Convergence of Augmented Lagrangian Methods for Composite Optimization Problems

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

Abstract. Local convergence analysis of the augmented Lagrangian method (ALM) is established for a large class of composite optimization problems with nonunique Lagrange multipliers under a second-order sufficient condition. We present a new second-order variational property, called the semi-stability of second subderivatives, and demonstrate that it is widely satisfied for numerous classes of functions, important for applications in constrained and composite optimization problems. Using the latter condition and a certain second-order sufficient condition, we are able to establish Q-linear convergence of the primal-dual sequence for an inexact version of the ALM for composite programs.

Keywords. augmented Lagrangian, nonunique Lagrange multiplier, second-order sufficient condition, $\mathcal{C}^2$-decomposable functions, Q-linear convergence.

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

1 Introduction

The augmented Lagrangian method, also known as the method of multipliers, was introduced and studied independently by Hestenes in [14] and Powell in [28] for nonlinear programming problems (NLPs) with equality constraints. It was later extended in [30] for problems of convex programming with inequality constraints by Rockafellar, where its convergence analysis was carried out by applying the proximal point algorithm to the dual of the augmented Lagrangian problem and ensuring the Q-linear convergence of the dual sequence and R-linear convergence of the primal sequence in the ALM. Most of the publications about the ALM afterwards exploited a different approach to establish the local convergence of this method, since the original idea in [30] relied heavily on duality, which was not available for nonconvex settings. The pioneering work of Fernández and Solodov in [10] was perhaps the culmination of those efforts over the last three decades to find conditions under which the Q-linear convergence of the primal-dual sequence, constructed by the ALM, can be achieved for NLPs with nonunique Lagrange multipliers. Instead of looking for duality in the augmented Lagrangian problem, the authors in [10] showed that the primal-dual iterates of the ALM are indeed a solution to a particular perturbation of the KKT system of the original problem. That paved the path for them to utilize a general framework by Fischer in [11] for solving generalized equations and demonstrate that the classical second-order sufficient condition alone suffices for local convergence analysis of the ALM and no constraint qualification is necessary for such a result.

Another possible approach to conduct local convergence analysis of the ALM for nonconvex optimization problems was recently developed by Rockafellar in [32], where he extended his original idea in [30] of using duality and the proximal point algorithm to obtain convergence of the dual sequence in the ALM. The principal idea therein was to assume the strong variational convexity (see page 166 in [31]) of the augmented Lagrangian function, which was characterized in [31, Theorem 5] to be equivalent to the *strong* second-order sufficient conditions for NLPs. While the approach in [32] provides a general framework for the local convergence analysis of

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

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

the ALM, it operates under a rather strong assumption, namely the strong second-order sufficient condition for constrained optimization problems. This, in particular, can be a challenging assumption to disentangle when we deal with important classes of optimization problems such as eigenvalue optimization problems.

Following the approach by Fernández and Solodov in [10], we aim to analyze the local convergence of the ALM for a class of composite optimization problems that have a representation of the form

$$\begin{cases} \text{minimize } \; \psi(x) \quad \text{subject to } \; x \in \Theta, \\ \text{with } \; \psi(x) := \varphi(x) + g(\Phi(x)), \end{cases} \tag{1.1}$$

where $\varphi : \mathbf{X} \to \mathbf{R}$ and $\Phi : \mathbf{X} \to \mathbf{Y}$ are twice continuously differentiable functions, $g : \mathbf{Y} \to \overline{\mathbf{R}} := [-\infty, \infty]$ is a proper lsc convex function, Θ is a polyhedral convex set, and $\mathbf{X}$ and $\mathbf{Y}$ are finite dimensional Hilbert spaces. The composite problem in (1.1) encompasses many important classes of constrained and composite optimization problems such as NLPs and nonlinear semidefinite programming problems (SDPs), convex piecewise linear-quadratic composite optimization problems, and eigenvalue related optimization problems. We reveal two major second-order variational conditions for the convex function g in (1.1), which together with the second-order sufficient condition allow us to establish Q-linear convergence of the primal-dual sequence of the ALM for (1.1). In particular, we do not assume any constraint qualification, and hence are able to deal with composite problems with nonunique Lagrange multipliers.

The main idea of the ALM is to smooth out the nondifferentiable parts in the composite function ψ in (1.1) and to minimize then the obtained augmented function for the next iterate of the method. To elaborate more, consider the augmented Lagrangian function $\mathscr{L} : \mathbf{X} \times \mathbf{Y} \times (0, \infty) \to \mathbf{R}$, defined by

$$\mathscr{L}(x, y, \rho) := \inf_{u \in \mathbf{Y}} \big\{ \psi(x, u) + \frac{1}{2\rho} \|u\|^2 - \langle y, u \rangle \big\}, \quad (x, y, \rho) \in \mathbf{X} \times \mathbf{Y} \times (0, \infty),$$

where $\psi(x, u)$ is a *partial* perturbation of ψ in (1.1), given by $\psi(x, u) := \varphi(x) + g(\Phi(x) + u)$ for any $(x, u) \in \mathbf{X} \times \mathbf{Y}$. Note that the full perturbation of ψ requires to replace ψ with $\psi + \delta_\Theta$ and consider then a second perturbation variable for Θ. This complicates our convergence analysis and seems unnecessary, however. A direct calculation then shows that for any $(x, y, \rho) \in \mathbf{X} \times \mathbf{Y} \times (0, \infty)$, the augment Lagrangian $\mathscr{L}$ can be equivalently expressed as

$$\mathscr{L}(x, y, \rho) = \varphi(x) + e_{1/\rho} g\big(\Phi(x) + \rho^{-1} y\big) - \tfrac{1}{2} \rho^{-1} \|y\|^2, \tag{1.2}$$

where $e_{1/\rho} g$ stands for the Moreau envelope of g, given by

$$(e_{1/\rho} g)(y) := \inf_{z \in \mathbf{Y}} \big\{ g(z) + \tfrac{1}{2} \rho \|y - z\|^2 \big\}, \quad y \in \mathbf{Y}.$$

Given the current triple $(x^k, y^k, \rho_k) \in \mathbf{X} \times \mathbf{Y} \times (0, \infty)$, the *exact* version of the ALM generates the next primal iterate x^{k+1} and the dual iterate y^{k+1}, respectively, by

$$x^{k+1} \in \operatorname*{arg\,min}_{x \in \Theta} \mathscr{L}(x, y^k, \rho_k) \quad \text{and} \quad y^{k+1} = y^k + \rho_k \nabla_y \mathscr{L}(x^{k+1}, y^k, \rho_k). \tag{1.3}$$

Since selecting x^{k+1} as an exact minimizer of the augmented Lagrangian function does not seem practical, we are going to consider a more realistic scenario and demand that the next primal iterate x^{k+1} be an *approximate* stationary solution to the constrained augmented problem

$$\text{minimize } \; \mathscr{L}(x, y^k, \rho_k) \quad \text{subject to} \quad x \in \Theta, \tag{1.4}$$

namely it satisfies the condition

$$\operatorname{dist}\big(-\nabla_x \mathscr{L}(x^{k+1}, y^k, \rho_k), N_\Theta(x^{k+1})\big) \le \epsilon_k, \tag{1.5}$$

where $\epsilon_k \ge 0$ is called the *tolerance* parameter. Note that an exact minimizer x^{k+1} as the one in (1.3) satisfies (1.5) with $\epsilon_k = 0$.

In our recent work [13], we extended the local convergence analysis of the ALM in [10] for a class of composite optimization problems that the modeling function therein, the function g in (1.1), was assumed to be convex piecewise linear-quadratic (CPLQ). In the present paper, our primary goal is twofold. First, we are going to demonstrate that a similar result can be obtained for composite optimization problem in (1.1) with the modeling function g belonging to a large class of functions, called $\mathcal{C}^2$-decomposable; see (3.6). Second, we provide two major conditions under which the local convergence of the ALM for (1.1) can be achieved. In doing so, we introduce two new concepts, the semi-strict graphical derivative and semi-strict second subderivative, and calculate them for some important classes of functions. We also introduce a new concept, called semi-stability of second subderivatives, and demonstrate that it holds for various important classes of functions in optimization. This property is then utilized to ensure a uniform version of the quadratic growth condition for the augmented Lagrangian function in (1.2), which is a major tool in our approach to analyze the convergence of the ALM.

The outline of the paper is as follows. We begin in Section 2 by recalling an iterative framework for our inexact ALM and reviewing our notation. The goal of Section 3 is to provide a general property under which we can ensure the metric subregualrity of the Karush-Kuhn-Tucker (KKT) system associated with (1.1). In Section 4, we introduce the concept of semi-stability of second subderivatives and use it to justify the quadratic growth condition for the augmented Lagrangian function. Section 5 is devoted to establishing the Q-linear convergence of the primal-dual sequence generated by the proposed inexact ALM. We summarize our results and discuss some open problems in Section 6. Finally, in Section 7, we present two independent results that are important for our developments in this paper.

2 Prelimenaries

2.1 Notation

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

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

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

Given a nonempty set $\Omega \subset \mathbf{X}$ with $\bar{x} \in \Omega$, the tangent cone to Ω at $\bar{x}$, denoted $T_\Omega(\bar{x})$, is defined by $T_\Omega(\bar{x}) = \limsup_{t \searrow 0} \frac{\Omega - \bar{x}}{t}$. The regular/Fréchet normal cone $\widehat{N}_\Omega(\bar{x})$ to Ω at $\bar{x}$ is defined by $\widehat{N}_\Omega(\bar{x}) = T_\Omega(\bar{x})^*$. For $x \notin \Omega$, we set $\widehat{N}_\Omega(x) = \emptyset$. The limiting/Mordukhovich normal cone $N_\Omega(\bar{x})$ to Ω at $\bar{x}$ is the set of all vectors $\bar{v} \in \mathbf{X}$ for which there exist sequences $\{x^k\}_{k\in\mathbb{N}}$ and $\{v^k\}_{k\in\mathbb{N}}$

with $v^k \in \widehat{N}_\Omega(x^k)$ such that $(x^k, v^k) \to (\bar{x}, \bar{v})$. When Ω is convex, both normal cones boil down to that of convex analysis. Given a function $f : \mathbf{X} \to \overline{\mathbf{R}}$ and a point $\bar{x} \in \mathbf{X}$ with $f(\bar{x})$ finite, the subderivative function $\mathrm{d}f(\bar{x})\colon \mathbf{X} \to \overline{\mathbf{R}}$ 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 $v \in \mathbf{X}$ is called a regular subgradient of f at $\bar{x}$ if $(v, -1) \in \widehat{N}_{\mathrm{epi}\, f}(\bar{x}, f(\bar{x}))$ with $\mathrm{epi}\, f = \{(x, \alpha) \in \mathbf{X} \times \mathbf{R} \,|\, f(x) \le \alpha\}$ being the epigraph of f. The set of all regular subgradients of f at $\bar{x}$ is denoted by $\widehat{\partial} f(\bar{x})$. Similarly, we can define $\partial f(\bar{x})$ using the limiting normal cone $N_{\mathrm{epi}\, f}(\bar{x}, f(\bar{x}))$. When f is a convex function, both sets reduce to the well-known subdifferential of convex functions. The critical cone of f at $\bar{x}$ for $\bar{v}$ with $\bar{v} \in \partial f(\bar{x})$ is defined by

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

When $f = \delta_\Omega$, where Ω is a nonempty subset of $\mathbf{X}$, the critical cone of δ_Ω at $\bar{x}$ for $\bar{v}$ is denoted by $K_\Omega(\bar{x}, \bar{v})$. In this case, the above definition of the critical cone of a function boils down to the well-known concept of the critical cone of a set (see [9, page 109]), namely $K_\Omega(\bar{x}, \bar{v}) = T_\Omega(\bar{x}) \cap [\bar{v}]^\perp$. The second subderivative of f at $\bar{x}$ for $\bar{v}$, denoted $\mathrm{d}^2 f(\bar{x}, \bar{v})$, is an extended-real-valued function defined by

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

where $\Delta_t^2 f(\bar{x}, \bar{v})$ is the parametric family of second-order difference quotients of f at $\bar{x}$ for $\bar{v}$ and is defined for any $w \in \mathbf{X}$ and $t > 0$ by

$$\Delta_t^2 f(\bar{x}, \bar{v})(w) = \frac{f(\bar{x} + tw) - f(\bar{x}) - t\langle \bar{v},\, w\rangle}{\frac{1}{2}t^2}. \tag{2.1}$$

The function f is called twice epi-differentiable at $\bar{x}$ for $\bar{v}$ if for any $w \in \mathbf{X}$ and any $t_k \searrow 0$, there exists $w^k \to w$ such that $\Delta_{t_k}^2 f(\bar{x}, \bar{v})(w^k) \to \mathrm{d}^2 f(\bar{x}, \bar{v})(w)$; see [33, Definition 13.6(b)]. For a set-valued mapping $F : \mathbf{X} \rightrightarrows \mathbf{Y}$, its graph and domain are defined, respectively, by $\mathrm{gph}\, F = \big\{(x, y) \in \mathbf{X} \times \mathbf{Y} \,\big|\, y \in F(x)\big\}$ and $\mathrm{dom}\, F = \big\{x \in \mathbf{X} \,\big|\, F(x) \neq \emptyset\big\}$. We say that F is calm at x for $y \in F(x)$ if there are neighborhoods U of x and V of y and a positive constant κ for which we have

$$F(x') \cap V \subset F(x) + \kappa\, \|x - x'\| \mathbb{B} \quad \text{for all } \ x' \in U.$$

The mapping F is called metrically subregular at x for $y \in F(x)$ if there are a neighborhood U of x and a positive constant ℓ such that $\mathrm{dist}(x', F^{-1}(y)) \le \ell\, \mathrm{dist}(y, F(x'))$ for any $x' \in U$.

2.2 An Iterative Framework for ALM

As shown later in this section, the primal-dual sequence $\{(x^k, y^k)\}_{k \in \mathbb{N}}$, generated by the ALM considered in this paper, are solutions to a particular perturbation of the KKT system associated with the composite optimization problem in (1.1), which is given by

$$0 \in \nabla_x L(x, y) + N_\Theta(x), \quad y \in \partial g(\Phi(x)), \tag{2.2}$$

where $L(x, y) = \varphi(x) + \langle y, \Phi(x) \rangle$ with $(x, y) \in \mathbf{X} \times \mathbf{Y}$ is the Lagrangian of (1.1) and where the functions g, Φ, and the set Θ are taken from (1.1). It is easy to see that (2.2) can be written as the generalized equation

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \nabla_x L(x, y) \\ -\Phi(x) \end{bmatrix} + \begin{bmatrix} N_\Theta(x) \\ (\partial g)^{-1}(y) \end{bmatrix}. \tag{2.3}$$

To elaborate more on the kind of perturbation we should consider for this generalized equation, we briefly recall an abstract iterative framework, suggested by Fischer in [11], for solving generalized equations with nonisolated solutions. Given the mappings $\Psi : \mathbf{H} \to \mathbf{H}'$ and $G : \mathbf{H} \rightrightarrows \mathbf{H}'$, where $\mathbf{H}$ and $\mathbf{H}'$ are finite dimensional Hilbert spaces, consider the generalized equation formulated by

$$0 \in \Psi(u) + G(u). \tag{2.4}$$

Define the solution mapping $\Upsilon : \mathbf{H}' \rightrightarrows \mathbf{H}$ to the canonical perturbation of (2.4) by

$$\Upsilon(v) = \big\{u \in \mathbf{H} |\, v \in \Psi(u) + G(u)\big\}, \quad v \in \mathbf{H}'. \tag{2.5}$$

Given a set of parameters $\mathbf{P}$, iterative algorithms for solving (2.4) often generate a sequence $\{u^k\}_{k\in\mathbb{N}}$ by iteratively solving subproblems of the form

$$0 \in \mathcal{A}(u, u^k, p_k) + G(u), \tag{2.6}$$

in which the single-valued part in (2.4) is replaced with a set-valued mapping $\mathcal{A} : \mathbf{H} \times \mathbf{H} \times \mathbf{P} \rightrightarrows \mathbf{H}'$, which gives an approximation of Ψ around the current iterate, while the set-valued part is kept unchanged. In order to ensure the convergence and to establish the rate of convergence of $\{u^k\}_{k\in\mathbb{N}}$, we have to choose a solution to (2.6) satisfying certain properties. Namely, given the current iterate u^k and a parameter p_k, we choose the next iterate u^{k+1} sufficiently close to u^k satisfying

$$u^{k+1} \in \big\{u \in \mathbf{H} \,\big|\, 0 \in \mathcal{A}(u, u^k, p_k) + G(u) \;\text{ and }\; \|u - u^k\| \le c\,\mathrm{dist}(u^k, \Upsilon(0))\big\}, \tag{2.7}$$

where $c > 0$ is arbitrary yet fixed, and $\Upsilon(0)$, taken from (2.5) with $v = 0$, is the set of solutions to (2.4). Below, we recall a result, established in [16, Theorem 4.1] (see also [18, Theorem 7.13]), in which the local convergence analysis of the sequence $\{u^k\}_{k\in\mathbb{N}}$ was obtained under rather mild assumptions. We should point out that assumption (a) in the theorem below was expressed in [16, Theorem 4.1] in a slightly stronger sense, namely the solution mapping Υ was assumed to be upper Lipschitzian instead of being calm. It is not hard to see from its proof, however, that calmness suffices for this result. The nonparametric version of this result was later appeared in [18, Theorem 7.13] with upper Lipschitzian replaced by calmness.

Theorem 2.1. *Let $\bar u \in \Upsilon(0)$, where Υ is the solution mapping from (2.5). Assume that $\Upsilon(0)$ is locally closed around $\bar u$ and that the following properties hold for some positive constant c:*

(a) (*calmness of solution mapping*) *the solution mapping Υ is calm at $\bar u$ for $0 \in \mathbf{H}'$ with constant $\ell_1 > 0$;*

(b) (*solvability of subproblems*) *there exists a positive constant ε_1 such that for any $\tilde u \in \mathbb{B}_{\varepsilon_1}(\bar u)$ and any $p \in \mathbf{P}$ the localized solution set*

$$\big\{u \in \mathbf{X} \,\big|\, 0 \in \mathcal{A}(u, \tilde u, p) + G(u) \;\; \textit{and} \;\; \|u - \tilde u\| \le c\,\mathrm{dist}(\tilde u, \Upsilon(0))\big\}$$

is nonempty;

(c) (*precision of approximation*) *there exist a positive constant ε_2 and a function $\omega : \mathbf{H} \times \mathbf{H} \times \mathbf{P} :\to \mathbf{R}_+$ such that*

$$\sup\big\{\omega(u, \tilde u, p) \,\big|\, \tilde u \in \mathbb{B}_{\varepsilon_2}(\bar u),\, \|u - \tilde u\| \le c\,\mathrm{dist}(\tilde u, \Upsilon(0)),\, p \in \mathbf{P}\big\} < 1/\ell_1,$$

where constant $\ell_1 > 0$ is taken from (a)*, and the estimate*

$$\sup\big\{\|w\| \,\big|\, w \in \Psi(u) - \mathcal{A}(u, \tilde u, p)\big\} \le \omega(u, \tilde u, p)\,\mathrm{dist}(\tilde u, \Upsilon(0))$$

holds for all $\tilde u \in \mathbb{B}_{\varepsilon_2}(\bar u)$, all $u \in \mathbf{H}$ with $\|u - \tilde u\| \le c\,\mathrm{dist}(\tilde u, \Upsilon(0))$, and all $p \in \mathbf{P}$.

Then there exists $\varepsilon_0 > 0$ such that for any starting point u^0 chosen from $\mathbb{B}_{\varepsilon_0}(\bar{u})$ and any sequence $\{p_k\}_{k\in\mathbb{N}} \subset \mathbf{P}$, the iterative scheme in (2.7) generates a sequence $\{u^k\}_{k\in\mathbb{N}}$ converging to some $\hat{u} \in \Upsilon(0)$. The rates of convergence of $\{u^k\}_{k\in\mathbb{N}}$ to $\hat{u}$ and of $\{\text{dist}(u^k, \Upsilon(0))\}_{k\in\mathbb{N}}$ to 0 are Q-linear, and they are Q-superlinear provided that $\omega(u^{k+1}, u^k, p_k) \to 0$ as $k \to \infty$. Moreover, for any $\varepsilon > 0$, we have $\|\hat{u} - \bar{u}\| < \varepsilon$ if u^0 is chosen sufficiently close to $\bar{u}$.

Our main objective in this paper is to show that if a second-order sufficient condition for optimality (see (3.18)) holds for the composite optimization problem in (1.1) and a certain calmness of the multiplier mapping associated with this problem (see (3.20)) is satisfied, then all assumptions (a)-(c) in Theorem 2.1 can be verified for many important classes of constrained and composite optimization problems.

Our final goal in this section is to show that the primal-dual sequence $\{(x^k, y^k)\}_{k\in\mathbb{N}}$, generated by the *inexact* ALM in this paper, can be fit into the iterative pattern described above. To this end, recall first that the proximal mapping of a convex function $g : \mathbf{Y} \to \overline{\mathbf{R}}$ is defined by

$$\text{prox}_{rg}(y) = \underset{z\in\mathbf{Y}}{\arg\min}\big\{g(z) + \tfrac{1}{2r}\|y - z\|^2\big\}, \quad y \in \mathbf{Y},$$

where r is a positive constant. In what follows, when $r = 1$, the proximal mapping of g will be denoted by prox_g. It follows from [33, Proposition 12.19] and [33, Theorem 2.26], respectively, that for any $y \in \mathbf{Y}$ we always have

$$\text{prox}_{rg}(y) = \big(I + r\partial g\big)^{-1}(y) \quad \text{and} \quad \nabla e_r g(y) = \big(rI + (\partial g)^{-1}\big)^{-1}(y), \tag{2.8}$$

where I stands for the identity mapping from $\mathbf{Y}$ onto $\mathbf{Y}$. Using these relationships, one can equivalently reformulate the dual update y^{k+1} in (1.3) as

$$y^{k+1} \in \partial g\big(\Phi(x^{k+1}) - \rho_k^{-1}(y^{k+1} - y^k)\big) \quad \text{or} \quad y^{k+1} = \nabla(e_{1/\rho_k} g)\big(\Phi(x^{k+1}) + \rho_k^{-1} y^k\big). \tag{2.9}$$

To express the KKT system in (2.2) in the form of the generalized equation in (2.4), we define the mappings $\Psi : \mathbf{X} \times \mathbf{Y} \to \mathbf{X} \times \mathbf{Y}$ and $G : \mathbf{X} \times \mathbf{Y} \rightrightarrows \mathbf{X} \times \mathbf{Y}$ by

$$\Psi(x, y) = \begin{bmatrix} \nabla_x L(x, y) \\ -\Phi(x) \end{bmatrix} \quad \text{and} \quad G(x, y) = \begin{bmatrix} N_\Theta(x) \\ (\partial g)^{-1}(y) \end{bmatrix}, \tag{2.10}$$

which clearly demonstrates that (2.2) can be covered by (2.4). Moreover, the inexact primal update x^{k+1}, satisfying (1.5), can be equivalently described via (2.9) by

$$\begin{aligned} 0 &\in \nabla_x \mathscr{L}(x^{k+1}, y^k, \rho_k) + \epsilon_k \mathbb{B} + N_\Theta(x^{k+1}) \\ &= \nabla\varphi(x^{k+1}) + \nabla\Phi(x^{k+1})^* \nabla(e_{1/\rho_k} g)\big(\Phi(x^{k+1}) + \rho_k^{-1} y^k\big) + \epsilon_k \mathbb{B} + N_\Theta(x^{k+1}) \\ &= \nabla_x L(x^{k+1}, y^{k+1}) + \epsilon_k \mathbb{B} + N_\Theta(x^{k+1}). \end{aligned}$$

In order to show that the primal-dual iterate (x^{k+1}, y^{k+1}) satisfies (2.6), we define the mapping $\mathcal{A} : \mathbf{X} \times \mathbf{Y} \times \mathbf{X} \times \mathbf{Y} \times \mathbf{R}_+ \times (0, \infty) \rightrightarrows \mathbf{X} \times \mathbf{Y}$ by

$$\mathcal{A}(x, y, \tilde{x}, \tilde{y}, \epsilon, \rho) := \begin{bmatrix} \nabla_x L(x, y) + \epsilon\mathbb{B} \\ -\Phi(x) + \rho^{-1}(y - \tilde{y}) \end{bmatrix} \tag{2.11}$$

for a given quadruple $(\tilde{x}, \tilde{y}, \epsilon, \rho) \in \mathbf{X} \times \mathbf{Y} \times \mathbf{R}_+ \times (0, \infty)$. Combining these with the first inclusion in (2.9) tells us that the primal-dual iterate (x^{k+1}, y^{k+1}) is a solution to the subproblem

$$(0, 0) \in \mathcal{A}(x, y, x^k, y^k, \epsilon_k, \rho_k) + G(x, y).$$

We close this section by recording some properties of the augmented Lagrangian function in (1.2), which directly result from (2.8). Note that for any $(x, \rho) \in \mathbf{X} \times (0, \infty)$, the mapping $y \mapsto \mathscr{L}(x, y, \rho)$ is $\mathcal{C}^1$ and its gradient can be calculated via (2.8) as

$$\nabla_y \mathscr{L}(x, y, \rho) = \rho^{-1}\big(\nabla(e_{1/\rho} g)\big(\Phi(x) + \rho^{-1} y\big) - y\big) = \Phi(x) - \text{prox}_{\rho^{-1} g}\big(\Phi(x) + \rho^{-1} y\big). \tag{2.12}$$

Proposition 2.2. *Let $(\bar{x}, \bar{y})$ be a solution to the KKT system in (2.2). Then for any $\rho > 0$, the following properties are satisfied.*

(a) $\mathscr{L}(\bar{x}, \bar{y}, \rho) = \psi(\bar{x}) = \varphi(\bar{x}) + g\big(\Phi(\bar{x})\big)$.

(b) $\nabla_x \mathscr{L}(\bar{x}, \bar{y}, \rho) = \nabla_x L(\bar{x}, \bar{y})$ *and* $\nabla_y \mathscr{L}(\bar{x}, \bar{y}, \rho) = 0$.

3 Metric Subregularity of KKT Systems

As shown in Theorem 2.1, the calmness of the solution mapping in (2.5) is required for local convergence analysis of the ALM for the composite optimization problem (1.1). In this section, we aim to provide a sufficient condition for an error bound estimate for the generalized equation in (2.3), which will be shown in Proposition 5.2 that is equivalent to the aforementioned calmness of (2.5). To achieve our goal, recall that the graphical derivative of a set-valued mapping $F : \mathbf{X} \rightrightarrows \mathbf{Y}$ at $\bar{x}$ for $\bar{y} \in F(\bar{x})$, denoted by $DF(\bar{x}, \bar{y})$, is a set-valued mapping from $\mathbf{X}$ into $\mathbf{Y}$, defined by

$$\operatorname{gph} DF(\bar{x}, \bar{y}) = T_{\operatorname{gph} F}(\bar{x}, \bar{y}).$$

Employing the definition of the tangent cone, one can easily obtain an equivalent sequential description of $\operatorname{gph} DF(\bar{x}, \bar{y})$. If, in addition, for any $\eta \in DF(\bar{x}, \bar{y})(w)$ and any choice of $t_k \searrow 0$, there exist sequences $w^k \to w$ and $\eta^k \to \eta$ with $\bar{y} + t_k \eta^k \in F(\bar{x} + t_k w^k)$, then F is said to be proto-differentiable at $\bar{x}$ for $\bar{y}$; see [33, page 331] for more details. One of the main source of proto-differentiability in variational analysis is subgradient mappings of various classes of functions including CPLQ functions; see [23, 24, 33] for more examples.

Definition 3.1 (semi-strict graphical derivatives)**.** The semi-strict graphical derivative of $F : \mathbf{X} \rightrightarrows \mathbf{Y}$ at $\bar{x}$ for $\bar{y} \in F(\bar{x})$, denoted by $\widehat{D}F(\bar{x}, \bar{y})$, is a set-valued mapping from $\mathbf{X}$ to $\mathbf{Y}$ defined by

$$\eta \in \widehat{D}F(\bar{x}, \bar{y})(w) \iff \begin{cases} \exists\, t_k \searrow 0,\ (x^k, y^k) \to (\bar{x}, \bar{y}),\ \widehat{y}^k \to \bar{y} \ \text{ with } y^k \in F(x^k),\ \widehat{y}^k \in F(\bar{x}) \\ \text{such that } \big(\frac{x^k - \bar{x}}{t_k}, \frac{y^k - \widehat{y}^k}{t_k}\big) \to (w, \eta). \end{cases}$$

Using the definition of the outer limit of a family of sets, one can see that the semi-strict graphical derivative of F can be equivalently expressed as

$$\widehat{D}F(\bar{x}, \bar{y})(w) = \limsup_{\substack{t \searrow 0,\, w' \to w \\ y \to \bar{y},\, y \in F(\bar{x})}} \frac{F(\bar{x} + t w') - y}{t}, \quad w \in \mathbf{X}. \tag{3.1}$$

By definition, we can immediately conclude that $DF(\bar{x}, \bar{y})(w) \subset \widehat{D}F(\bar{x}, \bar{y})(w)$ for any $w \in \mathbf{X}$. Our interest in the semi-strict graphical derivative resides in Theorem 3.9 in which we show using this notion that the second-order sufficient condition in (3.18) ensures an error bound for the KKT system of (1.1). In order to do so, we should investigate how $DF(\bar{x}, \bar{y})$ and $\widehat{D}F(\bar{x}, \bar{y})$ relate to each other when F is subgradient mappings. We begin with a result that can be used as our guide to pursue such a relationship. Recall first from [33, Definition 7.25] that an lsc function $f : \mathbf{X} \to \overline{\mathbf{R}}$ is said to be subdifferentially regular at $\bar{x} \in \mathbf{X}$ if $f(\bar{x})$ is finite and $\widehat{N}_{\operatorname{epi} f}(\bar{x}, f(\bar{x})) = N_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))$.

Proposition 3.2. *Assume that $f : \mathbf{X} \to \overline{\mathbf{R}}$ and $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$. If ∂f is proto-differentiable at $\bar{x}$ for $\bar{v}$, then we have*

$$\operatorname{cl}\big(D(\partial f)(\bar{x}, \bar{v})(w) - T_{\partial f(\bar{x})}(\bar{v})\big) \subset \widehat{D}(\partial f)(\bar{x}, \bar{v})(w) \quad \textit{for all} \quad w \in \mathbf{X}. \tag{3.2}$$

In addition, if f is subdifferentially regular at $\bar{x}$, then $T_{\partial f(\bar{x})}(\bar{v})$ in the left-hand side of (3.2) can be replaced with $K_f(\bar{x}, \bar{v})^*$.

Proof. Observe that the set $\widehat{D}(\partial f)(\bar{x},\bar{y})(w)$ is closed due to (3.1) and the fact that the outer limit of a sequence of sets is always closed (cf. [33, Prposition 4.4]). To prove (3.2), it suffices to take $\eta \in D(\partial f)(\bar{x},\bar{v})(w)$ and $\zeta \in T_{\partial f(\bar{x})}(\bar{v})$, and show that $\eta - \zeta \in \widehat{D}(\partial f)(\bar{x},\bar{v})(w)$. By $\zeta \in T_{\partial f(\bar{x})}(\bar{v})$, we find sequences $t_k \searrow 0$ and $\zeta^k \to \zeta$ with $\bar{v} + t_k\zeta^k \in \partial f(\bar{x})$. Moreover, proto-differentiability of f at $\bar{x}$ for $\bar{v}$ implies that there exist sequences $w^k \to w$ and $\eta^k \to \eta$ with $\bar{v} + t_k\eta^k \in \partial f(\bar{x} + t_k w^k)$. Putting all of these together, we have $(\bar{v} + t_k\zeta^k) + t_k(\eta^k - \zeta^k) \in \partial f(\bar{x} + t_k w^k)$, which leads us to $\eta - \zeta \in \widehat{D}(\partial f)(\bar{x},\bar{v})(w)$ and hence proves (3.2).

Assume now that f is subdifferentially regular at $\bar{x}$. According to [33, Theorem 8.30], we know that $\partial f(\bar{x})$ is closed and convex and that

$$K_f(\bar{x},\bar{v}) = N_{\partial f(\bar{x})}(\bar{v}), \tag{3.3}$$

which clearly justifies the final claim. □

We next examine whether the opposite inclusion in (3.2) holds for important classes of the modeling function g in the composite optimization problem (1.1). Before moving any further, let us show that if $f : \mathbf{X} \to \overline{\mathbf{R}}$ is convex and $(\bar{x},\bar{v}) \in \operatorname{gph}\partial f$, then we always have

$$\operatorname{dom}\widehat{D}(\partial f)(\bar{x},\bar{v}) \subset K_f(\bar{x},\bar{v}). \tag{3.4}$$

Indeed, assume that for $w \in \mathbf{X}$ there exists $\eta \in \widehat{D}(\partial f)(\bar{x},\bar{v})(w)$. By definition, we can select sequences $t_k \searrow 0$, $w^k \to w$, $v^k \in \partial f(\bar{x} + t_k w^k)$, and $\widehat{v}^k \to \bar{v}$ with $\widehat{v}^k \in \partial f(\bar{x})$ such that $(v^k - \widehat{v}^k)/t_k \to \eta$. In particular, we have $v^k \to \bar{v}$ and

$$\frac{f(\bar{x} + t_k w^k) - f(\bar{x})}{t_k} \le \langle v^k, w^k\rangle$$

for all k. Passing to the limit as $k \to \infty$ gives us $\mathrm{d}f(\bar{x})(w) \le \langle \bar{v}, w\rangle$ that actually holds as equality due to $\bar{v} \in \partial f(\bar{x})$ and [33, Exercise 8.4]. Thus, $w \in K_f(\bar{x},\bar{v})$, which proves (3.4).

Example 3.3. Assume that $f : \mathbf{X} \to \overline{\mathbf{R}}$ with $\mathbf{X} = \mathbf{R}^n$ is a CPLQ function. Recall that f is called piecewise linear-quadratic if $\operatorname{dom} f = \cup_{i=1}^s C_i$ with $s \in \mathbb{N}$ and C_i being polyhedral convex sets for $i = 1, \ldots, s$, and if f has a representation of the form

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

where A_i is an $n \times n$ symmetric matrix, $a_i \in \mathbf{R}^n$, and $\alpha_i \in \mathbf{R}$ for $i = 1, \ldots, s$. Pick $\bar{x} \in \operatorname{dom} f$ and $\bar{v} \in \partial f(\bar{x})$. Since f is convex, it is subdifferentially regular at $\bar{x}$, and ∂f is proto-differentiable at $\bar{x}$ for $\bar{v}$ according to [33, Proposition 13.9] and [33, Theorem 13.40]. Proposition 3.2 tells us that the inclusion in (3.2) holds. If $w \notin \operatorname{dom}\widehat{D}(\partial f)(\bar{x},\bar{v})$, then the opposite inclusion is trivial. We now verify the opposite inclusion for any $w \in \operatorname{dom}\widehat{D}(\partial f)(\bar{x},\bar{v})$. It follows from (3.4) that $w \in K_f(\bar{x},\bar{v})$. Take $\eta \in \widehat{D}(\partial f)(\bar{x},\bar{v})(w)$ and find, by definition, sequences $t_k \searrow 0$, $w^k \to w$, $v^k \in \partial f(\bar{x} + t_k w^k)$, and $\widehat{v}^k \to \bar{v}$ with $\widehat{v}^k \in \partial f(\bar{x})$ such that $(v^k - \widehat{v}^k)/t_k \to \eta$. We then have

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

It follows from the reduction lemma for CPLQ functions from [34, Theorem 2.3] that there exists a neighborhood $\mathcal{O}$ of $(0,0) \in \mathbf{R}^n \times \mathbf{R}^n$ for which we have

$$\big((\operatorname{gph}\partial f) - (\bar{x},\bar{v})\big) \cap \mathcal{O} = \big(\operatorname{gph} D(\partial f)(\bar{x},\bar{v})\big) \cap \mathcal{O}.$$

We can assume for any k sufficiently large that $(t_k w^k, v^k - \bar{v}) \in \mathcal{O}$ and therefore obtain $(w^k, (v^k - \bar{v})/t_k) \in \operatorname{gph} D(\partial g)(\bar{x}, \bar{v})$. By [34, Proposition 2.4(b)] and $w \in K_f(\bar{x}, \bar{v})$, there exists a constant $\tau \geq 0$ such that for any k sufficiently large we have

$$\frac{v^k - \bar{v}}{t_k} \in D(\partial f)(\bar{x}, \bar{v})(w^k) \subset D(\partial f)(\bar{x}, \bar{v})(w) + \tau \|w^k - w\| \mathbb{B}. \tag{3.5}$$

Note also that $\widehat{v}^k - \bar{v} \in \partial f(\bar{x}) - \bar{v} \subset T_{\partial f(\bar{x})}(\bar{v})$. This, coupled with (3.3), which holds for the CPLQ function f, brings us to $(\widehat{v}^k - \bar{v})/t_k \in K_f(\bar{x}, \bar{v})^*$ for all k. Using this and (3.5), we find $\eta^k \in D(\partial f)(\bar{x}, \bar{v})(w)$ and $\zeta^k \in \mathbb{B}$ such that

$$\frac{v^k - \widehat{v}^k}{t_k} - \tau \|w^k - w\| \zeta^k = \eta^k - \frac{\widehat{v}^k - \bar{v}}{t_k} \in D(\partial f)(\bar{x}, \bar{v})(w) - K_f(\bar{x}, \bar{v})^*$$

for any k sufficiently large. Passing to the limit as $k \to \infty$ and using the boundedness of ζ^k, we arrive at the inclusion

$$\eta \in \operatorname{cl}\big(D(\partial f)(\bar{x}, \bar{v})(w) - K_f(\bar{x}, \bar{v})^*\big),$$

which in turn demonstrates that the inclusion in (3.2) becomes equality for any $w \in \mathbf{R}^n$ for CPLQ functions. Note that in this framework, we can drop the closure in (3.2) to obtain

$$\widehat{D}(\partial f)(\bar{x}, \bar{v})(w) = D(\partial f)(\bar{x}, \bar{v})(w) - K_f(\bar{x}, \bar{v})^* \quad \text{for any } \ w \in \mathbf{X},$$

since the right-hand side is the sum of two polyhedral convex sets, which is known to be a polyhedral convex set, and hence is always closed.

Our next goal is to show that we should expect a similar result as the one in Example 3.3 outside the polyhedral framework. We begin with proving the following result, which is of its own interest. Recall from [35] that a function $g : \mathbf{Y} \to \overline{\mathbf{R}}$ is called $\mathcal{C}^2$-*decomposable* at $\bar{u} \in \mathbf{Y}$ if $g(\bar{u})$ is finite and g enjoys the composite representation

$$g(u) = g(\bar{u}) + \vartheta(\Xi(u)) \quad \text{ for } \quad u \in \mathcal{O}, \tag{3.6}$$

where $\mathcal{O} \subset \mathbf{Y}$ is an open neighborhood of $\bar{u}$, $\vartheta : \mathbf{Z} \to \overline{\mathbf{R}}$ is proper, lsc, and sublinear, and $\Xi : \mathcal{O} \to \mathbf{Z}$ is $\mathcal{C}^2$-smooth with $\bar{z} := \Xi(\bar{u}) = 0$, and where $\mathbf{Z}$ is a finite dimensional Hilbert space. Note that by [33, Definition 3.18 and Exercise 3.19], we always have $g(\bar{z}) = 0$, since g is proper and sublinear. As shown in [35, Example 2.4], the class of $\mathcal{C}^2$-decomposable functions is a generalization of $\mathcal{C}^2$-*cone reducible* sets in the sense of [6, Definition 3.135] through which many important constrained optimization problems can be studied. Recall that a closed convex set $C \subset \mathbf{Y}$ is $\mathcal{C}^2$-*cone reducible* at $\bar{u} \in \mathbf{Y}$ to the closed convex cone $\Theta \subset \mathbf{Z}$ if there exist a neighborhood $\mathcal{O} \subset \mathbf{Y}$ of $\bar{u}$ and a $\mathcal{C}^2$-smooth mapping $\Xi : \mathbf{Y} \to \mathbf{Z}$ such that

$$C \cap \mathcal{O} = \big\{u \in \mathcal{O} \,\big|\, \Xi(u) \in \Theta\big\}, \quad \Xi(\bar{u}) = 0, \quad \text{and} \quad \nabla \Xi(\bar{u}) : \mathbf{Y} \to \mathbf{Z} \ \text{ is surjective.} \tag{3.7}$$

Note that the surjectivity condition in (3.7) is not covered by (3.6) and has its own counterpart, called the *nondegeneracy* condition; see (3.13). Besides the indicator functions of $\mathcal{C}^2$-*cone reducible* sets, it was shown in [35, Examples 2.1 and 2.3] that polyhedral functions and the sum of the largest eigenvalues of a symmetric matrix are $\mathcal{C}^2$-decomposable. The later was extended for singular values of a matrix in [22, Example 5.3.18]. The readers can find more examples of $\mathcal{C}^2$-decomposable functions in [22, Section 5.3.3]. Note that CPLQ functions may not enjoy the composite representation in (3.6) in general; see Remark 3.7.

Suppose that the proper function $g:\mathbf{Y}\to\overline{\mathbf{R}}$ is $\mathcal{C}^2$-decomposable at $\bar u\in\mathbf{Y}$ with representation (3.6). Define the (Lagrange) multiplier mapping $M_{\bar u,g}:\mathbf{Y}\times\mathbf{Y}\rightrightarrows\mathbf{Z}$ by

$$M_{\bar u,g}(y,w)=\big\{\mu\in\mathbf{Z}\,\big|\,\nabla\Xi(\bar u)^*\mu=y,\ \mu\in\partial\vartheta(\Xi(\bar u)+w)\big\}. \tag{3.8}$$

Given $\bar y\in\partial g(\bar u)$, the set $M_{\bar u,g}(\bar y,0)$ is called the set of Lagrange multipliers associated with $(\bar u,\bar y)$. Below, we summerize some of the main properties of the composite representation in (3.6), used often in our proofs in this paper.

Proposition 3.4. *Assume that $g:\mathbf{Y}\to\overline{\mathbf{R}}$ is $\mathcal{C}^2$-decomposable at $\bar u\in\mathbf{Y}$ with representation (3.6) and that $\bar y\in\partial g(\bar u)$ and $\bar\mu\in M_{\bar u,g}(\bar y,0)$. Then the following properties hold.*

(a) *If the basic constraint qualification (BCQ) condition*

$$N_{\operatorname{dom}\vartheta}(\Xi(\bar u))\cap\ker\nabla\Xi(\bar u)^*=\{0\} \tag{3.9}$$

holds, then g is subdifferentially regular at $\bar u$.

(b) *For any $z\in\operatorname{dom}\vartheta$, it holds that $\partial\vartheta(z)\subset\partial\vartheta(\Xi(\bar u))$.*

(c) *It always holds that $K_\vartheta(\Xi(\bar u),\bar\mu)=N_{\partial\vartheta(\Xi(\bar u))}(\bar\mu)$. If, in addition, (3.9) is satisfied, then $K_g(\bar u,\bar y)=N_{\partial g(\bar u)}(\bar y)$.*

(d) *If the dual condition*

$$D(\partial\vartheta)(\Xi(\bar u),\bar\mu)(0)\cap\ker\nabla\Xi(\bar u)^*=\{0\} \tag{3.10}$$

holds, then $K_g(\bar u,\bar y)^*=\nabla\Xi(\bar x)^*K_\vartheta(\Xi(\bar u),\bar\mu)^*$.

(e) *If (3.9) is satisfied, the subgradient mapping ∂g is calm at $\bar u$ for $\bar y$.*

Proof. The conclusion in (a) is well-known; see [33, Exercise 10.25]. The inclusion in (b) results from [33, Corollary 8.25]. Both claims in (c) result from (a) and (3.3). To prove (d), the inclusion in (7.5) tells us that the dual condition in (3.10) yields (3.9). Thus, it follows from the chain rule for the subderivative in [33, Theorem 10.6] that

$$\begin{aligned}K_g(\bar u,\bar y)&=\big\{w\,\big|\,\mathrm{d}g(\bar u)(w)=\mathrm{d}\vartheta(\Xi(\bar u))(\nabla\Xi(\bar u)w)=\langle\bar y,w\rangle=\langle\bar\mu,\nabla\Xi(\bar x)w\rangle\big\}\\&=\big\{w\,\big|\,\nabla\Xi(\bar x)w\in K_\vartheta(\Xi(\bar u),\bar\mu)\big\}.\end{aligned} \tag{3.11}$$

Employing the chain rule for normal cones in [33, Exercise 10.26] allows us to conclude that

$$K_g(\bar u,\bar y)^*=N_{K_g(\bar u,\bar y)}(0)=\nabla\Xi(\bar u)^*N_{K_\vartheta(\Xi(\bar u),\bar\mu)}(0)=\nabla\Xi(\bar u)^*K_\vartheta(\Xi(\bar u),\bar\mu)^*,$$

which proves (d). The claim in (e) results from (b) and [32, Theorem 4.1]. □

As pointed out earlier, the equality in (3.2) plays a crucial role in our main result of this section. Below, we establish a chain rule for this property.

Proposition 3.5. *Assume that $g:\mathbf{Y}\to\overline{\mathbf{R}}$ is $\mathcal{C}^2$-decomposable at $\bar u\in\mathbf{Y}$ with representation (3.6) and that $\bar y\in\partial g(\bar u)$ and $\bar\mu\in M_{\bar u,g}(\bar y,0)$. If the outer function ϑ in (3.6) enjoys the property*

$$\widehat D(\partial\vartheta)(\Xi(\bar u),\bar\mu)(\xi)=\operatorname{cl}\big(D(\partial\vartheta)(\Xi(\bar u),\bar\mu)(\xi)-K_\vartheta(\Xi(\bar u),\bar\mu)^*\big)\quad\textit{for all}\quad\xi\in K_\vartheta(\Xi(\bar u),\bar\mu), \tag{3.12}$$

and the nondegeneracy condition

$$\operatorname{par}\{\partial\vartheta(\Xi(\bar u))\}\cap\ker\nabla\Xi(\bar u)^*=\{0\} \tag{3.13}$$

holds, then g satisfies the same property, meaning that

$$\widehat D(\partial g)(\bar u,\bar y)(w)=\operatorname{cl}\big(D(\partial g)(\bar u,\bar y)(w)-K_g(\bar u,\bar y)^*\big)\quad\textit{for all}\quad w\in\mathbf{Y}. \tag{3.14}$$

Proof. Pick $w \in \operatorname{dom} \widehat{D}(\partial g)(\bar u, \bar y)$ and $\eta \in \widehat{D}(\partial g)(\bar u, \bar y)(w)$. By definition, we find sequences $t_k \searrow 0$, $w^k \to w$, $y^k \in \partial g(\bar u + t_k w^k)$, and $\widehat{y}^k \to \bar y$ with $\widehat{y}^k \in \partial g(\bar u)$ such that $(y^k - \widehat{y}^k)/t_k \to \eta$. By the subdifferential chain rule from [33, Example 10.8] available due to the assumed nondegeneracy condition, we find $\mu^k \in \partial\vartheta(\Xi(\bar u + t_k w^k))$ and $\widehat{\mu}^k \in \partial\vartheta(\Xi(\bar u))$ such that $\nabla\Xi(\bar u + t_k w^k)^*\mu^k = y^k$ and $\nabla\Xi(\bar u)^*\widehat{\mu}^k = \widehat{y}^k$, respectively. Thus, we have $\mu^k \in M_{\bar u, g}(y^k, t_k\nabla\Xi(\bar u)w^k + o(t_k))$ and $\widehat{\mu}^k \in M_{\bar u, g}(\widehat{y}^k, 0)$. Note that it follows from (7.4) and Proposition 3.4(c) that

$$D(\partial\vartheta)(\Xi(\bar u), \bar\mu)(0) = K_\vartheta(\Xi(\bar u), \bar\mu)^* = T_{\partial\vartheta(\Xi(\bar u))}(\bar\mu) \subset \operatorname{par}\{\partial\vartheta(\Xi(\bar u))\}, \tag{3.15}$$

where the last inclusion results from the definition of the tangent cone. Thus, we conclude from the assumed nondegeneracy condition that the dual condition in (3.10) is satisfied. Appealing now to Theorem 7.1(d), we can conclude that both sequences $\{\mu^k\}_{k\in\mathbb{N}}$ and $\{\widehat{\mu}^k\}_{k\in\mathbb{N}}$ converge to $\bar\mu$ as $k \to \infty$. By passing to a subsequence, we can assume without loss of generality that the sequence $\{\nabla\Xi(\bar x)^*(\mu^k - \widehat{\mu}^k)/t_k\}_{k\in\mathbb{N}}$ is convergent, since

$$\begin{aligned}\frac{y^k - \widehat{y}^k}{t_k} &= \frac{\nabla\Xi(\bar u + t_k w^k)^*\mu^k - \nabla\Xi(\bar u)^*\widehat{\mu}^k}{t_k}\\ &= \Big(\frac{\nabla\Xi(\bar u + t_k w^k) - \Xi(\bar u)}{t_k}\Big)^*\mu^k + \nabla\Xi(\bar u)^*\frac{\mu^k - \widehat{\mu}^k}{t_k}.\end{aligned} \tag{3.16}$$

We claim now that the sequence $\{\|\mu^k - \widehat{\mu}^k\|/t_k\}_{k\in\mathbb{N}}$ is bounded. Otherwise, passing to a subsequence, if necessary, we can assume that $\|\mu^k - \widehat{\mu}^k\|/t_k \to \infty$, which implies that $\mu^k - \widehat{\mu}^k \neq 0$ for any k sufficiently large. Again, we can assume by passing to a subsequence, if necessary, that $\{(\mu^k - \widehat{\mu}^k)/\|\mu^k - \widehat{\mu}^k\|\}_{k\in\mathbb{N}}$ converges to some $\xi \neq 0$. Thus, we get $\xi \in \ker\nabla\Xi(\bar x)^*$, since

$$\nabla\Xi(\bar x)^*\xi = \lim_{k\to\infty}\nabla\Xi(\bar x)^*\frac{\mu^k - \widehat{\mu}^k}{\|\mu^k - \widehat{\mu}^k\|} = \lim_{k\to\infty}\nabla\Xi(\bar x)^*\frac{\mu^k - \widehat{\mu}^k}{t_k}\cdot\frac{t_k}{\|\mu^k - \widehat{\mu}^k\|} = 0.$$

On the other hand, it follows from Proposition 3.4(b) that $\mu^k \in \partial\vartheta(\Xi(\bar u + t_k w^k)) \subset \partial\vartheta(\Xi(\bar u))$, which in turn yields $(\mu^k - \widehat{\mu}^k)/\|\mu^k - \widehat{\mu}^k\| \in \operatorname{par}\{\partial\vartheta(\Xi(\bar u))\}$. Thus, we arrive at $\xi \in \operatorname{par}\{\partial\vartheta(\Xi(\bar u)\}$, which leads us to $\xi = 0$ due to (3.13), a contradiction. This tells us that $\{(\mu^k - \widehat{\mu}^k)/t_k\}_{k\in\mathbb{N}}$ is convergent to some $\zeta \in \mathbf{Z}$. Recall that $\mu^k \in \partial\vartheta(\Xi(\bar u + t_k w^k))$ and $\widehat{\mu}^k \in \partial\vartheta(\Xi(\bar u))$. Since $\big(\Xi(\bar u + t_k w^k) - \Xi(\bar u)\big)/t_k \to \nabla\Xi(\bar u)w$, it follows from Definition 3.1 that $\zeta \in \widehat{D}(\partial\vartheta)(\Xi(\bar u), \bar\mu)(\nabla\Xi(\bar u)w)$. Passing to the limit in (3.16) brings us to

$$\eta - \nabla^2\langle\bar\mu, \Xi\rangle(\bar u)w = \nabla\Xi(\bar u)^*\zeta \in \nabla\Xi(\bar u)^*\widehat{D}(\partial\vartheta)(\Xi(\bar u), \bar\mu)(\nabla\Xi(\bar x)w),$$

which in turn implies via (3.12), Theorem 7.2(b), and Proposition 3.4 that

$$\begin{aligned}\eta &\in \nabla^2\langle\bar\mu, \Xi\rangle(\bar u)w + \nabla\Xi(\bar u)^*\operatorname{cl}\big(\widehat{D}(\partial\vartheta)(\Xi(\bar u), \bar\mu)(\nabla\Xi(\bar u)w) - K_\vartheta(\Xi(\bar u), \bar\mu)^*\big)\\ &= \operatorname{cl}\big(\nabla^2\langle\bar\mu, \Xi\rangle(\bar u)w + \nabla\Xi(\bar u)^*\widehat{D}(\partial\vartheta)(\Xi(\bar u), \bar\mu)(\nabla\Xi(\bar u)w) - \nabla\Xi(\bar u)^*K_\vartheta(\Xi(\bar u), \bar\mu)^*\big)\\ &= \operatorname{cl}\big(D(\partial g)(\bar u, \bar y)(w) - K_g(\bar u, \bar y)^*\big).\end{aligned}$$

This proves the inclusion '$\subset$' in (3.14). To justify the opposite inclusion, observe from (7.5) and (3.13) that the BCQ in (3.9) is satisfied. By Proposition 3.4, g is subdifferentially regular at $\bar u$. Moreover, it follows from Theorem 7.2(b) that ∂g is proto-differentiable at $\bar u$ for $\bar y$. Thus, Proposition 3.2, together with (3.3), proves the inclusion '$\supset$' in (3.14) and hence completes the proof. $\square$

Using the established chain rule in the proposition above, we are going next to show that the inclusion in (3.2) becomes equality for the second-order cone.

Example 3.6. Let $\mathcal{Q}$ stand for the second-order cone in $\mathbf{R}^n$ described by

$$\mathcal{Q} := \big\{u = (u_0, u_r) \in \mathbf{R} \times \mathbf{R}^{n-1} \,\big|\, \|u_r\| \le u_0\big\}.$$

Our goal is to show that the inclusion in (3.2) becomes equality when $f = \delta_{\mathcal{Q}}$. To achieve this goal, we should consider there different cases for a point $\bar{u} \in \mathcal{Q}$. If $\bar{u} \in \operatorname{int} \mathcal{Q}$, it is easy to see that $\mathcal{Q}$ can be locally represented by the set $\{u \in \mathcal{O} \,|\, \Xi(u) \in \Theta\}$, where $\Xi : \mathbf{R}^n \to \{0\}$ and $\Theta = \{0\}$. If $\bar{u} \in \operatorname{bd} \mathcal{Q} \setminus \{0\}$, one can see that $\mathcal{Q}$ can be locally represented by $\{u \in \mathcal{O} \,|\, \Xi(u) \in \Theta\}$, where $\Xi : \mathbf{R}^n \to \mathbf{R}$ is defined by $\Xi(u) = \|u_r\| - u_0$ for any $u = (u_0, u_r) \in \mathbf{R} \times \mathbf{R}^{n-1}$ and $\Theta = \mathbf{R}_-$. Thus, in both cases, $\mathcal{Q}$ is $\mathcal{C}^2$-cone reducible to a polyhedral convex cone in the sense of (3.7) and hence $\delta_{\mathcal{Q}}$ is $\mathcal{C}^2$-decomposable with the outer function ϑ in the composite representation in (3.6) being either $\delta_{\{0\}}$ or $\delta_{\mathbf{R}_-}$. Since these functions are CPLQ, it follows from Proposition 3.5 and Example 3.3 that the inclusion in (3.2) becomes equality for $f = \delta_{\mathcal{Q}}$ at any nonzero $\bar{u} \in \mathcal{Q}$ and $\bar{y} \in N_{\mathcal{Q}}(\bar{u})$.

The last case to consider is $\bar{u} = 0$. Take $\bar{y} \in N_{\mathcal{Q}}(\bar{u})$ and observe from [12, Corollary 3.4] that $N_{\mathcal{Q}}$ is proto-differentiable at $\bar{u}$ for $\bar{y}$ and that $DN_{\mathcal{Q}}(\bar{u}, \bar{y})(w) = N_{K_{\mathcal{Q}}(\bar{u},\bar{y})}(w)$ for any $w \in \mathbf{R}^n$ with $K_{\mathcal{Q}}(\bar{u}, \bar{y}) = \mathcal{Q} \cap [\bar{y}]^\perp$. To achieve our goal, we only need to justify the inclusion

$$\widehat{D}N_{\mathcal{Q}}(\bar{u}, \bar{y})(w) \subset \operatorname{cl}\big(DN_{\mathcal{Q}}(\bar{u}, \bar{y})(w) - K_{\mathcal{Q}}(\bar{u}, \bar{y})^*\big) = \operatorname{cl}\big(N_{K_{\mathcal{Q}}(\bar{u},\bar{y})}(w) - K_{\mathcal{Q}}(\bar{u}, \bar{y})^*\big) \quad \text{for all } \; w \in \mathbf{R}^n,$$

since the opposite inclusion always holds according to Proposition 3.2. By [33, Corolalary 11.25(b)], we can equivalently express the latter as

$$\widehat{D}N_{\mathcal{Q}}(\bar{u}, \bar{y})(w) \subset \big(T_{K_{\mathcal{Q}}(\bar{u},\bar{y})}(w) \cap -K_{\mathcal{Q}}(\bar{u}, \bar{y})\big)^*. \tag{3.17}$$

To prove this inclusion, pick $w \in \operatorname{dom} \widehat{D}N_{\mathcal{Q}}(\bar{u}, \bar{y})$, $\eta \in \widehat{D}N_{\mathcal{Q}}(\bar{u}, \bar{y})(w)$ and conclude from (3.4) that $w \in K_{\mathcal{Q}}(\bar{u}, \bar{y})$. To verify (3.17), we must show that $\langle \eta, q\rangle \le 0$ for any $q \in C(w)$, where $C(w) := T_{K_{\mathcal{Q}}(\bar{u},\bar{y})}(w) \cap -K_{\mathcal{Q}}(\bar{u}, \bar{y})$. If $w = 0$, we clearly have

$$C(w) = K_{\mathcal{Q}}(\bar{u}, \bar{y}) \cap -K_{\mathcal{Q}}(\bar{u}, \bar{y}) = (\mathcal{Q} \cap [\bar{y}]^\perp) \cap -(\mathcal{Q} \cap [\bar{y}]^\perp) = \{0\},$$

which proves (3.17). If $w \ne 0$, it follows from $\eta \in \widehat{D}N_{\mathcal{Q}}(\bar{u}, \bar{y})(w)$ that there exist sequences $t_k \searrow 0$, $w^k \to w$, $y^k \in N_{\mathcal{Q}}(w^k)$, and $\widehat{y}^k \to \bar{y}$ with $\widehat{y}^k \in -\mathcal{Q}$ such that $(y^k - \widehat{y}^k)/t_k \to \eta$. We split our verification of (3.17) into three cases depending on the position of $\bar{y}$ in $\mathcal{Q}$:

(i) $\bar{y} \in -\operatorname{int} \mathcal{Q}$. In this case, we have $K_{\mathcal{Q}}(\bar{u}, \bar{y}) = \{0\}$. Since $w \in K_{\mathcal{Q}}(\bar{u}, \bar{y})$, we get $w = 0$, which is not possible.

(ii) $\bar{y} \in -(\operatorname{bd} \mathcal{Q}) \setminus \{0\}$. In this case, since $0 \ne w \in K_{\mathcal{Q}}(\bar{u}, \bar{y})$, it is not hard to see that $K_{\mathcal{Q}}(\bar{u}, \bar{y}) = \{tw|\; t \ge 0\}$. This leads us to

$$C(w) = T_{K_{\mathcal{Q}}(\bar{u},\bar{y})}(w) \cap -K_{\mathcal{Q}}(\bar{u}, \bar{y}) = [w] \cap \{tw|\; t \le 0\} = \{tw|\; t \le 0\}.$$

On the other hand, it follows from $y^k \in N_{\mathcal{Q}}(w^k)$ and $\widehat{y}^k \in -\mathcal{Q}$ that

$$\langle \frac{y^k - \widehat{y}^k}{t_k}, w^k\rangle = -\langle \frac{\widehat{y}^k}{t_k}, w^k\rangle \ge 0,$$

which in turn results in $\langle \eta, w\rangle \ge 0$. Pick now $q \in C(w)$. So, $q = tw$ for some $t \le 0$. Combining these shows $\langle \eta, q\rangle \le 0$, which again proves (3.17).

(iii) $\bar{y} = 0$. In this case, we have $K_{\mathcal{Q}}(\bar{u}, \bar{y}) = \mathcal{Q}$. Recall that $w \in K_{\mathcal{Q}}(\bar{u}, \bar{y})$. If $w \in \operatorname{int} \mathcal{Q}$, it follows from $w^k \to w$ that $w^k \in \operatorname{int} \mathcal{Q}$ for any k sufficiently large and therefore $y^k = 0$ for any such k. Thus, $-\widehat{y}^k/t_k \to \eta$, which, together with $\widehat{y}^k \in -\mathcal{Q}$, yields $\eta \in \mathcal{Q}$. Note that for such w, we have

$$C(w) = T_{K_{\mathcal{Q}}(\bar{u},\bar{y})}(w) \cap -K_{\mathcal{Q}}(\bar{u},\bar{y}) = \mathbf{R}^n \cap -\mathcal{Q} = -\mathcal{Q}.$$

Thus, we obtain $\eta \in C(w)^*$, which again proves (3.17). If $w \in \operatorname{bd} \mathcal{Q}$, we conclude that

$$C(w) = T_{K_{\mathcal{Q}}(\bar{u},\bar{y})}(w) \cap -K_{\mathcal{Q}}(\bar{u},\bar{y}) = T_{\mathcal{Q}}(w) \cap -\mathcal{Q} = \{tw \mid t \le 0\}.$$

Using the same argument as case (ii) shows that $\eta \in C(w)^*$ and hence proves (3.17). Combining these cases shows that the inclusion in (3.2) holds as equality for $f = \delta_{\mathcal{Q}}$ when $\bar{u} = 0$.

Remark 3.7. Note that CPLQ functions may not be $\mathcal{C}^2$-decomposable. To delineate this, it is important to mention that most functions, enjoying the composite representation in (3.6), automatically satisfy the nondegenracy condition in (3.13); see [35, Examples 2.1 and 2.3]. Assume that a function $g : \mathbf{Y} \to \overline{\mathbf{R}}$ is $\mathcal{C}^2$-decomposable at $\bar{u}$, that the nondegenracy condition in (3.13) holds, and that $\bar{y} \in \partial g(\bar{u})$. Employing now Theorem 7.2 and (7.2) tells us that the second subderivaive of g at $\bar{u}$ for $\bar{y}$ has a single quadratic piece, namely $\langle \bar{\mu}, \nabla^2 \Xi(\bar{u})(w,w)\rangle$, whenever $\nabla \Xi(\bar{u}) w \in K_{\vartheta}(\Xi(\bar{u}), \bar{y})$. If g were a CPLQ, it would follow from (4.2) that $\mathrm{d}^2 g(\bar{u}, \bar{y})$ may have several different quadratic pieces depending on how many polyhedral convex sets C_i exist in the representation of g; see the beginning of Example 3.3 for the definition of C_i. That contradicts our earlier observation about $\mathrm{d}^2 g(\bar{u}, \bar{y})$ and demonstrates that CPLQ functions cannot be $\mathcal{C}^2$-decomposable in general.

We are now going to establish an error bound for the KKT system in (2.2) associated with the composite optimization problem (1.1). Such a result was perviously established under the second-order sufficient condition (SOSC) for different classes of constrained optimization problems in [26, Theorem 5.9]. For a solution $(\bar{x}, \bar{y})$ to the KKT system (2.2), the latter result motivates us to consider the SOSC

$$\begin{cases} \langle \nabla^2_{xx} L(\bar{x},\bar{y}) w, w\rangle + \mathrm{d}^2 g(\Phi(\bar{x}),\bar{y})(\nabla \Phi(\bar{x}) w) > 0 \quad \text{for all } \ w \in \mathscr{D} \setminus \{0\}, \\ \text{where } \ \mathscr{D} := K_{\Theta}(\bar{x}, -\nabla_x L(\bar{x},\bar{y})) \cap \big\{ w \in \mathbf{X} \,\big|\, \nabla \Phi(\bar{x}) w \in K_g(\Phi(\bar{x}),\bar{y}) \big\}, \end{cases} \tag{3.18}$$

and investigate whether our desired error bound can be derived under this condition. Note that the SOSC in (3.18) boils down to the classical second-order sufficient condition formulated for different classes of constrained optimization problems including NLPs, second-order cone programming, and SDPs; see [23, Section 6] for a detailed discussion about the SOSC. For the composite programming problems in (1.1) with g therein being CPLQ, one can find recent developments about the SOSC in [13, 34].

It is well-known that the SOSC alone does not ensure the error bound for the KKT system in (2.2); see the discussion after [26, Theorem 5.9]. What is needed further is calmness of a multiplier mapping associated with the composite problem in (1.1), which is defined as the one for g in (3.8). Given $\bar{x} \in \mathbf{X}$ and ψ taken from (1.1), the *multiplier mapping* $M_{\bar{x},\psi} \colon \mathbf{X} \times \mathbf{Y} \rightrightarrows \mathbf{Y}$, associated with the canonically perturbed KKT system (2.2), is defined by

$$M_{\bar{x},\psi}(v,w) := \big\{ y \in \mathbf{Y} \,\big|\, v \in \nabla_x L(\bar{x}, y) + N_{\Theta}(\bar{x}),\ y \in \partial g\big(\Phi(\bar{x}) + w\big)\big\}, \quad (v,w) \in \mathbf{X} \times \mathbf{Y}. \tag{3.19}$$

It is easy to see that if $\bar{x}$ is a local minimum of the composite problem in (1.1), $M_{\bar{x},\psi}(0,0)$ reduces to the set of Lagrange multipliers associated with $\bar{x}$. Note also that for any $\bar{y} \in M_{\bar{x},\psi}(0,0)$, the pair $(\bar{x}, \bar{y})$ is a solution to the KKT system in (2.2). While, throughout this paper, we always assume that a Lagrange multiplier exists, it is well-known in general that the existence of such a Lagrange multiplier requires a *constraint qualification*. Getting into the question of which constraint qualification to choose is unnecessary for our developments in this paper, however.

Given a solution $(\bar{x}, \bar{y})$ to the KKT system (2.2), the calmness of the multiplier mapping $M_{\bar{x},\psi}$ in (3.19) at $(0, 0)$ for $\bar{y}$ amounts to the existence of a positive constant τ and neighborhoods U of $(0, 0)$ and V of $\bar{y}$ such that

$$M_{\bar{x},\psi}(v, w) \cap V \subset M_{\bar{x},\psi}(0, 0) + \tau\|(w, v)\|\mathbb{B} \quad \text{for all } \ (v, w) \in U.$$

It follows from [9, Theorem 3H.3] that this property is equivalent to the metric subregualrity of its inverse mapping $M^{-1}_{\bar{x},\psi}$ at $\bar{y}$ for $(0, 0)$, which means that there exist a positive constant ℓ and a neighborhood V of $\bar{y}$ for which the estimate

$$\operatorname{dist}\big(y, M_{\bar{x},\psi}(0, 0)\big) \le \ell\left(\operatorname{dist}(-\nabla_x L(\bar{x}, y), N_\Theta(\bar{x})) + \operatorname{dist}(\Phi(\bar{x}), (\partial g)^{-1}(y))\right) \quad \text{for all } y \in V \quad (3.20)$$

holds. Below, we record some conditions under which the clamness of the multiplier mapping $M_{\bar{x},\psi}$ is satisfied.

Proposition 3.8. *Assume that* $(\bar{x}, \bar{y})$ *is a solution to the KKT system* (2.2)*. Then the multiplier mapping* $M_{\bar{x},\psi}$ *is calm at* $(0, 0)$ *for* $\bar{y}$ *if one of the following conditions holds:*

(a) *the convex function* g *in* (1.1) *is CPLQ;*

(b) *the subgradient mapping* ∂g *is calm at* $\Phi(\bar{x})$ *for* $\bar{y}$ *and there exists* $\tilde{y} \in M_{\bar{x},\psi}(0, 0)$ *such that* $-\nabla_x L(\bar{x}, \tilde{y}) \in \operatorname{ri} N_\Theta(\bar{x})$ *and* $\tilde{y} \in \operatorname{ri} \partial g(\Phi(\bar{x}))$*.*

Proof. Observe first that the set $M_{\bar{x},\psi}(0, 0)$ can be equivalently written as

$$M_{\bar{x},\psi}(0, 0) = \Omega \cap \partial g(\Phi(\bar{x})) \quad \text{with} \quad \Omega := \big\{y \in \mathbf{Y} \;\big|\; 0 \in \nabla_x L(\bar{x}, y) + N_\Theta(\bar{x})\big\}.$$

If (a) holds, $\partial g(\Phi(\bar{x}))$ is a polyhedral convex set. By assumption, Ω is a polyhedral convex set as well. It follows then from [15, Theorem 8.35] that there is a constant $\rho \ge 0$ such that for any $y \in Y$ the estimate

$$\operatorname{dist}(y, M_{\bar{x},\psi}(0, 0)) \le \rho\big(\operatorname{dist}(y, \Omega) + \operatorname{dist}(y, \partial g(\Phi(\bar{x}))\big)$$

holds. If (b) holds, the same estimate holds for any y sufficiently close to $\bar{y}$ due to the existence of $\tilde{y}$, which can be equivalently translated as $\operatorname{ri}\Omega \cap \operatorname{ri}\partial g(\Phi(\bar{x})) \ne \emptyset$; see [3, Corollary 3]. Using the classical Hoffman lemma (cf. [9, Lemma 3C.4]) gives us a constant $\rho' \ge 0$ such that the estimate

$$\operatorname{dist}(y, \Omega) \le \rho' \operatorname{dist}\big(-\nabla_x L(\bar{x}, y), N_\Theta(\bar{x})\big)$$

is satisfied for any $y \in \mathbf{Y}$. If g is CPLQ, it follows from the proof of [33, Theorem 11.14(b)] that $\operatorname{gph}\partial g$ is a union of finitely many polyhedral convex sets. So, we conclude from [29, Proposition 1] (see also [9, Theorem 3D.1]) that the subgradient mapping ∂g is calm at $\Phi(\bar{x})$ for $\bar{y}$. Thus, if either (a) or (b) holds, $(\partial g)^{-1}$ is metrically subregular at $\bar{y}$ for $\Phi(\bar{x})$ (cf. [9, Theorem 3H.3]), meaning that there are a positive constant ρ'' and a neighborhood V of $\bar{y}$ for which the estimate

$$\operatorname{dist}\big(y, \partial g(\Phi(\bar{x}))\big) \le \rho'' \operatorname{dist}\big(\Phi(\bar{x}), (\partial g)^{-1}(y)\big)$$

holds for any $y \in V$. Combining all these estimates proves the calmness of the multiplier mapping $M_{\bar{x},\psi}$ at $(0, 0)$ for $\bar{y}$. □

It is worth mentioning that the existence of $\tilde{y}$ in Proposition 3.8(b) boils down to the classical *strict complementarity* assumption in case $\Theta = \mathbf{X}$ in the composite problem in (1.1), which reads as the existence of a multiplier $\tilde{y}$ such that $\nabla_x L(\bar{x}, \tilde{y}) = 0$ and $\tilde{y} \in \operatorname{ri}\partial g(\Phi(\bar{x}))$. Furthermore, according to [2, Theorem 3.3], the calmness of the subgradient mapping ∂g in Proposition 3.8(b) is equivalent to the conjugate function g^* satisfying the quadratic growth condition

$$g^*(y) \geq g^*(\bar{y}) + \langle \Phi(\bar{x}), y - \bar{y} \rangle + \kappa \, \mathrm{dist}\big(y, \partial g(\Phi(\bar{x}))\big) \quad \text{whenever } \; y \in V,$$

where κ is a positive constant and V is a neighborhood of $\bar{y}$. It is known that this calmness condition holds for the subgradient mapping of CPLQ functions (cf. [29, Proposition 1]) as well as the $\mathcal{C}^2$-decomposable functions under the condition (3.9) (cf. Propsoition 3.4(e)).

To proceed, define the *residual* function $r\colon \mathbf{X} \times \mathbf{Y} \to \mathbf{R}$ of the KKT system (2.2) by

$$r(x, y) := \mathrm{dist}(-\nabla_x L(x, y), N_\Theta(x)) + \|\Phi(x) - \mathrm{prox}_g\big(\Phi(x) + y\big)\|, \quad (x, y) \in \mathbf{X} \times \mathbf{Y}. \tag{3.21}$$

It is easy to see that a pair (x, y) is a solution to the KKT system in (2.2) if and only if $r(x, y) = 0$. The residual function r, indeed, measures the violation of the KKT system in (2.2) for a given pair (x, y) and plays an important role as a surrogate for the tolerance parameter ϵ_k in (1.5).

Theorem 3.9. *Assume that $(\bar{x}, \bar{y})$ is a solution to the KKT system (2.2) for which the SOSC in (3.18) holds. Assume further that one of the following conditions holds:*

(a) *the convex function g in (1.1) is $\mathcal{C}^2$-decomposable and the nondegeneracy condition in (3.13) is satisfied for $\bar{u} = \Phi(\bar{x})$;*

(b) *the convex function g in (1.1) is twice epi-differentiable at $\Phi(\bar{x})$ for $\bar{y}$ and the equality in (3.14) holds for $\bar{u} = \Phi(\bar{x})$ and $\bar{y}$.*

If the multiplier mapping $M_{\bar{x},\psi}$ in (3.19) is calm at $(0, 0) \in \mathbf{X} \times \mathbf{Y}$ for $\bar{y}$, then there exist constants $\gamma > 0$ and $\kappa \geq 0$ such that

$$\|x - \bar{x}\| + \mathrm{dist}\big(y, M_{\bar{x},\psi}(0, 0)\big) \leq \kappa \, r(x, y) \quad \textit{for all} \quad (x, y) \in \mathbb{B}_\gamma(\bar{x}, \bar{y}). \tag{3.22}$$

Proof. We begin by showing that that there exists a neighborhood U of $(\bar{x}, \bar{y})$ such that

$$\|x - \bar{x}\| = O\big(r(x, y)\big) \quad \text{for all} \quad (x, y) \in U. \tag{3.23}$$

If this claim fails, we find a sequence $(x^k, y^k) \to (\bar{x}, \bar{y})$ such that

$$t_k := \|x^k - \bar{x}\| > k \, r(x^k, y^k) \geq 0 \quad \text{for all } \; k \in \mathbb{N},$$

meaning that $r(x^k, y^k) = o(t_k)$. Thus, by (3.21), we find $\alpha^k \in \mathbf{X}$ and $\beta^k \in \mathbf{Y}$ such that $\alpha^k = o(t_k)$ and $\beta^k = o(t_k)$, where

$$-\nabla_x L(x^k, y^k) + \alpha^k \in N_\Theta(x^k) \quad \text{and} \quad \beta^k = \Phi\big(x^k\big) - \mathrm{prox}_g\big(\Phi(x^k) + y^k\big). \tag{3.24}$$

It then follows from the first identity in (2.8) that

$$y^k + \beta^k \in \partial g(\Phi(x^k) - \beta^k). \tag{3.25}$$

The calmness of multiplier mapping $M_{\bar{x},\psi}$ at $(0, 0)$ for $\bar{y}$ implies that there exist a positive constant ℓ and a neighborhood V of $\bar{y}$ for which the estimate in (3.20) holds. This allows us to conclude for any k sufficiently large that

$$\mathrm{dist}\big(y^k + \beta^k, M_{\bar{x},\psi}(0, 0)\big) \leq \ell\big(\mathrm{dist}(-\nabla_x L(\bar{x}, y^k + \beta^k), N_\Theta(\bar{x})) + \mathrm{dist}(\Phi(\bar{x}), (\partial g)^{-1}(y^k + \beta^k))\big).$$

Recall that Θ is a polyhedral set and that $x^k \to \bar{x}$ as $k \to \infty$. We then deduce that $N_\Theta(x^k) \subset N_\Theta(\bar{x})$ for k sufficiently large and get the estimates

$$\begin{aligned}\text{dist}(-\nabla_x L(\bar{x}, y^k + \beta^k), N_\Theta(\bar{x})) &\leq \text{dist}(-\nabla_x L(\bar{x}, y^k + \beta^k), N_\Theta(x^k)) \\ &\leq \| - \nabla_x L(\bar{x}, y^k + \beta^k) - (-\nabla_x L(x^k, y^k) + \alpha^k)\| \\ &\leq \|\nabla \varphi(x^k) - \nabla \varphi(\bar{x})\| + \|(\nabla \Phi(x^k) - \nabla \Phi(\bar{x}))^* y^k\| \\ &\quad + \|\nabla \Phi(\bar{x})^* \beta^k\| + \|\alpha^k\| = O(t_k),\end{aligned}$$

where the last equality comes from the Lipschitz continuity of mappings $\nabla\varphi$ and $\nabla\Phi$ around $\bar{x}$ and the estimates in (3.24). Moreover, by (3.25), we have $\Phi(x^k) - \beta^k \in (\partial g)^{-1}(\lambda^k + \beta^k)$, which coupled with the Lipschitz continuity of Φ around $\bar{x}$ and $\beta^k = o(t_k)$ leads us to

$$\text{dist}\big(\Phi(\bar{x}), (\partial g)^{-1}(\lambda^k + \beta^k)\big) \leq \|\Phi(x^k) - \beta^k - \Phi(\bar{x})\| = O(t_k)$$

for any k sufficiently large. Combining these estimates demonstrates that

$$\text{dist}\big(y^k + \beta^k, M_{\bar{x},\psi}(0,0)\big) = O(t_k)$$

for any k sufficiently large. Since $M_{\bar{x},\psi}(0,0)$ is a closed convex set, set $\widehat{y}^k := P_{M_{\bar{x},\psi}(0,0)}(y^k)$. This, the above estimate, and $\beta^k = o(t_k)$ allow us to obtain $y^k - \widehat{y}^k = O(t_k)$. Passing to subsequences if necessary, we can assume that

$$\frac{x^k - \bar{x}}{t_k} \to w \neq 0 \quad \text{and} \quad \frac{y^k - \widehat{y}^k}{t_k} \to \eta \;\text{ as }\; k \to \infty. \tag{3.26}$$

Recall from (2.2) that $-\nabla_x L(\bar{x}, \bar{y}) \in N_\Theta(\bar{x})$. This, coupled with the first inclusion in (3.24) and the reduction lemma for polyhedral convex sets (cf. [9, Lemma 2E.4]), allows us to conclude for any k sufficiently large that

$$-\big(\nabla_x L(x^k, y^k) - \nabla_x L(\bar{x}, \bar{y})\big) + \alpha^k \in N_{K_\Theta(\bar{x}, -\nabla_x L(\bar{x},\bar{y}))}(x^k - \bar{x}).$$

By the definition of $\widehat{y}^k$, we have $-\nabla_x L(\bar{x}, \widehat{y}^k) \in N_\Theta(\bar{x})$. Applying again the reduction lemma from [9, Lemma 2E.4] brings us to

$$-\big(\nabla_x L(\bar{x}, \widehat{y}^k) - \nabla_x L(\bar{x}, \bar{y})\big) \in N_{K_\Theta(\bar{x}, -\nabla_x L(\bar{x},\bar{y}))}(0) = K_\Theta(\bar{x}, -\nabla_x L(\bar{x}, \bar{y}))^*$$

for any k sufficiently large. On the other hand, we have

$$\begin{aligned}\nabla_x L(x^k, y^k) - \nabla_x L(\bar{x}, \bar{y}) &= \nabla_x L(x^k, \bar{y}) - \nabla_x L(\bar{x}, \bar{y}) + \nabla \Phi(x^k)^*(y^k - \bar{y}) \\ &= \nabla^2_{xx} L(\bar{x}, \bar{y})(x^k - \bar{x}) + \nabla \Phi(x^k)^*(y^k - \widehat{y}^k) + \nabla \Phi(\bar{x})^*(\widehat{y}^k - \bar{y}) + o(t_k),\end{aligned}$$

and

$$\nabla_x L(\bar{x}, \widehat{y}^k) - \nabla_x L(\bar{x}, \bar{y}) = \nabla \Phi(\bar{x})^*(\widehat{y}^k - \bar{y}).$$

Combining these and remembering that $\alpha^k = o(t_k)$ lead us to the inclusion

$$\begin{aligned}&\nabla^2_{xx} L(\bar{x}, \bar{y})(\frac{x^k - \bar{x}}{t_k}) + \nabla \Phi(x^k)^*(\frac{y^k - \widehat{y}^k}{t_k}) + \frac{o(t_k)}{t_k} \\ &\qquad \in K_\Theta(\bar{x}, -\nabla_x L(\bar{x}, \bar{\lambda}))^* - N_{K_\Theta(\bar{x}, -\nabla_x L(\bar{x},\bar{y}))}(\frac{x^k - \bar{x}}{t_k}) \\ &\qquad \subset K_\Theta(\bar{x}, -\nabla_x L(\bar{x}, \bar{\lambda}))^* - N_{K_\Theta(\bar{x}, -\nabla_x L(\bar{x},\bar{y}))}(w)\end{aligned}$$

for any k sufficiently large, where the last inclusion results from the first limit in (3.26) and $K_\Theta(\bar{x}, -\nabla_x L(\bar{x}, \bar{\lambda}))$ being a polyhedral convex set. The right-hand side of this inclusion is the sum of two polyhedral convex sets, which is again a polyhedral convex set and hence is closed. Passing to the limit brings us to

$$\nabla^2_{xx} L(\bar{x}, \bar{y})w + \nabla\Phi(\bar{x})^*\eta \in K_\Theta(\bar{x}, -\nabla_x L(\bar{x}, \bar{y}))^* - N_{K_\Theta(\bar{x}, -\nabla_x L(\bar{x}, \bar{y}))}(w).$$

This implies that $w \in K_\Theta(\bar{x}, -\nabla_x L(\bar{x}, \bar{y}))$ and that there are $\nu_1 \in K_\Theta(\bar{x}, -\nabla_x L(\bar{x}, \bar{y}))^*$ and $\nu_2 \in N_{K_\Theta(\bar{x}, -\nabla_x L(\bar{x}, \bar{y}))}(w) = K_\Theta(\bar{x}, -\nabla_x L(\bar{x}, \bar{y}))^* \cap [w]^\perp$ such that $\nabla^2_{xx} L(\bar{x}, \bar{y})w + \nabla\Phi(\bar{x})^*\eta = \nu_1 - \nu_2$, which yields

$$\langle \nabla^2_{xx} L(\bar{x}, \bar{y})w, w\rangle + \langle \eta, \nabla\Phi(\bar{x})w\rangle = \langle \nu_1, w\rangle - \langle \nu_2, w\rangle \le 0. \tag{3.27}$$

We claim now $\nabla\Phi(\bar{x})w \in K_g(\Phi(\bar{x}), \bar{y})$. It follows from the convexity of g, $\bar{y} \in \partial g(\Phi(\bar{x}))$ and [33, Theorem 8.30] that $\langle \bar{y}, \nabla\Phi(\bar{x})w\rangle \le \mathrm{d}g(\Phi(\bar{x}))(\nabla\Phi(\bar{x})w)$. To prove the opposite inequality, we deduce from (3.25) that

$$\begin{aligned}\langle y^k + \beta^k, \Phi(x^k) - \Phi(\bar{x}) - \beta^k\rangle &\ge g(\Phi(x^k) - \beta^k) - g(\Phi(\bar{x})) \\ &= g\big(\Phi(\bar{x}) + t_k \frac{\Phi(x^k) - \Phi(\bar{x}) - \beta^k}{t_k}\big) - g(\Phi(\bar{x})).\end{aligned}$$

Dividing both sides by t_k and passing then to the limit result in $\langle \bar{y}, \nabla\Phi(\bar{x})w\rangle \ge \mathrm{d}g(\Phi(\bar{x}))(\nabla\Phi(\bar{x})w)$. Thus we get $\langle \bar{y}, \nabla\Phi(\bar{x})w\rangle = \mathrm{d}g(\Phi(\bar{x}))(\nabla\Phi(\bar{x})w)$, which proves our claim. To arrive at a contradiction with the SOSC in (3.18), we need to prove that the second subderivative of g satisfies

$$\langle \eta, \nabla\Phi(\bar{x})w\rangle \ge \mathrm{d}^2 g(\Phi(\bar{x}), \bar{y})(\nabla\Phi(\bar{x})w). \tag{3.28}$$

Assume first that (b) holds, meaning that g satisfies in (3.14). By (3.26) and $\beta^k = o(t_k)$, we have

$$\big(\frac{\Phi(x^k) - \beta^k - \Phi(\bar{x})}{t_k}, \frac{y^k + \beta^k - \widehat{y}^k}{t_k}\big) \to (\nabla\Phi(\bar{x})w, \eta).$$

This, together with (3.25), the fact $\widehat{y}^k \in \partial g(\Phi(\bar{x}))$ and Definition 3.1, tells us that $\eta \in \widehat{D}(\partial g)(\Phi(\bar{x}), \bar{y})(\nabla\Phi(\bar{x})w)$. By (3.14), we find sequences $\{\eta_1^k\}_{k\in\mathbb{N}} \subset D(\partial g)(\Phi(\bar{x}), \bar{y})(\nabla\Phi(\bar{x})w)$ and $\{\eta_2^k\}_{k\in\mathbb{N}} \subset K_g(\Phi(\bar{x}), \bar{y})^*$ for which we have $\eta_1^k - \eta_2^k \to \eta$ as $k \to \infty$. It follows from [33, Theorem 13.40] and [8, Lemma 3.6] that $\mathrm{d}^2 g(\Phi(\bar{x}), \bar{y})(\nabla\Phi(\bar{x})w) = \langle \eta_1^k, \nabla\Phi(\bar{x})w\rangle$ and from $\nabla\Phi(\bar{x})w \in K_g(\Phi(\bar{x}), \bar{y})$ that $\langle \eta_2^k, \nabla\Phi(\bar{x})w\rangle \le 0$ for all $k \in \mathbb{N}$. So, we get

$$\langle \eta_1^k - \eta_2^k, \nabla\Phi(\bar{x})w\rangle \ge \mathrm{d}^2 g(\Phi(\bar{x}), \bar{y})(\nabla\Phi(\bar{x})w).$$

Passing to the limit then proves (3.28) when (b) holds. Assume now that (a) is satisfied. It follows from the nondegeneracy condition in (3.13) and the inclusions in (7.5) and (3.15) that the BCQ condition in (3.9) holds. This tells us that there is a neighborhood $\mathcal{O}$ of $\bar{u} = \Phi(\bar{x})$ such that for any $u \in \mathcal{O} \cap \operatorname{dom}\vartheta$, we have

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

So, for any k sufficiently large, we conclude from the chain rule for subdifferentials in [33, Theorem 10.6], $\widehat{y}^k \in \partial g(\Phi(\bar{x}))$, and (3.25) that there are $\widehat{\mu}^k \in \partial\vartheta(\Xi(\bar{u}))$ and $\mu^k \in \partial\vartheta(\Xi(u^k))$ with $u^k = \Phi(x^k) - \beta^k$ such that

$$\widehat{y}^k = \nabla\Xi(\bar{u})^*\widehat{\mu}^k \quad \text{and} \quad y^k + \beta^k = \nabla\Xi(u^k)^*\mu^k.$$

These allow us to obtain

$$\big\langle \frac{y^k + \beta^k - \widehat{y}^k}{t_k}, \frac{u^k - \bar{u}}{t_k}\big\rangle = \big\langle \frac{\nabla\Xi(\bar{u})^*(\mu^k - \widehat{\mu}^k)}{t_k}, \frac{u^k - \bar{u}}{t_k}\big\rangle + \big\langle \frac{(\nabla\Xi(u^k) - \nabla\Xi(\bar{u}))^*\mu^k}{t_k}, \frac{u^k - \bar{u}}{t_k}\big\rangle.$$

The first term in the right-hand side of the above equality can be estimated by

$$\begin{aligned}\Big\langle \frac{\nabla\Xi(\bar u)^*(\mu^k-\widehat{\mu}^k)}{t_k}, \frac{u^k-\bar u}{t_k}\Big\rangle &= \Big\langle \frac{\mu^k-\widehat{\mu}^k}{t_k}, \frac{\nabla\Xi(\bar u)(u^k-\bar u)}{t_k}\Big\rangle\\ &= \Big\langle \frac{\mu^k-\widehat{\mu}^k}{t_k}, \frac{\Xi(u^k)-\Xi(\bar u)}{t_k}\Big\rangle + \Big\langle \frac{\mu^k-\widehat{\mu}^k}{t_k}, \frac{o(\|u^k-\bar u\|)}{t_k}\Big\rangle\\ &\geq \Big\langle \frac{\mu^k-\widehat{\mu}^k}{t_k}, \frac{o(t_k)}{t_k}\Big\rangle,\end{aligned}$$

where the last inequality results from the monotonicity of the subgradient mapping $\partial\vartheta$ and the fact that $u^k-\bar u = O(t_k)$. Also, we have

$$\Big\langle \frac{(\nabla\Xi(u^k)-\nabla\Xi(\bar u))^*\mu^k}{t_k}, \frac{u^k-\bar u}{t_k}\Big\rangle = \Big\langle \frac{(\nabla^2\Xi(\bar u)(u^k-\bar u)+o(t_k))^*\mu^k}{t_k}, \frac{u^k-\bar u}{t_k}\Big\rangle.$$

Combining these results in the estimate

$$\Big\langle \frac{y^k+\beta^k-\widehat{y}^k}{t_k}, \frac{u^k-\bar u}{t_k}\Big\rangle \geq \Big\langle \frac{\mu^k-\widehat{\mu}^k}{t_k}, \frac{o(t_k)}{t_k}\Big\rangle + \Big\langle \frac{(\nabla^2\Xi(\bar u)(u^k-\bar u)+o(t_k))^*\mu^k}{t_k}, \frac{u^k-\bar u}{t_k}\Big\rangle. \quad (3.29)$$

By (3.26) and $\beta^k = o(t_k)$, we can conclude that the left-hand side in (3.29) converges to $\langle \eta, \nabla\Phi(\bar x)w\rangle$. Moreover, a similar argument as the one after (3.16) and the fact that $y^k-\widehat{y}^k = O(t_k)$ tell us that $\mu^k-\widehat{\mu}^k = O(t_k)$. Thus, the first term on the right-hand side in (3.29) converges to 0. Finally, the nondegeneracy condition in (3.13) and the inclusion in (7.5) implies that the dual condition in (3.10) is satisfied. By Theorem 7.1(d), we obtain that the sequence $\{\mu^k\}_{k\in\mathbb{N}}$ converges to $\bar\mu$ as $k\to\infty$, where $\bar\mu$ is the unique element in $M_{\bar u,g}(\bar y,0)$ with $M_{\bar u,g}(\bar y,0)$ taken from (3.8); see Theorem 7.1. Passing to the limit in (3.29) then shows that

$$\langle \eta, \nabla\Phi(\bar x)w\rangle \geq \big\langle \bar\mu, \nabla^2\Xi(\bar u)(\nabla\Phi(\bar x)w, \nabla\Phi(\bar x)w)\big\rangle.$$

Recall that $\nabla\Phi(\bar x)w \in K_g(\Phi(\bar x),\bar y)$, which together with (3.11) brings us to $\nabla\Xi(\bar u)(\nabla\Phi(\bar x)w) \in K_\vartheta(\Xi(\bar u),\bar\mu)$. Employing now the established formula for the second subderivative of g in Theorem 7.1(a) and using (7.2) result in

$$\big\langle \bar\mu, \nabla^2\Xi(\bar u)(\nabla\Phi(\bar x)w, \nabla\Phi(\bar x)w)\big\rangle = \mathrm{d}^2 g(\bar u,\bar y)(\nabla\Phi(\bar x)w).$$

Combining these two estimates proves (3.28) when (a) holds. By (3.28) and (3.27), we obtain

$$\langle \nabla^2_{xx}L(\bar x,\bar y)w,w\rangle + \mathrm{d}^2 g(\bar u,\bar y)(\nabla\Phi(\bar x)w) \leq \langle \nabla^2_{xx}L(\bar x,\bar y)w,w\rangle + \langle \eta, \nabla\Phi(\bar x)w\rangle \leq 0,$$

which contradicts the SOSC in (3.18), since $w\neq 0$ and $w\in\mathscr{D}$ with $\mathscr{D}$ taken from (3.18). This proves the estimate in (3.23).

To justify (3.22), we need to show that shrinking the neighborhood U from (3.23), if necessary, we have

$$\mathrm{dist}\big(y, M_{\bar x,\psi}(0,0)\big) = O\big(r(x,y)\big) \quad \text{for all} \quad (x,y)\in U. \quad (3.30)$$

To this end, pick (x,y) satisfying (3.23) and set $z := \mathrm{prox}_g(\Phi(x)+y)-\Phi(x)$, which results in $y-z\in\partial g(\Phi(x)+z)$ or $\Phi(x)+z\in(\partial g)^{-1}(y-z)$. It is easy to see that $z\to 0$ as $(x,y)\to(\bar x,\bar y)$. Thus, shrinking the neighborhood U from (3.23), if necessary, we can assume without loss of generality that $y-z\in V$, where V is taken from (3.20), and hence conclude from the estimate in (3.20) and the polyhedrality of Θ that

$$\begin{aligned}\mathrm{dist}(y-z, M_{\bar x,\psi}(0,0)) &\leq \ell\,\big(\mathrm{dist}(-\nabla_x L(\bar x,y-z), N_\Theta(\bar x)) + \mathrm{dist}(\Phi(\bar x),(\partial g)^{-1}(y-z))\big)\\ &\leq \ell\,\big(\|\nabla_x L(x,y)-\nabla_x L(\bar x,y-z)\| + \mathrm{dist}(-\nabla_x L(x,y); N_\Theta(x)) + \|\Phi(x)+z-\Phi(\bar x)\|\big)\\ &\leq O\big(\|x-\bar x\|+\|z\|\big) + \ell\,\mathrm{dist}(-\nabla_x L(x,y), N_\Theta(x))\\ &= O\big(r(x,y)\big),\end{aligned}$$

where the last equality comes from (3.23) and the definition of the residual function r. Since the distance function is Lipschitz continuous, we get

$$\operatorname{dist}\big(y, M_{\bar{x},\psi}(0,0)\big) - \operatorname{dist}\big(y-z, M_{\bar{x},\psi}(0,0)\big) = O(\|z\|) = O(\|\operatorname{prox}_g(\Phi(x)+y) - \Phi(x)\|) = O\big(r(x,y)\big).$$

Combining these estimates, we arrive at (3.30). This completes the proof. $\square$

The error bound in Theorem 3.9 was obtained previously in [26, Theorem 5.9] for $\mathcal{C}^2$-cone reducible constrained optimization problems using a rather lengthy reduction process, a path that was not followed here. It was extended for CPLQ composite optimization problems in [34, Theorem 3.6]. Theorem 3.9 provides a general property – condition (b) in this theorem– under which the error bound can be achieved. The main question that remains open here is whether $\mathcal{C}^2$-decomposable functions satisfy in (3.14). It is worth reiterating here that $\mathcal{C}^2$-decomposable functions do not contain CPLQ functions (see Remark 3.7) and hence condition (b) in Theorem 3.9 can't be covered by condition (a) therein.

4 Semi-Stability of Second Subderivatives

In parallel with the notion of semi-strict graphical derivative in Definition 3.1, we define in this section a semi-strict counterpart of the second subderivative of functions and use it to introduce a new regularity condition associated with the latter construction that plays a crucial role in our local convergence analysis of the ALM.

Definition 4.1. Assume that $f\colon \mathbf{X} \to \overline{\mathbf{R}}$, $\bar{x} \in \mathbf{X}$ with $f(\bar{x})$ finite, and $\bar{v} \in \partial f(\bar{x})$.

(a) The semi-strict second subderivative of f at $\bar{x}$ for $\bar{v}$ is defined by

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

(b) The second subderivative of f at $\bar{x}$ for $\bar{v}$ is said to be semi-stable if $\widehat{\mathrm{d}}^2 f(\bar{x},\bar{v})(w) = \mathrm{d}^2 f(\bar{x},\bar{v})(w)$ for any $w \in K_f(\bar{x},\bar{v})$.

As demonstrated below, the semistability of second subderivatives is satisfied for various classes of functions, important for modeling constrained and composite optimization problems. Note that the inequality $\widehat{\mathrm{d}}^2 f(\bar{x},\bar{v})(w) \le \mathrm{d}^2 f(\bar{x},\bar{v})(w)$ always holds for any $w \in \mathbf{X}$. Thus justifying semi-stability of second subderivatives reduces to investigating whether or not the latter inequality holds as equality.

Proposition 4.2. *Assume that $f\colon \mathbf{X} \to \overline{\mathbf{R}}$, $\bar{x} \in \mathbf{X}$ with $f(\bar{x})$ finite, and $\bar{v} \in \widehat{\partial} f(\bar{x})$. Then we have*

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

Proof. Take $w \in \operatorname{dom}\widehat{\mathrm{d}}^2 f(\bar{x},\bar{v})$ and conclude that $\widehat{\mathrm{d}}^2 f(\bar{x},\bar{v})(w) < \infty$. Since we obtain for any $t > 0$ that

$$\Delta_t^2 f(\bar{x},v)(w') = \frac{\Delta_t f(\bar{x})(w') - \langle v, w'\rangle}{\frac{1}{2}t} \quad \text{with } \Delta_t f(\bar{x})(w') := \frac{f(\bar{x}+tw') - f(\bar{x})}{t},$$

it follows from the definition of $\widehat{\mathrm{d}}^2 f(\bar{x},\bar{v})(w)$ that

$$\alpha := \liminf_{\substack{t\searrow 0,\, w'\to w\\ v\to\bar{v},\, v\in\partial f(\bar{x})}} \Delta_t f(\bar{x})(w') - \langle v, w'\rangle < \infty.$$

If $\alpha > 0$, we would get $\widehat{\mathrm{d}}^2 f(\bar{x}, \bar{v})(w) = \infty$, a contradiction. Thus, we must have $\alpha \le 0$, which clearly tells us that $\mathrm{d}f(\bar{x})(w) \le \langle \bar{v}, w \rangle$. Noting that the opposite inequality always holds because of $\bar{v} \in \widehat{\partial} f(\bar{x})$ (cf. [33, Exercise 8.4]), we arrive at the claimed inclusion. $\square$

Which classes of functions enjoy the semi-stability of second subserivatives? To answer this question, we begin with the following observation. Suppose that $f\colon \mathbf{X} \to \overline{\mathbf{R}}$ is a polyhedral function, meaning that its epigraph is a polyhedral convex set in $\mathbf{X} \times \mathbf{R}$. Given $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$, it is known (cf. [33, Proposition 13.9]) that

$$\mathrm{d}^2 f(\bar{x}, \bar{v}) = \delta_{K_f(\bar{x}, \bar{v})}.$$

It follows from the convexity of f that $\Delta_t^2 f(\bar{x}, v)(w') \ge 0$ for any $v \in \partial f(\bar{x})$, $w' \in \mathbf{X}$, and $t > 0$. Thus, we get $\widehat{\mathrm{d}}^2 f(\bar{x}, \bar{v})(w) \ge 0$ for any $w \in \mathbf{X}$ and conclude from the inequality $\widehat{\mathrm{d}}^2 f(\bar{x}, \bar{v})(w) \le \mathrm{d}^2 f(\bar{x}, \bar{v})(w)$, satisfied for any $w \in \mathbf{X}$, that $\widehat{\mathrm{d}}^2 f(\bar{x}, \bar{v})(w) = 0$ for any $w \in K_f(\bar{x}, \bar{v})$. Combining this and Proposition 4.2 confirms that

$$\widehat{\mathrm{d}}^2 f(\bar{x}, \bar{v}) = \delta_{K_f(\bar{x}, \bar{v})} = \mathrm{d}^2 f(\bar{x}, \bar{v}), \tag{4.1}$$

which means that the second subderivative of a polyhedral function is always semi-stable. This motivates us to explore further whether or not a similar result holds for CPLQ functions.

Example 4.3. Assume that the function $f : \mathbf{X} \to \overline{\mathbf{R}}$ with $\mathbf{X} = \mathbf{R}^n$ is CPLQ with the representation given in Example 3.3. It was proven in [33, Propsoition 13.9] that the second subderivative of f at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ can be calculated by

$$\mathrm{d}^2 f(\bar{x}, \bar{v})(w) = \begin{cases} \langle A_i w, w \rangle & \text{if } \ w \in K_{C_i}(\bar{x}, \bar{v}_i), \\ \infty & \text{otherwise}, \end{cases} \tag{4.2}$$

where $\bar{v}_i := \bar{v} - A_i \bar{x} - a_i$. We claim that $\widehat{\mathrm{d}}^2 f(\bar{x}, \bar{v})(w) = \mathrm{d}^2 f(\bar{x}, \bar{v})(w)$ for any $w \in \mathbf{X}$, which proves that the second subderivative of a CPLQ function is always semi-stable. Take $w \in K_f(\bar{x}, \bar{v})$ and select sequences $t_k \searrow 0$, $v^k \to \bar{v}$ with $v^k \in \partial f(\bar{x})$, and $w^k \to w$ such that

$$\widehat{\mathrm{d}}^2 f(\bar{x}, \bar{v})(w) = \lim_{k \to \infty} \Delta_{t_k}^2 f(\bar{x}, v^k)(w^k).$$

Since $\widehat{\mathrm{d}}^2 f(\bar{x}, \bar{v})(w) \le \mathrm{d}^2 f(\bar{x}, \bar{v})(w) < \infty$, we can assume without loss of generality that $\bar{x} + t_k w^k \in \operatorname{dom} f = \cup_{i=1}^s C_i$ for any k. Passing to a subsequence, if necessary, we find $j \in \{1, \ldots, s\}$ such that $\bar{x} + t_k w^k \in C_j$ for all k and therefore conclude that $w^k \in T_{C_j}(\bar{x})$ for all k. Clearly, we have $\bar{x} \in C_j$. Moreover, we derive from $w \in K_f(\bar{x}, \bar{v})$ and [34, Proposition 2.1(b)] that

$$\langle A_j \bar{x} + a_j, w \rangle = \mathrm{d}f(\bar{x})(w) = \langle \bar{v}, w \rangle,$$

which in turn brings us to $w \in K_{C_j}(\bar{x}, \bar{v}_j)$. On the other hand, it follows from [33, page 487] that

$$\partial f(\bar{x}) = \bigcap_{i \in I(\bar{x})} \big\{ v \in \mathbf{X} |\, v - A_i \bar{x} - a_i \in N_{C_i}(\bar{x}) \big\}, \tag{4.3}$$

where $I(\bar{x}) = \{ i \in \{1, \ldots, s\} \,|\, \bar{x} \in C_i \}$. By $v^k \in \partial f(\bar{x})$ and $j \in I(\bar{x})$, we get $v^k - A_j \bar{x} - a_j \in N_{C_j}(\bar{x})$. This, coupled with $w^k \in T_{C_j}(\bar{x})$, implies that

$$\langle w^k, v^k - A_j \bar{x} - a_j \rangle \le 0 \quad \text{for all } \ k.$$

Using this, we conclude from a simple calculation that

$$\begin{aligned}\Delta^2_{t_k} f(\bar{x}, v^k)(w^k) &= \frac{f(\bar{x}+t_k w^k) - f(\bar{x}) - t_k\langle v^k,\, w^k\rangle}{\frac{1}{2}t_k^2}\\ &= \frac{\frac{1}{2}\langle A_j(\bar{x}+t_k w^k), \bar{x}+t_k w^k\rangle + \langle a_j, \bar{x}+t_k w^k\rangle + \alpha_j - \frac{1}{2}\langle A_j\bar{x}, \bar{x}\rangle - \langle a_j, \bar{x}\rangle - \alpha_j - t_k\langle v^k,\, w_k\rangle}{\frac{1}{2}t_k^2}\\ &= \langle A_j w^k, w^k\rangle - \frac{\langle w^k, v^k - A_j\bar{x} - a_j\rangle}{\frac{1}{2}t_k} \ge \langle A_j w^k, w^k\rangle.\end{aligned}$$

This, combined with (4.2) and $w \in K_{C_j}(\bar{x}, \bar{v}_j)$, leads us to

$$\widehat{\mathrm{d}}^2 f(\bar{x}, \bar{v})(w) \ge \langle A_j w, w\rangle = \mathrm{d}^2 f(\bar{x}, \bar{v})(w).$$

Since the opposite inequality always holds, we arrive at

$$\widehat{\mathrm{d}}^2 f(\bar{x}, \bar{v})(w) = \mathrm{d}^2 f(\bar{x}, \bar{v})(w) \quad \text{for all} \;\; w \in K_f(\bar{x}, \bar{v}).$$

If $w \notin K_f(\bar{x}, \bar{v})$, we get from Proposition 4.2 and (4.2) that $\widehat{\mathrm{d}}^2 f(\bar{x}, \bar{v})(w) = \mathrm{d}^2 f(\bar{x}, \bar{v})(w) = \infty$, which proves our claim.

The next two examples reveal that the semi-stability of second subderivatives also holds for $\mathcal{C}^2$-decomposable functions and spectral functions, respectively.

Example 4.4. Assume that $g : \mathbf{Y} \to \overline{\mathbf{R}}$ is $\mathcal{C}^2$-decomposable at $\bar{u} \in \mathbf{Y}$ with representation (3.6) and that $\bar{y} \in \partial g(\bar{u})$ and $\bar{\mu} \in M_{\bar{u},g}(\bar{y}, 0)$. Assume further that the dual condition in (3.10) is satisfied. We claim that $\widehat{\mathrm{d}}^2 g(\bar{u}, \bar{y})(w) = \mathrm{d}^2 g(\bar{u}, \bar{y})(w)$ for any $w \in \mathbf{Y}$. To justify it, observe from $\bar{\mu} \in M_{\bar{u},g}(\bar{y}, 0)$ and (3.8) that $\bar{y} = \nabla\Xi(\bar{u})^*\bar{\mu}$ with $\bar{\mu} \in \partial\vartheta(\Xi(\bar{u}))$. It follows from the convexity of the sublinear function $\vartheta : \mathbf{Z} \to \overline{\mathbf{R}}$ that

$$0 \le \widehat{\mathrm{d}}^2\vartheta(\Xi(\bar{u}), \bar{\mu})(\xi) \le \mathrm{d}^2\vartheta(\Xi(\bar{u}), \bar{\mu})(\xi) = \delta_{K_\vartheta(\Xi(\bar{u}),\bar{\mu})}(\xi) \quad \text{for all} \;\; \xi \in \mathbf{Z},$$

where the last equality comes from (7.2). This implies that $\widehat{\mathrm{d}}^2\vartheta(\Xi(\bar{u}), \bar{\mu})(\xi) = \mathrm{d}^2\vartheta(\Xi(\bar{u}), \bar{\mu})(\xi)$ for any $\xi \in K_\vartheta(\Xi(\bar{u}), \bar{\mu})$. If $\xi \notin K_\vartheta(\Xi(\bar{u}), \bar{\mu})$, we infer from (7.2) and Proposition 4.2 that $\widehat{\mathrm{d}}^2\vartheta(\Xi(\bar{u}), \bar{\mu})(\xi) = \mathrm{d}^2\vartheta(\Xi(\bar{u}), \bar{\mu})(\xi) = \infty$. These prove that the second subderivative of ϑ at $\Xi(\bar{u})$ for $\bar{\mu}$ is semi-stable. To obtain the same conclusion for g, we first observe, for any $w \in \mathbf{Y}$, that

$$\begin{aligned}\widehat{\mathrm{d}}^2 g(\bar{u}, \bar{y})(w) &\le \mathrm{d}^2 g(\bar{u}, \bar{y})(w)\\ &= \langle\bar{\mu}, \nabla^2\Xi(\bar{u})(w, w)\rangle + \mathrm{d}^2\vartheta(\Xi(\bar{u}), \bar{\mu})(\nabla\Xi(\bar{u})w)\\ &= \langle\bar{\mu}, \nabla^2\Xi(\bar{u})(w, w)\rangle + \widehat{\mathrm{d}}^2\vartheta(\Xi(\bar{u}), \bar{\mu})(\nabla\Xi(\bar{u})w),\end{aligned}$$

where the penultimate step results from Theorem 7.2(a). Take $w \in \mathbf{Y}$ and select $t_k \searrow 0$, $y^k \to \bar{y}$ with $y^k \in \partial g(\bar{u})$, and $w^k \to w$ such that $\widehat{\mathrm{d}}^2 g(\bar{u}, \bar{y})(w) = \lim_{k\to\infty} \Delta^2_{t_k} g(\bar{u}, y^k)(w^k)$. We know from the dual condition (3.10) and (7.5) that the BCQ condition in (3.9) holds. Employing the chain rule for subdifferentials from [33, Theorem 10.6], we find $\mu^k \in \partial\vartheta(\Xi(\bar{u}))$ such that $y^k = \nabla\Xi(\bar{u})^*\mu^k$. We can also conclude from Theorem 7.1(d) that $\mu^k \to \bar{\mu}$ as $y \to \bar{y}$. Set $\xi^k := \Xi(\bar{u} + t_k w^k)/t_k$ and conclude from (3.6) that

$$\begin{aligned}\Delta^2_{t_k} g(\bar{u}, y^k)(w^k) &= \frac{\vartheta\big(\Xi(\bar{u}+t_k w^k)\big) - \vartheta\big(\Xi(\bar{u})\big) - t_k\langle\nabla\Xi(\bar{u})^*\mu^k, w^k\rangle}{\frac{1}{2}t_k^2}\\ &= \frac{\vartheta\big(\Xi(\bar{u}) + t_k\xi^k\big) - \vartheta\big(\Xi(\bar{u})\big) - t_k\langle\mu^k, \xi^k\rangle}{\frac{1}{2}t_k^2} + \frac{\langle\mu^k, \Xi(\bar{u}+t_k w^k) - t_k\nabla\Xi(\bar{u})w^k\rangle}{\frac{1}{2}t_k^2}\\ &= \Delta^2_{t_k}\vartheta(\Xi(\bar{u}), \mu^k)(\xi^k) + \langle\mu^k, \nabla^2\Xi(\bar{u})(w^k, w^k)\rangle + \frac{o(t_k^2)}{t_k^2}.\end{aligned}$$

Passing to the limit as $k \to \infty$ and using the facts that $\mu^k \to \bar{\mu}$ and $\xi^k \to \nabla\Xi(\bar{u})w$ lead us to

$$\widehat{\mathrm{d}}^2 g(\bar{u}, \bar{y})(w) \geq \widehat{\mathrm{d}}^2\vartheta(\Xi(\bar{u}), \bar{\mu})(\nabla\Xi(\bar{u})w) + \langle \bar{\mu}, \nabla^2\Xi(\bar{u})(w, w)\rangle.$$

Combining the above estimates tells us that $\widehat{\mathrm{d}}^2 g(\bar{u}, \bar{y})(w) = \mathrm{d}^2 g(\bar{u}, \bar{y})(w)$ for any $w \in \mathbf{X}$, which proves our claim.

Example 4.5. Our goal is to show that the second subderivative of spectral functions is semi-stable. Let $\mathbf{S^n}$ stand for the space of all real $n \times n$ symmetric matrices equipped with the inner product

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

The induced Frobenius norm of $X \in \mathbf{S^n}$ is defined by $\|X\| = \sqrt{\mathrm{tr}\,(X^2)}$. Recall that $f : \mathbf{S^n} \to \overline{\mathbf{R}}$ is a spectral function if it is orthogonally invariant, meaning that for any $X \in \mathbf{S^n}$ and any $n \times n$ orthogonal matrix U, we have $f(X) = f(U^\top XU)$. It follows from [21, Proposition 4] that any spectral function f can be expressed in the composite form

$$f(X) := (\theta \circ \lambda)(X), \quad X \in \mathbf{S^n}, \tag{4.4}$$

where $\theta : \mathbf{R}^n \to \overline{\mathbf{R}}$ is a permutation-invariant function on $\mathbf{R}^n$, called symmetric, and λ is the mapping assigning to each matrix $X \in \mathbf{S^n}$ the vector $\big(\lambda_1(X), \ldots, \lambda_n(X)\big)$ of its eigenvalues arranged in nonincreasing order. Before proceeding any further, we discuss some instances of eigenvalue functions that can be covered by spectral functions. To do so, pick $X \in \mathbf{S^n}$ and $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$. If $\theta(x) = \max\{x_1, \ldots, x_n\}$, the spectral function f in (4.4) boils down to the largest eigenvalue of X. If $\theta(x) = \sum_{i=1}^n x_i$, then f reduces to the sum of all eigenvalues of X. Considering $\theta(x) = \max\{x_i + x_j |\, i, j \in \{1, \ldots, n\},\ i \neq j\}$ allows us to cover the sum of two largest eigenvalues of X. If $\theta = \delta_{\mathbf{R}^n_+}$, then we get $f = \delta_{\mathbf{S^n_+}}$, where $\mathbf{S^n_+}$ stands for the cone of $n \times n$ positive semidefinite matrices. The readers can find more examples of eigenvalue functions, covered by (4.4), in [4, page 183].

Denote by $\mathbf{O}^n$ the set of all real $n \times n$ orthogonal matrices (those matrices U satisfying $U^\top U = I$). It is known that for any $X \in \mathbf{S^n}$, there exists $U \in \mathbf{O}$ for which we have

$$X = U\mathrm{Diag}\,(\lambda(X))U^\top \quad \text{with } \ \lambda(X) := \big(\lambda_1(X), \ldots, \lambda_n(X)\big). \tag{4.5}$$

For a given matrix $X \in \mathbf{S^n}$, we denote by $\mathbf{O}^n(X)$ the set of all orthogonal matrices U satisfying (4.5).

Suppose that $f : \mathbf{S^n} \to \overline{\mathbf{R}}$ is a proper convex spectral function and pick $(\overline{X}, \overline{Y}) \in \mathrm{gph}\,\partial f$. Assume that $\mu_1 > \cdots > \mu_r$ are distinct eigenvalues of $\overline{X}$ and define then the index sets

$$\alpha_m := \big\{i \in \{1, \ldots, n\} |\ \lambda_i(\overline{X}) = \mu_m\big\} \quad \text{for all } \ m = 1, \ldots, r.$$

Moreover, define ℓ_i for any $i \in \{1, \ldots, n\}$ to be the number of eigenvalues of $\overline{X}$ that are equal to $\lambda_i(\overline{X})$ but are ranked before $\lambda_i(\overline{X})$ including $\lambda_i(\overline{X})$. We claim that the second subderivative of f is semi-stable at $\overline{X}$ for $\overline{Y}$. To justify it, let $W \in \mathbf{S^n}$, $t > 0$, and $Y \in \partial f(\overline{X})$. Setting $\Delta_t\lambda(\overline{X})(W) := \big(\lambda(\overline{X} + tW) - \lambda(\overline{X})\big)/t$, we get

$$\begin{aligned}
\Delta_t^2 f\big(\overline{X}, Y\big)(W) &= \frac{\theta\big(\lambda(\overline{X} + tW)\big) - \theta\big(\lambda(\overline{X})\big) - t\big\langle Y, W\big\rangle}{\frac{1}{2}t^2} \\
&= \frac{\theta\big(\lambda(\overline{X}) + t\Delta_t\lambda(\overline{X})(W)\big) - \theta\big(\lambda(\overline{X})\big) - t\big\langle \lambda(Y), \Delta_t\lambda(\overline{X})(W)\big\rangle}{\frac{1}{2}t^2} \\
&\quad + \frac{\big\langle \lambda(Y), \Delta_t\lambda(\overline{X})(W)\big\rangle - \big\langle Y, W\big\rangle}{\frac{1}{2}t} \\
&= \Delta_t^2\theta\big(\lambda(\overline{X}), \lambda(Y)\big)\big(\Delta_t\lambda(\overline{X})(W)\big) + \frac{\big\langle \lambda(Y), \Delta_t\lambda(\overline{X})(W)\big\rangle - \big\langle Y, W\big\rangle}{\frac{1}{2}t}.
\end{aligned}$$

It follows from [7, Theorem 5.2.4] that $\lambda(Y) \in \partial\theta(\lambda(\overline{X}))$ and that there is $U \in \mathbf{O}(\overline{X}) \cap \mathbf{O}(Y)$ such that $\overline{X} = U\Lambda(\overline{X})U^\top$ and $Y = U\Lambda(Y)U^\top$, where $\Lambda(\overline{X}) = \mathrm{Diag}\,(\lambda(\overline{X}))$ and $\Lambda(Y) = \mathrm{Diag}\,(\lambda(Y))$. Using these yields

$$\langle Y, W\rangle = \langle U\Lambda(Y)U^\top, W\rangle = \langle \Lambda(Y), U^\top W U\rangle = \sum_{m=1}^{r} \langle \Lambda(Y)_{\alpha_m\alpha_m}, U_{\alpha_m}^\top W U_{\alpha_m}\rangle, \tag{4.6}$$

where $\Lambda(Y)_{\alpha_m\alpha_m}$ stands for the submatrix of $\Lambda(Y)$ obtained by removing all the rows and all the columns of $\Lambda(Y)$ not in α_m and where U_{α_m} stands for the submatrix of U obtained by removing all the columns of U not in α_m. On the other hand, it results from [25, Proposition 2.3] (see also the proof of [38, Theorem 1.5] and [38, Remark 4]) and Fan's inequality (cf. [7, Theorem 1.2.1]) that

$$\begin{aligned}
\langle \lambda(Y), \Delta_t\lambda(\overline{X})(W)\rangle &= \sum_{m=1}^{r}\sum_{j\in\alpha_m} \frac{\lambda_j(Y)\big(\lambda_j(\overline{X}+tW) - \lambda_j(\overline{X})\big)}{t}\\
&= \sum_{m=1}^{r}\sum_{j\in\alpha_m} \lambda_j(Y)\lambda_{l_j}\big(U_{\alpha_m}^\top W U_{\alpha_m} + tU_{\alpha_m}^\top W(\mu_m I - \overline{X})^\dagger W U_{\alpha_m}\big) + O(t^2)\\
&\ge \sum_{m=1}^{r} \langle \Lambda(Y)_{\alpha_m\alpha_m}, U_{\alpha_m}^\top W U_{\alpha_m} + tU_{\alpha_m}^\top W(\mu_m I - \overline{X})^\dagger W U_{\alpha_m}\rangle + O(t^2),
\end{aligned}$$

where $(\mu_m I - \overline{X})^\dagger$ stands for the pseudo-inverse of the matrix $\mu_m I - \overline{X}$. The latter, together with (4.6), implies that

$$\frac{\langle \lambda(Y), \Delta_t\lambda(\overline{X})(W)\rangle - \langle Y, W\rangle}{\frac{1}{2}t} \ge 2\sum_{m=1}^{r} \langle \Lambda(Y)_{\alpha_m\alpha_m}, U_{\alpha_m}^\top W(\mu_m I - \overline{X})^\dagger W U_{\alpha_m}\rangle + O(t).$$

Combining these estimates leads us to

$$\begin{aligned}
\Delta_t^2 f\big(\overline{X}, Y\big)(W) \ge\ & \Delta_t^2\theta\big(\lambda(\overline{X}), \lambda(Y)\big)\big(\Delta_t\lambda(\overline{X})(W)\big)\\
&+ 2\sum_{m=1}^{r} \langle \Lambda(Y)_{\alpha_m\alpha_m}, U_{\alpha_m}^\top W(\mu_m I - \overline{X})^\dagger W U_{\alpha_m}\rangle + O(t).
\end{aligned} \tag{4.7}$$

Take the sequences $W^k \to W$, $t_k \searrow 0$, and $Y^k \to \overline{Y}$ with $Y^k \in \partial f(\overline{X})$ such that

$$\Delta_{t_k}^2 f\big(\overline{X}, Y^k\big)(W^k) \to \widehat{\mathrm{d}}^2 f(\overline{X}, \overline{Y})(W) \quad \text{as } k \to \infty.$$

Employing the same argument as the one for (4.7) tells us that there exist a sequence $U^k \in \mathbf{O}(\overline{X}) \cap \mathbf{O}(Y^k)$ such that (4.7) holds when Y, t, U, and W are replaced with Y^k, t_k, U^k, and W^k, respectively. It follows from $Y^k \to \overline{Y}$ and the inequality

$$\|\lambda(Y^k) - \lambda(\overline{Y})\| \le \|Y^k - \overline{Y}\|$$

that $\lambda(Y^k) \to \lambda(\overline{Y})$. Moreover, since $\mathbf{O}$ is compact, we can assume without loss of generality that the sequence $\{U^k\}_{k\in\mathbb{N}}$ converges to some $U \in \mathbf{O}(\overline{X}) \cap \mathbf{O}(\overline{Y})$. Finally, it is known (cf. see [38, equation (4) and Remark 3]) that $\Delta_{t_k}\lambda(\overline{X})(W^k) \to \lambda'(\overline{X}; W)$, where $\lambda'(\overline{X}; W)$ stands for the classical directional derivative of the eigenvalue mapping $\lambda(\cdot)$ at $\overline{X}$ in the direction of W. Passing then to the limit in (4.7) allows us to conclude for any $W \in \mathbf{S}^\mathbf{n}$ that

$$\widehat{\mathrm{d}}^2 f(\overline{X}, \overline{Y})(W) \ge \widehat{\mathrm{d}}^2\theta\big(\lambda(\overline{X}), \lambda(\overline{Y})\big)\big(\lambda'(\overline{X}; W)\big) + 2\sum_{m=1}^{r} \langle \Lambda(\overline{Y})_{\alpha_m\alpha_m}, U_{\alpha_m}^\top W(\mu_m I - \overline{X})^\dagger W U_{\alpha_m}\rangle.$$

We should add here that the index sets α_m are fixed in this limiting process, since they only depend on $\overline{X}$. To obtain the opposite inequality, we assume further that the symmetric function θ is polyhedral. Employing now [25, Corollary 5.8] tells us that

$$\mathrm{d}^2 f(\overline{X},\overline{Y})(W) = \delta_{K_\theta(\lambda(\overline{X}),\lambda(\overline{Y}))}(\lambda'(\overline{X};W)) + 2\sum_{m=1}^{r} \langle \Lambda(\overline{Y})_{\alpha_m\alpha_m}, U_{\alpha_m}^\top W(\mu_m I - \overline{X})^\dagger W U_{\alpha_m}\rangle$$

for any $W \in \mathbf{S^n}$. By (4.1), we also have

$$\widehat{\mathrm{d}}^2\theta\big(\lambda(\overline{X}),\lambda(\overline{Y})\big)\big(\lambda'(\overline{X};W)\big) = \delta_{K_\theta(\lambda(\overline{X}),\lambda(\overline{Y}))}\big(\lambda'(\overline{X};W)\big).$$

Combining these and using the fact that the inequality $\widehat{\mathrm{d}}^2 f(\overline{X},\overline{Y})(W) \le \mathrm{d}^2 f(\overline{X},\overline{Y})(W)$ is always satisfied, we arrive at

$$\widehat{\mathrm{d}}^2 f(\overline{X},\overline{Y})(W) = \mathrm{d}^2 f(\overline{X},\overline{Y})(W) \quad \text{for all } \; W \in \mathbf{S^n},$$

which proves that the second subderivative of f is semi-stable at $\overline{X}$ for $\overline{Y}$. We should add that polyhedrality assumption on θ can be significantly weakened if one uses [25, Theorem 5.7] to obtain the calculation of $\mathrm{d}^2 f(\overline{X},\overline{Y})(W)$.

We are now in a position to establish our main result of this section, a uniform quadratic growth condition for the augmented Lagrangian in (1.2) when the SOSC (3.18) is satisfied. To provide a motivation for this result, we should remind the readers that it was proven in [13, Theorem 4.1] that for the composite optimization problem in (1.1) with g therein being CPLQ, the SOSC is equivalent to the quadratic growth condition for the augmented Lagrangian function. While the framework in [13] requires that g be CPLQ, it is not hard to see from the proofs therein that what actually is required is twice epi-differentiability in the sense of [33, Definition 13.6] of the convex function g in (1.1), a property that is satisfied for numerous classes of functions and sets; see [23, 24, 33]. Below, we record this characterization of the quadratic growth condition for the augmented Lagrangian function via the SOSC. Since its proof can be gleaned from [13, Theorem 4.1] and since it will not be exploited in this paper, we do not supply it with a proof.

Proposition 4.6. *Let $(\bar{x},\bar{y})$ be a solution to the KKT system (2.2). Then the following properties are equivalent:*

(a) *the SOSC in (3.18) holds at $(\bar{x},\bar{y})$;*

(b) *there exist positive constants $\bar{\rho}$, γ, and ℓ such that for any $\rho \ge \bar{\rho}$, the quadratic growth condition*

$$\mathscr{L}(x,\bar{y},\rho) \ge \mathscr{L}(\bar{x},\bar{y},\rho) + \ell\|x-\bar{x}\|^2 \quad \text{for all } \; x \in \mathbb{B}_\gamma(\bar{x}) \cap \Theta$$

holds.

While the characterization above seems appealing, we will see in the next section when establishing the error bound for consecutive terms of the ALM that the quadratic growth condition in Proposition 4.6(b) is not sufficient. In fact, what is needed is a stronger version of this condition, called the *uniform* quadratic growth condition, which allows for the Lagrangian multiplier $\bar{y}$ to move slightly and the same quadratic growth condition still holds. To achieve it, we begin with the calculation of the second-order difference quotient of the augmented Lagrangian (1.2). Note that $\bar{v}$ in the second-order difference quotient of f in (2.1) is taken from $\partial f(\bar{x})$. When f is $\mathcal{C}^1$, it is known that $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$. In such a case, we drop $\bar{v}$ from the notation of the second-order difference quotient of f in (2.1) and simply write $\Delta_t^2 f(\bar{x})(w)$. The second-order difference quotient of the Moreau envelope was calculated in the proof of [27, Theorem 6.1] (see equation (6.8) in [27]). For the augmented Lagrangian (1.2), a similar calculation was done in the proof of [19, Lemma 3.2]. Since the set Θ in (1.1) was not considered in [19], we provide a proof for the readers' convenience.

Lemma 4.7. *Let $(\bar{x}, \bar{y})$ be a solution to the KKT system (2.2) and set $\bar{v} := \nabla_x \mathscr{L}(\bar{x}, \bar{y}, \rho)$ and $\Delta_t^2 \mathscr{L}(\bar{x}, \bar{y}, \rho)(w) := (\mathscr{L}(\bar{x} + tw, \bar{y}, \rho) - \mathscr{L}(\bar{x}, \bar{y}, \rho) - \langle \bar{v}, w \rangle)/\frac{1}{2}t^2$ for $\rho > 0$. Then for any $\rho > 0$, $t > 0$, and $w \in \mathbf{X}$, we have*

$$\Delta_t^2 \mathscr{L}(\bar{x}, \bar{y}, \rho)(w) = \Delta_t^2 \varphi(\bar{x})(w) + \Delta_t^2 \langle \bar{y}, \Phi \rangle(\bar{x})(w) + \inf_{u \in \mathbf{Y}} \big\{ \Delta_t^2 g(\Phi(\bar{x}), \bar{y})(u) + \rho \|\Delta_t \Phi(\bar{x})(w) - u\|^2 \big\},$$

where $\Delta_t^2 \langle \bar{y}, \Phi \rangle(\bar{x})$ stands for the second-order difference quotient of the mapping $x \mapsto \langle \bar{y}, \Phi(x) \rangle$ at $\bar{x}$ and where $\Delta_t \Phi(\bar{x})(w) := (\Phi(\bar{x} + tw) - \Phi(\bar{x}))/t$ for any $t > 0$.

Proof. We can conclude from Proposition 2.2(a)-(b) that $\bar{v} = \nabla \varphi(\bar{x}) + \nabla \Phi(\bar{x})^* \bar{y}$ and

$$\begin{aligned}
&\tfrac{2}{t^2}\big(\mathscr{L}(\bar{x} + tw, \bar{y}, \rho) - \mathscr{L}(\bar{x}, \bar{y}, \rho) - t\langle \bar{v}, w \rangle\big) \\
&= \tfrac{2}{t^2}\big(\varphi(\bar{x} + tw) - \varphi(\bar{x}) + e_{1/\rho} g\big(\Phi(\bar{x} + tw) + \rho^{-1}\bar{y}\big) - g(\Phi(\bar{x})) - \tfrac{1}{2}\rho^{-1}\|\bar{y}\|^2 - t\langle \bar{v}, w \rangle\big) \\
&= \Delta_t^2 \varphi(\bar{x})(w) - \tfrac{2}{t}\langle \nabla \Phi(\bar{x})^* \bar{y}, w \rangle \\
&\quad + \tfrac{2}{t^2}\Big(\inf_{z \in \mathbf{Y}} \big\{ g(z) - g(\Phi(\bar{x})) - \langle \bar{y}, z - \Phi(\bar{x}) \rangle + \tfrac{1}{2}\rho\|z - \Phi(\bar{x}) - (\Phi(\bar{x} + tw) - \Phi(\bar{x}))\|^2 \\
&\quad + \tfrac{1}{2}\rho\|z - \Phi(\bar{x} + tw) - \rho^{-1}\bar{y}\|^2 - \tfrac{1}{2}\rho\|z - \Phi(\bar{x} + tw)\|^2 - \tfrac{1}{2}\rho^{-1}\|\bar{y}\|^2 + \langle \bar{y}, z - \Phi(\bar{x}) \rangle \big\} \Big) \\
&= \Delta_t^2 \varphi(\bar{x})(w) + \inf_{u \in \mathbf{Y}} \big\{ \Delta_t^2 g(\Phi(\bar{x}), \bar{y})(u) + \rho\|u - \Delta_t \Phi(\bar{x})(w)\|^2 \big\} \\
&\quad + \tfrac{2}{t^2}\langle \bar{y}, \Phi(\bar{x} + tw) - \Phi(\bar{x}) - t\nabla \Phi(\bar{x}) w \rangle \\
&= \Delta_t^2 \varphi(\bar{x})(w) + \Delta_t^2 \langle \bar{y}, \Phi \rangle(\bar{x})(w) + \inf_{u \in \mathbf{Y}} \big\{ \Delta_t^2 g(\Phi(\bar{x}), \bar{y})(u) + \rho\|u - \Delta_t \Phi(\bar{x})(w)\|^2 \big\}.
\end{aligned}$$

This clearly completes the proof. ☐

The main consequence of the semi-stability of second subderivatives resides in the following observation in which we ensure the uniform quadratic growth condition for the augmented Lagrangian function.

Theorem 4.8. *Let $(\bar{x}, \bar{y})$ be a solution to the KKT system (2.2). Assume that the second subderivative of g at $\Phi(\bar{x})$ for $\bar{y}$ is semi-stable and that there is a positive constant ℓ such that the SOSC*

$$\langle \nabla_{xx}^2 L(\bar{x}, \bar{\lambda}) w, w \rangle + \mathrm{d}^2 g(\Phi(\bar{x}), \bar{y})(\nabla \Phi(\bar{x}) w) \geq \ell \|w\|^2 \quad \textit{for all } \; w \in \mathscr{D} \tag{4.8}$$

holds, where $\mathscr{D}$ is taken from (3.18). Then, there exist positive constants $\bar{\rho}$ and γ for which the uniform quadratic growth condition

$$\begin{cases} \mathscr{L}(x, y, \rho) \geq \mathscr{L}(\bar{x}, y, \rho) + \frac{\kappa}{2}\|x - \bar{x}\|^2 \\ \textit{for all } \; x \in \Theta \cap \mathbb{B}_\gamma(\bar{x}), \; y \in M_{\bar{x},\psi}(0,0) \cap \mathbb{B}_\gamma(\bar{y}), \; \rho \geq \bar{\rho}, \end{cases} \tag{4.9}$$

is fulfilled for all $\kappa \in [0, \ell)$, where $M_{\bar{x},\psi}(0,0)$ is the set of Lagrange multipliers associated with $\bar{x}$ and is taken from (3.19).

Proof. Suppose by contradiction that (4.9) fails. Thus, there exist $x^k \to \bar{x}$ with $x^k \in \Theta$, $y^k \to \bar{y}$ with $y^k \in M_{\bar{x},\psi}(0,0)$, and $\rho_k \to \infty$ such that

$$\mathscr{L}(x^k, y^k, \rho_k) < \mathscr{L}(\bar{x}, y^k, \rho_k) + \frac{\kappa}{2}\|x^k - \bar{x}\|^2,$$

holds for some $\kappa \in [0,\ell)$. Set $t_k := \|x^k - \bar{x}\|$ and $w^k := (x^k - \bar{x})/t_k$. Passing to a subsequence if necessary, we can assume that $w^k \to w$ for some $w \in \mathbf{X}$ with $\|w\| = 1$. Substituting $x^k = \bar{x} + t_k w^k$ into the latter estimate, we arrive at

$$\Delta^2_{t_k} \mathscr{L}(\bar{x}, y^k, \rho_k)(w^k) + \frac{2}{t_k}\langle v^k, w^k\rangle = \frac{\mathscr{L}(\bar{x} + t_k w^k, y^k, \rho_k) - \mathscr{L}(\bar{x}, y^k, \rho_k)}{\frac{1}{2}t_k^2} < \kappa.$$

where $v^k := \nabla_x \mathscr{L}(\bar{x}, y^k, \rho_k)$. It follows from $y^k \in M_{\bar{x},\psi}(0,0)$ that $(\bar{x}, y^k)$ is a solution to the KKT system in (2.2). By Proposition 2.2(b), we have $-v^k = -(\nabla\varphi(\bar{x}) + \nabla\Phi(\bar{x})^* y^k) \in N_\Theta(\bar{x})$. Set $\alpha_k := \Delta^2_{t_k}\varphi(\bar{x})(w^k) + \Delta^2_{t_k}\langle y^k, \Phi\rangle(\bar{x})(w^k)$ and observe that

$$\alpha_k \to \langle \nabla^2_{xx} L(\bar{x}, \bar{y})w, w\rangle \quad \text{ and } \quad v^k \to \nabla_x L(\bar{x}, \bar{y}) \quad \text{as } \ k \to \infty. \tag{4.10}$$

It follows from Lemma 4.7 that

$$\alpha_k + \frac{2}{t_k}\langle v^k, w^k\rangle + \inf_{u \in \mathbf{Y}} \left\{\Delta^2_{t_k} g(\Phi(\bar{x}), y^k)(u) + \rho_k \|u - \Delta_{t_k}\Phi(\bar{x})(w^k)\|^2\right\} < \kappa.$$

Since g is convex and $y^k \in \partial g(\Phi(\bar{x}))$, we get that $\Delta^2_{t_k} g(\Phi(\bar{x}), y^k)(\cdot)$ is nonnegative. So, for any $k \in \mathbb{N}$, we find $u^k \in \mathbf{Y}$ such that

$$\alpha_k + \frac{2}{t_k}\langle v^k, w^k\rangle + \Delta^2_{t_k} g(\Phi(\bar{x}), y^k)(u^k) + \rho_k\|u^k - \Delta_{t_k}\Phi(\bar{x})(w^k)\|^2 < \kappa + \frac{1}{k}. \tag{4.11}$$

It also follows from the convexity of Θ that $\frac{2}{t_k}\langle v^k, w^k\rangle = \Delta^2_{t_k}\delta_\Theta(\bar{x}, -v^k)(w^k) \geq 0$. Using these and (4.11), we obtain

$$\|\Delta_{t_k}\Phi(\bar{x})(w^k) - u^k\|^2 \leq \frac{1}{\rho_k}\big(\kappa + \tfrac{1}{k} - \alpha_k\big),$$

which clearly yields $\Delta_{t_k}\Phi(\bar{x})(w^k) - u^k \to 0$ as $k \to \infty$. Combining this with $\Delta_{t_k}\Phi(\bar{x})(w^k) \to \nabla\Phi(\bar{x})w$, we arrive at $u^k \to \nabla\Phi(\bar{x})w$. Employing (4.11) again, we conclude that

$$\frac{2}{t_k}\big(\Delta_{t_k} g(\Phi(\bar{x}))(u^k) - \langle y^k, u^k\rangle\big) = \Delta^2_{t_k} g(\Phi(\bar{x}), y^k)(u^k) < \kappa + \frac{1}{k} - \alpha_k,$$

which implies that

$$\liminf_{k\to\infty} \frac{2}{t_k}\big(\Delta_{t_k} g(\Phi(\bar{x}))(u^k) - \langle y^k, u^k\rangle\big) < \infty.$$

In particular, it holds that $\liminf_{k\to\infty} \Delta_{t_k} g(\Phi(\bar{x}))(u^k) - \langle y^k, u^k\rangle \leq 0$. This, coupled with $u^k \to \nabla\Phi(\bar{x})w$, confirms that $\mathrm{d}g(\Phi(\bar{x}))(\nabla\Phi(\bar{x})w) \leq \langle \bar{y}, \nabla\Phi(\bar{x})w\rangle$. It follows from the convexity of g and $\bar{y} \in \partial g(\Phi(\bar{x}))$ that the opposite inequality $\mathrm{d}g(\Phi(\bar{x}))(\nabla\Phi(\bar{x})w) \geq \langle \bar{y}, \nabla\Phi(\bar{x})w\rangle$ always holds. Thus, we get $\nabla\Phi(\bar{x})w \in K_g(\Phi(\bar{x}), \bar{y})$. Repeating the above arguments for the second-order quotient of δ_Θ at $\bar{x}$ for $-v^k$ and w^k yields $w \in K_\Theta(\bar{x}, -\nabla_x L(\bar{x}, \bar{y}))$. Thus, it results from the definition of $\mathscr{D}$ in (3.18) that $w \in \mathscr{D}$. Observe again that (4.11) gives us the estimate

$$\alpha_k + \Delta^2_{t_k} g(\Phi(\bar{x}), y^k)(u^k) + \Delta^2_{t_k}\delta_\Theta(\bar{x}, -v^k)(w^k) < \kappa + \frac{1}{k}.$$

Passing to the limit, using (4.10), and remembering that $u^k \to \nabla\Phi(\bar{x})w$ demonstrate that

$$\langle \nabla^2_{xx} L(\bar{x}, \bar{y})w, w\rangle + \widehat{\mathrm{d}}^2 g(\Phi(\bar{x}), \bar{y})(\nabla\Phi(\bar{x})w) + \widehat{\mathrm{d}}^2\delta_\Theta\big(\bar{x}, -\nabla_x L(\bar{x}, \bar{y})\big)(w) \leq \kappa. \tag{4.12}$$

Since Θ is a polyhedral convex set, δ_Θ is CPLQ. By Example 4.3, the second subderivative of δ_Θ is semi-stable at $\bar{x}$ for $-\nabla_x L(\bar{x}, \bar{y})$. Moreover, (4.1) and $w \in K_\Theta(\bar{x}, -\nabla_x L(\bar{x}, \bar{y}))$ imply that

$$\widehat{\mathrm{d}}^2\delta_\Theta(\bar{x}, -\nabla_x L(\bar{x}, \bar{y}))(w) = \delta_{K_\Theta(\bar{x}, -\nabla_x L(\bar{x}, \bar{y}))}(w) = 0,$$

which, coupled with semi-stability of the second subderivative of g at $\Phi(\bar{x})$ for $\bar{y}$, results in

$$\langle \nabla^2_{xx} L(\bar{x}, \bar{y})w, w\rangle + \mathrm{d}^2 g(\Phi(\bar{x}), \bar{y})(\nabla\Phi(\bar{x})w) \leq \kappa,$$

a contradiction with (4.8), since $\kappa \in [0, \ell)$, $\|w\| = 1$, and $w \in \mathscr{D}$. This completes the proof. $\square$

It is worth adding here that the SOSCs in (3.18) and (4.8) are equivalent. This, in fact, results from the lower semicontinuity of the second subderivative function $\mathrm{d}^2 g(\Phi(\bar{x}), \bar{y})$ therein, which is due to [33, Propsoition 13.5]. Note that the proof of Theorem 4.8 is inspired by that of [36, Theorem 3.4(b)], which was replicated in [20, Proposition 3.3].

The uniform quadratic growth condition for augmented Lagrangian functions was first established in [10, Proposition 3.1] for NLPs without appealing to the concept of the second subderivative. It was generalized in [13, Theorem 4.3] for the composite problem in (1.1) with the modeling function g therein being CPLQ via a different approach. It was recently generalized in [39, Theorem 1] for SDPs when the multiplier $\bar{y}$ is taken from the relative interior of the set of Lagrange multipliers. Theorem 4.8 goes far beyond these frameworks and provides a relatively easy proof of this important result, which is based on the new concept of semi-stability of second subderivatives, introduced in this section. We should also add here that the uniform quadratic growth condition in (4.9) implies the validity of the SOSC in (4.8) for some $\ell > 0$. This can be proven using Proposition 4.6. Since such a result will not be used in our convergence analysis and since it requires calculation of the second subderivative of the augmented Lagrangian function, we did not state it in Theorem 4.8.

5 Local Convergence Analysis of ALM

After some preparations in Sections 3 and 4, we are ready to analyze the local convergence of an inexact version of the ALM described below.

Algorithm 5.1 (ALM)**.** Choose $(x^0, y^0) \in \mathbf{X} \times \mathbf{Y}$, $\bar{\rho} > 0$, and $\hat{c} > 0$. Pick a sequence $\{\rho_k\}_{k\geq 0}$ with $\rho_k \geq \bar{\rho}$ for all k and a function $\epsilon : \mathbf{R}_+ \to \mathbf{R}_+$ satisfying $\epsilon(t) = o(t)$ and set $k := 0$. Then

(1) if (x^k, y^k) satisfies a suitable termination criterion, stop;

(2) otherwise, set $\epsilon_k := \epsilon(r(x^k, \lambda^k))$ with the residual function r taken from (3.21) and choose the primal-dual update (x^{k+1}, y^{k+1}) according to (1.5) and (2.9) such that

$$\|x^{k+1} - x^k\| + \|y^{k+1} - y^k\| \leq \hat{c}\, r(x^k, y^k); \tag{5.1}$$

(3) set $k \leftarrow k + 1$ and go to Step 1.

The roadmap to ensure the convergence of the sequence $\{(x^k, y^k)\}_{k\in\mathbb{N}}$, constructed by Algorithm 5.1, was already established in Theorem 2.1. To this end, we are going to demonstrate that assumptions (a)-(c) in the latter result are satisfied under the same assumptions utilized in Theorem 3.9. To begin, define the solution mapping $\Upsilon : \mathbf{X} \times \mathbf{Y} \to \mathbf{X} \times \mathbf{Y}$ to the canonical perturbation of the generalized equation in (2.3) by

$$\Upsilon(v, w) := \big\{(x, y) \in \mathbf{X} \times \mathbf{Y} \,\big|\, (v, w) \in \Psi(x, y) + G(x, y)\big\}, \;\; (v, w) \in \mathbf{X} \times \mathbf{Y}, \tag{5.2}$$

where Ψ and G are taken from (2.10). Observe that assumption (a) in Theorem 2.1 requires that Υ enjoy a calmness property. Below, we show that a certain calmness property of Υ is equivalent to the error bound estimate in (3.22).

Proposition 5.2. *Let $(\bar{x}, \bar{y})$ be a solution to the KKT system in (2.2). Then, the following properties are equivalent.*

(a) *There exist constants* $\gamma > 0$ *and* $\kappa \geq 0$ *for which the error bound estimate in* (3.22) *holds.*
(b) *There are positive constants* δ' *and* ℓ *for which the solution mapping* Υ *from* (5.2) *enjoys the calmness property*

$$\Upsilon(v,w) \cap \mathbb{B}_{\delta'}(\bar{x},\bar{y}) \subset \big(\{\bar{x}\} \times M_{\bar{x},\psi}(0,0)\big) + \ell\,\|(v,w)\|\mathbb{B} \quad \text{for all } (v,w) \in \delta'\mathbb{B}, \tag{5.3}$$

where $M_{\bar{x},\psi}(0,0)$ *is the set of Lagrange multipliers associated with* $\bar{x}$ *and is defined by* (3.19).

Proof. The equivalence of (a) and (b) was established in [34, Proposition 3.8] for the composite problem in (1.1) with g therein being CPLQ. A close look into its proof, however, tells us that the latter assumption on g was not exploited and the given argument works for any convex function g. □

We now proceed with an elaboration of assumption (b) in Theorem 2.1 for the inexact ALM from Algorithm 5.1, which consists of two steps: (1) verifying the solvability of the subproblem in (1.4) and (2) establishing an error bound estimate for consecutive iterates of the ALM. We begin with solvability of subproblems. This was already justified in [13, Propsoition 5.2] when the convex function g in (1.1) is CPLQ. The given proof therein did not utilize such an assumption on g and indeed works for any convex function g. Below, we record this result and skip its proof.

Proposition 5.3. *Assume that* $(\bar{x},\bar{y})$ *is a solution to the KKT system* (2.2) *such that the SOSC in* (4.8) *hold at* $(\bar{x},\bar{y})$. *Assume further that the second subderivative of* g *at* $\Phi(\bar{x})$ *for* $\bar{y}$ *be semi-stable. Take the positive constants* γ *and* $\bar{\rho}$ *from Theorem* 4.8. *Then, there exist positive constants* $\hat{\ell}$ *and* $\hat{\gamma} \in (0,\gamma)$ *such that for any* $\rho \in [\bar{\rho},\infty)$, *the optimal solution mapping* $S_\rho : \mathbf{Y} \rightrightarrows \mathbf{X}$, *defined by*

$$S_\rho(y) := \operatorname{arg\,min}\big\{\mathscr{L}(x,y,\rho)\;\big|\; x \in \Theta \cap \mathbb{B}_{\hat{\gamma}}(\bar{x})\big\}, \quad y \in \mathbf{Y}, \tag{5.4}$$

enjoys the uniform isolated calmness property

$$S_\rho(y) \subset \{\bar{x}\} + \hat{\ell}\|y - \bar{y}\|\mathbb{B} \tag{5.5}$$

and satisfies the inclusion $\emptyset \neq S_\rho(y) \subset \operatorname{int}\mathbb{B}_{\hat{\gamma}}(\bar{x})$ *for all* $y \in \mathbb{B}_{\hat{\gamma}/(2\hat{\ell})}(\bar{y})$.

Next, we address the second property, required in assumption (b) in Theorem 2.1, which is an error bound estimate for consecutive terms of our inexact ALM. While the proof of the result below mostly uses a similar argument in our recent work in [13, Theorem 5.7], it differs from the proof of the latter result in the second half part. In fact, some parts of the proof of [13, Theorem 5.7] depend heavily on that fact that g from (1.1) was assumed to be CPLQ. In those parts, we proceed with a new idea, used recently in [39, Proposition 5].

Proposition 5.4. *Assume that* $(\bar{x},\bar{y})$ *is a solution to the KKT system in* (2.2) *for which all the assumptions in Theorem* 3.9 *hold and that the second subderivative of* g *at* $\Phi(\bar{x})$ *for* $\bar{y}$ *is semi-stable. Take* $\bar{\rho}$ *from Theorem* 4.8 *and* $\hat{\gamma}$ *from Proposition* 5.3. *Then, there exist positive constants* $\hat{\varepsilon}$ *and* $\hat{c}$ *such that for any* $\rho \geq \bar{\rho}$, *any* $(x,y) \in \mathbb{B}_{\hat{\varepsilon}}(\bar{x},\bar{y})$ *satisfying* $r(x,y) > 0$, *and any optimal solution* s *to the regularized problem*

$$\text{minimize } \mathscr{L}(u,y,\rho) \quad \text{subject to} \quad u \in \Theta \cap \mathbb{B}_{\hat{\gamma}}(\bar{x}) \tag{5.6}$$

the error bound estimate

$$\|s - x\| + \|y_s - y\| \leq \hat{c}\, r(x,y) \quad \text{with } y_s := y + \rho\nabla_y\mathscr{L}(s,y,\rho) \tag{5.7}$$

holds, where the residual function r *is defined by* (3.21).

Proof. Since the SOSC in (3.18) holds, it follows from lower semicontinuity of $\mathrm{d}^2 g(\Phi(\bar{x}), \bar{y})$ therein (cf. [33, Proposition 13.5]) that there is a positive constant ℓ for which the SOSC in (4.8) is satisfied. By Theorem 4.8, we find positive constants $\bar{\rho}$, γ, and κ for which the uniform quadratic growth condition in (4.9) holds for all $x \in \Theta \cap \mathbb{B}_\gamma(\bar{x})$, all $y \in M_{\bar{x},\psi}(0,0) \cap \mathbb{B}_\gamma(\bar{y})$, and all $\rho \geq \bar{\rho}$. According to Proposition 5.3, the solution mapping S_ρ enjoys the uniform isolated calmness property in (5.5) and $S_\rho(\tilde{y}) \subset \operatorname{int} \mathbb{B}_{\hat{\gamma}}(\bar{x})$ for all $\tilde{y} \in \mathbb{B}_{\hat{\gamma}/(2\hat{\ell})}(\bar{y})$ and all $\rho \geq \bar{\rho}$, where both positive constants $\hat{\ell}$ and $\hat{\gamma}$ are taken from this proposition. Thus, for every $\tilde{y} \in \mathbb{B}_{\hat{\gamma}/(2\hat{\ell})}(\bar{y})$ and every $\rho \geq \bar{\rho}$, any optimal solution s to (5.6) satisfies the first-order optimality condition

$$0 \in \nabla_x \mathscr{L}(s, \tilde{y}, \rho) + N_\Theta(s). \tag{5.8}$$

To justify (5.7), assume by contradiction that it does not hold. Thus, we find a sequence $\{(x^k, y^k, \rho_k)\}_{k \in \mathbb{N}} \subset \mathbf{X} \times \mathbf{Y} \times [\bar{\rho}, \infty)$ with $(x^k, y^k) \to (\bar{x}, \bar{y})$ and an optimal solution s^k to (5.6) associated with $(y, \rho) = (y^k, \rho_k)$ such that

$$\|s^k - x^k\| + \|q^k - y^k\| > k\, r_k \quad \text{with } \; q^k := y^k + \rho_k \nabla_y \mathscr{L}(s^k, y^k, \rho_k), \tag{5.9}$$

where $r_k := r(x^k, y^k) > 0$. The latter particularly tells us that r_k is finite, which yields $x^k \in \Theta$ for all k due to (3.21). Denoting by β_k the left-hand side of (5.9), we get $r_k = o(\beta_k)$. The definition of the residual function r in (3.21) then leads us to

$$-\nabla_x L(x^k, y^k) + o(\beta_k) \in N_\Theta(x^k) \quad \text{and} \quad \Phi(x^k) + o(\beta_k) = \operatorname{prox}_g\big(\Phi(x^k) + y^k\big). \tag{5.10}$$

Using the definition of β_k and passing to a subsequence, if necessary, we can find $(\zeta, \eta) \in \mathbf{X} \times \mathbf{Y}$ such that

$$\frac{s^k - x^k}{\beta_k} \to \zeta \quad \text{and} \quad \frac{q^k - y^k}{\beta_k} \to \eta \quad \text{with } \; (\zeta, \eta) \neq 0. \tag{5.11}$$

Since the set of Lagrange multiplier $M_{\bar{x},\psi}(0,0)$ from (3.19) is convex and closed, $P_{M_{\bar{x},\psi}(0,0))}(y^k)$ exists and is a singleton. Set $\tilde{y}^k := P_{M_{\bar{x},\psi}(0,0))}(y^k)$. By Theorem 3.9, the estimate in (3.22) is satisfied, which allows us to conclude that $x^k - \bar{x} = O(r_k)$ and $y^k - \tilde{y}^k = O(r_k)$ for all k sufficiently large. Thus, we have

$$x^k - \bar{x} = o(\beta_k) \quad \text{and} \quad y^k - \tilde{y}^k = o(\beta_k) \quad \text{as } k \to \infty, \tag{5.12}$$

which together with $y^k \to \bar{y}$ yields $\tilde{y}^k \to \bar{y}$ and hence $\tilde{y}^k \in M_{\bar{x},\psi}(0,0) \cap \mathbb{B}_\gamma(\bar{y})$ for all k sufficiently large. Combining the latter with $s^k \in \Theta \cap \mathbb{B}_\gamma(\bar{x})$, $\rho_k \geq \bar{\rho}$, and (4.9) brings us to

$$\begin{aligned}
\|s^k - \bar{x}\|^2 &\leq \frac{2}{\kappa}\big(\mathscr{L}(s^k, \tilde{y}^k, \rho_k) - \mathscr{L}(\bar{x}, \tilde{y}^k, \rho_k)\big) \\
&\leq \frac{2}{\kappa}\big(\mathscr{L}(s^k, y^k, \rho_k) + \langle \nabla_y \mathscr{L}(s^k, y^k, \rho_k), \tilde{y}^k - y^k \rangle - \mathscr{L}(\bar{x}, \tilde{y}^k, \rho_k)\big),
\end{aligned}$$

where the last inequality results from the fact that the mapping $y \mapsto \mathscr{L}(s^k, y, \rho_k)$ is $\mathcal{C}^1$ and concave; see [33, Exercise 11.56]. Since $s^k \in S_{\rho_k}(y^k)$, we can conclude from the definition of the augmented Lagrangian function in (1.2) that

$$\mathscr{L}(s^k, y^k, \rho_k) \leq \mathscr{L}(\bar{x}, y^k, \rho_k) \leq \varphi(\bar{x}) + g(\Phi(\bar{x})) = \mathscr{L}(\bar{x}, \tilde{y}^k, \rho_k),$$

where the equality comes from Proposition 2.2(a) and $\tilde{y}^k \in M_{\bar{x},\psi}(0,0)$. Combining the last two estimates and using then (5.9) lead us to

$$\|s^k - \bar{x}\|^2 \leq \frac{2}{\kappa \rho_k}\langle q^k - y^k, \tilde{y}^k - y^k \rangle \leq \frac{2}{\kappa \rho_k}\|q^k - y^k\| \cdot \|\tilde{y}^k - y^k\|. \tag{5.13}$$

Claim I. We have $s^k - \bar{x} = o(\beta_k)$, $\rho_k\|s^k - \bar{x}\|^2 = o(\beta_k)$, and $\rho_k\|s^k - \bar{x}\|^2 = o(\beta_k^2)$.

To prove the first estimate, we use (5.13) together with (5.11)-(5.12) and $\rho_k \geq \bar{\rho}$ to obtain

$$\frac{\|s^k - \bar{x}\|^2}{{\beta_k}^2} \leq \frac{2}{\kappa\rho_k}\frac{\|q^k - y^k\|}{\beta_k}\cdot\frac{\|\tilde{y}^k - y^k\|}{\beta_k} \leq \frac{2}{\kappa\bar{\rho}}\frac{\|q^k - y^k\|}{\beta_k}\cdot\frac{\|\tilde{y}^k - y^k\|}{\beta_k} \to 0.$$

The remaining estimates can be verified by a similar argument via (5.11)-(5.13).

Claim II. We have $s^k \to \bar{x}$ as $k \to \infty$.

To prove this claim, we infer from $\tilde{y}^k \in M_{\bar{x},\psi}(0,0)$ that $\tilde{y}^k \in \partial g(\Phi(\bar{x}))$, and hence $\Phi(\bar{x}) = \operatorname{prox}_{\rho^{-1}g}\big(\Phi(\bar{x}) + \rho^{-1}\tilde{y}^k\big)$ for any $\rho > 0$ due to (2.8). Thus, we derive from (5.9), (5.13), (2.12), and $\rho_k \geq \bar{\rho}$ that

$$\begin{aligned}
\|s^k - \bar{x}\|^2 &\leq \frac{2}{\kappa}\|\Phi(s^k) - \operatorname{prox}_{\rho_k^{-1}g}\big(\Phi(s^k) + \rho_k^{-1}y^k\big)\|\cdot\|\tilde{y}^k - y^k\| \\
&= \frac{2}{\kappa}\big\|\Phi(s^k) - \Phi(\bar{x}) + \operatorname{prox}_{\rho_k^{-1}g}\big(\Phi(\bar{x}) + \rho_k^{-1}\tilde{y}^k\big) - \operatorname{prox}_{\rho_k^{-1}g}\big(\Phi(s^k) + \rho_k^{-1}y^k\big)\big\|\cdot\|\tilde{y}^k - y^k\| \\
&\leq \frac{2}{\kappa}\big(2\|\Phi(s^k) - \Phi(\bar{x})\| + (\bar{\rho})^{-1}\|\tilde{y}^k - y^k\|\big)\|\tilde{y}^k - y^k\| \\
&= \big(O(\|s^k - \bar{x}\|) + \tfrac{2}{\kappa}(\bar{\rho})^{-1}\|\tilde{y}^k - y^k\|\big)\|\tilde{y}^k - y^k\|, \qquad (5.14)
\end{aligned}$$

where the inequality is due to the fact that the proximal mapping $\operatorname{prox}_{\rho^{-1}g}$ is nonexpansive (cf. [33, Proposition 12.19]). Since $\|s^k - \bar{x}\|$ is bounded and $\|y^k - \tilde{y}^k\| \to 0$, we deduce from (5.14) that $\|s^k - \bar{x}\| \to 0$, which proves the claim.

To proceed, we use the first estimate in Claim I and (5.12) to get

$$\zeta = \lim_{k\to\infty}\frac{s^k - x^k}{\beta_k} = \lim_{k\to\infty}\frac{s^k - \bar{x}}{\beta_k} - \lim_{k\to\infty}\frac{x^k - \bar{x}}{\beta_k} = 0 - 0 = 0.$$

Our goal is to show that $\eta = 0$, which together with $\zeta = 0$ leads us to a contradiction with (5.11). To this end, we proceed with considering two cases. Assume first that either $\{\rho_k\}_{k\in\mathbb{N}}$ or $\{\rho_k/\beta_k\}_{k\in\mathbb{N}}$ is bounded. Using a similar argument as the one for (5.14), we obtain via (5.9) and (2.12) that

$$\frac{\|q^k - y^k\|}{\beta_k} = \frac{\rho_k}{\beta_k}\big\|\Phi(s^k) - \operatorname{prox}_{\rho_k^{-1}g}\big(\Phi(s^k) + \rho_k^{-1}y^k\big)\big\| \leq O\big(\frac{\rho_k}{\beta_k}\|s^k - \bar{x}\|\big) + \frac{\|\tilde{y}^k - y^k\|}{\beta_k}.$$

It also follows from $s^k - \bar{x} = o(\beta_k)$ that $\rho_k\|s^k - \bar{x}\|/\beta_k \to 0$. Using this and (5.12) and passing to the limit in the inequality above, we get $\eta = 0$, a contradiction with (5.11).

Assume now that both sequences $\{\rho_k\}_{k\in\mathbb{N}}$ and $\{\rho_k/\beta_k\}_{k\in\mathbb{N}}$ are unbounded. We can assume by passing to a subsequence if necessary that

$$\rho_k \to \infty \quad \text{and} \quad \frac{\rho_k}{\beta_k} \to \infty \text{ as } k \to \infty. \qquad (5.15)$$

Since s^k is an optimal solution to (5.6) associated with (y^k, ρ_k), we deduce from (5.8) that

$$0 \in \nabla_x\mathscr{L}(s^k, y^k, \rho_k) + N_\Theta(s^k) = \nabla\varphi(s^k) + \nabla\Phi(s^k)^*q^k + N_\Theta(s^k),$$

where the equality comes from (2.12) and the definition of q^k from (5.9). By the definition of the normal cone, we have

$$\begin{aligned}
0 &\leq \langle\nabla\varphi(s^k) + \nabla\Phi(s^k)^*q^k, \bar{x} - s^k\rangle = \langle\nabla\varphi(s^k), \bar{x} - s^k\rangle - \langle q^k, \nabla\Phi(s^k)(s^k - \bar{x})\rangle \\
&= \langle\nabla\varphi(s^k), \bar{x} - s^k\rangle - \langle q^k, \nabla\Phi(\bar{x})(s^k - \bar{x})\rangle - \langle q^k, (\nabla\Phi(s^k) - \nabla\Phi(\bar{x}))(s^k - \bar{x})\rangle. \qquad (5.16)
\end{aligned}$$

Recall also that $\tilde{y}^k \in M_{\bar{x},\psi}(0,0)$. By (3.19), the latter implies that $\tilde{y}^k \in \partial g(\Phi(\bar{x}))$ and

$$0 \in \nabla_x L(\bar{x}, \tilde{y}^k) + N_\Theta(\bar{x}) = \nabla\varphi(\bar{x}) + \nabla\Phi(\bar{x})^*\tilde{y}^k + N_\Theta(\bar{x}).$$

Again, the definition of the normal cone leads us to

$$0 \le \langle \nabla\varphi(\bar{x}) + \nabla\Phi(\bar{x})^*\tilde{y}^k, s^k - \bar{x}\rangle = \langle \nabla\varphi(\bar{x}), s^k - \bar{x}\rangle + \langle \tilde{y}^k, \nabla\Phi(\bar{x})(s^k - \bar{x})\rangle. \tag{5.17}$$

Adding both sides of (5.16) and (5.17) brings us to

$$\begin{aligned}\langle q^k - \tilde{y}^k, \nabla\Phi(\bar{x})(s^k - \bar{x})\rangle &\le -\langle \nabla\varphi(s^k) - \nabla\varphi(\bar{x}), s^k - \bar{x}\rangle - \langle q^k, (\nabla\Phi(s^k) - \nabla\Phi(\bar{x}))(s^k - \bar{x})\rangle \\ &\le (1 + \|q^k\|)O(\|s^k - \bar{x}\|^2). \end{aligned}\tag{5.18}$$

On the other hand, it follows from (2.9) and the definition of q^k in (5.9) that $q^k \in \partial g(z^k)$ where $z^k := \Phi(s^k) - \rho_k^{-1}(q^k - y^k)$. Recall that $\tilde{y}^k \in \partial g(\Phi(\bar{x}))$. Thus, we get from the monotonicity of ∂g that

$$\begin{aligned}0 &\le \langle q^k - \tilde{y}^k, \Phi(s^k) - \rho_k^{-1}(q^k - y^k) - \Phi(\bar{x})\rangle \\ &= \langle q^k - \tilde{y}^k, \nabla\Phi(\bar{x})(s^k - \bar{x}) + O(\|s^k - \bar{x}\|^2) - \rho_k^{-1}(q^k - y^k)\rangle \\ &\le \langle q^k - \tilde{y}^k, \nabla\Phi(\bar{x})(s^k - \bar{x})\rangle - \rho_k^{-1}\|q^k - y^k\|^2 \\ &\quad + \|q^k - \tilde{y}^k\|O(\|s^k - \bar{x}\|^2) + {\rho_k}^{-1}\|y^k - \tilde{y}^k\| \cdot \|q^k - y^k\|. \end{aligned}\tag{5.19}$$

Combining (5.18) and (5.19) brings us to

$$\begin{aligned}\|q^k - y^k\|^2 &\le \rho_k(1 + \|q^k\| + \|q^k - \tilde{y}^k\|)O(\|s^k - \bar{x}\|^2) + \|y^k - \tilde{y}^k\| \cdot \|q^k - y^k\| \\ &\le (1 + \|y^k\|)O(\rho_k\|s^k - \bar{x}\|^2) + (2\|q^k - y^k\| + \|y^k - \tilde{y}^k\|)O(\rho_k\|s^k - \bar{x}\|^2) \\ &\quad + \|y^k - \tilde{y}^k\| \cdot \|q^k - y^k\| \\ &= o(\beta_k^2) + O(\beta_k)o(\beta_k) + o(\beta_k)O(\beta_k) = o(\beta_k^2), \end{aligned}$$

where the last estimate comes from Claim I, (5.11), (5.12), and boundedness of the sequence $\{y^k\}$. Dividing both sides by β_k^2 and passing then to the limit tell us via (5.11) that $\|\eta\|^2 = 0$, implying that $\eta = 0$. This is a contradiction with $(\zeta, \eta) \neq (0,0)$ and thus completes the proof. $\square$

To finish our discussion about assumption (b) in Theorem 2.1, we should point out that one more step is required to be taken. Comparing the right-hand side of (5.7) with the estimate in assumption (b) indicates that the residual function r in (5.7) should be replaced with the distance function to the solution set of the KKT system in (2.2). That requires to assume that Θ in (1.1) be an affine set, as recorded below.

Proposition 5.5. *Assume that $(\bar{x}, \bar{y})$ is a solution to the KKT system (2.2) with Θ therein being an affine set. Then there exist a positive constant κ and a neighborhood U of $(\bar{x}, \bar{y})$ such that*

$$r(x, \lambda) \le \kappa\big(\|x - \bar{x}\| + \operatorname{dist}(y, M_{\bar{x},\psi}(0,0))\big) \quad \textit{for all} \quad (x, y) \in U, \tag{5.20}$$

where $M_{\bar{x},\psi}(0,0)$ is defined in (3.19).

Proof. The claimed estimate was recently justified for the KKT system in (2.2) with g being CPLQ. Its proof, however, did not use the latter condition on g and works for any convex function g. Thus, we omit the proof and refer the readers to the latter result. $\square$

What remains is to show that assumption (c) of Theorem 2.1 holds automatically for the inexact ALM proposed in Algorithm 5.1. That will be done inside the proof of the next result in which we establish the wellposedness and local convergence of our inexact ALM.

Theorem 5.6. *Assume that $(\bar{x}, \bar{y})$ is a solution to the KKT system in (2.2) for which all the assumptions in Theorem 3.9 hold and that Θ in (1.1) is an affine set. Assume further that the second subderivative of g at $\Phi(\bar{x})$ for $\bar{y}$ is semi-stable. Then, there exist positive constants $\hat{c}$, $\bar{\rho}$, and ε_0 such that for any starting point $(x^0, y^0) \in \mathbb{B}_{\varepsilon_0}(\bar{x}, \bar{y})$ and any sequence $\{\rho_k\}_{k\in\mathbb{N}}$ with $\rho_k \geq \bar{\rho}$, there is a primal-dual sequence $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ satisfying (1.5) with $\epsilon_k = o(r(x^k, y^k))$ and the estimate (5.1), where r is the residual function defined by (3.21). Moreover, any such a sequence converges to $(\bar{x}, \widehat{y})$ for some $\widehat{y} \in M_{\bar{x},\psi}(0,0)$, and the rates of convergence of $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ to $(\bar{x}, \widehat{y})$ and of $\{\text{dist}\big((x^k, y^k), \{\bar{x}\} \times M_{\bar{x},\psi}(0,0)\big)\}_{k\in\mathbb{N}}$ to zero are Q-linear. Furthermore, if $\rho_k \to \infty$, the rates of convergence of both sequences are Q-superlinear.*

Proof. To justify the claims, we need to show that assumptions (a)-(c) in Theorem 2.1 are satisfied. We begin with assumption (a), namely the calmness of the solution mapping Υ from (5.2) at $(0,0)$ for $(\bar{x}, \bar{y})$. To achieve it, observe from Theorem 3.9 that the error-bound estimate in (3.22) holds. Take the neighborhood U of $(\bar{x}, \bar{y})$ from (3.22) and choose $\delta > 0$ such that $\mathbb{B}_\delta(\bar{x}, \bar{y}) \subset U$. By (3.22), we get

$$\Upsilon(0,0) \cap \mathbb{B}_\delta(\bar{x}, \bar{y}) = \big(\{\bar{x}\} \times M_{\bar{x},\psi}(0,0)\big) \cap \mathbb{B}_\delta(\bar{x}, \bar{y}). \tag{5.21}$$

Moreover, it follows from Proposition 5.2 that there are positive constants δ' and ℓ for which (5.3) holds. Shrinking δ' if necessary, we can assume that $\max\{\delta', \ell\delta'\} \leq \delta/2$. Thus, it results from (5.3) that

$$\begin{aligned}
\Upsilon(v,w) \cap \mathbb{B}_{\delta'}(\bar{x}, \bar{y}) &\subset \Big(\big(\{\bar{x}\} \times M_{\bar{x},\psi}(0,0)\big) + \ell\,\|(v,w)\|\mathbb{B}\Big) \cap \mathbb{B}_{\delta/2}(\bar{x}, \bar{y}) \\
&\subset \big(\big(\{\bar{x}\} \times M_{\bar{x},\psi}(0,0)\big) \cap \mathbb{B}_\delta(\bar{x}, \bar{y}) + \ell\,\|(v,w)\|\mathbb{B} \\
&= \Upsilon(0,0) \cap \mathbb{B}_\delta(\bar{x}, \bar{y}) + \ell\,\|(v,w)\|\mathbb{B} \subset \Upsilon(0,0) + \ell\,\|(v,w)\|\mathbb{B},
\end{aligned}$$

for all $(v,w) \in \delta'\mathbb{B}$, which confirms the calmness of Υ at $(0,0)$ for $(\bar{x}, \bar{y})$.

We turn next to assumption (b) in Theorem 2.1. Take the positive constants $\hat{\varepsilon}$ and $\hat{c}$ from Proposition 5.4 and assume without loss of generality that $\mathbb{B}_{\hat{\varepsilon}}(\bar{x}, \bar{y}) \subset U$ and $\hat{\varepsilon} \leq 2\delta$, where U is taken from (5.20) and δ comes from (5.21). Combining now the estimates in (5.7) and (5.20) implies that

$$\begin{aligned}
\|s - x\| + \|y_s - y\| \leq \hat{c}\, r(x,y) &\leq \hat{c}\kappa\big(\|x - \bar{x}\| + \text{dist}(y, M_{\bar{x},\psi}(0,0))\big) \\
&\leq 2\hat{c}\kappa\text{dist}((x,y), \{\bar{x}\} \times M_{\bar{x},\psi}(0,0))
\end{aligned}$$

for all $(x,y) \in \mathbb{B}_{\hat{\varepsilon}}(\bar{x}, \bar{y})$ such that $r(x,y) > 0$. If $(x,y) \in \mathbb{B}_{\hat{\varepsilon}}(\bar{x}, \bar{y})$ and $r(x,y) = 0$, one can set $s := x$, $y_s := y$ and observes that (s, y_s) is a solution to the KKT system in (2.2). It then follows from Proposition 2.2(b) that s and y_s satisfy (1.5) and (2.9), respectively. Combining this with (5.21) allows us to conclude that for any $(x,y) \in \mathbb{B}_{\hat{\varepsilon}}(\bar{x}, \bar{y})$ there is a pair (s, y_s) satisfying (1.5), (2.9), and the estimate

$$\|s - x\| + \|y_s - y\| \leq 2\hat{c}\kappa\text{dist}((x,y), \Upsilon(0,0)),$$

which confirms the validity of assumption (b) in Theorem 2.1.

Finally, to justify assumption (c) in Theorem 2.1, pick the positive constant $\bar{\rho}$ from Proposition 5.4, take any $(\tilde{x}, \tilde{y}) \in \mathbb{B}_{\hat{\varepsilon}}(\bar{x}, \bar{y})$, any $(x,y) \in \mathbf{X} \times \mathbf{Y}$ with $\|(x,y) - (\tilde{x}, \tilde{y})\| \leq \hat{c}\,\text{dist}((x,y), \Upsilon(0,0))$, and any $\rho \geq \bar{\rho}$. Choose also the tolerance parameter ϵ such that $\epsilon = o(r(x,y))$. For any

$(w_1, w_2) \in \Psi(x, y) - \mathcal{A}(x, y, \tilde{x}, \tilde{y}, \epsilon, \rho)$, where Ψ and $\mathcal{A}$ are taken from (2.10) and (2.11), we find $b \in \mathbb{B}$ such that

$$\begin{aligned}
\|(w_1, w_2)\| &= \|(\epsilon b, \rho^{-1}(y - \tilde{y})\| \le \epsilon + \rho^{-1}\|y - \tilde{y}\| \\
&\le o(r(x,y)) + \rho^{-1}\hat{c}\,\mathrm{dist}((x,y), \Upsilon(0,0)) \\
&= \frac{o(r(x,y))}{r(x,y)} r(x,y) + \rho^{-1}\hat{c}\,\mathrm{dist}((x,y), \Upsilon(0,0)) \\
&\le \frac{o(r(x,y))}{r(x,y)} \kappa\big(\|x - \bar{x}\| + \mathrm{dist}(y, M_{\bar{x},\psi}(0,0))\big) + \rho^{-1}\hat{c}\,\mathrm{dist}((x,y), \Upsilon(0,0)) \\
&\le \frac{o(r(x,y))}{r(x,y)} 2\kappa\big(\mathrm{dist}((x,y), \{\bar{x}\} \times M_{\bar{x},\psi}(0,0))\big) + \rho^{-1}\hat{c}\,\mathrm{dist}((x,y), \Upsilon(0,0)) \\
&= \big(\frac{2\kappa\, o(r(x,y))}{r(x,y)} + \rho^{-1}\hat{c}\big)\mathrm{dist}((x,y), \Upsilon(0,0)),
\end{aligned}$$

where the second inequality results from (5.20) and the last equality comes from (5.21). Defining the function $\omega : \mathbf{X} \times \mathbf{Y} \times \mathbf{X} \times \mathbf{Y} \times (0, \infty) \to \mathbf{R}_+$ by $\omega(x, y, \tilde{x}, \tilde{y}, \rho) = 2\kappa\, o(r(x,y))/r(x,y) + \rho^{-1}\hat{c}$, we obtain the second estimate in assumption (c) in Theorem 2.1. Since $\omega(x, y, \tilde{x}, \tilde{y}, \rho) \to 0$ as $(x, y, \tilde{x}, \tilde{y}, \rho) \to (\bar{x}, \bar{y}, \bar{x}, \bar{y}, \infty)$, the first estimate in assumption (c) also holds. Appealing now to the latter theorem proves all of convergence claims and hence completes the proof. □

We close this section by commenting on the established local convergence in Theorem 5.6. For the case of NLPs, this result reduces to [10, Theorem 3.4], where such a result was achieved for the first time for NLPs. Note that the calmness of multiplier mapping, assumed in Theorem 5.6, is automatically satisfied by Proposition 3.8 for NLPs. When g in (1.1) is CPLQ, Theorem 5.6 covers our recent result in [13, Theorem 5.7]. When $g = \delta_{\mathbf{S}^{\mathbf{n}}_+}$ and $\Theta = \mathbf{S}^{\mathbf{n}}$ in (1.1), which allow us to cover SDPs, the Q-linear convergence of the primal-dual sequence from the exact ALM was justified in [37] when the nondegeneracy condition and the strong second-order sufficient condition are satisfied. Moreover, it was recently shown in [39, Theorem 2] that if the multiplier $\bar{y}$ is taken from the relative interior of the set of Lagrange multipliers and the SOSC is satisfied, the inexact version of the ALM is Q-linearly convergent. Using a stronger version of the SOSC, the authors in [40] obtained the R-linear convergence for the primal sequence and Q-linear convergence for the dual sequence of the ALM without assuming the strict complementarity condition. Theorem 5.6 improves all of these for nonlinear SDPs by demonstrating that the SOSC alone suffices to achieve the Q-linear convergence of the primal-dual sequence for the inexact ALM. Finally, we should add that local convergence of the ALM was established recently in [32, Theorem 3.1] under the strong variational convexity, a condition that is equivalent to the strong second-order condition for NLPs, for the composite optimization problem in (1.1) with $\Theta = \mathbf{X}$. First, our analysis uses a weaker version of the SOSC. Indeed, the strong version of the SOSC amounts to assuming the classical SOSC in a neighborhood of a solution to the KKT system, which is much stronger than the version we used in this paper. Second, while we establish Q-linear convergence of the primal-dual sequence from the ALM, the latter result in [32] only presents the R-linear convergence of the primal sequence and Q-linear convergence of the dual sequence from this method. It is not hard to see that our primal-dual Q-linear convergence implies the R-linear convergence of the primal and dual sequences in the ALM. We should also mention that a modified version of the ALM, called the *safeguarded* ALM in which the y^k in (1.4) is replaced with a certain vector w^k chosen from a bounded set, was studied in [1, 5, 20], where it was shown that this modification has an interesting global convergence. The main motivation for this modification comes from the fact that the dual sequence, constructed by the ALM, may be unbounded in general.

6 Conclusion and Open Problems

Using the tools of second-order variational analysis, we were able to identify two main properties under which local convergence analysis of an inexact version of the ALM can be carried out. The first one demands a certain relationship between the graphical and semi-strict graphical derivatives of subgradient mappings, namely the equation in (3.14). The second is the semi-stability of second subderivatives, which was shown in Section 4 to be satisfied for various important classes of functions in variational analysis. While we proved that the equation in (3.14) holds when g therein is CPLQ or $\delta_{\mathcal{Q}}$, it remains as an open question whether the same conclusion holds for other classes of functions. In particular, it is important to explore it for $\mathcal{C}^2$-decomposable functions. A positive answer to the latter allows us to shorten the proof of Theorem 3.9 considerably. Another possibility for improving our results is to investigate whether the SOSC can be replaced with the *noncriticality of multipliers*; see [34, Definition 3.1]. Note that the noncriticality of multipliers is strictly weaker than the SOSC. It was proven in [17] that such an improvement is possible for NLPs provided that the strict complementarity condition is satisfied.

7 Appendix

In this section, we aim to justify two independent results from our local convergence analysis of the ALM, which play an important role in the process of justifying some results in Sections 3 and 4. We begin with a useful characterization of the *outer Lipschitzian* property of the multiplier mapping $M_{\bar{u},g}$ from (3.8) via the dual condition (3.10).

Proposition 7.1. *Assume that $g : \mathbf{Y} \to \overline{\mathbf{R}}$ is $\mathcal{C}^2$-decomposable at $\bar{u} \in \mathbf{Y}$ with representation (3.6) and that $\bar{y} \in \partial g(\bar{u})$ and $\bar{\mu} \in M_{\bar{u},g}(\bar{y}, 0)$, where $M_{\bar{u},g}$ is defined by (3.8). Then the following properties are equivalent.*

(a) *The set of Lagrange multipliers $M_{\bar{u},g}(\bar{y}, 0)$ is a singleton and there exist constants $\ell \geq 0$ and $\varepsilon > 0$ ensuring the error bound estimate*

$$\|\mu - \bar{\mu}\| \leq \ell\big(\|\nabla\Xi(\bar{u})^*\mu - \bar{y}\| + \operatorname{dist}\big(\Xi(\bar{u}), (\partial\vartheta)^{-1}(\mu)\big)\big) \ \textit{for all} \ \ \mu \in \mathbb{B}_\varepsilon(\bar{\mu}).$$

(b) *The dual condition (3.10) is satisfied.*

(c) *There exist positive numbers γ and κ such that the inclusion*

$$M_{\bar{u},g}(y, w) \cap \mathbb{B}_\gamma(\bar{\mu}) \subset \{\bar{\mu}\} + \kappa(\|y - \bar{y}\| + \|w\|)\mathbb{B} \ \textit{for all} \ \ (y, w) \in \mathbb{B}_\gamma(\bar{y}, 0)$$

holds.

(d) *There exist positive numbers γ and κ such that the inclusion*

$$M_{\bar{u},g}(y, w) \subset \{\bar{\mu}\} + \kappa(\|y - \bar{y}\| + \|w\|)\mathbb{B} \ \textit{for all} \ \ (y, w) \in \mathbb{B}_\gamma(\bar{y}, 0)$$

holds.

Proof. The equivalence of (a), (b), and (c) was established in [26, Theorem 4.1] when the convex function ϑ was the indicator function of a convex set. A closer look into the proof of the latter result shows that a similar argument works for ϑ. Clearly, we have the implication (d) $\Longrightarrow$ (c). Assume now that (b) holds. To prove (d), suppose by contradiction that there exist sequences $(y^k, w^k) \to (\bar{y}, 0)$ as $k \to \infty$ and the corresponding multipliers $\mu^k \in M_{\bar{u},g}(y^k, w^k)$ satisfying the inequality

$$\|\mu^k - \bar{\mu}\| > k(\|y^k - \bar{y}\| + \|w^k\|) \ \text{ for all } \ k \in \mathbb{N}. \tag{7.1}$$

Note that the equivalence of (a) and (b) yields $M_{\bar{u},g}(\bar{y},0)=\{\bar{\mu}\}$. Set $t_k:=\|\mu^k-\bar{\mu}\|$. We claim that the sequence $\{\mu^k\}_{k\in\mathbb{N}}$ is bounded. Indeed, it follows from [22, page 138] that ϑ is twice epi-differentiable at $\Xi(\bar{x})$ for $\bar{\mu}$ in the sense of [33, Definition 13.6(b)] and its second subderivative has a representation in the form

$$\mathrm{d}^2\vartheta(\Xi(\bar{u}),\bar{\mu})=\delta_{K_\vartheta(\Xi(\bar{u}),\bar{\mu})}. \tag{7.2}$$

Employing [33, Theorem 13.40], we arrive at

$$D(\partial\vartheta)(\Xi(\bar{u}),\bar{\mu})(w)=\partial\big(\tfrac{1}{2}\mathrm{d}^2\vartheta(\Xi(\bar{u}),\bar{\mu})\big)(w)\quad\text{for all}\ \ w\in K_\vartheta(\Xi(\bar{u}),\bar{\mu}). \tag{7.3}$$

In particular, this tells us that

$$D(\partial\vartheta)(\Xi(\bar{u}),\bar{\mu})(0)=\partial\big(\tfrac{1}{2}\mathrm{d}^2\vartheta(\Xi(\bar{u}),\bar{\mu})\big)(0)=N_{K_\vartheta(\Xi(\bar{u}),\bar{\mu})}(0)=K_\vartheta(\Xi(\bar{u}),\bar{\mu})^*. \tag{7.4}$$

Take $w\in K_\vartheta(\Xi(\bar{u}),\bar{\mu})$ and observe that $w\in\operatorname{dom}\mathrm{d}\vartheta(\Xi(\bar{u}))$. Since the inclusion $\operatorname{dom}\mathrm{d}\vartheta(\Xi(\bar{u}))\subset T_{\operatorname{dom}\vartheta}(\Xi(\bar{u}))$ always holds, we arrive at $K_\vartheta(\Xi(\bar{u}),\bar{\mu})\subset T_{\operatorname{dom}\vartheta}(\Xi(\bar{u}))$. This, coupled with (7.4), leads us to

$$N_{\operatorname{dom}\vartheta}(\Xi(\bar{u}))\subset K_\vartheta(\Xi(\bar{u}),\bar{\mu})^*=D(\partial\vartheta)(\Xi(\bar{u}),\bar{\mu})(0), \tag{7.5}$$

which implies via (3.10) that the BCQ in (3.9) holds. To justify the boundedness of $\{\mu^k\}_{k\in\mathbb{N}}$, assume by contradiction that it is unbounded. In this case, we can pass to a subsequence if necessary to ensure that the sequence $\{\mu^k/\|\mu^k\|\}_{k\in\mathbb{N}}$ converges to some ξ that $\|\xi\|=1$. Since $\mu^k\in M_{\bar{u},g}(y^k,w^k)$, it follows from (3.8) that $\nabla\Xi(\bar{u})^*\mu^k=y^k$, which in turn implies that $\xi\in\ker\nabla\Xi(\bar{u})^*$. Moreover, we have $\mu^k\in\partial\vartheta(\Xi(\bar{u})+w^k)\subset\partial\vartheta(0)$, by Proposition 3.4(b). Take $w\in\operatorname{dom}\vartheta$ and conclude via the definition of the subdifferential of convex functions that

$$\langle\mu^k,w\rangle\le\vartheta(w)-\vartheta(0)=\vartheta(w).$$

Dividing both sides by $\|\mu^k\|$ and passing to the limit, we obtain $\langle\mu,w\rangle\le 0$ for any $w\in\operatorname{dom}\vartheta$. This clearly tells us that $\xi\in N_{\operatorname{dom}\vartheta}(0)$, which together with $\xi\in\ker\nabla\Xi(\bar{u})^*$ contradicts (3.9) and hence proves the boundedness of $\{\mu^k\}_{k\in\mathbb{N}}$. Recall that $M_{\bar{u},g}(\bar{y},0)=\{\bar{\mu}\}$. Since $\{\mu^k\}_{k\in\mathbb{N}}$ is bounded, we can assume that it is convergent by passing to a subsequence if necessary. Thus, it follows from $\mu^k\in M_{\bar{u},g}(y^k,w^k)$ and $M_{\bar{u},g}(\bar{y},0)=\{\bar{\mu}\}$ that $\mu^k\to\bar{\mu}$. This immediately implies that $t_k=\|\mu^k-\bar{\mu}\|\to 0$. Observe then from (7.1) that $\|y^k-\bar{y}\|=o(t_k)$ and $\|w^k\|=o(t_k)$. Set $\eta^k:=(\mu^k-\bar{\mu})/t_k$ and assume without loss of generality that $\eta^k\to\eta$ with $\|\eta\|=1$. Thus, we have $\eta^k\in(\partial\vartheta(\Xi(\bar{u})+t_kw^k/t_k)-\bar{\mu})/t_k$. Passing to the limit as $k\to\infty$ leads us to $\eta\in D(\partial\vartheta)(\Xi(\bar{u}),\bar{\mu})(0)$, due to the fact that $w^k=o(t_k)$. Moreover, we have $\nabla\Xi(\bar{u})^*(\mu^k-\bar{\mu})=y^k-\bar{y}$, which together with $y^k-\bar{y}=o(t_k)$ results in $\eta\in\ker\nabla\Xi(\bar{u})^*$. The latter together contradict (3.10), since $\eta\neq 0$, and thus proves (d). □

Recall from [33, Definition 13.65] that a proper function $f:\mathbf{X}\to\overline{\mathbf{R}}$ is parabolically regular at $\bar{x}\in\operatorname{dom}f$ for $\bar{v}\in\partial f(\bar{x})$ if for any w such that $\mathrm{d}^2f(\bar{x},\bar{v})(w)<\infty$, there exist, among the sequences $t_k\searrow 0$ and $w_k\to w$ with $\Delta^2_{t_k}f(\bar{x},\bar{v})(w_k)\to\mathrm{d}^2f(\bar{x},\bar{v})(w)$, those with the additional property that $\limsup_{k\to\infty}\|w_k-w\|/t_k<\infty$. Parabolic regularity was recently studied extensively in [23–25] for different classes of sets and functions. Below, we show that $\mathcal{C}^2$-decomposable functions enjoy this property when the dual condition in (3.10) is satisfied. To this end, we need to recall the concept of parabolically epi-differentiability of functions from [33, Definition 13.59]. Given $f:\mathbf{X}\to\overline{\mathbf{R}}$, $\bar{x}\in\mathbf{X}$ with $f(\bar{x})$ finite, and $w\in\mathbf{X}$ with $\mathrm{d}f(\bar{x})(w)$ finite, the *parabolic subderivative* of f at $\bar{x}$ for w with respect to z is defined by

$$\mathrm{d}^2f(\bar{x})(w\,|\,z)=\liminf_{\substack{t\searrow 0\\ z'\to z}}\Delta^2_tf(\bar{x})(w\,|\,z'),$$

where $\Delta_t^2 f(\bar{x})(w \,|\, z') := (f(\bar{x}+tw+\frac{1}{2}t^2 z') - f(\bar{x}) - t\mathrm{d}f(\bar{x})(w))/\frac{1}{2}t^2$. The function f is called *parabolically epi-differentiable* at $\bar{x}$ for w if

$$\operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \,|\, \cdot) = \big\{z \in \mathbf{X} \,|\, \mathrm{d}^2 f(\bar{x})(w \,|\, z) < \infty\big\} \neq \emptyset,$$

and for every $z \in \mathbf{X}$ and every sequence $t_k \searrow 0$ there exists a sequences $z^k \to z$ such that $\Delta_{t_k}^2 f(\bar{x})(w \,|\, z^k) \to \mathrm{d}^2 f(\bar{x})(w \,|\, z)$.

Theorem 7.2. *Assume that $g : \mathbf{Y} \to \overline{\mathbf{R}}$ is $\mathcal{C}^2$-decomposable at $\bar{u} \in \mathbf{Y}$ with representation (3.6) and that $\bar{y} \in \partial g(\bar{u})$, $\bar{\mu} \in M_{\bar{u},g}(\bar{y}, 0)$, where $M_{\bar{u},g}$ is defined by (3.8), and the dual condition in (3.10) is satisfied. Then, the following properties hold.*

(a) *The function g is parabolically regular at $\bar{u}$ for $\bar{y}$ and its second subderivative can be calculated by*

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

(b) *The mapping ∂g is proto-differentiable at $\bar{u}$ for $\bar{y}$ and its proto-derivative can be calculated by*

$$D(\partial g)(\bar{u}, \bar{y})(w) = \nabla^2 \langle \bar{\mu}, \Xi\rangle(\bar{x}) w + \nabla \Xi(\bar{u})^* D(\partial \vartheta)(\Xi(\bar{u}), \bar{y})(\nabla \Xi(\bar{u}) w) \quad \text{for all } w \in \mathbf{Y}.$$

Proof. To prove (a), we first show that ϑ is parabolically regular at $\Xi(\bar{u}) = 0$ for $\bar{\mu}$, where ϑ and Ξ are taken from (3.6). To this end, take $w \in \mathbf{Z}$ such that $\mathrm{d}^2 \vartheta(\Xi(\bar{u}), \bar{\mu})(w) < \infty$, which via (7.2) reads as $w \in K_\vartheta(\Xi(\bar{u}), \bar{\mu})$. By the sublinearity of ϑ, we can conclude from the latter inclusion that $\langle \bar{\mu}, w\rangle = \mathrm{d}\vartheta(\Xi(\bar{u}))(w) = \vartheta(w)$. This yields

$$\lim_{t \searrow 0} \Delta_t^2 \vartheta(\Xi(\bar{u}), \bar{\mu})(w) = \lim_{t \searrow 0} \frac{t(\vartheta(w) - \langle \bar{\mu}, w\rangle)}{\frac{1}{2}t^2} = 0 = \mathrm{d}^2 \vartheta(\Xi(\bar{u}), \bar{\mu})(w).$$

Take $t_k \searrow 0$ and set $w^k := w$ for any $k \in \mathbb{N}$. According to the above relationships, we have $\Delta_{t_k}^2 \vartheta(\Xi(\bar{u}), \bar{\mu})(w^k) \to \mathrm{d}^2 \vartheta(\Xi(\bar{u}), \bar{\mu})(w)$, which proves that ϑ is parabolically regular at $\Xi(\bar{u})$ for $\bar{\mu}$. We are going next to show ϑ is parabolically epi-differentiable at $\Xi(\bar{u})$ for any $w \in K_\vartheta(\Xi(\bar{u}), \bar{\mu})$. To achieve it, take $w \in K_\vartheta(\Xi(\bar{u}), \bar{\mu})$ and conclude for any $z \in \mathbf{Z}$ that

$$\Delta_t^2 \vartheta(\Xi(\bar{u}))(w \,|\, z) = \frac{\vartheta(tw + \frac{1}{2}t^2 z) - t\mathrm{d}\vartheta(\Xi(\bar{u}))(w)}{\frac{1}{2}t^2} = \frac{\vartheta(w + \frac{1}{2}tz) - \vartheta(w)}{\frac{1}{2}t} = \Delta_{t/2} \vartheta(w)(z). \quad (7.6)$$

Since ϑ is proper and convex, it follows from [33, Example 7.27] that it is always epi-differentiable in the sense of [33, Definition 7.23]. Thus, for any $z \in \mathbf{Z}$ and $t_k \searrow 0$, there exists $z^k \to z$ such that $\Delta_{t_k} \vartheta(w)(z^k) \to \mathrm{d}\vartheta(w)(z)$. Combining this with (7.6) tells us that for any $z \in \mathbf{Z}$ and $t_k \searrow 0$, we can find $z^k \to z$ such that $\Delta_{t_k}^2 \vartheta(\Xi(\bar{u}))(w \,|\, z^k) \to \mathrm{d}^2 \vartheta(\Xi(\bar{u}))(w \,|\, z)$. Moreover, it is not hard to see that $\mathrm{d}\vartheta(w)(0) < \infty$. This, coupled with (7.6), leads us to $0 \in \operatorname{dom} \mathrm{d}^2 \vartheta(\Xi(\bar{u}))(w \,|\, \cdot)$, since

$$\operatorname{dom} \mathrm{d}^2 \vartheta(\Xi(\bar{u}))(w \,|\, \cdot) = \big\{z \in \mathbf{Z} \,|\, \mathrm{d}^2 \vartheta(\Xi(\bar{u}))(w \,|\, z) < \infty\big\} = \big\{z \in \mathbf{Z} \,|\, \mathrm{d}\vartheta(w)(z) < \infty\big\}.$$

These demonstrate that ϑ is parabolically epi-differentiable at $\Xi(\bar{u})$ for $w \in K_\vartheta(\Xi(\bar{u}), \bar{\mu})$.

From the proof of Proposition 7.1, we know that the dual condition in (3.10) implies the BCQ condition in (3.9) and uniqueness of the Lagrange multiplier $\bar{\mu}$ in $M_{\bar{u},g}(\bar{y}, 0)$. Appealing now to [24, Theorem 5.4 and Remark 5.3], we obtain both assertions in (a). To prove (b), observe from (7.2) that $\operatorname{dom} \mathrm{d}^2 \vartheta(\Xi(\bar{u}), \bar{\mu}) = K_\vartheta(\Xi(\bar{u}), \bar{\mu})$ and from (7.4) that the dual condition in (3.10) can be equivalently expressed as

$$N_{K_\vartheta(\Xi(\bar{u}), \bar{\mu})}(0) \cap \ker \nabla \Xi(\bar{u})^* = \{0\}.$$

Thus, it follows from the chain rule for subdifferentials from [33, Theorem 10.6] that, for any $w \in \mathbf{Y}$ such that $\nabla\Xi(\bar{u})w \in K_\vartheta(\Xi(\bar{u}), \bar{\mu})$, we have

$$\begin{aligned}\partial_w\big(\tfrac{1}{2}\mathrm{d}^2\vartheta(\Xi(\bar{u}), \bar{\mu})\big)(\nabla\Xi(\bar{u})w) &= \nabla\Xi(\bar{u})^*\partial\big(\tfrac{1}{2}\mathrm{d}^2\vartheta(\Xi(\bar{u}), \bar{\mu})\big)(\nabla\Xi(\bar{u})w)\\ &= \nabla\Xi(\bar{u})^* D(\partial\vartheta)(\Xi(\bar{u}), \bar{\mu})(\nabla\Xi(\bar{u})w),\end{aligned}$$

where the last equality comes from (7.3). This, coupled with [24, Corollary 3.9] and the formula for $\mathrm{d}^2 g(\bar{u}, \bar{y})$ from (a), implies that ∂g is proto-differentiable at $\bar{u}$ for $\bar{y}$ and that

$$\begin{aligned}D(\partial g)(\bar{u}, \bar{y})(w) = \partial\big(\tfrac{1}{2}\mathrm{d}^2 g(\bar{u}, \bar{y})\big)(w) &= \tfrac{1}{2}\nabla_w\langle\bar{\mu}, \nabla^2\Xi(\bar{u})(w, w)\rangle + \partial_w\big(\tfrac{1}{2}\mathrm{d}^2\vartheta(\Xi(\bar{u}), \bar{\mu})\big)(\nabla\Xi(\bar{u})w)\\ &= \nabla^2\langle\bar{\mu}, \Xi\rangle(\bar{x})w + \nabla\Xi(\bar{u})^* D(\partial\vartheta)(\Xi(\bar{u}), \bar{y})(\nabla\Xi(\bar{u})w),\end{aligned}$$

for all $w \in \mathbf{Y}$, which proves (b) and hence ends the proof. □

References

[1] Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: On augmented Lagrangian methods with general lower-level constraints. SIAM J. Optim. 18, 1286–1309 (2007)

[2] Aragón Artacho, F.J., Geoffroy, M.H.: Characterization of metric regularity of subdifferentials. J. Convex Anal. 15, 365–380 (2008)

[3] Bauschke, H.H., Borwein, J.M., Li, W.: Strong conical hull intersection property, bounded linear regularity, Jameson's property (G), and error bounds in convex optimization. Math. Program. 86, 135–160 (1999)

[4] Beck, A.: First-Order Methods in Optimization. MOS-SIAM, Philadelphia (2017)

[5] Birgin, E.G., Haeser. G., Ramos, A.: Augmented Lagrangians with constrained subproblems and convergence to second-order stationary points. Comput. Optim. Appl. 69, 51–75 (2018)

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

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

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

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

[10] Fernández, D., Solodov, M.V.: Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition. SIAM J. Optim. 22, 384–407 (2012)

[11] Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. 94, 91–124 (2002)

[12] Hang, N.T.V., Mordukhovich, B.S., Sarabi, M.E.: Second-order variational analysis in second-order cone programming. Math. Program. 180, 75–106 (2020)

[13] 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, 2665–2694 (2021)

[14] Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. 4, 303–320 (1969)

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

[16] Izmailov, A.F., Kurennoy, A.S.: Abstract Newtonian frameworks and their applications. SIAM J. Optim. 23, 2369–2396 (2013)

[17] Izmailov, A.F., Kurennoy, A.S., Solodov, M.V.: Local convergence of the method of multipliers for variational and optimization problems under the noncriticality assumption. Comput. Optim. Appl. 60, 111–140 (2015)

[18] Izmailov, A.F., Solodov, M.V.: Newton-Type Methods for Optimization and Variational Problems. Springer Cham, Switzerland (2014)

[19] Kan, C., Song, W.: Augmented Lagrangian duality for composite optimization problems. J. Optim. Theory Appl. 165, 763–784 (2015)

[20] Kanzow, C., Steck, D.: Improved local convergence results for augmented Lagrangian methods in $\mathcal{C}^2$-cone reducible constrained optimization. Math. Program. 177, 425–438 (2019)

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

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

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

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

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

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

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

[28] Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) Optimization, pp. 283–298. Academic Press, New York (1969)

[29] Robinson, S.M.: Some continuity properties of polyhedral multifunctions. Math. Program. Stud. 14, 206–214, (1981)

[30] Rockafellar, R.T.: A dual approach to solving nonlinear programming problems by unconstrained optimization. Math. Program. 5, 354–373 (1973)

[31] Rockafellar, R.T.: Augmented Lagrangians and hidden convexity in sufficient conditions for local optimality. Math. Program. 198, 159–194 (2023)

[32] Rockafellar, R.T.: Convergence of augmented Lagrangian methods in extensions beyond nonlinear programming. Math. Program. 199, 375–420 (2023)

[33] Rockafellar, R.T., Wets, R.J-B.: Variational Analysis. Springer-Verlag 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)

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

[36] Shapiro, A., Sun, J.: Some properties of the augmented Lagrangian in cone constrained optimization. Math. Oper. Res. 29, 479–491 (2004)

[37] Sun, D., Sun, J., Zhang, L.: The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. Math. Program. 114, 349–391 (2008)

[38] Torki, M.: Second-order directional derivatives of all eigenvalues of a symmetric matrix. Nonlinear Anal. 46, 1133–1150 (2001)

[39] Wang, S., Ding, C.: Local convergence analysis of augmented Lagrangian method for nonlinear semidefinite programming. Arxiv preprint arXiv:2110.10594 (2022)

[40] Wang, S., Ding, C., Zhang, Y., Zhao, X.: Strong variational sufficiency for nonlinear semidefinite programming and its implications. Arxiv preprint arXiv:2210.04448 (2022)