AUGMENTED LAGRANGIAN METHOD FOR SECOND-ORDER CONE PROGRAMS UNDER SECOND-ORDER SUFFICIENCY

NGUYEN T. V. HANG[1], BORIS S. MORDUKHOVICH[2] and M. EBRAHIM SARABI[3]

Abstract. This paper addresses problems of second-order cone programming important in optimization theory and applications. The main attention is paid to the augmented Lagrangian method (ALM) for such problems considered in both exact and inexact forms. Using generalized differential tools of second-order variational analysis, we formulate the corresponding version of second-order sufficiency and use it to establish, among other results, the uniform second-order growth condition for the augmented Lagrangian. The latter allows us to justify the solvability of subproblems in the ALM and to prove the linear primal-dual convergence of this method.
Key words. Augmented Lagrangian method, second-order cone programming, second-order sufficiency, variational analysis, linear convergence
Mathematics Subject Classification (2000) 90C99, 49J52, 49J53

1 Introduction

In this paper we consider the class of constrained optimization problems belonging to *second-order cone programming* (SOCP) that are given in the form:

$$\text{minimize} \ \ f(x) \ \ \text{subject to} \ \ \Phi(x) \in \mathcal{Q}, \tag{1.1}$$

where both mappings $f \colon \mathbb{R}^n \to \mathbb{R}$ and $\Phi \colon \mathbb{R}^n \to \mathbb{R}^{m+1}$ are twice continuously differentiable ($\mathcal{C}^2$-smooth) around the reference points, and where the underlying set $\mathcal{Q}$ is the second-order (Lorentz, ice-cream) cone in $\mathbb{R}^{m+1}$ defined by

$$\mathcal{Q} := \big\{ y = (y_0, y_r) \in \mathbb{R} \times \mathbb{R}^m \,\big|\, \|y_r\| \leq y_0 \big\}. \tag{1.2}$$

Problems of this type (SOCPs for brevity) constitute a remarkable subclass of nonpolyhedral conic programs that has been well recognized in constrained optimization and various applications; see, e.g., [5, 9, 10, 33] and the references therein.

Our main attention is paid to developing a numerical method to solve (1.1) that involves the *augmented Lagrangian* $\mathscr{L} \colon \mathbb{R}^n \times \mathbb{R}^{m+1} \times (0, \infty) \to \mathbb{R}$ associated with this problem, which is defined by

$$\mathscr{L}(x, \lambda, \rho) := f(x) + \frac{\rho}{2} \text{dist}^2 \big(\Phi(x) + \rho^{-1}\lambda; \mathcal{Q} \big) - \frac{1}{2}\rho^{-1}\|\lambda\|^2, \ \ (x, \lambda, \rho) \in \mathbb{R}^n \times \mathbb{R}^{m+1} \times (0, \infty), \tag{1.3}$$

where $\lambda \in \mathbb{R}^{m+1}$ is a (vector) multiplier, and where $\rho > 0$ is a penalty parameter of $\mathscr{L}$. The principal idea of the *augmented Lagrangian method* (ALM) for (1.1) is to solve a sequence of

---

unconstrained problems which objectives are defined by the augmented Lagrangian (1.3) at a given multiplier-parameter pair (λ, ρ); namely,

$$\text{minimize} \quad \mathscr{L}(x, \lambda, \rho) \quad \text{over} \quad x \in \mathbb{R}^n. \tag{1.4}$$

This means that, given a multiplier λ and a penalty parameter ρ, the ALM solves the unconstrained problem (1.4) for the primal variable x and uses the obtained value to update both the multiplier and penalty parameter in the next iteration.

The ALM was first proposed independently by Hestenes and Powell for nonlinear programming problems (NLPs) with equality constraints [21,36] and was originally known as the method of multipliers. For the latter framework, Powell observed in [36] that the ALM converges locally with an arbitrarily linear rate if one started the method with a sufficiently high penalty factor (but without the requirement of driving the penalty parameter to infinity) and from a point sufficiently close to a primal-dual pair that satisfies the standard second-order sufficient conditions (SOSC). This is an appealing feature of the ALM, since it provides a numerical stability that cannot be achieved in the usual smooth penalty method.

The ALM was largely extended to various settings of NLPs as well as convex programming with both equality and inequality constraints by Rockafellar [38–40]; see also the monographs [7, 35,45] and the references therein. Further theoretical and practical developments for this method were achieved in [1–4] for NLPs. Furthermore, a modified version of the ALM was suggested in [1] to alleviate the possibility of unboundedness of dual sequences in the standard ALM; see [8] for a detailed discussion on this modification of the ALM and its successful implementation.

The classical results for the linear convergence of the ALM in NLP framework impose the SOSC, the linear independence constraint qualification (LICQ), and the strict complementarity condition, which all together guarantee the uniqueness of the primal solution as well as the corresponding dual solution/multiplier.

More recently, the study of the ALM has been growing with important theoretical developments. On one hand, various attempts have been made to relax the restrictive assumptions for the convergence of this method in the NLP settings. In such a framework, Fernández and Solodov achieved in [15] a remarkable progress for NLPs by proving that the linear convergence of the primal-dual sequence in the ALM can be ensured if the SOSC alone is satisfied. This result significantly improved the classical ones for NLPs by verifying that neither the LICQ nor the strict complementarity condition is required for local convergence analysis of the ALM. A further improvement was obtained in Izmailov et al. [23] by showing that the conventional SOSC utilized in [15] can be replaced by the noncriticality of Lagrange multipliers for problems with equality constraints. On the other hand, the ALM has been studied for other major classes of constrained optimization including SOCPs [27] and semidefinite programming problems (SDPs) [47]. For $\mathcal{C}^2$-cone reducible problems of conic programming (in the sense of Bonnans and Shapiro [10]), Kanzow and Steck [25,26] established the linear convergence of the primal-dual sequence generated by modified versions of the ALM under the SOSC and strong Robinson constraint qualification; the latter yields that the Lagrange multiplier is unique. However, the solvability of subproblems in the ALM was not addressed in these papers. We also refer the reader to the paper by Cui et al. [11] and the bibliography therein for recent developments on the ALM for particular classes of convex composite problems of conic programming.

The major goal of this paper is to develop both *exact* and *inexact* versions of the ALM for SOCPs under fairly *mild assumptions*. We aim first at establishing the *solvability* and *Lipschitzian stability* of the ALM subproblems by imposing merely the corresponding *SOSC* for (1.1) in the general case of *nonunique* Lagrange multipliers. Having this, we verify a local *primal-dual convergence* of iterates with an arbitrary *linear rate* by assuming in addition the *uniqueness* of multipliers. Similarly to Fernández and Solodov [15], our approach revolves around

the *second-order growth condition* for the augmented Lagrangian (1.3). To the best of our knowledge, the origin of such a second-order growth condition for NLPs goes back to Rockafellar in [43, Theorem 7.4] from which [15] significantly benefits. However, in contrast to [43], [15] as well as to the vast majority of other publications on numerical optimization, we achieve our goal for (1.1) by employing the concepts of the *second subderivative* and *twice epi-differentiability* of extended-real-valued functions in the framework of second-order variational analysis.

The outline of the paper is as follows. In Section 2 we recall the notions of variational analysis and generalized differentiation with some *preliminary material* used in the paper.

The main result of Section 3, which is of its own interest, provides an *error bound* estimate for the canonically perturbed KKT system associated with (1.1) under the SOSC and a certain *calmness* property of the multiplier mapping with respect to perturbations that automatically holds for NLPs. We also present here an example showing that the imposed calmness property is essential for the validity of the error bound in the SOCP setting and then discuss efficient conditions ensuring the fulfillment of this calmness for nonpolyhedral SOCPs.

Section 4 conducts a comprehensive *second-order variational analysis* of the augmented Lagrangian (1.3) associated with the second-order cone program (1.1). Based on the obtained precise computation of the second subderivative of (1.3), we characterize here the *second-order growth* condition for (1.3) via the SOSC and then establish its *uniform* counterpart needed in the general case of nonunique Lagrange multipliers.

The concluding Section 5 provides a detailed *solvability, stability,* and *local convergence analysis* of the suggested ALM algorithm for SOCPs that strongly exploits the SOSC and obtained second-order growth conditions. Our analysis includes the proof of solvability of the ALM subproblems in both exact and inexact versions and then establishes the linear convergence of primal-dual iterates to the designated solution of the KKT systems under the SOSC by using the established *robust isolated calmness* and *upper Lipschitzian* properties of the corresponding perturbed multiplier mappings. In this way we obtain explicit relationships between the constants involved in the algorithm and the imposed assumptions on the given data.

Throughout the paper we use the standard notation and terminology of variational analysis and conic programming; see, e.g., [10, 32, 44]. Recall that $\mathbb{B}$ and $\mathbb{S}$ stand for the closed unit ball and unit sphere, respectively, of the space in question, that $\mathbb{B}_\gamma(x) := x + \gamma\mathbb{B}$ is the closed ball centered at x with radius $\gamma > 0$, that $\mathbb{N} := \{1, 2, \ldots\}$, and that A^* indicates the matrix transposition. Given a nonempty set $\Omega \subset \mathbb{R}^n$, the symbols $\operatorname{int}\Omega$, $\operatorname{ri}\Omega$, $\operatorname{bd}\Omega$, and $\Omega^\perp$ signify its interior, relative interior, boundary, and orthogonal complement space, respectively. The indicator function of Ω is defined by $\delta_\Omega(x) := 0$ for $x \in \Omega$ and $\delta_\Omega(x) := \infty$ otherwise, $\operatorname{dist}(x; \Omega)$ signifies the distance between $x \in \mathbb{R}^n$ and the set Ω, and the projection of x onto Ω is denoted by $\Pi_\Omega(x)$. As in (1.2), we often decompose a vector $y \in \mathcal{Q} \subset \mathbb{R}^{m+1}$ into $y = (y_0, y_r)$ with $y_0 \in \mathbb{R}$ and $y_r \in \mathbb{R}^m$. Taking this decomposition into account, denote $\widetilde{y} := (-y_0, y_r)$. Similarly, for the mapping $\Phi\colon \mathbb{R}^n \to \mathbb{R}^{m+1}$ with $\Phi = (\Phi_0, \ldots, \Phi_m)$, we denote by $\widetilde{\Phi}(x)$ the vector $(-\Phi_0(x), \Phi_r(x))$ for any $x \in \mathbb{R}^n$.

2 Preliminaries from Variational Analysis

We start this section with recalling those constructions of variational analysis and generalized differentiation, which are broadly employed in what follows; see [10, 32, 44] for more details and references. Given a set $\Theta \subset \mathbb{R}^n$ with $\bar{x} \in \Theta$, the *tangent cone* to Θ at $\bar{x}$ is defined by

$$T_\Theta(\bar{x}) := \big\{w \in \mathbb{R}^n \mid \exists\, t_k \downarrow 0,\ w^k \to w \ \text{ as } \ k \to \infty \ \text{ with } \ \bar{x} + t_k w^k \in \Theta\big\}. \tag{2.1}$$

If Θ is convex, the *normal cone* to Θ at $\bar{x} \in \Theta$ in the sense of convex analysis is

$$N_\Theta(\bar{x}) := \big\{v \in \mathbb{R}^n \mid \langle v, x - \bar{x}\rangle \leq 0 \ \text{ for all } x \in \Theta\big\}.$$

3

In the case where $\Theta = \mathcal{Q}$, the second-order cone (1.2), we get, respectively, the expressions

$$T_{\mathcal{Q}}(y) = \begin{cases} \mathbb{R}^{m+1} & \text{if } y \in \text{int } \mathcal{Q}, \\ \mathcal{Q} & \text{if } y = 0, \\ \{y' \in \mathbb{R}^{m+1} \mid \langle \widetilde{y}, y' \rangle \leq 0\} & \text{if } y \in (\text{bd } \mathcal{Q}) \setminus \{0\}, \end{cases}$$

$$N_{\mathcal{Q}}(y) = \begin{cases} \{0\} & \text{if } y \in \text{int } \mathcal{Q}, \\ -\mathcal{Q} & \text{if } y = 0, \\ \mathbb{R}_+ \widetilde{y} & \text{if } y \in (\text{bd } \mathcal{Q}) \setminus \{0\}. \end{cases} \tag{2.2}$$

Given further an extended-real-valued function $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}} := (-\infty, \infty]$, its domain and epigraph are defined, respectively, by

$$\text{dom } \varphi := \{x \in \mathbb{R}^n \mid \varphi(x) < \infty\} \quad \text{and} \quad \text{epi } \varphi := \{(x, \mu) \in \mathbb{R}^{n+1} \mid \mu \geq \varphi(x)\}.$$

Considering next a set-valued mapping $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ with its domain and graph given by

$$\text{dom } F := \{x \in \mathbb{R}^n \mid F(x) \neq \emptyset\} \quad \text{and} \quad \text{gph } F := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid y \in F(x)\},$$

we define the *graphical derivative* of F at $(\bar{x}, \bar{y}) \in \text{gph } F$ via the tangent cone (2.1) by

$$DF(\bar{x}, \bar{y})(u) := \{v \in \mathbb{R}^m \mid (u, v) \in T_{\text{gph } F}(\bar{x}, \bar{y})\}, \quad u \in \mathbb{R}^n. \tag{2.3}$$

A mapping $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is called *calm* at $(\bar{x}, \bar{y}) \in \text{gph } F$ if there exist $\ell \geq 0$ and neighborhoods U of $\bar{x}$ and V of $\bar{y}$ for which

$$F(x) \cap V \subset F(\bar{x}) + \ell \|x - \bar{x}\| \mathbb{B} \quad \text{whenever } x \in U. \tag{2.4}$$

It is said that F has the *isolated calmness* property at $(\bar{x}, \bar{y}) \in \text{gph } F$ if (2.4) holds with the replacement of $F(\bar{x})$ by $\{\bar{y}\}$ on the right-hand side therein. Furthermore, F has the *robust isolated calmness* property at $(\bar{x}, \bar{y})$ if

$$F(x) \cap V \subset \{\bar{y}\} + \ell \|x - \bar{x}\| \mathbb{B} \quad \text{with } F(x) \cap V \neq \emptyset \text{ for all } x \in U. \tag{2.5}$$

Properties of this type go back to Robinson [37] who introduced the *upper Lipschitzian* version of calmness corresponding to (2.4) with $V = \mathbb{R}^m$. Similarly to (2.5), we say that F has the *robust isolated upper Lipschitzian* property if (2.5) holds with $V = \mathbb{R}^m$. It is well known that (2.4) is equivalent to the metric subregularity of the inverse mapping F^{-1} at $(\bar{y}, \bar{x})$. These "one point" properties are more subtle and essentially less investigated than their robust "two-points" counterparts (as metric regularity and Lipschitz-like/Aubin ones), while their importance for optimization theory, numerical algorithms, and applications has been broadly recognized in the literature; see, e.g., [11–13, 17, 20, 24, 30, 32, 48] with the references and discussions therein.

Turning now to the constructions of second-order variational analysis, for a function $\varphi\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, define the parametric family of *second-order difference quotients* at $\bar{x}$ for $\bar{v} \in \mathbb{R}^n$ by

$$\Delta_t^2 \varphi(\bar{x}, \bar{v})(w) = \frac{\varphi(\bar{x} + tw) - \varphi(\bar{x}) - t\langle \bar{v}, w \rangle}{\frac{1}{2} t^2} \quad \text{with } w \in \mathbb{R}^n, \ t > 0.$$

If $\varphi(\bar{x})$ is finite, the *second subderivative* of φ at $\bar{x}$ for $\bar{v}$ and w is defined by

$$\text{d}^2 \varphi(\bar{x}, \bar{v})(w) = \liminf_{\substack{t \downarrow 0 \\ w' \to w}} \Delta_t^2 \varphi(\bar{x}, \bar{v})(w'). \tag{2.6}$$

4

Following [44, Definition 13.6], a function $\varphi \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be *twice epi-differentiable* at $\bar{x}$ for $\bar{v}$ if the sets $\mathrm{epi}\,\Delta_t^2 \varphi(\bar{x}, \bar{v})$ converge to $\mathrm{epi}\,\mathrm{d}^2\varphi(\bar{x}, \bar{v})$ as $t \downarrow 0$. If in addition the second subderivative is a proper function (i.e., does not take the value $-\infty$ and is finite at some point), then we say that φ is *properly* twice epi-differentiable at $\bar{x}$ for $\bar{v}$. The twice epi-differentiability of φ at $\bar{x}$ for $\bar{v}$ can be understood equivalently by [44, Proposition 7.2] as that for every $w \in \mathbb{R}^n$ and every sequence $t_k \downarrow 0$ there exists a sequence $w^k \to w$ with

$$\Delta_{t_k}^2 \varphi(\bar{x}, \bar{v})(w^k) \to \mathrm{d}^2\varphi(\bar{x}, \bar{v})(w).$$

Twice epi-differentiability, together with a precise calculation of the second subderivative (2.6) of the augmented Lagrangian (1.3) associated with (1.1), plays a major role in our developments. This property was introduced by Rockafellar in [41] who verified it for *fully amenable* compositions. Quite recently [28, 29, 31], the class of extended-real-valued functions satisfying this property has been dramatically enlarged by showing that twice epi-differentiability holds under *parabolic regularity*, which covers the SOCP setting; see more details in the cited papers.

When φ is the indicator function to the second-order cone $\mathcal{Q}$, it is shown in [19, Theorem 3.1] that $\varphi = \delta_{\mathcal{Q}}$ is properly twice epi-differentiable at any $\bar{x} \in \mathcal{Q}$ for every $\bar{v} \in N_{\mathcal{Q}}(\bar{x})$ and its second subderivative (2.6) is calculated by the precise formula

$$\mathrm{d}^2\delta_{\mathcal{Q}}(\bar{x}, \bar{v})(w) = \begin{cases} \delta_{K_{\mathcal{Q}}(\bar{x}, \bar{v})}(w) & \text{if } \bar{x} \in (\mathrm{int}\,\mathcal{Q}) \cup \{0\}, \\ \frac{\|\bar{v}\|}{\|\bar{x}\|}\left(\|w_r\|^2 - w_0^2\right) + \delta_{K_{\mathcal{Q}}(\bar{x}, \bar{v})}(w) & \text{if } \bar{x} \in (\mathrm{bd}\,\mathcal{Q}) \setminus \{0\}, \end{cases} \tag{2.7}$$

where $w = (w_0, w_r) \in \mathbb{R} \times \mathbb{R}^m$, and where $K_{\mathcal{Q}}(\bar{x}, \bar{v}) := T_{\mathcal{Q}}(\bar{x}) \cap \{\bar{v}\}^\perp$ stands for the *critical cone* to the set $\mathcal{Q}$ at $\bar{x} \in \mathcal{Q}$ for any normal direction $\bar{v} \in N_{\mathcal{Q}}(\bar{x})$.

The *Karush-Kuhn-Tucker* (KKT) optimality system associated with (1.1) is given by

$$\nabla_x L(x, \lambda) = \nabla f(x) + \nabla\Phi(x)^* \lambda = 0, \quad \lambda \in N_{\mathcal{Q}}\big(\Phi(x)\big), \tag{2.8}$$

where $L(x, \lambda) := f(x) + \langle \lambda, \Phi(x) \rangle$ is the (standard) *Lagrangian* of problem (1.1) with $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^{m+1}$. For any $\bar{x} \in \mathbb{R}^n$, define the set of *Lagrange multipliers* associated with $\bar{x}$ by

$$\Lambda(\bar{x}) := \left\{ \lambda \in \mathbb{R}^{m+1} \,\big|\, \nabla_x L(\bar{x}, \lambda) = 0,\ \lambda \in N_{\mathcal{Q}}\big(\Phi(\bar{x})\big) \right\}. \tag{2.9}$$

Our major attention to the second subderivative (2.6) in this paper is due to its ability to characterize the second-order growth condition in (1.1) and thus to provide a second-order sufficient condition for strict local minimizers of this problem. To this end, we recall the corresponding result from [29, Proposition 7.3] justifying such an application for SOCPs.

Proposition 2.1 (SOSC yields second-order growth). *Let $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^{m+1}$ be a solution to the KKT system* (2.8), *and let the second-order sufficient condition*

$$\begin{cases} \langle \nabla_{xx}^2 L(\bar{x}, \bar{\lambda}) w, w \rangle + \mathrm{d}^2 \delta_{\mathcal{Q}}\big(\Phi(\bar{x}), \bar{\lambda}\big)\big(\nabla\Phi(\bar{x}) w\big) > 0 \\ \text{for all } w \in \mathbb{R}^n \setminus \{0\} \text{ with } \nabla\Phi(\bar{x}) w \in K_{\mathcal{Q}}\big(\Phi(\bar{x}), \bar{\lambda}\big) \end{cases} \tag{2.10}$$

hold. Then there exist positive numbers ℓ, γ such that the second-order growth condition

$$f(x) \geq f(\bar{x}) + \tfrac{1}{2}\ell\|x - \bar{x}\|^2 \quad \text{for all } x \in \mathbb{B}_\gamma(\bar{x}) \text{ with } \Phi(x) \in \mathcal{Q} \tag{2.11}$$

is satisfied for the second-order cone program (1.1).

Observe that the presented SOSC (2.10) is equivalent to the second-order conditions used for SOCPs in other publications [9, 19, 25]. This indeed follows from the second subderivative formula (2.7). Note also that SOSC (2.10) is stronger than the conventional second-order sufficient condition for (1.1), the latter requires the supremum of the quadratic term in (2.10) over all the Lagrange multipliers from (2.9) be positive. This stronger condition is in fact equivalent to the second-order growth (2.11) under an appropriate constraint qualification; see [29, Theorem 7.2]. Let us now provide an equivalent version of SOSC (2.10) that is often used in what follows.

Remark 2.2 (equivalent version of SOSC). It is not hard to check that the formulated SOSC (2.10) amounts to saying that there exists a number $\bar{\ell} > 0$ such that we have

$$\begin{cases} \langle \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w, w \rangle + \mathrm{d}^2 \delta_{\mathcal{Q}}\big(\Phi(\bar{x}), \bar{\lambda}\big)\big(\nabla \Phi(\bar{x})w\big) \geq \bar{\ell}\, \|w\|^2 \\ \text{for all}\ \ w \in \mathbb{R}^n\ \ \text{with}\ \ \nabla \Phi(\bar{x})w \in K_{\mathcal{Q}}\big(\Phi(\bar{x}), \bar{\lambda}\big). \end{cases} \tag{2.12}$$

Conversely, the fulfillment of (2.12) at $(\bar{x}, \bar{\lambda})$ ensures that for any $\ell \in (0, \bar{\ell})$ there exists a positive number γ such that the second-order growth condition (2.11) is satisfied at $\bar{x}$.

We conclude this section by recalling some properties of the augmented Lagrangian (1.3) that are used below; see, e.g., [44, Exercise 11.56].

Proposition 2.3 (properties of the augmented Lagrangian). *For* (1.3) *with* $(x, \lambda, \rho) \in \mathbb{R}^n \times \mathbb{R}^{m+1} \times (0, \infty)$ *the following hold:*

(i) *The function* $\rho \mapsto \mathscr{L}(x, \lambda, \rho)$ *is nondecreasing.*

(ii) *The function* $\lambda \mapsto \mathscr{L}(x, \lambda, \rho)$ *is concave.*

It follows from the direct differentiation of (1.3) that for any $\rho > 0$ we have

$$\begin{aligned} \nabla_x \mathscr{L}(x, \lambda, \rho) &= \nabla f(x) + \nabla \Phi(x)^* \Pi_{-\mathcal{Q}}\big(\rho \Phi(x) + \lambda)\big), \\ \nabla_\lambda \mathscr{L}(x, \lambda, \rho) &= \rho^{-1}\big[\Pi_{-\mathcal{Q}}\big(\rho \Phi(x) + \lambda\big) - \lambda\big], \end{aligned} \tag{2.13}$$

which allows us to readily deduce that $(\bar{x}, \bar{\lambda})$ is a solution to the KKT system (2.8) if and only if for any $\rho > 0$ this pair satisfies the equation

$$\big(\nabla_x \mathscr{L}(x, \lambda, \rho), \nabla_\lambda \mathscr{L}(x, \lambda, \rho)\big) = (0, 0). \tag{2.14}$$

Finally in this section, recall some properties of the projection mapping for the second-order cone $\mathcal{Q}$ that are extensively exploited throughout the paper:

(P1) $p = \Pi_{\mathcal{Q}}(y)$ if and only if $p \in \mathcal{Q}$, $y - p \in -\mathcal{Q}$, and $\langle y - p, p \rangle = 0$.

(P2) For every $y \in \mathbb{R}^{m+1}$ we have $y = \Pi_{\mathcal{Q}}(y) + \Pi_{-\mathcal{Q}}(y)$.

(P3) For every $y \in \mathbb{R}^{m+1}$ we have $\big\langle \Pi_{\mathcal{Q}}(y), \Pi_{-\mathcal{Q}}(y) \big\rangle = 0$.

(P4) $\lambda \in N_{\mathcal{Q}}(y)$ if and only if $\Pi_{\mathcal{Q}}(y + \lambda) = y$.

3 Error Bounds for Perturbed KKT Systems of SOCPs

Here we derive an efficient error bound estimate for the KKT system of problem (1.1) under the validity of SOSC (2.10). This is highly important for the subsequent results of the paper.

A crucial role of error bounds in convergence analysis of major numerical algorithms has been well understood in optimization theory; see, e.g., the books [14, 24]. To the best of our knowledge, the first error bound estimate for KKT systems of NLPs under the classical second-order sufficient condition alone was derived in Hager and Gowda [18, Lemma 2] and then was improved by Izmailov [22] who replaced the conventional SOSC with the weaker *noncriticality* of Lagrange multipliers introduced therein. It has been recently observed by Mordukhovich and Sarabi [34]

that similar results for nonpolyhedral conic programs require an additional assumption of the *calmness* of Lagrange multiplier mappings associated with canonically perturbed KKT systems. The latter assumption automatically holds for NLPs.

For any fixed $\bar{x} \in \mathbb{R}^n$ the *multiplier mapping* $M_{\bar{x}} \colon \mathbb{R}^n \times \mathbb{R}^{m+1} \rightrightarrows \mathbb{R}^{m+1}$, associated with the canonically perturbed KKT system (2.8) of (1.1), is defined by

$$M_{\bar{x}}(v, w) := \left\{ \lambda \in \mathbb{R}^{m+1} \mid \nabla_x L(\bar{x}, \lambda) = v, \ \lambda \in N_{\mathcal{Q}}\big(\Phi(\bar{x}) + w\big) \right\}, \quad (v, w) \in \mathbb{R}^n \times \mathbb{R}^{m+1}. \quad (3.1)$$

It is easy to see that $M_{\bar{x}}(0, 0)$ reduces to the set of Lagrange multipliers $\Lambda(\bar{x})$ of the unperturbed system (2.9). Given a solution $(\bar{x}, \bar{\lambda})$ to the KKT system (2.8), the calmness condition (2.4) for $M_{\bar{x}}$ at $((0, 0), \bar{\lambda})$ reads as the existence of positive constants τ and γ such that

$$M_{\bar{x}}(v, w) \cap \mathbb{B}_\gamma(\bar{\lambda}) \subset \Lambda(\bar{x}) + \tau\big(\|w\| + \|v\|\big)\mathbb{B} \ \text{ whenever } \ (v, w) \in \mathbb{B}_\gamma(0, 0).$$

This can be equivalently rewritten as the existence of $\tau, \gamma > 0$ such that the estimate

$$\operatorname{dist}\big(\lambda; \Lambda(\bar{x})\big) \le \tau\big(\|\nabla_x L(\bar{x}, \lambda)\| + \operatorname{dist}\big(\Phi(\bar{x}); N_{\mathcal{Q}}^{-1}(\lambda)\big)\big) \quad (3.2)$$

holds for all $\lambda \in \mathbb{B}_\gamma(\bar{\lambda})$. We can easily check that for (polyhedral) NLPs the calmness of the multiplier mapping follows automatically from the classical Hoffman lemma. Efficient conditions for the calmness of (3.1) in the SOCP framework (1.1) are presented at the end of this section.

Now we are ready to derive the main result of this section ensuring the aforementioned error bound estimate. Define the *residual function* $\sigma \colon \mathbb{R}^n \times \mathbb{R}^{m+1} \to \mathbb{R}$ of the KKT system (2.8) by

$$\sigma(x, \lambda) := \|\nabla_x L(x, \lambda)\| + \|\Phi(x) - \Pi_{\mathcal{Q}}\big(\Phi(x) + \lambda\big)\|, \quad (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^{m+1}. \quad (3.3)$$

It is easy to see that if $(\bar{x}, \bar{\lambda})$ is a solution to the KKT system (2.8), then it follows from property (P4) of the projection mapping that $\sigma(\bar{x}, \bar{\lambda}) = 0$. Using this and the Lipschitz continuity of σ with respect to both x and λ around $(\bar{x}, \bar{\lambda})$, we can find constants $\gamma_2 > 0$ and $\kappa_2 \ge 0$ such that

$$\sigma(x, \lambda) \le \kappa_2\big(\|x - \bar{x}\| + \operatorname{dist}(\lambda; \Lambda(\bar{x}))\big) \quad \text{for all } (x, \lambda) \in \mathbb{B}_{\gamma_2}(\bar{x}, \bar{\lambda}). \quad (3.4)$$

Below we show that the opposite inequality in (3.4), which is crucial for our subsequent developments of the ALM, can be achieved if in addition both SOSC (2.10) and the calmness of the multiplier mapping are satisfied. The provided proof, being strongly based on the geometry of the second-order cone (1.2), is much simpler than the one given recently in [34, Theorem 5.9] for $\mathcal{C}^2$-cone reducible cone programs that is based on a highly involved reduction technique.

Theorem 3.1 (error bound for SOCPs under calmness and SOSC). *Let $(\bar{x}, \bar{\lambda})$ be a solution to the KKT system (2.8), and let SOSC (2.10) hold at $(\bar{x}, \bar{\lambda})$. If the multiplier mapping $M_{\bar{x}}$ in (3.1) is calm at $((0, 0), \bar{\lambda})$, then there exist constants $\gamma_1 > 0$ and $\kappa_1 \ge 0$ such that*

$$\|x - \bar{x}\| + \operatorname{dist}\big(\lambda; \Lambda(\bar{x})\big) \le \kappa_1 \sigma(x, \lambda) \quad \text{for all } (x, \lambda) \in \mathbb{B}_{\gamma_1}(\bar{x}, \bar{\lambda}), \quad (3.5)$$

where the residual function σ is taken from (3.3).

Proof. Observe that if $x = \bar{x}$ and $\lambda \in \Lambda(\bar{x})$, then (3.5) holds since both sides are equal to 0. Let us now verify (3.5) while assuming that either $x \ne \bar{x}$ or $\lambda \notin \Lambda(\bar{x})$. We first show that

$$\|x - \bar{x}\| = O\big(\sigma(x, \lambda)\big) \quad \text{as} \quad (x, \lambda) \to (\bar{x}, \bar{\lambda}). \quad (3.6)$$

Arguing by contradiction, suppose that there exists a sequence $(x^k, \lambda^k) \to (\bar{x}, \bar{\lambda})$ with either $x^k \ne \bar{x}$ or $\lambda^k \notin \Lambda(\bar{x})$ satisfying the strict inequalities

$$\|x^k - \bar{x}\| > k\, \sigma(x^k, \lambda^k) > 0 \quad \text{for all } k \in \mathbb{N},$$

which imply that $\sigma(x^k, \lambda^k) = o(\|x^k - \bar{x}\|)$. By the definition of σ the latter means that

$$\nabla_x L(x^k, \lambda^k) = o(\|x^k - \bar{x}\|) \quad \text{and} \quad \alpha^k := \Phi(x^k) - \Pi_{\mathcal{Q}}(\Phi(x^k) + \lambda^k) = o(\|x^k - \bar{x}\|). \qquad (3.7)$$

Using the second equality in (3.7) combined with property (P1), we get the relationships

$$\Phi(x^k) - \alpha^k \in \mathcal{Q}, \quad \lambda^k + \alpha^k \in -\mathcal{Q}, \quad \text{and} \quad \langle \Phi(x^k) - \alpha^k, \lambda^k + \alpha^k \rangle = 0, \qquad (3.8)$$

which in turn bring us to the inclusion

$$\lambda^k + \alpha^k \in N_{\mathcal{Q}}(\Phi(x^k) - \alpha^k). \qquad (3.9)$$

It follows from the calmness estimate (3.2) that

$$\text{dist}(\lambda^k + \alpha^k; \Lambda(\bar{x})) \leq \tau(\|\nabla_x L(\bar{x}, \lambda^k + \alpha^k)\| + \text{dist}(\Phi(\bar{x}); N_{\mathcal{Q}}^{-1}(\lambda^k + \alpha^k)))$$

for all $k \in \mathbb{N}$ sufficiently large. Since the gradient ∇f and Jacobian $\nabla \Phi$ mappings are Lipschitz continuous around $\bar{x}$, we always have the estimate

$$\begin{aligned}
\|\nabla_x L(\bar{x}, \lambda^k + \alpha^k)\| &\leq \|\nabla f(x^k) - \nabla f(\bar{x})\| + \|\nabla_x L(x^k, \lambda^k)\| + \|(\nabla \Phi(x^k) - \nabla \Phi(\bar{x}))^* \lambda^k\| \\
&\quad + \|\nabla \Phi(\bar{x})^* \alpha^k\| = O(\|x^k - \bar{x}\|).
\end{aligned}$$

On the other hand, it follows from (3.9) that $\Phi(x^k) - \alpha^k \in N_{\mathcal{Q}}^{-1}(\lambda^k + \alpha^k)$, and hence

$$\text{dist}(\Phi(\bar{x}); N_{\mathcal{Q}}^{-1}(\lambda^k + \alpha^k)) \leq \|\Phi(x^k) - \alpha^k - \Phi(\bar{x})\| = O(\|x^k - \bar{x}\|),$$

where the last equality comes from the Lipschitz continuity of Φ around $\bar{x}$ and the condition $\alpha^k = o(\|x^k - \bar{x}\|)$. This ensures in turn that $\lambda^k - \widehat{\lambda}^k = O(\|x^k - \bar{x}\|)$, where $\widehat{\lambda}^k := \Pi_{\Lambda(\bar{x})}(\lambda^k)$. Passing to subsequences if necessary gives us

$$\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \to \xi \neq 0 \quad \text{and} \quad \frac{\lambda^k - \widehat{\lambda}^k}{\|x^k - \bar{x}\|} \to \eta \quad \text{as} \quad k \to \infty. \qquad (3.10)$$

Appealing now to the first estimate in (3.7), we arrive at the equalities

$$\begin{aligned}
o(\|x^k - \bar{x}\|) &= \nabla_x L(x^k, \lambda^k) = \nabla_x L(x^k, \bar{\lambda}) + \nabla \Phi(x^k)^*(\lambda^k - \bar{\lambda}) \\
&= \nabla_x L(\bar{x}, \bar{\lambda}) + \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \\
&\quad + \nabla \Phi(x^k)^*(\lambda^k - \widehat{\lambda}^k) + (\nabla \Phi(x^k) - \nabla \Phi(\bar{x}))^*(\widehat{\lambda}^k - \bar{\lambda}) \\
&= \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})(x^k - \bar{x}) + \nabla \Phi(x^k)^*(\lambda^k - \widehat{\lambda}^k) \\
&\quad + (\nabla^2 \Phi(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|))^*(\widehat{\lambda}^k - \bar{\lambda}) + o(\|x^k - \bar{x}\|).
\end{aligned}$$

Dividing both sides by $\|x^k - \bar{x}\|$ and then passing to the limit as $k \to \infty$ show that

$$0 = \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})\xi + \nabla \Phi(\bar{x})^* \eta. \qquad (3.11)$$

Let us now verify the inclusion $\nabla \Phi(\bar{x})\xi \in K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda}) = T_{\mathcal{Q}}(\Phi(\bar{x})) \cap \{\bar{\lambda}\}^\perp$. Indeed, using the first relation in (3.8) yields

$$\mathcal{Q} \ni \Phi(x^k) - \alpha^k = \Phi(\bar{x}) + \|x^k - \bar{x}\| \left[\nabla \Phi(\bar{x}) \left(\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \right) + \frac{o(\|x^k - \bar{x}\|)}{\|x^k - \bar{x}\|} \right],$$

which tells us that $\nabla \Phi(\bar{x})\xi \in T_{\mathcal{Q}}(\Phi(\bar{x}))$. Combining this with $\bar{\lambda} \in N_{\mathcal{Q}}(\Phi(\bar{x}))$, we obtain $\langle \bar{\lambda}, \nabla \Phi(\bar{x})\xi \rangle \leq 0$. To prove the equality therein, deduce from (3.9) that

$$0 \geq \langle \lambda^k + \alpha^k, \Phi(\bar{x}) - \Phi(x^k) + \alpha^k \rangle = -\left\langle \lambda^k + \alpha^k, \|x^k - \bar{x}\| \left[\nabla \Phi(\bar{x}) \left(\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \right) + \frac{o(\|x^k - \bar{x}\|)}{\|x^k - \bar{x}\|} \right] \right\rangle.$$

Dividing both sides by $\|x^k - \bar{x}\|$ and then passing to the limit as $k \to \infty$ verify that $\langle \bar{\lambda}, \nabla \Phi(\bar{x})\xi \rangle \geq 0$. Thus we get $\langle \bar{\lambda}, \nabla \Phi(\bar{x})\xi \rangle = 0$ and hence arrive at $\nabla \Phi(\bar{x})\xi \in K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})$.

Our next step is to prove the following inequality involving the second subderivative (2.6):

$$\langle \nabla \Phi(\bar{x})\xi, \eta \rangle \geq \mathrm{d}^2 \delta_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})(\nabla \Phi(\bar{x})\xi). \tag{3.12}$$

To proceed, remember that $\widehat{\lambda}^k \in N_{\mathcal{Q}}(\Phi(\bar{x}))$. Using (3.9) and the monotonicity of the normal cone mapping to a convex set, we get

$$\begin{aligned}
0 & \leq & \langle \Phi(x^k) - \Phi(\bar{x}) - \alpha^k, \lambda^k - \widehat{\lambda}^k + \alpha^k \rangle \\
& = & \langle \nabla \Phi(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \lambda^k - \widehat{\lambda}^k + o(\|x^k - \bar{x}\|) \rangle.
\end{aligned}$$

Dividing both sides by $\|x^k - \bar{x}\|^2$ and passing to the limit as $k \to \infty$ give us

$$\langle \nabla \Phi(\bar{x})\xi, \eta \rangle \geq 0.$$

This combined with (2.7) verifies (3.12) if either $\Phi(\bar{x}) = 0$, $\Phi(x) \in \mathrm{int}\,\mathcal{Q}$, or $\bar{\lambda} = 0$.

It remains to validate (3.12) in the case where $\Phi(\bar{x}) \in (\mathrm{bd}\,\mathcal{Q}) \setminus \{0\}$ and $\bar{\lambda} \neq 0$. Then (3.9) and the normal cone representation (2.2) allow us to find $t_k \in \mathbb{R}_+$ and $\widehat{t}_k \in \mathbb{R}_+$ such that $\lambda^k + \alpha^k = t_k(\widetilde{\Phi}(x^k) - \widetilde{\alpha}^k)$ and $\widehat{\lambda}^k = \widehat{t}_k \widetilde{\Phi}(\bar{x})$ for large $k \in \mathbb{N}$. We clearly have $\lim_{k \to \infty} t_k = \lim_{k \to \infty} \widehat{t}_k = \|\bar{\lambda}\|/\|\Phi(\bar{x})\|$. Passing to a subsequence if necessary, assume without loss of generality that either $t_k \geq \widehat{t}_k$ or $t_k \leq \widehat{t}_k$ for all $k \in \mathbb{N}$. If the former holds, then

$$\begin{aligned}
& & \langle \Phi(x^k) - \Phi(\bar{x}) - \alpha^k, \lambda^k - \widehat{\lambda}^k + \alpha^k \rangle = \langle \Phi(x^k) - \Phi(\bar{x}) - \alpha^k, t_k \widetilde{\Phi}(x^k) - \widehat{t}_k \widetilde{\Phi}(\bar{x}) - t_k \widetilde{\alpha}^k \rangle \\
& = & \widehat{t}_k \langle \Phi(x^k) - \Phi(\bar{x}) - \alpha^k, \widetilde{\Phi}(x^k) - \widetilde{\Phi}(\bar{x}) - \widetilde{\alpha}^k \rangle + (t_k - \widehat{t}_k) \langle \Phi(x^k) - \Phi(\bar{x}) - \alpha^k, \widetilde{\Phi}(x^k) - \widetilde{\alpha}^k \rangle \\
& = & \widehat{t}_k \langle \nabla \Phi(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \nabla \widetilde{\Phi}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \rangle - (t_k - \widehat{t}_k) \langle \Phi(\bar{x}), \widetilde{\Phi}(x^k) - \widetilde{\alpha}^k \rangle \\
& \geq & \widehat{t}_k \langle \nabla \Phi(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \nabla \widetilde{\Phi}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \rangle,
\end{aligned}$$

where the third equality comes from $\Phi(x^k) - \alpha^k \in \mathrm{bd}\,\mathcal{Q}$ and the last inequality is due to $\Phi(\bar{x}) \in \mathcal{Q}$ while $\widetilde{\Phi}(x^k) - \widetilde{\alpha}^k \in -\mathcal{Q}$. If the latter holds, a similar argument brings us to

$$\begin{aligned}
& & \langle \Phi(x^k) - \Phi(\bar{x}) - \alpha^k, \lambda^k - \widehat{\lambda}^k + \alpha^k \rangle = \langle \Phi(x^k) - \Phi(\bar{x}) - \alpha^k, t_k \widetilde{\Phi}(x^k) - \widehat{t}_k \widetilde{\Phi}(\bar{x}) - t_k \widetilde{\alpha}^k \rangle \\
& = & t_k \langle \Phi(x^k) - \Phi(\bar{x}) - \alpha^k, \widetilde{\Phi}(x^k) - \widetilde{\Phi}(\bar{x}) - \widetilde{\alpha}^k \rangle + (t_k - \widehat{t}_k) \langle \Phi(x^k) - \Phi(\bar{x}) - \alpha^k, \widetilde{\Phi}(\bar{x}) \rangle \\
& = & t_k \langle \nabla \Phi(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \nabla \widetilde{\Phi}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \rangle + (t_k - \widehat{t}_k) \langle \Phi(x^k) - \alpha^k, \widetilde{\Phi}(\bar{x}) \rangle \\
& \geq & t_k \langle \nabla \Phi(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|), \nabla \widetilde{\Phi}(\bar{x})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \rangle.
\end{aligned}$$

Dividing these estimates by $\|x^k - \bar{x}\|^2$ and passing to the limit as $k \to \infty$ result in

$$\langle \nabla \Phi(\bar{x})\xi, \eta \rangle \geq \frac{\|\bar{\lambda}\|}{\|\Phi(\bar{x})\|} \langle \nabla \Phi(\bar{x})\xi, \nabla \widetilde{\Phi}(\bar{x})\xi \rangle = \mathrm{d}^2 \delta_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})(\nabla \Phi(\bar{x})\xi),$$

where the last equality is taken from (2.7). This fully justifies (3.12).

Combining now (3.12) with (3.11) implies that

$$\langle \xi, \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})\xi \rangle + \mathrm{d}^2 \delta_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})(\nabla \Phi(\bar{x})\xi) \leq \langle \xi, \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})\xi \rangle + \langle \nabla \Phi(\bar{x})\xi, \eta \rangle = 0,$$

which contradicts the second-order sufficient condition (2.10) since $\Phi(\bar{x})\xi \in T_{\mathcal{Q}}(\Phi(\bar{x}))$ and $\xi \neq 0$, and thus verifies estimate (3.6).

To finish the proof of the claimed error bound (3.5), it remains to show that

$$\mathrm{dist}(\lambda; \Lambda(\bar{x})) = O(\sigma(x, \lambda)) \quad \text{as} \quad (x, \lambda) \to (\bar{x}, \bar{\lambda}). \tag{3.13}$$

To proceed, pick (x, λ) satisfying (3.6) and denote $y := \Pi_{\mathcal{Q}}(\Phi(x) + \lambda) - \Phi(x)$. Thus we get $\lambda - y \in N_{\mathcal{Q}}(\Phi(x) + y)$. Moreover, since $(x, \lambda) \to (\bar{x}, \bar{\lambda})$, we get $y \to 0$. Combining the latter with (3.2) readily yields the relationships

$$
\begin{aligned}
\operatorname{dist}\big(\lambda - y; \Lambda(\bar{x})\big) &= O\big(\|\nabla_x L(\bar{x}, \lambda - y)\| + \operatorname{dist}(\Phi(\bar{x}); N_{\mathcal{Q}}^{-1}(\lambda - y))\big) \\
&= O\big(\|\nabla_x L(x, \lambda)\| + \|y\| + \|x - \bar{x}\|\big) = O\big(\sigma(x, \lambda)\big),
\end{aligned}
$$

where the last equality comes from (3.6). Since

$$
\operatorname{dist}\big(\lambda; \Lambda(\bar{x})\big) - \operatorname{dist}\big(\lambda - y; \Lambda(\bar{x})\big) = O(\|y\|) = O\big(\sigma(x, \lambda)\big),
$$

we arrive at (3.13). The error bound (3.5) follows from the combination of (3.6) and (3.13), and hence completes the proof of the theorem. □

Next we present an example showing that the assumed calmness of the multiplier mapping in Theorem 3.1 is essential for the validity of the error bound (3.5). In fact, the following example demonstrates more: not only does the *primal-dual* error bound (3.5) fail without the calmness assumption on (3.1), but even the *primal estimate* (3.6) is violated in the absence of calmness. This illustrates a striking difference between NLPs and nonpolyhedral SOCPs.

Example 3.2 (failure of error bound in the absence of calmness of multiplier mappings). Consider SOCP (1.1) with the data $f : \mathbb{R}^2 \to \mathbb{R}$ and $\Phi : \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$
f(x) := x_2^2 \quad \text{and} \quad \Phi(x) := (-x_1^2 + x_2, x_2, 0) \quad \text{with } x = (x_1, x_2) \in \mathbb{R}^2.
$$

Take $\bar{x} := (0, 0)$ and observe that $\Phi(\bar{x}) = 0$ and that

$$
\nabla f(\bar{x}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \nabla \Phi(\bar{x})^* = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}, \quad \Lambda(\bar{x}) = -\mathcal{Q} \cap \{(1, 1, 0)\}^{\perp} = \mathbb{R}_+(-1, 1, 0).
$$

Letting $\bar{\lambda} := (-1, 1, 0) \in \Lambda(\bar{x})$, we conclude that the pair $(\bar{x}, \bar{\lambda})$ satisfies the KKT system (2.8). It follows from the equality

$$
\nabla_{xx}^2 L(\bar{x}, \bar{\lambda}) = \nabla^2 f(\bar{x}) + \nabla^2 \langle \bar{\lambda}, \Phi \rangle(\bar{x}) = 2I_2,
$$

with I_2 standing for the 2×2 identity matrix, that SOSC (2.10) holds at $(\bar{x}, \bar{\lambda})$. To show now that the multiplier mapping $M_{\bar{x}}$ from (3.1) is not calm at $((0, 0), \bar{\lambda})$, select $\lambda^k := \big(-1, t_k, \sqrt{1 - t_k^2}\big)$ with $t_k \uparrow 1$ as $k \to \infty$, which yields $\lambda^k \to \bar{\lambda}$ as $k \to \infty$ and $\lambda^k \in -\mathcal{Q}$ for all $k \in \mathbb{N}$. Direct calculations give us the expressions

$$
\operatorname{dist}^2\big(\lambda^k; \Lambda(\bar{x})\big) = \left\| \lambda^k - \frac{\langle \lambda^k, \bar{\lambda} \rangle}{\|\bar{\lambda}\|^2} \bar{\lambda} \right\|^2 = \frac{3 - 2t_k - t_k^2}{2} \quad \text{and}
$$

$$
\|\nabla f(\bar{x}) + \nabla \Phi(\bar{x})^* \lambda^k\|^2 = (t_k - 1)^2,
$$

which lead us to the limit calculations

$$
\lim_{k \to \infty} \frac{\operatorname{dist}^2\big(\lambda^k; \Lambda(\bar{x})\big)}{\|\nabla f(\bar{x}) + \nabla \Phi(\bar{x})^* \lambda^k\|^2} = \frac{1}{2} \lim_{k \to \infty} \frac{3 - 2t_k - t_k^2}{(t_k - 1)^2} = \infty.
$$

This tells us that the multiplier mapping $M_{\bar{x}}$ is not calm at $((0, 0), \bar{\lambda})$.

Next we check that the primal estimate (3.6) fails in this example. To proceed, take $x^k := (0, \alpha_k)$ with $\alpha_k := -(t_k - 1)/2$ and observe that $(x^k, \lambda^k) \to (\bar{x}, \bar{\lambda})$ as $k \to \infty$. This yields

$$
\nabla f(x^k) + \nabla \Phi(x^k)^* \lambda^k = \begin{pmatrix} 0 \\ 2\alpha_k \end{pmatrix} + \begin{pmatrix} 0 \\ t_k - 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = o(\|x^k - \bar{x}\|). \tag{3.14}
$$

10

On the other hand, since $\Phi(x^k)$ is a nonzero point on the boundary of $\mathcal{Q}$ and λ^k is a nonzero point on the boundary of $-\mathcal{Q}$, it follows that

$$\Phi(x^k) + \lambda^k = \alpha_k(1,1,0) + \left(-1, t_k, \sqrt{1-t_k^2}\right) \notin \mathcal{Q} \cup -\mathcal{Q}.$$

Letting $y^k := \Phi(x^k) + \lambda^k$, we calculate that

$$\Pi_{\mathcal{Q}}(y^k) = \frac{1}{2}\left(y_0^k + \|y_r^k\|\right)\left(1, \frac{y_r^k}{\|y_r^k\|}\right)$$

and then easily check as $k \to \infty$ that

$$\left(1, \frac{y_r^k}{\|y_r^k\|}\right) \to (1,1,0) \quad \text{and} \quad \lim_{k\to\infty} \frac{y_0^k + \|y_r^k\|}{\|x^k - \bar{x}\|} = \lim_{k\to\infty} \frac{\alpha_k - 1 + \sqrt{\alpha_k^2 + 2\alpha_k t_k + 1}}{\alpha_k} = 2.$$

This allows us to compute the limits

$$\begin{aligned}
\lim_{k\to\infty} \frac{\|\Phi(x^k) - \Pi_{\mathcal{Q}}(\Phi(x^k) + \lambda^k)\|}{\|x^k - \bar{x}\|} &= \lim_{k\to\infty} \left\| \frac{\Phi(x^k)}{\alpha_k} - \frac{y_0^k + \|y_r^k\|}{2\|x^k - \bar{x}\|}\left(1, \frac{y_r^k}{\|y_r^k\|}\right)\right\| \\
&= \|(1,1,0) - (1,1,0)\| = 0.
\end{aligned}$$

Combining the latter with (3.14) demonstrates that the primal estimate (3.6) and hence the error bound (3.5) both fail in this simple example.

Let us now turn our attention to efficient conditions that ensure the fulfillment of the imposed calmness of the multiplier mapping (3.1). First we provide an improvement of a result established recently in [34, Theorem 4.1], which gives a complete characterization of the calmness property of (3.1) together with the *uniqueness* of Lagrange multipliers in terms of the *dual qualification condition* that involves the graphical derivative (2.3) of the normal cone mapping for (1.2). To proceed, consider the *fully perturbed* set of Lagrange multipliers $M \colon \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{m+1} \rightrightarrows \mathbb{R}^{m+1}$, where—in contrast to $M_{\bar{x}}(v,w)$ in (3.1)—the decision variable x is also included in the perturbation procedure. We define this mapping by

$$M(x,v,w) := \left\{\lambda \in \mathbb{R}^{m+1} \mid \nabla_x L(x,\lambda) = v, \ \lambda \in N_{\mathcal{Q}}(\Phi(x) + w)\right\} \tag{3.15}$$

for $(x,v,w) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{m+1}$ and observe that $M(\bar{x},0,0) = M_{\bar{x}}(0,0) = \Lambda(\bar{x})$. The next proposition provides a full characterization of the *upper Lipschitzian* property of the fully perturbed multiplier mapping M via the dual qualification condition, which plays a key role in the convergent analysis of the ALM for SOCP (1.1).

Proposition 3.3 (calmness and uniqueness of Lagrange multipliers). *Let $(\bar{x}, \bar{\lambda})$ be a solution to the KKT system (2.8). Then the following assertions are equivalent:*

(i) *The multiplier mapping $M_{\bar{x}}$ is calm at $((0,0), \bar{\lambda})$, and $\Lambda(\bar{x}) = \{\bar{\lambda}\}$, i.e., the mapping $M_{\bar{x}}$ has the isolated calmness property at $(\bar{x}, \bar{\lambda})$.*

(ii) *We have the dual qualification condition*

$$DN_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})(0) \cap \ker \nabla\Phi(\bar{x})^* = \{0\}. \tag{3.16}$$

(iii) *There exist positive numbers γ_3 and κ_3 such that the upper Lipschitzian estimate*

$$M(x,v,w) \subset \{\bar{\lambda}\} + \kappa_3(\|x - \bar{x}\| + \|v\| + \|w\|)\mathbb{B} \ \text{ for all } \ (x,v,w) \in \mathbb{B}_{\gamma_3}(\bar{x},0,0) \tag{3.17}$$

holds for the fully perturbed multiplier mapping (3.15).

Proof. The equivalence between (i) and (ii) was established in [34, Theorem 4.1]. Also it is not hard to see that (iii) implies (i) since $M(\bar{x}, 0, 0) = \Lambda(\bar{x})$. Thus it remains to verify the last implication (ii) $\Longrightarrow$ (iii). Observe to this end due to [19, Corollary 3.4] that

$$DN_{\mathcal{Q}}\big(\Phi(\bar{x}), \bar{\lambda}\big)(0) = N_{K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})}(0) = K_{\mathcal{Q}}\big(\Phi(\bar{x}), \bar{\lambda}\big)^* = \big(T_{\mathcal{Q}}\big(\Phi(\bar{x})\big) \cap \{\bar{\lambda}\}^\perp\big)^*,$$

which in turn yields the inclusion

$$N_{\mathcal{Q}}\big(\Phi(\bar{x})\big) = T_{\mathcal{Q}}\big(\Phi(\bar{x})\big)^* \subset DN_{\mathcal{Q}}\big(\Phi(\bar{x}), \bar{\lambda}\big)(0).$$

Then the dual qualification (3.3) ensures the fulfillment of the basic constraint qualification

$$N_{\mathcal{Q}}\big(\Phi(\bar{x})\big) \cap \ker \nabla\Phi(\bar{x})^* = \{0\},$$

which implies that the Lagrange multiplier sets $M(x, v, w)$ are uniformly bounded for all (x, v, w) in some neighborhood $\mathcal{U}$ of the nominal triple $(\bar{x}, 0, 0)$.

Having this in hand and arguing by contraposition, suppose on the contrary that the upper Lipschitzian property (3.17) fails. The equivalence between (i) and (ii) readily implies that $M(\bar{x}, 0, 0) = \Lambda(\bar{x}) = \{\bar{\lambda}\}$. Thus it follows from the contraposition assumption that there exist sequences of $(x^k, v^k, w^k) \to (\bar{x}, 0, 0)$ as $k \to \infty$ and of the corresponding multipliers $\lambda^k \in M(x^k, v^k, w^k)$ satisfying the inequality

$$\|\lambda^k - \bar{\lambda}\| > k(\|x^k - \bar{x}\| + \|v^k\| + \|w^k\|) \quad \text{whenever} \quad k \in \mathbb{N}. \tag{3.18}$$

Suppose without loss of generality that $(x^k, v^k, w^k) \in \mathcal{U}$ for all $k \in \mathbb{N}$. Hence the sequence $\{\lambda^k\}$ is bounded, and so it has a limiting point $\widehat{\lambda}$. Taking into account the robustness (closed graph property) of the normal cone mapping $N_{\mathcal{Q}}$ with respect to perturbations of the initial point, the continuity of the mappings $\Phi, \nabla f$, and $\nabla\Phi$ as well as the convergence $(x^k, v^k, w^k) \to (\bar{x}, 0, 0)$, we arrive at $\widehat{\lambda} \in \Lambda(\bar{x}) = \{\bar{\lambda}\}$, which tells us that $\lambda^k \to \bar{\lambda}$ as $k \to \infty$. Letting now $t_k := \|\lambda^k - \bar{\lambda}\|$ ensures that $t_k \downarrow 0$ and allows us to conclude by (3.18) that

$$x^k - \bar{x} = o(t_k), v^k = o(t_k), \quad \text{and} \quad w^k = o(t_k) \quad \text{as} \quad k \to \infty. \tag{3.19}$$

Furthermore, the passage to a subsequence if necessary gives us a vector $\eta \in \mathbb{R}^{m+1} \setminus \{0\}$ such that $\dfrac{\lambda^k - \bar{\lambda}}{t_k} \to \eta$. Recalling that $\lambda^k \in M(x^k, v^k, w^k)$, we get

$$
\begin{aligned}
o(t_k) = v^k &= \nabla f(x^k) + \nabla\Phi(x^k)^* \lambda^k \\
&= \nabla f(x^k) - \nabla f(\bar{x}) + \nabla\Phi(x^k)^* \lambda^k - \nabla\Phi(\bar{x})^* \bar{\lambda} \\
&= \nabla f(x^k) - \nabla f(\bar{x}) + \big(\nabla\Phi(x^k) - \nabla\Phi(\bar{x})\big)^* \lambda^k + \nabla\Phi(\bar{x})^*(\lambda^k - \bar{\lambda}) \\
&= o(t_k) + \nabla\Phi(\bar{x})^*(\lambda^k - \bar{\lambda}),
\end{aligned}
$$

where the verification of the last equality uses the Lipschitz continuity of ∇f and $\nabla\Phi$ around $\bar{x}$, the boundedness of $\{\lambda^k\}$, and the first estimate in (3.19). Dividing both sides of the latter by t_k and passing to the limit as $k \to \infty$ result in $\eta \in \ker \nabla\Phi(\bar{x})^*$. On the other hand, we have

$$\left(\frac{\Phi(x^k) + w^k - \Phi(\bar{x})}{t_k}, \frac{\lambda^k - \bar{\lambda}}{t_k}\right) = \frac{\big(\Phi(x^k) + w^k, \lambda^k\big) - \big(\Phi(\bar{x}), \bar{\lambda}\big)}{t_k} \in \frac{\text{gph } N_{\mathcal{Q}} - \big(\Phi(\bar{x}), \bar{\lambda}\big)}{t_k},$$

which yields $(0, \eta) \in T_{\text{gph } N_{\mathcal{Q}}}(\Phi(\bar{x}), \bar{\lambda})$ and hence verifies the condition

$$\eta \in DN_{\mathcal{Q}}\big(\Phi(\bar{x}, \bar{\lambda})\big)(0) \cap \ker \nabla\Phi(\bar{x})^*.$$

Since $\eta \neq 0$, the latter contradicts (3.16) and thus justifies the claimed estimate (3.17). $\quad\square$

12

A different sufficient condition for the upper Lipschitzian property (3.17) was obtained in [10, Proposition 4.47] by using a condition called the "strict constraint qualification." This condition is strictly more restrictive than the dual qualification (3.16), which—as shown in Proposition 3.3—is indeed equivalent to the upper Lipschitzian estimate in (3.17).

Our next goal is to provide a more detailed analysis of the calmness of the multiplier mapping for (1.1) entirely via the given SOCP data at the fixed solution $(\bar{x}, \bar{\lambda})$ to the KKT system (2.8). Consider all the possible cases. If $\Phi(\bar{x}) \in \operatorname{int} \mathcal{Q}$, then it follows from the normal cone representation (2.2) that $\Lambda(\bar{x}) = \{0\}$ for the set of Lagrange multipliers in (2.9). Since $M_{\bar{x}}(0,0) = \Lambda(\bar{x})$ and since $M_{\bar{x}}(u, v) = \{0\}$ whenever the pair (u, v) is sufficiently close to $(0,0)$, we surely get the calmness of the multiplier mapping at $((0,0), \bar{\lambda})$ with $\bar{\lambda} = 0$ in this case. If further $\Phi(\bar{x}) \in (\operatorname{bd} \mathcal{Q}) \setminus \{0\}$, then it follows from (2.2) that $\Lambda(\bar{x})$ is the intersection of two polyhedral convex sets. Employing the classical Hoffman lemma ensures that

$$\operatorname{dist}\big(\lambda; \Lambda(\bar{x})\big) = O\big(\|\nabla_x L(\bar{x}, \lambda)\| + \operatorname{dist}(\lambda; N_{\mathcal{Q}}(\Phi(\bar{x})))\big) = O\big(\|\nabla_x L(\bar{x}, \lambda)\| + \operatorname{dist}(\Phi(\bar{x}); N_{\mathcal{Q}}^{-1}(\lambda))\big)$$

for all λ close enough to $\bar{\lambda} \in N_{\mathcal{Q}}(\Phi(\bar{x}))$, where the last equality comes from the fact that the mapping $N_{\mathcal{Q}}$ is clearly calm at $(\Phi(\bar{x}), \bar{\lambda})$ in this case. This again verifies the calmness property of the multiplier mapping (3.1) at $((0,0), \bar{\lambda})$.

Considering further the remaining case where $\Phi(\bar{x}) = 0$, we deduce from [19, Proposition 4.1] that the set of Lagrange multipliers $\Lambda(\bar{x})$ admits one of the following representations:

(a) The strict complementarity holds for $\Lambda(\bar{x})$, i.e., $\Lambda(\bar{x})$ contains an interior point of $-\mathcal{Q}$.

(b) $\Lambda(\bar{x}) = \{0\}$.

(c) $\Lambda(\bar{x}) = \{\bar{\lambda}\}$ and $\bar{\lambda} \in \operatorname{bd}(-\mathcal{Q}) \setminus \{0\}$.

(d) $\Lambda(\bar{x}) = \mathbb{R}_+ \bar{\lambda}$ and $\bar{\lambda} \in \operatorname{bd}(-\mathcal{Q}) \setminus \{0\}$.

The next proposition describes the calmness of multipliers for (1.1) when $\Phi(\bar{x}) = 0$.

Proposition 3.4 (calmness of SOCP multipliers at vertex). *Let $(\bar{x}, \bar{\lambda})$ be a solution for the generalized KKT system (2.8), and let $\Phi(\bar{x}) = 0$. The following hold:*

(i) *In cases (a) and (b) for $\Lambda(\bar{x})$ the multiplier mapping $M_{\bar{x}}$ is calm at $((0,0), \bar{\lambda})$.*

(ii) *In case (c) for $\Lambda(\bar{x})$ the calmness of $M_{\bar{x}}$ at $((0,0), \bar{\lambda})$ is equivalent to the full rank of $\nabla \Phi(\bar{x})$.*

Proof. In case (a) we get from [19, Proposition 4.1] that estimate (3.2) is satisfied, which verifies the claimed calmness property of the multiplier mapping. In case (b) it follows from (2.9) that $\nabla f(\bar{x}) = 0$, which yields the equalities

$$-\mathcal{Q} \cap \ker \nabla \Phi(\bar{x})^* = N_{\mathcal{Q}}\big(\Phi(\bar{x})\big) \cap \ker \nabla \Phi(\bar{x})^* = \Lambda(\bar{x}) = \{0\}, \tag{3.20}$$

and so $\bar{\lambda} = 0$ and $K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda}) = T_{\mathcal{Q}}(0) = \mathcal{Q}$. By [19, Corollary 3.4] we have

$$DN_{\mathcal{Q}}\big(\Phi(\bar{x}), \bar{\lambda}\big)(0) = N_{K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})}(0) = -\mathcal{Q}.$$

This together with (3.20) tells us the dual qualification condition (3.16) holds in this case. Employing Proposition 3.3 confirms the calmness of the multiplier mapping $M_{\bar{x}}$ at $((0,0), \bar{\lambda})$.

Finally, consider case (c). If $\nabla \Phi(\bar{x})$ has full rank, then the dual qualification condition (3.16) is satisfied. Hence Proposition 3.3 ensures that the multiplier mapping $M_{\bar{x}}$ is calm at $((0,0), \bar{\lambda})$. Conversely, the validity of the calmness property for $M_{\bar{x}}$ in the framework of (c) implies by Proposition 3.3 that the dual qualification condition (3.16) holds. Combining this with the fact that $\bar{\lambda} \in \operatorname{bd}(-\mathcal{Q}) \setminus \{0\}$ in (c) confirms that the matrix $\nabla \Phi(\bar{x})$ has full rank; see [19, Theorem 4.5] for the verification of this claim. This completes the proof of the proposition. $\square$

The above discussions paint a clear picture for the calmness of the multiplier mapping in all the possible cases but (d). It has not been clarified at this stage how to provide verifiable conditions ensuring the calmness property of $M_{\bar{x}}$ in case (d).

4 Second-Order Variational Analysis of Augmented Lagrangians

This section aims at providing characterizations of the second-order growth condition for the penalized problem (1.4). Our main device to obtain such characterizations is the *second subderivative*. As observed by Rockafellar [42, Theorem 2.2], the second-order growth condition for a proper extended-real-valued function can be characterized via its second subderivative. Using this rather simple albeit powerful result for the penalized problem (1.4) requires the *calculation* of the second subderivative of the augmented Lagrangian (1.3).

We begin with the following assertion that calculates the second subderivative of the Moreau envelope of a convex function. Given $\varphi \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\rho > 0$, recall that the *Moreau envelope* of φ relative to ρ is defined by the infimal convolution

$$(e_{1/\rho}\varphi)(x) := \inf_w \left\{ \varphi(w) + \tfrac{1}{2}\rho\|w - x\|^2 \right\}, \quad x \in \mathbb{R}^n. \tag{4.1}$$

Proposition 4.1 (second subderivatives of Moreau envelopes). *Let $\varphi \colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper, lower semicontinuous (l.s.c.), and convex function, and let $\bar{v} \in \partial\varphi(\bar{x})$. If φ is twice epi-differentiable at $\bar{x}$ for $\bar{v}$, then for any $\rho > 0$ the Moreau envelope $e_{1/\rho}\varphi$ is properly twice epi-differentiable at $\bar{x} + \rho^{-1}\bar{v}$ for $\bar{v}$ and its second subderivative at this point is calculated by*

$$\mathrm{d}^2(e_{1/\rho}\varphi)(\bar{x} + \rho^{-1}\bar{v}, \bar{v})(w) = e_{1/2\rho}\big(\mathrm{d}^2\varphi(\bar{x}, \bar{v})\big)(w) \text{ for all } w \in \mathbb{R}^n. \tag{4.2}$$

Proof. Fix $\rho > 0$. It follows from [44, Theorem 11.23] that

$$(e_{1/\rho}\varphi)^*(z) = \varphi^*(z) + \tfrac{1}{2}\rho^{-1}\|z\|^2 \text{ for all } z \in \mathbb{R}^n, \tag{4.3}$$

where '$*$' signifies the Fenchel conjugate in the sense of convex analysis. Because φ is proper, convex, and twice epi-differentiable at $\bar{x}$ for $\bar{v}$, we deduce from [44, Proposition 13.20] that $\mathrm{d}^2\varphi(\bar{x}, \bar{v})$ is proper, l.s.c., and convex as well. Furthermore, it follows from [44, Theorem 13.21] that the proper twice epi-differentiability of φ at $\bar{x}$ for $\bar{v}$ yields this property for the conjugate function φ^* at $\bar{v}$ for $\bar{x}$. Employing [44, Proposition 12.19] tells us that the inclusion $\bar{v} \in \partial\varphi(\bar{x})$ ensures that $\nabla(e_{1/\rho}\varphi)(\bar{x} + \rho^{-1}\bar{v}) = \bar{v}$. Combining these facts with (4.3) and the sum rule for twice epi-differentiability from [44, Exercise 13.18] implies that $(e_{1/\rho}\varphi)^*$ is properly twice epi-differentiable at $\bar{v}$ for $\bar{x} + \rho^{-1}\bar{v}$ and that its second subderivative is given by

$$\mathrm{d}^2(e_{1/\rho}\varphi)^*(\bar{v}, \bar{x} + \rho^{-1}\bar{v})(w) = \mathrm{d}^2\varphi^*(\bar{v}, \bar{x})(w) + \rho^{-1}\|w\|^2 \text{ for all } w \in \mathbb{R}^n. \tag{4.4}$$

This together with [44, Theorem 13.21] yields the proper twice epi-differentiability of $e_{1/\rho}\varphi$ at $\bar{x} + \rho^{-1}\bar{v}$ for $\bar{v}$. Thus the second subderivative of the latter function can be calculated by

$$\begin{aligned}
\tfrac{1}{2}\mathrm{d}^2(e_{1/\rho}\varphi)(\bar{x} + \rho^{-1}\bar{v}, \bar{v})(w) &= \left(\tfrac{1}{2}\mathrm{d}^2(e_{1/\rho}\varphi)^*(\bar{v}, \bar{x} + \rho^{-1}\bar{v}) \right)^*(w) \\
&= \inf_{u \in \mathbb{R}^n} \left\{ \left(\tfrac{1}{2}\mathrm{d}^2\varphi^*(\bar{v}, \bar{x}) \right)^*(u) + \tfrac{1}{2}\rho\|u - w\|^2 \right\} \\
&= \inf_{u \in \mathbb{R}^n} \left\{ \tfrac{1}{2}\mathrm{d}^2\varphi(\bar{x}, \bar{v})(u) + \tfrac{1}{2}\rho\|u - w\|^2 \right\},
\end{aligned}$$

where the first equality comes from [44, Theorem 13.21], the second one is due to (4.4) and [6, Proposition 14.1(i)], and the last equality follows from [44, Theorem 13.21]. This readily justifies the claimed formula for the second subderivative of $e_{1/\rho}\varphi$ at $\bar{x} + \rho^{-1}\bar{v}$ for $\bar{v}$. $\qquad\square$

The second subderivative of the Moreau envelope for general prox-regular functions was established in [44, Exercise 13.45]. However, there are several differences between the latter result and Proposition 4.1. Firstly, the result of [44] was obtained for $\bar{v} = 0$ and $\rho > 0$ sufficiently large.

Our result does not demand neither of these requirements. Secondly, there is the coefficient $1/2$ in [44, Exercise 13.45], which does not appear in (4.2). The price for a nicer formula, however, is confining ourselves to the framework to convex functions.

Proposition 4.1 allows us to obtain the required calculation of the second subderivative of the augmented Lagrangian (1.3).

Theorem 4.2 (second subderivatives of augmented Lagrangians). *Let $(\bar{x}, \bar{\lambda})$ be a solution to the KKT system (2.8). Then for any $\rho > 0$ the function $x \mapsto \mathscr{L}(x, \bar{\lambda}, \rho)$ defined via the augmented Lagrangian (1.3) is twice epi-differentiable at $\bar{x}$ for 0 and its second subderivative is given by*

$$\mathrm{d}_x^2 \mathscr{L}\big((\bar{x}, \bar{\lambda}, \rho), 0\big)(w) = \big\langle w, \nabla_{xx}^2 L(\bar{x}, \bar{\lambda}) w \big\rangle + Q_{\bar{x}, \bar{\lambda}, \rho}(\nabla \Phi(\bar{x}) w) + \rho \operatorname{dist}^2\big(\nabla \Phi(\bar{x}) w; K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})\big), \tag{4.5}$$

for $w \in \mathbb{R}^n$, where the quadratic function $Q_{\bar{x}, \bar{\lambda}, \rho} : \mathbb{R}^{m+1} \to \mathbb{R}$ is defined by

$$Q_{\bar{x}, \bar{\lambda}, \rho}(v) := \begin{cases} 0 & \text{if } \Phi(\bar{x}) \in (\operatorname{int} \mathcal{Q}) \cup \{0\} \text{ or } \bar{\lambda} = 0, \\ \dfrac{\rho \|\bar{\lambda}\|}{\rho \|\Phi(\bar{x})\| + \|\bar{\lambda}\|} \Big(\|v_r\|^2 - \dfrac{\langle \bar{\lambda}_r, v_r \rangle^2}{\|\bar{\lambda}_r\|^2} \Big) & \text{if } \Phi(\bar{x}) \in (\operatorname{bd} \mathcal{Q}) \setminus \{0\} \text{ and } \bar{\lambda} \neq 0. \end{cases} \tag{4.6}$$

Proof. Since $(\bar{x}, \bar{\lambda})$ is a solution to the KKT system (2.8), we have $\nabla_x \mathscr{L}(\bar{x}, \bar{\lambda}, \rho) = 0$, where $\nabla_x \mathscr{L}$ is calculated in (2.13). The twice epi-differentiability of the function $x \mapsto \mathscr{L}(x, \bar{\lambda}, \rho)$ at $\bar{x}$ for $\bar{v} = 0$ follows from [29, Theorem 8.3(i)]. Let us proceed with the second subderivative calculation for the latter function. If either $\Phi(\bar{x}) \in (\operatorname{int} \mathcal{Q}) \cup \{0\}$ or $\bar{\lambda} = 0$, then by (2.7) we get

$$\mathrm{d}^2 \delta_{\mathcal{Q}}\big(\Phi(\bar{x}), \bar{\lambda}\big)(w) = \delta_{K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})}(w) \quad \text{whenever} \quad w \in \mathbb{R}^{m+1}.$$

Employing again [29, Theorem 8.3(i,iii)] and the second subderivative calculation (4.2) from Proposition 4.1 for the Moreau envelope (4.1) of $\varphi = \delta_{\mathcal{Q}}$ tells us that

$$\begin{aligned} \mathrm{d}_x^2 \mathscr{L}\big((\bar{x}, \bar{\lambda}, \rho), 0\big)(w) &= \big\langle w, \nabla_{xx}^2 L(\bar{x}, \bar{\lambda}) w \big\rangle + e_{1/2\rho}\big(\mathrm{d}^2 \delta_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})\big)(w) \\ &= \big\langle w, \nabla_{xx}^2 L(\bar{x}, \bar{\lambda}) w \big\rangle + \inf_{u \in \mathbb{R}^{m+1}} \big\{ \delta_{K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})}(u) + \rho \|u - \nabla \Phi(\bar{x}) w\|^2 \big\} \\ &= \big\langle w, \nabla_{xx}^2 L(\bar{x}, \bar{\lambda}) w \big\rangle + \rho \operatorname{dist}^2\big(\nabla \Phi(\bar{x}) w; K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})\big), \end{aligned}$$

which verifies formula (4.5) with $Q_{\bar{x}, \bar{\lambda}, \rho}(\nabla \Phi(\bar{x}) w) = 0$ from (4.6) in this case. Assuming next that $\Phi(\bar{x}) \in (\operatorname{bd} \mathcal{Q}) \setminus \{0\}$ and $\bar{\lambda} \neq 0$, define the function $\theta(y) := \frac{1}{2} \operatorname{dist}^2(y; \mathcal{Q})$ for $y \in \mathbb{R}^{m+1}$. It is well known that θ is continuously differentiable on $\mathbb{R}^{m+1}$ and its gradient is given by

$$\nabla \theta(y) = \Pi_{-\mathcal{Q}}(y) \quad \text{whenever} \quad y \in \mathbb{R}^{m+1}.$$

Since $\bar{\lambda} \in N_{\mathcal{Q}}(\Phi(\bar{x}))$ and since $\Phi(\bar{x}) \in (\operatorname{bd} \mathcal{Q}) \setminus \{0\}$ with $\bar{\lambda} \neq 0$, we get $\bar{y} \notin \mathcal{Q} \cup (-\mathcal{Q})$ with $\bar{y} = (\bar{y}_0, \bar{y}_r) := \Phi(\bar{x}) + \rho^{-1} \bar{\lambda}$. This clearly yields $\|\bar{y}_r\| > 0$, and so we arrive at

$$\nabla \theta(y) = \Pi_{-\mathcal{Q}}(y) = \frac{1}{2} \Big(1 - \frac{y_0}{\|y_r\|} \Big) \big(- \|y_r\|, y_r \big) = \frac{1}{2} \Big(y_0 - \|y_r\|, y_r - y_0 \frac{y_r}{\|y_r\|} \Big)$$

for all y close to $\bar{y}$. This confirms, in particular, that θ is $\mathcal{C}^2$-smooth around $\bar{y}$ with

$$\nabla^2 \theta(\bar{y}) = \nabla \Pi_{-\mathcal{Q}}(\bar{y}) = \frac{1}{2} \begin{pmatrix} 1 & -\dfrac{\bar{y}_r^*}{\|\bar{y}_r\|} \\ -\dfrac{\bar{y}_r}{\|\bar{y}_r\|} & I_m - \dfrac{\bar{y}_0}{\|\bar{y}_r\|} I_m + \dfrac{\bar{y}_0}{\|\bar{y}_r\|} \dfrac{\bar{y}_r \bar{y}_r^*}{\|\bar{y}_r\|^2} \end{pmatrix}, \tag{4.7}$$

where I_m the $m \times m$ identity matrix, and where $\bar{y}_r^*$ stands for the corresponding vector row. Since $\Phi(\bar{x}) \in (\operatorname{bd} Q) \setminus \{0\}$ and $\bar{\lambda} \in N_Q(\Phi(\bar{x})) \setminus \{0\}$, it follows that $\bar{\lambda} = t\widetilde{\Phi}(\bar{x}) = t(-\Phi_0(\bar{x}), \Phi_r(\bar{x}))$ for some $t > 0$ and $\bar{\lambda}_0 = -\|\bar{\lambda}_r\|$. Thus we have

$$\bar{y} = \Phi(\bar{x}) + \rho^{-1}\bar{\lambda} = \frac{1}{t}\big(-\bar{\lambda}_0, \bar{\lambda}_r\big) + \frac{1}{\rho}(\bar{\lambda}_0, \bar{\lambda}_r) = \left(\frac{t-\rho}{t\rho}\bar{\lambda}_0, \frac{t+\rho}{t\rho}\bar{\lambda}_r\right) = \left(\frac{\rho-t}{t\rho}\|\bar{\lambda}_r\|, \frac{\rho+t}{t\rho}\bar{\lambda}_r\right).$$

Plugging the latter into (4.7) gives us the gradient formula

$$\nabla\Pi_{-Q}(\bar{y}) = \frac{1}{2}\begin{pmatrix} 1 & -\dfrac{\bar{\lambda}_r^*}{\|\bar{\lambda}_r\|} \\ -\dfrac{\bar{\lambda}_r}{\|\bar{\lambda}_r\|} & \dfrac{2t}{\rho+t}I_m + \dfrac{\rho-t}{\rho+t}\dfrac{\bar{\lambda}_r\bar{\lambda}_r^*}{\|\bar{\lambda}_r\|^2} \end{pmatrix},$$

which being combined with (4.7) and $\bar{\lambda}_0 = -\|\bar{\lambda}_r\|$ results in

$$\begin{aligned}
\langle\nabla^2\theta(\bar{y})v, v\rangle &= \frac{1}{2}\left(v_0^2 - \frac{2v_0}{\|\bar{\lambda}_r\|}\langle\bar{\lambda}_r, v_r\rangle + \frac{2t}{\rho+t}\|v_r\|^2 + \frac{\rho-t}{\rho+t}\frac{\langle\bar{\lambda}_r, v_r\rangle^2}{\|\bar{\lambda}_r\|^2}\right) \\
&= \frac{1}{2}\left[v_0^2 - 2v_0\frac{\langle\bar{\lambda}_r, v_r\rangle}{\|\bar{\lambda}_r\|} + \left(\frac{\langle\bar{\lambda}_r, v_r\rangle}{\|\bar{\lambda}_r\|}\right)^2\right] + \frac{t}{\rho+t}\left(\|v_r\|^2 - \frac{\langle\bar{\lambda}_r, v_r\rangle^2}{\|\bar{\lambda}_r\|^2}\right) \\
&= \frac{(\bar{\lambda}_0 v_0)^2 + 2\widehat{\lambda}_0 v_0\langle\bar{\lambda}_r, v_r\rangle + \langle\bar{\lambda}_r, v_r\rangle^2}{2\|\bar{\lambda}_r\|^2} + \frac{t}{\rho+t}\left(\|v_r\|^2 - \frac{\langle\bar{\lambda}_r, v_r\rangle^2}{\|\bar{\lambda}_r\|^2}\right) \\
&= \frac{\langle\bar{\lambda}, v\rangle^2}{\|\bar{\lambda}\|^2} + \frac{t}{\rho+t}\left(\|v_r\|^2 - \frac{\langle\bar{\lambda}_r, v_r\rangle^2}{\|\bar{\lambda}_r\|^2}\right) \\
&= \operatorname{dist}^2\big(v; K_Q(\Phi(\bar{x}), \bar{\lambda})\big) + \frac{\|\bar{\lambda}\|}{\rho\|\Phi(\bar{x})\| + \|\bar{\lambda}\|}\left(\|v_r\|^2 - \frac{\langle\bar{\lambda}_r, v_r\rangle^2}{\|\bar{\lambda}_r\|^2}\right), \qquad (4.8)
\end{aligned}$$

for all $v = (v_0, v_r) \in \mathbb{R}^{m+1}$. In the last equality we use the fact that $K_Q(\Phi(\bar{x}), \bar{\lambda}) = \{\bar{\lambda}\}^\perp$ and $\|\bar{\lambda}\| = t\|\Phi(\bar{x})\|$. It follows from the twice differentiability of θ at $\bar{y}$ that the function $x \mapsto \mathscr{L}(x, \bar{\lambda}, \rho)$ is twice differentiable at $\bar{x}$ with its second subderivative computed by

$$\mathrm{d}_x^2\mathscr{L}((\bar{x}, \bar{\lambda}, \rho), 0)(w) = \langle\nabla_{xx}^2\mathscr{L}(\bar{x}, \bar{\lambda}, \rho)w, w\rangle = \langle w, \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w\rangle + \rho\langle\nabla^2\theta(\bar{y})v, v\rangle$$

with $v = \nabla\Phi(\bar{x})w$. Combining this and (4.8) gives us the claimed second subderivative formula in this case and thus finishes the proof of the theorem. $\square$

Now we are ready to establish complete pointwise characterizations of the second-order growth condition for the penalized problem (1.4) in terms of SOSC (2.10) and the second subderivative of the augmented Lagrangian (1.3).

Theorem 4.3 (characterizations of second-order growth condition for augmented Lagrangians). *Let $(\bar{x}, \bar{\lambda})$ be a solution to the KKT system (2.8) for SOCP (1.1). Then the following assertions are equivalent:*

(i) *The second-order sufficient condition (2.10) holds at $(\bar{x}, \bar{\lambda})$.*

(ii) *There exists a constant $\rho_{\bar{\lambda}} > 0$ such that for any $\rho \geq \rho_{\bar{\lambda}}$ we have*

$$\mathrm{d}_x^2\mathscr{L}((\bar{x}, \bar{\lambda}, \rho), 0)(w) > 0 \quad \text{whenever} \quad w \in \mathbb{R}^n \setminus \{0\}. \qquad (4.9)$$

(iii) *There exist positive constants $\rho_{\bar{\lambda}}, \gamma_{\bar{\lambda}},$ and $\ell_{\bar{\lambda}}$ such that for any $\rho \geq \rho_{\bar{\lambda}}$ we have*

$$\mathscr{L}(x, \bar{\lambda}, \rho) \geq f(\bar{x}) + \ell_{\bar{\lambda}}\|x - \bar{x}\|^2 \quad \text{for all} \quad x \in \mathbb{B}_{\gamma_{\bar{\lambda}}}(\bar{x}). \qquad (4.10)$$

Proof. Since $(\bar{x}, \bar{\lambda})$ is a solution to the KKT system (2.8), for all $\rho > 0$ we have $\mathscr{L}(\bar{x}, \bar{\lambda}, \rho) = f(\bar{x})$ and $\nabla_x \mathscr{L}(\bar{x}, \bar{\lambda}, \rho) = 0$. Assuming that (ii) holds, deduce from [44, Theorem 13.24] that the second-order growth condition (4.10) for $\rho = \rho_{\bar{\lambda}}$ follows from (4.9) with the same constant ρ. Appealing now to Proposition 2.3(i) tells us that

$$\mathscr{L}(x, \lambda, \rho) \geq \mathscr{L}(x, \lambda, \rho_{\bar{\lambda}}) \quad \text{whenever} \quad \rho \geq \rho_{\bar{\lambda}}.$$

This combined with (4.10) for $\rho = \rho_{\bar{\lambda}}$ justifies the second-order growth condition for any $\rho \geq \rho_{\bar{\lambda}}$ and thus verifies (iii). The opposite implication (iii) $\Longrightarrow$ (ii) follows directly from the definition of the second subderivative.

Assume now that (ii) holds and let $\rho \geq \rho_{\bar{\lambda}}$. To justify (i), pick $w \in \mathbb{R}^n \setminus \{0\}$ with $v := \nabla\Phi(\bar{x})w \in K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})$. We next show that

$$\mathrm{d}^2 \delta_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})(v) \geq Q_{\bar{x}, \bar{\lambda}, \rho}(v) + \rho \, \mathrm{dist}^2(v; K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})) \tag{4.11}$$

for all $\rho > 0$. If either $\Phi(\bar{x}) \in (\mathrm{int}\,\mathcal{Q}) \cup \{0\}$ or $\bar{\lambda} = 0$, we get from (2.7) and (4.6) that

$$\mathrm{d}^2 \delta_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})(v) = Q_{\bar{x}, \bar{\lambda}, \rho}(v) + \rho \, \mathrm{dist}^2(v; K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})) = 0, \tag{4.12}$$

where the last equality stems from the fact that $\nabla\Phi(\bar{x})w \in K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})$. Otherwise, if $\Phi(\bar{x}) \in (\mathrm{bd}\,\mathcal{Q}) \setminus \{0\}$ and $\bar{\lambda} \neq 0$, then we get that $K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda}) = \{\bar{\lambda}\}^{\perp}$. It follows from $v \in K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})$ and $\bar{\lambda} \in \mathrm{bd}\,(-\mathcal{Q}) \setminus \{0\}$ that

$$\langle v_r, \bar{\lambda}_r \rangle^2 = v_0^2 \bar{\lambda}_0^2 = v_0^2 \|\bar{\lambda}_r\|^2 \quad \text{and} \quad \|v_r\|^2 \geq v_0^2. \tag{4.13}$$

We then deduce from (2.7) and (4.6) that

$$
\begin{aligned}
& \mathrm{d}^2 \delta_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})(v) - Q_{\bar{x}, \bar{\lambda}, \rho}(v) - \rho \, \mathrm{dist}^2(v; K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})) \\
= \quad & \frac{\|\bar{\lambda}\|}{\|\Phi(\bar{x})\|}\big(\|v_r\|^2 - v_0^2\big) - \frac{\rho\|\bar{\lambda}\|}{\rho\|\Phi(\bar{x})\| + \|\bar{\lambda}\|}\left(\|v_r\|^2 - \frac{\langle\bar{\lambda}_r, v_r\rangle^2}{\|\bar{\lambda}_r\|^2}\right) \\
= \quad & \frac{\|\bar{\lambda}\|^2}{\|\Phi(\bar{x})\|(\rho\|\Phi(\bar{x})\| + \|\bar{\lambda}\|)}\big(\|v_r\|^2 - v_0^2\big) - \frac{\rho\|\bar{\lambda}\|}{\rho\|\Phi(\bar{x})\| + \|\bar{\lambda}\|}\left(v_0^2 - \frac{\langle\bar{\lambda}_r, v_r\rangle^2}{\|\bar{\lambda}_r\|^2}\right) \\
\geq \quad & 0,
\end{aligned}
\tag{4.14}
$$

where the last inequality is due to estimates in (4.13). Thus, we justify (4.11) for $v \in K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})$. Note that (4.11) is obvious if $v \notin K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda}) = \mathrm{dom}\,\mathrm{d}^2\delta_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})$. Referring to (4.5), we get SOSC (2.10) from (4.9) and (4.11). Thus we are done with (ii) $\Longrightarrow$ (i).

To complete the proof of the theorem, it remains to verify implication (i) $\Longrightarrow$ (ii). Since the second subderivative is positive homogenous of degree 2, to prove (4.9) it is neccesary and sufficient to verify the condition: for all $\rho > 0$ sufficiently large we get

$$\mathrm{d}_x^2 \mathscr{L}((\bar{x}, \bar{\lambda}, \rho), 0)(w) > 0 \quad \text{whenever} \quad w \in \mathbb{S}. \tag{4.15}$$

Assuming that (i) holds, we first justify the claim that (4.15) holds for all $w \in \mathbb{S}$ with $v := \nabla\Phi(\bar{x})w \in K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})$. It is worth mentioning that the quadratic function (in w) on the left-hand side of SOSC (2.10) must attain its minimum value on the compact set $\mathbb{S}$. Let ℓ_0 denote such a value, then by (2.10) we have $\ell_0 > 0$. We now show that

$$Q_{\bar{x}, \bar{\lambda}, \rho}(v) + \rho \, \mathrm{dist}^2(v; K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})) \geq \mathrm{d}^2\delta_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})(v) - \frac{\ell_0}{2} \tag{4.16}$$

for all $\rho > 0$ sufficiently large. In the above proof of the implication (ii) $\Longrightarrow$ (i), it is proved that the latter holds for all $\rho > 0$ whenever $\Phi(\bar{x}) \in (\operatorname{int} \mathcal{Q}) \cup \{0\}$ or $\bar{\lambda} = 0$, see (4.12). Turning now to the remaining case with $\Phi(\bar{x}) \in (\operatorname{bd} \mathcal{Q}) \setminus \{0\}$ and $\bar{\lambda} \neq 0$. Recall from (4.13) and (4.14) that

$$
\begin{aligned}
& Q_{\bar{x},\bar{\lambda},\rho}(v) + \rho \operatorname{dist}^2\big(v; K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})\big) - \mathrm{d}^2 \delta_{\mathcal{Q}}\big(\Phi(\bar{x}), \bar{\lambda}\big)(v) \\
= & -\frac{\|\bar{\lambda}\|^2}{\|\Phi(\bar{x})\|(\rho\|\Phi(\bar{x})\| + \|\bar{\lambda}\|)}\big(\|v_r\|^2 - v_0^2\big) \geq -\frac{\|\bar{\lambda}\|^2 \|v\|^2}{\|\Phi(\bar{x})\|(\rho\|\Phi(\bar{x})\| + \|\bar{\lambda}\|)} \\
\geq & -\frac{\|\bar{\lambda}\|^2 \|\nabla\Phi(\bar{x})\|^2}{\|\Phi(\bar{x})\|(\rho\|\Phi(\bar{x})\| + \|\bar{\lambda}\|)},
\end{aligned}
$$

where, in the last equality, we use the fact that $v = \nabla\Phi(\bar{x})w$ with $\|w\| = 1$. Pick $\varrho_0 > 0$ such that the condition

$$
\frac{\big(\|\nabla\Phi(\bar{x})\| \cdot \|\bar{\lambda}\|\big)^2}{\|\Phi(\bar{x})\|(\rho\|\Phi(\bar{x})\| + \|\bar{\lambda}\|)} \leq \frac{\ell_0}{2} \quad \text{for all } \rho \geq \varrho_0 \tag{4.17}
$$

is fulfilled. Then (4.16) is satisfied for the case with $\Phi(\bar{x}) \in (\operatorname{bd} \mathcal{Q}) \setminus \{0\}$ and $\bar{\lambda} \neq 0$, and therefore, for all possible position of $\Phi(\bar{x}) \in \mathcal{Q}$ and $\bar{\lambda} \in N_{\mathcal{Q}}(\Phi(\bar{x}))$ whenever $\rho \geq \varrho_0$. Referring to (4.5) and SOSC (2.10), we get by (4.16) that

$$
\mathrm{d}_x^2 \mathscr{L}\big((\bar{x}, \bar{\lambda}, \rho), 0\big)(w) \geq \big\langle w, \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w\big\rangle + \mathrm{d}^2 \delta_{\mathcal{Q}}\big(\Phi(\bar{x}), \bar{\lambda}\big)(v) - \frac{\ell_0}{2} \geq \frac{\ell_0}{2} \tag{4.18}
$$

for all $w \in \mathbb{S}$ with $v = \nabla\Phi(\bar{x})w \in K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})$ and for all $\rho \geq \varrho_0$, which just completes the verification of (4.15) for such w.

Next we decompose the unit sphere into the two pieces:

$$
\mathbb{S}_+ := \big\{w \in \mathbb{S} \,\big|\, \big\langle w, \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w\big\rangle + Q_{\bar{x},\bar{\lambda},\varrho_0}(v) > 0\big\}
$$

and

$$
\mathbb{S}_- := \big\{w \in \mathbb{S} \,\big|\, \big\langle w, \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w\big\rangle + Q_{\bar{x},\bar{\lambda},\varrho_0}(v) \leq 0\big\},
$$

where ϱ_0 is taken from (4.17). We see from (4.6) that the function $\rho \mapsto Q_{\bar{x},\bar{\lambda},\rho}(v)$ is nondecreasing on $\mathbb{R}_+$, then by (4.5) the estimate (4.15) is satisfied for any $w \in \mathbb{S}_+$ and any $\rho \geq \varrho_0$. Define the function $\vartheta: \mathbb{S}_- \to \mathbb{R}$ by

$$
\vartheta(w) := -\frac{\big\langle \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w, w\big\rangle}{\operatorname{dist}^2\big(\nabla\Phi(\bar{x})w; K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})\big)}, \quad w \in \mathbb{S}_-.
$$

Picking an arbitrary vector $w \in \mathbb{S}_-$, we conclude from the just proved claim that $\nabla\Phi(\bar{x})w \notin K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})$. This confirms that $\operatorname{dist}\big(\nabla\Phi(\bar{x})w; K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})\big) > 0$. Also we get by (4.6) that $Q_{\bar{x},\bar{\lambda},\varrho_0}(v)$ is always positive, then $w \in \mathbb{S}_-$ implies that $\big\langle \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w, w\big\rangle \leq 0$. Thus the function ϑ is continuous and nonnegative on the compact set $\mathbb{S}_-$, and hence its maximum value over this set, denoted by ϱ_1, is finite and nonnegative. This demonstrates that for any $\rho > \varrho_1$ we have the estimate

$$
\big\langle \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w, w\big\rangle + \rho \operatorname{dist}^2\big(\nabla\Phi(\bar{x})w; K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})\big) > 0 \quad \text{whenever } w \in \mathbb{S}_-.
$$

This together with the above estimate for the case of $w \in \mathbb{S}_+$ and $\rho > \varrho_0$ verifies (4.9) for all $w \in \mathbb{R}^n \setminus \{0\}$ and $\rho \geq \rho_{\bar{\lambda}} > \max\{\varrho_0, \varrho_1\}$ and thus completes the proof of the theorem. $\square$

Implication (i) $\Longrightarrow$ (iii) in Theorem 4.3 was established by Rockafellar in [43, Theorem 7.4] for nonlinear programming problems. His proof strongly exploits the geometry of NLPs and

does not appeal to the second subderivative as in our proof. For the second-order cone programming problem (1.1), the aforementioned implication, not the established equivalencies in Theorem 4.3, was obtained in [27, Proposition 10], where in addition the strict complementarity and nondegeneracy conditions were imposed.

To proceed further, observe that both constants $\ell_{\bar{\lambda}}$ and $\gamma_{\bar{\lambda}}$ in (4.10) depend on $\bar{\lambda}$. Now we are going to find additional assumptions that allow us to justify the second-order growth condition (4.10) for *all* $\lambda \in \Lambda(\bar{x})$ sufficiently close to $\bar{\lambda}$, where the aforementioned constants do not depend on λ. This is crucial for the convergence analysis of the ALM in the case of *nonunique* Lagrange multipliers. The rest of this section is mainly focusing on achieving such a *uniform second-order growth condition* for the augmented Lagrangian (1.3).

We begin with the following lemma, which provides a common constant $\ell_{\bar{\lambda}}$ that works for all λ sufficiently close to $\bar{\lambda}$. Then we derive a similar result for $\gamma_{\bar{\lambda}}$ in the proof of the next theorem.

Lemma 4.4 (uniform estimate for second subderivatives of augmented Lagrangians).
Let $(\bar{x}, \bar{\lambda})$ be a solution to the KKT system (2.8), and let SOSC (2.10) hold at $(\bar{x}, \bar{\lambda})$. Then there exist positive constants $\rho_{\bar{\lambda}}$, ℓ_1, ε_0 such that for all $\rho \geq \rho_{\bar{\lambda}}$ and $\lambda \in \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_0}(\bar{\lambda})$ we have

$$\mathrm{d}_x^2 \mathscr{L}((\bar{x}, \lambda, \rho), 0)(w) \geq \tfrac{1}{2}\ell_1 \|w\|^2 \quad \text{whenever } w \in \mathbb{R}^n. \tag{4.19}$$

Proof. Theorem 4.3 gives us a constant $\rho_{\bar{\lambda}} > 0$ for which condition (4.9) holds when $\rho \geq \rho_{\bar{\lambda}}$. Recall that the second subderivative is l.s.c. and positive homogenous of degree 2. Owing to (4.5), condition (4.9) amounts to the existence of a constant $\ell_1 > 0$ such that

$$\mathrm{d}_x^2 \mathscr{L}((\bar{x}, \bar{\lambda}, \rho_{\bar{\lambda}}), 0)(w) = \langle w, \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w \rangle + Q_{\bar{x}, \bar{\lambda}, \rho_{\bar{\lambda}}}(v) + \rho_{\bar{\lambda}} \operatorname{dist}^2\big(v; K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})\big) \geq \ell_1 \tag{4.20}$$

for all w from the unit sphere $\mathbb{S} \subset \mathbb{R}^n$ and $v = \nabla\Phi(\bar{x})w$, where the quadratic form $Q_{\bar{x}, \bar{\lambda}, \rho_{\bar{\lambda}}}(\cdot)$ is taken from (4.6) with $\rho = \rho_{\bar{\lambda}}$. Let us now verify the existence of $\varepsilon_0 > 0$ so that for any $\lambda \in \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_0}(\bar{\lambda})$ we have

$$\mathrm{d}_x^2 \mathscr{L}((\bar{x}, \lambda, \rho_{\bar{\lambda}}), 0)(w) = \langle w, \nabla_{xx}^2 L(\bar{x}, \lambda)w \rangle + Q_{\bar{x}, \lambda, \rho_{\bar{\lambda}}}(v) + \rho_{\bar{\lambda}} \operatorname{dist}^2\big(v; K_{\mathcal{Q}}(\Phi(\bar{x}), \bar{\lambda})\big) \geq \frac{\ell_1}{2}, \quad w \in \mathbb{S}, \tag{4.21}$$

where $Q_{\bar{x}, \lambda, \rho_{\bar{\lambda}}}(\cdot)$ is taken from (4.6) with replacing $\bar{\lambda}$ by λ and ρ by $\rho_{\bar{\lambda}}$. We first observe that $\langle w, \nabla_{xx}^2 L(\bar{x}, \lambda)w \rangle$ converges to $\langle w, \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w \rangle$ as $\lambda \to \bar{\lambda}$ with $\lambda \in \Lambda(\bar{x})$ uniformly for all $w \in \mathbb{S}$ due to the following estimate

$$|\langle w, \nabla_{xx}^2 L(\bar{x}, \lambda)w \rangle - \langle w, \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w \rangle| \leq \|\nabla^2 \Phi(\bar{x})\| \cdot \|\lambda - \bar{\lambda}\|.$$

We now prove the uniform convergence of $Q_{\bar{x}, \lambda, \rho_{\bar{\lambda}}}(v) \to Q_{\bar{x}, \bar{\lambda}, \rho_{\bar{\lambda}}}(v)$ as $\lambda \to \bar{\lambda}$ with $\lambda \in \Lambda(\bar{x})$ for all $v \in \nabla\Phi(\bar{x})(\mathbb{S})$. It is obvious for the case with $\Phi(\bar{x}) \in (\operatorname{int} \mathcal{Q}) \cup \{0\}$, since quadratic forms reduce to 0 by (4.6). Assume that $\Phi(\bar{x}) \in (\operatorname{bd} \mathcal{Q}) \setminus \{0\}$. If $\bar{\lambda} = 0$, $\lambda \to \bar{\lambda}$ with $\lambda \in \Lambda(\bar{x})$, then it follows from (4.6) that

$$\begin{aligned}
|Q_{\bar{x}, \lambda, \rho_{\bar{\lambda}}}(w) - Q_{\bar{x}, \bar{\lambda}, \rho_{\bar{\lambda}}}(w)| &= \frac{\rho_{\bar{\lambda}} \|\lambda\|}{\rho_{\bar{\lambda}} \|\Phi(\bar{x})\| + \|\lambda\|} \left(\|v_r\|^2 - \frac{\langle \lambda_r, v_r \rangle^2}{\|\lambda_r\|^2} \right) \\
&\leq \frac{\|\lambda\|}{\|\Phi(\bar{x})\|} \|v\|^2 \leq \frac{\|\nabla\Phi(\bar{x})\|^2}{\|\Phi(\bar{x})\|} \|\lambda - \bar{\lambda}\|,
\end{aligned}$$

which justifies the claimed uniform convergence in this case as well. Finally, assume that $\bar{\lambda} \neq 0$ and $\lambda \to \bar{\lambda}$ with $\lambda \in \Lambda(\bar{x})$ and suppose without loss of generality that $\lambda \neq 0$. Since $\lambda, \bar{\lambda} \in \Lambda(\bar{x})$

and $\Phi(\bar{x}) \in (\mathrm{bd}\,\mathcal{Q}) \setminus \{0\}$, it follows from (2.2) that there exist positive constants t and $\bar{t}$ such that $\lambda = t\widetilde{\Phi}(\bar{x})$ and $\bar{\lambda} = \bar{t}\widetilde{\Phi}(\bar{x})$. These relationships result in the equality

$$\frac{\langle \lambda_r, v_r \rangle^2}{\|\lambda_r\|^2} = \frac{\langle \bar{\lambda}_r, v_r \rangle^2}{\|\bar{\lambda}_r\|^2}.$$

Using this together with (4.6) brings us to the estimates

$$\begin{aligned}
|Q_{\bar{x},\lambda,\rho_{\bar{\lambda}}}(v) - Q_{\bar{x},\bar{\lambda},\rho_{\bar{\lambda}}}(v)| &= \left| \frac{\rho_{\bar{\lambda}}\|\lambda\|}{\rho_{\bar{\lambda}}\|\Phi(\bar{x})\| + \|\lambda\|} - \frac{\rho_{\bar{\lambda}}\|\bar{\lambda}\|}{\rho_{\bar{\lambda}}\|\Phi(\bar{x})\| + \|\bar{\lambda}\|} \right| \left(\|v_r\|^2 - \frac{\langle \bar{\lambda}_r, v_r \rangle^2}{\|\bar{\lambda}_r\|^2} \right) \\
&\leq |t - \bar{t}| \|v\|^2 = \frac{\|\lambda - \bar{\lambda}\|}{\|\Phi(\bar{x})\|} \|v\|^2 \leq \frac{\|\nabla\Phi(\bar{x})\|^2}{\Phi(\bar{x})} \|\lambda - \bar{\lambda}\|,
\end{aligned}$$

which again justify the claimed uniform convergence in this last case. Thus we find a number $\varepsilon_1 > 0$ ensuring the uniform condition

$$\langle w, \nabla^2_{xx}L(\bar{x},\lambda)w \rangle + Q_{\bar{x},\lambda,\rho_{\bar{\lambda}}}(v) \geq \langle w, \nabla^2_{xx}L(\bar{x},\bar{\lambda})w + Q_{\bar{x},\bar{\lambda},\rho_{\bar{\lambda}}}(v) - \frac{\ell_1}{4} \qquad (4.22)$$

whenever $w \in \mathbb{S}$ and $\lambda \in \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_1}(\bar{\lambda})$. Next we intend to verify the existence of $\varepsilon_2 > 0$ such that

$$\begin{cases} \mathrm{dist}^2\big(\nabla\Phi(\bar{x})w; K_{\mathcal{Q}}(\Phi(\bar{x}),\lambda)\big) - \mathrm{dist}^2\big(\nabla\Phi(\bar{x})w; K_{\mathcal{Q}}(\Phi(\bar{x}),\bar{\lambda})\big) \geq -\frac{\ell_1}{4\rho_{\bar{\lambda}}} \\ \text{for all } w \in \mathbb{S} \text{ and all } \lambda \in \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_2}(\bar{\lambda}). \end{cases} \qquad (4.23)$$

To proceed, consider the following four possible locations of $\bar{\lambda}$ in $-\mathcal{Q}$:

(a) $\bar{\lambda} = 0$. In this case we have

$$K_{\mathcal{Q}}\big(\Phi(\bar{x}),\bar{\lambda}\big) = T_{\mathcal{Q}}\big(\Phi(\bar{x})\big) \supset K_{\mathcal{Q}}\big(\Phi(\bar{x}),\lambda\big)$$

for all $\lambda \in \Lambda(\bar{x})$, which verifies the fulfillment of (4.23).

(b) $\bar{\lambda} \in \mathrm{int}\,(-\mathcal{Q})$ with $\Phi(\bar{x}) = 0$. If λ is sufficiently close to $\bar{\lambda}$, then $\lambda \in \mathrm{int}\,(-\mathcal{Q})$. This yields

$$K_{\mathcal{Q}}\big(\Phi(\bar{x}),\lambda\big) = K_{\mathcal{Q}}\big(\Phi(\bar{x}),\bar{\lambda}\big) = \{0\},$$

which immediately ensures that (4.23) holds.

(c) $\bar{\lambda} \in \mathrm{bd}\,(-\mathcal{Q}) \setminus \{0\}$ with $\Phi(\bar{x}) \in \mathrm{bd}\,(\mathcal{Q}) \setminus \{0\}$. If $\lambda \to \bar{\lambda}$ with $\lambda \in \Lambda(\bar{x})$, we get $\lambda = t\bar{\lambda}$ for some $t > 0$, which confirms that

$$K_{\mathcal{Q}}(\Phi(\bar{x}),\lambda) = K_{\mathcal{Q}}\big(\Phi(\bar{x}),\bar{\lambda}\big).$$

This clearly justifies the claimed estimate (4.23).

(d) $\bar{\lambda} \in \mathrm{bd}\,(-\mathcal{Q}) \setminus \{0\}$ with $\Phi(\bar{x}) = 0$. In this case, we have for all $\lambda \in \Lambda(\bar{x}) \setminus \{0\}$ that

$$K_{\mathcal{Q}}\big(\Phi(\bar{x}),\lambda\big) = \begin{cases} \mathbb{R}_+\widetilde{\lambda} & \text{if } \lambda \in \mathrm{bd}\,(-\mathcal{Q}) \setminus \{0\}, \\ \{0\} & \text{if } \lambda \in \mathrm{int}\,(-\mathcal{Q}), \end{cases}$$

where the tilde-notation for the ice-cream cone is defined at the end of Section 1. Then (4.23) is obviously satisfied when $\lambda \in \Lambda(\bar{x}) \cap \mathrm{int}\,(-\mathcal{Q})$. Assume now that $\lambda \in [\Lambda(\bar{x}) \cap \mathrm{bd}\,(-\mathcal{Q})] \setminus \{0\}$. It is not hard to verify that for any such a λ we get

$$\mathrm{dist}^2\big(v; K_{\mathcal{Q}}(\Phi(\bar{x}),\lambda)\big) = \|v\|^2 - \frac{1}{\|\lambda\|^2}\big(\max\{0, \langle \widetilde{\lambda}, v \rangle\}\big)^2.$$

It is worth mentioning that the function $\big(\max(0,t)\big)^2$ is $\mathcal{C}^1$ on the whole real line. It follows from the latter formula that $\lambda \in [\Lambda(\bar{x}) \cap \mathrm{bd}\,(-\mathcal{Q})] \setminus \{0\} \mapsto \mathrm{dist}^2\big(v; K_{\mathcal{Q}}(\Phi(\bar{x}),\lambda)\big)$ is a $\mathcal{C}^1$ function relative to the set $[\Lambda(\bar{x}) \cap \mathrm{bd}\,(-\mathcal{Q})] \setminus \{0\}$. Taking this into account and choosing λ to be sufficiently close to $\bar{\lambda}$ ensure the existence of $\varepsilon_2 > 0$ for which the uniform estimate (4.23) is guaranteed. This completes the justification of (4.23) for all the possible cases.

Finally, denote $\varepsilon_0 := \min\{\varepsilon_1, \varepsilon_2\}$ with ε_1 and ε_2 taken from (4.22) and (4.23), respectively. Combining (4.20), (4.22), and (4.23) tells us that estimate (4.21) is satisfied for any $\lambda \in \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_0}(\bar{\lambda})$. Thus for any such a multiplier λ we have

$$\mathrm{d}_x^2 \mathscr{L}\big((\bar{x}, \lambda, \rho_{\bar{\lambda}}), 0\big)(w) \geq \tfrac{1}{2}\ell_1 \|w\|^2 \ \ \text{whenever} \ \ w \in \mathbb{R}^n.$$

This together with (4.5) and the fact that $\rho \mapsto Q_{\bar{x}, \bar{\lambda}, \rho}(v)$ is nondecreasing on $\mathbb{R}_+$ implies for any $\lambda \in \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_0}(\bar{\lambda})$ that

$$\mathrm{d}_x^2 \mathscr{L}\big((\bar{x}, \lambda, \rho), 0\big)(w) \geq \tfrac{1}{2}\ell_1 \|w\|^2 \ \ \text{for all} \ \ w \in \mathbb{R}^n \ \ \text{and all} \ \ \rho \geq \rho_{\bar{\lambda}},$$

which therefore completes the proof of the lemma. $\qquad\square$

Now we are ready to derive a uniform version of the second-order growth condition for (1.3).

Theorem 4.5 (uniform second-order growth condition for augmented Lagrangians). *Let $(\bar{x}, \bar{\lambda})$ be a solution to the KKT system (2.8), and let SOSC (2.10) hold at $(\bar{x}, \bar{\lambda})$. Assume in addition that the Lagrange multiplier set $\Lambda(\bar{x})$ in (2.9) is either a polyhedron, or that the multiplier $\bar{\lambda}$ belongs to the interior of $-\mathcal{Q}$. Then there are positive constants $\rho_{\bar{\lambda}}, \gamma_{\bar{\lambda}}, \varepsilon_{\bar{\lambda}}, \ell_{\bar{\lambda}}$ such that for all $\lambda \in \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_{\bar{\lambda}}}(\bar{\lambda})$ and $\rho \geq \rho_{\bar{\lambda}}$ we have the uniform second-order growth condition*

$$\mathscr{L}(x, \lambda, \rho) \geq f(\bar{x}) + \ell_{\bar{\lambda}} \|x - \bar{x}\|^2 \ \ \text{whenever} \ \ x \in \mathbb{B}_{\gamma_{\bar{\lambda}}}(\bar{x}). \tag{4.24}$$

Proof. Take the positive constants ℓ_1, ε_0, and $\rho_{\bar{\lambda}}$ from Lemma 4.4 for which (4.19) holds whenever $\lambda \in \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_0}(\bar{\lambda})$ and $\rho \geq \rho_{\bar{\lambda}}$. Using [44, Theorem 13.24] and remembering that $\mathscr{L}(\bar{x}, \lambda, \rho_{\bar{\lambda}}) = f(\bar{x})$ for all $\lambda \in \Lambda(\bar{x})$, we deduce from (4.19) that for any $\lambda \in \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_0}(\bar{\lambda})$ there exists $\gamma_\lambda > 0$ ensuring the estimate

$$\mathscr{L}(x, \lambda, \rho_{\bar{\lambda}}) \geq f(\bar{x}) + \frac{\ell_1}{4}\|x - \bar{x}\|^2 \ \ \text{whenever} \ \ x \in \mathbb{B}_{\gamma_\lambda}(\bar{x}), \tag{4.25}$$

where the constant $\ell_1/4$ can be chosen the same for all the multipliers $\lambda \in \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_0}(\bar{\lambda})$. This comes from (4.19) and the proof of [44, Theorem 13.24]; see also Remark 2.2 for a similar discussion. However, the radii of the balls centered at $\bar{x}$ in (4.25) depend on λ. It is shown below that we can find a common radius for all the multipliers $\lambda \in \Lambda(\bar{x})$ that are sufficiently close to $\bar{\lambda}$. To proceed, define the function $\varphi \colon \mathbb{R}^{m+1} \to \overline{\mathbb{R}}$ by

$$\varphi(\lambda) := \sup_{x \in \mathbb{B}_{\gamma_{\bar{\lambda}}}(\bar{x})} \frac{f(\bar{x}) - \mathscr{L}(x, \lambda, \rho_{\bar{\lambda}})}{\|x - \bar{x}\|^2} + \delta_{\Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_0}(\bar{\lambda})}(\lambda), \ \ \lambda \in \mathbb{R}^{m+1}. \tag{4.26}$$

Proposition 2.3(ii) tells us that the function $\lambda \mapsto \mathscr{L}(x, \lambda, \rho_{\bar{\lambda}})$ is concave. This together with the convexity of the set $\Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_0}(\bar{\lambda})$ ensures that φ in (4.26) is a convex function. Let us now verify that for any $\lambda \in \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_0}(\bar{\lambda})$ the value $\varphi(\lambda)$ is finite. To this end, pick such a multiplier λ and observe that if $\gamma_\lambda \geq \gamma_{\bar{\lambda}}$ we get by (4.25) the estimates

$$\varphi(\lambda) \leq \sup_{x \in \mathbb{B}_{\gamma_\lambda}(\bar{x})} \frac{f(\bar{x}) - \mathscr{L}(x, \lambda, \rho_{\bar{\lambda}})}{\|x - \bar{x}\|^2} \leq -\frac{\ell_1}{4}.$$

In particular, this implies that $\varphi(\bar{\lambda}) \leq -\ell_1/4$. If $\gamma_\lambda < \gamma_{\bar{\lambda}}$, then

$$\varphi(\lambda) \leq \max\left\{\sup_{x \in \mathbb{B}_{\gamma_\lambda}(\bar{x})} \frac{f(\bar{x}) - \mathscr{L}(x, \lambda, \rho_{\bar{\lambda}})}{\|x - \bar{x}\|^2}, \max_{\gamma_\lambda \leq \|x - \bar{x}\| \leq \gamma_{\bar{\lambda}}} \frac{f(\bar{x}) - \mathscr{L}(x, \lambda, \rho_{\bar{\lambda}})}{\|x - \bar{x}\|^2}\right\} < \infty,$$

where the first term inside the maximum does not exceed $-\ell_1/4$ because of (4.25), and where the second term is finite since it is the maximum of a continuous function over a compact set. This implies that $\varphi(\lambda)$ is finite for all $\lambda \in \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_0}(\bar{\lambda})$, which ensures that

$$\operatorname{dom}\varphi = \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_0}(\bar{\lambda}).$$

If $\bar{\lambda} \in \operatorname{int}(-\mathcal{Q})$, we get $\bar{\lambda} \in \operatorname{ri}\Lambda(\bar{x})$, which clearly implies that $\bar{\lambda} \in \operatorname{ri}(\operatorname{dom}\varphi)$. Since φ is convex, it is continuous at $\bar{\lambda}$ relative to its domain. Hence we find $\varepsilon_{\bar{\lambda}} \in (0, \varepsilon_0]$ such that

$$\varphi(\lambda) \leq \varphi(\bar{\lambda}) + \frac{\ell_1}{8} \leq -\frac{\ell_1}{8} \quad \text{for all} \quad \lambda \in \operatorname{dom}\varphi \cap \mathbb{B}_{\varepsilon_{\bar{\lambda}}}(\bar{\lambda}) = \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_{\bar{\lambda}}}(\bar{\lambda}). \tag{4.27}$$

Next we proceed to achieve a similar result when $\Lambda(\bar{x})$ is a polyhedral convex set. In this case the collection of Lagrange multipliers $\Lambda(\bar{x})$ is either a ray on the boundary of $-\mathcal{Q}$, or a singleton. If the latter holds, we obtain $\Lambda(\bar{x}) = \{\bar{\lambda}\}$, and hence the uniform growth condition (4.24) follows directly from (4.10). If $\Lambda(\bar{x})$ is a ray on the boundary of $-\mathcal{Q}$, then $\Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_0}(\bar{\lambda})$ is a segment. If now $\bar{\lambda} \neq 0$, then we get $\bar{\lambda} \in \operatorname{ri}[\Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_0}(\bar{\lambda})] = \operatorname{ri}(\operatorname{dom}\varphi)$. Arguing as above leads us to (4.27) in this case. Otherwise, $\bar{\lambda}$ is an endpoint of the aforementioned segment, and thus $\bar{\lambda} = 0$. Let λ_e be the other endpoint. If $\varphi(\lambda_e) \leq \varphi(\bar{\lambda}) + \ell_1/8$, then (4.27) holds for $\varepsilon_{\bar{\lambda}} := \varepsilon_0$, which follows from the convexity of φ. Otherwise, we have that $\varphi(\lambda_e) > \varphi(\bar{\lambda}) + \ell_1/8$. Denote

$$\bar{t} := \frac{\ell_1}{8(\varphi(\lambda_e) - \varphi(\bar{\lambda}))} \in (0, 1) \quad \text{and} \quad \lambda_{\bar{t}} := (1 - \bar{t})\bar{\lambda} + \bar{t}\lambda_e.$$

Then using the convexity of φ tells us that

$$\varphi(\lambda_{\bar{t}}) \leq (1 - \bar{t})\varphi(\bar{\lambda}) + \bar{t}\varphi(\lambda_e) = \varphi(\bar{\lambda}) + \bar{t}(\varphi(\lambda_e) - \varphi(\bar{\lambda})) = \varphi(\bar{\lambda}) + \frac{\ell_1}{8} \leq -\frac{\ell_1}{8},$$

which readily yields (4.27) with $\varepsilon_{\bar{\lambda}} := \|\lambda_{\bar{t}} - \bar{\lambda}\| \in (0, \varepsilon_0]$. This completes the verification of (4.27) with some constant $\varepsilon_{\bar{\lambda}} \in (0, \varepsilon_0]$ if either $\Lambda(\bar{x})$ is a polyhedral convex set, or $\bar{\lambda} \in \operatorname{int}(-\mathcal{Q})$. Consequently, it follows from (4.26) and (4.27) that

$$\mathscr{L}(x, \lambda, \rho_{\bar{\lambda}}) \geq f(\bar{x}) + \frac{\ell_1}{8}\|x - \bar{x}\|^2 \quad \text{for all} \quad x \in \mathbb{B}_{\gamma_{\bar{\lambda}}}(\bar{x}) \quad \text{and} \quad \lambda \in \Lambda(\bar{x}) \cap \mathbb{B}_{\varepsilon_{\bar{\lambda}}}(\bar{\lambda}). \tag{4.28}$$

Employing now Proposition 2.3(i) gives us the inequality

$$\mathscr{L}(x, \lambda, \rho) \geq \mathscr{L}(x, \lambda, \rho_{\bar{\lambda}}) \quad \text{for all} \quad \rho \geq \rho_{\bar{\lambda}}.$$

Combining this with (4.28) and setting $\ell_{\bar{\lambda}} := \ell_1/8$ verify the uniform growth condition (4.24). $\square$

A similar result to Theorem 4.5 was derived in [15, Proposition 3.1] for NLPs. We are not familiar with any previous results on the uniform second-order growth condition (4.24) for SOCPs. As shown in the next section, the second-order growth conditions obtained above are crucial for developing the augmented Lagrangian method for this class of optimization problems.

5 Well-Posedness and Convergence Analysis of ALM for SOCPs

In this concluding section of the paper we apply the suggested approach and results of second-order variational analysis (which are undoubtedly of their independent interest) to the convergence analysis of the *augmented Lagrangian method* for solving SOCPs (1.1).

The principal idea of the ALM for (1.1) is to solve a sequence of *unconstrained minimization* problems for which the objective functions, at each iteration, are approximations of the augmented Lagrangian (1.3). Namely, given the current iteration (x^k, λ^k, ρ_k), the ALM solves the following unconstrained problem (called a *subproblem*):

$$\text{minimize } \mathscr{L}(x, \lambda^k, \rho_k) \text{ for } x \in \mathbb{R}^n \tag{5.1}$$

for next primal iterate x^{k+1} and then use it to construct the next dual iterate λ^{k+1}. More specifically, we aim at solving the *stationary equation*

$$\nabla_x \mathscr{L}(x, \lambda^k, \rho_k) = 0 \tag{5.2}$$

for x^{k+1} and then to update the corresponding multiplier by $\lambda^{k+1} := \Pi_{-\mathcal{Q}}(\rho_k \Phi(x^{k+1}) + \lambda^k)$.

Since solving (5.2) is not easy in practice, it is more convenient to choose an approximate solution x^{k+1} satisfying the *approximate stationary condition*

$$\|\nabla_x \mathscr{L}(x^{k+1}, \lambda^k, \rho_k)\| \leq \varepsilon_k \tag{5.3}$$

with a given accuracy/tolerance $\varepsilon_k \geq 0$. Following the conventional terminology of nonlinear programming, we say that the ALM is *exact* of $\varepsilon_k = 0$, i.e., the exact stationary equation (5.2) is used, and *inexact* if (5.3) with $\varepsilon_k > 0$ is under consideration. In this paper we deal with both exact and inexact versions of the ALM by choosing an arbitrary accuracy $\varepsilon_k \geq 0$ sufficiently small. The ALM for (1.1) is described as follows.

Algorithm 5.1 (augmented Lagrangian method for SOCPs). Choose $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^{m+1}$ and $\bar{\rho} > 0$. Pick $\varepsilon_k \to 0$ as $k \to \infty$ and ρ_k with $\rho_k \geq \bar{\rho}$ for all k and set $k := 0$. Then:

(1) If (x^k, λ^k) satisfies a suitable termination criterion, stop.

(2) Otherwise, find x^{k+1} satisfying (5.3) and update the Lagrange multiplier by

$$\lambda^{k+1} := \Pi_{-\mathcal{Q}}\big(\rho_k \Phi(x^{k+1}) + \lambda^k\big). \tag{5.4}$$

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

To perform the well-posedness and convergence analysis of Algorithm 5.1, we need to make sure first of all that the ALM is *well-defined*, i.e., its subproblems constructed in (5.1) are *solvable*. The following theorem reveals that the optimal solution mappings to subproblems (5.1) enjoy the robust isolated upper Lipschitzian property uniformly in ρ. This confirms, in particular, that subproblems (5.1) always admit a local optimal solution. Note that the developed proof of the theorem requires only the second-order growth condition (4.10), which is based on SOSC (2.10), without any additional assumptions.

Theorem 5.2 (solvability and robust stability of subproblems in ALM). *Let $\rho_{\bar{\lambda}}$, $\gamma_{\bar{\lambda}}$, and $\ell_{\bar{\lambda}}$ be positive constants for which the second-order growth condition (4.10) holds whenever $\rho \geq \rho_{\bar{\lambda}}$. Then there exist constants $\ell > 0$, $\widehat{\gamma} \in (0, \gamma_{\bar{\lambda}}]$, and $\varepsilon > 0$ such that the local optimal solution mapping $S_\rho \colon \mathbb{R}^{m+1} \to \mathbb{R}^n$ defined by*

$$S_\rho(\lambda) := \operatorname{argmin}\big\{\mathscr{L}(x, \lambda, \rho) \,\big|\, x \in \mathbb{B}_{\widehat{\gamma}}(\bar{x})\big\}, \quad \lambda \in \mathbb{R}^{m+1}, \tag{5.5}$$

satisfies, for all $\lambda \in \mathbb{B}_\varepsilon(\bar{\lambda})$ *and all* $\rho \in [\rho_{\bar{\lambda}}, \infty)$, *the inclusions*

$$S_\rho(\lambda) \subset \{\bar{x}\} + \ell\|\lambda - \bar{\lambda}\|\mathbb{B} \quad and \quad \emptyset \neq S_\rho(\lambda) \subset \operatorname{int}\mathbb{B}_{\widehat{\gamma}}(\bar{x}), \tag{5.6}$$

which mean that the mapping S_ρ *enjoys the strengthened robust isolated upper Lipschitzian property at* $(\bar{x}, \bar{\lambda})$ *uniformly in* ρ *on the interval* $[\rho_{\bar{\lambda}}, \infty)$.

Proof. Since Φ is twice differentiable at $\bar{x}$, there are constants $\widehat{\gamma} \in (0, \gamma_{\bar{\lambda}}]$ and $\kappa > 0$ with

$$\|\Phi(x) - \Phi(\bar{x})\| \leq \kappa\|x - \bar{x}\| \quad \text{for all} \quad x \in \mathbb{B}_{\widehat{\gamma}}(\bar{x}). \tag{5.7}$$

Employing the second-order growth condition (4.10) tells us that $S_\rho(\bar{\lambda}) = \{\bar{x}\}$ for all $\rho \geq \rho_{\bar{\lambda}}$. Define now the the positive constant

$$\ell := \frac{\kappa}{\ell_{\bar{\lambda}}} + \sqrt{\frac{\kappa^2}{\ell_{\bar{\lambda}}^2} + \frac{1}{\ell_{\bar{\lambda}}\rho_{\bar{\lambda}}}}, \tag{5.8}$$

select a positive number $\varepsilon < \ell^{-1}\widehat{\gamma}$, and then pick any $\lambda \in \mathbb{B}_\varepsilon(\bar{\lambda})$ and $\rho \geq \rho_{\bar{\lambda}}$. Observe further that for all such λ and ρ we have $S_\rho(\lambda) \neq \emptyset$, since the optimization problem in (5.5) admits an optimal solution by the classical Weierstrass theorem. Fix any $u \in S_\rho(\lambda)$ and recall from Proposition 2.3(ii) that the function $\lambda \mapsto \mathscr{L}(u, \lambda, \rho)$ is concave. This together with (2.13) yields

$$\begin{aligned} \mathscr{L}(u, \lambda, \rho) &\geq \mathscr{L}(u, \bar{\lambda}, \rho) - \langle \nabla_\lambda \mathscr{L}(u, \lambda, \rho), \bar{\lambda} - \lambda \rangle \\ &= \mathscr{L}(u, \bar{\lambda}, \rho) - \rho^{-1} \langle \Pi_{-\mathcal{Q}}(\rho\Phi(u) + \lambda) - \lambda, \bar{\lambda} - \lambda \rangle \\ &\geq f(\bar{x}) + \ell_{\bar{\lambda}}\|u - \bar{x}\|^2 - \rho^{-1} \langle \Pi_{-\mathcal{Q}}(\rho\Phi(u) + \lambda) - \lambda, \bar{\lambda} - \lambda \rangle, \end{aligned} \tag{5.9}$$

where we use (4.10) for the last inequality. It follows from the optimality of u that

$$\mathscr{L}(u, \lambda, \rho) \leq \mathscr{L}(\bar{x}, \lambda, \rho) = f(\bar{x}) + \frac{\rho}{2}\operatorname{dist}^2(\Phi(\bar{x}) + \rho^{-1}\lambda; \mathcal{Q}) - \frac{1}{2}\rho^{-1}\|\lambda\|^2 \leq f(\bar{x}),$$

which together with (5.9) brings us to the estimate

$$\|u - \bar{x}\|^2 \leq \frac{1}{\rho\ell_{\bar{\lambda}}} \langle \Pi_{-\mathcal{Q}}(\rho\Phi(u) + \lambda) - \lambda, \bar{\lambda} - \lambda \rangle. \tag{5.10}$$

Employing the projection properties (P2) and (P4) from Section 2, we get

$$\begin{aligned} \|\Pi_{-\mathcal{Q}}(\rho\Phi(u) + \lambda) - \lambda\| &= \|\rho\Phi(u) - \Pi_{\mathcal{Q}}(\rho\Phi(u) + \lambda)\| \\ &= \|\rho(\Phi(u) - \Phi(\bar{x})) + \Pi_{\mathcal{Q}}(\rho\Phi(\bar{x}) + \bar{\lambda}) - \Pi_{\mathcal{Q}}(\rho\Phi(u) + \lambda)\| \\ &\leq \rho\|\Phi(u) - \Phi(\bar{x})\| + \rho\|\Phi(u) - \Phi(\bar{x})\| + \|\bar{\lambda} - \lambda\| \\ &\leq 2\rho\kappa\|u - \bar{x}\| + \|\bar{\lambda} - \lambda\|, \end{aligned}$$

where the last inequality comes from (5.7). Using this and (5.10) tells us that

$$\|u - \bar{x}\|^2 \leq \frac{1}{\rho\ell_{\bar{\lambda}}}\Big(2\rho\kappa\|u - \bar{x}\| + \|\lambda - \bar{\lambda}\|\Big)\|\lambda - \bar{\lambda}\|,$$

which can be written in the equivalent form as

$$\ell_{\bar{\lambda}}\|u - \bar{x}\|^2 - 2\kappa\|\lambda - \bar{\lambda}\| \cdot \|u - \bar{x}\| - \frac{\|\lambda - \bar{\lambda}\|^2}{\rho} \leq 0.$$

This in turn gives us the estimate

$$\|u - \bar{x}\| \leq \left(\frac{\kappa}{\ell_{\bar{\lambda}}} + \sqrt{\frac{\kappa^2}{\ell_{\bar{\lambda}}^2} + \frac{1}{\ell_{\bar{\lambda}}\rho}}\right)\|\lambda - \bar{\lambda}\| \leq \ell\|\lambda - \bar{\lambda}\| \leq \ell\varepsilon < \widehat{\gamma},$$

which simultaneously verifies both inclusions in (5.6) and thus completes the proof. □

It follows from Theorem 5.2 that, at each iteration k, the condition $\lambda^k \in \mathbb{B}_\varepsilon(\bar{\lambda})$ on the current multiplier in Algorithm 5.1 allows us to find an exact local solution to the optimization problem (5.1) such that $\|u^k - \bar{x}\| \le \ell\|\lambda^k - \bar{\lambda}\|$. Then the Lipschitz continuity of $\nabla_x \mathscr{L}(\cdot, \lambda^k, \rho_k)$ around u^k ensures that for any $\varepsilon_k \ge 0$ we can get an ε_k-solution x^{k+1} satisfying both the approximate stationary condition (5.3) and the same estimate

$$\|x^{k+1} - \bar{x}\| \le \ell\|\lambda^k - \bar{\lambda}\| \tag{5.11}$$

as the exact solution u^k to the optimization problem (5.1) under consideration.

Now we are ready to proceed with local convergence analysis of Algorithm 5.1, which mainly exploits the two major ingredients and the corresponding results developed above: **(1)** *SOSC* (2.10) at $(\bar{x}, \bar{\lambda})$ and the associated *second-order growth* of the augmented Lagrangian, and **(2)** the *calmness* of the multiplier mapping. In addition, we assume that the set of Lagrange multipliers is a *singleton*. The main reason for imposing this restriction is that the convergent analysis of the general case, namely when the set of Lagrange multipliers is not a singleton, is conducted by using an iterative framework proposed by Fischer in [16, Theorem 1]. However, the latter result demands an error bound estimate the for consecutive terms of the ALM. Deriving such an estimate for SOCPs is our ongoing research project.

The following theorem establishes the linear convergence of Algorithm 5.1 in both exact and inexact frameworks of the ALM with an arbitrarily chosen tolerance in (5.3) in the form $\varepsilon_k = o\big(\sigma(x^k, \lambda^k)\big)$, where $\sigma(x, \lambda)$ is the error bound from (3.3).

Theorem 5.3 (primal-dual convergence of ALM). *Let $(\bar{x}, \bar{\lambda})$ be a solution to the KKT system* (2.8), *let SOSC* (2.10) *hold at $(\bar{x}, \bar{\lambda})$, and let the multiplier mapping $M_{\bar{x}}$ from* (3.1) *be calm at $((0,0), \bar{\lambda})$ and $\Lambda(\bar{x}) = \{\bar{\lambda}\}$. Then there exist positive numbers $\bar{\gamma}$ and $\bar{\rho}$ ensuring the following: for any starting point $(x^0, \lambda^0) \in \mathbb{B}_{\bar{\gamma}}(\bar{x}, \bar{\lambda})$ and any $\rho_k \ge \bar{\rho}$, Algorithm* 5.1 *generates a sequence of iterates (x^k, λ^k) with a tolerance in* (5.3) *arbitrary chosen as $\varepsilon_k = o(\sigma(x^k, \lambda^k))$ such that (x^k, λ^k) converges to $(\bar{x}, \bar{\lambda})$ as $k \to \infty$, and the rate of this convergence is linear.*

Proof. Let $\rho_{\bar{\lambda}}, \gamma_{\bar{\lambda}}, \ell_{\bar{\lambda}}$ be the positive constants taken from Theorem 4.3(iii), and let κ_i and γ_i for $i = 1, 2, 3$ be positive constants taken from the Lipschitzian estimates (3.5), (3.4), and (3.17), respectively. Picking the positive constants κ and $\widehat{\gamma}$ from (5.7), ℓ from (5.8), and ε from Theorem 5.2, define the positive numbers

$$\widehat{\gamma}_1 := \min\{\gamma_1, \widehat{\gamma}\}, \quad \gamma_{1,2} := \max\{\gamma_1, \gamma_2\}, \quad \gamma := \min\left\{\gamma_3, \frac{\widehat{\gamma}_1}{2\kappa_3}\right\}, \tag{5.12}$$

$$\bar{\rho} := \max\{\rho_{\bar{\lambda}}, 2\kappa_1, 8\kappa_1^2\kappa_2\}, \text{ and } \bar{\gamma} := \min\left\{\widehat{\gamma}_1, \gamma_2, \varepsilon, \frac{\bar{\rho}\gamma}{2\sqrt{10}}, \frac{\widehat{\gamma}_1}{2\ell}, \frac{\gamma}{2\ell(\kappa+1)}\right\}. \tag{5.13}$$

Assume also without loss of generality that

$$o\big(\sigma(x, \lambda)\big) \le \min\left\{\frac{1}{\kappa_2\bar{\rho}}, \frac{1}{8\kappa_1\kappa_2}\right\}\sigma(x, \lambda) \text{ whenever } (x, \lambda) \in \mathbb{B}_{\gamma_{1,2}}(\bar{x}, \bar{\lambda}) \tag{5.14}$$

and then show that for any starting point $(x^0, \lambda^0) \in \mathbb{B}_{\bar{\gamma}}(\bar{x}, \bar{\lambda})$ there exists a sequence $\{(x^k, \lambda^k)\}$ generated by Algorithm 5.1 with any $\rho_k \ge \bar{\rho}$ such that

$$(x^k, \lambda^k) \in \mathbb{B}_{\bar{\gamma}}(\bar{x}, \bar{\lambda}) \text{ for all } k \in \mathbb{N} \cup \{0\}. \tag{5.15}$$

Arguing by induction, observe that (5.15) obviously holds for $k = 0$ and suppose that (5.15) is satisfied for some $k \in \mathbb{N}$ with $\rho_k \ge \bar{\rho}$. We are going to verify that (5.15) fulfills for $k + 1$. To

25

furnish this, deduce first from (5.13) that $\|\lambda^k - \bar{\lambda}\| \leq \varepsilon$. This together with the remark after the proof of Theorem 5.2 ensures the existence of an approximate solution x^{k+1} with

$$\|\nabla_x \mathscr{L}(x^{k+1}, \lambda^k, \rho_k)\| \leq \varepsilon_k = o(\sigma(x^k, \lambda^k)),$$

where $\varepsilon_k \geq 0$ can be chosen arbitrary in this form. It follows from (5.11) that the obtained ε_k-solution satisfies the estimates

$$\|x^{k+1} - \bar{x}\| \leq \ell\|\lambda^k - \bar{\lambda}\| \leq \ell\bar{\gamma} \leq \frac{\widehat{\gamma}_1}{2}, \tag{5.16}$$

where the last inequality comes from (5.13). We proceed now to establish a similar estimate for the dual iterate λ^{k+1}. Using (5.4) and the projection property (P4) yields $\lambda^{k+1} \in N_{\mathcal{Q}}(\Phi(x^{k+1}) + \rho_k^{-1}(\lambda^k - \lambda^{k+1}))$ and hence $\lambda^{k+1} \in M(x^{k+1}, v^{k+1}, w^{k+1})$ with $w^{k+1} := \dfrac{\lambda^k - \lambda^{k+1}}{\rho_k}$ and

$$v^{k+1} := \nabla_x L(x^{k+1}, \lambda^{k+1}) = \nabla_x \mathscr{L}(x^{k+1}, \lambda^k, \rho_k) = o(\sigma(x^k, \lambda^k)). \tag{5.17}$$

The inclusion $(x^k, \lambda^k) \in \mathbb{B}_{\gamma_2}(\bar{x}, \bar{\lambda})$ allows us to deduce from (3.4), (5.14), and (5.17) that

$$\|v^{k+1}\| \leq \frac{\sigma(x^k, \lambda^k)}{\kappa_2 \bar{\rho}} \leq \frac{\|x^k - \bar{x}\| + \|\lambda^k - \bar{\lambda}\|}{\bar{\rho}}. \tag{5.18}$$

Employing again the updating scheme (5.4), we arrive at the relationships

$$
\begin{aligned}
\|w^{k+1}\| &= \|\rho_k^{-1}\lambda^k - \Pi_{-\mathcal{Q}}(\Phi(x^{k+1}) + \rho_k^{-1}\lambda^k)\| \\
&\leq \rho_k^{-1}\|\lambda^k - \bar{\lambda}\| + \|\Pi_{-\mathcal{Q}}(\Phi(x^{k+1}) + \rho_k^{-1}\lambda^k) - \Pi_{-\mathcal{Q}}(\Phi(\bar{x}) + \rho_k^{-1}\bar{\lambda})\| \\
&\leq 2\rho_k^{-1}\|\lambda^k - \bar{\lambda}\| + \|\Phi(x^{k+1}) - \Phi(\bar{x})\| \\
&\leq 2\bar{\rho}^{-1}\|\lambda^k - \bar{\lambda}\| + \kappa\|x^{k+1} - \bar{x}\|
\end{aligned}
$$

with the last estimate coming from (5.7) and $x^{k+1} \in \mathbb{B}_{\widehat{\gamma}}(\bar{x})$. Thus (5.16) and (5.18) bring us to

$$
\begin{aligned}
\|x^{k+1} - \bar{x}\| + \|v^{k+1}\| + \|w^{k+1}\| &\leq (\kappa+1)\|x^{k+1} - \bar{x}\| + \bar{\rho}^{-1}\|x^k - \bar{x}\| + 3\bar{\rho}^{-1}\|\lambda^k - \bar{\lambda}\| \\
&\leq \ell(\kappa+1)\|\lambda^k - \bar{\lambda}\| + \sqrt{10}\bar{\rho}^{-1}\|(x^k, \lambda^k) - (\bar{x}, \bar{\lambda})\| \leq \gamma,
\end{aligned}
$$

where the last inequality employs the induction assumption (5.15) together with (5.13). This along with $\gamma \leq \gamma_3$ due to (5.12) ensures that $(x^{k+1}, v^{k+1}, w^{k+1}) \in \mathbb{B}_{\gamma_3}(\bar{x}, 0, 0)$. Hence we deduce from the upper Lipschitzian property in (3.17) and the definition of γ in (5.12) that

$$\|\lambda^{k+1} - \bar{\lambda}\| \leq \kappa_3(\|x^{k+1} - \bar{x}\| + \|v^{k+1}\| + \|w^{k+1}\|) \leq \kappa_3\gamma \leq \frac{\widehat{\gamma}_1}{2}$$

verifying therefore the promised estimate for the dual iterate λ^{k+1}. This together with (5.16) shows that $(x^{k+1}, \lambda^{k+1}) \in \mathbb{B}_{\widehat{\gamma}_1}(\bar{x}, \bar{\lambda})$. Using the latter, the imposed SOSC (2.10), and the calmness of the multiplier mappings $M_{\bar{x}}$ from (3.1), we conclude from Theorem 3.1 that

$$\|x^{k+1} - \bar{x}\| + \|\lambda^{k+1} - \bar{\lambda}\| \leq \kappa_1\sigma_{k+1} \quad \text{with}$$

$$\sigma_{k+1} = \|\nabla_x L(x^{k+1}, \lambda^{k+1})\| + \|\Phi(x^{k+1}) - \Pi_{\mathcal{Q}}(\Phi(x^{k+1}) + \lambda^{k+1})\|. \tag{5.19}$$

Define further the projection vector

$$p^{k+1} := \Pi_{\mathcal{Q}}(\Phi(x^{k+1}) + \rho_k^{-1}\lambda^k)$$

and deduce from the updating scheme (5.4) that

$$\Phi(x^{k+1}) - p^{k+1} = \frac{\lambda^{k+1} - \lambda^k}{\rho_k}. \tag{5.20}$$

Employing the projection properties (P1) and (P2) results in $\langle p^{k+1}, \lambda^{k+1} \rangle = 0$ due to

$$\rho_k^{-1}\lambda^{k+1} = \Pi_{-\mathcal{Q}}\big(\Phi(x^{k+1}) + \rho_k^{-1}\lambda^k\big) = \Phi(x^{k+1}) + \rho_k^{-1}\lambda^k - p^{k+1},$$

which together with $p^{k+1} \in \mathcal{Q}$ yields $\lambda^{k+1} \in N_{\mathcal{Q}}(p^{k+1})$. Hence $p^{k+1} = \Pi_{\mathcal{Q}}(p^{k+1} + \lambda^{k+1})$ by property (P4). Since the mapping $y \mapsto y - \Pi_{\mathcal{Q}}(y + \lambda^{k+1}) = \Pi_{-\mathcal{Q}}(y + \lambda^{k+1}) - \lambda^{k+1}$ is clearly nonexpansive, we arrive at the relationships

$$
\begin{aligned}
&\big\|\Phi(x^{k+1}) - \Pi_{\mathcal{Q}}\big(\Phi(x^{k+1}) + \lambda^{k+1}\big)\big\| \\
={}& \big\|\Phi(x^{k+1}) - \Pi_{\mathcal{Q}}\big(\Phi(x^{k+1}) + \lambda^{k+1}\big)\big\| - \big\|p^{k+1} - \Pi_{\mathcal{Q}}(p^{k+1} + \lambda^{k+1})\big\| \\
\leq{}& \big\|\Phi(x^{k+1}) - \Pi_{\mathcal{Q}}\big(\Phi(x^{k+1}) + \lambda^{k+1}\big) - \big(p^{k+1} - \Pi_{\mathcal{Q}}(p^{k+1} + \lambda^{k+1})\big)\big\| \\
\leq{}& \big\|\Phi(x^{k+1}) - p^{k+1}\big\| \\
\leq{}& \rho_k^{-1}\|\lambda^{k+1} - \lambda^k\| \qquad \text{(by (5.20))} \\
\leq{}& \rho_k^{-1}\big(\|\lambda^{k+1} - \bar\lambda\| + \|\lambda^k - \bar\lambda\|\big) \\
\leq{}& \kappa_1\rho_k^{-1}(\sigma_{k+1} + \sigma_k).
\end{aligned}
$$

Using this together with (5.17) and (5.19) leads us to the estimates

$$\sigma_{k+1} \leq \varepsilon_k + \big\|\Phi(x^{k+1}) - \Pi_{\mathcal{Q}}\big(\Phi(x^{k+1}) + \lambda^{k+1}\big)\big\| \leq \varepsilon_k + \frac{\kappa_1}{\rho_k}\Big(\sigma_{k+1} + \sigma_k\Big),$$

which can be equivalently rewritten as

$$\Big(1 - \frac{\kappa_1}{\rho_k}\Big)\sigma_{k+1} \leq \varepsilon_k + \frac{\kappa_1}{\rho_k}\sigma_k.$$

Since $\rho_k \geq \bar\rho$, by (5.13), we get $1 - \dfrac{\kappa_1}{\rho_k} > \dfrac{1}{2}$, which ensures that

$$\sigma_{k+1} \leq 2\sigma_k\left(\frac{\varepsilon_k}{\sigma_k} + \frac{\kappa_1}{\rho_k}\right).$$

Applying finally the error bounds (3.5) and (3.4) and then appealing to (5.13) and (5.14) yields

$$
\begin{aligned}
\|x^{k+1} - \bar x\| + \|\lambda^{k+1} - \bar\lambda\| &\leq \kappa_1\sigma_{k+1} \leq 2\kappa_1\left(\frac{\varepsilon_k}{\sigma_k} + \frac{\kappa_1}{\rho_k}\right)\sigma_k \\
&\leq 2\kappa_1\kappa_2\left(\frac{\varepsilon_k}{\sigma_k} + \frac{\kappa_1}{\rho_k}\right)\big(\|x^k - \bar x\| + \|\lambda^k - \bar\lambda\|\big) \\
&\leq \frac{1}{2}\big(\|x^k - \bar x\| + \|\lambda^k - \bar\lambda\|\big), \tag{5.21}
\end{aligned}
$$

which together with the induction assumption (5.15) brings us to

$$(x^{k+1}, \lambda^{k+1}) \in \mathbb{B}_{\bar\gamma}(\bar x, \bar\lambda).$$

This finishes our induction argument to justify (5.15) for all $k \in \mathbb{N}$. Observe that the latter inclusion along with (5.13) implies that $\|\lambda^{k+1} - \bar\lambda\| \leq \varepsilon$ while allowing us to use Theorem 5.2 to construct the next primal iterate x^{k+2}. Since (5.21) holds for all $k \in \mathbb{N}$, we clearly get that $(x^k, \lambda^k) \to (\bar x, \bar\lambda)$ as $k \to \infty$. Furthermore, the obtained estimate tells us that rate of convergence of (x^k, λ^k) to $(\bar x, \bar\lambda)$ is linear, which therefore completes the proof of the theorem. $\qquad\square$

To conclude the paper, let us compare the convergence analysis of the ALM given in Theorem 5.3 with the one provided recently by Kanzow and Steck [25, 26] for the class of $\mathcal{C}^2$-cone reducible conic programs that includes SOCPs. These publications were devoted to convergence analysis of a modified version of the ALM, called the *safeguarded augmented Lagrangian methods* therein, in which the λ^k in the formation of subproblems (5.1) is replaced with a certain vector w^k chosen from a bounded set or obtained from the projection of λ^k onto a bounded set. The main motivation for this modification comes from the fact that the dual sequence, constructed by the standard ALM, may be unbounded in general. As has been extensively documented in [1–3, 8], this modification has a remarkable global convergence theory and was successfully implemented in the ALGENCAN software; see [8, Chapter 10] for a through discussion about the different aspects of this implementation. Also, note that while the safeguarded ALM in [25, 26] uses a particular updating scheme for the penalty parameter ρ_k, our approach reveals that there is no need to confine the convergence analysis of the standard ALM to a particular updating scheme for ρ_k as long as we keep it sufficiently large. Also, as mentioned in Section 1, the solvability of the standard ALM subproblems (5.1) was not addressed in [25, 26]. Let us finally emphasize that the progress achieved in this paper is largely based on the application and development of powerful tools of second-order variational analysis and generalized differentiation.

Data Availability Statement: Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

References

1. Andreani, R., Birgin, E., Martínez, J.M., Schuverdt, M.L.: On Augmented Lagrangian methods with general lower-level constraints. SIAM Journal on Optimization 18, 1286–1309, (2007)

2. Andreani, R., Martínez, J.M., Schuverdt, M.L.: On Second-Order Optimality Conditions for Nonlinear Programming. Optimization 56, 529–542 (2007)

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

4. Andreani, R., Haeser. G., Silva, P.S., Ramos, A.: A second-order sequential optimality condition associated to the convergence of optimization algorithms. IMA J. Numer. Anal. 37, 1902–1929 (2017)

5. Alizadeh, F., Goldfarb, D.: Second-order cone programming. Math. Program. 95, 3–51 (2003)

6. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)

7. Bertsekas, D.P.: Constrained Optimization and Lagrangian Multiplier Method. Academic Press, New York (1982)

8. Birgin, E.G., Martńez, J. M.: Practical Augmented Lagrangian Methods for Constrained Optimization. SIAM, Philadelphia (2014)

9. Bonnans, J.F., Ramírez, H.C.: Perturbation analysis of second-order cone programming problems. Math. Program. 104, 205–227 (2005)

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

11. Cui, Y., Sun, D., Toh, K.-C.: On the R-superlinear convergence of the KKT residues generated by the augmented Lagrangian method for convex composite conic programming. Math. Program. 178, 381–415 (2019)

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

13. Drusvyatskiy, D., Lewis, A.S.: Error bounds, quadratic growth, and linear convergence of proximal methods. Math. Oper. Res. 43, 919–948, (2018)

14. Facchinei, F., Pang, J.-C.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer, New York (2003)

15. 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)

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

17. Gfrerer, H.: First-order and second-order characterizations of metric subregularity and calmness of constraint set mappings. SIAM J. Optim. 21, 1439–1474 (2011)

18. Hager, W., Gowda, M.S.: Stability in the presence of degeneracy and error estimation. Math. Program. 85, 181–192 (1999)

19. Hang, N.T.V., Mordukhovich, B.S., Sarabi, E.M.: Second-order variational analysis in second-order cone programming. Math. Program. 180, 75–116 (2020)

20. Henrion, R., Outrata, J.V.: Calmness of constraint systems with applications. Math. Program. 104, 437–464 (2005)

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

22. Izmailov, A.F.: On the analytical and numerical stability of critical Lagrange multipliers. Comput. Math. Math. Phys. 45, 930–946 (2005)

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

24. Izmailov A.F., Solodov, M.S.: Newton-Type Methods for Optimization and Variational Problems. Springer, New York (2014)

25. Kanzow, C., Steck, D.: On error bounds and multiplier methods for variational problems in Banach spaces. SIAM J. Optim. 56, 1716–1738 (2018)

26. 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)

27. Liua, Y.J., Zhang, L.: Convergence analysis of the augmented Lagrangian method for nonlinear second-order cone optimization problems. Nonlin. Anal. 67, 1359–1373 (2007)

28. Mohammadi, A., Mordukhovich, B.S., Sarabi, E.M.: Variational analysis of composite models with applications to continuous optimization. Math. Oper. Res. https://doi.org/10.1287/moor.2020.1074 (2021)

29. Mohammadi, A., Mordukhovich, B.S., Sarabi, E.M.: Parabolic regularity in geometric variational analysis. Trans. Amer. Math. Soc. 374, 1711–1763 (2021)

30. Mohammadi, A., Mordukhovich, B.S., Sarabi, M.E.: Superlinear Convergence of the Sequential Quadratic Method in Constrained Optimization. J. Optim. Theory Appl. 186, 731–758 (2020)

31. 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)

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

33. Mordukhovich, B.S., Outrata, J.V., Sarabi, M.E.: Full stability of locally optimal solution in second-order cone programming. SIAM J. Optim. 24, 1581–1613 (2014)

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

35. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer, New York (2006)

36. 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)

37. Robinson, S.M.: Stability theory for systems of inequalities, II: differentiable nonlinear systems. SIAM J. Numer. Anal. 13, 497–513 (1976)

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

39. Rockafellar, R.T.: The multiplier method of Hestenes and Powell applied to convex programming. J. Optim. Theory Appl. 12, 555–562 (1973)

40. Rockafellar, R.T.: Augmented Lagrangian multiplier functions and duality in nonconvex programming. SIAM J. Control 12, 268–285 (1974)

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

42. Rockafellar, R.T.: Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives. Math. Oper. Res. 14, 462–484 (1989)

43. Rockafellar, R.T.: Lagrange multipliers and optimality. SIAM Rev. 35, 183–238 (1993)

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

45. Ruszczynski, A.: Nonlinear Optimization. Princeton University Press, Princeton, NJ (2006)

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

47. 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)

48. Ye J.J., Ye, X.Y.: Necessary optimality conditions for optimization problems with variational inequality constraints. Math. Oper. Res. 22, 977–997 (2007)