On the Weak Second-order Optimality Condition for Nonlinear Semidefinite and Second-order Cone Programming

Ellen H. Fukuda[1] · Gabriel Haeser[2] · Leonardo M. Mito[2]

Abstract

Second-order necessary optimality conditions for nonlinear conic programming problems that depend on a single Lagrange multiplier are usually built under nondegeneracy and strict complementarity. In this paper we establish a condition of such type for two classes of nonlinear conic problems, namely semidefinite and second-order cone programming, assuming Robinson's constraint qualification and a weak constant rank-type property which are, together, strictly weaker than nondegeneracy. Our approach is done via a penalty-based strategy, which is aimed at providing strong global convergence results for first- and second-order algorithms. Since we are not assuming strict complementarity, the critical cone does not reduce to a subspace, thus, the second-order condition we arrive at is defined in terms of the lineality space of the critical cone. In the case of nonlinear programming, this condition reduces to the standard second-order condition widely used as second-order stationarity measure in the algorithmic practice.

Keywords Optimality conditions · Second-order conditions · Weak constant rank · Semidefinite programming · Second-order cone programming

Mathematics Subject Classification (2010) 90C46 · 90C30 · 90C26 · 90C22

Ellen H. Fukuda, Gabriel Haeser, and Leonardo M. Mito contributed equally to this work.

✉ Leonardo M. Mito
  leokoto@ime.usp.br

  Ellen H. Fukuda
  ellen@i.kyoto-u.ac.jp

  Gabriel Haeser
  ghaeser@ime.usp.br

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

[2] Department of Applied Mathematics, University of São Paulo, São Paulo, Brazil

🍂 Springer

1 Introduction

Consider the following *nonlinear conic programming* (NCP) problem in standard form:

$$\begin{aligned}
&\underset{x \in \mathbb{R}^n}{\text{Minimize}} && f(x), \\
&\text{subject to} && g(x) \in \mathcal{K}, \\
&&& h(x) = 0,
\end{aligned} \tag{NCP}$$

where $f \colon \mathbb{R}^n \to \mathbb{R}$, $g \colon \mathbb{R}^n \to \mathbb{E}$ and $h \colon \mathbb{R}^n \to \mathbb{R}^p$ are twice continuously differentiable functions, $\mathbb{E}$ is a finite-dimensional linear space equipped with an inner product $\langle \cdot, \cdot \rangle$ and the norm $|| \cdot ||$ induced by it, and $\mathcal{K} \subseteq \mathbb{E}$ is a closed convex cone that is assumed to be *self-dual*, which means $\mathcal{K} = \mathcal{K}^* \doteq \{w \in \mathbb{E} \colon \forall y \in \mathcal{K}, \langle w, y \rangle \geqslant 0\}$.

We are primarily interested in second-order necessary optimality conditions for two well-established particular cases of (NCP):

- *Nonlinear second-order cone programming* (NSOCP), which is obtained when $\mathbb{E} = \mathbb{R}^m$ and $\mathcal{K}$ is the so-called (Lorentz) *second-order cone*, defined as $\mathbb{L}^m \doteq \{(w_0, \bar{w}) \in \mathbb{R} \times \mathbb{R}^{m-1} \colon w_0 \geqslant ||\bar{w}||_2\}$ when $m > 1$ and $\mathbb{L}^1 \doteq \{w \in \mathbb{R} \colon w \geqslant 0\}$, or the Cartesian product of r second-order cones in $\mathbb{R}^{m_i}$, with $i \in \{1, \ldots, r\}$ and $m_1 + \cdots + m_r = m$;
- *Nonlinear semidefinite programming* (NSDP), which is obtained when $\mathbb{E} = \mathbb{S}^m$ is the space of all $m \times m$ real symmetric matrices and $\mathcal{K}$ is the cone $\mathbb{S}^m_+ \doteq \{W \in \mathbb{S}^m \colon \forall d \in \mathbb{R}^m, d^\top W d \geqslant 0\}$ of all positive semidefinite matrices, or a Cartesian product in the form $\mathcal{K} = \mathbb{S}^{m_i}_+ \times \cdots \times \mathbb{S}^{m_r}_+$, with $m_1 + \cdots + m_r = m$.

Both fields have grown independently and accumulated a large set of applications over the years, for example, in robust control [33, 34], passive reduced-order modelling [37], structural optimization [50, 55], the sphere covering problem [24], and others (see [19, 56, 70] for a vast collection of examples). In conjunction, several algorithms have been developed for them, such as interior-point methods [23, 48, 71], sequential quadratic programming methods [52, 53], Newton-type methods [39, 51], and augmented Lagrangian methods [4, 15, 73], to name a few (see Yamashita and Yabe [72] for more details). Consequently, some theoretical aspects of NSOCP and NSDP, such as optimality conditions and regularity, have gained much relevance in the community as well. In particular, necessary optimality conditions are especially useful for giving theoretical global convergence support for iterative algorithms, in the sense that every feasible limit point of a given algorithm can be proven to satisfy some necessary optimality condition under a set of hypotheses. In fact, the reliability of an algorithm is deeply related with the strength of the optimality condition that supports its global convergence theory. From this point of view, second-order necessary optimality conditions improve the first-order ones by considering the curvature of the problem data over the set of directions where first-order information has little meaning, which is usually called *cone of critical directions* (or *critical cone*). Note that this kind of convergence theory is different from what is usually done for convex optimization problems, where *second-order sufficient conditions* are used as convergence hypotheses. In the nonconvex case, the latter results in a local convergence analysis. Since the results of this paper are meant to be used in the aid of global convergence, we focus on necessary optimality conditions.

It is worth mentioning that second-order analysis in non-polyhedral conic contexts, such as NSOCP and NSDP, is considerably more intricate than in polyhedral contexts, such as in *nonlinear programming* (NLP). This is justified by the fact that the curvature of $\mathcal{K}$ must be taken into account, besides the curvature of the functions defining the problem. The initial efforts to characterize this curvature were done by Kawasaki [54], whose results were

generalized and refined by Cominetti [31], and later completed by Bonnans, Cominetti, and Shapiro [27] with the notion of *second-order regularity*. Then, Shapiro [67] obtained a specialized statement for it in the context of NSDP, that was later re-discovered by Forsgren [36], Jarre [49], and Lourenço, Fukuda, and Fukushima [57], using distinct nontrivial techniques that make each proof interesting on its own. For NSOCP, second-order necessary optimality conditions were first characterized by Bonnans and Ramírez [28], and later studied by Fukuda and Fukushima [38] who also presented sufficient conditions. In recent years, significant advances were obtained were obtained for very general classes of problems that have NSOCP and NSDP among their particular cases; see, for instance, the papers of Chieu et al. [30] and Mohammadi, Mordukhovich, and Sarabi [58], both working under very weak assumptions, but while the former uses the classical notion of *cone reducibility* introduced by [27], the latter employs a new and more general concept called *parabolic regularity* which allows obtaining second-order conditions by directly differentiating the indicator function in a particular sense instead of reducing the problem to remove its curvature at the point of interest. Thus, it is possible to say that the motivation for studying alternative ways of deriving second-order conditions for conic problems, in particular NSDP and NSOCP, has gone far beyond practical usage, but nevertheless we believe practice should not be ignored.

With this in mind, some useful tools for proving new first- and second-order optimality conditions for optimization problems, which are deeply connected to the algorithmic approach, are the so-called *sequential optimality conditions*. They were introduced in NLP, and later extended to NSOCP and NSDP, as KKT variants designed for building convergence theory of iterative algorithms (for details, we refer to the work of Andreani et al. [4, 6, 7, 15, 17]) and they gained some attention for being able to sharpen most convergence results for them in a general and unified manner (see, for instance, [14, Section 5.2]). Also, a second-order sequential condition has recently appeared in the work of Andreani et al. [13] for NLP, which not only provided an ideal way of incorporating second-order information in numerical methods, but also an intuitive strategy for building second-order analysis under weaker hypotheses than the traditional *linear independence constraint qualification* (LICQ). These improvements were obtained by considering a somewhat "weak" second-order necessary optimality condition in the sense that only the lineality space of the critical cone is taken into account in their results. However, as it is well-known in NLP, this "weak" condition is the most suitable second-order condition for global convergence analysis of algorithms, since the stronger conditions that deal with the whole critical cone are not guaranteed to be fulfilled at the convergence points of a large class of algorithms, such as barrier-type methods [41] and augmented Lagrangian-type methods [18], even under very strong hypotheses. Besides, checking the validity of the "strong" second-order condition is an NP-hard class problem, whereas checking the "weak" condition is of polynomial class. Nevertheless, as far as we know, the latter condition has never received due attention in nonconvex conic contexts other than NLP.

Inspired by [13], we prove that every local minimizer satisfies the weaker version of the second-order necessary condition, for NSOCP and NSDP, but under weaker assumptions than all previous related works. In fact, the meaning of our results lies in the fact we assume neither *nondegeneracy* nor *strict complementarity*, since under these hypotheses the "weak" and "strong" second-order conditions are equivalent. Our approach is based on sequential conditions, which suggests that our results may be useful for proving convergence of algorithms to second-order stationary points of NSOCP and NSDP problems. We stress that even though NSOCP can be represented in terms of NSDP, it is interesting to discriminate them since the numerical methods designed to solve each problem might have different performances

in practice [1]. Also, it is not straightforward to derive second-order results for NSOCP only based on the NSDP results.

This paper is structured as follows: we begin by reviewing some classical results on first- and second-order optimality conditions for (NCP) and its particular cases with some degree of details, in Section 2. Then, we present our second-order analysis for NSOCP in Section 3, and for NSDP in Section 4. At last, in Section 5 we give some final considerations about this paper and related works.

2 Technical Background

In this section, we introduce our notation and present some results from the literature that are directly related to ours. We also review in details some classical results on first- and second-order optimality conditions for NSOCP and NSDP.

We consider the standard inner product in $\mathbb{R}^n$, given by $\langle a, b \rangle \doteq \sum_{i=1}^n a_i b_i$, and the Euclidean norm, given by $||a||_2 \doteq \sqrt{\langle a, a \rangle}$, for every $a, b \in \mathbb{R}^n$. The terms $\mathrm{int}(\mathcal{K})$, $\mathrm{bd}(\mathcal{K})$, and $\mathrm{bd}^+(\mathcal{K})$ stand for the *interior, boundary*, and *boundary excluding the origin* of $\mathcal{K}$, respectively. Also, for any closed convex cone C, $\mathrm{lin}(C) \doteq C \cap (-C)$ denotes its *lineality space*, which is the largest subspace contained in C.

For a given finite indexed set $\{a_i : i \in \{1, \ldots, k\}\} \subset \mathbb{R}$, we denote the array that has a_i in its i-th position by $[a_i]_{i \in \{1, \ldots, k\}} \in \mathbb{R}^k$ and, analogously, the matrix whose entries are the elements of $\{b_{ij} : i, j \in \{1, \ldots, k\}\} \subset \mathbb{R}$ is denoted by $[b_{ij}]_{i,j \in \{1, \ldots, k\}} \in \mathbb{R}^{k \times k}$. The identity matrix of $\mathbb{R}^{n \times n}$ is denoted by $\mathbb{I}_n$. The *gradient* and the *Hessian* of a function $f \colon \mathbb{R}^n \to \mathbb{R}$ at an arbitrary point $x \in \mathbb{R}^n$ are represented by $\nabla f(x)$ and $\nabla^2 f(x)$, respectively, and the first derivative of $g \colon \mathbb{R}^n \to \mathbb{E}$ at x is the linear mapping $Dg(x)[\cdot] \colon \mathbb{R}^n \to \mathbb{E}$ defined by the action

$$Dg(x)[h] \doteq \sum_{i=1}^n \partial_i g(x) h_i$$

for every $h \in \mathbb{R}^n$, where $\partial_i g(x) \in \mathbb{E}$ is the partial derivative of g in the i-th variable, at x. In particular, if $\mathbb{E} = \mathbb{R}^m$ then $Dg(x)$ is exactly the Jacobian matrix of g at x, in the canonical basis of $\mathbb{R}^m$; for instance, in this case the i-th row of $Dg(x)$ is given by the *transpose* of $\nabla g_i(x)$, which is denoted by $\nabla g_i(x)^\top$, where $i \in \{1, \ldots, m\}$. The *adjoint* of $Dg(x)$ is the linear mapping $Dg(x)^*[\cdot] \colon \mathbb{E} \to \mathbb{R}^n$ such that $\langle Dg(x)[h], w \rangle = \langle h, Dg(x)^*[w] \rangle$ holds for every $h \in \mathbb{R}^n$ and every $w \in \mathbb{E}$, hence

$$Dg(x)^*[w] = [\langle \partial_i g(x), w \rangle]_{i \in \{1, \ldots, n\}},$$

for every $w \in \mathbb{E}$ and, if $\mathbb{E} = \mathbb{R}^m$ then $Dg(x)^* = Dg(x)^\top$. Similarly, we define the action of the linear mapping $D^2 g(x)^*[\cdot] \colon \mathbb{E} \to \mathbb{R}^{n \times n}$ by

$$D^2 g(x)^*[w] \doteq [\langle \partial_i \partial_j g(x), w \rangle]_{i,j \in \{1, \ldots, n\}},$$

for every $w \in \mathbb{E}$.

The *orthogonal projection* of $w \in \mathbb{E}$ onto $\mathcal{K}$ is the point $\Pi_{\mathcal{K}}(w) \in \mathcal{K}$ such that

$$||w - \Pi_{\mathcal{K}}(w)|| = \min\{||w - z|| \colon z \in \mathcal{K}\}.$$

Note that $\Pi_{\mathcal{K}}(w)$ is well-defined as a convex function of w since $\mathcal{K}$ is closed and convex. Also, a very useful fact is that every $w \in \mathbb{E}$ can be written as

$$w = \Pi_{\mathcal{K}}(w) - \Pi_{\mathcal{K}}(-w),$$

with $\langle \Pi_{\mathcal{K}}(w), \Pi_{\mathcal{K}}(-w) \rangle = 0$. This is commonly called the *Moreau's decomposition* of w and one of its many consequences is that $w \in \mathcal{K}$ if, and only if, $\Pi_{\mathcal{K}}(-w) = 0$. Hence, the function

$$\Phi(x) \doteq \frac{1}{2} \left(||h(x)||_2^2 + ||\Pi_{\mathcal{K}}(-g(x))||^2 \right)$$

can be used as a measure of violation of the constraints of (NCP), that is, a measure of infeasibility. A result by Fitzpatrick and Phelps [35, Thm. 2.2] can be employed to derive an expression for the gradient of Φ at x:

Theorem 1 *For every $x \in \mathbb{R}^n$, we have*

$$\nabla \Phi(x) = Dh(x)^\top h(x) - Dg(x)^*[\Pi_{\mathcal{K}}(-g(x))].$$

Also, we observe that $\nabla \Phi$ is a Lipschitz function, but it is not differentiable everywhere. In our analyses, we make use of its second derivative, which must be taken in the nonsmooth sense.

2.1 Some Elements of Nonsmooth Analysis

Let X and Y be finite-dimensional normed linear spaces over $\mathbb{R}$. Let $F \colon X \to Y$ be a locally Lipschitz function and denote the set in which it is differentiable by $\mathcal{D}(F)$. The so-called *B-subdifferential* of F at a point $x \in X$, is the set of all limiting derivatives of F at x, denoted by

$$\partial_B F(x) \doteq \left\{ V \in \mathcal{L}(X, Y) \colon \exists \{x^k\}_{k \in \mathbb{N}} \subset \mathcal{D}(F), \; x^k \to x, \; DF(x^k) \to V \right\},$$

where $\mathcal{L}(X, Y)$ denotes the set of all linear mappings from X to Y, and similarly to the previous section, $DF(x^k) \in \mathcal{L}(X, Y)$ denotes the first derivative of F at x^k, for every $x^k \in \mathcal{D}(F)$. Evidently, for every x, the set $\partial_B F(x)$ is compact, and it is a singleton when $x \in \mathcal{D}(F)$, but it is not convex in general. Then, we also define the *Clarke subdifferential* of F at x, denoted by $\partial F(x)$, as the *convex hull* of $\partial_B F(x)$, that is,

$$\partial F(x) \doteq \text{Conv}(\partial_B F(x)).$$

In particular, when $X = \mathbb{R}^n$ and $Y = \mathbb{R}$, the *generalized Hessian* of F at x is defined as

$$\partial^2 F(x) \doteq \partial \nabla F(x),$$

which is the convex hull of the set of all limiting Hessian matrices of F at x. Following Hiriart-Urruty et al. [46], the second-order necessary optimality condition for unconstrained minimizers of F when it is differentiable, is the following:

Theorem 2 *If $x^\star$ is a local minimizer of a differentiable function $F \colon \mathbb{R}^n \to \mathbb{R}$ such that ∇F is locally Lipschitz, then $\nabla F(x^\star) = 0$, and for each $d \in \mathbb{R}^n$, there exists some $M \in \partial^2 F(x^\star)$ such that $d^\top M d \geqslant 0$. In other words,*

$$\limsup_{d \in \mathbb{R}^n} d^\top \partial^2 F(x^\star) d \geqslant 0.$$

We refer to [46, Thm. 3.1] for a proof. As observed by Hiriart-Urruty et al., it is not true that $d^\top M d \geqslant 0$ for all $M \in \partial^2 F(x^\star)$, in general, and not even this holds for some fixed M and all d. We employ this result to analyse the second derivative of Φ, which is a *nonsmooth-smooth* composition, so a chain rule is also required. There are several different extensions of the chain rule for subdifferentials, but the following result, by Páles and Zeidan [61], is enough for our purposes:

Theorem 3 *Fix some $x \in X$ and let $G \colon X \to Y$ and $F \colon Y \to Y$ be functions such that G is continuously differentiable at x, and F is Lipschitz in a neighborhood of $G(x)$. Then, we have*

$$\partial(F \circ G)(x) \subseteq \partial F(G(x)) \circ DG(x),$$

where $\partial F(G(x)) \circ DG(x) \doteq \{V \circ DG(x) \colon V \in \partial F(G(x))\}$.

Proof The result follows from [61, Thm. 5.1] since it was originally proved for Banach spaces that satisfy the Radon-Nikodým property, which holds for every reflexive space, and every finite-dimensional space with a norm is reflexive.

Specificities about the subdifferential of the orthogonal projection onto the second-order and the semidefinite cone will be given in their respective sections.

2.2 Necessary Optimality Conditions and Constraint Qualifications

A *constraint qualification* (CQ) is any assumption over the constraints at a feasible point x, that implies that the feasible set is similar to its first-order approximation around x. For instance, one of the most relevant ones is *Robinson's CQ* [65], that holds at a feasible point x when $Dh(x)$ has full row rank and there exists some $d \in \mathbb{R}^n$ such that[1]

$$g(x) + Dg(x)[d] \in \text{int}(\mathcal{K})$$
$$\text{and}$$
$$Dh(x)d = 0.$$

It is widely known that Robinson's CQ is a generalization of the classical *Mangasarian-Fromovitz constraint qualification* (MFCQ) from NLP. Such regularity condition allows one to study the optimality of a point in terms of the first-order approximation of the problem around it, that is, it is possible to prove that for every local solution $x^\star$ of (NCP) that satisfies Robinson's CQ, there exists some $\omega^\star \in \mathcal{K}$ and some $\mu^\star \in \mathbb{R}^p$ such that

$$\nabla_x L(x^\star, \omega^\star, \mu^\star) = 0, \tag{1}$$

and

$$\langle g(x^\star), \omega^\star \rangle = 0, \tag{2}$$

where

$$L(x, \omega, \mu) \doteq f(x) - \langle g(x), \omega \rangle + \langle h(x), \mu \rangle$$

is the *Lagrangian* function of (NCP) and

$$\nabla_x L(x, \omega, \mu) \doteq \nabla f(x) - Dg(x)^*[\omega] + Dh(x)^\top \mu$$

is the gradient of $L(x, \omega, \mu)$ with respect to x. Equations (1) and (2) compose the so-called *Karush-Kuhn-Tucker* (KKT) conditions and, in this context, $\omega^\star$ and $\mu^\star$ are *Lagrange multipliers* associated with $x^\star$. Points that satisfy the KKT conditions are often called *first-order stationary* or *KKT points*. Condition (2) is often called *complementarity*, and when additionally $-\omega^\star$ belongs to the relative interior of $T_{\mathcal{K}}(g(x^\star))^\circ$, where

$$T_{\mathcal{K}}(g(x^\star)) \doteq \left\{ d \in \mathbb{E} \colon \begin{array}{l} \exists \{d^k\}_{k \in \mathbb{N}} \to d, \ \exists \{\alpha^k\}_{k \in \mathbb{N}} \to 0, \\ \forall k \in \mathbb{N}, \ \alpha^k > 0, \ g(x^\star) + \alpha^k d^k \in \mathcal{K} \end{array} \right\}$$

---

[1] We use this characterization of Robinson's CQ as a definition because $\mathcal{K}$ is assumed to be self-dual, and consequently, to have nonempty interior.

is the (*Bouligand*) *tangent cone* to $\mathcal{K}$ at $g(x^\star)$, we say that *strict complementarity* holds at the pair $(x^\star, \omega^\star)$ [29, Def. 4.74]. A relevant implication of Robinson's CQ is the boundedness of the set $\mathcal{M}(x^\star)$ of all Lagrange multipliers associated with a local solution $x^\star$.

Second-order optimality conditions give extra information over the set of directions where first-order information is not meaningful. That is, we are interested in the set

$$C(x^\star) \doteq \{d \in \mathbb{R}^n \colon \langle \nabla f(x^\star), d \rangle = 0, \ Dh(x^\star)^\top d = 0, \ Dg(x^\star)[d] \in T_\mathcal{K}(g(x^\star))\},$$

which is the critical cone of (NCP) at $x^\star$. If Robinson's CQ holds at a local minimizer $x^\star$ of (NCP), then besides KKT, it also satisfies the *basic second-order necessary condition* (BSOC), that is, for every $d \in C(x^\star)$ there are Lagrange multipliers $\omega_d^\star \in \mathcal{K}$ and $\mu_d^\star \in \mathbb{R}^p$ such that (1), (2), and

$$d^\top (\nabla_x^2 L(x^\star, \omega_d^\star, \mu_d^\star) + \sigma(x^\star, \omega_d^\star))d \geqslant 0 \tag{3}$$

hold, where

$$\nabla_x^2 L(x, \omega, \mu) \doteq \nabla^2 f(x) - D^2 g(x)^*[\omega] + \sum_{i=1}^p \mu_i \nabla^2 h_i(x)$$

and $\sigma(x, \omega)$ is the so-called "*sigma-term*", as presented by Cominetti [31, Thm. 4.1]. In that paper, the author builds second-order conditions for (NCP) based on the *second-order tangent set* of $\mathcal{K}$ at $g(x)$ along $Dg(x)[d]$, that may be denoted by $T_\mathcal{K}^2(g(x), Dg(x)[d])$, and then establishes a "dual form" for it using the support function of $T_\mathcal{K}^2(g(x), Dg(x)[d])$, which is precisely the sigma-term. Hence, the sigma-term represents a possible curvature of $\mathcal{K}$ at $g(x)$, to some extent, and it can be proved that $\sigma(x, \omega) = 0$ when $\mathcal{K}$ is *polyhedral*, such as in NLP (for details, see [31]). In fact, the difficulty of second-order analysis in contexts more general than NLP lies almost entirely on the characterization of the sigma-term, which can be a very challenging task.

One of the major practical drawbacks of BSOC is that in order to verify whether it holds or not at a given point x, one must know the whole set $\mathcal{M}(x)$, which is not always possible. The stronger optimality condition where inequality (3) holds for every $d \in C(x^\star)$, for some pair of multipliers $(\omega^\star, \mu^\star)$ (not depending on d), which is sometimes called the *semi-strong necessary optimality condition*, does not present such a drawback. However, deciding the positivity of a matrix over a cone is an NP-hard class problem [60], and so is checking the semi-strong condition.

A more practical alternative to BSOC and the semi-strong condition is the so-called *weak second-order necessary condition* (WSOC), which is defined as follows:

Definition 1 Let $x^\star$ be a KKT point associated with some Lagrange multipliers $\omega^\star \in \mathcal{K}$ and $\mu^\star \in \mathbb{R}^p$. We say that WSOC holds at $x^\star$ when

$$d^\top (\nabla_x^2 L(x^\star, \omega^\star, \mu^\star) + \sigma(x^\star, \omega^\star))d \geqslant 0, \tag{4}$$

for every $d \in S(x^\star) \doteq \mathrm{lin}(C(x^\star))$, which is the largest subspace contained in $C(x^\star)$.

Note that in Definition 1 we only take directions in the subspace $S(x^\star)$, called the *critical subspace* of (NCP) at $x^\star$, which coincides with $C(x^\star)$ under strict complementarity[2]. At first

---

[2] We will give a short proof for the fact $C(x^\star) = \mathrm{lin}(C(x^\star))$ under strict complementarity, for completeness: Let $d \in C(x^\star)$ and suppose that there exists a Lagrange multiplier $-\omega^\star$ in the relative interior of $T_\mathcal{K}(g(x^\star))^\circ$. Note that $\langle \nabla f(x^\star), d \rangle = \langle \omega^\star, Dg(x^\star)[d] \rangle = 0$ by the KKT conditions. Hence, $-\omega^\star \in T_\mathcal{K}(g(x^\star))^\circ \cap \{Dg(x^\star)[d]\}^\perp$, which implies $T_\mathcal{K}(g(x^\star))^\circ \subseteq \{Dg(x^\star)[d]\}^\perp$ and, consequently, $\mathrm{span}(Dg(x^\star)[d]) \subseteq T_\mathcal{K}(g(x^\star))$. Then, $Dg(x^\star)[d] \in \mathrm{lin}(T_\mathcal{K}(g(x^\star)))$, so $d \in \mathrm{lin}(C(x^\star))$.

sight, a second-order condition that only covers $S(x^\star)$ instead of the whole $C(x^\star)$ may seem disadvantageous in comparison with the semi-strong condition. In fact, the semi-strong condition implies WSOC. However, there are strong evidences that suggest that it is unlikely that BSOC or the semi-strong condition can be used to support the global convergence theory of any practical algorithm, unless $C(x^\star) = S(x^\star)$. In fact, for the particular case of NLP, Gould and Toint [41] presented a simple counterexample, with a quadratic objective function and a constraint of the form $x \geqslant 0$, for which a large class of barrier-type methods may produce an output sequence whose limit points fail to satisfy both BSOC and the semi-strong condition, even when every iterate of such sequence satisfies the second-order sufficient condition for its respective penalized problem. Later, Andreani and Secchin [18] made a small modification in Gould and Toint's counterexample to obtain the same conclusion for augmented Lagrangian-type algorithms. WSOC, on the other hand, is guaranteed to be fulfilled under weak assumptions for some variants of the two methods we mentioned above [2, 59], and also for a regularized SQP method for NLP [40]. The negative conclusions regarding BSOC and the semi-strong condition have led some authors to doubt the existence of an algorithm that could be associated with a second-order condition that takes the whole critical cone into consideration. Following this discussion, Andreani et al. [13] managed to characterize the weakest second-order constraint qualification that could guarantee the fulfilment of the semi-strong condition at the limit points of a large class of penalization-type algorithms that encompasses, for instance, all the aforementioned ones. However, such a constraint qualification was proven not to imply nor to be implied by LICQ [13, Ex. 4.5 and 4.6], and to be violated even for box constraints.

Despite the good algorithmic advantages of WSOC, Robinson's CQ alone is not enough to guarantee its fulfilment at local minimizers – see, for instance, the counterexample by Baccari [21, Section 3] or the discussion in [22]. Instead, the existing results on WSOC usually require a stronger CQ called *nondegeneracy* (or *transversality*), which holds at a feasible point x when

$$\mathbb{E} \times \mathbb{R}^p = (\text{lin}(T_\mathcal{K}(g(x))) + \text{Im}(Dg(x))) \times \text{Im}(Dh(x)). \tag{5}$$

It was translated from differential equations to optimization by Shapiro and Fan [68] and it is well-known that, for every nondegenerate solution $x^\star$ of (NCP), the set $\mathcal{M}(x^\star)$ is a singleton, what resembles the effects of LICQ in NLP. Thus, nondegeneracy is analogous to LICQ, in this sense.

Theorem 4 *If $x^\star$ is a local minimizer of (NCP) that satisfies nondegeneracy, then the KKT conditions hold at $x^\star$ for some Lagrange multipliers $\omega^\star \in \mathcal{K}$ and $\mu^\star \in \mathbb{R}^p$ and, moreover, WSOC holds with respect to these multipliers.*

Note that Theorem 4 is simply a rephrasing of the necessity of BSOC after assuming uniqueness of the Lagrange multiplier (nondegeneracy), but we stated it as it is for comparison purposes since our main results consist of proving of Theorem 4 under less demanding conditions.

In the context of NLP, Andreani, Martínez, and Schuverdt [16] were able to prove Theorem 4 replacing nondegeneracy (LICQ) with only MFCQ together with the so-called *weak constant rank* (WCR) property, which holds at a feasible point $x^\star$ when there exists a neighborhood $\mathcal{N}$ of $x^\star$ such that

$$\{\nabla g_i(x)\}_{i:\ g_i(x^\star)=0} \cup \{\nabla h_j(x)\}_{j \in \{1,\ldots,p\}} \tag{6}$$

has the same rank for every $x \in \mathcal{N}$. It is worth mentioning that WCR is not a CQ on its own [16, Ex. 5.1] and that the joint condition "MFCQ+WCR" was proven to be strictly weaker than

LICQ [16, Ex. 5.2]. Later, a simpler proof of this result was presented by Andreani et al. [13, Crlr. 4.3 and Thm. 4.1], using sequential optimality conditions. In the following sections, we generalize the WCR property and the result of [16] for NSOCP and NSDP, using an approach similar to [13].

As a matter of fact, Andreani, Echagüe, and Schuverdt [3] presented a result similar to Theorem 4, but under Janin's *constant rank constraint qualification* (CRCQ) [47], which is also weaker than LICQ and independent of "MFCQ+WCR". However, extending constant rank-type CQs to conic contexts is not easy, and finding an extension that preserves all of its interesting properties is even more difficult. In fact, there is a series of papers by Andreani et al. [8, 10–12] presenting distinct extensions of CRCQ for NSDP and NSOCP that suit distinct applications. For instance, [10, 12] deal with convergence of algorithms to first-order stationary points but no second-order properties were proven, whereas [11] presents a more geometric approach with some interesting theoretical properties but no application towards algorithms was provided. We should mention, nevertheless, that the extension of WCR presented this paper is not a particular case of any of the conditions from the aforementioned papers.

3 Second-order Cone Programming

The standard NSOCP problem can be seen as a particular case of (NCP) where $\mathbb{E} = \mathbb{R}^m$ and $\mathcal{K} \doteq \mathcal{K}_1 \times \cdots \times \mathcal{K}_r$ is a Cartesian product of Lorentz cones, that is, $\mathcal{K}_i \doteq \mathbb{L}^{m_i}$ for all $i \in \{1, \ldots, r\}$, where $m_1 + \cdots + m_r = m$ and $\mathcal{K}_i \subset \mathbb{R}^{m_i}$. In this section, we consider $\mathbb{R}^m$ with its standard inner product and the Euclidean norm. The notation $w = (w_0, \bar{w})$ refers to a partition of $w \in \mathbb{R}^{m_i}$ where $w_0 \in \mathbb{R}$ is its first entry and $\bar{w} \in \mathbb{R}^{m_i - 1}$ is the subvector with the remaining entries. To make the NSOCP problem explicit, define $g \doteq (g_1, \ldots, g_r)$ with $g_i : \mathbb{R}^n \to \mathbb{R}^{m_i}$ for every $i \in \{1, \ldots, r\}$, and obtain

$$
\begin{aligned}
\underset{x \in \mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\
\text{subject to} \quad & g_i(x) \in \mathcal{K}_i, \forall i \in \{1, \ldots, r\} \\
& h(x) = 0.
\end{aligned}
\tag{NSOCP}
$$

As usual in the study of (NSOCP), given a feasible point x, we define the following sets of indices, which constitute a partition of $\{1, \ldots, r\}$:

$$
\begin{aligned}
I_0(x) &\doteq \{i \in \{1, \ldots, r\} : g_i(x) = 0\}, \\
I_B(x) &\doteq \{i \in \{1, \ldots, r\} : g_i(x) \in \text{bd}^+(\mathcal{K}_i)\}, \\
I_I(x) &\doteq \{i \in \{1, \ldots, r\} : g_i(x) \in \text{int}(\mathcal{K}_i)\}.
\end{aligned}
\tag{7}
$$

Moreover, when we are dealing with a KKT point $x^\star$ associated with Lagrange multipliers $\omega^\star \in \mathcal{K}$ and $\mu^\star \in \mathbb{R}^p$, we consider the subset of $I_B(x^\star)$ given by

$$
I_{BB}(x^\star, \omega^\star) \doteq \{i \in I_B(x^\star) : \omega_i^\star \in \text{bd}^+(\mathcal{K}_i)\}
$$

and the critical subspace of (NSOCP) at $x^\star$ can be written in terms of such indices, as follows:

$$
S(x^\star) = \left\{ d \in \mathbb{R}^n : \begin{array}{l} Dh(x^\star)d = 0; \ Dg_i(x^\star)d = 0, i \in I_0(x^\star); \\ g_i(x^\star)^\top \Gamma_i Dg_i(x^\star)d = 0, i \in I_B(x^\star) \end{array} \right\},
\tag{8}
$$

where

$$\Gamma_i \doteq \begin{bmatrix} 1 & 0^\top \\ 0 & -\mathbb{I}_{m_i-1} \end{bmatrix} \in \mathbb{R}^{m_i \times m_i} \quad \text{for all } i \in \{1, \ldots, r\}. \tag{9}$$

The sigma-term at $x^\star$, when specialized to (NSOCP), can be written as

$$\sigma(x^\star, \omega^\star) = \sum_{i \in I_{BB}(x^\star, \omega^\star)} \sigma_i(x^\star, \omega^\star),$$

where

$$\sigma_i(x^\star, \omega^\star) = -\frac{[\omega_i^\star]_0}{[g_i(x^\star)]_0} Dg_i(x^\star)^\top \Gamma_i Dg_i(x^\star), \quad \text{for all } i \in I_{BB}(x^\star, \omega^\star). \tag{10}$$

We refer to [38] for details.

Also, the specialized characterization of the nondegeneracy condition in NSOCP, following Bonnans and Ramírez [28, Prop. 19], can be written as follows:

Proposition 5 *Let $x^\star \in \mathbb{R}^n$ be a feasible point of* (NSOCP). *The nondegeneracy condition holds at $x^\star$ if, and only if, the set*

$$\left\{\nabla h_i(x^\star)\right\}_{i \in \{1, \ldots, p\}} \bigcup \left\{\nabla g_{ij}(x^\star)\right\}_{\substack{i \in I_0(x^\star) \\ j \in \{1, \ldots, m_i\}}} \bigcup \left\{Dg_i(x^\star)^\top \Gamma_i \tilde{g}_i(x^\star)\right\}_{i \in I_B(x^\star)} \tag{11}$$

is linearly independent, where $\nabla g_{ij}(x)$ denotes the transpose of the j-th row of $Dg_i(x)$ and

$$\tilde{g}_i(x) \doteq (||\overline{g_i(x)}||_2, \overline{g_i(x)}). \tag{12}$$

In [4, Def. 3.3], the authors extend a sequential optimality condition called *Approximate-KKT* (AKKT) from NLP [7] to the NSOCP context. In short, AKKT is a punctual necessary optimality condition that also incorporates a bit of local information. That is, every point $x^\star$ that satisfies AKKT (though not necessarily KKT) is accompanied by a sequence $\{x^k\}_{k \in \mathbb{N}} \to x^\star$ such that each x^k approximately satisfies the KKT conditions with some approximate Lagrange multipliers ω^k and μ^k. Since our analyses are based on AKKT, we now recall its definition and some of its properties.

Definition 2 [AKKT for NSOCP] A feasible point $x^\star$ of (NSOCP) satisfies the AKKT condition when there exist sequences $\{x^k\}_{k \in \mathbb{N}} \to x^\star$, $\{\omega^k\}_{k \in \mathbb{N}} \subset \mathcal{K}$, and $\{\mu^k\}_{k \in \mathbb{N}} \subset \mathbb{R}^p$ such that

$$\nabla_x L(x^k, \omega^k, \mu^k) \to 0 \tag{13}$$

and

$$i \in I_I(x^\star) \Rightarrow \omega_i^k \to 0,$$

$$i \in I_B(x^\star) \Rightarrow \omega_i^k \to 0 \text{ or } \omega_i^k \in \text{bd}^+(\mathcal{K}_i) \text{ with } \frac{\bar{\omega}_i^k}{||\bar{\omega}_i^k||_2} \to \frac{g_i(\bar{x}^\star)}{||g_i(\bar{x}^\star)||_2}. \tag{14}$$

It was proved in [4, Thm. 3.1] that AKKT is indeed a genuine necessary optimality condition independently of CQs, in contrast with KKT. Also, their proof is constructive, which means it tells us how to obtain the sequences of perturbed KKT points and multipliers. Next, we state their result with a slightly different phrasing, in order to highlight such construction.

Theorem 6 *Let $x^\star$ be a local minimizer of* (NSOCP). *Then, for any given sequence* $\{\rho_k\}_{k\in\mathbb{N}} \to +\infty$, *there exists a sequence* $\{x^k\}_{k\in\mathbb{N}} \to x^\star$, *such that each x^k is a local minimizer of the regularized penalty function*

$$F_k(x) \doteq f(x) + \frac{1}{4}||x - x^\star||_2^4 + \frac{\rho_k}{2}\left(\sum_{i=1}^r ||\Pi_{\mathcal{K}_i}(-g_i(x))||_2^2 + ||h(x)||_2^2\right).$$

Also, the multiplier sequences given by $\omega_i^k \doteq \rho_k \Pi_{\mathcal{K}_i}(-g_i(x^k))$ *for all $i \in \{1, \ldots, r\}$ and* $\mu^k \doteq \rho_k h(x^k)$ *satisfy* (13) *and* (14) *together with* $\{x^k\}_{k\in\mathbb{N}}$. *Consequently, $x^\star$ satisfies AKKT.*

A key property of AKKT, as stated in [4, Thm. 3.1], is that the sequences of multipliers from Definition 2 must be bounded when $x^\star$ satisfies Robinson's CQ. Hence, AKKT implies KKT under Robinson's CQ. Also, in the same paper the authors present a variant of the classical *Powell-Hestenes-Rockafellar* (PHR) Augmented Lagrangian method (see [45, 63, 66]) and prove that its output sequences can be fully described by AKKT.

3.1 Second-order Optimality Conditions

Here, we build second-order analysis for (NCP) primarily under Robinson's CQ instead of nondegeneracy and strict complementarity, but since Robinson's CQ alone is not enough to complete that task [21], we also introduce a generalized version of the WCR property.

Definition 3 (WCR for NSOCP) We say that the *weak constant rank* property is satisfied at a feasible point $x^\star$ of (NSOCP) if there exists a neighborhood $\mathcal{N}$ of $x^\star$ such that the set

$$\left\{\nabla h_i(x)\right\}_{i\in\{1,\ldots,p\}} \bigcup \left\{\nabla g_{ij}(x)\right\}_{\substack{i\in I_0(x^\star)\\j\in\{1,\ldots,m_i\}}} \bigcup \left\{Dg_i(x)^\top \Gamma_i \tilde{g}_i(x)\right\}_{i\in I_B(x^\star)} \tag{15}$$

has the same rank, for all $x \in \mathcal{N}$.

In view of the characterization of nondegeneracy for NSOCP provided by Proposition 5, we see that nondegeneracy implies both Robinson's CQ and WCR in this context, just as in the NLP case. On the other hand, [16, Ex. 5.2] exhibits a point that satisfies MFCQ and WCR, but not LICQ. Hence, the joint condition "Robinson's CQ+WCR" is strictly weaker than nondegeneracy.

The main feature of the WCR property in NLP is its effect on the continuity of perturbations of the critical subspace around a feasible point $x^\star$. Next, we prove that this property is maintained in (NSOCP).

Lemma 1 *Let $x^\star \in \mathbb{R}^n$ be a feasible point of* (NSOCP). *Then, the WCR property holds at $x^\star$ if, and only if, the set-valued mapping $x \mapsto S(x, x^\star)$ is inner semicontinuous at $x^\star$, where*

$$S(x, x^\star) \doteq \left\{d \in \mathbb{R}^n : \begin{array}{l} Dh(x)d = 0; \ \forall i \in I_0(x^\star), Dg_i(x)d = 0; \\ \forall i \in I_B(x^\star), \tilde{g}_i(x)^\top \Gamma_i Dg_i(x)d = 0 \end{array}\right\}, \tag{16}$$

and $\tilde{g}_i$ is defined in (12).

Proof Following the steps of the proof of [43, Prop. 2], we see that [20, Thm. 1.1.8] tells us that $x \mapsto S(x, x^\star)$ is inner semicontinuous at $x^\star$ if, and only if, the set-valued mapping

$x \mapsto S(x, x^\star)^\circ$ is outer semicontinuous at $x^\star$, where $S(x, x^\star)^\circ \doteq -S(x, x^\star)^*$ denotes the *polar* of $S(x, x^\star)$. Since in this case we have

$$S(x, x^\star)^\circ = \left\{ \begin{array}{c} \sum\limits_{i \in I_0(x^\star)} Dg_i(x)^\top a_i + \sum\limits_{i=1}^p \nabla h_i(x)b_i + \\ + \sum\limits_{i \in I_B(x^\star)} Dg_i(x)^\top \Gamma_i \tilde{g}_i(x)c_i \end{array} : a_i \in \mathbb{R}^{m_i}, \ b_i, c_i \in \mathbb{R} \right\},$$

the result follows directly from [32, Prop. 3.2.9].

As in NLP, the subspace $S(x, x^\star)$ may be called *perturbed critical subspace* of (NSOCP) at x, around $x^\star$. The last ingredient we need for the main theorem of this section is an explicit characterization of the subdifferential of the projection onto $\mathcal{K}_i$. In order to present that, for each $i \in \{1, \ldots, r\}$, let $M_i : \mathbb{R} \times \mathbb{R}^{m_i-1} \to \mathbb{R}^{m_i \times m_i}$ be defined as

$$M_i(\xi, w) \doteq \frac{1}{2} \begin{bmatrix} 1 & w^\top \\ w & (1+\xi)\mathbb{I}_{m_i-1} - \xi w w^\top \end{bmatrix}$$

and observe that the matrix $M_i(\xi, u)$ is symmetric positive semidefinite whenever $|\xi| \leq 1$ and $||w||_2 = 1$ [51, Lem. 2.8].

The following lemma, that can be found in [62, Lem. 14] and [44, Prop. 4.8], provides a description of the B-subdifferential of the projection onto $\mathcal{K}_i$, in terms of $M_i(\xi, w)$.

Lemma 2 *The B-subdifferential $\partial_B \Pi_{\mathcal{K}_i}(z)$ of the orthogonal projection onto $\mathcal{K}_i$ at $z \in \mathbb{R}^{m_i}$ is given as follows:*

(a) If $z \in \mathrm{int}(-\mathcal{K}_i)$, then $\partial_B \Pi_{\mathcal{K}_i}(z) = \{0\}$;
(b) If $z \in \mathrm{int}(\mathcal{K}_i)$, then $\partial_B \Pi_{\mathcal{K}_i}(z) = \{\mathbb{I}_{m_i}\}$;
(c) If $z \notin \mathcal{K}_i \cup (-\mathcal{K}_i)$, then $\partial_B \Pi_{\mathcal{K}_i}(z) = \left\{ M_i\left(\dfrac{z_0}{||\bar{z}||_2}, \dfrac{\bar{z}}{||\bar{z}||_2} \right) \right\}$;
(d) If $z \in \mathrm{bd}^+(\mathcal{K}_i)$, then $\partial_B \Pi_{\mathcal{K}_i}(z) = \left\{ \mathbb{I}_{m_i}, M_i\left(1, \dfrac{\bar{z}}{||\bar{z}||_2} \right) \right\}$;
(e) If $z \in \mathrm{bd}^+(-\mathcal{K}_i)$, then $\partial_B \Pi_{\mathcal{K}_i}(z) = \left\{ 0, M_i\left(-1, \dfrac{\bar{z}}{||\bar{z}||_2} \right) \right\}$;
(f) If $z = 0$, then $\partial_B \Pi_{\mathcal{K}_i}(z) = \{0, \mathbb{I}_{m_i}\} \cup \{M_i(\xi, w) : |\xi| \leq 1, ||w||_2 = 1\}$.

To the best of our knowledge, the first specialized study on second-order necessary conditions for (NSOCP) is credited to Bonnans and Ramírez [28, Thm. 30], where they assume nondegeneracy and the so-called *second-order growth condition* (or *uniform growth condition*). Fukuda and Fukushima [38, Thm. 4.5] also developed second-order conditions via squared slack variables, under nondegeneracy and strict complementarity. Our contribution to this discussion is to draw attention to the fact that the nondegeneracy assumption can be strictly weakened and that strict complementarity is not necessary when considering WSOC, which is also the main result of this section.

Theorem 7 *Let $x^\star$ be a local minimizer of (NSOCP) satisfying Robinson's CQ and the WCR property. Then, there are some Lagrange multipliers $\omega^\star \in \mathcal{K}$ and $\mu^\star \in \mathbb{R}^p$ such that the KKT conditions and WSOC hold.*

Proof Let $x^\star$ be a local minimizer of (NSOCP). Then, by Theorem 6, for any given $\{\rho_k\}_{k\in\mathbb{N}} \to +\infty$, there exists a sequence $\{x^k\}_{k\in\mathbb{N}} \to x^\star$ such that x^k is a local minimizer of $F_k(x)$ for each k, where

$$F_k(x) = f(x) + \frac{1}{4}\|x - x^\star\|_2^4 + \frac{\rho_k}{2}\left(\sum_{i=1}^{r}\|\Pi_{\mathcal{K}_i}(-g_i(x))\|_2^2 + \|h(x)\|_2^2\right).$$

From the local optimality of x^k, we obtain

$$\nabla F_k(x^k) = \nabla f(x^k) + \|x^k - x^\star\|_2^2(x^k - x^\star)$$
$$-\sum_{i=1}^{r}Dg_i(x^k)^\top\rho_k\Pi_{\mathcal{K}_i}(-g_i(x^k)) + \rho_k Dh(x^k)^\top h(x^k) = 0$$

and by Theorem 2, for every $d \in \mathbb{R}^n$ and every $i \in \{1,\ldots,r\}$, there exists some $\chi_i^k \in \partial(\Pi_{\mathcal{K}_i}\circ -g_i)(x^k)$ such that $d^\top\nabla^2 F_k(x^k)d \geqslant 0$, where we denote by $\nabla^2 F_k(x^k)$ the element of the generalized Hessian of F_k at x^k that is defined in terms of χ_i^k, by an abuse of notation. That is,

$$\nabla^2 F_k(x^k) \doteq \nabla^2 f(x^k) + \|x^k - x^\star\|_2^2\,\mathbb{I}_n + 2(x^k - x^\star)(x^k - x^\star)^\top$$
$$-\sum_{i=1}^{r}\left(\sum_{j=1}^{m_i}(\rho_k\Pi_{\mathcal{K}_i}(-g_i(x^k)))_j\nabla^2 g_{ij}(x^k) - \rho_k Dg_i(x^k)^\top\chi_i^k\right)$$
$$+\sum_{j=1}^{p}\left(\rho_k h_j(x^k)\nabla^2 h_j(x^k) + \rho_k\nabla h_j(x^k)\nabla h_j(x^k)^\top\right).$$

Following Theorem 6, we define $\omega_i^k \doteq \rho_k\Pi_{\mathcal{K}_i}(-g_i(x^k))$ for all $i \in \{1,\ldots,r\}$, and $\mu^k \doteq \rho_k h(x^k)$ for every $k \in \mathbb{N}$, which satisfy (13). Also, it follows from Theorem 3 that there exists some $V_i^k \in \partial\Pi_{\mathcal{K}_i}(-g_i(x^k))$ such that

$$\chi_i^k = V_i^k \circ -Dg_i(x^k) = -V_i^k Dg_i(x^k),$$

where $\circ$ denotes a composition of linear operators. Hence, the expression $d^\top\nabla^2 F_k(x^k)d \geqslant 0$ can be rewritten as

$$d^\top\left(\nabla_x^2 L(x^k,\omega^k,\mu^k) + \rho_k\sum_{i=1}^{r}Dg_i(x^k)^\top V_i^k Dg_i(x^k)\right.$$
$$\left. + \rho_k\sum_{j=1}^{p}\nabla h_j(x^k)\nabla h_j(x^k)^\top\right)d \geqslant -d^\top\Delta^k d, \qquad (17)$$

where $\Delta^k \doteq \|x^k - x^\star\|_2^2\mathbb{I}_n + 2(x^k - x^\star)(x^k - x^\star)^\top \to 0$.

Under Robinson's CQ, the sequence $\{(\omega^k,\mu^k)\}_{k\in\mathbb{N}}$ is bounded (see the proof of [4, Thm. 3.3]). Then, for every limit point $(\omega^\star,\mu^\star)$ of $\{(\omega^k,\mu^k)\}_{k\in\mathbb{N}}$, note that $x^\star$ satisfies the KKT conditions. Without loss of generality, we assume $\{(\omega^k,\mu^k)\}_{k\in\mathbb{N}} \to (\omega^\star,\mu^\star)$. Now, from WCR and Lemma 1, we know that the mapping $x \mapsto S(x,x^\star)$ as in (16) is inner semi-continuous at $x^\star$, then for each $d \in S(x^\star)$ there exists a sequence $\{d^k\}_{k\in\mathbb{N}} \to d$ such that $d^k \in S(x^k,x^\star)$ for all $k \in \mathbb{N}$.

For each $i \in \{1,\ldots,r\}$, define

$$u_i^k = \left([u_i^k]_0, \overline{u_i^k}\right) \doteq Dg_i(x^k)d^k.$$

Our next step is to compute $\rho_k(u_i^k)^\top V_i^k u_i^k$ and its limit points in three independent cases:

1. If $i \in I_I(x^\star)$, we have $g_i(x^k) \in \text{int}(\mathcal{K}_i)$ for all k sufficiently large. Then, from Lemma 2 item (a), $V_i^k = 0$ and $\rho_k(u_i^k)^\top V_i^k u_i^k = 0$ for such k;
2. If $i \in I_0(x^\star)$, recalling that $d^k \in S(x^k, x^\star)$, we have $u_i^k = 0$ for all $k \in \mathbb{N}$, which means $\rho_k(u_i^k)^\top V_i^k u_i^k = 0$ in this case as well;
3. If $i \in I_B(x^\star)$, the sequence $\{g_i(x^k)\}_{k \in \mathbb{N}}$ can be essentially split into three subsequences, which have distinct influences over $\rho_k(u_i^k)^\top V_i^k u_i^k$. Hence, they are separately analysed below, where N_1, N_2, and N_3 constitute a partition of $\mathbb{N}$:

   (i) $\{g_i(x^k)\}_{k \in N_1} \subset \text{int}(\mathcal{K}_i)$. Here, $\omega_i^k = 0$ for every $k \in N_1$. Also, by item (a) of Lemma 2, $V_i^k = 0$ and $\rho_k(u_i^k)^\top V_i^k u_i^k = 0$ for every $k \in N_1$;
   (ii) $\{g_i(x^k)\}_{k \in N_2} \subset \mathbb{R}^m \setminus (\mathcal{K}_i \cup -\mathcal{K}_i)$. From Lemma 2 item (c) we obtain

$$V_i^k = M_i \left(-\frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2}, -\frac{\overline{g_i(x^k)}}{||\overline{g_i(x^k)}||_2} \right)$$

which can be explicitly written as

$$V_i^k = \frac{1}{2} \begin{bmatrix} 1 & -\frac{\overline{g_i(x^k)}}{||\overline{g_i(x^k)}||_2}^\top \\ -\frac{\overline{g_i(x^k)}}{||\overline{g_i(x^k)}||_2} & Z^k \end{bmatrix},$$

where

$$Z^k \doteq \left(1 - \frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2} \right) \mathbb{I}_{m_i - 1} + \left(\frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2} \right) \frac{\overline{g_i(x^k)} \; \overline{g_i(x^k)}^\top}{||\overline{g_i(x^k)}||_2^2},$$

and it is elementary to see that

$$(u_i^k)^\top V_i^k u_i^k = \frac{1}{2} \left([u_i^k]_0^2 - \frac{2[u_i^k]_0 \overline{g_i(x^k)}^\top \overline{u_i^k}}{||\overline{g_i(x^k)}||_2} + \left(1 - \frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2} \right) ||\overline{u_i^k}||_2^2 \right.$$
$$\left. + \frac{[g_i(x^k)]_0 (\overline{g_i(x^k)}^\top \overline{u_i^k})^2}{||\overline{g_i(x^k)}||_2^3} \right).$$

Also, since $d^k \in S(x^k, x^\star)$ and $i \in I_B(x^\star)$ we have $\tilde{g}_i(x^k)^\top \Gamma_i u_i^k = 0$, or equivalently, $\overline{g_i(x^k)}^\top \overline{u_i^k} = ||\overline{g_i(x^k)}||_2 [u_i^k]_0$. Replacing this in the above expression, we obtain:

$$(u_i^k)^\top V_i^k u_i^k$$
$$= \frac{1}{2} \left([u_i^k]_0^2 - 2[u_i^k]_0^2 + \left(1 - \frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2} \right) ||\overline{u_i^k}||_2^2 + \frac{[u_i^k]_0^2 [g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2} \right)$$
$$= -\frac{1}{2} \left(1 - \frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2} \right) \left([u_i^k]_0^2 - ||\overline{u_i^k}||_2^2 \right). \tag{18}$$

It follows from our specific choice of approximate multiplier that

$$\omega_i^k = \rho_k \Pi_{\mathcal{K}_i}(-g_i(x^k)) = \rho_k \frac{||\overline{g_i(x^k)}||_2 - [g_i(x^k)]_0}{2} \left(1, -\frac{\overline{g_i(x^k)}}{||\overline{g_i(x^k)}||_2}\right).$$

Hence, we have

$$\frac{[\omega_i^k]_0}{||\overline{g_i(x^k)}||_2} = \frac{\rho_k}{2}\left(1 - \frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2}\right)$$

and from (18), we obtain

$$\rho_k(u_i^k)^\top V_i^k u_i^k = -\frac{[\omega_i^k]_0}{||\overline{g_i(x^k)}||_2}\left([u_i^k]_0^2 - ||\overline{u_i^k}||_2^2\right) = -\frac{[\omega_i^k]_0}{||\overline{g_i(x^k)}||_2}(u_i^k)^\top \Gamma_i u_i^k. \quad (19)$$

(iii) $\{g_i(x^k)\}_{k \in N_3} \subset \mathrm{bd}^+(\mathcal{K}_i)$. For every $k \in N_3$, we have $\omega_i^k = 0$. Also, for all such k, Lemma 2 item (e) implies

$$V_i^k = \tau M_i\left(-1, -\frac{\overline{g_i(x^k)}}{||\overline{g_i(x^k)}||_2}\right),$$

for some $\tau \in [0, 1]$. Then, note that

$$M_i\left(-1, -\frac{\overline{g_i(x^k)}}{||\overline{g_i(x^k)}||_2}\right) = M_i\left(-\frac{[g_i(x^k)]_0}{||\overline{g_i(x^k)}||_2}, -\frac{\overline{g_i(x^k)}}{||\overline{g_i(x^k)}||_2}\right),$$

which means that simply multiplying (19) by τ is enough to obtain $\rho_k(u_i^k)^\top V_i^k u_i^k = 0$ as well, since $[\omega_i^k]_0 = 0$.

Considering exclusively any infinite subsequence indexed by N_1, N_2, or N_3, based on our previous analyses we observe that, for k sufficiently large, (17) implies

$$\liminf_{k \to \infty} (d^k)^\top \left(\nabla_x^2 L(x^k, \omega^k, \mu^k) - \sum_{i \in I_B(x^\star)} \frac{[\omega_i^k]_0}{||\overline{g_i(x^k)}||_2} Dg_i(x^k)^\top \Gamma_i Dg_i(x^k)\right) d^k \geq 0.$$

Since $x^k \to x^\star$, $\omega^k \to \omega^\star$, $\mu^k \to \mu^\star$, $d^k \to d \in S(x^\star)$ and $||\overline{g_i(x^k)}||_2 \to ||\overline{g_i(x^\star)}||_2 = [g_i(x^\star)]_0$ when $i \in I_B(x^\star)$, we conclude that

$$d^\top \left(\nabla_x^2 L(x^\star, \omega^\star, \mu^\star) + \sum_{i \in I_{BB}(x^\star)} \sigma_i(x^\star, \omega^\star)\right) d \geq 0 \quad \text{for all } d \in S(x^\star),$$

where $\sigma_i(x^\star, \omega^\star)$ is defined as in (10). Therefore, $x^\star$ satisfies WSOC.

Note that Theorem 7 contains a proof for the fact that every feasible limit point of any sequence $\{x^k\}_{k \in \mathbb{N}}$ generated by an *external penalty method* must satisfy WSOC if it satisfies Robinson's CQ and WCR. Moreover, with minor adaptations, it is possible to prove that the same holds for every feasible limit point of a modified extension of the *augmented Lagrangian method* for NLP considered in [25]. And finally, we remark that if $m_1 = \ldots = m_r = 1$, that is, if (NSOCP) reduces to a NLP problem, then Theorem 7 recovers a result by Andreani et al. [13, Crlr. 4.2 and Crlr. 4.3] with an alternative proof.

4 Semidefinite Programming

In this section, $\mathbb{S}^m$ is the linear space of all $m \times m$ symmetric matrices with real entries, equipped with the (*Frobenius*) inner product given by $\langle M, N \rangle \doteq \text{trace}(MN)$ and the norm $||M||_F \doteq \sqrt{\langle M, M \rangle}$, for every $M, N \in \mathbb{S}^m$. We define $M \odot N$ as the (*Hadamard*) entry-wise product between M and N. Also, the cone of all symmetric positive semidefinite matrices is denoted by $\mathbb{S}^m_+$ and $\succeq$ is the partial order induced by it, that is, $M \succeq N$ if, and only if, $M - N \in \mathbb{S}^m_+$. Similarly, $M \succ N$ when $M - N \in \text{int}(\mathbb{S}^m_+)$.

Recall that every $M \in \mathbb{S}^m$ has a *spectral decomposition* in the form $M = U \Lambda U^\top$, where U is an orthogonal matrix whose columns are eigenvectors of M and $\Lambda = \text{Diag}(\lambda_1^U(M), \ldots, \lambda_m^U(M))$ is a diagonal matrix whose entries are the eigenvalues of M respective to the columns of U. It is well-known that the orthogonal projection of M onto $\mathbb{S}^m_+$ under $|| \cdot ||_F$ is given by

$$\Pi_{\mathbb{S}^m_+}(M) \doteq U \text{Diag}(\max\{0, \lambda_1^U(M)\}, \ldots, \max\{0, \lambda_m^U(M)\}) U^\top.$$

The specialization of (NCP) to an NSDP is obtained by setting $\mathbb{E} = \mathbb{S}^m$ and $\mathcal{K} = \mathbb{S}^m_+$, and it is often stated in the form

$$
\begin{aligned}
& \underset{x \in \mathbb{R}^n}{\text{Minimize}} && f(x), \\
& \text{subject to} && g(x) \succeq 0, \\
& && h(x) = 0.
\end{aligned}
\tag{NSDP}
$$

Here, for simplicity, we consider a single conic constraint since it is enough to cover all major aspects of the problem and the notation would be unnecessarily heavy otherwise. Similarly to the NSOCP case, several concepts of general conic programming can be specialized and explicitly characterized here, for example, the tangent cone to $\mathbb{S}^m_+$ at some $M \in \mathbb{S}^m_+$ can be written as

$$T_{\mathbb{S}^m_+}(M) = \{E \in \mathbb{S}^m : V^\top E V \in \mathbb{S}^{|\beta|}_+\},$$

where $V \in \mathbb{R}^{m \times |\beta|}$ is any matrix with orthonormal columns that form a basis for $\text{Ker}(M)$ and $|\beta|$ is its dimension (see [67] for details).

Let x be a feasible point of (NSDP). In this section we always consider spectral decompositions of $g(x)$ that keep zero and nonzero eigenvalues separated, for example,

$$g(x) = U \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} U^\top,$$

where $\mathbb{S}^\alpha \ni \Lambda \succ 0$ and $\alpha \doteq \alpha(x)$ is the set of indices of the positive eigenvalues of $g(x)$. Let $\beta \doteq \beta(x)$ be the set of indices of the null eigenvalues of $g(x)$ and partition the columns U with respect to α and β as follows: $U \doteq [U_\alpha, U_\beta]$. For every $d \in \mathbb{R}^n$, define $D\tilde{g}(x)[d] \doteq U^\top Dg(x)[d]U$ as a reverse conjugation of $Dg(x)[d]$ around $g(x)$ and set

$$D\tilde{g}(x)[d] = \begin{bmatrix} (D\tilde{g}(x)[d])_{\alpha\alpha} & (D\tilde{g}(x)[d])_{\alpha\beta} \\ (D\tilde{g}(x)[d])_{\alpha\beta}^\top & (D\tilde{g}(x)[d])_{\beta\beta} \end{bmatrix}$$

as a partition of $D\tilde{g}(x)[d]$ with respect to α and β. Note that since U is an orthogonal matrix, the inner product is invariant to reverse conjugation in terms of U, that is,

$$\langle A, B \rangle = \text{trace}(AB) = \text{trace}(UU^\top A U U^\top B) = \text{trace}(U^\top A U U^\top B U) = \langle \tilde{A}, \tilde{B} \rangle$$

for all $A, B \in \mathbb{S}^m$.

The critical cone of (NSDP) at a feasible point $x^\star$ is given by

$$C(x^\star) = \{d \in \mathbb{R}^n : \nabla f(x^\star)^\top d = 0, Dh(x^\star)d = 0, (D\tilde{g}(x^\star)[d])_{\beta\beta} \succeq 0\}.$$

Under Robinson's CQ, if $x^\star$ is a KKT point associated with some Lagrange multipliers $\omega^\star \in \mathbb{S}^m_+$ and $\mu^\star \in \mathbb{R}^p$ that satisfy strict complementarity, then the critical cone becomes equal to the critical subspace

$$S(x^\star) = \{d \in \mathbb{R}^n : Dh(x^\star)d = 0, (D\tilde{g}(x^\star)[d])_{\beta\beta} = 0\}$$

since $\tilde{\omega}^\star_{\alpha\alpha} = 0$ and $\nabla f(x^\star)^\top d = \langle Dg(x^\star)[d], \omega^\star \rangle - \langle Dh(x^\star)d, \mu^\star \rangle = 0$ for every $d \in \mathbb{R}^n$.

In [15], Andreani, Haeser, and Viana proposed an extension of the AKKT condition from NLP to (NSDP) as well. We state it as follows:

Definition 4 [AKKT for NSDP] A feasible point $x^\star$ of (NSDP) satisfies the AKKT condition when there are sequences $\{x^k\}_{k\in\mathbb{N}} \to x^\star$, $\{\omega^k\}_{k\in\mathbb{N}} \subset \mathbb{S}^m_+$, and $\{\mu^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^p$ such that

$$\nabla_x L(x^k, \omega^k, \mu^k) \to 0 \tag{20}$$

and

$$\lambda^U_i(g(x^\star)) > 0 \Rightarrow \lambda^{S^k}_i(\omega^k) = 0, \tag{21}$$

for sufficiently large k, where U diagonalizes $g(x^\star)$, S^k diagonalizes ω^k for each k, and $S^k \to U$.

If $x^\star$ is a KKT point of (NSDP) associated with multipliers $\omega^\star \in \mathbb{S}^m_+$ and $\mu^\star \in \mathbb{R}^p$, note that the complementarity condition $\langle g(x^\star), \omega^\star \rangle = 0$ holds for $g(x^\star), \omega^\star \in \mathbb{S}^m_+$ if, and only if, $g(x^\star)\omega^\star = 0$, then it is elementary to check that $g(x^\star)$ and $\omega^\star$ must be simultaneously diagonalizable (i.e. they commute) in this case. In light of this, note that Definition 4 relaxes the commutativity between $g(x^\star)$ and $\omega^\star$ by requiring $S^k \to U$.

Also in [15], the authors prove that AKKT as in Definition 4 is a necessary optimality condition, independently of the fulfilment of constraint qualifications. We state it below in the same form as Theorem 6, with some emphasis on how the sequences that compose it are generated.

Theorem 8 *Let $x^\star$ be a local minimizer of (NSDP). Then, for any sequence $\{\rho_k\}_{k\in\mathbb{N}} \to +\infty$, there exists some $\{x^k\}_{k\in\mathbb{N}} \to x^\star$, such that for every k, x^k is a local minimizer of the regularized penalty function*

$$F_k(x) \doteq f(x) + \frac{1}{4}||x - x^\star||^4_2 + \frac{\rho_k}{2}\left(||\Pi_{\mathbb{S}^m_+}(-g(x))||^2_F + ||h(x)||^2_2\right).$$

Also, the multiplier sequences given by $\omega^k \doteq \rho_k \Pi_{\mathbb{S}^m_+}(-g(x^k))$ and $\mu^k \doteq \rho_k h(x^k)$ satisfy (20) and (21) with $\{x^k\}_{k\in\mathbb{N}}$. Consequently, since ω^k and $g(x^k)$ are simultaneously diagonalizable in this case for every $k \in \mathbb{N}$, $x^\star$ satisfies AKKT.

Under Robinson's CQ, the sequences $\{\omega^k\}_{k\in\mathbb{N}}$ and $\{\mu^k\}_{k\in\mathbb{N}}$ are bounded, and also all limit points of those sequences are Lagrange multipliers associated with $x^\star$ [15, Thm. 6.1]. That is, AKKT implies KKT in this case. An augmented Lagrangian algorithm is also presented in [15] for NSDP, whose global convergence theory is built around AKKT. Such results were sharpened in [5] and further extended in [6], for the general (NCP).

4.1 Second-order Analysis

As mentioned before, there are many different works that deal with a specialized second-order analysis for (NSDP), which mainly differ in the assumptions required for it and the techniques employed to characterize the sigma-term. As far as we know, the first work on this topic is due to Shapiro [67, Section 4], who obtained the very useful and practical expression

$$\sigma(x^\star, \omega^\star) = [2\langle \omega^\star, \partial_i g(x^\star) g(x^\star)^\dagger \partial_j g(x^\star)\rangle]_{i,j \in \{1,\dots,n\}}, \tag{22}$$

where $g(x^\star)^\dagger$ is the *Moore-Penrose pseudoinverse* of $g(x^\star)$. Shapiro's idea was to write the semidefinite cone using the second-order directional derivative of the least eigenvalue function $\lambda_{\min}$, as follows:

$$\mathbb{S}^m_+ = \{M \in \mathbb{S}^m \colon \lambda_{\min}(M) \geqslant 0\},$$

and the expression of the sigma-term comes from the expression of the second-order directional derivative of $\lambda_{\min}$. Moreover, his second-order analysis was based on the uniqueness of the Lagrange multiplier, via nondegeneracy, and strict complementarity (Theorem 4). Then, Jarre [49, Thm. 2] presented another way of achieving Shapiro's characterization of the sigma-term, and consequently an alternative proof for Theorem 4, using a locally equivalent formulation of (NSDP) based on the Schur complement of $g(x^\star)_{\alpha\alpha}$, which turned out to be a more elementary proof. Later, Lourenço, Fukuda, and Fukushima [57, Props. 5.1 and 5.2], studied a characterization of the semidefinite cone with squared slack variables

$$\mathbb{S}^m_+ = \{M \in \mathbb{S}^m \colon \exists Z \in \mathbb{S}^m, \ M - ZZ = 0\},$$

which induces a reformulation of (NSDP) as a NLP problem. Then, the authors related the classical second-order conditions for NLP with the second-order conditions for (NSDP) (with the curvature term), under the same hypotheses as Shapiro and Jarre. Forsgren [36, Thm. 2], on the other hand, proved that strict complementarity was not needed for WSOC when assuming a different notion of regularity that treats structural sparsity and also uses Schur complements. In this section, we use the characterization

$$\mathbb{S}^m_+ = \{M \in \mathbb{S}^m \colon ||\Pi_{\mathbb{S}^m_+}(-M)||^2 = 0\},$$

and the generalized derivative of the orthogonal projection, to obtain second-order results that do not require uniqueness of multipliers, nor strict complementarity.

Our approach is based on extracting second-order information from AKKT and Theorem 8 and we do this in a similar manner of the previous section, which means we begin by exhibiting a characterization of the derivative of $\Pi_{\mathbb{S}^m_+}$, then we extend the WCR condition from NLP to NSDP and, at last, we compute the sigma-term using the second-order (generalized) derivative of F_k.

Based on the works of Bonnans et al. [26] and Pang et al. [62], Sun [69] characterized the B-subdifferential of the projection onto the semidefinite cone. To make a proper reference, we define the following matrix:

$$\mathcal{B}(\lambda^U(M)) \doteq \left[\frac{\max\{\lambda^U_i(M), 0\} + \max\{\lambda^U_j(M), 0\}}{|\lambda^U_i(M)| + |\lambda^U_j(M)|} \right]_{i,j \in \{1,\dots,m\}},$$

where $0/0$ is set as 1 and U is an orthogonal matrix that diagonalizes M. Next, we make a slightly adapted transcription of a proposition by Qi [64, Prop. 2.5] summarizing Sun's result:

Proposition 9 *Suppose that* $M = U\Lambda U^\top$ *is the spectral decomposition of* $M \in \mathbb{S}^m$ *and let* α, β *and* γ *be the sets of indices of the positive, zero and negative eigenvalues of* M, *respectively. Without loss of generality, assume those three blocks are separated and that* $U \doteq [U_\alpha, U_\beta, U_\gamma]$. *Then, for any* $V \in \partial_B \Pi_{\mathbb{S}^m_+}(-M)$ *there exists a* $V_{|\beta|} \in \partial_B \Pi_{\mathbb{S}^{|\beta|}_+}(0)$ *such that*

$$V[H] = U \begin{bmatrix} 0 & 0 & \tilde{H}_{\alpha\gamma} \odot \mathcal{B}(\lambda^U(M))_{\alpha\gamma} \\ 0 & V_{|\beta|}[\tilde{H}_{\beta\beta}] & \tilde{H}_{\beta\gamma} \\ \tilde{H}^\top_{\alpha\gamma} \odot \mathcal{B}(\lambda^U(M))^\top_{\alpha\gamma} & \tilde{H}^\top_{\beta\gamma} & \tilde{H}_{\gamma\gamma} \end{bmatrix} U^\top \quad (23)$$

for every $H \in \mathbb{S}^m$, *where* $\tilde{H} \doteq U^\top HU$. *Conversely, for every* $V_{|\beta|} \in \partial_B \Pi_{\mathbb{S}^{|\beta|}_+}(0)$, *there exists some* $V \in \partial_B \Pi_{\mathbb{S}^m_+}(-M)$ *such that* (23) *holds.*

Even though we assume the eigenvalues are separated by sign, the ordering inside each partition is not relevant. Note that Proposition 9 is still true if we replace the B-subdifferential for the Clarke subdifferential.

Corollary 1 *Under the hypotheses of* Proposition 9, *for any* $V \in \partial \Pi_{\mathbb{S}^m_+}(-M)$ *there exists a* $V_{|\beta|} \in \partial \Pi_{\mathbb{S}^{|\beta|}_+}(0)$ *such that* (23) *holds. Conversely, for every* $V_{|\beta|} \in \partial \Pi_{\mathbb{S}^{|\beta|}_+}(0)$, *there exists some* $V \in \partial \Pi_{\mathbb{S}^m_+}(-M)$ *such that* (23) *holds.*

Proof Let $V \in \partial \Pi_{\mathbb{S}^m_+}(-M)$. Then, $V = \sum_{i=1}^s a_i V^i$, for some $s \in \mathbb{N}$, some $a_i \geqslant 0$, and some $V^i \in \partial_B \Pi_{\mathbb{S}^m_+}(-M)$, $i \in \{1, \ldots, s\}$, with $\sum_{i=1}^s a_i = 1$. This means there are $V^i_{|\beta|} \in \partial_B \Pi_{\mathbb{S}^{|\beta|}_+}(0)$, $i \in \{1, \ldots, s\}$, such that (23) holds. Hence, for every $H \in \mathbb{S}^m$, we have $V[H] = \sum_{i=1}^s a_i V^i[H]$ and the proof is over, because $\sum_{i=1}^s a_i V^i_{|\beta|}[\tilde{H}_{\beta\beta}] \in \partial \Pi_{\mathbb{S}^{|\beta|}_+}(0)$. The converse is analogous.

In order to study perturbations of the critical subspace around a given point $x^\star$ via WCR, let $\bar{\alpha}$ and $\bar{\beta}$ represent the indices of positive and zero eigenvalues of $g(x^\star)$, respectively, regarding the decomposition

$$g(x^\star) = U\mathrm{Diag}(\lambda^U(g(x^\star)))U^\top, \quad (24)$$

where $U \doteq [U_{\bar{\alpha}}, U_{\bar{\beta}}]$ is a matrix whose columns are eigenvectors of $g(x^\star)$ and, in particular, the columns of $U_{\bar{\beta}}$ form a basis for $\mathrm{Ker}(g(x^\star))$. Moreover, we will use a construction from Bonnans and Shapiro's book [29, Ex. 3.98 and Ex. 3.140], which will be stated as a lemma below:

Lemma 3 *Let* $M^\star \in \mathbb{S}^m$, *set* $\bar{\beta}$ *as the indices of zero eigenvalues of* M, *and let* $U_{\bar{\beta}}$ *be a matrix with orthonormal columns that span* $\mathrm{Ker}(M^\star)$. *There exists a neighborhood* $\mathcal{N}$ *of* $M^\star$ *and an analytic matrix function* $\mathcal{U}_{\bar{\beta}} : \mathcal{N} \to \mathbb{R}^{m \times |\bar{\beta}|}$ *such that* $\mathcal{U}_{\bar{\beta}}(M^\star) = U_{\bar{\beta}}$ *and, for every* $M \in \mathcal{N}$, *the columns of* $\mathcal{U}_{\bar{\beta}}(M)$ *form an orthonormal basis for the space spanned by the eigenvectors associated with the* $|\bar{\beta}|$ *smallest eigenvalues of* M.

This construction allows us to approximate the critical subspace around $x^\star$. Indeed, let $\mathcal{N}$ be the neighborhood of $g(x^\star)$ and $\mathcal{U}_{\bar{\beta}} : \mathcal{N} \to \mathbb{R}^{m \times |\bar{\beta}|}$ be the function given by Lemma 3 such that $\mathcal{U}_{\bar{\beta}}(g(x^\star)) = U_{\bar{\beta}}$. Then, for every x close enough to $x^\star$ so that $g(x) \in \mathcal{N}$, consider the following set:

$$S(x, x^\star) = \{d \in \mathbb{R}^n : \mathcal{U}_{\bar{\beta}}(g(x))^\top Dg(x)[d]\mathcal{U}_{\bar{\beta}}(g(x)) = 0, Dh(x)d = 0\},$$

which will be called *perturbed critical subspace* at x, centered at $x^\star$.

Extending WCR from NLP to (NSDP) is not a trivial task because the notion of "rank" of the three-dimensional tensor $Dg(x)$ may have multiple meanings. Fortunately, there is a useful characterization of nondegeneracy by Shapiro and Fan [68], which provides some insight on how to talk about rank in NSDP. Next, we make a transcription of this result as stated in [67, Prop. 6], for completeness.

Proposition 10 *Suppose that the dimension of* $\operatorname{Ker}(g(x^\star))$ *is* $|\bar{\beta}|$ *and let* $U_{\bar{\beta}} \doteq [u_1, \ldots, u_{|\bar{\beta}|}] \in \mathbb{R}^{m \times |\bar{\beta}|}$ *be a matrix whose columns form a basis for* $\operatorname{Ker}(g(x^\star))$. *Then, nondegeneracy holds at a feasible point* $x^\star$ *of* (NSDP) *if, and only if, the set of n-dimensional vectors* $\{v_{ij} \colon 1 \leqslant i \leqslant j \leqslant |\bar{\beta}|\} \cup \{\nabla h_i(x^\star) \colon i \in \{1, \ldots, p\}\}$ *is linearly independent, where* $v_{ij} \doteq [u_i^\top \partial_\ell g(x^\star) u_j]_{\ell \in \{1, \ldots, n\}}$.

Inspired by this characterization, we define WCR as follows:

Definition 5 [WCR for NSDP] Let $x^\star$ be a feasible point of (NSDP) and let (24) be a spectral decomposition of $g(x^\star)$. We say that $x^\star$ satisfies the weak constant rank (WCR) property when there exists a neightborhood $\mathcal{N}$ of $x^\star$ such that the set

$$\{\bar{v}_{ij}(x) \colon 1 \leqslant i \leqslant j \leqslant |\bar{\beta}|\} \cup \{\nabla h_i(x) \colon i \in \{1, \ldots, p\}\}$$

has the same rank for every $x \in \mathcal{N}$, where

$$\bar{v}_{ij}(x) \doteq [\bar{u}_i(x)^\top \partial_\ell g(x) \bar{u}_j(x)]_{\ell \in \{1, \ldots, n\}}$$

and $\bar{u}_1(x), \ldots, \bar{u}_{|\bar{\beta}|}(x) \in \mathbb{R}^m$ denote the columns of $\mathcal{U}_{\bar{\beta}}(g(x))$.

Also, in the following lemma we prove that WCR as in Definition 5 is equivalent to the inner semicontinuity of the mapping $x \mapsto S(x, x^\star)$ at $x^\star$.

Lemma 4 *A feasible point* $x^\star$ *satisfies WCR if, and only if, the set-valued mapping* $x \mapsto S(x, x^\star)$ *is inner semicontinuous at* $x^\star$.

Proof First, we shall prove that, for every $x \in \mathbb{R}^n$,

$$S(x, x^\star) = \left\{ d \in \mathbb{R}^n \colon \bar{v}_{ij}(x)^\top d = 0, \ 1 \leqslant i \leqslant j \leqslant |\bar{\beta}|; \ Dh(x)d = 0 \right\}. \tag{25}$$

Note that for each $\ell \in \{1, \ldots, n\}$ we have

$$\mathcal{U}_{\bar{\beta}}(g(x^\star))^\top \partial_\ell g(x) \mathcal{U}_{\bar{\beta}}(g(x^\star)) = \begin{bmatrix} \bar{u}_1(x)^\top \partial_\ell g(x) \bar{u}_1(x) & \cdots & \bar{u}_1(x)^\top \partial_\ell g(x) \bar{u}_{|\bar{\beta}|}(x) \\ \vdots & \ddots & \vdots \\ \bar{u}_{|\bar{\beta}|}(x)^\top \partial_\ell g(x) \bar{u}_1(x) & \cdots & \bar{u}_{|\bar{\beta}|}(x)^\top \partial_\ell g(x) \bar{u}_{|\bar{\beta}|}(x) \end{bmatrix},$$

hence, considering that $\bar{u}_j(x)^\top \partial_\ell g(x) \bar{u}_i(x) = \bar{u}_i(x)^\top \partial_\ell g(x) \bar{u}_j(x) = (\bar{v}_{ij}(x))_\ell$ for every i, j, ℓ, we obtain

$$\mathcal{U}_{\bar{\beta}}(g(x^\star))^\top Dg(x)[d] \mathcal{U}_{\bar{\beta}}(g(x^\star)) = \begin{bmatrix} \bar{v}_{11}(x)^\top d & \cdots & \bar{v}_{1|\bar{\beta}|}(x)^\top d \\ \vdots & \ddots & \vdots \\ \bar{v}_{1|\bar{\beta}|}(x)^\top d & \cdots & \bar{v}_{|\bar{\beta}||\bar{\beta}|}(x)^\top d \end{bmatrix},$$

whence follows (25).

Now, similarly to the NSOCP case, since $\partial_\ell g$ and $\mathcal{U}_{\bar\beta}$ are continuous, $\bar{v}_{ij}$ is also continuous, then a result from Facchinei and Pang [32, Prop. 3.2.9] tells us that WCR is equivalent to the outer semicontinuity of the mapping $x \mapsto S(x, x^\star)^\circ$ at $x^\star$, where

$$S(x, x^\star)^\circ = \left\{ \sum_{1 \leqslant i \leqslant j \leqslant |\bar\beta|} a_{ij} \bar{v}_{ij}(x) + \sum_{i=1}^{p} b_i \nabla h_i(x) \,:\, a_{ij} \in \mathbb{R}, b_i \in \mathbb{R} \right\}$$

using the characterization in (25). Then, the desired result follows from [20, Thm. 1.1.8], which states that the inner semicontinuity of a set-valued mapping at a given point is equivalent to the outer semicontinuity of its polar at that point.

Clearly, WCR as in Definition 5 is implied by nondegeneracy, in view of Proposition 10. Also, let us assume for a moment that $g(x)$ is a structurally diagonal matrix constraint whose diagonal elements are denoted by $g_1(x), \ldots, g_m(x)$, and let $x^\star$ be such that $g(x) \in \mathbb{S}_+^m$. Without loss of generality, let us assume that $g_1(x^\star) > \ldots > g_{|\bar\alpha|}(x^\star) > g_{|\bar\alpha|+1}(x^\star) = \ldots = g_{|\bar\alpha|+|\bar\beta|}(x^\star) = 0$. Then, we can take

$$\mathcal{U}_{\bar\beta}(g(x)) \doteq \begin{bmatrix} 0 \\ \mathbb{I}_{|\bar\beta|} \end{bmatrix}$$

as a constant function to obtain that $\bar{v}_{ii}(x) = \nabla g_i(x)$ for every $i \in \{1, \ldots, |\bar\beta|\}$ and $\bar{v}_{ij}(x) = 0$ when $i \neq j$. That is, the WCR condition as in Definition 5 recovers the NLP definition of WCR when such NLP constraints are modelled as a single structurally diagonal matrix constraint, with this choice of $\mathcal{U}_{\bar\beta}$. It is important to keep in mind, however, that even if the constraints $g_1(x) \geqslant 0, \ldots, g_m(x) \geqslant 0$ satisfy LICQ at $x^\star$, nondegeneracy may not hold at $x^\star$, as observed by Shapiro in [67, p. 309]. The converse, on the other hand, is true. Now, recall that [16, Ex. 5.2] exhibits an NLP problem with a feasible point that satisfies "MFCQ+WCR", but not LICQ, and the above discussion tells us that it can be used again to prove that nondegeneracy is strictly stronger than "Robinson's CQ+WCR". Moreover, nondegeneracy and LICQ are equivalent when considering multiple unidimensional semidefinite constraints, and so are Definition 5 and the NLP version of WCR. Furthermore, Forsgren [36, Section 2.3] and Andreani et al. [9, Def. 3.2] considered regularity notions different from nondegeneracy, that also recover the standard LICQ in NLP. Thus, in all cases, regardless of modelling, the example of [16, Ex. 5.2] can be used to conclude that "Robinson's CQ+WCR" is strictly weaker than all existing notions of nondegeneracy.

With this in mind, we proceed to the main result of this section:

Theorem 11 *If $x^\star$ is a local minimizer such that Robinson's CQ and the WCR property hold, then there are some Lagrange multipliers $\omega^\star \in \mathbb{S}_+^m$ and $\mu^\star \in \mathbb{R}^p$ such that the KKT conditions and WSOC hold for this pair of multipliers.*

Proof If $x^\star$ is a local minimizer of (NSDP), Theorem 8 tells us that for any given $\{\rho_k\}_{k \in \mathbb{N}} \to +\infty$, there is some sequence $\{x^k\}_{k \in \mathbb{N}} \to x^\star$ such that, for every $k \in \mathbb{N}$, x^k is a local minimizer of the penalty function

$$F_k(x) = f(x) + \frac{1}{4}\|x - x^\star\|_2^4 + \frac{\rho_k}{2}\left(\|\Pi_{\mathbb{S}_+^m}(-g(x))\|_F^2 + \|h(x)\|_2^2\right).$$

Hence, it satisfies the first-order stationarity condition

$$\nabla F(x^k) = \nabla f(x^k) + \|x^k - x^\star\|_2^2(x^k - x^\star) + Dh(x^k)^\top(\rho_k h(x^k))$$
$$- Dg(x^k)^*[\rho_k \Pi_{\mathbb{S}_+^m}(-g(x^k))] = 0.$$

Springer

Setting the approximate multipliers $\omega^k \doteq \rho_k \Pi_{\mathbb{S}_+^m}(-g(x^k))$ and $\mu^k \doteq \rho_k h(x^k)$, we obtain (20) and (21) due to Theorem 8. Also, x^k is second-order stationary in the nonsmooth sense (see Theorem 2), which means that, for each unitary vector $d \in \mathbb{R}^n$, there exists some $\chi^k \in \partial(\Pi_{\mathbb{S}_+^m} \circ -g)(x^k)$ such that

$$
\begin{aligned}
& d^\top \nabla^2 F(x^k) d \\
&= d^\top \left(\nabla^2 f(x^k) - D^2 g(x^k)^* [\rho_k \Pi_{\mathbb{S}_+^m}(-g(x^k))] + \sum_{i=1}^p \mu_i^k \nabla^2 h(x^k) \right) d + \\
&\quad + \rho_k (Dh(x^k)d)^\top Dh(x^k)d - d^\top \left(Dg(x^k)^* \left[\rho_k \chi^k[d] \right] \right) + d^\top \Delta^k d \\
&\geqslant 0
\end{aligned}
\tag{26}
$$

where $\Delta^k \doteq \|x^k - x^\star\|_2^2 \mathbb{I}_n + 2(x^k - x^\star)(x^k - x^\star)^\top$ and $\nabla^2 F(x^k)$ denotes the element of $\partial^2 F_k(x^k)$ that is defined in terms of χ^k, as an abuse of notation. By Theorem 3, there exists some $V^k \in \partial\Pi_{\mathbb{S}_+^m}(-g(x^k))$, such that

$$
\chi^k = V^k \circ -Dg(x^k),
$$

for every $k \in \mathbb{N}$.

Under Robinson's CQ, the sequences $\{\omega^k\}_{k \in \mathbb{N}}$ and $\{\mu^k\}_{k \in \mathbb{N}}$ are bounded, so they have convergent subsequences which we will consider to be themselves from now on, without loss of generality. Denote their limits by $\omega^\star$ and $\mu^\star$, respectively. In [15, Thm. 6.1], the authors also prove that $\omega^\star$ and $\mu^\star$ are Lagrange multipliers associated with $x^\star$.

Now, let $d \in S(x^\star)$. By WCR there is a sequence $\{d^k\}_{k \in \mathbb{N}} \to d$ such that $d^k \in S(x^k, x^\star)$ for every k. Rewriting (26) in terms of d^k, V^k, ω^k, and μ^k, we obtain

$$
\begin{aligned}
(d^k)^\top \nabla_x^2 L(x^k, \omega^k, \mu^k) d^k + \rho_k(Dh(x^k)d^k)^\top Dh(x^k)d^k & \\
+ \rho_k \left\langle Dg(x^k)[d^k], V^k \left[Dg(x^k)[d^k] \right] \right\rangle &\geqslant -\delta^k,
\end{aligned}
\tag{27}
$$

where $\delta^k \doteq (d^k)^\top \Delta^k d^k \to 0$. The following paragraphs prove that (27) implies

$$
d^\top \nabla_x^2 L(x^\star, \omega^\star, \mu^\star) d + 2 \left\langle Dg(x^\star)[d], \omega^\star Dg(x^\star)[d]g(x^\star)^\dagger \right\rangle \geqslant 0,
\tag{28}
$$

which is enough to complete the proof since

$$
2 \left\langle \omega^\star, Dg(x^\star)[d]g(x^\star)^\dagger Dg(x^\star)[d] \right\rangle = d^\top \sigma(x^\star, \omega^\star) d,
$$

for every $d \in \mathbb{R}^n$ due to (22).

To complete that task, we proceed to analyse the behaviour of the sequence $\{\rho_k \langle Dg(x^k)[d^k], V^k[Dg(x^k)[d^k]] \rangle\}_{k \in \mathbb{N}}$ in distinct cases. In the following paragraphs, we let $\alpha \doteq \alpha(x^k)$, $\beta \doteq \beta(x^k)$, and $\gamma \doteq \gamma(x^k)$ be the sets of indices of the positive, zero and negative eigenvalues of $g(x^k)$, respectively, regarding the spectral decomposition

$$
g(x^k) = S^k \mathrm{Diag}(\lambda^{S^k}(g(x^k)))(S^k)^\top
$$

with $S^k \to U$. Recall that, by construction, the columns of $\mathcal{U}_{\bar{\beta}}(g(x^k))$ span the eigenspace associated with the $\bar{\beta}$ smallest eigenvalues of $g(x^k)$, for all k sufficiently large. Denote the submatrix of S^k that has the eigenvectors associated with the $\bar{\beta}$ smallest eigenvectors of $g(x^k)$ in its columns by $S_{\bar{\beta}}^k$, and since $d^k \in S(x^k, x^\star)$, we have $(S_{\bar{\beta}}^k)^\top Dg(x^k)[d^k]S_{\bar{\beta}}^k = 0$ for every k large enough. We proceed by analysing a few cases:

1. If $g(x^\star) \succ 0$, then $-g(x^k) \prec 0$ for k sufficiently large. For such k, since $\gamma(x^k) = \beta(x^k) = \emptyset$ and $\alpha(x^k) = \{1, \ldots, m\}$, we obtain and $V^k[Dg(x^k)[d^k]] = 0$ from Proposition 9, which implies

$$\rho_k \langle Dg(x^k)[d^k], \; V^k[Dg(x^k)[d^k]] \rangle = 0.$$

   Also, note that $\sigma(x^\star, \omega^\star) = 0$ in this case, because $\langle g(x^\star), \omega^\star \rangle = 0$ implies $\omega^\star = 0$.

2. If $g(x^\star) = 0$, then $g(x^\star)^\dagger = 0$ and $\sigma(x^\star, \omega^\star) = 0$ as well. On the other hand, note that

$$\langle Dg(x^k)[d^k], \; V^k[Dg(x^k)[d^k]] \rangle = \langle D\tilde{g}(x^k)[d^k], \; (U^k)^\top V^k[Dg(x^k)[d^k]]U^k \rangle$$
$$= 0,$$

   because $\bar{\beta} = \{1, \ldots, m\}$ and $d^k \in S(x^k, x^\star)$ implies

$$D\tilde{g}(x^k)[d^k] = (S^k)^\top Dg(x^k)[d^k]S^k = 0$$

   in this case.

3. If $g(x^\star) \succeq 0$, but $g(x^\star) \neq 0$, assume the diagonalization is taken such that nonzero eigenvalues are separated from the others and the common zero eigenvalues between $g(x^\star)$ and $\omega^\star$ are discriminated, that is,

$$g(x^\star) = U \begin{bmatrix} \Lambda & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} U^\top \quad \text{and} \quad \omega^\star = U \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Gamma \end{bmatrix} U^\top,$$

where $\mathbb{S}^{|\bar{\alpha}|} \ni \Lambda \succ 0$ and $\mathbb{S}^{|\bar{\gamma}|} \ni \Gamma \succ 0$ are diagonal matrices, $\bar{\kappa} \cup \bar{\gamma}$ is a partition of $\bar{\beta}$, and U is orthogonal. Denoting $H \doteq Dg(x^\star)[d]$, since $d \in S(x^\star)$ we get $\tilde{H}_{\bar{\beta}\bar{\beta}} = 0$ and

$$g(x^\star)^\dagger H \omega^\star = U \begin{bmatrix} \Lambda^{-1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} U^\top U \begin{bmatrix} \tilde{H}_{\bar{\alpha}\bar{\alpha}} & \tilde{H}_{\bar{\alpha}\bar{\kappa}} & \tilde{H}_{\bar{\alpha}\bar{\gamma}} \\ \tilde{H}_{\bar{\kappa}\bar{\alpha}} & 0 & 0 \\ \tilde{H}_{\bar{\gamma}\bar{\alpha}} & 0 & 0 \end{bmatrix} U^\top U \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Gamma \end{bmatrix} U^\top$$
$$= U \begin{bmatrix} 0 & 0 & \Lambda^{-1}\tilde{H}_{\bar{\alpha}\bar{\gamma}}\Gamma \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} U^\top.$$

Conveniently,

$$\Lambda^{-1}\tilde{H}_{\bar{\alpha}\bar{\gamma}}\Gamma = \left[\frac{\lambda_j^U(\omega^\star)}{\lambda_i^U(g(x^\star))} \tilde{H}_{ij} \right]_{i \in \bar{\alpha}, j \in \bar{\gamma}} \doteq A \odot \tilde{H}_{\bar{\alpha}\bar{\gamma}}, \tag{29}$$

where $\mathbb{R}^{|\bar{\alpha}| \times |\bar{\gamma}|} \ni A \doteq \left[\lambda_j^U(\omega^\star)\lambda_i^U(g(x^\star))^{-1} \right]_{i \in \bar{\alpha}, j \in \bar{\gamma}}$. Also, note that

$$\langle H, g(x^\star)^\dagger H \omega^\star \rangle = \left\langle \tilde{H}, \frac{1}{2} \begin{bmatrix} 0 & 0 & A \odot \tilde{H}_{\bar{\alpha}\bar{\gamma}} \\ 0 & 0 & 0 \\ (A \odot \tilde{H}_{\bar{\alpha}\bar{\gamma}})^\top & 0 & 0 \end{bmatrix} \right\rangle.$$

In view of this characterization of the sigma-term over d, its relation with (27) can be made explicit. Consider the following spectral decomposition of $g(x^k)$:

$$g(x^k) = S^k \begin{bmatrix} \Lambda_+^k & 0 & 0 & 0 \\ 0 & \Lambda_-^k & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Psi^k \end{bmatrix} (S^k)^\top, \tag{30}$$

where we separate the eigenvalues of $g(x^k)$ primarily by their sign and, secondarily, by their limit points. For instance, $\Lambda_+^k \in \mathbb{S}^{|\alpha_+|}$ are the positive ones that converge to Λ, while $\Lambda_-^k \in \mathbb{S}^{|\alpha_-|}$ are the positive ones that converge to zero. The squared block of zeros in the diagonal of (30) is of dimension β and $\Psi^k \in \mathbb{S}^{|\gamma|}$ contains the negative eigenvalues of $g(x^k)$. Also, $|\alpha_+| + |\alpha_-| + |\beta| + |\gamma| = m$. Recall that S^k simultaneously diagonalizes $g(x^k)$ and ω^k, by definition of ω^k. In order to simplify the notation, define

$$H^k \doteq Dg(x^k)[d^k]$$

and

$$B_{\alpha\gamma}^k \doteq \tilde{H}_{\alpha\gamma}^k \odot \mathcal{B}(\lambda^{S^k}(-g(x^k)))_{\alpha\gamma}.$$

Using the characterization of V^k provided in (23) from Proposition 9 (and Corollary 1), we obtain

$$V^k[H^k] = S^k \begin{bmatrix} 0 & 0 & 0 & B_{\alpha_+\gamma}^k \\ 0 & 0 & 0 & B_{\alpha_-\gamma}^k \\ 0 & 0 & V_{|\beta|}[\tilde{H}_{\beta\beta}] & \tilde{H}_{\beta\gamma}^k \\ (B_{\alpha_+\gamma}^k)^\top & (B_{\alpha_-\gamma}^k)^\top & (\tilde{H}_{\beta\gamma}^k)^\top & \tilde{H}_{\gamma\gamma}^k \end{bmatrix} (S^k)^\top.$$

Since $\rho_k \langle H^k, V^k[H^k] \rangle = \langle \tilde{H}^k, \rho_k V^k[\tilde{H}^k] \rangle$, it is fundamental to note that for every $i \in \alpha_+$ and $j \in \gamma$,

$$\left(\rho_k \mathcal{B}(\lambda^{S^k}(-g(x^k))) \right)_{ij} = \frac{\rho_k \lambda_j^{S^k}(-g(x^k))}{\lambda_j^{S^k}(-g(x^k)) - \lambda_i^{S^k}(-g(x^k))} \rightarrow \frac{\lambda_j^U(\omega^\star)}{\lambda_i^U(g(x^\star))}, \quad (31)$$

because $\rho_k \lambda_j^{S^k}(-g(x^k)) = \lambda_j^{S^k}(\omega^k)$ and $\lambda_j^{S^k}(-g(x^k)) \rightarrow 0$. Also, keep in mind that

$$\lambda_i^{S^k}(-g(x^k)) = -\lambda_i^{S^k}(g(x^k)).$$

The blocks indexed by α_-, β, and γ, are all blocks of zeros because if k is large enough, we must have $|\alpha_-| + |\beta| + |\gamma| = \bar{\beta}$ and, on the other hand, since $d^k \in S(x^k, x^\star)$ we also have that $\tilde{H}_{\bar{\beta}\bar{\beta}} \doteq (S_{\bar{\beta}}^k)^\top Dg(x^k)[d^k]S_{\bar{\beta}}^k = 0$. Similarly, $Dh(x^k)d^k = 0$. Thus

$$\lim_{k \to \infty} \rho_k \langle H^k, V^k[H^k] \rangle = 2\langle H, g(x^\star)^\dagger H\omega^\star \rangle$$

and, consequently, (27) implies (28), which means $x^\star$ satisfies the WSOC with the multiplier $\omega^\star$.

In the presence of nondegeneracy, the set of Lagrange multipliers is a singleton and Theorem 11 recovers the classical result of [67], but even without assuming uniqueness of the Lagrange multiplier it ensures there will be at least one multiplier satisfying WSOC. Moreover, in contrast with [49, 57, 67], our proof does not require strict complementarity; but nevertheless, if it does hold, then the proof of Theorem 11 can be significantly simplified, since in this case the sequence $\{g(x^k)\}_{k \in \mathbb{N}}$ is nonsingular and we can avoid the use of subdifferentials.

5 Conclusion

In this paper, we proved that every local minimizer of a nonlinear semidefinite program or a nonlinear second-order cone program satisfies the weak second-order necessary optimality

condition under Robinson's constraint qualification and the so-called weak constant rank property (WCR), which was extended from NLP [16]. This joint condition is strictly weaker than nondegeneracy in NLP, NSOCP, and NSDP. We also stress that we do not assume strict complementarity, which is common in second-order analyses for conic programming. In contrast, our second-order necessary condition is based on the lineality space of the critical cone, and not the critical cone itself. This is consistent with the algorithmic practice of second-order algorithms as no algorithm is known to achieve a stronger second-order necessary optimality condition (see the extended version of [22] for details).

In the context of conic programming, several different approaches are known for obtaining second-order necessary optimality conditions [28, 31, 36, 38, 49, 57, 67]. We present a novel approach by extending the existing theory of first-order sequential optimality conditions to the second-order context. In particular, it is remarkable to see the appearance of the *sigma-term* in such a variety of approaches, which contributes to the understanding of this concept.

Our approach has a heavy algorithmic taste, as our proof is based on the construction of a sequence of approximate solutions of penalized subproblems, very similarly to a sequence generated by practical algorithms. In particular, a similar first-order approach has recently led to several improvements of global convergence theory of augmented Lagrangian methods in conic contexts [4–6, 15].

Thus, this paper opens the path to the development of second-order algorithms in conic optimization, which, as far as we know, has not been considered yet in the literature. In particular, augmented Lagrangian and interior point methods [25, 42] are expected to be well suited to the techniques we develop here. In this context, the joint condition "Robinson's CQ+WCR" is the natural candidate for a condition to guarantee global convergence to a second-order stationary point.

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

Declarations

Competing Interest The authors have no relevant financial or non-financial interests to disclose.

References

1. Alizadeh, F., Goldfarb, D.: Second-order cone programming. Mathematical Programming Series B **95**, 3–51 (2003)
2. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: Second-order negative-curvature methods for box-constrained and general constrained optimization. Computational Optimization and Applications **45**, 209–236 (2010)
3. Andreani, R., Echagüe, C.E., Schuverdt, M.L.: Constant-rank condition and second-order constraint qualification. Journal of Optimization theory and Applications **146**, 255–266 (2010)
4. Andreani, R., Fukuda, E.H., Haeser, G., Santos, D.O., Secchin, L.D.: Optimality conditions for nonlinear second-order cone programming and symmetric cone programming. Technical report (2019). Available at Optimization Online: http://www.optimization-online.org/DB_HTML/2019/10/7436.html. Accessed 18 Oct 2019

5. 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. Computational Optimization and Applications, 79 (2021)

6. Andreani, R., Gómez, W., Haeser, G., Mito, L.M., Ramos, A.: On optimality conditions for nonlinear conic programming. To appear in Mathematics of Operations Research (2022). https://doi.org/10.1287/moor.2021.1203

7. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**, 627–641 (2011)

8. Andreani, R., Haeser, G., Mito, L.M., Ramırez, H., Santos, D.O., Silveira, T.P.: Naive constant rank-type constraint qualifications for multifold second-order cone programming and semidefinite programming. To appear in Optimization Letters (2021). https://doi.org/10.1007/s11590-021-01737-w

9. R. Andreani, G. Haeser, L. M. Mito, and H. Ramírez C. Weak notions of nondegeneracy in nonlinear semidefinite programming. Technical report, 2020. Available at ArXiv:2012.14810

10. R. Andreani, G. Haeser, L. M. Mito, and H. Ramírez C. Sequential constant rank constraint qualifications for nonlinear semidefinite programming with applications. Technical report, 2021. Available at ArXiv:2106.00775v2

11. R. Andreani, G. Haeser, L M. Mito, H. Ramírez C., and T. P. Silveira. First- and second-order optimality conditions for second-order cone and semidefinite programming under a constant rank condition. Technical report, 2021. Available at ArXiv:2107.04693v1

12. R. Andreani, G. Haeser, L M. Mito, H. Ramírez C., and T. P. Silveira. Sequential constant rank for nonlinear second-order cone programming problems. Technical report, 2021

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

14. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two new weak constraint qualifications and applications. SIAM Journal on Optimization **22**, 1109–1135 (2012)

15. Andreani, R., Haeser, G., Viana, D.S.: Optimality conditions and global convergence for nonlinear semidefinite programming. Mathematical Programming **180**(1–2), 203–235 (2020)

16. Andreani, R., Martínez, J.M., Schuverdt, M.L.: On second-order optimality conditions for nonlinear programming. Optimization **56**, 529–542 (2007)

17. Andreani, R., Martínez, J.M., Svaiter, B.F.: A new sequential optimality condition for constrained optimization and algorithmic consequences. SIAM Journal on Optimization **20**, 3533–3554 (2010)

18. Roberto Andreani and Leonardo D Secchin. A note on the convergence of an augmented lagrangian algorithm to second-order stationary points. Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, 6(1), 2018

19. M. F. Anjos and J. B. Lassere (eds.). Handbook on Semidefinite, Conic and Polynomial Optimization. International Series in Operations Research and Management Science. Springer US. New York, 2012

20. J. P. Aubin and H. Frankowska. Set-Valued Analysis. Birkhauser, 1990

21. Baccari, A.: On the Classical Necessary Second-Order Optimality Conditions. Journal of Optimization Theory and Applications **123**(1), 213–221 (2004)

22. R. Behling, G. Haeser, A. Ramos, and D. S. Viana. On a conjecture in second-order optimality conditions. Journal of Optimization Theory and Applications, 176(3):625–633, 2018. Extended version at arXiv:1706.07833

23. Benson, H.Y., Vanderbei, R.J.: Solving problems with semidefinite and related constraints using interior-point methods for nonlinear programming. Mathematical Programming **95**, 279–302 (2003)

24. E. G. Birgin, W. Gómez, G. Haeser, L. M. Mito, and D. O. Santos. An Augmented Lagrangian algorithm for nonlinear semidefinite programming applied to the covering problem. Computational and Applied Mathematics, 39(10), 2020

25. Birgin, E.G., Haeser, G., Ramos, A.: Augmented Lagrangians with constrained subproblems and convergence to second-order stationary points. Computational Optimization and Applications **69**(1), 51–75 (2018)

26. Bonnans, J.F., Cominetti, R., Shapiro, A.: Sensitivity analysis of optimization problems under second order regular constraints. Mathematics of Operations Research **23**(4), 806–831 (1998)

27. Bonnans, J.F., Cominetti, R., Shapiro, A.: Second order optimality conditions based on parabolic second order tangent sets. SIAM Journal on Optimization **9**(2), 466–492 (1999)

28. J. F. Bonnans and H. Ramírez C. Perturbation analysis of second-order cone programming problems. Mathematical Programming, 104:205-227, 2005

29. Bonnans, J.F., Shapiro, A.: Pertubation Analysis of Optimization Problems. Springer-Verlag, Berlin (2000)

30. N. H. Chieu, L. V. Hien, T. T. A. Nghia, H. A. Tuan. Second order optimality conditions for strong local minimizers via subgradient graphical derivative. Technical report, 2019. Available at ArXiv:1903.05746

31. Cominetti, R.: Metric regularity, tangent sets, and second-order optimality conditions. Applied Mathematics and Optimization **287**, 265–287 (1990)
32. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, vol. I. Springer Series in Operations Research. Springer-Verlag, New York (2003)
33. Fares, B., Apkarian, P., Noll, D.: An augmented Lagrangian method for a class of LMI-constrained problems in robust control theory. International Journal of Control **74**(4), 348–360 (2001)
34. Fares, B., Noll, D., Apkarian, P.: Robust control via sequential semidefinite programming. SIAM Journal on Control and Optimization **40**, 1791–1820 (2002)
35. Fitzpatrick, S., Phelps, R.R.: Differentiability of the metric projection in Hilbert space. Transactions of the American Mathematical Society **270**(2), 483–501 (1982)
36. Forsgren, A.: Optimality conditions for nonconvex semidefinite programming. Mathematical Programming **88**, 105–128 (2000)
37. Freund, R.W., Jarre, F., Vogelbusch, C.H.: Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling. Mathematical Programming **109**, 581–611 (2007)
38. Fukuda, E.H., Fukushima, M.: The use of squared slack variables in nonlinear second-order cone programming. Journal of Optimization Theory and Applications **170**(2), 394–418 (2016)
39. Fukuda, E.H., Silva, P.J.S., Fukushima, M.: Differentiable exact penalty functions for nonlinear second-order cone programs. SIAM Journal on Optimization **22**(4), 1607–1633 (2012)
40. Gill, P.E., Kungurtsev, V., Robinson, D.P.: A stabilized SQP method: global convergence. IMA Journal of Numerical Analysis (2016). https://doi.org/10.1093/imanum/drw004
41. Gould, N.I.M., Toint, P.L.: A note on the convergence of barrier algorithms to second-order necessary points. Mathematical programming **85**(2), 433–438 (1999)
42. Haeser, G.: A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms. Computational Optimization and Applications **70**(2), 615–639 (2018)
43. Haeser, G.: Some theoretical limitations of second-order algorithms for smooth constrained optimization. Operations Research Letters **46**(3), 295–299 (2018)
44. Hayashi, S., Yamashita, N., Fukushima, M.: A combined smoothing and regularization method for monotone second-order cone complementarity problems. SIAM Journal on Optimization **15**, 593–615 (2005)
45. Hestenes, M.R.: Multiplier and gradient methods. Journal of Optimization Theory and Applications **4**, 303–320 (1969)
46. Hiriart-Urruty, J.-B., Strodiot, J.-J., Nguyen, V.H.: Generalized Hessian matrix and second-order optimality conditions for problems with C1,1 data. Applied Mathematics and Optimization **11**, 43–56 (1984)
47. Janin, R.: Direction derivative of the marginal function in nonlinear programming. Mathematical Programming Studies **21**, 127–138 (1984)
48. Jarre, F.: An interior method for nonconvex semidefinite programs. Optimization and Engineering **1**, 347–372 (2000)
49. F. Jarre. Elementary optimality conditions for nonlinear SDPs. In M. F. Anjos and J. B. Lassere, editors, Handbook on Semidefinite, Conic and Polynomial Optimization. International Series in Operations Research and Management Science, 2012
50. Kanno, Y., Takewaki, I.: Sequential semidefinite program for maximum robustness design of structures under load uncertainty. Journal of Optimization Theory and Applications **130**, 265–287 (2006)
51. 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 Journal on Optimization **20**, 297–320 (2009)
52. Kanzow, C., Nagel, C., Kato, H., Fukushima, M.: Successive linearization methods for nonlinear semidefinite programs. Computational Optimization and Applications **31**, 251–273 (2005)
53. Kato, H., Fukushima, M.: An SQP-type algorithm for nonlinear second-order cone programs. Optimization Letters **1**, 129–144 (2007)
54. Kawasaki, H.: An envelope-like effect of infinitely many inequality constraints on second-order necessary conditions for minimization problems. Mathematical Programming **41**, 73–96 (1988)
55. Kočvara, M., Stingl, M.: Solving nonconvex SDP problems of structural optimization with stability control. Optimization Methods and Software **19**(5), 595–609 (2004)
56. Lobo, M.S., Vandenberghe, L., Boyd, S., Lebret, H.: Applications of second-order cone programming. Linear Algebra and Applications **284**, 193–228 (1998)
57. Lourenço, B.F., Fukuda, E.H., Fukushima, M.: Optimality conditions for nonlinear semidefinite programming via squared slack variables. Mathematical Programming **168**(1–2), 177–200 (2018)
58. Mohammadi, A., Mordukhovich, B.S., Sarabi, M.E.: Parabolic regularity in geometric variational analysis. Transactions of the American Mathematical Society **374**, 1711–1763 (2021)

59. Morguerza, J.M., Prieto, F.J.: An augmented Lagrangian interior-point method using directions of negative curvature. Mathematical Programming **95**(3), 573–616 (2003)
60. Murty, K.G., Kabadi, S.N.: Some NP-complete problems in quadratic and nonlinear programming. Mathematical Programming **39**(2), 117–129 (1987)
61. Páles, Z., Zeidan, V.: Generalized Jacobian for functions with infinite dimensional range and domain. Set-Valued Analysis **15**, 331–375 (2007)
62. Pang, J.S., Sun, D., Sun, J.: Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems. Mathematics of Operations Research **28**(1), 39–63 (2003)
63. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. Academic Press, New York (1982)
64. Qi, H.D.: Local duality of nonlinear semidefinite programming. Mathematics of Operations Research **34**(1), 124–141 (2009)
65. Robinson, S.M.: First-order conditions for general nonlinear optimization. SIAM Journal on Applied Mathematics **30**(4), 597–610 (1976)
66. Rockafellar, R.T.: Lagrange multipliers and optimality. SIAM Review **35**, 183–238 (1993)
67. Shapiro, A.: First and second order analysis of nonlinear semidefinite programs. Mathematical Programming, Series B **77**(2), 301–320 (1997)
68. Shapiro, A., Fan, M.K.H.: On Eigenvalue optimization. SIAM Journal on Optimization **5**(3), 552–569 (1995)
69. Sun, D.: The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. Mathematics of Operations Research **31**(4), 761–776 (2006)
70. H. Wolkowicz, R. Saigal, and L. (eds.) Vandenberghe. Handbook of Semidefinite Programming: Theory, Algorithms, an Applications. International Series in Operations Research and Management Science. Springer US. New York, 2000
71. Yamashita, H., Yabe, H.: A primal-dual interior point method for nonlinear optimization over second-order cones. Optimization Methods and Software **24**, 407–426 (2009)
72. Yamashita, H., Yabe, H.: A survey of numerical methods for nonlinear semidefinite programming. Journal of the Operations Research Society of Japan **58**(1), 24–60 (2015)
73. Zhao, X.Y., Sun, D., Toh, K.C.: A Newton-CG augmented Lagrangian method for semidefinite programming. SIAM Journal on Optimization **20**(4), 1737–1765 (2010)