*u,$$

which confirms the claimed formula for $\nabla\sigma(y)$ and hence completes the proof. □

Theorem 5.6 can be viewed as an extension of the classical inverse mapping theorem for generalized equations. This well-known result ensures under the nonsingularity of the Jacobian matrix that the inverse of a $\mathcal{C}^1$ function has a Lipschitz continuous single-valued localization, which is continuously differentiable. Robinson, in his landmark paper [24], showed that for generalized equations, one can expect under appropriate conditions that their solution mappings have a Lipschitz continuous single-valued localization. Theorem 5.6 demonstrates that for nondegenerate solutions to some particular class of generalized equations such a localization can be continuously differentiable as well.

We should also add that one can show that the continuous differentiability of ψ from (1.1) can be weakened to strict differentiability at a given solution $\bar{x}$ to (1.1) in all the results in this section with no harm. In such a case, we can only expect that the function σ in Theorem 5.6 be strictly differentiable at 0.

We close this section with an application of our main result in studying regularity properties of the solution mapping to the KKT system of the composite minimization problem

$$\text{minimize } \varphi(x) + \big(g \circ \Phi\big)(x), \quad \text{subject to } x \in \mathbb{R}^n, \tag{5.10}$$

where $\varphi : \mathbb{R}^n \to \mathbb{R}$ and $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ are $\mathcal{C}^2$ functions and $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ is a polyhedral function. The KKT system associated with the composite problem (5.10) is given by

$$0 = \nabla_x L(x, \lambda), \quad \lambda \in \partial g(\Phi(x)), \tag{5.11}$$

where $L(x, \lambda) := \varphi(x) + \langle \lambda, \Phi(x) \rangle$ with $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$ is the Lagrangian of (5.10). A pair $(\bar{x}, \bar{\lambda})$ is called a *KKT point* of (5.10) provided that it satisfies the KKT system (5.11). Define the mapping $\Psi : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$ by

$$\Psi(x, \lambda) := \begin{bmatrix} \nabla_x L(x, \lambda) \\ -\Phi(x) \end{bmatrix} + \begin{bmatrix} 0 \\ \partial g^*(\lambda) \end{bmatrix} \tag{5.12}$$

and observe that $(\bar{x}, \bar{\lambda})$ is a KKT point if and only if $(0, 0) \in \Psi(\bar{x}, \bar{\lambda})$. We aim at finding conditions under which the solution mapping to the canonical perturbed of the KKT system (5.11), defined by

$$S(p, q) := \Psi^{-1}(p, q) = \big\{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \,\big|\, (p, q) \in \Psi(x, \lambda)\big\},$$

has a Lipschitz continuous single-valued localization. As shown below, this can be distilled from Theorem 5.6.

Theorem 5.7. *Let $(\bar{x}, \bar{\lambda})$ be a KKT point of (5.10) with $\bar{\lambda} \in \operatorname{ri} \partial g(\Phi(\bar{x}))$. Then the following properties are equivalent:*

(a) *the mapping Ψ is metrically regular at $(\bar{x}, \bar{\lambda})$ for $(0, 0)$;*
(b) *the mapping Ψ is strongly metrically regular at $(\bar{x}, \bar{\lambda})$ for $(0, 0)$;*
(c) *the solution mapping S has a Lipschitz continuous single-valued localization around $(\bar{x}, \bar{\lambda})$ for $(0, 0)$, which is $\mathcal{C}^1$ in a neighborhood of $(\bar{x}, \bar{\lambda})$;*
(d) *the implication*

$$\begin{cases} \nabla^2_{xx} L(\bar{x}, \bar{\lambda})w + \nabla \Phi(\bar{x})^* w' = 0, \\ \nabla \Phi(\bar{x}) w \in K_g(\Phi(\bar{x}), \bar{\lambda}), \ w' \in K_g(\Phi(\bar{x}), \bar{\lambda})^{\perp} \end{cases} \implies (w, w') = (0, 0)$$

holds.

Proof. Set

$$\psi(x, \lambda) := \begin{bmatrix} \nabla_x L(x, \lambda) \\ -\Phi(x) \end{bmatrix} \quad \text{and} \quad \hat{g}(x, \lambda) := g^*(\lambda), \ \ (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m,$$

and observe that the KKT system (5.11) can be written as the generalized equation

$$(0,0) \in \Psi(x,\lambda) = \psi(x,\lambda) + \partial\hat{g}(x,\lambda). \tag{5.13}$$

It follows from [33, Theorem 11.14(a)] that g^* is a polyhedral function and so is $\hat{g}$. Also, we deduce from [33, Proposition 10.5] that $\partial\hat{g}(x,\lambda) = \{0\} \times \partial g^*(\lambda)$. Moreover, by (3.6), the condition $\bar{\lambda} \in \operatorname{ri}\partial g(\Phi(\bar{x}))$ is equivalent to $\Phi(\bar{x}) \in \operatorname{ri}\partial g^*(\bar{\lambda})$. Combining these tells us that $\bar{\lambda} \in \operatorname{ri}\partial g(\Phi(\bar{x}))$ amounts to the condition $-\psi(\bar{x},\bar{\lambda}) \in \operatorname{ri}\partial\hat{g}(\bar{x},\bar{\lambda})$. The equivalence of (a)-(c) comes directly from Theorems 5.4 and 5.6. Part (d) is an adaptation of the property in Theorem 5.2(f) for the generalized equation (5.13). To elaborate more, we can use [33, Proposition 10.5] to conclude that $\mathrm{d}\hat{g}(\bar{x},\bar{\lambda})(w,w') = \mathrm{d}g^*(\bar{\lambda})(w')$ for any $(w,w') \in \mathbb{R}^n \times \mathbb{R}^m$. This, together with the definition of the critical cone, shows that

$$K_{\hat{g}}\big((\bar{x},\bar{\lambda}), -\psi(\bar{x},\bar{\lambda})\big) = \mathbb{R}^n \times K_{g^*}(\Phi(\bar{x}),\bar{\lambda}) = \mathbb{R}^n \times K_g(\Phi(\bar{x}),\bar{\lambda})^* = \mathbb{R}^n \times K_g(\Phi(\bar{x}),\bar{\lambda})^{\perp}.$$

This yields $K_{\hat{g}}\big((\bar{x},\bar{\lambda}), -\psi(\bar{x},\bar{\lambda})\big)^{\perp} = \{0\} \times K_g(\Phi(\bar{x}),\bar{\lambda})$ and

$$\nabla\psi(\bar{x},\bar{\lambda})^*(w,w') \in K_{\hat{g}}\big((\bar{x},\bar{\lambda}), -\psi(\bar{x},\bar{\lambda})\big)^{\perp} \iff \begin{cases} \nabla^2_{xx}L(\bar{x},\bar{\lambda})w - \nabla\Phi(\bar{x})^*w' = 0, \\ \nabla\Phi(\bar{x})w \in K_g(\Phi(\bar{x}),\bar{\lambda}). \end{cases}$$

Combining this with Theorem 5.2 confirms that (d) and (a) are equivalent and hence completes the proof. □

For classical nonlinear programming problems (NLPs), it is well-known that metric regularity and strong metric regularity of KKT systems are equivalent; see [2, Theorem 4I.2] and [11, Section 7.5]. By using a new approach, Theorem 5.7 extends this result for the composite problem (5.10) under an extra condition, called the *strict complementarity condition*. This extra condition allows us to demonstrate further that the Lipschitz continuous single-valued localization of the solution mapping to the KKT system of (5.10) is continuously differentiable. This can be viewed as an extension of Fiacco and McCormick's result in [5] for NLPs, which was achieved under the classical second-order sufficient condition, strict complementarity condition, and linear independence constraint qualification.

Acknowledgments

The first author would like to thank Vietnam Institute for Advanced Study in Mathematics for hospitality during her post-doctoral fellowship of the Institute in 2021–2022.

References

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

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

[3] Dontchev, A.L., Rockafellar, R.T.: Characterizations of strong regularity for variational inequalities over polyhedral convex sets. SIAM J. Optim. **6**(4), 1087–1105 (1996)

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

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

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

[7] Hang, N.T.V., Sarabi, M.E.: Local convergence analysis of augmented Lagrangian methods for piecewise linear-quadratic composite optimization problems. SIAM J. Optim. **31**(4), 2665–2694 (2021)

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

[9] Hiriart-Urruty, J.-B.: At what points is the projection mapping differentiable? Amer. Math. Monthly **89**(7), 456–458 (1982)

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

[11] Klatte, D., Kummer, B.: Nonsmooth equations in optimization. Regularity, calculus, methods and applications. Kluwer, Dordrecht, (2002)

[12] Lemaréchal, C., Oustry, F., Sagastizábal, C.: The $\mathcal{U}$-Lagrangian of a convex function. Trans. Amer. Math. Soc. **352** (2), 711–729 (2000)

[13] Lemaréchal, C.,Sagastizábal, C.: More than first-order developments of convex functions: primal-dual relations. J. Convex Anal. **3**(2), 255–268 (1996)

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

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

[16] Mohammadi, A., Sarabi, M.E.: Twice epi-differentiability of extended-real-valued functions with applications in composite optimization. SIAM J. Optim. **30**(3), 2379–2409 (2020)

[17] Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation, I: Basic Theory; II: Applications. Springer Berlin, Heidelberg (2006)

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

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

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

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

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

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

[24] Robinson, S.M.: Strongly regular generalized equations. Math. Oper. Res. **5**(1), 43–62 (1980)

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

[26] Robinson, S.M.: An implicit-function theorem for a class of nonsmooth functions. Math. Oper. Res. **16**(2), 292–309 (1991)

[27] Robinson, S.M.: Normal maps induced by linear transformations. Math. Oper. Res. **17**(3), 691–714 (1992)

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

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

[30] Rockafellar, R.T.: First- and second-order epi-differentiability in nonlinear programming. Trans. Amer. Math. Soc. **307**(1), 75–108 (1988)

[31] Rockafellar, R.T.: Generalized second derivatives of convex functions and saddle functions. Trans. Amer. Math. Soc. **322**(1), 51–77 (1990)

[32] Rockafellar, R.T.: Augmented Lagrangians and hidden convexity in sufficient conditions for local optimality. Math. Program. (2022). doi.org/10.1007/s10107-022-01768-w

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

[34] Sarabi, M.E.: Primal superlinear convergence of SQP methods for piecewise linear-quadratic composite optimization problems. Set-Valued Var. Anal. **30**, 1–37 (2022)