Journal of Optimization Theory and Applications
https://doi.org/10.1007/s10957-023-02338-6

Optimality Conditions for Nonlinear Second-Order Cone Programming and Symmetric Cone Programming

Roberto Andreani[1] · Ellen H. Fukuda[2] · Gabriel Haeser[3] · Daiana O. Santos[4] · Leonardo D. Secchin[5]

/ Accepted: 21 October 2023

Abstract
Nonlinear symmetric cone programming (NSCP) generalizes important optimization problems such as nonlinear programming, nonlinear semi-definite programming and nonlinear second-order cone programming (NSOCP). In this work, we present two new optimality conditions for NSCP without constraint qualifications, which implies the Karush–Kuhn–Tucker conditions under a condition weaker than Robinson's constraint qualification. In addition, we show the relationship of both optimality conditions in the context of NSOCP, where we also present an augmented Lagrangian method with global convergence to a KKT point under a condition weaker than Robinson's constraint qualification.

Keywords Second-order cones · Symmetric cones · Optimality conditions · Constraint qualifications · Augmented Lagrangian method

1 Introduction

The *nonlinear symmetric cone programming* (NSCP) problem is an optimization problem where the constraints are defined on a general symmetric cone. In recent years, the interest in NSCP has grown considerably. The reason is that many well-known

Communicated by Florian Potra.

✉ Gabriel Haeser
ghaeser@ime.usp.br

[1] Department of Applied Mathematics, University of Campinas, Campinas, SP, Brazil

[2] Graduate School of Informatics, Kyoto University, Kyoto, Japan

[3] Department of Applied Mathematics, University of São Paulo, São Paulo, SP, Brazil

[4] Paulista School of Politics, Economics and Business, Federal University of São Paulo, Osasco, SP, Brazil

[5] Department of Applied Mathematics, Federal University of Espírito Santo, São Mateus, ES, Brazil

Published online: 07 December 2023

optimization problems are particular cases of NSCP, including the *nonlinear programming* (NLP), the *nonlinear second-order cone programming* (NSOCP) and *nonlinear semidefinite programming* (NSDP) problems. The theoretical tool that allows us to unify the study of all these problems is the notion of Euclidean Jordan algebras. In particular, Faraut and Korányi studied in [27] the concept of cone of squares, where it is shown that every symmetric cone can be represented as a cone of squares of some Euclidean Jordan algebra. Although the interest in optimization problems with conic constraints is increasing, the studies related to these problems are still insufficient.

For the NLP case, sequential optimality conditions have been considered very useful in the past years due to the possibility of unifying and extending results of global convergence for several algorithms [5]. These necessary optimality conditions appeared in NLP as an alternative for conditions of the form “KKT or not-CQ,” which means that every local optimizer is either a *Karush–Kuhn–Tucker* (KKT) point, or it does not satisfy a *constraint qualification* (CQ). In fact, sequential optimality conditions are necessary for optimality without requiring a CQ. Besides, they imply a condition of the form “KKT or not-CQ” under so-called *strict* constraint qualifications [14] in place of “CQ,” which are stronger than the ones appearing in more theoretical venues but, more interestingly, are weaker than CQs usually employed in global convergence analysis (such as Mangasarian–Fromovitz CQ/Robinson’s CQ). Due to their relevance, sequential optimality conditions were extended to different classes of optimization problems such as multiobjective optimization [28, 32], NSDP [12], NLP with complementarity constraints [11, 46], generalized Nash equilibrium problems [25], optimization of discontinuous functions [21], optimization in Banach spaces [38], variational and quasi-variational inequalities [35, 39] and quasi-equilibrium problems [24].

One of the most relevant sequential optimality conditions is the so-called *Approximate-Karush–Kuhn–Tucker* (AKKT) condition, which is related to several first- and second-order algorithms. For more details, see [10, 13–15, 19]. In this paper, we consider an extension of the AKKT condition for NSCP problems. For measuring the complementarity condition, we use the eigenvalues of the constraint functions and of the approximate Lagrange multipliers. We also introduce another optimality condition that is more suited to the conic framework, by measuring the complementarity with the Jordan product. In the NLP case, it coincides with the stronger condition known as *Complementary-AKKT* (CAKKT) [16]. A detailed study for NSOCP is presented, where we show that CAKKT is in fact stronger than AKKT. We also present the global convergence of an augmented Lagrangian algorithm to points satisfying AKKT and, under an additional smoothness assumption, CAKKT. These optimality conditions are strictly better than Fritz-John’s, usually employed in this context.

This paper is organized as follows. In Sect. 2, we recall some theoretical results on Euclidean Jordan algebras. In Sect. 3, we propose the sequential optimality conditions for NSCP and present some of their properties. In Sect. 4, we analyze in detail the case of NSOCP. In Sect. 5, we introduce the augmented Lagrangian algorithm for NSOCP, and we show that the limit points of a sequence generated by the method satisfy our optimality conditions. An illustrative numerical experiment is also conducted. Finally, some conclusions are presented in Sect. 6.

The following notations will be adopted in this paper. For any matrix $A \in \mathbb{R}^{n\times\ell}$, its transpose is denoted by $A^T \in \mathbb{R}^{\ell\times n}$. A vector $z \in \mathbb{R}^{\ell}$ can be written as $(z_0, \bar{z}) \in$

$\mathbb{R} \times \mathbb{R}^{\ell-1}$. Let $h \colon \mathbb{R}^n \to \mathbb{R}^\ell$ be a function with components $h_i \colon \mathbb{R}^n \to \mathbb{R}$, with $i = 1, \ldots, \ell$. Then, for any $x \in \mathbb{R}^n$, we can write $h(x) = \left([h(x)]_0, \overline{h(x)}\right) \in \mathbb{R} \times \mathbb{R}^{\ell-1}$. Also, the Jacobian of h at $x \in \mathbb{R}^n$ is denoted by $Jh(x) \in \mathbb{R}^{\ell \times n}$, and the gradient of h_i at x is written as $\nabla h_i(x)$. In $\mathbb{R}^\ell$, the Euclidean norm and the Euclidean inner product will be denoted by $\| \cdot \|$ and $\langle \cdot, \cdot \rangle$, respectively. If $\mathcal{J}$ is a finite-dimensional inner product space and $P \colon \mathbb{R}^n \to \mathcal{J}$ is an operator, then the adjoint of P with respect to the Euclidean inner product in $\mathbb{R}^n$ is given by $P^* \colon \mathcal{J} \to \mathbb{R}^n$. We denote by $\operatorname{int}(C)$ the interior of a set C. The matrix I_ℓ is the $\ell \times \ell$ identity matrix. The ℓ-dimensional second-order cone is given by

$$K_\ell := \left\{ z = (z_0, \bar{z}) \in \mathbb{R} \times \mathbb{R}^{\ell-1} \;\middle|\; \begin{array}{ll} \|\bar{z}\| \leq z_0 & \text{for } \ell \geq 2 \\ z \geq 0 & \text{for } \ell = 1 \end{array} \right\}.$$

2 Symmetric Cones and Euclidean Jordan Algebras

In this section, we review some results on symmetric cones and Euclidean Jordan algebras. We refer to [18, 27] for more details. Here, let $\mathcal{E}$ be a finite-dimensional real vector space with an inner product $\langle \cdot, \cdot \rangle$. In order to define symmetric cones, we need the concept of self-dual and homogeneous cones. The dual of a cone $\mathcal{K} \subseteq \mathcal{E}$ is defined by $\mathcal{K}^* := \{u \in \mathcal{E} \mid \langle u, v \rangle \geq 0 \text{ for all } v \in \mathcal{K}\}$. Moreover, $\mathcal{K}$ is *self-dual* if $\mathcal{K} = \mathcal{K}^*$. Furthermore, the cone $\mathcal{K}$ is *homogeneous* if for each $u, v \in \operatorname{int}(\mathcal{K})$, there exists a linear bijection T such that $T(u) = v$ and $T(\mathcal{K}) = \mathcal{K}$.

Definition 2.1 The cone $\mathcal{K}$ is *symmetric* if it is self-dual, homogeneous and has nonempty interior.

Some well-known optimization problems are defined on symmetric cones, as in the case of: the nonlinear programming (NLP) problem on the nonnegative orthant, the nonlinear second-order cone programming (NSOCP) problem on the second-order cone (or Lorentz cone), the nonlinear semidefinite programming (NSDP) problem on the positive semidefinite cone, and others. The concept of Euclidean Jordan algebra allows us to unify the study of NLP, NSOCP, NSDP, and other problems obtained through a mix of different conic constraints by means of a symmetric conic constraint. The definition is given as follows.

Definition 2.2 Let $\mathcal{E}$ be a finite-dimensional real vector space with an inner product $\langle \cdot, \cdot \rangle$ and a bilinear operator $\circ : \mathcal{E} \times \mathcal{E} \to \mathcal{E}$. Then, we say that $(\mathcal{E}, \circ)$ is an *Euclidean Jordan algebra* if for all $u, v, w \in \mathcal{E}$:

(i) $u \circ v = v \circ u$,
(ii) $u \circ (u^2 \circ v) = u^2 \circ (u \circ v)$, where $u^2 = u \circ u$,
(iii) $\langle v \circ w, u \rangle = \langle v, w \circ u \rangle$.

It is well known that an Euclidean Jordan algebra $\mathcal{E}$ can be decomposed as a direct sum of *simple* Euclidean Jordan algebras

$$\mathcal{E} = \mathcal{E}_1 \oplus \mathcal{E}_2 \oplus \cdots \oplus \mathcal{E}_s, \tag{1}$$

which induces a decomposition of a symmetric cone $\mathcal{K}$ as a direct sum of symmetric cones, i.e.,

$$\mathcal{K} = \mathcal{K}_1 \oplus \mathcal{K}_2 \oplus \cdots \oplus \mathcal{K}_s,$$

where each algebra $(\mathcal{E}_i, \circ)$ is such that it is not possible to obtain nonzero sub-algebras U_i, V_i of $\mathcal{E}_i$ such that $\mathcal{E}_i = U_i \oplus V_i$. Let us now present a characterization of symmetric cones in terms of cones of squares.

Theorem 2.1 ([27, Theorem III.2.1]). *Let $\mathcal{E}$ be a finite-dimensional inner product space. A cone $\mathcal{K} \subseteq \mathcal{E}$ is symmetric if and only if $\mathcal{K}$ is a cone of squares of some Euclidean Jordan algebra $(\mathcal{E}, \circ)$, i.e., $\mathcal{K} = \{u \circ u \mid u \in \mathcal{E}\}$.*

With the above theorem, we can use Euclidean Jordan algebra theory to study properties of symmetric cones. In particular, given a symmetric cone $\mathcal{K}$ we will consider its natural Euclidean Jordan algebra such that $\mathcal{K}$ is its cone of squares. For example, we can characterize the second-order cone $K_\ell \subset \mathbb{R}^\ell$ as a cone of squares defining the Jordan product

$$(z_0, \overline{z}) \circ (y_0, \overline{y}) = (\langle z, y\rangle, z_0\overline{y} + y_0\overline{z}), \tag{2}$$

for all $z = (z_0, \overline{z}) \in \mathbb{R} \times \mathbb{R}^{\ell-1}$ and $y = (y_0, \overline{y}) \in \mathbb{R} \times \mathbb{R}^{\ell-1}$. Moreover, to characterize the positive semidefinite cone $\mathbb{S}^m_+$ as a cone of squares, we define $X \circ Y = (YX + XY)/2$ for $X, Y \in \mathbb{S}^m$, where $\mathbb{S}^m$ is the linear space of all real symmetric matrices with dimension $m \times m$ equipped with the inner product $\langle X, Y\rangle := \operatorname{tr}(XY)$. Considering the nonnegative orthant $\mathbb{R}^m_+$ equipped with the Euclidean inner product, the Jordan product is the usual Hadamard product. Now, we present the notion of a spectral decomposition in Euclidean Jordan algebras, which is a natural extension of the usual decomposition for symmetric matrices.

Theorem 2.2 ([27, Theorem III.1.2]). *Let $(\mathcal{E}, \circ)$ be an Euclidean Jordan algebra and $u \in \mathcal{E}$. Then, there exist the so-called idempotents $0 \neq c_i(u) \in \mathcal{E}$, $i = 1, \ldots, r$ satisfying*

$$c_i(u) \circ c_j(u) = 0, \quad i \neq j, \tag{3}$$

$$c_i(u) \circ c_i(u) = c_i(u), \quad i = 1, \ldots, r,$$

$$c_1(u) + \cdots + c_r(u) = \boldsymbol{e}, \quad i = 1, \ldots, r, \tag{4}$$

and the so-called eigenvalues $\lambda_i(u) \in \mathbb{R}$ with $i = 1, \ldots, r$ such that

$$u = \lambda_1(u)c_1(u) + \cdots + \lambda_r(u)c_r(u),$$

where $\boldsymbol{e}$ satisfies $u \circ \boldsymbol{e} = \boldsymbol{e} \circ u = u$ for all u.

The number r, which is the rank of the algebra, and the identity element $\mathbf{e}$ in the above theorem are uniquely defined. We also say that $c_i(u)$, $i = 1, \ldots, r$, in the previous theorem form a Jordan frame for u. We can see in [27, Proposition

II.2.] that the functions λ_i, when the order is fixed, are all continuous over $\mathcal{E}$ and uniquely determined by u. By (4) and Definition 2.2 (iii), we have that $\langle c_i(u), c_j(u)\rangle = \langle c_i(u)\circ c_i(u), c_j(u)\rangle = \langle c_i(u), c_i(u)\circ c_j(u)\rangle$ which is equal to zero by (3) when $i \neq j$, and hence, the idempotents are orthogonal.

The elements $u \in \mathcal{E}$ and $v \in \mathcal{E}$ are said to *operator commute* if they share a Jordan frame, that is, if there exists a common Jordan frame $\{c_1, \dots, c_r\}$ such that

$$u = \lambda_1(u)c_1 + \cdots + \lambda_r(u)c_r \quad \text{and} \quad v = \lambda_1(v)c_1 + \cdots + \lambda_r(v)c_r.$$

For a symmetric matrix, its spectral decomposition coincides with the classic eigenvalue decomposition, the idempotents being the outer product of the eigenvectors, namely $c_i(u) = q_i(u)q_i(u)^T$, where $q_i(u)$ is the corresponding eigenvector of the symmetric matrix u. For the particular case of the Euclidean Jordan algebra associated with the second-order cone $K_\ell \subset \mathbb{R}^\ell$, the spectral decomposition of $z \in \mathbb{R}^\ell$ is given by $z = \lambda_-(z)c_-(z) + \lambda_+(z)c_+(z)$ where $\lambda_-(z) = z_0 - \|\overline{z}\|$, $\lambda_+(z) = z_0 + \|\overline{z}\|$ and

$$c_-(z) = \begin{cases} \frac{1}{2}\left(1, -\frac{\overline{z}}{\|\overline{z}\|}\right) & \text{if } \overline{z} \neq 0, \\ \frac{1}{2}(1, -\overline{w}) & \text{if } \overline{z} = 0, \end{cases} \qquad c_+(z) = \begin{cases} \frac{1}{2}\left(1, \frac{\overline{z}}{\|\overline{z}\|}\right) & \text{if } \overline{z} \neq 0, \\ \frac{1}{2}(1, \overline{w}) & \text{if } \overline{z} = 0, \end{cases} \tag{5}$$

where $\overline{w}$ is any vector in $\mathbb{R}^{\ell-1}$ such that $\|\overline{w}\| = 1$. See, for instance, [31] for spectral properties of the second-order cone.

The following result summarizes some interesting properties that relate the inner product, the Jordan product and the eigenvalues. See, for instance, [43, Proposition 2.1] and [27, Theorem III.4.1].

Proposition 2.1 *Let $(\mathcal{E}, \circ)$ be an Euclidean Jordan algebra, with $\mathcal{K}$ as its cone of squares. For $u, v \in \mathcal{E}$, the following properties hold.*

(i) $u \in \mathcal{K}$ $(u \in \operatorname{int}(\mathcal{K}))$ *if and only if the eigenvalues of u are nonnegative (positive);*
(ii) *if $u, v \in \mathcal{K}$, then $u \circ v = 0$ if and only if $\langle u, v\rangle = 0$;*
(iii) *if $u, v \in \mathcal{K}$ is such that $\langle u, v\rangle = 0$, then u and v operator commute,*
(iv) *when the algebra is simple, $\langle u, v\rangle = \theta \operatorname{tr}(u \circ v)$ where $\operatorname{tr}(u) = \sum_{i=1}^r \lambda_i(u)$ and θ is constant.*

In item (iv), $\theta = 1/2$ for a second-order cone of dimension at least 2, and $\theta = 1$ for the semidefinite cone. The projection onto a symmetric cone $\mathcal{K}$ is also easily computed when the spectral decomposition is available.

Proposition 2.2 [43, Proposition 2.5] *Let $(\mathcal{E}, \circ)$ be an Euclidean Jordan algebra, with $\mathcal{K}$ as its cone of squares. Let $u = \sum_{i=1}^r \lambda_i(u)c_i(u)$ be a spectral decomposition of $u \in \mathcal{E}$. Then, the projection of u onto $\mathcal{K}$ is given by*

$$[u]_+ = \sum_{i=1}^r \max\{0, \lambda_i(u)\}c_i(u).$$

The following proposition will be relevant in our analysis:

Proposition 2.3 *Let $(\mathcal{E}, \circ)$ be an Euclidean Jordan algebra of rank r and $\{u^k\} \subset \mathcal{E}$ with $u^k \to u$. Then, there exists a subsequence of the Jordan frames $\{c_i(u^k)\}_{i=1}^r$ of u^k converging to a Jordan frame $\{c_i(u)\}_{i=1}^r$ of u.*

Proof It follows from the boundedness of the sequence of Jordan frames, continuity of eigenvalues and continuity of $\circ$. See, for instance, the proof of Theorem 2.7.25 in [17]. □

3 Nonlinear Symmetric Cone Programming

The *nonlinear symmetric cone programming* (NSCP) problem is defined as follows:

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & F(x), \\ \text{subject to} \quad & G(x) \in \mathcal{K}, \end{aligned} \tag{NSCP}$$

where $F: \mathbb{R}^n \to \mathbb{R}$ and $G: \mathbb{R}^n \to \mathcal{E}$ are continuously differentiable functions, $\mathcal{K} \subseteq \mathcal{E}$ is a symmetric cone, and $\mathcal{E}$ is a finite-dimensional inner product space. The Lagrangian function $\mathcal{L}: \mathbb{R}^n \times \mathcal{E} \to \mathbb{R}$ of (NSCP) is defined by

$$\mathcal{L}(x, \sigma) := F(x) - \langle G(x), \sigma \rangle.$$

We say that $(x, \sigma) \in \mathbb{R}^n \times \mathcal{E}$ is a KKT pair for (NSCP) if the following conditions are satisfied:

$$\begin{aligned} \nabla \mathcal{L}(x, \sigma) = \nabla F(x) - JG(x)^* \sigma &= 0, \\ \langle G(x), \sigma \rangle &= 0, \\ G(x) &\in \mathcal{K}, \\ \sigma &\in \mathcal{K}, \end{aligned} \tag{6}$$

where $JG(x): \mathbb{R}^n \to \mathcal{E}$ is the Jacobian of G at x and $JG(x)^*$ is its adjoint. By considering the Euclidean Jordan algebra $(\mathcal{E}, \circ)$ such that $\mathcal{K}$ is its cone of squares, by Proposition 2.1 we can replace condition (6) by $G(x) \circ \sigma = 0$. Since this implies that $G(x)$ and σ operator commute, it is easy to see that condition (6) can also be replaced by

$$\lambda_i(G(x))\lambda_i(\sigma) = 0, \quad i = 1, \ldots, r,$$

where the ordering of the eigenvalues is the order given by their common Jordan frame.

It is known that some constraint qualification is required in order to make KKT conditions hold at local optimal solutions. As well-known constraint qualifications for (NSCP), we can cite nondegeneracy and Robinson's CQ.

Definition 3.1 We say that $x \in \mathbb{R}^n$ satisfies the *nondegeneracy* condition if

$$\operatorname{Im} JG(x) + \mathcal{T}_{\mathcal{K}}^{\mathrm{lin}}(G(x)) = \mathcal{E},$$

where $\mathcal{T}_{\mathcal{K}}^{\mathrm{lin}}(G(x))$ is the lineality space of the tangent cone of $\mathcal{K}$ at $G(x)$ and $\operatorname{Im} JG(x)$ denotes the image of the linear mapping $JG(x)$.

Nondegeneracy was studied by Bonnans and Shapiro in [22], which is also related to the so-called transversality condition. When nondegeneracy is satisfied at a point, the associated Lagrange multiplier is unique. When $\mathcal{T}_{\mathcal{K}}^{\mathrm{lin}}(G(x))$ is replaced by the tangent cone of $\mathcal{K}$ at $G(x)$ in the definition of nondegeneracy, we arrive at Robinson's CQ, which is clearly weaker than nondegeneracy and can also be defined in the following equivalent way:

Definition 3.2 We say that $x \in \mathbb{R}^n$ satisfies *Robinson's CQ* if there exists $d \in \mathbb{R}^n$ such that

$$G(x) + JG(x)d \in \operatorname{int}(\mathcal{K}).$$

When Robinson's CQ is satisfied, the set of Lagrange multipliers is nonempty and bounded.

3.1 New Optimality Conditions

As previously stated, in general, most optimality conditions are of the form "KKT or not-CQ." An alternative to this type of optimality conditions, that do not require any CQ, is a so-called *sequential optimality condition*. The sequence needed for checking this condition is usually the one generated by standard algorithms. This provides global convergence results stronger than the usual ones. In this section, we present an extension of the condition called Approximate-Karush–Kuhn–Tucker (AKKT) that was studied in [5, 45] for NLP and [12] for NSDP. In NLP, the AKKT condition is a strong optimality condition satisfied by limit points of many first- and second-order methods. For more details, see [20, 21, 26, 33–35, 44, 47]. We will also introduce a generalization of the Complementary-AKKT (CAKKT) condition introduced for NLP in [16], which is stronger than AKKT. This study was initiated in [12] but the notion of the Jordan product simplifies this task. In [12], the notion of a Trace-AKKT (TAKKT) point was introduced as a substitute for CAKKT, but here we will shed a light in this discussion by the introduction of CAKKT while clarifying the relationship between AKKT and TAKKT, which was an open question in [12]. Let us start by presenting our definition of AKKT for NSCP.

Definition 3.3 Let $x^* \in \mathbb{R}^n$ be a feasible point. We say that x^* is an *Approximate-Karush–Kuhn–Tucker* (AKKT) point for (NSCP) if there exist sequences $\{x^k\} \subset \mathbb{R}^n$

and $\{\sigma^k\} \subset \mathcal{K}$ with $x^k \to x^*$ such that

$$\lim_{k\to\infty} \nabla\mathcal{L}(x^k, \sigma^k) = 0, \tag{7}$$

$$\text{if } \lambda_i(G(x^*)) > 0 \text{ then } \lambda_i(\sigma^k) = 0 \quad \text{for sufficiently large } k, \tag{8}$$

$$\lim_{k\to\infty} c_i(\sigma^k) = c_i(G(x^*)), \tag{9}$$

for all $i = 1, \ldots, r$, where

$$G(x^*) = \lambda_1(G(x^*))c_1(G(x^*)) + \cdots + \lambda_r(G(x^*))c_r(G(x^*)), \tag{10}$$

$$\sigma^k = \lambda_1(\sigma^k)c_1(\sigma^k) + \cdots + \lambda_r(\sigma^k)c_r(\sigma^k), \tag{11}$$

are spectral decompositions of $G(x^*)$ and σ^k, respectively.

Notice that the definition of the AKKT condition is independent of the choices of $c_i(G(x^*))$ and $c_i(\sigma^k)$ for $i = 1, \ldots, r$. We recall that in the definition of AKKT for NLP, the Lagrange multipliers associated with inactive constraints at x^* are taken equal to zero. In NSCP, we need a notion that gives us a way to associate the "inactive" eigenvalues of $G(x^*)$ with the zero eigenvalues of σ^k. In the above definition, the relation (9) provides the necessary condition for pairing the eigenvalues correctly. The equivalence below provides a way of detecting an AKKT sequence in terms of a sequence of tolerances $\{\varepsilon_k\}$ that converges to zero.

Lemma 3.1 *A feasible point $x^* \in \mathbb{R}^n$ satisfies the AKKT condition for (NSCP) if, and only if, there are sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\sigma^k\} \subset \mathcal{K}$, $\{\epsilon_k\} \subset \mathbb{R}_+$ with $x^k \to x^*$, $\epsilon_k \to 0$, such that*

$$\|\nabla\mathcal{L}(x^k, \sigma^k)\| \le \epsilon_k, \tag{12}$$

$$\|[-G(x^k)]_+\| \le \epsilon_k, \tag{13}$$

$$\lambda_i(G(x^k)) > \epsilon_k \Rightarrow \lambda_i(\sigma^k) = 0, \quad \textit{for sufficiently large } k, \tag{14}$$

$$\|c_i(\sigma^k) - c_i(G(x^k))\| \le \epsilon_k, \tag{15}$$

for all $i = 1, \ldots, r$, where

$$G(x^k) = \lambda_1(G(x^k))c_1(G(x^k)) + \cdots + \lambda_r(G(x^k))c_r(G(x^k)),$$
$$\sigma^k = \lambda_1(\sigma^k)c_1(\sigma^k) + \cdots + \lambda_r(\sigma^k)c_r(\sigma^k),$$

are spectral decompositions of $G(x^k)$ and σ^k, respectively.

Proof Let us assume that $x^* \in \mathbb{R}^n$ satisfies the AKKT condition. By the definition of AKKT, we can take the spectral decompositions for $G(x^*)$ and σ^k given by (10) and (11) with $c_i(G(x^k)) \to c_i(G(x^*))$ and $c_i(\sigma^k) \to c_i(G(x^*))$ for $i = 1, \ldots, r$, such that (7)–(9) hold. Now, note that

$$\|[-G(x^k)]_+\| = \left\| \sum_{i=1}^{r} \max\{0, -\lambda_i(G(x^k))\} c_i(G(x^k)) \right\|$$
$$\to \left\| \sum_{i=1}^{r} \max\{0, -\lambda_i(G(x^*))\} c_i(G(x^*)) \right\| = 0, \tag{16}$$

because x^* is feasible. Define the sequence $\{\epsilon_k\} \subset \mathbb{R}_+$ as follows:

$$\epsilon_k := \max \Big\{ \|\nabla \mathcal{L}(x^k,\ \sigma^k)\|,\ \|[-G(x^k)]_+\|,\ \lambda_i(G(x^k)) : i \in I(x^*),$$
$$\|c_i(\sigma^k) - c_i(G(x^k))\| : i = 1, \ldots, r \Big\},$$

where $I(x^*) := \{i \mid \lambda_i(G(x^*)) = 0\}$. Observe that the limit for $k \to \infty$ of each term inside the above maximum is zero from (7), (9) and (16). By the continuity of the involved functions, we have that $\epsilon_k \to 0$. Hence, (12), (13) and (15) hold. Now, let k be sufficiently large. To prove (14), note that for $j \in \{1, \ldots, r\}$ such that $\lambda_j(G(x^k)) > \epsilon_k$ we have that

$$\lambda_j(G(x^k)) > \epsilon_k \geq \lambda_i(G(x^k)) \quad \text{for all} \quad i \in I(x^*).$$

In particular, $j \notin I(x^*)$, that is, $\lambda_j(G(x^*)) > 0$. Hence, from (8), $\lambda_j(\sigma^k) = 0$, and so (14) holds.

Let us now assume that there are $\{x^k\}$, $\{\sigma^k\}$, $\{\epsilon_k\}$ satisfying $x^k \to x^*$, $\epsilon_k \to 0$ and (12)–(15). The continuity of the involved functions and (13) ensure that x^* is feasible. The limit (7) follows trivially from (12). Since $\{c_i(\sigma^k)\}$ and $\{c_i(G(x^k))\}$ are bounded for all $i = 1, \ldots, r$, we may take a subsequence if necessary such that $c_i(G(x^k)) \to c_i(G(x^*))$ for all $i = 1, \ldots, r$. Hence, (9) follows from (15). Now, if we suppose that $\lambda_i(G(x^*)) > 0$, then $\lambda_i(G(x^k)) > \epsilon_k$ for k large enough. Thus, by (14) we have that $\lambda_i(\sigma^k) = 0$, which means that (8) holds. Therefore, $x^* \in \mathbb{R}^n$ is an AKKT point. $\square$

We can use the previous lemma to define a simple stopping criterion for algorithms for NSCP. Let $\epsilon_{\text{opt}}, \epsilon_{\text{feas}}, \epsilon_{\text{compl}}$ and ϵ_{spec} be small tolerances associated with optimality, feasibility, complementarity and spectral decomposition, respectively. Then, an algorithm for solving NSCP that generates an AKKT sequence $\{x^k\} \subset \mathbb{R}^n$ and a dual sequence $\{\sigma^k\} \subset \mathcal{K}$ can be safely stopped when:

$$\|\nabla \mathcal{L}(x^k, \sigma^k)\| \leq \epsilon_{\text{opt}},$$
$$\|[-G(x^k)]_+\| \leq \epsilon_{\text{feas}},$$
$$\lambda_i(G(x^k)) > \epsilon_{\text{compl}} \quad \Rightarrow \quad \lambda_i(\sigma^k) = 0 \quad \text{for all } i,$$
$$\|c_i(\sigma^k) - c_i(G(x^k))\| \leq \epsilon_{\text{spec}} \quad \text{for all } i.$$

In order to prove that the AKKT condition necessarily holds at local minimizers of an NSCP problem, we will use the standard external penalty technique. More precisely, consider the problem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & F(x), \\ \text{subject to} \quad & G(x) \in \mathcal{K},\ x \in \Omega, \end{aligned} \tag{17}$$

where $\Omega \subseteq \mathbb{R}^n$ is a nonempty closed set. For this problem, the function $P(x) := \| [-G(x)]_+ \|^2$ is a measure of infeasibility, in the sense that $P(\cdot)$ is continuous and nonnegative, and x is feasible for (17) if, and only if, $P(x) = 0$ and $x \in \Omega$. This function can be used for constructing a penalized problem that has the property described below.

Lemma 3.2 [29], [19, pp. 16, 26] *Choose a sequence $\{\rho_k\} \subset \mathbb{R}$ with $\rho_k \to +\infty$. For each k, let x^k be a solution, if it exists, for the following penalized problem*

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\mathit{Minimize}} \quad & F(x) + \rho_k P(x), \\ \mathit{subject\ to} \quad & x \in \Omega. \end{aligned}$$

Then, all limit points of $\{x^k\}$ are global minimizers of (17).

Theorem 3.1 *Let $x^* \in \mathbb{R}^n$ be a local minimizer of (NSCP). Then, x^* satisfies the AKKT condition.*

Proof Consider the problem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & F(x) + \frac{1}{2}\|x - x^*\|^2, \\ \text{subject to} \quad & G(x) \in \mathcal{K}, \\ & x \in \mathcal{B}(x^*, \delta), \end{aligned} \tag{18}$$

where $\mathcal{B}(x^*, \delta) := \{x \in \mathbb{R}^n \mid \|x - x^*\| \leq \delta\}$ for $\delta > 0$ such that x^* is the unique solution of (18). Let $x^k \in \mathbb{R}^n$ be a solution of

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & F(x) + \frac{1}{2}\|x - x^*\|^2 + \frac{\rho_k}{2}\|[-G(x)]_+\|^2, \\ \text{subject to} \quad & x \in \mathcal{B}(x^*, \delta). \end{aligned} \tag{19}$$

By the continuity of the involved functions and the compactness of $\mathcal{B}(x^*, \delta)$, the sequence $\{x^k\}$ is well defined for all k.

In addition, the set $\mathcal{B}(x^*, \delta)$ is nonempty and compact, hence, by Lemma 3.2 we have that all limit points of $\{x^k\}$ are global solutions of (18) and $x^k \to x^*$. Now, for large enough k, we have that x^k is a local minimizer of $F(x)+\frac{1}{2}\|x-x^*\|^2+\frac{\rho_k}{2}\|[-G(x)]_+\|^2$. Then, we obtain

$$\nabla F(x^k) + (x^k - x^*) - \rho_k JG(x^k)^*[-G(x^k)]_+ = 0. \tag{20}$$

Let us define for all k, $\sigma^k := \rho_k[-G(x^k)]_+ \in \mathcal{K}$. Let us also consider the spectral decompositions

$$G(x^k) = \lambda_1(G(x^k))c_1(G(x^k)) + \cdots + \lambda_r(G(x^k))c_r(G(x^k)),$$
$$\sigma^k = \rho_k \max\{0, \lambda_1(-G(x^k))\}c_1(G(x^k)) + \cdots + \rho_k \max\{0, \lambda_r(-G(x^k))\}c_r(G(x^k)).$$

Note that $-\lambda_i(G(x^k)) = \lambda_i(-G(x^k))$ for all $i = 1, \ldots, r$. If $\lambda_i(G(x^*)) > 0$, then $\lambda_i(G(x^k)) > 0$ for k large enough. Thus, $\lambda_i(\sigma^k) = \rho_k \max\{0, \lambda_i(-G(x^k))\} = 0$. Moreover, $c_i(\sigma^k) = c_i(G(x^k)) \to c_i(G(x^*))$ is trivially satisfied. Taking the limit when $k \to \infty$ in (20), we have

$$\lim_{k\to\infty} \nabla\mathcal{L}(x^k, \sigma^k) = \lim_{k\to\infty} \nabla F(x^k) - JG(x^k)^*\sigma^k = 0.$$

Hence, x^* satisfies the AKKT condition. □

There are many different and equivalent ways to measure complementarity in the KKT conditions. However, they give rise to different sequential optimality conditions. We introduced the AKKT condition by using eigenvalues to measure complementarity, and we now present a definition of CAKKT in the context of NSCP by measuring complementarity with the Jordan product. This is a natural way of avoiding the computation of eigenvalues in the context of NSCP. Also, our definition will coincide with the well-known CAKKT for NLP in the particular case of the nonnegative orthant. The definition is as follows:

Definition 3.4 Let $x^* \in \mathbb{R}^n$ be a feasible point. We say that x^* is a *Complementary–Approximate–Karush–Kuhn–Tucker* (CAKKT) point for (NSCP) if there exist sequences $\{x^k\} \subset \mathbb{R}^n$ and $\{\sigma^k\} \subset \mathcal{K}$ with $x^k \to x^*$ such that

$$\nabla\mathcal{L}(x^k, \sigma^k) \to 0, \tag{21}$$
$$G(x^k) \circ \sigma^k \to 0. \tag{22}$$

Theorem 3.2 *Let $x^* \in \mathbb{R}^n$ be a local minimizer of* (NSCP)*. Then, x^* satisfies the CAKKT condition.*

Proof Considering that the sequences $\{x^k\}$ and $\{\sigma^k\}$ built in the proof of Theorem 3.1, it remains to prove (22). We start by noting that since x^* is a feasible point of (19) for all k, and x^k is the corresponding global minimizer, the following holds for all k:

$$F(x^k) + \frac{1}{2}\|x^k - x^*\|^2 + \frac{\rho_k}{2}\|[-G(x^k)]_+\|^2 \le F(x^*).$$

This implies that $\rho_k\|[-G(x^k)]_+\|^2 \to 0$. However, we have

$$\begin{aligned}\|[-G(x^k)]_+\|^2 &= \langle[-G(x^k)]_+, [-G(x^k)]_+\rangle \\ &= \sum_{i=1}^{r} \|c_i(G(x^k))\|^2 \max\{0, -\lambda_i(G(x^k))\}^2,\end{aligned}$$

where $G(x^k) = \sum_{i=1}^r \lambda_i(G(x^k))c_i(G(x^k))$.

It follows that $\rho_k \max\{0, -\lambda_i(G(x^k))\}^2 \to 0$ for all $i = 1, \ldots, r$.

Now, since $\max\{0, -\lambda_i(G(x^k))\}^2 = -\lambda_i(G(x^k)) \max\{0, -\lambda_i(G(x^k))\}$ and $\{c_i(x^k)\}$ are bounded for all i, we conclude that

$$G(x^k) \circ \sigma^k = \sum_{i=1}^r \rho_k \lambda_i(G(x^k)) \max\{0, -\lambda_i(G(x^k))\} c_i(G(x^k)) \to 0,$$

which completes the proof. □

Note that in the case of NLP where the cone is the nonnegative orthant, since the Jordan product is the Hadamard product, (21)–(22) reduce to the usual CAKKT condition [16]. Also, since $\langle G_i(x^k), \sigma_i^k \rangle = \theta_i \operatorname{tr}(G_i(x^k) \circ \sigma_i^k)$, $i = 1, \ldots s$ (considering the decomposition (1) and Proposition 2.1 item iv), one can easily extend the definition of Trace-AKKT (TAKKT) from NSDP [12] to general NSCP by replacing (22) by the weaker statement

$$\langle G(x^k), \sigma^k \rangle \to 0,$$

that is, CAKKT implies TAKKT. Hence, by Theorem 3.2, TAKKT is also a necessary optimality condition for NSCP.

In the context of NSDP [12], for avoiding eigenvalues, it was natural to consider the inner product, giving rise to the TAKKT condition. However, TAKKT is somewhat not natural in the context of NLP. In fact, in [12], the relationship of TAKKT and AKKT was not known. The following example shows that these conditions are independent conditions.

Example 3.1 (TAKKT does not imply AKKT) Consider the following NLP problem:

$$\begin{aligned}
\underset{x \in \mathbb{R}^3}{\text{Minimize}} \quad & F(x) := (x_2 - 2)^2/2 + x_3/2, \\
\text{subject to} \quad & x_1^2 \geq 0, \\
& -x_1 x_2 \geq 0, \\
& -x_1^2 x_2^2 + e^{x_3} \geq 0.
\end{aligned}$$

Let us prove that TAKKT holds at $x^* := (0, 1, 0)$. Let the primal sequence $\{x^k\}$ be defined by $x^k := (1/k, 1, 0)$ and the dual sequence $\{\sigma^k\}$ be defined by $\sigma^k := (k^2/2, k, 1/2)$ for all k. Hence, $\nabla \mathcal{L}(x^k, \sigma^k) = (1/k, 1/k^2, 0) \to 0$ and

$$\left\langle \sigma^k, ((x_1^k)^2, x_1^k x_2^k, (x_1^k)^2 (x_2^k)^2 - e^{x_3^k}) \right\rangle = 1/(2k^2) \to 0,$$

which implies that TAKKT holds. Note, however, that any sequences $\{\bar{x}^k\}$, $\{\bar{\sigma}^k\}$ such that $\nabla \mathcal{L}(\bar{x}^k, \bar{\sigma}^k) \to 0$ are such that the approximate Lagrange multiplier $\bar{\sigma}_3^k$ associated with the third constraint converges to $1/2$. Since the third constraint is inactive at x^*, this shows that AKKT does not hold.

Since it is known [12] that AKKT does not imply TAKKT, the example above shows that these conditions are independent. This is somewhat surprising as AKKT is considered to be the simplest condition, at least in the context of NLP. Despite TAKKT not implying even AKKT, that is, an approximate Lagrange multiplier for TAKKT may fail in detecting inactive constraints, it is remarkable that TAKKT is a necessary optimality condition still strictly stronger than Fritz-John's optimality condition. In this context, our extension of CAKKT fits much better our purpose of presenting an optimality condition free of eigenvalue computations. The strength of CAKKT will be shown in the context of NSOCP by showing that it implies AKKT.

Now, to fix ideas, let us revisit [12, Example 1], where our optimality conditions are verified at a local minimizer of a NSDP that does not satisfy the KKT conditions.

Example 3.2 Consider the NSDP problem:

$$\begin{aligned} \underset{x\in\mathbb{R}}{\text{Minimize}} \quad & 2x, \\ \text{subject to} \quad & G(x) := \begin{bmatrix} 0 & -x \\ -x & 1 \end{bmatrix} \in \mathbb{S}^2_+ \end{aligned}$$

at its unique global minimizer $x^* := 0$. Since there is no $d \in \mathbb{R}$ such that

$$G(x^*) + JG(x^*)d := \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + d \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \in \operatorname{int}(\mathbb{S}^2_+),$$

Robinson's CQ fails. (Hence, Fritz-John's condition holds.) The sequences $\{x^k\}$ and $\{\sigma^k\}$ defined by

$$x^k := -\frac{k}{k^2-1} \quad \text{and} \quad \sigma^k := \begin{bmatrix} k & -1 \\ -1 & 1/k \end{bmatrix} \in \mathbb{S}^2_+$$

attest that x^* is an AKKT point (see [12]). To see that CAKKT holds at x^* (thus, also TAKKT holds), it is enough to compute $G(x^k) \circ \sigma^k := \frac{1}{2}(G(x^k)\sigma^k + \sigma^k G(x^k))$ and note that it converges to zero. A simple inspection shows that the KKT conditions do not hold [12].

3.2 The Strength of the Optimality Conditions

In this section, we will measure the strength of AKKT and CAKKT for NSCP in comparison with an optimality condition of the type "KKT or not-CQ." Let us show that our necessary optimality condition is at least as good as "KKT or not-Robinson's CQ" (also called Fritz-John's condition).

Theorem 3.3 *Let $x^* \in \mathbb{R}^n$ be an AKKT or CAKKT point that satisfies Robinson's CQ. Then, x^* satisfies the KKT condition.*

Proof We will prove only the case where AKKT holds. The case where CAKKT holds can be deduced easily from the proof below. Let $\{x^k\} \subset \mathbb{R}^n$ and $\{\sigma^k\} \subset \mathcal{K}$ with $x^k \to x^*$ be such that

$$\lim_{k\to\infty} \nabla\mathcal{L}(x^k, \sigma^k) = 0, \tag{23}$$

$$\text{If } \lambda_i(G(x^*)) > 0 \text{ then } \lambda_i(\sigma^k) = 0 \text{ for sufficiently large } k, \tag{24}$$

$$\lim_{k\to\infty} c_i(\sigma^k) = c_i(G(x^*)), \tag{25}$$

for all $i = 1, \ldots, r$, where

$$\begin{aligned} G(x^*) &= \lambda_1(G(x^*))c_1(G(x^*)) + \cdots + \lambda_r(G(x^*))c_r(G(x^*)), \\ \sigma^k &= \lambda_1(\sigma^k)c_1(\sigma^k) + \cdots + \lambda_r(\sigma^k)c_r(\sigma^k), \end{aligned} \tag{26}$$

are spectral decompositions of $G(x^*)$ and σ^k, respectively.

If $\{\sigma^k\}$ is contained in a compact set, there exists $\sigma \in \mathcal{K}$ such that, taking a subsequence if necessary, $\sigma^k \to \sigma$. Then, by (23) we have that

$$\nabla F(x^*) - JG(x^*)^*\sigma = 0.$$

Now, consider the index sets

$$I_1 := \{i \mid \lambda_i(G(x^*)) > 0\} \quad \text{and} \quad I_2 := \{i \mid \lambda_i(G(x^*)) = 0\}.$$

Note that, by (25), $G(x^*)$ and σ operator commute. Then, we obtain

$$G(x^*) \circ \sigma = \sum_{i\in I_1} \lambda_i(G(x^*))\lambda_i(\sigma)c_i(G(x^*)) + \sum_{i\in I_2} \lambda_i(G(x^*))\lambda_i(\sigma)c_i(G(x^*)) = 0,$$

since $\lambda_i(G(x^*)) = 0$ for $i \in I_2$ and $\lambda_i(\sigma) = 0$ for $i \in I_1$, which implies KKT. Now, suppose that $\{\sigma^k\}$ is not contained in a compact set. Let us consider a subsequence such that $t_k := \|\sigma^k\| \to \infty$. Then, $\dfrac{\sigma^k}{t_k} \to \sigma^* \neq 0$ for some $\sigma^* \in \mathcal{K}$. Then, by (23) and (24) we have that

$$\lim_{k\to\infty} \frac{\nabla F(x^k)}{t_k} - JG(x^k)^*\frac{\sigma^k}{t_k} = -JG(x^*)^*\sigma^* = 0, \tag{27}$$

and by taking the limit in (26),

$$\begin{aligned} G(x^*) \circ \sigma^* = &\sum_{i\in I_1} \lambda_i(G(x^*))\lambda_i(\sigma^*)c_i(G(x^*)) \\ &+ \sum_{i\in I_2} \lambda_i(G(x^*))\lambda_i(\sigma^*)c_i(G(x^*)) = 0, \end{aligned} \tag{28}$$

To see that (27) and (28) contradict Robinson's condition, let $d \in \mathbb{R}^n$ be such that

$$G(x^*) + JG(x^*)d \in \text{int}(\mathcal{K}).$$

Thus, we have

$$\begin{aligned} 0 &= \langle G(x^*), \sigma^* \rangle + \langle JG(x^*)^* \sigma^*, d \rangle \\ &= \langle G(x^*) + JG(x^*)d, \sigma^* \rangle. \end{aligned}$$

Since $0 \neq \sigma^* \in \mathcal{K}$ and $w := G(x^*) + JG(x^*)d \in \text{int}(\mathcal{K})$, we have that $\sigma^* - \varepsilon w \in \text{int}(\mathcal{K})$ for some $\varepsilon > 0$, and hence, $\langle \sigma^* - \varepsilon w, w \rangle < 0$, which contradicts the self-duality of $\mathcal{K}$. □

Even in the context of NLP, it is well known that AKKT or CAKKT is strictly stronger than "KKT or not-Robinson's CQ" (see [5, 16]). That is, both AKKT or CAKKT may be able to detect nonoptimality, while Fritz-John's condition may fail. This is the main reason we are interested in global convergence results based on (C)AKKT.

We finish this section by noting that one can define a constraint qualification (AKKT-regularity), strictly weaker than Robinson's CQ, imposing that the constraint set is such that for every objective function such that AKKT holds at x^*, the KKT conditions also hold, which can also be defined in a geometric way in terms of the continuity of a point-to-set KKT cone. The same holds true for CAKKT and TAKKT. We do not formalize these conditions here since they are direct extensions of what was done in [12]. Also, based on a first version of this paper, new constant rank constraint qualifications which imply (C)AKKT-regularity have been defined in the context of NSOCP [8, 9]. See also [6, 7] for the definition of similar CQs in the context of NSDP which are also related with AKKT.

In the next section, we particularize our definitions to the context of NSOCP, where the KKT conditions have a particular structure, and we characterize AKKT as a perturbation of these conditions. We also show that, similarly to the case of NLP, CAKKT is strictly stronger than AKKT in the context of NSOCP. After a first version of this paper has appeared, this proof for NSDP has been done in [4]. The fact that AKKT does not imply CAKKT is known from [16].

4 Nonlinear Second-Order Cone Programming

Besides the nonnegative orthant and the positive semidefinite cone, an important example of symmetric cone is the second-order cone. In this section, we will study the nonlinear second-order cone programming (NSOCP) problem, which is a particular case of NSCP. Since the second-order cone has a very particular structure, we present the AKKT and CAKKT conditions in a more specific form.

The nonlinear second-order cone programming that we are interested is given below:

$$\begin{aligned} &\underset{x\in\mathbb{R}^n}{\text{Minimize}} && f(x), \\ &\text{subject to} && g(x) \in K, \end{aligned} \tag{NSOCP}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $g\colon \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable functions and $K := K_{m_1} \times \cdots \times K_{m_r}$ is the Cartesian product of second-order cones K_{m_i}, $i = 1, \ldots, r$, such that $m_1 + \cdots + m_r = m$. Let us define the feasible set of (NSOCP) by

$$F := \{x \in \mathbb{R}^n \mid g_i(x) \in K_{m_i} \ \text{ for all } \ i = 1, \ldots, r\}.$$

The topological interior and boundary of the cone K_{m_i} are characterized, respectively, by

$$\begin{aligned} \operatorname{int}(K_{m_i}) &:= \{(x_0, \overline{x}) \in \mathbb{R} \times \mathbb{R}^{m_i-1} \mid \|\overline{x}\| < x_0\} \ \text{ and} \\ \operatorname{bd}(K_{m_i}) &:= \{(x_0, \overline{x}) \in \mathbb{R} \times \mathbb{R}^{m_i-1} \mid \|\overline{x}\| = x_0\}. \end{aligned}$$

The boundary of K_{m_i} excluding the null vector will be denoted by $\operatorname{bd}^+(K_{m_i})$. For $x \in F$, we define the following index sets:

$$\begin{aligned} I_I(x) &:= \{i = 1, \ldots, r \mid g_i(x) \in \operatorname{int}(K_{m_i})\}, \\ I_B(x) &:= \{i = 1, \ldots, r \mid g_i(x) \in \operatorname{bd}^+(K_{m_i})\}, \\ I_0(x) &:= \{i = 1, \ldots, r \mid g_i(x) = 0\}. \end{aligned}$$

Now, consider the Lagrangian function $L\colon \mathbb{R}^n \times \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_r} \to \mathbb{R}$ given by

$$L(x, \mu_1, \ldots, \mu_r) := f(x) - \sum_{i=1}^{r} \langle g_i(x), \mu_i \rangle.$$

We say that $x \in \mathbb{R}^n$ is a KKT point for (NSOCP) if there exist multipliers $\mu_i \in \mathbb{R}^{m_i}$, $i = 1, \ldots, r$, such that

$$\begin{aligned} \nabla_x L(x, \mu_1, \ldots, \mu_r) &= 0, && \\ g_i(x) \circ \mu_i &= 0, && i = 1, \ldots, r, \\ g_i(x) &\in K_{m_i}, && i = 1, \ldots, r, \\ \mu_i &\in K_{m_i}, && i = 1, \ldots, r, \end{aligned} \tag{29}$$

where the Jordan product $\circ$ is defined in (2). As it can be seen in [1, Lemma 15], the complementarity condition (29) is equivalent to the following conditions:

$$\begin{aligned} i \in I_I(x) &\Rightarrow \mu_i = 0, \\ i \in I_B(x) &\Rightarrow \mu_i = 0 \text{ or } \mu_i = \left([\mu_i]_0, -\alpha_i(x)\overline{g_i(x)}\right) \text{ with } [\mu_i]_0 \neq 0, \end{aligned} \tag{30}$$

where $\alpha_i(x) = [\mu_i]_0/[g_i(x)]_0$ for $i = 1, \ldots, r$. The relations $i \in I_I(x)$ and $i \in I_B(x)$ can be written in terms of the eigenvalues of $g_i(x)$ as follows:

$$g_i(x) \in \text{int}(K_{m_i}) \Leftrightarrow \lambda_-(g_i(x)) > 0,\ \lambda_+(g_i(x)) > 0, \tag{31}$$

$$g_i(x) \in \text{bd}^+(K_{m_i}) \Leftrightarrow \lambda_-(g_i(x)) = 0,\ \lambda_+(g_i(x)) > 0. \tag{32}$$

In the case of (30), due to the expression of μ_i and the definition of $c_-(\cdot)$ and $c_+(\cdot)$ in (5), an important observation is that μ_i and $g_i(x)$ operator commute but $c_-(\mu_i) = c_+(g_i(x))$ and $c_+(\mu_i) = c_-(g_i(x))$.

4.1 Optimality Conditions

Let us now rewrite our optimality conditions in the case that the symmetric cone $\mathcal{K}$ is the second-order cone K.

Theorem 4.1 *A feasible point $x^* \in \mathbb{R}^n$ satisfies the AKKT condition for (NSOCP) if, and only if, there exist sequences $\{x^k\} \subset \mathbb{R}^n$ and $\{\mu_i^k\} \subset K_{m_i}$ for all i with $x^k \to x^*$ such that*

$$\lim_{k\to\infty} \nabla f(x^k) - \sum_{i=1}^{r} J g_i(x^k)^T \mu_i^k = 0, \tag{33}$$

$$i \in I_I(x^*) \Rightarrow \mu_i^k = 0 \text{ for sufficiently large } k, \tag{34}$$

$$i \in I_B(x^*) \Rightarrow \mu_i^k = 0\ \forall k \text{ or } \mu_i^k \in bd^+(K_{m_i}) \text{ with } -\frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|} \to \frac{\overline{g_i(x^*)}}{\|\overline{g_i(x^*)}\|}. \tag{35}$$

Proof By the definition of AKKT for (NSCP), we need only to verify the relation between the complementarity condition requested in AKKT for NSCP (8) and (9) and the complementarity conditions given by (34) and (35).

Assume that the feasible point $x^* \in \mathbb{R}^n$ satisfies the AKKT condition for NSCP. We have the following spectral decomposition for $g_i(x^*) \in K_{m_i}$:

$$g_i(x^*) = \lambda_1(g_i(x^*))c_1(g_i(x^*)) + \lambda_2(g_i(x^*))c_2(g_i(x^*)),$$

where $\lambda_1(g_i(x^*)) := \lambda_-(g_i(x^*))$ and $\lambda_2(g_i(x^*)) := \lambda_+(g_i(x^*))$. We have two cases to consider:

(i) If $i \in I_I(x^*)$, then, from (31), $\lambda_1(g_i(x^*)) > 0$ and $\lambda_2(g_i(x^*)) > 0$. By AKKT for NSCP, we have that $\lambda_1(\mu_i^k) = \lambda_2(\mu_i^k) = 0$ and therefore, $\mu_i^k = 0$.

(ii) If $i \in I_B(x^*)$, then, from (32), $\lambda_1(g_i(x^*)) = 0$ and $\lambda_2(g_i(x^*)) > 0$. By AKKT for NSCP, we have that $\lambda_2(\mu_i^k) = 0$. Note that we have two options for $\lambda_2(\mu_i^k)$:

(a$_1$) If $\lambda_2(\mu_i^k) = \lambda_+(\mu_i^k) = [\mu_i^k]_0 + \|\overline{\mu_i^k}\|$ for infinitely many k, then $\lambda_2(\mu_i^k) = 0$ implies $\mu_i^k = 0$ and we may relabel the sequence such that $\mu_i^k = 0$ for all k.

(a$_2$) Otherwise, $\lambda_2(\mu_i^k) = \lambda_-(\mu_i^k) = [\mu_i^k]_0 - \|\overline{\mu_i^k}\|$ for sufficiently large k, with $\lambda_1(\mu_i^k) = [\mu_i^k]_0 + \|\overline{\mu_i^k}\| > 0$. Hence, $\lambda_2(\mu_i^k) = 0$ implies $\mu_i^k \in \mathrm{bd}^+(K_{m_i})$. Since $c_2(\mu_i^k) = c_-(\mu_i^k)$ and $c_2(\mu_i^k) \to c_2(g_i(x^*))$, we get that

$$-\frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|} \to \frac{\overline{g_i(x^*)}}{\|\overline{g_i(x^*)}\|}.$$

Now, let us prove the converse. Note that if $\lambda_2(g_i(x^*)) = 0$, then $\lambda_1(g_i(x^*)) = 0$ and we do not need to check $\lambda_1(\mu_i^k)$ and $\lambda_2(\mu_i^k)$. Consider the following cases:

(i) Assume that $\lambda_1(g_i(x^*)) > 0$ and $\lambda_2(g_i(x^*)) > 0$. Thus, $i \in I_I(x^*)$ and then, by AKKT for NSOCP, $\mu_i^k = 0$. Hence, $\lambda_1(\mu_i^k) = \lambda_2(\mu_i^k) = 0$ and $c_1(\mu_i^k) \to c_1(g_i(x^*))$ and $c_2(\mu_i^k) \to c_2(g_i(x^*))$ are trivially satisfied.

(ii) Assume that $\lambda_1(g_i(x^*)) = 0$ and $\lambda_2(g_i(x^*)) > 0$. Thus, $i \in I_B(x^*)$. By AKKT for NSOCP, $\mu_i^k = 0$ for all k or $\mu_i^k \in \mathrm{bd}^+(K_{m_i})$ with $-\frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|} \to \frac{\overline{g_i(x^*)}}{\|\overline{g_i(x^*)}\|}$.

(b$_1$) If $\mu_i^k = 0$ for all k, then $\lambda_1(\mu_i^k) = \lambda_2(\mu_i^k) = 0$, $c_1(\mu_i^k) \to c_1(g_i(x^*))$ and $c_2(\mu_i^k) \to c_2(g_i(x^*))$ are trivially satisfied.

(b$_2$) If $\mu_i^k \in \mathrm{bd}^+(K_{m_i})$ with $-\frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|} \to \frac{\overline{g_i(x^*)}}{\|\overline{g_i(x^*)}\|}$, we have that $\lambda_-(\mu_i^k) = [\mu_i^k]_0 - \|\overline{\mu_i^k}\| = 0$ with $\lambda_+(\mu_i^k) = [\mu_i^k]_0 + \|\overline{\mu_i^k}\| > 0$.
This limit implies that $c_+(\mu_i^k) \to c_-(g_i(x^*))$ and $c_-(\mu_i^k) \to c_+(g_i(x^*))$. Since $\lambda_2(g_i(x^k)) = \lambda_+(g_i(x^k))$, we have that $\lambda_2(\mu_i^k) = \lambda_-(\mu_i^k)$ which was shown to be zero.

Note that the convergence (9) also holds for $i \in I_I(x^*)$ or $i \in I_0(x^*)$ due to the freedom allowed in choosing $c_\pm(z)$ when $\overline{z} = 0$. This completes the proof. □

To illustrate the optimality condition, we present two examples where we build an AKKT sequence around a minimizer that does not satisfy the KKT conditions.

Example 4.1 (AKKT sequence at a non-KKT solution—linear problem). Consider the following linear second-order cone problem from [2]:

$$\begin{aligned} &\underset{x\in\mathbb{R}^2}{\text{Minimize}} && -x_2, \\ &\text{subject to} && g(x) := (x_1, x_1, x_2) \in K_3, \end{aligned}$$

at an optimal solution $x^* = 0$.

(i) x^* is not a KKT point. Note that the only nonempty index set is $I_0(x^*)$. Let $\mu = (\mu_0, \mu_1, \mu_2) \in K_3$, and we have

$$\nabla f(x^*) - Jg(x^*)^T \mu = 0 \Leftrightarrow \begin{pmatrix} 0 \\ -1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \mu_0 - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \mu_1 - \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mu_2 = 0.$$

This implies that $\mu_2 = -1$ and $\mu_1 = -\mu_0$. The fact that $\mu \in K_3$ gives the contradiction $\mu_0 \geq \sqrt{1 + \mu_0^2}$. Therefore, x^* is not a KKT point.

(ii) x^* is an AKKT point. Define $\{x^k\}$ as any sequence such that $x^k \to x^*$ and $\mu^k := (\sqrt{k^2 + 1}, -k, -1) \in K_3$ for all k. Then, it is easy to see that $\nabla f(x^k) - Jg(x^k)^T \mu^k \to 0$.

Example 4.2 (AKKT sequence at a non-KKT solution). Consider the following NSOCP problem:

$$\begin{aligned} \underset{x \in \mathbb{R}}{\text{Minimize}} \quad & x, \\ \text{subject to} \quad & g_1(x) = (-x, 0) \in K_2, \quad g_2(x) = (0, x^2) \in K_2, \\ & g_3(x) = (1, x) \in K_2, \quad g_4(x) = (1 + x, 1 + x) \in K_2. \end{aligned}$$

The point $x^* = 0$ is the optimal solution of the above problem.

(i) x^* is not a KKT point. First, note that $I_I(x^*) = \{3\}$, $I_B(x^*) = \{4\}$ and $I_0(x^*) = \{1, 2\}$. Let $\mu_1 = ([\mu_1]_0, \overline{\mu_1})$, $\mu_2 = ([\mu_2]_0, \overline{\mu_2})$, $\mu_3 = ([\mu_3]_0, \overline{\mu_3})$ and $\mu_4 = ([\mu_4]_0, \overline{\mu_4})$. We have

$$\nabla f(x^*) - \sum_{i=1}^{4} Jg_i(x^*)^T \mu_i = 0 \Rightarrow 1 + [\mu_1]_0 - \overline{\mu_3} - ([\mu_4]_0 + \overline{\mu_4}) = 0. \quad (36)$$

Complementarity implies that $\mu_3 = 0$ and, from (30), $\mu_4 = 0$ or $\mu_4 = ([\mu_4]_0, -[\mu_4]_0)$. In any case, we have $[\mu_4]_0 + \overline{\mu_4} = 0$. Thus, by (36) we have that $[\mu_1]_0 = -1$, which makes $\mu_1 \notin K_2$. Therefore, x^* is not a KKT point.

(ii) x^* is an AKKT point. Define $x^k = \dfrac{1}{2k}$, $\mu_1^k = \left(\dfrac{1}{k}, \dfrac{1}{k}\right)$, $\mu_2^k = (k, k)$, $\mu_3^k = (0, 0)$ and $\mu_4^k = (1, -1)$. Then,

$$\nabla f(x^k) - \sum_{i=1}^{4} Jg_i(x^k)^T \mu_i^k = 1 + \frac{1}{k} - 2\frac{1}{2k}k = \frac{1}{k} \to 0.$$

Since $\mu_3^k = (0, 0)$ and $\mu_4^k = (1, -1) \in \text{bd}^+(K_2)$ is such that $-\dfrac{\overline{\mu_4^k}}{\|\overline{\mu_4^k}\|} = \dfrac{\overline{g_4(x^*)}}{\|\overline{g_4(x^*)}\|}$, complementarity is fulfilled and x^* is an AKKT point.

In the context of NLP, it is easy to note that CAKKT implies AKKT. Let us now show that this implication is true in the context of NSOCP.

Theorem 4.2 *Let x^* satisfy CAKKT for* (NSOCP). *Then, x^* satisfies AKKT for* (NSOCP).

Proof Let $\{x^k\} \subset \mathbb{R}^n$ and $\{\mu_i^k\} \subset K_{m_i}$, $i = 1, \ldots, r$, be given by Definition 3.4 such that (33) holds and $g_i(x^k) \circ \mu_i^k \to 0$, $i = 1, \ldots, r$. In order to prove that AKKT holds, it is enough to prove (34) and (35). By the definition of the Jordan product (2), for all $i = 1, \ldots, r$ we have that

$$g_i(x^k) \circ \mu_i^k := \left(\langle g_i(x^k), \mu_i^k\rangle, [g_i(x^k)]_0 \overline{\mu_i^k} + [\mu_i^k]_0 \overline{g_i(x^k)}\right) \to 0.$$

Hence, since $\langle g_i(x^k), \mu_i^k\rangle = [g_i(x^k)]_0[\mu_i^k]_0 + \overline{g_i(x^k)}^T \overline{\mu_i^k}$ and $[g_i(x^k)]_0 \to [g_i(x^*)]_0$, we have that

$$[g_i(x^k)]_0^2[\mu_i^k]_0 + [g_i(x^k)]_0 \overline{g_i(x^k)}^T \overline{\mu_i^k} \to 0.$$

Also, since $[g_i(x^k)]_0 \overline{\mu_i^k} + [\mu_i^k]_0 \overline{g_i(x^k)} \to 0$ and $\overline{g_i(x^k)} \to \overline{g_i(x^*)}$, we obtain

$$[g_i(x^k)]_0 \overline{g_i(x^k)}^T \overline{\mu_i^k} + [\mu_i^k]_0 \|\overline{g_i(x^k)}\|^2 \to 0.$$

Thus,

$$[\mu_i^k]_0 \left([g_i(x^k)]_0^2 - \|\overline{g_i(x^k)}\|^2\right) \to 0. \tag{37}$$

This implies that $\mu_i^k \to 0$ for all $i \in I_I(x^*)$ and hence (34) holds replacing μ_i^k by zero, while (33) still holds by continuity. To prove (35), let $i \in I_B(x^*)$. If there is a subsequence such that $\mu_i^k \to 0$, one can replace μ_i^k by zero and the result follows. Otherwise, $\{[\mu_i^k]_0\}$ is bounded away from zero. Since $\{[g_i(x^k)]_0\}$ is also bounded away from zero,

$$[g_i(x^k)]_0 \overline{\mu_i^k} + [\mu_i^k]_0 \overline{g_i(x^k)} \to 0 \quad \Rightarrow \quad \frac{\overline{\mu_i^k}}{[\mu_i^k]_0} + \frac{\overline{g_i(x^k)}}{[g_i(x^k)]_0} \to 0.$$

It remains to prove that μ_i^k can be chosen in the nonzero boundary of K_{m_i}. From (37), we conclude that

$$[\mu_i^k]_0 \left([g_i(x^k)]_0 - \|\overline{g_i(x^k)}\|\right) \to 0, \tag{38}$$

as $[g_i(x^k)]_0 + \|\overline{g_i(x^k)}\|$ converges to some positive number. Since

$$[g_i(x^k)]_0[\mu_i^k]_0 - \|\overline{g_i(x^k)}\| \|\overline{\mu_i^k}\| \le [g_i(x^k)]_0[\mu_i^k]_0 + \overline{g_i(x^k)}^T \overline{\mu_i^k} \to 0,$$

using (38) and dividing by $[g_i(x^k)]_0$, we conclude that $\limsup_{k\to+\infty}([\mu_i^k]_0 - \|\overline{\mu_i^k}\|) \leq 0$. Since $\mu_i^k \in K_{m_i}$, thus, $[\mu_i^k]_0 - \|\overline{\mu_i^k}\| \geq 0$, we have that $\lambda_-(\mu_i^k) = [\mu_i^k]_0 - \|\overline{\mu_i^k}\| \to 0$. Hence, by replacing $\lambda_-(\mu_i^k)$ by zero in the spectral decomposition of μ_i^k, the limit (33) is still valid and AKKT holds. □

Note that similarly to the case of NLP, in the definition of AKKT for NSCP, the complementarity (8) can be replaced by the looser one $\lambda_i(\sigma^k) \to 0$ whenever $\lambda_i(G(x^*)) > 0$, for all $i = 1, \ldots, r$. In this case, in the equivalence of Theorem 4.1, the condition that $\mu_i^k \in \mathrm{bd}^+(K_{m_i})$ would be replaced by its approximate version $[\mu_i^k]_0 - \|\overline{\mu_i^k}\| \to 0$. These conditions are in fact equivalent as when $\lambda_i(\sigma^k) \to 0$, one may replace these eigenvalues by zero in the spectral decomposition of σ^k while keeping the other limits. This definition is more natural in the context of CAKKT, since the dual CAKKT sequence is such that the eigenvalues only converge to zero and are not necessarily equal to zero. Also, this allows simpler proofs that algorithms generate AKKT sequences without replacing the multipliers. However, the definition in the current form is inspired by the augmented Lagrangian method, which generates dual sequences with this particular format. We keep the definition in the current form to be consistent with previous definitions in other contexts.

5 An Algorithm that Satisfies the New Optimality Conditions

The interest in NSOCPs comes from the variety of problems that can be formulated with them (see, e.g., [1, 23, 42]), in particular, in the linear case. Several algorithms for the nonlinear case have been proposed in the literature [3, 30, 37, 40, 41, 48]. In this section, we will analyze an augmented Lagrangian algorithm, proposed in [41], which extends the usual augmented Lagrangian algorithm for NLP to NSOCP, where we will show better global convergence results.

5.1 Augmented Lagrangian Method

In this subsection, we will prove that the augmented Lagrangian method for NSOCP, proposed by Liu and Zhang [41], generates sequences that satisfy AKKT for NSOCP without any constraint qualification, whereas in [41] the authors rely on the nondegeneracy condition in order to prove that the KKT conditions hold. The augmented Lagrangian function used is the Powell–Hestenes–Rockafellar (PHR) function defined as follows:

$$L_\rho(x, \mu_1, \ldots, \mu_r) := f(x) + \frac{1}{2\rho}\sum_{i=1}^{r}\left(\|[\mu_i - \rho g_i(x)]_+\|^2 - \|\mu_i\|^2\right),$$

where $\rho > 0$ is the penalty parameter. The gradient with respect to x of the above augmented Lagrangian function is given by

$$\nabla_x L_\rho(x, \mu_1, \ldots, \mu_r) = \nabla f(x) - \sum_{i=1}^{r} J g_i(x)^T [\mu_i - \rho g_i(x)]_+.$$

The formal statement of the algorithm is as follows:

Algorithm 5.1 Augmented Lagrangian Algorithm for NSOCP

Let $\tau \in (0, 1)$, $\gamma > 1$, $\rho_1 > 0$ and $\mu^0 \in K$. Take a sequence of tolerances $\{\epsilon_k\} \subset \mathbb{R}_+$ such that $\epsilon_k \to 0$. Define $\hat{\mu}^1 \in K$. Choose an arbitrary starting point $x^0 \in \mathbb{R}^n$. Initialize $k := 1$ and $\|v^0\| := +\infty$.

Step 1: Find an approximate minimizer x^k of $L_{\rho_k}(x, \hat{\mu}_1^k, \ldots, \hat{\mu}_r^k)$. That is, find x^k satisfying

$$\|\nabla_x L_{\rho_k}(x^k, \hat{\mu}_1^k, \ldots, \hat{\mu}_r^k)\| \le \epsilon_k.$$

Step 2: Define $v^k = (v_1^k, \ldots, v_r^k)$ with

$$v_i^k := \left[\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right]_+ - \frac{\hat{\mu}_i^k}{\rho_k}, \text{ for all } i.$$

If $\|v^k\| \le \tau \|v^{k-1}\|$, define $\rho_{k+1} := \rho_k$, otherwise define $\rho_{k+1} := \gamma \rho_k$.
Step 3: Compute

$$\mu_i^k := [\hat{\mu}_i^k - \rho_k g_i(x^k)]_+, \text{ for all } i$$

and define $\hat{\mu}^{k+1} \in K$ such that $\{\hat{\mu}^k\}$ is bounded.
Step 4: Set $k := k + 1$, and go to Step 1.

Now, we proceed to prove our results. We start by showing that the algorithm tends to find feasible points, in the following sense:

Theorem 5.1 *Let $x^* \in \mathbb{R}^n$ be a limit point of a sequence $\{x^k\}$ generated by Algorithm 5.1. Then, x^* is a stationary point for the following problem*

$$\textit{Minimize} \quad P(x) := \|[-g_i(x)]_+\|^2. \tag{39}$$

Proof Let us consider the following two cases:

(i) Assume that $\{\rho_k\}$ is bounded. Then, there exists k_0 such that for $k \ge k_0$ we have $\rho_k = \rho_{k_0}$. Thus, from Step 3 of Algorithm 5.1, for all i, $v_i^k \to 0$. In a subsequence where $\{\hat{\mu}_i^k\}$ converges to some $\hat{\mu}_i \in K_{m_i}$, we have

$$\lim_{k\to\infty} \mu_i^k = \lim_{k\to\infty} \left[\hat{\mu}_i^k - \rho_{k_0} g_i(x^k)\right]_+ = \left[\hat{\mu}_i - \rho_{k_0} g_i(x^*)\right]_+ = \lim_{k\to\infty} \hat{\mu}_i^k = \hat{\mu}_i.$$

Now, if $\hat{\mu}_i - \rho_{k_0} g_i(x^*) = \lambda_1 c_1 + \lambda_2 c_2$ is a spectral decomposition of $\hat{\mu}_i - \rho_{k_0} g_i(x^*)$, we can write:

$$\hat{\mu}_i = [\hat{\mu}_i - \rho_{k_0} g_i(x^*)]_+ = \max\{0, \lambda_1\}c_1 + \max\{0, \lambda_2\}c_2.$$

This gives the following spectral decomposition

$$g_i(x^*) = (1/\rho_{k_0}) \left((\max\{0, \lambda_1\} - \lambda_1)c_1 + (\max\{0, \lambda_2\} - \lambda_2)c_2\right),$$

where the eigenvalues are $\max\{0, \lambda_1\} - \lambda_1 \geq 0$ and $\max\{0, \lambda_2\} - \lambda_2 \geq 0$. Hence, $g_i(x^*) \in K_{m_i}$, which means that $P(x^*) = 0$. Therefore, x^* is a global minimizer of (39).

(ii) Assume that $\{\rho_k\}$ is unbounded. Let us define $\delta^k := \nabla f(x^k) - \sum_{i=1}^r Jg_i(x^k)^T \mu_i^k$ where $\mu_i^k := [\hat{\mu}_i^k - \rho_k g_i(x^k)]_+$. By Step 1 of Algorithm 5.1, we have that $\|\delta^k\| \leq \epsilon_k$. Thus,

$$\frac{\delta^k}{\rho_k} = \frac{\nabla f(x^k)}{\rho_k} - \sum_{i=1}^r Jg_i(x^k)^T \left[\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right]_+ \to 0.$$

Since $\hat{\mu}_i^k$ is bounded and all involved functions are continuous, we have that $\nabla P(x^*) = 2\sum_{i=1}^r Jg_i(x^*)^T \left[-g_i(x^*)\right]_+ = 0$. Therefore, x^* is a stationary point of (39).

□

Theorem 5.2 *Assume that $x^* \in \mathbb{R}^n$ is a feasible limit point of a sequence $\{x^k\}$ generated by Algorithm 5.1. Then, x^* is an AKKT point.*

Proof Assume without loss of generality that $x^k \to x^*$. From Step 1 of Algorithm 5.1, we have that

$$\left\| \nabla f(x^k) - \sum_{i=1}^r Jg_i(x^k)^T [\hat{\mu}_i^k - \rho g_i(x^k)]_+ \right\| \leq \epsilon_k \Rightarrow$$

$$\lim_{k\to\infty} \nabla f(x^k) - \sum_{i=1}^r Jg_i(x^k)^T \mu_i^k = 0,$$

where $\mu_i^k = [\hat{\mu}_i^k - \rho g_i(x^k)]_+$. Now, we will prove that (34) and (35) hold. Similarly to Theorem 5.1, we have two cases to analyze: The sequence $\{\rho_k\}$ is bounded or unbounded.

(i) If the sequence $\{\rho_k\}$ is bounded, then for some k_0 and all $k \geq k_0$, we have $\rho_k = \rho_{k_0}$. Taking the spectral decomposition for μ_i^k as follows

$$\mu_i^k = [\hat{\mu}_i^k - \rho_{k_0} g_i(x^k)]_+ = \max\{0, \lambda_1^k\}c_1^k + \max\{0, \lambda_2^k\}c_2^k,$$

where $c_i^k \to c_i, \lambda_i^k \to \lambda_i, i = 1, 2$, we can choose them such that, similarly to item (i) of Theorem 5.1,

$$g_i(x^*) = \frac{1}{\rho_{k_0}} \left((\max\{0, \lambda_1\} - \lambda_1)c_1 + (\max\{0, \lambda_2\} - \lambda_2)c_2\right). \tag{40}$$

which gives the eigenvalues

$$\lambda_1(g_i(x^*)) = \frac{\max\{0, \lambda_1\} - \lambda_1}{\rho_{k_0}} > 0 \quad \text{and}$$
$$\lambda_2(g_i(x^*)) = \frac{\max\{0, \lambda_2\} - \lambda_2}{\rho_{k_0}} > 0.$$

Now, let us consider that $i \in I_I(x^*)$. Thus, $\max\{0, \lambda_1\} > \lambda_1$ and $\max\{0, \lambda_2\} > \lambda_2$, which imply $\lambda_1 < 0$, $\lambda_2 < 0$ and $\lambda_1^k < 0$, $\lambda_2^k < 0$ for all k large enough. Moreover, $\lambda_1(\mu_i^k) = \max\{0, \lambda_1^k\} = 0$ and $\lambda_2(\mu_i^k) = \max\{0, \lambda_2^k\} = 0$ for all k large enough. Thus, $\mu_i^k = 0$ for all sufficiently large k. Let us consider that $i \in I_B(x^*)$. Considering (32) and (40), we can assume without loss of generality that the eigenvalues of $g_i(x^*)$ are such that

$$\lambda_1(g_i(x^*)) = \lambda_-(g_i(x^*)) = 0 \quad \text{and} \quad \lambda_2(g_i(x^*)) = \lambda_+(g_i(x^*)) > 0.$$

With an argument similar to the previous case, $\lambda_2(g_i(x^*)) = \dfrac{\max\{0, \lambda_2\} - \lambda_2}{\rho_{k_0}} > 0$ implies $\lambda_2(\mu_i^k) = \max\{0, \lambda_2^k\} = 0$ for all k large enough. Thus, we have two possibilities. If for an infinite subset of indexes we have

$$\lambda_2(\mu_i^k) = \lambda_+(\mu_i^k) := [\mu_i^k]_0 + \|\overline{\mu_i^k}\| = 0 \quad \text{and}$$
$$\lambda_1(\mu_i^k) = \lambda_-(\mu_i^k) := [\mu_i^k]_0 - \|\overline{\mu_i^k}\| = 0,$$

then we can define a subsequence such that $\mu_i^k = 0$ for all k. Otherwise, for k large enough, we have

$$\lambda_2(\mu_i^k) = \lambda_-(\mu_i^k) := [\mu_i^k]_0 - \|\overline{\mu_i^k}\| = 0 \quad \text{and}$$
$$\lambda_1(\mu_i^k) = \lambda_+(\mu_i^k) := [\mu_i^k]_0 + \|\overline{\mu_i^k}\| > 0, \tag{41}$$

which implies that $\mu_i^k \in \mathrm{bd}^+(K_{m_i})$. In addition, we have

$$c_1 = c_-(g_i(x^*)) := \left(\frac{1}{2}, -\frac{\overline{g_i(x^*)}}{2\|\overline{g_i(x^*)}\|}\right) \quad \text{and}$$
$$c_2 = c_+(g_i(x^*)) := \left(\frac{1}{2}, \frac{\overline{g_i(x^*)}}{2\|\overline{g_i(x^*)}\|}\right).$$

Moreover, we get by (41) that our decomposition of μ_i^k is such that

$$c_1^k = c_+(\mu_i^k) := \left(\frac{1}{2}, \frac{\overline{\mu_i^k}}{2\|\overline{\mu_i^k}\|}\right) \quad \text{and} \quad c_2^k = c_-(\mu_i^k) := \left(\frac{1}{2}, -\frac{\overline{\mu_i^k}}{2\|\overline{\mu_i^k}\|}\right).$$

Thus, since $c_i^k \to c_i$ for $i = 1, 2$, we have $-\frac{\overline{\mu_i^k}}{\|\overline{\mu_i^k}\|} \to \frac{\overline{g_i(x^*)}}{\|\overline{g_i(x^*)}\|}$.

(ii) Let us consider that $\{\rho_k\}$ is unbounded. Since the sequence $\{\hat{\mu}_i^k\}$ is bounded, we have that $\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k) \to -g_i(x^*)$. Let us take spectral decompositions

$$\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k) = \lambda_1^k c_1^k + \lambda_2^k c_2^k \tag{42}$$

and

$$\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k) \to -g_i(x^*) = \lambda_1 c_1 + \lambda_2 c_2, \tag{43}$$

with $c_1^k \to c_1$, $c_2^k \to c_2$, $\lambda_1^k \to \lambda_1$ and $\lambda_2^k \to \lambda_2$. In addition, we can get a spectral decomposition for μ_i^k using (42) given by

$$\mu_i^k = [\hat{\mu}_i^k - \rho_k g_i(x^k)]_+ = \rho_k \max\{0, \lambda_1^k\} c_1^k + \rho_k \max\{0, \lambda_2^k\} c_2^k. \tag{44}$$

Then, if $i \in I_I(x^*)$ we have that $\lambda_1 < 0$ and $\lambda_2 < 0$ since these are eigenvalues of $-g_i(x^*)$. Thus, $\lambda_1^k < 0$ and $\lambda_2^k < 0$ for all sufficiently large k, which implies that $\lambda_1(\mu_i^k) = \rho_k \max\{0, \lambda_1^k\} = 0$ and $\lambda_2(\mu_i^k) = \rho_k \max\{0, \lambda_2^k\} = 0$; hence, $\mu_i^k = 0$ for all k large enough.
Now, if $i \in I_B(x^*)$, we can choose without loss of generality $-\lambda_1 = \lambda_1(g_i(x^*)) = \lambda_-(g_i(x^*)) = 0$ and $-\lambda_2 = \lambda_2(g_i(x^*)) = \lambda_+(g_i(x^*)) > 0$. Thus, $\lambda_2^k < 0$ for all sufficiently large k, implying that $\lambda_2(\mu_i^k) = \rho_k \max\{0, \lambda_2^k\} = 0$. The proof now follows analogously to the case where $\{\rho_k\}$ is bounded and $i \in I_B(x^*)$.

□

Now, we proceed to prove that the algorithm actually generates CAKKT sequences. For this, the following assumption on the smoothness of the function $g(\cdot)$ is needed. See [16].

Assumption 1 Let x^* be a feasible limit point of a sequence $\{x^k\}$ generated by Algorithm 5.1. The following generalized Lojasiewicz inequality holds: There is $\delta > 0$ and a function $\psi : B(x^*, \delta) \to \mathbb{R}$, where $B(x^*, \delta)$ is the closed ball centered at x^* with radius $\delta > 0$, such that $\psi(x) \to 0$ when $x \to x^*$ and, for all $x \in B(x^*, \delta)$,

$$P(x) \leq \psi(x)\|\nabla P(x)\|,$$

where $P(x) := \|[-g(x)]_+\|^2$.

The following lemma is an extension to second-order cones of Weyl's Lemma [36] for symmetric matrices.

Lemma 5.1 *Let $z, y \in K_{m_i}$. Then, the following inequalities hold:*

$$\lambda_-(z) + \lambda_-(y) \leq \lambda_-(z+y) \leq \lambda_-(z) + \lambda_+(y) \leq \lambda_+(z+y) \leq \lambda_+(z) + \lambda_+(y).$$

Proof It is a trivial computation from the expression of the eigenvalues. □

Theorem 5.3 *Let us suppose that $x^* \in \mathbb{R}^n$ is a feasible limit point of a sequence $\{x^k\}$ generated by Algorithm 5.1 such that Assumption 1 holds. Then, x^* is a CAKKT point.*

Proof Similarly to Theorem 5.2, let us assume that a feasible $x^* \in \mathbb{R}^n$ is a limit point of $\{x^k\}$ generated by Algorithm 5.1. Taking a subsequence if necessary, we can assume that $x^k \to x^*$. From Step 1, we have that

$$\lim_{k\to\infty} \nabla f(x^k) - \sum_{i=1}^{r} J g_i(x^k)^T \mu_i^k = 0,$$

where $\mu_i^k = [\hat{\mu}_i^k - \rho_k g_i(x^k)]_+$. Now, it remains to prove that $g_i(x^k) \circ \mu_i^k \to 0$ for all $i = 1, \ldots, r$. In the sequel, we will analyze the case where $\{\rho_k\}$ is bounded.

Let us take a decomposition for μ_i^k as in the case (i) of Theorem 5.1. Thus, for $k \geq k_0$ we have $\rho_k = \rho_{k_0}$,

$$\hat{\mu}_i = [\hat{\mu}_i - \rho_{k_0} g_i(x^*)]_+ = \max\{0, \lambda_1\}c_1 + \max\{0, \lambda_2\}c_2$$

and

$$\begin{aligned} &\hat{\mu}_i - \rho_{k_0} g_i(x^*) = \lambda_1 c_1 + \lambda_2 c_2 \Rightarrow \\ &g_i(x^*) = (1/\rho_{k_0})\left((\max\{0, \lambda_1\} - \lambda_1)c_1 + (\max\{0, \lambda_2\} - \lambda_2)c_2\right). \end{aligned}$$

Hence, using the fact that $c_1 \circ c_2 = 0$, $c_1^2 = c_1$, and $c_2^2 = c_2$, we obtain

$$\begin{aligned} g_i(x^k) \circ \mu_i^k \to g_i(x^*) \circ \hat{\mu}_i = &\frac{1}{\rho_{k_0}} (\max\{0, \lambda_1\}(\max\{0, \lambda_1\} - \lambda_1)c_1 + \\ &\max\{0, \lambda_2\}(\max\{0, \lambda_2\} - \lambda_2)c_2) = 0. \end{aligned}$$

Now, let us consider the case where the sequence $\{\rho_k\}$ is unbounded. If $i \in I_I(x^*)$, we have by Theorem 5.2 that $\mu_i^k = 0$ for all k, and then, $g_i(x^k) \circ \mu_i^k = 0$ for all k. Otherwise, since $\{\nabla_x L_{\rho_k}(x^k, \hat{\mu}_1^k, \ldots, \hat{\mu}_r^k)\}$ is bounded by Step 1 of the algorithm and $\dfrac{\hat{\mu}_i^k}{\rho_k} \to 0$, there is $M > 0$ such that $\rho_k \|\nabla P(x^k)\| \leq M$. Moreover, we have by Assumption 1 that

$$\rho_k P(x^k) \leq \psi(x^k) \|\rho_k \nabla P(x^k)\| \leq \psi(x^k) M.$$

Taking the limit in k, we get

$$\rho_k \left([\lambda_1(-g_i(x^k))]_+^2 + [\lambda_2(-g_i(x^k))]_+^2 \right) = \rho_k \|[-g_i(x^k)]_+\|^2 \to 0. \tag{45}$$

Now, let us consider decompositions similar to (42), (43) and (44) in the proof of Theorem 5.2. That is, $g_i(x^k) = -(\lambda_1^k c_1^k + \lambda_2^k c_2^k) + \dfrac{\hat{\mu}_i^k}{\rho_k}$, $-g_i(x^*) = \lambda_1 c_1 + \lambda_2 c_2$ and

$$\mu_i^k = [\hat{\mu}_i^k - \rho_k g_i(x^k)]_+ = \rho_k [\lambda_1^k]_+ c_1^k + \rho_k [\lambda_2^k]_+ c_2^k,$$

where $c_1^k \to c_1$, $c_2^k \to c_2$, $\lambda_1^k \to \lambda_1$ and $\lambda_2^k \to \lambda_2$. Here, we use the notation $[t]_+ := \max\{0, t\}$ for a real number t. Thus,

$$g_i(x^k) \circ \mu_i^k = -\rho_k(\lambda_1^k [\lambda_1^k]_+ c_1^k + \lambda_2^k [\lambda_2^k]_+ c_2^k) + \frac{\hat{\mu}_i^k}{\rho_k} \circ \mu_i^k.$$

Note that

$$\frac{\hat{\mu}_i^k}{\rho_k} \circ \mu_i^k = \hat{\mu}_i^k \circ \frac{\mu_i^k}{\rho_k} = \hat{\mu}_i^k \circ \left[\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k) \right]_+.$$

Since $\{\hat{\mu}_i^k\}$ is bounded and $\dfrac{\mu_i^k}{\rho_k} \to 0$, we obtain that $\dfrac{\hat{\mu}_i^k}{\rho_k} \circ \mu_i^k \to 0$. It remains to prove that $\rho_k(\lambda_1^k [\lambda_1^k]_+ c_1^k + \lambda_2^k [\lambda_2^k]_+ c_2^k) \to 0$ and for this it is enough to prove that

$$\lambda_\pm \left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k) \right) \left[\lambda_\pm(\hat{\mu}_i^k - \rho_k g_i(x^k)) \right]_+ \to 0.$$

From Lemma 5.1, we have

$$\lambda_-(\hat{\mu}_i^k) + \rho_k \lambda_-(-g_i(x^k)) \le \lambda_-(\hat{\mu}_i^k - \rho_k g_i(x^k)) \le \lambda_+(\hat{\mu}_i^k) + \rho_k \lambda_-(-g_i(x^k)). \tag{46}$$

Now, observe that $\lambda_-(-g_i(x^*)) \le 0$, because we are considering cases $i \in I_B(x^*)$ or $i \in I_0(x^*)$, and $\lambda_-(-g_i(x^*)) = -\lambda_+(g_i(x^*))$. Note that if $\lambda_-(-g_i(x^*)) < 0$, since $\lambda_-(-g_i(x^k)) \to \lambda_-(-g_i(x^*))$, the sequence $\{\hat{\mu}_i^k\}$ is bounded and $\rho_k \to +\infty$, we have that $\lambda_-(\hat{\mu}_i^k - \rho_k g_i(x^k)) < 0$ for k large enough. Hence,

$$\lambda_- \left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k) \right) \left[\lambda_-(\hat{\mu}_i^k - \rho_k g_i(x^k)) \right]_+ \to 0. \tag{47}$$

If $\lambda_-(-g_i(x^*)) = 0$ but $\lambda_-(-g_i(x^k)) \leq 0$ for an infinite subset of indices, we have for this subsequence, using (46), that

$$\begin{aligned}
0 \leq [\lambda_-(\hat{\mu}_i^k - \rho_k g_i(x^k))]_+ &= \max\{0, \lambda_-(\hat{\mu}_i^k - \rho_k g_i(x^k))\} \\
&\leq \max\{0, \lambda_+(\hat{\mu}_i^k) + \rho_k \lambda_-(-g_i(x^k)))\} \\
&\leq \max\{0, \lambda_+(\hat{\mu}_i^k)\} + \max\{0, \rho_k \lambda_-(-g_i(x^k)))\}, \\
&= \max\{0, \lambda_+(\hat{\mu}_i^k)\}.
\end{aligned}$$

Then, $\{[\lambda_-(\hat{\mu}_i^k - \rho_k g_i(x^k))]_+\}$ is bounded. Since $\lambda_-\left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right) \to 0$, we have that (47) also holds. If $\lambda_-(-g_i(x^*)) = 0$ but $\lambda_-(-g_i(x^k)) > 0$ for all k large enough, we can multiply (46) by $\lambda_-\left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right) > 0$ to arrive at

$$\begin{aligned}
&\lambda_-(\hat{\mu}_i^k)\lambda_-\left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right) + \rho_k \lambda_-(-g_i(x^k))\lambda_-\left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right) \\
&\quad \leq \lambda_-(\hat{\mu}_i^k - \rho_k g_i(x^k))\lambda_-\left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right)
\end{aligned}$$

and

$$\begin{aligned}
&\lambda_-(\hat{\mu}_i^k - \rho_k g_i(x^k))\lambda_-\left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right) \\
&\quad \leq \lambda_+(\hat{\mu}_i^k)\lambda_-\left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right) + \rho_k \lambda_-(-g_i(x^k))\lambda_-\left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right).
\end{aligned}$$

Note that $\lambda_-(\hat{\mu}_i^k)\lambda_-\left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right) \to 0$ and $\lambda_+(\hat{\mu}_i^k)\lambda_-\left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right) \to 0$, once again because $\{\hat{\mu}_i^k\}$ is bounded and $\lambda_-(-g_i(x^k)) \to \lambda_-(-g_i(x^*)) = 0$. Then, we need to show that $\rho_k \lambda_-(-g_i(x^k))\lambda_-\left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right) \to 0$. To see this, it is enough to multiply (46) by $\lambda_-(-g_i(x^k))$. In fact, by doing this, we get

$$0 < \rho_k \lambda_-(-g_i(x^k))\lambda_-\left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right) \leq \lambda_+(\hat{\mu}_i^k)\lambda_-(-g_i(x^x)) + \rho_k \lambda_-(-g_i(x^k))^2.$$

Using again the fact that $\{\hat{\mu}_i^k\}$ is bounded and $\lambda_-(-g_i(x^k)) \to 0$, as well as $\rho_k \lambda_-(-g_i(x^k))^2 \to 0$ (by (45)), we conclude that the above left-hand side term

Table 1 Value of the correspondent measure on the final three iterations of Algorithm 5.1

Problem	Complementarity in CAKKT	Complementarity in AKKT	Gradient of the Lagrangian
P1	5.5e−03 1.1e−04 1.5e−06	1.0e−03 2.1e−05 2.8e−07	4.7e−03 2.1e−04 1.3e−06
P2(5,10)	4.0e−06 2.8e−06 9.4e−07	2.5e−05 1.8e−05 2.6e−06	4.0e−05 3.5e−05 9.6e−06
P2(10,10)	1.3e−06 1.7e−07 6.9e−08	8.0e−05 9.2e−06 3.7e−06	2.6e−05 1.7e−05 5.8e−06
P2(25,50)	1.6e−05 5.6e−06 2.7e−07	9.4e−06 3.0e−06 1.5e−07	7.8e−05 2.5e−05 1.6e−06
P2(50,50)	4.7e−05 1.0e−05 8.3e−07	5.8e−05 1.2e−05 1.0e−06	7.4e−04 7.9e−05 9.1e−06
P2(50,100)	6.1e−06 2.5e−06 5.7e−07	6.1e−06 3.4e−06 7.1e−07	8.9e−05 8.3e−05 8.7e−06
P2(100,100)	4.1e−07 1.3e−07 4.3e−08	3.7e−07 3.4e−07 1.8e−08	1.5e−05 1.1e−05 7.4e−07
P3(10,10)	4.8e−05 1.4e−05 4.1e−06	4.0e−06 1.1e−06 2.9e−07	4.0e−05 1.1e−05 3.1e−06
P3(50,50)	1.5e−04 8.8e−05 9.4e−06	1.7e−07 7.3e−08 7.5e−09	5.8e−05 2.0e−05 3.3e−06
P3(100,100)	1.8e−04 6.9e−05 3.7e−05	4.0e−08 1.2e−08 3.2e−09	4.2e−05 1.7e−05 2.6e−06

converges to zero. Therefore, (47) also holds in this case. The proof that

$$\lambda_+\left(\frac{\hat{\mu}_i^k}{\rho_k} - g_i(x^k)\right)\left[\lambda_+(\hat{\mu}_i^k - \rho_k g_i(x^k))\right]_+ \to 0$$

is done analogously. We conclude that $g_i(x^k) \circ \mu_i^k \to 0$ for all $i = 1, \ldots, r$. □

To finish the paper and illustrate our results, we check whether Algorithm 5.1 in fact generate AKKT or CAKKT sequences in a practical implementation. To see this, we implemented the algorithm in MATLAB R2023a, under Ubuntu 20.04. We chose three different problems from the literature:

- P1: Example 3.6 of [37], a nonconvex NSOCP with $n = m = 3$.
- P2(n, m): Experiment 1 of [40], a convex NSOCP with arbitrary n and m.
- P3(n, m): Experiment 2 of [40], a nonconvex NSOCP with arbitrary $n = m$.

For simplicity, we consider a unique second-order cone constraint for the random problems P2 and P3. The subproblems were solved with `fminunc` function, with a trust region algorithm. We also took the subproblem's tolerance as $\epsilon_0 = 10^{-5}$, and $\epsilon_{k+1} = \epsilon_k/k$, the tolerance for stopping the algorithm as 10^{-5}, and all other parameters were chosen similarly to ALGENCAN (see [19]).

The results of the experiments are shown in Table 1. In particular, we analyze the results for problem P1, 6 instances for problem P2 with $n = m$ and $m = 2n$ and 3 instances for problem P3. Table 1 shows the norm of the gradient of the Lagrangian that appears in both AKKT and CAKKT conditions, as well as the norm of the corresponding complementarity condition, of the last 3 iterations.

First we observe that whenever the algorithm stops with optimality, it can solve all the considered problems in few iterations. The nonconvex problem P3 with the largest dimension ($n = m = 100$), for instance, took 32 outer iterations and 193 inner iterations. Moreover, in all cases, we clearly observe decrease of the complementarity measures and the gradient of the Lagrangian. Therefore, we can conclude that Algorithm 5.1 in fact can generate AKKT and CAKKT sequences numerically.

6 Conclusions

In this paper, we extended the optimality conditions AKKT and CAKKT to the context of symmetric cone programming. These optimality conditions were shown to be strictly stronger than Fritz-John's condition. When specializing to second-order cones, an explicit characterization of AKKT is presented, which takes into account the particular form of the KKT conditions in this context, not relying on a spectral decomposition. The definition of CAKKT is very natural in the context of symmetric cones, as it does not rely on a spectral decomposition and uses the Jordan product to measure complementarity. In the context of second-order cones, we have shown that CAKKT is stronger than AKKT, while we also show that a previous attempt of avoiding eigenvalues by using the inner product turns out to give a condition weaker than CAKKT and independent of AKKT.

For second-order cones, we showed that an augmented Lagrangian algorithm generates AKKT sequences, while under an additional smoothness assumption it generates CAKKT sequences. This gives a global convergence result not relying on Robinson's constraint qualification. There are several ways in which we expect to continue our research on this topic. For instance, one shall prove that this algorithm generates AKKT and CAKKT sequences for general symmetric cones; also, it would be interesting to understand the relationship of CAKKT and AKKT for general symmetric cones. Other classes of algorithms, such as interior point methods, can probably also generate CAKKT sequences. These possibilities will be subject of further research.

Acknowledgements This work was supported by the Grant-in-Aid for Scientific Research (C) (19K11840) from Japan Society for the Promotion of Science, FAPESP grants 2013/07375-0, 2017/18308-2 and 2018/24293-0, FAPES grant 116/2019, CNPq grants 309136/2021-0 and 306988/2021-6, CAPES, USP-Santander and PRONEX - CNPq/FAPERJ (grant E-26/010.001247/2016). We also would like to take the opportunity to thank an anonymous referee for insightful comments and suggestions.

References

1. Alizadeh, F., Goldfarb, D.: Second-order cone programming. Math. Program. **95**(1), 3–51 (2003)
2. Andersen, E.D., Roos, C., Terlaky, T.: Notes on duality in second order and p-order cone optimization. Optimization **51**(4), 627–643 (2002)
3. Andersen, E.D., Roos, C., Terlaky, T.: On implementation a primal-dual interior-point method for conic quadratic programming. Math. Program. **95**, 249–277 (2003)
4. Andreani, R., Fukuda, E.H., Haeser, G., Santos, D.O., Secchin, L.D.: On the use of Jordan algebras for improving global convergence of an Augmented Lagrangian method in nonlinear semidefinite programming. Comput. Optim. Appl. **79**, 633–648 (2021)
5. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**(5), 627–641 (2011)
6. Andreani, R., Haeser, G., Mito, L.M., RamírezC H.: Sequential constant rank constraint qualifications for nonlinear semidefinite programming with applications. To appear in Set-Valued and Variational Analysis, (2022)
7. Andreani, R., Haeser, G., Mito, L.M., RamírezC H.: Weak notions of nondegeneracy in nonlinear semidefinite programming. arXiv:2012.14810 (2022)
8. Andreani, R., Haeser, G., Mito, L.M., Ramírez C, H., Santos, D.O., Silveira, T.P.: Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming. Optim. Lett. **16**, 589–610 (2022)
9. Andreani, R., Haeser, G., Mito, L.M., Ramírez C, H., Silveira, T.P.: Global convergence of algorithms under constant rank conditions for nonlinear second-order cone programming. J. Optim. Theory Appl. **195**, 42–78 (2022)
10. Andreani, R., Haeser, G., Ramos, A., Silva, P.J.S.: A second-order sequential optimality condition associated to the convergence of algorithms. IMA J. Numer. Anal. **37**(4), 1902–1929 (2017)
11. Andreani, R., Haeser, G., Secchin, L.D., Silva, P.J.S.: New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences. SIAM J. Optim. **29**(4), 3201–3230 (2019)
12. Andreani, R., Haeser, G., Viana, D.S.: Optimality conditions and global convergence for nonlinear semidefinite programming. Math. Program. **180**, 203–235 (2020)
13. Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: A cone-continuity constraint qualification and algorithmic consequences. SIAM J. Optim. **26**(1), 96–110 (2016)
14. Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: Strict constraint qualifications and sequential optimality conditions for constrained optimization. Math. Oper. Res. **43**(3), 693–717 (2018)
15. Andreani, R., Martínez, J.M., Santos, L.T.: Newton's method may fail to recognize proximity to optimal points in constrained optimization. Math. Program. **160**, 547–555 (2016)

16. Andreani, R., Martínez, J.M., Svaiter, B.F.: A new sequencial optimality condition for constrained optimization and algorithmic consequences. SIAM J. Optim. **20**(6), 3533–3554 (2010)
17. Baes, M.: Spectral Functions and Smoothing Techniques on Jordan Algebras. PhD thesis, Université Catholique de Louvain, France (2006)
18. Baes, M.: Convexity and differentiability properties of spectral functions and spectral mappings on Euclidean Jordan algebras. Linear Algebra Appl. **422**(2), 664–700 (2007)
19. Birgin, E., Martínez, J.M.: Practical Augmented Lagrangian Methods for Constrained Optimization. SIAM Publications, New Delhi (2014)
20. Birgin, E.G., Haeser, G., Ramos, A.: Augmented Lagrangians with constrained subproblems and convergence to second-order stationary points. Comput. Optim. Appl. **69**(1), 51–75 (2018)
21. Birgin, E.G., Krejić, N., Martínez, J.M.: On the minimization of possibly discontinuous functions by means of pointwise approximations. Optim. Lett. **11**(8), 1623–1637 (2017)
22. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York (2000)
23. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
24. Bueno, L.F., Haeser, G., Lara, F., Rojas, F.N.: An Augmented Lagrangian method for quasi-equilibrium problems. Comput. Optim. Appl. **76**(3), 737–766 (2020)
25. Bueno, L.F., Haeser, G., Rojas, F.N.: Optimality conditions and constraint qualifications for generalized Nash equilibrium problems and their practical implications. SIAM J. Optim. **29**(1), 31–54 (2019)
26. Dutta, J., Deb, K., Tulshyan, R., Arora, R.: Approximate KKT points and a proximity measure for termination. J. Global Optim. **56**(4), 1463–1499 (2013)
27. Faraut, J., Korányi, A.: Analysis on Symmetric Cones. Oxford Mathematical Monographs. Clarendon Press, Oxford (1994)
28. Feng, M., Li, S.: An approximate strong KKT condition for multiobjective optimization. Top **26**(3), 489–509 (2018)
29. Fiacco, A.V., McCormick, G.P.: Nonlinear Programming Sequential Unconstrained Minimization Techniques. Wiley, New York (1968)
30. Fukuda, E.H., Silva, P.J.S., Fukushima, M.: Differentiable exact penalty functions for nonlinear second-order cone programs. SIAM J. Optim. **22**(4), 1607–1633 (2012)
31. Fukushima, M., Luo, Z.-Q., Tseng, P.: Smoothing functions for second-order-cone complementarity problems. SIAM J. Optim. **12**(2), 436–460 (2001)
32. Giorgi, G., Jiménez, B., Novo, V.: Approximate Karush–Kuhn–Tucker condition in multiobjective optimization. J. Optim. Theory Appl. **171**(1), 70–89 (2016)
33. Haeser, G.: A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms. Comput. Optim. Appl. **70**(2), 615–639 (2018)
34. Haeser, G., Melo, V.V.: Convergence detection for optimization algorithms: approximate-KKT stopping criterion when Lagrange multipliers are not available. Oper. Res. Lett. **43**(5), 484–488 (2015)
35. Haeser, G., Schuverdt, M.L.: On approximate KKT condition and its extension to continuous variational inequalities. J. Optim. Theory Appl. **149**(3), 528–539 (2011)
36. Horn, R.A., Johnson, C.R.: Topics in Matrix Analysis. Cambridge University Press, Cambridge (1991)
37. Kanzow, C., Ferenczi, I., Fukushima, M.: On the local convergence of semismooth Newton methods for linear and nonlinear second-order cone programs without strict complementarity. SIAM J. Optim. **20**(41), 297–320 (2009)
38. Kanzow, C., Steck, D., Wachsmuth, D.: An augmented Lagrangian methods for optimization problems in Banach spaces. SIAM J. Control Optim. **56**(1), 272–291 (2018)
39. Kanzow, D., Steck, D.: Approximated Lagrangian and exact penalty methods for quasi-variational inequalities. Comput. Optim. Appl. **60**(3), 801–824 (2018)
40. Kato, H., Fukushima, M.: An SQP-type algorithm for nonlinear second-order cone programs. Optim. Lett. **1**(2), 129–144 (2007)
41. Liu, Y.J., Zhang, L.W.: Convergence of the augmented Lagrangian method for nonlinear optimization problems over second-order cone. J. Optim. Theory Appl. **139**(3), 557–575 (2008)
42. Lobo, M.S., Vandenberghe, L., Boyd, S., Lebret, H.: Applications of second-order cone programming. Linear Algebra Appl. **284**(1–3), 193–228 (1998)
43. Lourenço, B.F., Fukuda, E.H., Fukushima, M.: Optimality conditions for problems over symmetric cones and a simple augmented Lagrangian method. Math. Oper. Res. **43**, 1233–1251 (2018)
44. Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optimization **60**(4), 429–440 (2011)

45. Qi, L., Wei, Z.: On the constant positive linear dependence conditions and its application to SQP methods. SIAM J. Optim. **10**(4), 963–981 (2000)
46. Ramos, A.: Mathematical programs with equilibrium constraints: a sequential optimality condition, new constraint qualifications and algorithmic consequences. Optim. Methods Softw. **36**, 45–81 (2019)
47. Tuyen, N.V., Yao, J., Wen, C.: A note on approximate Karush–Kuhn–Tucker conditions in locally Lipschitz multiobjective optimization. Optim. Lett. **13**, 163–174 (2019)
48. Yamashita, H., Yabe, H.: A primal-dual interior point method for nonlinear optimization over second-order cones. Optim. Methods Softw. **24**(3), 407–426 (2009)