Mathematical Programming (2020) 180:203–235
https://doi.org/10.1007/s10107-018-1354-5

FULL LENGTH PAPER
Series A

Optimality conditions and global convergence for nonlinear semidefinite programming

Roberto Andreani[1] · Gabriel Haeser[2] · Daiana S. Viana[2,3]

Received: 13 March 2018 / Accepted: 28 November 2018 / Published online: 7 December 2018

Abstract

Sequential optimality conditions have played a major role in unifying and extending global convergence results for several classes of algorithms for general nonlinear optimization. In this paper, we extend theses concepts for nonlinear semidefinite programming. We define two sequential optimality conditions for nonlinear semidefinite programming. The first is a natural extension of the so-called Approximate-Karush–Kuhn–Tucker (AKKT), well known in nonlinear optimization. The second one, called Trace-AKKT, is more natural in the context of semidefinite programming as the computation of eigenvalues is avoided. We propose an augmented Lagrangian algorithm that generates these types of sequences and new constraint qualifications are proposed, weaker than previously considered ones, which are sufficient for the global convergence of the algorithm to a stationary point.

Keywords Nonlinear semidefinite programming · Optimality conditions · Constraint qualifications · Practical algorithms

Mathematics Subject Classification 90C22 · 90C46 · 90C30

This work was supported by FAPESP (Grants 2013/05475-7 and 2017/18308-2), CNPq and CAPES.

✉ Gabriel Haeser
ghaeser@ime.usp.br

Roberto Andreani
andreani@ime.unicamp.br

Daiana S. Viana
daiana@ime.usp.br

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

2 Department of Applied Mathematics, University of São Paulo, São Paulo, SP, Brazil

3 Center of Exact and Technological Sciences, Federal University of Acre, Rio Branco, AC, Brazil

1 Introduction

Nonlinear semidefinite programming (NLSDP) is a generalization of the usual nonlinear programming problem, where the inequality constraints are replaced by a conic constraint defined by the matrix of constraints being negative semidefinite. The study of nonlinear semidefinite programming problems has grown a great deal in recent years from several application fields, such as control theory, structural optimization, material optimization, eigenvalue problems and others. See [22,23,26,39,40,48,54,61] and references therein. Theoretical issues such as optimality conditions, duality and nondegeneracy were also studied and several algorithms for solving NLSDPs have been proposed. The interested reader may see, for instance, [1,13,19,28,35,44,51,52,64,66]. The particular case of (linear) semidefinite programming, which generalizes linear programming, has seen a variety of applications in several fields, and is an important tool in numerical analysis [25,37,58,60].

For nonlinear programming (NLP), a useful concept is the notion of sequential optimality conditions [4]. These conditions are genuine necessary optimality conditions, independently of the fulfillment of any constraint qualification, such as the linear independence of the gradients of active constraints (LICQ), the Mangasarian-Fromovitz constraint qualification (MFCQ), or the constant positive linear dependence (CPLD). Besides that, the sequences generated by several classes of algorithms (as Augmented Lagrangians, interior point methods, sequential quadratic programming and inexact restoration methods; see [6]), are precisely the sequences required for verifying the sequential optimality condition. This property makes sequential optimality conditions useful tools for naturally providing a perturbed optimality condition, which is suitable for the definition of stopping criteria and complexity analysis for several algorithms. Also, a careful study of the relation of sequential optimality conditions with classical stationarity measures under a constraint qualification, yields global convergence results under weak assumptions [3,8,9].

Sequential optimality conditions have been shown to be important tools for extending and unifying global convergence results for NLP algorithms. In NLSDP, most of the methods proposed need constraint qualifications to guarantee global convergence results. Both [53] and [62] present an Augmented Lagrangian method for solving NLSDPs. Although [62] uses quadratic penalty, and [53] uses modified barrier functions in their Augmented Lagrangian methods, both need constraint qualifications such as CPLD, MFCQ or Nondegeneracy. With sequential optimality conditions, we introduce a very natural optimality condition for NLSDP, without constraint qualifications, that is fulfilled by the limit points of an Augmented Lagrangian algorithm. With this approach, we have also been able to introduce a companion constraint qualification, weaker than the ones previously considered for NLSDP, in order to provide a classical global convergence result to a KKT point.

In this article we extend the Approximate-Karush–Kuhn–Tucker (AKKT) condition, known in nonlinear programming, to the context of NLSDP. We also introduce a new sequential optimality condition that we call Trace-AKKT (TAKKT). We will show that these conditions are genuine necessary optimality conditions for nonlinear semidefinite programming and we will show that an Augmented Lagrangian method produces AKKT and TAKKT sequences.

This paper is organized as follows: In Sect. 2, we formally define the nonlinear semidefinite programming, the notation, and we present some preliminary results for NLSDP needed in the remaining sections. In Sect. 3, we define the AKKT condition and we prove that it is a genuine optimality condition without constraint qualifications. In Sect. 4, we show that an Augmented Lagrangian method, based on the quadratic penalization, generates a sequence whose feasible limit points satisfy AKKT. In Sect. 5, we introduce a new sequential optimality condition called TAKKT, which is more suitable for nonlinear semidefinite programming. In Sect. 6, we introduce new constraint qualifications, weaker than the ones previously considered for global convergence results of nonlinear semidefinite programming algorithms, which are suitable for our global convergence analysis. In Sect. 7, we present our conclusions and final remarks.

2 Preliminaries

In this section we introduce the main notations adopted throughout the paper, together with known optimality conditions for nonlinear semidefinite programming. We will consider the following problem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Minimize}} \quad & f(x), \\ \text{subject to} \quad & G(x) \preceq 0, \end{aligned} \tag{NLSDP}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{S}^m$ are continuously differentiable functions. For simplicity, we do not take equality constraints into consideration. We denote by $\mathbb{S}^m$ the set of symmetric $m \times m$ matrices equipped with the inner product $\langle A, B\rangle := \mathrm{tr}(AB)$, $A, B \in \mathbb{S}^m$, where $\mathrm{tr}(A)$ denotes the trace of the matrix A. The set $\mathscr{F}$ will denote the feasible set $\mathscr{F} := \{x \in \mathbb{R}^n \mid G(x) \preceq 0\}$. The notation $G(x) \preceq 0$ means that $G(x)$ is a symmetric negative semidefinite matrix. The set of symmetric positive semidefinite matrices will be denoted by $\mathbb{S}^m_+$, while the set of symmetric negative semidefinite matrices will be denoted by $\mathbb{S}^m_-$. Given a matrix $A \in \mathbb{S}^m$ and an orthogonal diagonalization $A = U\Lambda U^T$ of A, we denote by $\lambda_i^U(A)$ the eigenvalue of A at position i on the diagonal matrix Λ, that is, $\lambda_i^U(A) = \Lambda_{ii}$, $i = 1, \ldots, m$. We omit U when dealing with a diagonalization that places the eigenvalues of A in ascending order in Λ, namely, $\lambda_1(A) \le \cdots \le \lambda_m(A)$.

The Frobenius norm of a matrix $A \in \mathbb{S}^m$ is the norm associated with the companion inner product, that is,

$$\|A\| := \sqrt{\langle A, A\rangle} = \sqrt{\sum_{i,j=1}^{m} A_{i,j}^2} = \sqrt{\sum_{i=1}^{m} \lambda_i(A)^2}.$$

Given $A \in \mathbb{S}^m$, we denote by $[A]_+$ the projection of A onto $\mathbb{S}^m_+$. Namely, if

$$A = U\mathrm{diag}(\lambda_1^U(A), \ldots, \lambda_m^U(A))U^T$$

is an orthogonal decomposition of A, then

$$[A]_+ := U\mathrm{diag}([\lambda_1^U(A)]_+, \ldots, [\lambda_m^U(A)]_+)U^T,$$

where $[v]_+ := \max\{0, v\}$, $v \in \mathbb{R}$. It is clear that $[A]_+$ does not depend on the choice of U. The following lemma provides an upper bound on the inner product of two matrices in terms of the inner product of their ordered eigenvalues. The subsequent lemma provides additional information when the matrices are positive/negative semidefinite.

Lemma 1 (von Neumann–Theobald [57]) *Let $A, B \in \mathbb{S}^m$ be given. Then,*

$$\langle A, B\rangle \leq \sum_{i=1}^{m} \lambda_i(A)\lambda_i(B),$$

where equality holds if and only if A and B are simultaneously diagonalizable with eigenvalues in ascending order.

Lemma 2 ([53]) *Let $A \in \mathbb{S}^m_+$ and $B \in \mathbb{S}^m_-$ be given. Then, the following conditions are equivalent:*

a) $\langle A, B\rangle = 0$,
b) $AB = 0$,
c) A and B are simultaneously diagonalizable by an orthogonal matrix U and $\lambda_i^U(A)\lambda_i^U(B) = 0$ for all $i = 1, \ldots, m$,
d) A and B are simultaneously diagonalizable in ascending order and $\lambda_i(A)\lambda_i(B) = 0$ for all $i = 1, \ldots, m$.

Note that with $A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$, $\langle A, B\rangle \neq 0$, but $\lambda_i(A)\lambda_i(B) = 0, i = 1, 2$, even though A and B are simultaneously diagonalizable by the identity matrix I. Note however that, as claimed by Lemma 2, it is not the case that $\lambda_i^I(A)\lambda_i^I(B) = 0$ for all $i = 1, 2$. The notation λ_i^U corrects the statement of this lemma in [53], with a similar proof, which we omit. Note also that these matrices are not simultaneously diagonalizable in ascending order.

Another useful result is the Weyl Lemma below. A lower and upper bound for the eigenvalues of the sum of two symmetric matrices are obtained, by means of the sum of the eigenvalues of these matrices.

Lemma 3 (Weyl [32]) *Let $A, B \in \mathbb{S}^m$ be given matrices. For $k = 1, \ldots, m$,*

$$\lambda_1(A) + \lambda_k(B) \leq \lambda_k(A + B) \leq \lambda_m(A) + \lambda_k(B).$$

We proceed to define the main concepts associated with problem (NLSDP). We define the Lagrangian function $L : \mathbb{R}^n \times \mathbb{S}^m \to \mathbb{R}$ associated with (NLSDP) by

$$L(x, \Omega) := f(x) + \langle G(x), \Omega\rangle.$$

The derivative of the mapping $G : \mathbb{R}^n \to \mathbb{S}^m$ at a point $x \in \mathbb{R}^n$ is given by $DG(x) : \mathbb{R}^n \to \mathbb{S}^m$, defined by

$$DG(x)h := \sum_{i=1}^{n} G_i(x)h_i, \ h \in \mathbb{R}^n,$$

where $G_i(x) := \frac{\partial G(x)}{\partial x_i} \in \mathbb{S}^m$, $i = 1, \ldots, m$, are the partial derivative matrices with respect to x_i. Also, we define the adjoint operator $DG(x)^* : \mathbb{S}^m \to \mathbb{R}^n$ by

$$DG(x)^*\Omega := (\langle G_1(x), \Omega\rangle, \ldots, \langle G_n(x), \Omega\rangle)^T, \ \Omega \in \mathbb{S}^m,$$

where it is easy to see that $\langle DG(x)h, \Omega\rangle = \langle h, DG(x)^*\Omega\rangle$ for all $h \in \mathbb{R}^n$, $x \in \mathbb{R}^n$ and $\Omega \in \mathbb{S}^m$.

Definition 1 We say that $\Omega \in \mathbb{S}^m$ is a Lagrange multiplier associated with $\bar{x} \in \mathbb{R}^n$ if the following Karush–Kuhn–Tucker (KKT) conditions hold:

$$\nabla_x L(\bar{x}, \Omega) = \nabla f(\bar{x}) + DG(\bar{x})^*\Omega = 0, \tag{1}$$

$$\langle G(\bar{x}), \Omega\rangle = 0, \tag{2}$$

$$G(\bar{x}) \in \mathbb{S}^m_-, \tag{3}$$

$$\Omega \in \mathbb{S}^m_+. \tag{4}$$

When there exists a Lagrange multiplier associated with $\bar{x}$, we say that $\bar{x}$ is a KKT point. Condition (2) is known as the complementarity condition, and it can be equivalently replaced by $\lambda_i^U(G(\bar{x}))\lambda_i^U(\Omega) = 0$ for all $i = 1, \ldots, m$, where $G(\bar{x})$ and Ω are simultaneously diagonalizable by U. See Lemma 2.

In order to prove the validity of the KKT conditions at a local minimizer of (NLSDP), a constraint qualification is needed. In general, more stringent assumptions are needed in order to prove global convergence results of an algorithm, that is, that every feasible accumulation point of a sequence generated by the algorithm is a KKT point. The most common assumptions of these kind are the nondegeneracy constraint qualification and the Mangasarian-Fromovitz constraint qualification (MFCQ) that we define below. More details can be found, e.g., in [19,51].

Definition 2 We say that the feasible point $\bar{x} \in \mathscr{F}$ satisfies the nondegeneracy constraint qualification if the n-dimensional vectors $(e_i^T G_1(\bar{x})e_j, \ldots, e_i^T G_n(\bar{x})e_j)$, $i, j = 1, \ldots, m - r$ are linearly independent, where $r = \operatorname{rank} G(\bar{x})$ and the vectors $e_1, \ldots, e_{m-r}$ form a basis of the null space of the matrix $G(\bar{x})$.

Definition 3 We say that the feasible point $\bar{x} \in \mathscr{F}$ satisfies the Mangasarian-Fromovitz constraint qualification if there is a vector $h \in \mathbb{R}^n$ such that $G(\bar{x}) + DG(\bar{x})h$ is negative definite.

As in the nonlinear programming case, the Mangasarian-Fromovitz constraint qualification is equivalent to the non-emptiness and boundedness of the set of

Lagrange multipliers. Similarly, the more stringent nondegeneracy condition (also called transversality condition), implies the uniqueness of the Lagrange multiplier, so this condition is treated in the literature as an analogous of the linear independence constraint qualification for nonlinear programming. This analogy is not perfect, as it does not reduce to the classical linear independence constraint qualification when the matrix $G(x)$ is diagonal. However, Lourenço, Fukuda and Fukushima [42], showed that if one reformulates the NLSDP such as an NLP using squared slack variables, where the conic constraint is replaced by the equality constraints $G(x) + Y^2 = 0$ with $Y \in \mathbb{S}^m$, if $(x, Y) \in \mathbb{R}^n \times \mathbb{S}^m$ satisfies the linear independence constraint qualification for this new NLP, then x satisfies the nondegeneracy condition. On the other hand, if x satisfies the nondegeneracy condition, then (x, Y) satisfies the linear independence constraint qualification for the NLP, where Y is the square root of the positive semidefinite matrix $-G(x)$. In this sense, the nondegeneracy condition is essentially the linear independence constraint qualification for NLSDP. Finally, we note that since $G(\bar{x})\Omega = 0$, rank $G(\bar{x})$+rank $\Omega \leq m$ always holds at a KKT point. When the equality holds, we say that $\bar{x}$ and Ω satisfy the strict complementarity.

In the survey paper [64], the two types of Augmented Lagrangian methods are discussed: one with quadratic penalization (by [62]) and another with a modified barrier function (by [53]), where both need a strict complementarity assumption for the local convergence results. In [63] and [65], the authors present an interior point method for NLSDP where for the convergence results, the first one adopts MFCQ while the second one assumes nondegeneracy and strict complementarity. In [20], the global convergence of a sequential SDP (SSDP), which is a method based on sequential quadratic programming (SQP) for NLP, has been proved under MFCQ. Another global convergent method presented in [28] applies a filter algorithm to SSDP where MFCQ is also assumed. In constrast with most of the literature, we will prove global convergence for an Augmented Lagrangian method without assuming strict complementarity and under a constraint qualification weaker than MFCQ, that we define in Sect. 4. In particular, our results allow the algorithm to generate an unbounded sequence of Lagrange multipliers, without hampering the global convergence results to a KKT point.

3 Sequential optimality condition for NLSDPs

In this section we will define the notion of an Approximate-KKT (AKKT) point for nonlinear semidefinite programming. This concept has been defined for nonlinear programming in [4,49]. Many first- and second-order global convergence proofs of algorithms have been done, under weak assumptions, based on this and similar notions [2,3,5–9,11,12,14–17,21,29–31,46,47,50,59].

The Approximate-KKT (AKKT) condition is the most natural sequential optimality condition in nonlinear programming. Most algorithms generate this type of sequence [1,6,11,14,45,49]. The following is a natural generalization of AKKT to the context of NLSDP.

Definition 4 We say that $\bar{x} \in \mathbb{R}^n$ satisfies the Approximate-KKT (AKKT) condition if $G(\bar{x}) \preceq 0$ and there exist sequences $x^k \to \bar{x}$ and $\{\Omega^k\} \subset \mathbb{S}^m_+$ such that

$$\lim_{k\to\infty} \nabla f(x^k) + DG(x^k)^* \Omega^k = 0, \tag{5}$$

$$\lambda_i^U(G(\bar{x})) < 0 \Rightarrow \lambda_i^{S_k}(\Omega^k) = 0, \text{ for all } i = 1, \ldots, m \text{ and sufficiently large } k, \tag{6}$$

where $G(\bar{x}) = U\mathrm{diag}(\lambda_1^U(G(\bar{x})), \ldots, \lambda_m^U(G(\bar{x})))U^T$, $\Omega^k = S_k\mathrm{diag}(\lambda_1^{S_k}(\Omega^k), \ldots, \lambda_m^{S_k}(\Omega^k))S_k^T$, where U and S_k are orthogonal matrices with $S_k \to U$.

All KKT points satisfy AKKT by considering constant sequences. We refer to Sect. 6 concerning the reciprocal implication.

Note that the definition of AKKT is independent of the choices of U and of the sequence $\{S_k\}$. Note also that given any orthogonal S_k that diagonalizes Ω^k, one can take a convergent subsequence. Hence, the definition above restricts the choice of $\{\Omega^k\}$, in such a way that the limit of its eigenvectors coincides with the eigenvectors of $G(\bar{x})$. Note however that the corresponding eigenvalues of $\{\Omega^k\}$ may be unbounded.

One of the reasons the extension of AKKT from NLP to NLSDP is not straightforward lies in the fact that in NLSDP the concept of active/inactive constraint is not apparent, which plays a key role in the definition of AKKT for NLP. In NLSDP, the eigenvalues less than zero play the role of the inactive constraints, hence, the "corresponding" Lagrange multiplier should vanish. However, it is not clear how to "correspond" an inactive eigenvalue of the constraint with an eigenvalue of the Lagrange multiplier matrix to vanish (in the definition of a KKT point, this correspondence is made by the assumption of simultaneous diagonalization). In this sense, the relation given by $S_k \to U$ provides the necessary notion for pairing the eigenvalues, which makes it a natural assumption for defining AKKT. In some sense, the definition says that the matrices must be approximately simultaneously diagonalizable. Note that when considering NLP as a particular case of an NLSDP with diagonal matrix, by considering a diagonal Lagrange multiplier matrix, both matrices are simultaneously diagonalizable by the identity, and we arrive at the usual AKKT concept for NLP.

When $\{x^k\}$ is a sequence as in the above definition, we say that $\{x^k\}$ is an AKKT sequence. Note that the sequence does not have to be formed by feasible points. The sequence $\{\Omega^k\}$ will be called the corresponding dual sequence.

Let us see that AKKT is closely related to a natural stopping criterion for numerical algorithms:

Lemma 4 *A point $\bar{x} \in \mathbb{R}^n$ satisfies AKKT if, and only if, there are sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\Omega^k\} \subset \mathbb{S}^m_+$, $\{\varepsilon_k\} \subset \mathbb{R}$ with $x^k \to \bar{x}$, $\varepsilon_k \to 0^+$ such that for all k,*

$$\|\nabla f(x^k) + DG(x^k)^* \Omega^k\| \leq \varepsilon_k, \tag{7}$$

$$\|[G(x^k)]_+\| \leq \varepsilon_k, \tag{8}$$

$$\textit{for all } i = 1, \ldots, m \textit{ and sufficiently large } k, \quad \lambda_i^{U_k}(G(x^k)) < -\varepsilon_k \Rightarrow \lambda_i^{S_k}(\Omega^k) = 0, \tag{9}$$

$$\|U_k - S_k\| \le \varepsilon_k, \tag{10}$$

where $G(x^k) = U_k diag(\lambda_1^{U_k}(G(x^k)), \ldots, \lambda_m^{U_k}(G(x^k)))U_k^T$, $\Omega^k = S_k diag(\lambda_1^{S_k}(\Omega^k), \ldots, \lambda_m^{S_k}(\Omega^k))S_k^T$ *are orthogonal diagonalizations of* $G(x^k)$ *and* Ω^k *for all* k.

Proof Suppose that $\bar{x} \in \mathbb{R}^n$ satisfies AKKT. Let us take diagonalizations U_k of $G(x^k)$, S_k of Ω^k and U of $G(\bar{x})$ in such a way that, for an appropriate subsequence, $U_k \to U$, $S_k \to U$ and (5) and (6) hold. Let us define the sequence $\{\varepsilon_k\} \subset \mathbb{R}$ in order to satisfy (7–10). Define

$$\varepsilon_k := \max\Big\{\|\nabla f(x^k) + DG(x^k)^*\Omega^k\|, \|[G(x^k)]_+\|, \|U_k - S_k\|, \\ -\lambda_i^{U_k}(G(x^k)) : i \in I(\bar{x})\Big\},$$

where $I(\bar{x})$ is the set of all $i \in \{1, \ldots, m\}$ such that $\lambda_i^U(G(\bar{x})) = 0$. Hence, (7), (8) and (10) hold. To prove (9), note that if $j_0 \in \{1, \ldots, m\}$ is such that $\lambda_{j_0}^{U_k}(G(x^k)) < -\varepsilon_k$, then $-\lambda_{j_0}^{U_k}(G(x^k)) > \varepsilon_k \ge -\lambda_j^{U_k}(G(x^k))$ for all $j \in I(\bar{x})$. In particular, $j_0 \notin I(\bar{x})$. Therefore, by the definition of AKKT, $\lambda_{j_0}^{S_k}(\Omega^k) = 0$. The definition of AKKT ensures also that $\varepsilon_k \to 0^+$.

Suppose now that conditions (7–10) are valid. Since $\{U_k\}$ and $\{S_k\}$ are bounded, by (10) we may take a subsequence and U, that diagonalizes $G(\bar{x})$, such that $U_k \to U$ and $S_k \to U$. The limit $\nabla f(x^k) + DG(x^k)^*\Omega^k \to 0$ follows trivially, while the continuity of the functions involved ensures that $\bar{x} \in \mathscr{F}$. Now, suppose that $\lambda_i^U(G(\bar{x})) < 0$. Then, since $\lambda_i^{U_k}(G(x^k)) \to \lambda_i^U(G(\bar{x}))$, for k large enough $\lambda_i^{U_k}(G(x^k)) < -\varepsilon_k$. Therefore, $\lambda_i^{S_k}(\Omega^k) = 0$ and AKKT is satisfied. □

Lemma 4 provides a natural stopping criterion associated with AKKT. Given small tolerances $\varepsilon_{\text{opt}} > 0$, $\varepsilon_{\text{feas}} > 0$, $\varepsilon_{\text{diag}} > 0$ and $\varepsilon_{\text{compl}} > 0$ associated with optimality, feasibility simultaneous diagonalization and complementarity, respectively, an algorithm that aims at solving NLSDP and generates an AKKT sequence $\{x^k\} \subset \mathbb{R}^n$ together with a dual sequence $\{\Omega^k\} \subset \mathbb{S}_+^m$ should be stopped at iteration k when

$$\|\nabla f(x^k) + DG(x^k)^*\Omega^k\| \le \varepsilon_{\text{opt}}, \tag{11}$$

$$\|[G(x^k)]_+\| \le \varepsilon_{\text{feas}}, \tag{12}$$

$$\|U_k - S_k\| \le \varepsilon_{\text{diag}}, \tag{13}$$

$$\lambda_i^{U_k}(G(x^k)) < -\varepsilon_{\text{compl}} \Rightarrow \lambda_i^{S_k}(\Omega^k) = 0. \tag{14}$$

As it is usual in the literature of sequential optimality conditions, three properties must be satisfied: (i) It must be a genuine necessary optimality condition, independently of the fulfillment of constraint qualifications. This gives a meaning with respect to optimality to the stopping criterion (11–14), independently of constraint qualifications. (ii) The condition must be satisfied by limit points of sequences generated by relevant algorithms. Note that this simplifies the usual global convergence proofs to

KKT points, also providing a guide to how new algorithms could be proposed. (iii) It must be a strong optimality condition. This is shown by defining a weak constraint qualification that makes every point that satisfies the condition a true KKT point.

We are going to show that AKKT satisfies all these requirement. In the remainder of this section we show that it satisfies the first requirement, that is, it is indeed a necessary optimality condition. For this, let us first consider the external penalty algorithm for the problem:

$$\text{Minimize } f(x), \text{ subject to } G(x) \preceq 0, \; x \in \Sigma, \tag{15}$$

where $\Sigma \subseteq \mathbb{R}^n$ is a non-empty closed set.

Theorem 1 *Choose a sequence $\{\rho_k\} \subset \mathbb{R}$ with $\rho_k \to +\infty$ and for each k, let x^k be the global solution, if it exists, of the problem*

$$\begin{aligned} \textit{Minimize } & f(x) + \frac{\rho_k}{2} P(x), \\ & \textit{subject to } x \in \Sigma, \end{aligned}$$

where $P(x) := \operatorname{tr}\left([G(x)]_+^2\right) = \|\,[G(x)]_+\,\|^2$ is the penalty function. Then, any limit point of this sequence, if any exist, is a solution of problem (15) provided that its feasible region is non-empty.

Proof See [24]. □

The penalization function $x \mapsto P(x)$ is a measure of infeasibility of x, as $P(x) \geq 0$ is continuous, and x is feasible if, and only if, $P(x) = 0$. We will also need the following result about the derivative of $P(x)$:

Lemma 5 *Let $G : \mathbb{R}^n \to \mathbb{S}^m$ be a differentiable function and $P(x) := \operatorname{tr}\left([G(x)]_+^2\right)$. Then, the gradient of P at $x \in \mathbb{R}^n$ is given by*

$$\nabla P(x) = 2DG(x)^*[G(x)]_+.$$

Proof It follows from [41, Corollary 3.2]. □

We are now ready to prove the main result of this section.

Theorem 2 *Let $\bar{x}$ be a local minimizer of (NLSDP). Then, $\bar{x}$ satisfies AKKT.*

Proof To show that AKKT is an optimality condition we will apply the external penalty algorithm to the following problem

$$\text{Minimize } f(x) + \frac{1}{2}\|x - \bar{x}\|^2 \text{ subject to } G(x) \preceq 0, \; x \in B(\bar{x}, \delta), \tag{16}$$

where $\delta > 0$ is small enough and $B(\bar{x}, \delta)$ denotes the closed ball of center $\bar{x}$ and radius δ. Clearly, δ can be chosen such that $\bar{x}$ is the unique solution of (16). Let $\rho_k \to +\infty$ and for each k, let x^k be a solution of

$$\text{Minimize } f(x) + \frac{1}{2}\|x - \bar{x}\|^2 + \frac{\rho_k}{2} \operatorname{tr}\left([G(x)]_+^2\right) \text{ subject to } x \in B(\bar{x}, \delta). \quad (17)$$

By the non-emptiness and compactness of $B(\bar{x}, \delta)$, and continuity of the objective function, x^k is well defined for all $k \in \mathbb{N}$. By the boundedness of $\{x^k\}$, the uniqueness of the solution of problem (16), and Theorem 1, $x^k \to \bar{x}$. Furthermore, note that x^k is in the interior of $B(\bar{x}, \delta)$ for k large enough; hence, the gradient of $f(x) + \frac{1}{2}\|x - \bar{x}\|^2 + \frac{\rho_k}{2} P(x)$ vanishes at x^k, that is, from Lemma 5, we have

$$\nabla f(x^k) + x^k - \bar{x} + \rho_k DG(x^k)^* \left[G(x^k)\right]_+ = 0.$$

Defining $\Omega^k := \rho_k \left[G(x^k)\right]_+$, and taking the limit when k tends to infinity in the above equality we obtain

$$\lim_{k\to\infty} \nabla f(x^k) + DG(x^k)^* \Omega^k = 0.$$

Since $G(x^k)$, Ω^k, and $G(\bar{x})$ are symmetric matrices we can diagonalize

$$G(x^k) = U_k \operatorname{diag}(\lambda_1^{U_k}(G(x^k)), \ldots, \lambda_m^{U_k}(G(x^k)))U_k^T,$$
$$\Omega^k = \rho_k \left[G(x^k)\right]_+ = U_k \operatorname{diag}(\rho_k[\lambda_1^{U_k}(G(x^k))]_+, \ldots, \rho_k[\lambda_m^{U_k}(G(x^k))]_+)U_k^T,$$

and

$$G(\bar{x}) = U \operatorname{diag}(\lambda_1^U(G(\bar{x})), \ldots, \lambda_m^U(G(\bar{x})))U^T,$$

where $U_k U_k^T = I$, $UU^T = I$ and $U_k \to U$. By the definition of Ω^k, the matrices $G(x^k)$ and Ω^k are simultaneously diagonalizable. Thus, the matrix S_k in the definition of AKKT is taken coinciding with U_k. Now, if $\lambda_i^U(G(\bar{x})) < 0$ then $\lambda_i^{U_k}(G(x^k)) < 0$ for all sufficiently large k; which implies that $\lambda_i^{U_k}(\Omega^k) = \rho_k[\lambda_i^{U_k}(G(x^k))]_+ = 0$. □

In the next example, we consider a local minimizer that does not satisfy the KKT conditions. Let us then build an AKKT sequence guaranteed to exist by Theorem 2.

Example 1 (AKKT sequence at a non-KKT solution) Consider the following problem

$$\begin{aligned} &\underset{x\in\mathbb{R}}{\text{Minimize}} \quad 2x, \\ &\text{Subject to} \quad G(x) := \begin{bmatrix} 0 & x \\ x & -1 \end{bmatrix} \preceq 0. \end{aligned}$$

Since the unique feasible point is $\bar{x} := 0$, this is the unique global minimizer.

(i) $\bar{x} = 0$ is not a KKT point. Let $\Omega := \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{12} & \mu_{22} \end{bmatrix}$. We have

$$\nabla f(\bar{x}) + DG(\bar{x})^* \Omega = 2 + 2\mu_{12} = 0 \Rightarrow \mu_{12} = -1,$$

and

$$\langle G(\bar{x}), \Omega \rangle = 0 \Rightarrow \mu_{22} = 0.$$

Thus, $\Omega = \begin{bmatrix} \mu_{11} & -1 \\ -1 & 0 \end{bmatrix}$, which is an indefinite matrix, regardless of μ_{11}. Therefore, $\bar{x}$ can not be a KKT point.

(ii) $\bar{x} = 0$ is an AKKT point. Define $x^k := -\dfrac{k}{k^2-1}$ and $\Omega^k := \begin{bmatrix} k & -1 \\ -1 & 1/k \end{bmatrix}$. Then,

$$\nabla f(x^k) + DG(x^k)^* \Omega^k = 2 - 2 = 0.$$

Note that, in the way that we have defined x^k and Ω^k, we have that $\Omega^k G(x^k) = G(x^k)\Omega^k$, which is a well known necessary and sufficient condition for simultaneous diagonalization. Computing the eigenvectors, $G(x^k)$ and Ω^k share a common basis of eigenvectors given by $u_1^k := \left(\frac{1}{\sqrt{1+k^2}}, \frac{k}{\sqrt{1+k^2}}\right)$ and $u_2^k := \left(\frac{-k}{\sqrt{1+k^2}}, \frac{1}{\sqrt{1+k^2}}\right)$, with corresponding eigenvalues

$$\lambda_1^{U_k}(G(x^k)) = \frac{-k^2}{k^2-1}, \lambda_2^{U_k}(G(x^k)) = \frac{1}{k^2-1}, \lambda_1^{S_k}(\Omega^k) = 0, \lambda_2^{S_k}(\Omega^k) = \frac{1}{k} + k,$$

where $U_k = S_k$ is the matrix with first and second columns given by u_1^k and u_2^k, respectively. Let $U := \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ be the limit of $\{S_k\}$. Now, since $\lambda_1^U(G(\bar{x})) = -1 < 0$ and $\lambda_2^U(G(\bar{x})) = 0$, AKKT holds since the eigenvalue of Ω^k corresponding to the negative eigenvalue is $\lambda_1^{S_k}(\Omega^k)$, which is zero for all k.

4 An augmented Lagrangian algorithm that generates AKKT sequences

In this section we will present an Augmented Lagrangian method for NLSDP based on the quadratic penalty function, which we will prove to generate an AKKT sequence. This is one of the most popular methods for solving NLSDPs [55]. For more details, see [1,33,36,38,56,62,64]. Our algorithm is inspired by the Augmented Lagrangian method with safeguards for nonlinear programming, introduced in [1,2] (see also [14]), which turns out to be the same as the one from [62] for NLSDP. Let us discuss in some details the contributions of [62], which is most related to our approach. In [62], the authors studied global convergence properties of four Augmented Lagrangian methods using different algorithmic strategies. For their study, the following NLSDP is considered:

$$\begin{aligned} &\underset{x\in\mathbb{R}^n}{\text{Minimize}} && f(x), \\ &\text{subject to} && G(x) \preceq 0, h(x) = 0, x \in \mathscr{V}, \end{aligned} \tag{18}$$

where $\mathscr{V} := \{x \in \mathbb{R}^n \mid g(x) \leq 0\}$ and $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^p$, $g : \mathbb{R}^n \to \mathbb{R}^q$ and $G : \mathbb{R}^n \to \mathbb{S}^m$ are continuously differentiable functions. Note that this problem differs from ours by the presence of additional standard nonlinear constraints. For (18), an Augmented Lagrangian method is presented using the safeguarded technique from [14]. The constraints $G(x) \preceq 0$ and $h(x) = 0$ are penalized (upper-level constraints) with the use of the Powell–Hestenes–Rockafellar (PHR) Augmented Lagrangian function for NLSDP, while the constraints in $\mathscr{V}$ are kept in the subproblems (lower-level constraints). The proposed algorithm then finds an approximate minimizer of the Augmented Lagrangian function over the set $\mathscr{V}$, followed by standard updates of the penalty parameter and safeguarded multipliers approximations. They then proceed to investigate global convergence results in a similar fashion as in [2,7]. In one of their main results, the authors show that a limit point is either feasible or, if the constant positive linear dependence (CPLD) constraint qualification with respect to $\mathscr{V}$ holds, then it is a KKT point for an infeasibility problem with respect to the upper-level constraints. They then resort to MFCQ in order to prove that a feasible limit point is a KKT point of (18). That is, they only consider a weak constraint qualification (CPLD) for the standard nonlinear programming constraints in $\mathscr{V}$, but not for the conic constraints $G(x) \preceq 0$. We will provide global convergence results for this Augmented Lagrangian method for (NLSDP), which can be formulated with a new constraint qualification weaker than MFCQ. That is, for simplicity, we will consider problems only with conic constraints and no additional nonlinear constraints. Any additional nonlinear constraints could also be penalized or incorporated in the conic constraints in a standard way. Although we consider only unconstrained Augmented Lagrangian subproblems, the theory we develop is general enough in order to consider even subproblems with conic constraints, with an approach similar to what is done in [1,16]. We leave the details of this more general approach for a later study, while we now focus on an Augmented Lagrangian method that penalizes all constraints and solves unconstrained subproblems.

Given a penalty parameter $\rho > 0$, the Augmented Lagrangian function $L_\rho : \mathbb{R}^n \times \mathbb{S}^m \to \mathbb{R}$, associated with (NLSDP), is defined by:

$$L_\rho(x, \Omega) := f(x) + \frac{1}{2\rho}\left\{\|[\Omega + \rho G(x)]_+\|^2 - \|\Omega\|^2\right\}. \tag{19}$$

This function is a natural extension of the Augmented Lagrangian function for nonlinear programming.

Similarly to the external penalty method, our goal is to solve (NLSDP) by solving a sequence of unconstrained minimization problems with respect to x, obtaining at each iteration k a primal iterate x^k, where the objective function is given by (19), and $\Omega := \bar{\Omega}^k$ and $\rho := \rho_k$ are iteratively updated according to some performance criterion. We expect that, for suitable choices of the parameters, the sequence of solutions generated will converge to the solution of problem (NLSDP). From the definition of the Augmented Lagrangian function above, we have that its gradient, with respect to x, is given by:

$$\nabla_x L_\rho(x, \Omega) = \nabla f(x) + DG(x)^* [\Omega + \rho G(x)]_+ .$$

Therefore, a natural update rule for the dual sequence is:

$$\Omega \leftarrow [\Omega + \rho G(x)]_+ ,$$

however, to avoid solving the next subproblem with a Lagrange multiplier approximation too large (given that we do not assume MFCQ, and Lagrange multipliers may be unbounded), we take the Lagrange multiplier approximation for the next subproblem as the projection of the natural update above onto a safeguarded box as in [14]. This means that when the penalty parameter is large, the method reduces to the external penalty method. To update the penalty parameter, we will rely on a joint measure of feasibility and complementarity. The formal statement of the algorithm is as follows:

Algorithm 1 Augmented Lagrangian Algorithm

STEP 0 (Initialization): Let $\tau \in (0, 1)$, $\gamma > 1$, $\rho_1 > 0$ and $\Omega^{\max} \in \mathbb{S}^m_+$. Take $\{\varepsilon_k\} \subset \mathbb{R}$ a sequence of positive scalars such that $\lim \varepsilon_k = 0$. Define $0 \preceq \bar{\Omega}^1 \preceq \Omega^{\max}$. Choose $x^0 \in \mathbb{R}^n$ an arbitrary starting point. Initialize $k := 1$.

STEP 1 (Solve subproblem): Use x^{k-1} to find an approximate minimizer x^k of $L_{\rho_k}(x, \bar{\Omega}^k)$, that is, a point x^k such that

$$\|\nabla_x L_{\rho_k}(x^k, \bar{\Omega}^k)\| \leq \varepsilon_k.$$

STEP 2 (Penalty parameter update): Define

$$V^k := \left[\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right]_+ - \frac{\bar{\Omega}^k}{\rho_k}.$$

If

$$\|V^k\| \leq \tau \|V^{k-1}\|,$$

define

$$\rho_{k+1} := \rho_k,$$

otherwise, define

$$\rho_{k+1} := \gamma \rho_k.$$

STEP 3 (Multiplier update): Compute

$$\Omega^k := \left[\bar{\Omega}^k + \rho_k G(x^k)\right]_+ ,$$

and define $\bar{\Omega}^{k+1} := proj_S(\Omega^k)$, the orthogonal projection of Ω^k onto S, where $S := \{X \in \mathbb{S}^m | 0 \preceq X \preceq \Omega^{\max}\}$. Set $k := k + 1$, and go to Step 1.

Note that $V^k = 0$ if, and only if x^k is feasible and complementarity holds. Indeed,

$$V^k = 0 \Rightarrow \left[\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right]_+ = \frac{\bar{\Omega}^k}{\rho_k} \Rightarrow \bar{\Omega}^k = \left[\bar{\Omega}^k + \rho_k G(x^k)\right]_+,$$

and then, computing the orthogonal decomposition, we have

$$\bar{\Omega}^k + \rho_k G(x^k) = U_k \mathrm{diag}(\lambda_1^{U_k}, \ldots, \lambda_m^{U_k})U_k^T \text{ and}$$
$$\bar{\Omega}^k = [\bar{\Omega}^k + \rho_k G(x^k)]_+ = U_k \mathrm{diag}([\lambda_1^{U_k}]_+, \ldots, [\lambda_m^{U_k}]_+)U_k^T.$$

Hence,

$$G(x^k) = (1/\rho_k)U_k \mathrm{diag}(\lambda_1^{U_k} - [\lambda_1^{U_k}]_+, \ldots, \lambda_m^{U_k} - [\lambda_m^{U_k}]_+)U_k^T.$$

As $\lambda_i^{U_k} - [\lambda_i^{U_k}]_+ \leq 0$ for all $i = 1, \ldots, m$, we have $G(x^k) \preceq 0$. Also, the multiplier $\bar{\Omega}^k$ is such that, if $\lambda_i^{U_k}(G(x^k)) < 0$ then $\lambda_i^{U_k} < [\lambda_i^{U_k}]_+$ and thus, $\lambda_i^{U_k}(\bar{\Omega}^k) = [\lambda_i^{U_k}]_+ = \max\{0, \lambda_i^{U_k}\} = 0$. The reciprocal implication follows similarly. Thus, the algorithm keeps the previous penalty parameter unchanged if $\|V^k\|$ is sufficiently reduced, otherwise, the penalty parameter is increased to force feasibility and complementarity.

Note also that we consider $\{\Omega^k\}$ as the dual sequence generated by the algorithm in order to check a stopping criterion, since this sequence together with $\{x^k\}$ fulfills the stopping criterion related to optimality (11) when $\varepsilon_k \leq \varepsilon_{\text{opt}}$.

The following result shows that Algorithm 1 finds stationary points of an infeasibility measure. This shows that the algorithm tends to find feasible points, which are global minimizers of the infeasibility measure, whenever the feasible region is non-empty.

Theorem 3 *Let $\bar{x} \in \mathbb{R}^n$ be a limit point of a sequence $\{x^k\}$ generate by Algorithm 1. Then, $\bar{x}$ is a stationary point for the optimization problem*

$$\underset{x \in \mathbb{R}^n}{\mathit{Minimize}} \quad P(x) := \mathrm{tr}\left([G(x)]_+^2\right). \tag{20}$$

Proof If $\{\rho_k\}$ is bounded, that is, for $k \geq k_0$ the penalty parameter remains unchanged, we have that $V^k \to 0$. The sequence $\{\bar{\Omega}^k\}$ is bounded, then, there is an infinite subset $K_1 \underset{\infty}{\subset} \mathbb{N}$ such that $\lim_{k \in K_1} \bar{\Omega}^k = \bar{\Omega}$. Now, from $V^k \to 0$ we get

$$\bar{\Omega} = \lim_{k \in K_1} \bar{\Omega}^k = \lim_{k \in K_1} \left[\bar{\Omega}^k + \rho_{k_0} G(x^k)\right]_+ = \lim_{k \in K_1} \Omega^k = \left[\bar{\Omega} + \rho_{k_0} G(\bar{x})\right]_+.$$

The computation is now similar to our previous discussion where V^k was assumed to be zero. Writing the orthogonal decomposition of the matrix $\bar{\Omega}$, we have

$$\bar{\Omega} = [\bar{\Omega} + \rho_{k_0} G(\bar{x})]_+ = U \mathrm{diag}([\lambda_1^U]_+, \ldots, [\lambda_m^U]_+)U^T,$$

with $UU^T = I$. Moreover,

$$\bar{\Omega} + \rho_{k_0} G(\bar{x}) = U\text{diag}(\lambda_1^U, \ldots, \lambda_m^U)U^T.$$

In this way,

$$G(\bar{x}) = (1/\rho_{k_0})U\text{diag}((\lambda_1^U - [\lambda_1^U]_+), \ldots, (\lambda_m^U - [\lambda_m^U]_+))U^T,$$

thus, $G(\bar{x}) \preceq 0$. Therefore, $\bar{x}$ is a global minimizer of the optimization problem (20). If $\{\rho_k\}$ is unbounded, let us define

$$\delta^k := \nabla f(x^k) + DG(x^k)^* \Omega^k,$$

where $\Omega^k := [\bar{\Omega}^k + \rho_k G(x^k)]_+$. Clearly, from Step 1 of the algorithm, we have that $\|\delta^k\| \leq \varepsilon_k$. Dividing δ^k by ρ_k,

$$\frac{\delta^k}{\rho_k} = \frac{\nabla f(x^k)}{\rho_k} + DG(x^k)^* \left[\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right]_+,$$

since $\bar{\Omega}^k$ is bounded and $\nabla f(x^k) \to \nabla f(\bar{x})$, we have that $DG(\bar{x})^*[G(\bar{x})]_+ = 0$. Thus, the result follows from Lemma 5.

□

Next, we will show that a feasible limit point of a sequence $\{x^k\}$ generated by the algorithm is an AKKT point. In fact, $\{x^k\}$ is an associated AKKT sequence and $\{\Omega^k\}$ is the corresponding dual sequence.

Theorem 4 *Assume that $\bar{x} \in \mathbb{R}^n$ is a feasible limit point of a sequence $\{x^k\}$ generated by Algorithm 1. Then, $\bar{x}$ is an AKKT point.*

Proof Let $\bar{x} \in \mathscr{F}$ be a limit point of a sequence $\{x^k\}$ generated by Algorithm 1. Let us assume without loss of generality that $x^k \to \bar{x}$. From Step 1 of the algorithm we have that

$$\|\nabla f(x^k) + DG(x^k)^* \Omega^k\| \leq \varepsilon_k \Rightarrow \lim_{k\to\infty} \nabla f(x^k) + DG(x^k)^* \Omega^k = 0,$$

where

$$\Omega^k = \left[\bar{\Omega}^k + \rho_k G(x^k)\right]_+.$$

Now, let us prove that for appropriate matrices U and $S_k \to U$, we have that for all $i = 1, \ldots, m$, if $\lambda_i^U(G(\bar{x})) < 0$ then $\lambda_i^{S_k}(\Omega^k) = 0$ for all sufficiently large k. We have two cases to analyze:

(i) If $\rho_k \to +\infty$, since the sequence $\{\bar{\Omega}^k\}$ is bounded we have that $\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) \to G(\bar{x})$. Let us take a diagonalization

$$\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) = S_k \mathrm{diag}(\lambda_1^k, \ldots, \lambda_m^k) S_k^T,$$

where $S_k S_k^T = I$. Taking a subsequence if necessary, let us take a diagonalization

$$G(\bar{x}) = U \mathrm{diag}(\lambda_1, \ldots, \lambda_m) U^T,$$

where $UU^T = I$ with $S_k \to U$, $\lambda_i^k \to \lambda_i$ for all i. Then,

$$\Omega^k = [\bar{\Omega}^k + \rho_k G(x^k)]_+ = S_k \mathrm{diag}(\rho_k[\lambda_1^k]_+, \ldots, \rho_k[\lambda_m^k]_+) S_k^T.$$

Now, assume that $\lambda_i^U(G(\bar{x})) := \lambda_i < 0$. Then, $\lambda_i^k < 0$ for all sufficiently large k; which implies that $\lambda_i^{S_k}(\Omega^k) := \rho_k[\lambda_i^k]_+ = 0$.

(ii) If $\{\rho_k\}$ is bounded, similarly to what was done in Theorem 3, for $k \geq k_0$ we have $\rho_k = \rho_{k_0}$; hence, $V^k \to 0$. Thus, taking a subsequence if necessary, we have

$$\bar{\Omega} = \lim\ \bar{\Omega}^k = \lim \left[\bar{\Omega}^k + \rho_{k_0} G(x^k)\right]_+ = \lim\ \Omega^k = \left[\bar{\Omega} + \rho_{k_0} G(\bar{x})\right]_+.$$

Let us take an orthogonal decomposition of the matrix $\bar{\Omega}^k + \rho_{k_0} G(x^k)$, that is,

$$\bar{\Omega}^k + \rho_{k_0} G(x^k) = S_k \mathrm{diag}(\lambda_1^{S_k}, \ldots, \lambda_m^{S_k}) S_k^T,$$

and let us take a subsequence such that $\{S_k\}$ converges to some orthogonal matrix U. Then,

$$\bar{\Omega} + \rho_{k_0} G(\bar{x}) = \lim\ \bar{\Omega}^k + \rho_{k_0} G(x^k) = U \mathrm{diag}(\lambda_1^U, \ldots, \lambda_m^U) U^T,$$

where $\lambda_i^{S_k} \to \lambda_i^U$ for all i. In this way, since

$$\bar{\Omega} = U \mathrm{diag}([\lambda_1^U]_+, \ldots, [\lambda_m^U]_+) U^T,$$

we have that

$$G(\bar{x}) = (1/\rho_{k_0}) U \mathrm{diag}((\lambda_1^U - [\lambda_1^U]_+), \ldots, (\lambda_m^U - [\lambda_m^U]_+)) U^T,$$

and then, $\lambda_i^U(G(\bar{x})) = \dfrac{\lambda_i^U - [\lambda_i^U]_+}{\rho_{k_0}}$. Assuming that it is negative, we have that $\lambda_i^U < [\lambda_i^U]_+$; hence, $\lambda_i^U < 0$. Then, $\lambda_i^{S_k} < 0$ for all sufficiently large k; which implies that $\lambda_i^{S_k}(\Omega^k) = [\lambda_i^{S_k}]_+ = 0$.

□

5 A new sequential optimality condition for NLSDP

In Sect. 3, we presented an extension of the classical AKKT sequential optimality condition known for NLP. However, for NLSDP, it turns out that we can define a much more natural and simpler sequential optimality condition, which is new even in the context of NLP. The new condition does not rely on eigenvalue computations for treating the complementarity, which are replaced by a simpler inner product of the constraint matrix and the dual matrix. The new condition also does not require the convergence $S_k \to U$, which relates the eigenvectors of the dual sequence with the ones of the limit of the constraint matrix. This new condition is called Trace-Approximate-KKT (TAKKT) and is defined below.

Definition 5 We say that a point $\bar{x} \in \mathbb{R}^n$ satisfies the Trace-Approximate-KKT (TAKKT) condition if $G(\bar{x}) \preceq 0$ and there exist sequences $x^k \to \bar{x}$ and $\{\Omega^k\} \subset \mathbb{S}^m_+$ such that

$$\lim_{k\to\infty} \nabla f(x^k) + DG(x^k)^* \Omega^k = 0, \tag{21}$$

$$\lim_{k\to\infty} \langle \Omega^k, G(x^k)\rangle = 0. \tag{22}$$

Remark 1 Note that, $\lim_{k\to\infty}\langle \Omega^k, G(x^k)\rangle = 0$ does not imply that Ω^k and $G(x^k)$ are (approximately) simultaneously diagonalizable. Indeed, consider the matrices

$$G(x^k) := \begin{bmatrix} 1/k & 0 \\ 0 & -1/k \end{bmatrix} \in \mathbb{S}^2 \text{ and } \Omega^k := \begin{bmatrix} k & k \\ k & k \end{bmatrix} \in \mathbb{S}^2_+.$$

Then, $\lim_{k\to\infty}\langle \Omega^k, G(x^k)\rangle = 0$ but these matrices are not simultaneously diagonalizable since $\Omega^k G(x^k) \neq G(x^k)\Omega^k$. More precisely, let us consider the orthogonal diagonalizations

$$\begin{aligned} \Omega^k &= S_k \mathrm{diag}(\lambda_1^{S_k}(\Omega^k), \lambda_2^{S_k}(\Omega^k)) S_k^T \\ &= \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 2k \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \end{aligned}$$

and

$$G(x^k) = U_k \mathrm{diag}(\lambda_1^{U_k}(G(x^k)), \lambda_2^{U_k}(G(x^k))) U_k^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/k & 0 \\ 0 & -1/k \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Thus, no matter how the eigenvalues are ordered, we always have that $\lambda_i^{U_k}(G(x^k))\lambda_i^{S_k}(\Omega^k) \not\to 0$ for some i.

In this section, we will show that TAKKT is indeed an optimality condition and that the Augmented Lagrangian algorithm generates TAKKT sequences under the generalized Lojasiewicz inequality. In some sense, TAKKT plays the role of the

Complementarity-AKKT (CAKKT) condition known for NLP [10] in the global convergence analysis of NLSDP algorithms, under a very natural stopping criterion for NLSDP, which is based on the simple lemma below. More remarks about the relationship with CAKKT for NLP will follow at the end of this section.

Lemma 6 *The point $\bar{x} \in \mathbb{R}^n$ satisfies TAKKT if, and only if, there exist sequences $x^k \to \bar{x}$, $\{\Omega^k\} \subset \mathbb{S}^m_+$, and $\{\varepsilon_k\} \subset \mathbb{R}_+$ such that $x^k \to \bar{x}$, $\varepsilon_k \to 0^+$ and for all $k \in \mathbb{N}$,*

$$\|\nabla f(x^k) + DG(x^k)^* \Omega^k\| \leq \varepsilon_k, \tag{23}$$

$$\|[G(x^k)]_+\| \leq \varepsilon_k, \tag{24}$$

$$|\langle \Omega^k, G(x^k)\rangle| \leq \varepsilon_k. \tag{25}$$

To prove that TAKKT is an optimality condition for NLSDP, it is sufficient to note that in the proof of Theorem 2, where we proved that AKKT is an optimality condition, the dual sequence is defined as $\Omega^k := \rho_k[G(x^k)]_+$. We also note that from the definition of $\{x^k\}$ in the proof of Theorem 2 as the global solution of problem (17), the additional property holds:

$$f(x^k) - f(\bar{x}) + \frac{1}{2}\|x^k - \bar{x}\|^2 + \frac{\rho_k}{2} \operatorname{tr}\left(\left[G(x^k)\right]_+^2\right) \leq 0. \tag{26}$$

Let us see that these observations are sufficient to prove that TAKKT is an optimality condition.

Theorem 5 *Let $\bar{x}$ be a local minimizer of (NLSDP). Then $\bar{x}$ satisfies TAKKT.*

Proof Let $\{x^k\}$ be the AKKT sequence defined in the proof of Theorem 2. In particular, the dual sequence is defined for all k as $\Omega^k := \rho_k[G(x^k)]_+$ and (26) holds with some $\rho_k > 0$. It remains to prove that $\langle G(x^k), \Omega^k\rangle \to 0$.

Taking the limit in (26) we have that

$$\langle [G(x^k)]_+, \Omega^k\rangle = \sum_{i=1}^m \rho_k[\lambda_i(G(x^k))]_+^2 = \rho_k \operatorname{tr}\left(\left[G(x^k)\right]_+^2\right) \to 0.$$

Since

$$\lambda_i(G(x^k))[\lambda_i(G(x^k))]_+ = [\lambda_i(G(x^k))]_+^2 \text{ for all } i = 1, \ldots, m,$$

it follows that $\langle G(x^k), \Omega^k\rangle = \sum_{i=1}^m \rho_k \lambda_i(G(x^k))[\lambda_i(G(x^k))]_+ \to 0$. Therefore, $\bar{x}$ is a TAKKT point. □

Note that in Example 1, where a local minimizer that is not a KKT point is presented, the same sequence used to verify that AKKT holds can be used to check that TAKKT holds. Let us see an additional example from [43, Example 6.3.6] that illustrates the same thing, but where there is a non-zero duality gap.

Example 2 (TAKKT sequence at a non-KKT solution) Consider the following problem

$$
\begin{aligned}
&\underset{x\in\mathbb{R}^2}{\text{Minimize}} \quad x_1,\\
&\text{Subject to} \quad G(x) := \begin{bmatrix} 0 & -x_1 & 0 \\ -x_1 & -x_2 & 0 \\ 0 & 0 & -1-x_1 \end{bmatrix} \preceq 0.
\end{aligned}
$$

The point $\bar{x} := (0, 0)$ is a global minimizer, with an optimal value of zero. The dual problem can be stated as

$$
\begin{aligned}
&\underset{\Omega\in\mathbb{S}^3}{\text{Maximize}} \quad -\mu_{33},\\
&\text{Subject to} \quad \Omega := \begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} \\ \mu_{12} & \mu_{22} & \mu_{23} \\ \mu_{13} & \mu_{23} & \mu_{33} \end{bmatrix} \succeq 0,\\
&\qquad\qquad \mu_{33} + 2\mu_{12} = 1,\\
&\qquad\qquad \mu_{22} = 0,
\end{aligned}
$$

where the conic constraint implies $\mu_{12} = 0$ and hence the optimal value is -1.

Now, it is easy to check that the KKT condition does not hold at $\bar{x}$, but TAKKT holds with the sequence $x^k := (\frac{1}{k}, \frac{1}{k})$ and $\Omega^k := \begin{bmatrix} (k-1)^2 & \frac{k-1}{2k} & 0 \\ \frac{k-1}{2k} & \frac{1}{4k^2} & 0 \\ 0 & 0 & \frac{1}{k} \end{bmatrix} \in \mathbb{S}^3_+$.

In the following example we can see that AKKT does not imply TAKKT.

Example 3 (AKKT does not imply TAKKT)

Consider the following problem

$$
\begin{aligned}
&\underset{x\in\mathbb{R}^2}{\text{Minimize}} \quad x_2,\\
&\text{Subject to} \quad G(x) := \begin{bmatrix} x_1 & 0 & 0 \\ 0 & -x_1 & 0 \\ 0 & 0 & x_1 - x_1^2 x_2 \end{bmatrix} \preceq 0.
\end{aligned}
$$

(i) The feasible point $\bar{x} := (0, 1)$ is an AKKT point. Define $x_1^k := 1/k$, $x_2^k := 1$ and $\Omega^k := \begin{bmatrix} \mu_{11}^k & 0 & 0 \\ 0 & \mu_{22}^k & 0 \\ 0 & 0 & \mu_{33}^k \end{bmatrix} \in \mathbb{S}^3_+$ with $\mu_{11}^k := 2k$, $\mu_{22}^k := k^2$, and $\mu_{33}^k := k^2$. Also, we have

$$
\nabla f(x^k) + DG(x^k)^*\Omega^k = (\mu_{11}^k - \mu_{22}^k + \mu_{33}^k - 2x_1^k x_2^k \mu_{33}^k, 1 - (x_1^k)^2\mu_{33}^k) = (0, 0).
$$

Note that $G(x^k)$ and Ω^k are simultaneously diagonalizable by the identity matrix. Since all eigenvalues of $G(\bar{x})$ are zero, there is nothing else to check.

(ii) TAKKT does not hold at $\bar{x}$. First, note that since $G(x^k)$ is diagonal we can take Ω^k a diagonal matrix without loss of generality. Now, suppose that there exist sequences $x^k \to \bar{x}$ and $\Omega^k := \begin{bmatrix} \mu_{11}^k & 0 & 0 \\ 0 & \mu_{22}^k & 0 \\ 0 & 0 & \mu_{33}^k \end{bmatrix} \in \mathbb{S}^3_+$ such that

$$\nabla f(x^k) + DG(x^k)^* \Omega^k = (\mu_{11}^k - \mu_{22}^k + \mu_{33}^k - 2x_1^k x_2^k \mu_{33}^k, 1 - (x_1^k)^2 \mu_{33}^k) \to 0,$$

and

$$\begin{aligned} \langle G(x^k), \Omega^k \rangle &= x_1^k(\mu_{11}^k - \mu_{22}^k + \mu_{33}^k) - (x_1^k)^2 x_2^k \mu_{33}^k \\ &= x_1^k(\mu_{11}^k - \mu_{22}^k + \mu_{33}^k - 2x_1^k x_2^k \mu_{33}^k) + (x_1^k)^2 x_2^k \mu_{33}^k \to 0. \end{aligned}$$

Note that, since $\mu_{11}^k - \mu_{22}^k + \mu_{33}^k - 2x_1^k x_2^k \mu_{33}^k \to 0$, we conclude that $(x_1^k)^2 x_2^k \mu_{33}^k \to 0$. Since $x_2^k \to 1$, this yields $(x_1^k)^2 \mu_{33}^k \to 0$, which constradicts $1 - (x_1^k)^2 \mu_{33}^k \to 0$. Thus, TAKKT does not hold at $\bar{x} = (0, 1)$.

Note that in the previous example, the point $\bar{x} = (0, 1)$ is an AKKT point that is not a solution. Hence, following our global convergence result given by Theorem 4, we can not rule out the possibility of the augmented Lagrangian algorithm converging to this undesirable point. However, let us show that indeed the algorithm avoids the point $\bar{x}$ in this example. For this, let us prove the new global convergence result of the algorithm, that is, that it generates TAKKT sequences. We will need the following smoothness assumption, which is known as the generalized Lojasiewicz inequality [18]:

Assumption 1 Every feasible limit point $\bar{x}$ of a sequence $\{x^k\}$ generated by Algorithm 1 satisfies the generalized Lojasiewicz inequality below:

There is $\delta > 0$ and a function $\varphi : B(\bar{x}, \delta) \to \mathbb{R}$, where $B(\bar{x}, \delta)$ is the closed ball centered at $\bar{x}$ and radius δ, with $\varphi(x) \to 0$ when $x \to \bar{x}$ such that for all $x \in B(\bar{x}, \delta)$,

$$|P(x) - P(\bar{x})| \leq \varphi(x) \|\nabla P(x)\|,$$

where $P(x) := \operatorname{tr}\left([G(x)]_+^2\right)$.

Note that, by Lemma 5, for all $x \in \mathbb{R}^n$ we have $\nabla P(x) = 2DG(x)^*[G(x)]_+$.

Assumption 1 is a natural extension of the smoothness assumption that is required in nonlinear programming to prove convergence of the Augmented Lagrangian method to CAKKT points [10]. Note that this is an assumption about the smoothness of the function $G(\cdot)$, which holds, for instance, when $G(\cdot)$ is an analytic function (see [10]). More importantly, this is not an assumption about how the function $G(\cdot)$ behaves in the conic constraint $G(x) \preceq 0$, such as in constraint qualification assumptions. We are now ready to prove the result.

Theorem 6 *Let Assumption 1 hold. Assume that $\bar{x} \in \mathbb{R}^n$ is a feasible limit point of a sequence $\{x^k\}$ generated by Algorithm 1. Then $\bar{x}$ is a TAKKT point.*

Proof By the proof of Theorem 4, it is sufficient to verify that $\lim_{k\to\infty}\langle \Omega^k, G(x^k)\rangle = 0$. Let us consider the following decomposition:

$$\frac{\bar{\Omega}^k}{\rho_k} + G(x^k) = S_k \mathrm{diag}(\lambda_1^k, \ldots, \lambda_m^k) S_k^T \Rightarrow \Omega^k \\ = S_k \mathrm{diag}(\rho_k[\lambda_1^k]_+, \ldots, \rho_k[\lambda_m^k]_+) S_k^T, \tag{27}$$

and

$$G(x^k) = U_k \mathrm{diag}(\lambda_1^{U_k}(G(x^k)), \ldots, \lambda_m^{U_k}(G(x^k))) U_k^T,$$

where, for all k, S_k and U_k are orthogonal matrices such that

$$\lambda_1^k \leq \lambda_2^k \leq \cdots \leq \lambda_m^k \text{ and } \lambda_1^{U_k}(G(x^k)) \leq \lambda_2^{U_k}(G(x^k)) \leq \cdots \leq \lambda_m^{U_k}(G(x^k)). \tag{28}$$

In addition, taking a subsequence if necessary, we have

$$G(\bar{x}) = U \mathrm{diag}(\lambda_1^U(G(\bar{x})), \ldots, \lambda_m^U(G(\bar{x}))) U^T,$$

where U is orthogonal and $\lambda_1^U(G(x^k)) \leq \lambda_2^U(G(\bar{x})) \leq \cdots \leq \lambda_m^U(G(\bar{x}))$ with $U_k \to U$.

(i) Assume that $\rho_k \to +\infty$. By Step 1 of the algorithm, the sequence $\{\nabla_x L_{\rho_k}(x^k, \Omega^k)\}$ is bounded. Also, as

$$\nabla_x L_{\rho_k}(x^k, \Omega^k) = \nabla f(x^k) + DG(x^k)^*[\bar{\Omega}^k + \rho_k G(x^k)]_+ \\ = \nabla f(x^k) + \rho_k DG(x^k)^* \left[\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right]_+,$$

and since $\dfrac{\bar{\Omega}^k}{\rho_k} \to 0$, there is some $M > 0$ such that

$$\rho_k \|\nabla P(x^k)\| = \left\| \rho_k DG(x^k)^*[G(x^k)]_+ \right\| \leq M$$

for all k. Taking the function φ given by Assumption 1 and using the fact that $P(\bar{x}) = 0$, we have for all k:

$$\left|\rho_k P(x^k)\right| \leq \varphi(x^k) \left\| \rho_k \nabla P(x^k) \right\| \leq M\varphi(x^k).$$

Taking the limit in k we have:

$$\lim_{k\to\infty} \rho_k P(x^k) = 0 \Rightarrow \lim_{k\to\infty} \rho_k \,\mathrm{tr}([G(x^k)]_+^2) = 0. \tag{29}$$

From (27), we get

$$G(x^k) = S_k \mathrm{diag}(\lambda_1^k, \ldots, \lambda_m^k) S_k^T - \frac{\bar{\Omega}^k}{\rho_k}.$$

Now,

$$\begin{aligned}\langle \Omega^k, G(x^k)\rangle &= \left\langle \Omega^k, S_k \mathrm{diag}(\lambda_1^k, \ldots, \lambda_m^k) S_k^T - \frac{\bar{\Omega}^k}{\rho_k} \right\rangle \\ &= \left\langle \Omega^k, S_k \mathrm{diag}(\lambda_1^k, \ldots, \lambda_m^k) S_k^T \right\rangle - \left\langle \Omega^k, \frac{\bar{\Omega}^k}{\rho_k} \right\rangle. \qquad (30)\end{aligned}$$

Let us show that $\langle \Omega^k, G(x^k)\rangle \to 0$ in an infinite subsequence $k \in K_1 \subset \mathbb{N}$. For this, let us show that

$$\left\langle \Omega^k, S_k \mathrm{diag}(\lambda_1^k, \ldots, \lambda_m^k) S_k^T \right\rangle \to 0 \text{ and } \left\langle \Omega^k, \frac{\bar{\Omega}^k}{\rho_k} \right\rangle \to 0.$$

Firstly,

$$\left\langle \Omega^k, \frac{\bar{\Omega}^k}{\rho_k} \right\rangle = \left\langle \left[\bar{\Omega}^k + \rho_k G(x^k)\right]_+, \frac{\bar{\Omega}^k}{\rho_k} \right\rangle = \left\langle \left[\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right]_+, \bar{\Omega}^k \right\rangle. \quad (31)$$

Thus, by the boundedness of $\{\bar{\Omega}^k\}$ and since $\rho_k \to +\infty$ and $\left[G(x^k)\right]_+ \to 0$, we have that $\left[\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right]_+ \to 0$; therefore, $\left\langle \Omega^k, \frac{\bar{\Omega}^k}{\rho_k} \right\rangle \to 0$. Furthermore,

$$\begin{aligned}\left\langle \Omega^k, S_k \mathrm{diag}(\lambda_1^k, \ldots, \lambda_m^k) S_k^T \right\rangle &= \sum_{i=1}^m \rho_k [\lambda_i^k]_+ \lambda_i^k \\ &= \sum_{i=1}^m \left[\lambda_i^{S_k}\left(\bar{\Omega}^k + \rho_k G(x^k)\right)\right]_+ \lambda_i^{S_k}\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right).\end{aligned}$$

Let us now show that $\left[\lambda_i^{S_k}\left(\bar{\Omega}^k + \rho_k G(x^k)\right)\right]_+ \lambda_i^{S_k}\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right) \to 0$ for all $i = 1, \ldots, m$. From the order given in (28), we can apply Lemma 3 to get

$$\lambda_1(\bar{\Omega}^k) + \rho_k \lambda_i^{U_k}(G(x^k)) \le \lambda_i^{S_k}(\bar{\Omega}^k + \rho_k G(x^k)) \le \lambda_m(\bar{\Omega}^k) + \rho_k \lambda_i^{U_k}(G(x^k)). \tag{32}$$

From now on, since the diagonalizations were taken in such a way that the eigenvalues of all the matrices considered are ordered, we can omit the superindexes in $\lambda_i^{U_k}(\cdot)$, $\lambda_i^{S_k}(\cdot)$, and $\lambda_i^{U}(\cdot)$. Now, if $\lambda_i(G(\bar{x})) < 0$, we have from

(32), since $\lambda_i(G(x^k)) \to \lambda_i(G(\bar{x}))$, $\{\bar{\Omega}^k\}$ is bounded, and $\rho_k \to +\infty$, that $\lambda_i(\bar{\Omega}^k + \rho_k G(x^k)) < 0$ for all sufficiently large k. Hence, for sufficiently large k,

$$\left[\lambda_i\left(\bar{\Omega}^k + \rho_k G(x^k)\right)\right]_+ \lambda_i\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right) = 0.$$

Now, if $\lambda_i(G(\bar{x})) = 0$ but $\lambda_i(G(x^k)) \leq 0$ for an infinite set of indices $K_1 \subset \mathbb{N}$, we have that

$$\begin{aligned} 0 \leq \lambda_i(\Omega^k) &= \max\{0, \lambda_i(\bar{\Omega}^k + \rho_k G(x^k))\} \\ &\leq \max\{0, \lambda_m(\bar{\Omega}^k) + \rho_k \lambda_i(G(x^k))\} \\ &\leq \max\{0, \lambda_m(\bar{\Omega}^k)\} + \rho_k \max\{0, \lambda_i(G(x^k))\} \\ &= \lambda_m(\bar{\Omega}^k). \end{aligned} \tag{33}$$

Thus, for $k \in K_1$, $\{\lambda_i(\Omega^k)\}$ is bounded. Since $\lambda_i\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right) \to 0$, we have

$$\lim_{k \in K_1} \left[\lambda_i\left(\bar{\Omega}^k + \rho_k G(x^k)\right)\right]_+ \lambda_i\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right) = 0.$$

Finally, if $\lambda_i(G(\bar{x})) = 0$ and $\lambda_i(G(x^k)) > 0$ for all sufficiently large k, let us multiply (32) by $\lambda_i\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right) > 0$ to arrive at

$$\begin{aligned} &\lambda_1(\bar{\Omega}^k)\lambda_i\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right) + \rho_k \lambda_i(G(x^k))\lambda_i\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right) \\ &\quad \leq \lambda_i(\bar{\Omega}^k + \rho_k G(x^k))\lambda_i\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right) \end{aligned} \tag{34}$$

and

$$\begin{aligned} &\lambda_i(\bar{\Omega}^k + \rho_k G(x^k))\lambda_i\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right) \\ &\quad \leq \lambda_m(\bar{\Omega}^k)\lambda_i\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right) + \rho_k \lambda_i(G(x^k))\lambda_i\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right). \end{aligned} \tag{35}$$

Note that

$$\lambda_1(\bar{\Omega}^k)\lambda_i\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right) \to 0 \text{ and } \lambda_m(\bar{\Omega}^k)\lambda_i\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right) \to 0,$$

then, it remains to show that

$$\rho_k \lambda_i(G(x^k))\lambda_i\left(\frac{\bar{\Omega}^k}{\rho_k} + G(x^k)\right) = \lambda_i(G(x^k))\lambda_i\left(\bar{\Omega}^k + \rho_k G(x^k)\right) \to 0.$$

Since $\lambda_i(G(x^k)) \to 0$, $\{\bar{\Omega}^k\}$ is bounded, and $\rho_k \lambda_i(G(x^k))^2 \to 0$ (by (29)), this follows by multiplying (32) by $\lambda_i(G(x^k))$:

$$\begin{aligned}\lambda_1(\bar{\Omega}^k)\lambda_i(G(x^k)) + \rho_k \lambda_i(G(x^k))^2 &\leq \lambda_i(\bar{\Omega}^k + \rho_k G(x^k))\lambda_i(G(x^k)) \\ &\leq \lambda_m(\bar{\Omega}^k)\lambda_i(G(x^k)) + \rho_k \lambda_i(G(x^k))^2. \end{aligned} \tag{36}$$

Thus, we conclude that $\lim_{k\to\infty}\langle \Omega^k, G(x^k)\rangle = 0$.

(ii) If $\{\rho_k\}$ is bounded, the proof follows as in the proof of Theorem 4, item (ii), by noting that

$$\begin{aligned}\lim_{k\to\infty}\langle \Omega^k, G(x^k)\rangle &= \sum_{i=1}^m \lambda_i^U(\bar{\Omega})\lambda_i^U(G(\bar{x})) \\ &= \sum_{i=1}^m \frac{1}{\rho_{k_0}}[\lambda_i^U]_+(\lambda_i^U - [\lambda_i^U]_+) = 0.\end{aligned}$$

□

To end this section, we discuss about the relevance of TAKKT in the context of sequential optimality conditions for NLP. For this, let us consider $G(x) := \operatorname{diag}(g_1(x), \ldots, g_m(x))$ and $\Omega^k := \operatorname{diag}(\mu_1^k, \ldots, \mu_m^k) \succeq 0$. Thus, a feasible point $\bar{x}$ satisfies TAKKT if, and only if,

$$\lim_{k\to\infty} \nabla f(x^k) + DG(x^k)^*\Omega^k = \lim_{k\to\infty} \nabla f(x^k) + \sum_{i=1}^m \mu_i^k \nabla g_i(x^k) = 0, \tag{37}$$

$$\lim_{k\to\infty}\langle \Omega^k, G(x^k)\rangle = \lim_{k\to\infty} \sum_{i=1}^m \mu_i^k g_i(x^k) = 0, \tag{38}$$

while for verifying AKKT, condition (38) is replaced by

$$g_i(\bar{x}) < 0 \Rightarrow \mu_i^k = 0, \text{ for all sufficiently large } k \text{ and } i = 1, \ldots, m,$$

which, for the sake of this discussion, we may consider its equivalent form

$$g_i(\bar{x}) < 0 \Rightarrow \mu_i^k \to 0^+, \text{ for all } i = 1, \ldots, m. \tag{39}$$

That is, the AKKT condition (37) + (39) says that the dual sequence $\{\mu^k\}$ may be unbounded, but the corresponding gradient of the Lagrangian function should vanish in the limit. Also, the gradient of an inactive constraint at the limit should not be considered, eventually, in the computation of the gradient of the Lagrangian. AKKT is typically considered as the most natural (and weakest) sequential optimality condition, as these properties are the least one would expect to hold at KKT-related conditions.

The complementarity condition imposed by TAKKT, namely, condition (38), has not been considered before in the context of NLP, as it is not natural in this context.

The issue is that it does not immediately imply that $\mu_i^k \to 0^+$ when $g_i(\bar{x}) < 0$. Namely, (38) may not be adequate to assert complementarity in the limit, as one could satisfy it, for instance, with an inactive constraint at the limit but with a corresponding Lagrange multiplier sequence bounded away from zero. For instance, with constraints $(g_1(x^k), g_2(x^k)) := (-1+\frac{1}{k}, \frac{1}{k}) \to (-1, 0)$ and $(\mu_1^k, \mu_2^k) := (1, k)$. That is, TAKKT may not correctly identify the active constraints in the verification of (37). Surprisingly, as we will show later, under MFCQ this issue is not present, as a TAKKT point is in fact a KKT point. The notion of a TAKKT point for NLP is closely related to the so-called Sum Converging to Zero (SCZ) concept defined in [27]; however, a thorough investigation of the relevance of TAKKT in NLP, together with its relations with other sequential optimality conditions, would be in order, but out of the scope of this paper. For example, even for NLP, although we conjecture that TAKKT would not imply AKKT, we were not able to find an example. That is, it is not clear if the concern raised above can actually happen.

Another known way to measure complementarity, in the case of NLP, is the so-called Complementarity-Approximate-KKT (CAKKT) condition defined in [10], where (38) is replaced by the complementarity measurement:

$$\lim_{k\to\infty} g_i(x^k)\mu_i^k = 0, \text{ for all } i = 1, \ldots, m. \tag{40}$$

Clearly, (40) implies (38) and (39), hence CAKKT is stronger than TAKKT and AKKT for NLP. Note that since these are necessary optimality conditions, a stronger condition is more desirable than a weaker one. For inactive constraints, (40) and (39) impose the same condition on the multiplier, namely, that it should converge to zero. The difference is with respect to an active constraint $g_i(\bar{x}) = 0$: while AKKT allows the corresponding Lagrange multiplier μ_i^k to diverge to infinity arbitrarily fast, CAKKT imposes a bound on its rate of growth, namely, it should go to infinity slower than $\frac{1}{|g_i(x^k)|}$ (see [29] for further discussions under this point of view). The CAKKT condition is natural in the context of interior point methods [29] and it is also generated by augmented Lagrangian methods under the generalized Lojasiewicz inequality [10]. In [10] it was also shown that CAKKT is able to detect non-optimality in some cases where AKKT does not detect it.

Hence, another interesting endeavour would be to define an extension of CAKKT to NLSDP. Similarly to what was done for AKKT, such an extension would need to take into account the "correspondence" among the eigenvalues of Ω^k and $G(x^k)$, in such a way that the condition would reduce to CAKKT for NLP in the diagonal case. A first idea would be to define CAKKT for NLSDP by replacing the complementarity measurement in AKKT by the stronger one

$$\lim_{k\to\infty} \lambda_i^{U_k}(G(x^k))\lambda_i^{S_k}(\Omega^k) = 0, \text{ for all } i = 1, \ldots, m,$$

where U_k and S_k are orthogonal matrices that diagonalize $G(x^k)$ and Ω^k, respectively, such that $S_k \to U$ and $U_k \to U$. Note that in the definition of AKKT we asked that $S_k \to U$ to obtain a correspondence between an inactive eigenvalue of the constraint matrix with an eigenvalue of the Lagrange multiplier matrix to vanish. In addition, for

CAKKT we would ask that $U_k \to U$ to ensure that $\lambda_i^{U_k}(G(x^k)) \to \lambda_i^U(G(\bar{x}))$ for all $i = 1, \ldots, m$.

Even though CAKKT defined in this way would be an optimality condition, we were not able to prove that the augmented Lagrangian algorithm generates this type of sequences under the generalized Lojasiewicz inequality. The crucial difficulty is that, as in the proof that the algorithm generates TAKKT sequences, it is important to take orthogonal diagonalizations of $G(x^k)$ and Ω^k in such a way that the eigenvalues are ordered. This is not an issue in TAKKT since it does not require any correspondence among the eigenvalues; however, the correspondence imposed by the definition of CAKKT does not allow us to arbitrarily order the eigenvalues of $G(x^k)$ and Ω^k.

In the next section we will show that our global convergence results to AKKT or TAKKT points are strictly stronger than standard global convergence results to a KKT point under MFCQ. In particular, our results do not require boundedness of Lagrange multipliers in order to assert that a feasible limit point of the algorithm is a KKT point.

6 Strength of the sequential optimality conditions

At this point, we have not yet mentioned that the optimality conditions that we have defined are strong in any sense. Namely, they could be satisfied at any feasible point, which would make our results up to now meaningless. Let us show that this is not the case. We start by discussing the strength of AKKT.

First, let us show that the global convergence results that we have proved are at least as good as standard global convergence results to a KKT point under MFCQ.

Theorem 7 *Let $\bar{x} \in \mathscr{F}$ be a feasible point that satisfies MFCQ. Then, for any objective function f in (NLSDP) such that $\bar{x}$ satisfies AKKT, $\bar{x}$ satisfies in addition the KKT conditions for this problem.*

Proof From the definition of AKKT, there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $x^k \to \bar{x}$ and $\{\Omega^k\} \subset \mathbb{S}^m_+$, together with orthogonal matrices S_k and U that diagonalize Ω^k for each k and $G(\bar{x})$, respectively, with $S_k \to U$, such that

$$\lim_{k\to\infty} \nabla f(x^k) + DG(x^k)^*\Omega^k = 0, \tag{41}$$

and

$$\lambda_i^U(G(\bar{x})) < 0 \Rightarrow \lambda_i^{S_k}(\Omega^k) = 0. \tag{42}$$

If $\{\Omega^k\}$ is contained in a compact set, there is $K \subset \mathbb{N}$, $\Omega \in \mathbb{S}^m_+$, such that $\lim_{k\in K} \Omega^k = \Omega$. Let us consider the partition of $\{1, \ldots, m\}$ into the index sets:

$$I_1 := \{i \mid \lambda_i^U(G(\bar{x})) < 0\} \text{ and } I_2 := \{i \mid \lambda_i^U(G(\bar{x})) = 0\}.$$

From (41), we have that

$$\nabla f(\bar{x}) + DG(\bar{x})^*\Omega = 0,$$

and by noting that $G(\bar{x})$ and Ω are simultaneously diagonalizable, we have that

$$\lim_{k\to\infty}\langle \Omega^k, G(x^k)\rangle = \langle \Omega, G(\bar{x})\rangle = \sum_{i\in I_1}\lambda_i^U(\Omega)\lambda_i^U(G(\bar{x})) + \sum_{i\in I_2}\lambda_i^U(\Omega)\lambda_i^U(G(\bar{x})) = 0,$$

where the previous equality is due to the fact that $\lambda_i^U(\Omega) = 0$ for $i \in I_1$ and $\lambda_i^U(G(\bar{x})) = 0$ for $i \in I_2$. Therefore, $\bar{x}$ satisfies the KKT conditions.

Now, let us consider a subsequence such that $t_k := \|\Omega^k\| \to \infty$. Thus, let us take $K_1 \subset \mathbb{N}$ such that $\lim_{k\in K_1} \dfrac{\Omega^k}{t_k} = \bar{\Omega} \neq 0$ for some $\bar{\Omega} \in \mathbb{S}^m_+$ simultaneously diagonalizable with $G(\bar{x})$. Then, from (41) and (42) we have that

$$\lim_{k\in K_1} \frac{\nabla f(x^k)}{t_k} + DG(x^k)^* \frac{\Omega^k}{t_k} = DG(\bar{x})^*\bar{\Omega} = 0, \text{ and} \tag{43}$$

$$\begin{aligned}\lim_{k\in K_1}\left\langle \frac{\Omega^k}{t_k}, G(x^k)\right\rangle = \langle \bar{\Omega}, G(\bar{x})\rangle &= \sum_{i\in I_1}\lambda_i^U(\bar{\Omega})\lambda_i^U(G(\bar{x}))\\ &+ \sum_{i\in I_2}\lambda_i^U(\bar{\Omega})\lambda_i^U(G(\bar{x})) = 0.\end{aligned} \tag{44}$$

To see that (43)–(44) contradicts MFCQ, let $d \in \mathbb{R}^n$ be such that $G(\bar{x}) + DG(\bar{x})d$ is negative definite. Thus,

$$\begin{aligned}0 &= \langle \bar{\Omega}, G(\bar{x})\rangle + \langle DG(\bar{x})^*\bar{\Omega}, d\rangle\\ &= \langle \bar{\Omega}, G(\bar{x})\rangle + \langle \bar{\Omega}, DG(\bar{x})d\rangle\\ &= \langle \bar{\Omega}, G(\bar{x}) + DG(\bar{x})d\rangle,\end{aligned}$$

which implies by Lemma 2 that $\bar{\Omega} = 0$, contradicting the definition of $\bar{\Omega}$.

□

A result similar to Theorem 7 holds for TAKKT. The proof is very similar to the proof of Theorem 7 and is omitted.

Theorem 8 *Let $\bar{x} \in \mathscr{F}$ be a feasible point that satisfies MFCQ. Then, for any objective function f in (NLSDP) such that $\bar{x}$ satisfies TAKKT, $\bar{x}$ satisfies in addition the KKT conditions for this problem.*

One may view Theorem 7 as strenghtening the simple fact that MFCQ is a constraint qualification, that is, MFCQ implies that the KKT condition holds not only at local minimizers, but also at AKKT points (which includes local minimizers, as shown in Theorem 2). This has been called a *strict* constraint qualification in [9]. We use Theorem 7 to inspire the definition of a new constraint qualification: note that the property satisfied by points fulfilling MFCQ, given by Theorem 7, is a property of the feasible set of (NLSDP), independently of the objective function. Also, since AKKT is an optimality condition, it is clear that this property is a constraint qualification (weaker than MFCQ). The same holds true with respect to the property stated in Theorem 8. Let us make these definitions precise.

Definition 6 (*AKKT/TAKKT-regularity*) We say that a feasible point $\bar{x} \in \mathscr{F}$ satisfies AKKT-regularity (TAKKT-regularity) if for any objective function f in (NLSDP) such that $\bar{x}$ satisfies AKKT (TAKKT, respectively), $\bar{x}$ satisfies in addition the KKT conditions for this problem.

Thus, the global convergence result presented in Theorem 4 implies that under AKKT-regularity, feasible limit points of the algorithm are KKT points. Analogously, Theorem 6 shows the global convergence of the algorithm to a KKT point under TAKKT-regularity. This gives a more standard constraint qualification formulation of our global convergence results, which can be compared with other ones in the literature.

Let us now show that our global convergence results are strictly stronger than the more standard one based on MFCQ. We do this by showing that neither AKKT-regularity nor TAKKT-regularity imply MFCQ (note that Theorems 7 and 8 show that AKKT/TAKKT-regularity are weaker than MFCQ). The following example serve this purpose:

Example 4 (AKKT-regularity and TAKKT-regularity do not imply MFCQ) Consider $\bar{x} := 0$ and the feasible set defined by

$$G(x) := \begin{bmatrix} x & 0 \\ 0 & -x \end{bmatrix} \preceq 0.$$

(i) ($\bar{x}$ does not satisfies MFCQ.) Indeed,

$$G(\bar{x}) + DG(\bar{x})h = \begin{bmatrix} h & 0 \\ 0 & -h \end{bmatrix}.$$

Therefore, there is no $h \in \mathbb{R}$ such that $G(\bar{x}) + DG(\bar{x})h$ is negative definite.

(ii) ($\bar{x}$ satisfies AKKT-regularity and TAKKT-regularity.) This is trivially true as, independently of the objective function, $\bar{x} = 0$ is a KKT point.

The previous example is very simple, as it reduces trivially to a nonlinear programming problem. However one could arrive at exactly the same conclusions by considering a feasible set defined by $G(x_1, x_2) := \begin{bmatrix} x_1 & x_2 \\ x_2 & -x_1 \end{bmatrix} \preceq 0.$

Let us note that by revisiting the constraints in Example 3 in more details, that is, without a fixed objective function, we can see that it actually shows that TAKKT-regularity does not imply AKKT-regularity. Note also that in Example 1, AKKT-regularity does not hold, given that AKKT holds but KKT does not hold. This shows that AKKT-regularity may fail for linear constraints (which does not occur for AKKT-regularity in NLP). Similarly, Example 2 shows that TAKKT-regularity may fail for linear constraints.

Similarly to [9], we can provide a geometric interpretation of AKKT/TAKKT-regularity by means of the outer semicontinuity of a point-to-set mapping. We present this interpretation for TAKKT-regularity, but a similar definition can be made for AKKT-regularity.

Theorem 9 *A feasible point $\bar{x}$ satisfies TAKKT-regularity if, and only if, the following point-to-set mapping is outer semicontinuous at $(\bar{x}, 0)$:*

$$K^{\mathrm{TAKKT}}(x, r) := \{DG(x)^* \Omega : |\langle \Omega, G(x)\rangle| \le r, \ \Omega \in \mathbb{S}^m_+\},$$

that is, $\limsup_{(x,r)\to(\bar{x},0)} K^{\mathrm{TAKKT}}(x, r) \subset K^{\mathrm{TAKKT}}(\bar{x}, 0)$.

Proof Let f be a smooth objective function such that TAKKT holds at $\bar{x}$. Then, by the definition, there exist sequences $\{x^k\} \subset \mathbb{R}^n$, $\{\Omega^k\} \in \mathbb{S}^m_+$, $\{\varepsilon_k\} \subset \mathbb{R}^n$, and $\{r^k\} \subset \mathbb{R}_+$ such that

$$\lim_{k\to\infty} x^k = \bar{x}, \varepsilon_k := \nabla f(x^k) + DG(x^k)^* \Omega^k \to 0, \text{ and } r^k := |\langle \Omega^k, G(x^k)\rangle| \to 0.$$

Define $w^k := DG(x^k)^* \Omega^k$. Clearly, the sequence w^k satisfies

$$w^k \in K^{\mathrm{TAKKT}}(x^k, r^k) \text{ and } w^k \to -\nabla f(\bar{x}).$$

From the definition of lim sup for the point-to-set mapping K^{TAKKT}, we have

$$-\nabla f(\bar{x}) \in \limsup_{(x,r)\to(\bar{x},0)} K^{\mathrm{TAKKT}}(x, r) \subset K^{\mathrm{TAKKT}}(\bar{x}, 0),$$

which, by the definition of $K^{\mathrm{TAKKT}}(\bar{x}, 0)$, implies that the KKT conditions hold. This proves that $\bar{x}$ satisfies TAKKT-regularity. Now, let us assume TAKKT-regularity at $\bar{x}$ and let us prove that

$$\limsup_{(x,r)\to(\bar{x},0)} K^{\mathrm{TAKKT}}(x, r) \subset K^{\mathrm{TAKKT}}(\bar{x}, 0).$$

Take $\overline{w} \in \limsup_{(x,r)\to(\bar{x},0)} K^{\mathrm{TAKKT}}(x, r)$. So, there are sequences $\{x^k\}$, $\{w^k\}$ and $\{r^k\}$ such that $x^k \to \bar{x}$, $w^k \to \overline{w}$, $r^k \to 0$ and $w^k \in K^{\mathrm{TAKKT}}(x^k, r^k)$ for all k. Now, let us define the linear function $f(x) := -\langle \overline{w}, x\rangle$ for all $x \in \mathbb{R}^n$. Let us see that TAKKT holds at $\bar{x}$ with this choice of f. Since $w^k \in K^{\mathrm{TAKKT}}(x^k, r^k)$, there is a multiplier $\Omega^k \in \mathbb{S}^m_+$ such that

$$w^k = DG(x^k)^* \Omega^k \text{ and } |\langle \Omega^k, G(x^k)\rangle| \le r^k.$$

Since $r^k \to 0$, we have $\lim_{k\to\infty}\langle \Omega^k, G(x^k)\rangle = 0$. Also, since $w^k \to \overline{w}$, we have $\varepsilon_k := \nabla f(x^k) + DG(x^k)^* \Omega^k = -\overline{w} + w^k \to 0$. Thus, TAKKT holds at $\bar{x}$, which implies by TAKKT-regularity that the objective function defined is such that the KKT conditions hold at $\bar{x}$. This can be written as $-\nabla f(\bar{x}) = \overline{w} \in K^{\mathrm{TAKKT}}(\bar{x}, 0)$, which concludes the proof. $\square$

For NLP, several constraint qualifications strictly weaker than MFCQ that still imply AKKT-regularity have been proposed (such as CPLD [49], RCPLD [7], CRSC [6] and CPG [6] – or CRCQ [34] and RCRCQ [47], which are independent of MFCQ). These weak constraint qualifications help in ensuring AKKT-regularity in a more tractable way by means of properties of the derivatives of the constraints in a neighborhood of the point of interest. We expect that extensions of these concepts to NLSDP would be relevant in characterizing global convergence of algorithms for NLSDP, while also showing that AKKT-regularity is much weaker than MFCQ, way beyond what our simple Example 4 suggests.

7 Conclusions

In the past ten years, sequential optimality conditions have played an increasing role in global convergence analysis of algorithms for nonlinear programming problems. Without a constraint qualification, the fact that the KKT conditions hold approximately at any local minimizer justifies the numerical practice of not verifying constraint qualifications at all when deciding when to stop the execution of an algorithm.

It can be conjectured that the stronger the theoretical properties of limit points of a sequence generated by an algorithm, the better the algorithm will behave in practice. This indicates the practical relevance of proving global convergence of algorithms under weak assumptions.

In this paper, we extended the NLP concept of Approximate-KKT point to nonlinear semidefinite programming. This extension is not straightforward as the KKT conditions for NLSDP require that the constraint matrix and the Lagrange multiplier matrix to be simultaneously diagonalizable; hence, it is not straightforward to define the notion of “satisfying the KKT conditions approximately”. Surprisingly, we were able to define a sequential optimality condition, new even in the context of NLP, that is naturally presented in NLSDP, as it does not rely on eigenvalue computations, which we call Trace-Approximate-KKT.

These tools allowed us to define an Augmented Lagrangian algorithm whose global convergence analysis can be done without relying on the Mangasarian-Fromovitz constraint qualification, that is, a limit point of a sequence generated by an algorithm may be proven to be a KKT point even though this point may have an unbounded Lagrange multiplier set.

Further research is needed to access the practical performance of the algorithm. We also envision that the theory presented can be generalized by considering an algorithm that solves constrained augmented Lagrangian subproblems, or by solving the subproblems up to second-order. Also, we expect other sequential optimality conditions for NLP to be generalized to NLSDP, such as the Approximate-Gradient-Projection [46], which can be relevant for algorithms that do not produce a Lagrange multiplier approximation.

References

1. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: On augmented Lagrangian methods with general lower-level constraint. SIAM J. Optim. **18**(4), 1286–1309 (2007)
2. Andreani, R., Birgin, E.G., Martínez, J.M., Schuverdt, M.L.: Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. Math. Program. **111**, 5–32 (2008)
3. Andreani, R., Fazzio, N.S., Schuverdt, M.L., Secchin, L.D.: A sequential optimality condition related to the quasinormality constraint qualification and its algorithmic consequences. Optimization online (2017). http://www.optimization-online.org/DB_HTML/2017/09/6194.html
4. Andreani, R., Haeser, G., Martínez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization **60**(5), 627–641 (2011)
5. Andreani, R., Haeser, G., Ramos, A., Silva, P.J.S.: A second-order sequential optimality condition associated to the convergence of algorithms. IMA J. Numer. Anal. **37**(4), 1902–1929 (2017)
6. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: Two new weak constraint qualifications and applications. SIAM J. Optim. **22**(3), 1109–1135 (2012)
7. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: A relaxed constant positive linear dependence constraint qualification and applications. Math. Program. **135**(1–2), 255–273 (2012)
8. Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: A cone-continuity constraint qualification and algorithmic consequences. SIAM J. Optim. **26**(1), 96–110 (2016)
9. Andreani, R., Martínez, J.M., Ramos, A., Silva, P.J.S.: Strict constraint qualifications and sequential optimality conditions for constrained optimization. Math. Oper. Res. **43**(3), 693–717 (2018)
10. Andreani, R., Martínez, J.M., Svaiter, B.F.: A new sequencial optimality condition for constrained optimization and algorithmic consequences. SIAM J. Optim. **20**(6), 3533–3554 (2010)
11. Andreani, R., Martínez, J.M., Santos, L.T.: Newton's method may fail to recognize proximity to optimal points in constrained optimization. Math. Program. **160**, 547–555 (2016)
12. Andreani, R., Secchin, L.D., Silva, P.J.S.: Convergence properties of a second order augmented Lagrangian method for mathematical programs with complementarity constraints. SIAM J. Optim. **28**(3), 2574–2600 (2018)
13. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Practical Methods of Optimization: Theory and Algorithms. Wiley, NJ (2006)
14. Birgin, E., Martínez, J.M.: Practical Augmented Lagrangian Methods for Constrained Optimization. SIAM, Philadelphia (2014)
15. Birgin, E.G., Gardenghi, J.L., Martínez, J.M., Santos, S.A., Toint, PhL: Evaluation complexity for nonlinear constrained optimization using unscaled KKT conditions and high-order models. SIAM J. Optim. **26**, 951–967 (2016)
16. Birgin, E.G., Haeser, G., Ramos, A.: Augmented Lagrangians with constrained subproblems and convergence to second-order stationary points. Comput. Optim. Appl. **69**(1), 51–75 (2018)
17. Birgin, E.G., Krejic, N., Martínez, J.M.: On the minimization of possibly discontinuous functions by means of pointwise approximations. Optim. Lett. **11**(8), 1623–1637 (2017)
18. Bolte, J., Daniilidis, A., Lewis, A.S.: The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim. **17**(4), 1205–1223 (2007)
19. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York (2000)
20. Correa, R., Ramírez, H.: A global algorithm for nonlinear semidefinite programming. SIAM J. Optim. **15**(1), 303–318 (2004)
21. Dutta, J., Deb, K., Tulshyan, R., Arora, R.: Approximate KKT points and a proximity measure for termination. J. Glob. Optim. **56**(4), 1463–1499 (2013)
22. Fares, B., Apkarian, P., Noll, D.: An augmented Lagrangian method for a class of LMI-constrained problems in robust control theory. Int. J. Control **74**(4), 348–360 (2001)
23. Fares, B., Noll, D., Apkarian, P.: Robust control via sequential semidefinite programming. SIAM J. Control Optim. **40**(6), 1791–1820 (2002)
24. Fiacco, A.V., McCormick, G.P.: Nonlinear Programming Sequential Unconstrained Minimization Techniques. Wiley, New York (1968)
25. Forsgren, A.: Optimality conditions for nonconvex semidefinite programming. Math. Program. **88**(1), 105–128 (2000)
26. Freund, R.W., Jarre, F., Vogelbusch, C.H.: Nonlinear semidefinite programming: sensitivity, convergence, and an application in passive reduced-order modeling. Math. Program. **109**, 581–611 (2007)

27. Giorgi, G., Jiménez, B., Novo, V.: Approximate Karush–Kuhn–Tucker condition in multiobjective optimization. J. Optim. Theory Appl. **171**(1), 70–89 (2016)
28. Gómez, W., Ramírez, H.: A filter algorithm for nonlinear semidefinite programming. Comput. Appl. Math. **29**(2), 297–328 (2010)
29. Haeser, G.: A second-order optimality condition with first- and second-order complementarity associated with global convergence of algorithms. Comput. Optim. Appl. **70**(2), 615–639 (2018)
30. Haeser, G., Melo, V.V.: Convergence detection for optimization algorithms: approximate-KKT stopping criterion when Lagrange multipliers are not available. Oper. Res. Lett. **43**(5), 484–488 (2015)
31. Haeser, G., Schuverdt, M.L.: On approximate KKT condition and its extension to continuous variational inequalities. J. Optim. Theory Appl. **149**(3), 528–539 (2011)
32. Horn, R.A., Johnson, C.R.: Topics in Matrix Analysis. Cambridge University Press, Cambridge (1991)
33. Huang, X.X., Teo, K.L., Yang, X.Q.: Approximate augmented Lagrangian functions and nonlinear semidefinite programs. Acta Math. Sin. **22**(5), 1283–1296 (2006)
34. Janin, R.: Directional Derivative of the Marginal Function in Nonlinear Programming, pp. 110–126. Springer, Berlin (1984)
35. Jarre, F.: Elementary optimality conditions for nonlinear SDPs. In: Handbook on Semidefinite, Conic and Polynomial Optimization. International Series in Operations Research & Management Science (2012)
36. Kočvara, M., Stingl, M.: PENNON—a generalized augmented Lagrangian method for semidefinite programming. In: Di Pillo, G., Murli, A. (eds.) High Performance Algorithms and Software for Nonlinear Optimization, pp. 297–315. Kluwer, Dordrecht (2003)
37. Kočvara, M., Stingl, M.: On the solution of large-scale SDP problems by the modified barrier method using iterative solvers. Math. Program. **109**, 413–444 (2007)
38. Kočvara, M., Stingl, M.: PENNON—a code for convex nonlinear and semidefinite programming. Optim. Methods Softw. **18**(3), 317–333 (2010)
39. Kanno, Y., Takewaki, I.: Sequential semidefinite program for maximum robustness design of structures under load uncertainty. J. Optim. Theory Appl. **130**, 265–287 (2006)
40. Konno, H., Kawadai, N., Wu, D.: Estimation of failure probability using semi-definite Logit model. Comput. Manag. Sci. **1**(1), 59–73 (2003)
41. Lewis, A.S.: Convex analysis on the Hermitian matrices. SIAM J. Optim. **6**(1), 164–177 (1993)
42. Lourenço, B.F., Fukuda, E.H., Fukushima, M.: Optimality conditions for nonlinear semidefinite programming via squared slack variables. Math. Program. **166**, 1–24 (2016)
43. Lovász, L.: Semidefinite Programs and Combinatorial Optimization, pp. 137–194. Springer, New York (2003)
44. Luo, H.Z., Wu, H.X., Chen, G.T.: On the convergence of augmented Lagrangian methods for nonlinear semidefinite programming. J. Glob. Optim. **54**(3), 599–618 (2012)
45. Martínez, J.M., Pilotta, E.A.: Inexact restoration algorithm for constrained optimization. J. Optim. Theory Appl. **104**(1), 135–163 (2000)
46. Martínez, J.M., Svaiter, B.F.: A practical optimality condition without constraint qualifications for nonlinear programming. J. Optim. Theory Appl. **118**(1), 117–133 (2003)
47. Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optimization **60**(4), 429–440 (2011)
48. Qi, H., Sun, D.: A quadratically convergent newton method for computing the nearest correlation matrix. SIAM J. Matrix Anal. Appl. **28**(2), 360–385 (2006)
49. Qi, L., Wei, Z.: On the constant positive linear dependence conditions and its application to SQP methods. SIAM J. Optim. **10**(4), 963–981 (2000)
50. Ramos, A.: Mathematical programs with equilibrium constraints: a sequential optimality condition, new constraint qualifications and algorithmic consequences. Optimization online (2016). http://www.optimization-online.org/DB_HTML/2016/04/5423.html
51. Shapiro, A.: First and second order analysis of nonlinear semidefinite programs. SIAM J. Optim. **77**(1), 301–320 (1997)
52. Shapiro, A., Sun, J.: Some properties of the augmented Lagrangian in cone constrained optimization. Math. Oper. Res. **29**, 479–491 (2004)
53. Stingl, M.: On the Solution of Nonlinear Semidefinite Programs by Augmented Lagrangian Methods. PhD thesis, University of Erlangen (2005)
54. Stingl, M., Kočvara, M., Leugering, G.: A sequential convex semidefinite programming algorithm with an application to multiple-load free material optimization. SIAM J. Optim. **20**(1), 130–155 (2009)

55. Sun, D., Sun, J., Zhang, L.: The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. Math. Program. **114**(2), 349–391 (2008)
56. Sun, J., Zhang, L.W., Wu, Y.: Properties of the augmented Lagrangian in nonlinear semidefinite optimization. J. Optim. Theory Appl. **12**(3), 437–456 (2006)
57. Theobald, C.M.: An inequality for the trace of the product of two symmetric matrices. Math. Proc. Camb. Philos. Soc. **77**(2), 265–267 (1975)
58. Todd, M.J.: Semidefinite optimization. Acta Numer. **10**, 515–560 (2003)
59. Tuyen, N.V., Yao, J., Wen, C.: A Note on Approximate Karush–Kuhn–Tucker Conditions in Locally Lipschitz Multiobjective Optimization. ArXiv:1711.08551 (2017)
60. Vandenberghe, L., Boyd, S.: Semidefinite programming. SIAM Rev. **38**(1), 549–95 (1996)
61. Vandenberghe, L., Boyd, S., Wu, S.P.: Determinant maximization with linear matrix inequality constraints. SIAM J. Matrix Anal. Appl. **19**(2), 499–533 (1998)
62. Wu, H., Luo, H., Ding, X., Chen, G.: Global convergence of modified augmented Lagrangian methods for nonlinear semidefinite programmings. Comput. Optim. Appl. **56**(3), 531–558 (2013)
63. Yamashita, H., Yabe, H.: Local and superlinear convergence of a primal-dual interior point method for nonlinear semidefinite programming. Math. Program. **132**(1–2), 1–30 (2012)
64. Yamashita, H., Yabe, H.: A survey of numerical methods for nonlinear semidefinite programming. J. Oper. Res. Soc. Jpn. **58**(1), 24–60 (2015)
65. Yamashita, H., Yabe, H., Harada, K.: A primal-dual interior point method for nonlinear semidefinite programming. Math. Program. **135**(1–2), 89–121 (2012)
66. Zhu, Z.B., Zhu, H.L.: A filter method for nonlinear semidefinite programming with global convergence. Acta Math. Sin. **30**(10), 1810–1826 (2014)