

RESEARCH

Open Access



# A superlinearly convergent SSDP algorithm for nonlinear semidefinite programming

Jian Ling Li<sup>1\*</sup> and Hui Zhang<sup>1</sup>

\*Correspondence:  
[jianlingli@126.com](mailto:jianlingli@126.com)

<sup>1</sup>College of Mathematics and Information Science, Guangxi University, Nanning, China

## Abstract

In this paper, we present a sequential semidefinite programming (SSDP) algorithm for nonlinear semidefinite programming. At each iteration, a linear semidefinite programming subproblem and a modified quadratic semidefinite programming subproblem are solved to generate a master search direction. In order to avoid Maratos effect, a second-order correction direction is determined by solving a new quadratic programming. And then a penalty function is used as a merit function for arc search. The superlinear convergence is shown under the strict complementarity and the strong second-order sufficient conditions with the sigma term. Finally, some preliminary numerical results are reported.

**Keywords:** Nonlinear semidefinite programming; Penalty function; Sequential semidefinite programming; Global convergence; Superlinear convergence

## 1 Introduction

Consider the following nonlinear semidefinite programming (NLSDP) with a negative semidefinite matrix constraint:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & \mathcal{A}(x) \preceq 0, \end{aligned} \tag{1.1}$$

where  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $\mathcal{A}: \mathbb{R}^n \rightarrow \mathbb{S}^m$ ,  $\mathbb{S}^m$  is the set of  $m$ -order symmetric matrix and  $\mathbb{S}_+^m$  ( $\mathbb{S}_-^m$ ) is the set of  $m$ -order positive (negative) semidefinite matrix.  $\mathcal{A}(x) \preceq 0$  means that  $\mathcal{A}(x)$  is a negative semidefinite matrix.

Nonlinear semidefinite programming has many applications both in theory and in the real world. Many convex optimization problems, such as variational inequality problems, fixed point problems [1–3], can be reformulated as convex NLSDP. Robust control problems, optimal structural design, and truss design problems can be reformulated as NLSDP (see [4–6]). There are a lot of literature for NLSDP on algorithms, for example, the augmented Lagrangian method [7–12], primal-dual interior point method [13, 14], and sequential semidefinite programming (SSDP) method [15–21]. Our research focus is on the SSDP method.

SSDP method, which is a generalization of SQP method for classic nonlinear programming, is one of effective methods for nonlinear semidefinite programming. For example,

Correa and Ramirez [16] proposed a global SSDP algorithm for NLSDP. Recently, as illustrated by the extensive numerical experiments in [20, 21], SSDP algorithm has performed very well in finding a solution to NLSDP. At each iteration of SSDP method, a special quadratic semidefinite programming subproblem is solved to generate a search direction. However, just as traditional SQP method, most of existing SSDP methods also have some inherent pitfalls, e.g., (1) the first direction finding subproblem (DFP for short), namely a quadratic semidefinite programming (QSDP for short), is not ensured to be consistent. The algorithm in [21] is based on the assumption that the optimal solution of the first DFP exists. The algorithm in [20] directly goes to feasibility restoration phase when the first DFP is inconsistent. As we know, feasibility restoration phase will increase the computational cost. (2) The optimal solution to the first DFP is not ensured to be an improving direction, so it is possible that Maratos effect occurs. As a result, the superlinear convergence is not guaranteed to obtain.

Since NLSDP contains a negative semidefinite matrix constraint, it is more difficult to deal with these drawbacks comparing with SQP method for classic nonlinear programming. In this paper, we borrow the ideas of modified strategy of quadratic programming subproblem for nonlinear programming from [22]. We first construct a linear semidefinite programming (LSDP for short), and then by means of the solution of the LSDP we construct a special QSDP to yield the master search direction, which is ensured to be consistent. In order to avoid the Maratos effect, a second-order correction direction is introduced which is determined by solving a new quadratic programming. A penalty function is used as a merit function for arc search. The proposed algorithm possesses superlinear convergence under the strict complementarity and the strong second-order sufficient conditions with the sigma term.

The paper is organized as follows. Some notations and preliminaries are described in the next section. In Sect. 3, we present our algorithm in detail and analyze its feasibility. Under some mild conditions, the global convergence and superlinear convergence are shown in Sect. 4 and Sect. 5, respectively. In Sect. 6, the preliminary numerical results are reported. Some concluding remarks are given in the last section.

## 2 Preliminaries

In this section, for the sake of convenience, some definitions, notations, and results for NLSDP are introduced.

The differential operator  $D\mathcal{A}(x) : \mathbb{R}^n \rightarrow \mathbb{S}^m$  is defined by

$$D\mathcal{A}(x)d := \sum_{i=1}^n d_i \frac{\partial \mathcal{A}(x)}{\partial x_i}, \quad \forall d \in \mathbb{R}^n. \tag{2.1}$$

The adjoint operator  $D\mathcal{A}(x)^*$  of  $D\mathcal{A}(x)$  is defined by

$$D\mathcal{A}(x)^*Z = \left( \left\langle \frac{\partial \mathcal{A}(x)}{\partial x_1}, Z \right\rangle, \dots, \left\langle \frac{\partial \mathcal{A}(x)}{\partial x_n}, Z \right\rangle \right)^T, \quad \forall Z \in \mathbb{S}^m, \tag{2.2}$$

where  $\langle A, B \rangle$  means the inner product of the matrix  $A$  and  $B$ , which is defined by  $\langle A, B \rangle = \text{Tr}(AB)$  for any  $A, B \in \mathbb{S}^m$ ,  $\text{Tr}(\cdot)$  is the trace of a matrix.

The operator  $D^2\mathcal{A}(x) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{S}^m$  is defined by

$$d^T D^2\mathcal{A}(x)\bar{d} := \sum_{i,j=1}^n d_i \bar{d}_j \frac{\partial^2 \mathcal{A}(x)}{\partial x_i \partial x_j}, \quad \forall d, \bar{d} \in \mathbb{R}^n. \tag{2.3}$$

**Definition 2.1** Given  $x \in \mathbb{R}^n$ , if there exists a matrix  $\Lambda \in \mathbb{S}_+^m$  such that

$$\nabla f(x) + D\mathcal{A}(x)^* \Lambda = 0, \tag{2.4a}$$

$$\mathcal{A}(x) \leq 0, \tag{2.4b}$$

$$\text{Tr}(\Lambda \mathcal{A}(x)) = 0, \tag{2.4c}$$

then  $x$  is called a KKT point of NLSDP (1.1), the matrix  $\Lambda$  is called the Lagrangian multiplier, (2.4a)–(2.4c) is called the KKT conditions of NLSDP (1.1).

Let  $\lambda_1, \dots, \lambda_n$  be the eigenvalues of  $A \in \mathbb{R}^{n \times n}$ , and let  $\lambda_1(A)$  be the largest eigenvalue of  $A$ . The following results will be used in the subsequent analysis.

**Lemma 2.1** ([23]) For any  $A, B \in \mathbb{S}^m$ , the following inequality is true:

$$\text{Tr}(AB) \leq \sum_{i=1}^m \lambda_i(A)\lambda_i(B), \tag{2.5}$$

the equality holds if and only if there exists an invertible matrix  $P$  such that  $P^{-1}AP$  and  $P^{-1}BP$  are diagonal.

Based on Lemma 2.1, the following result is obvious.

**Lemma 2.2** For any  $A \in \mathbb{S}^m, B \in \mathbb{S}_+^m$ , the following inequality is true:

$$\text{Tr}(AB) \leq \lambda_1(A) \text{Tr}(B). \tag{2.6}$$

**Lemma 2.3** ([24] (Weyl’s inequality)) Suppose  $A, B \in \mathbb{S}^m$ , then  $\lambda_1(A + B) \leq \lambda_1(A) + \lambda_1(B)$ .

**Lemma 2.4** For any  $A, B \in \mathbb{S}^m$ , if  $\lambda_1(A + B) < \lambda_1(A)$ , then the following inequality is true:

$$\lambda_1(A + \eta B) \leq \lambda_1(A), \quad \forall \eta \in (0, 1).$$

*Proof* If  $\lambda_1(B) \leq 0$ , then it follows from Lemma 2.3 that

$$\lambda_1(A + \eta B) \leq \lambda_1(A) + \lambda_1(\eta B) \leq \lambda_1(A),$$

that is, the result is true.

If  $\lambda_1(B) > 0$ , note that  $\lambda_1(A + B) < \lambda_1(A)$  and  $\eta \in (0, 1)$ , then it follows from Lemma 2.3 that

$$\begin{aligned} \lambda_1(A + \eta B) &= \lambda_1(A + B + (\eta - 1)B) \\ &\leq \lambda_1(A + B) + \lambda_1((\eta - 1)B) \end{aligned}$$

$$\begin{aligned} &\leq \lambda_1(A) + (\eta - 1)\lambda_1(B) \\ &\leq \lambda_1(A), \end{aligned}$$

so the result is proved. □

Let  $x^k$  be the current iterative point, motivated by the idea in [22], we construct a linear SDP subproblem (LSDP) as follows:

$$\begin{aligned} \text{LSDP}(x^k) \quad \min \quad & z \\ \text{s.t.} \quad & \mathcal{A}(x^k) + D\mathcal{A}(x^k)d \preceq zE_m, \\ & z \geq 0, \quad \|d\| \leq 1, \end{aligned} \tag{2.7}$$

where  $E_m$  is the  $m$ -order identity matrix.

It is known that  $\text{LSDP}(x^k)$  (2.7) has optimal solutions. Let  $((\widehat{d}^k)^T, z_k)^T$  be an optimal solution of  $\text{LSDP}(x^k)$  (2.7). Now we construct a quadratic semidefinite programming (QSDP( $x^k, H_k$ )) by means of  $z_k$  as follows:

$$\begin{aligned} \text{QSDP}(x^k, H_k) \quad \min \quad & \nabla f(x^k)^T d + \frac{1}{2}d^T H_k d \\ \text{s.t.} \quad & \mathcal{A}(x^k) + D\mathcal{A}(x^k)d \preceq z_k E_m, \end{aligned} \tag{2.8}$$

where  $H_k \in \mathbb{S}^n$  is the Hesse matrix or an approximation of the Hesse matrix of the Lagrangian function of NLSDP (1.1) at  $x^k$ .

Generally, the optimal solution  $d^k$  to  $\text{QSDP}(x^k, H_k)$  (2.8) cannot be guaranteed to avoid the Maratos effect and get superlinear convergence, so it needs a modification. To this end, motivated by the ideas in [20], we introduce a second-order correction direction by solving the following subproblem:

$$\begin{aligned} \min \quad & \nabla f(x^k)^T (d^k + d) + \frac{1}{2}(d^k + d)^T H_k (d^k + d) \\ \text{s.t.} \quad & \bar{N}_k^T (\mathcal{A}(x^k + d^k) + D\mathcal{A}(x^k)d) \bar{N}_k = -\|d^k\|^\varrho E_{m-r}, \end{aligned} \tag{2.9}$$

where  $\varrho \in (2, 3)$ ,  $r = \text{rank}(\mathcal{A}(x^k) + D\mathcal{A}(x^k)d^k)$ ,  $\bar{N}_k = (p_1^k, p_2^k, \dots, p_{m-r}^k) \in \mathbb{R}^{m \times (m-r)}$ , and  $\{p_1^k, p_2^k, \dots, p_{m-r}^k\}$  is an orthogonal basis for the null space of the matrix  $\mathcal{A}(x^k) + D\mathcal{A}(x^k)d^k$ .

The following basic assumptions are required.

**A 1** The functions  $f(x)$  and  $\mathcal{A}(x)$  are continuously differentiable.

**A 2** There exist two positive constants  $a$  and  $\bar{a}$  such that

$$a\|d\|^2 \leq d^T H_k d \leq \bar{a}\|d\|^2, \quad \forall d \in \mathbb{R}^n.$$

Under Assumptions A1–A2, the following lemma follows.

**Lemma 2.5** *Suppose that Assumptions A1–A2 hold. Then subproblem  $\text{QSDP}(x^k, H_k)$  (2.8) has a unique solution  $d^k$ , and there exists a matrix  $\Lambda_k \in \mathbb{S}_+^m$  satisfying the KKT conditions*

of QSDP( $x^k, H_k$ ) (2.8), i.e.,

$$\nabla f(x^k) + D\mathcal{A}(x^k)^* \Lambda_k + H_k d^k = 0, \tag{2.10a}$$

$$\mathcal{A}(x^k) + D\mathcal{A}(x^k)d^k \leq z_k E_m, \tag{2.10b}$$

$$\text{Tr}(\Lambda_k(\mathcal{A}(x^k) + D\mathcal{A}(x^k)d^k - z_k E_m)) = 0. \tag{2.10c}$$

Define a measure of constraint violation for NLSDP (1.1) as follows:

$$P(x) = \lambda_1(\mathcal{A}(x))_+, \tag{2.11}$$

where  $\lambda_1(\mathcal{A}(x))_+ = \max\{\lambda_1(\mathcal{A}(x)), 0\}$ . Obviously,  $P(x) = 0$  if and only if  $x$  is a feasible point of NLSDP (1.1).

By means of  $P(x)$ , we define a penalty function as a merit function for arc search:

$$\theta_\alpha(x) = f(x) + \alpha P(x) = f(x) + \alpha \lambda_1(\mathcal{A}(x))_+, \tag{2.12}$$

where  $\alpha > 0$  is a penalty parameter. The function  $\theta_\alpha(x)$  comes from the Han penalty function for nonlinear programming.

**Lemma 2.6** *Suppose that Assumptions A1–A2 hold,  $d^k$  is the optimal solution of QSDP( $x^k, H_k$ ) (2.8). Then the directional derivative  $\theta'_\alpha(x^k; d^k)$  satisfies the following inequality:*

$$\begin{aligned} \theta'_\alpha(x^k; d^k) &\leq \nabla f(x^k)^\top d^k - \alpha(\lambda_1(\mathcal{A}(x^k))_+ - \lambda_1(z_k E_m)_+) \end{aligned} \tag{2.13}$$

$$\leq -(d^k)^\top H_k d^k + \text{Tr}(\Lambda_k(\mathcal{A}(x^k) - z_k E_m)) - \alpha(\lambda_1(\mathcal{A}(x^k))_+ - \lambda_1(z_k E_m)_+), \tag{2.14}$$

where  $z_k$  is the optimal value of LSDP( $x^k$ ) (2.7),  $\Lambda_k$  is a KKT multiplier of QSDP( $x^k, H_k$ ) (2.8) corresponding to the constraint.

*Proof* First, it follows from (2.10a) that

$$\nabla f(x^k)^\top d^k = - \sum_{i=1}^n d_i^k \text{Tr}\left(\Lambda_k \frac{\partial \mathcal{A}(x^k)}{\partial x_i}\right) - (d^k)^\top H_k d^k. \tag{2.15}$$

Since  $\lambda_1(\cdot)_+$  is convex, we obtain

$$\begin{aligned} &\lambda_1\left(\mathcal{A}(x^k) + t \sum_{i=1}^n d_i^k \frac{\partial \mathcal{A}(x^k)}{\partial x_i}\right)_+ \\ &\leq (1-t)\lambda_1(\mathcal{A}(x^k))_+ + t\lambda_1\left(\mathcal{A}(x^k) + \sum_{i=1}^n d_i^k \frac{\partial \mathcal{A}(x^k)}{\partial x_i}\right)_+ \\ &\leq (1-t)\lambda_1(\mathcal{A}(x^k))_+ + t\lambda_1(z_k E_m)_+, \end{aligned}$$

the last inequality above is due to (2.10b). By the definition of directional derivative and the above inequality, we get

$$\begin{aligned}
 & [\lambda_1(\cdot)_+]'(\mathcal{A}(x^k); D\mathcal{A}(x^k)d^k) \\
 &= \lim_{t \rightarrow 0^+} t^{-1} \left[ \lambda_1 \left( \mathcal{A}(x^k) + t \sum_{i=1}^n d_i^k \frac{\partial \mathcal{A}(x^k)}{\partial x_i} \right)_+ - \lambda_1(\mathcal{A}(x^k))_+ \right] \\
 &\leq \lim_{t \rightarrow 0^+} t^{-1} [(1-t)\lambda_1(\mathcal{A}(x^k))_+ + t\lambda_1(z_k E_m)_+ - \lambda_1(\mathcal{A}(x^k))_+] \\
 &= -(\lambda_1(\mathcal{A}(x^k))_+ - \lambda_1(z_k E_m)_+). \tag{2.16}
 \end{aligned}$$

Combining with the definition of the directional derivative  $\theta'_\alpha(x^k; d^k)$  and (2.16), we have

$$\begin{aligned}
 \theta'_\alpha(x^k; d^k) &= \nabla f(x^k)^T d^k + \alpha([\lambda_1(\cdot)_+]'(\mathcal{A}(x^k); D\mathcal{A}(x^k)d^k)) \\
 &\leq \nabla f(x^k)^T d^k - \alpha(\lambda_1(\mathcal{A}(x^k))_+ - \lambda_1(z_k E_m)_+),
 \end{aligned}$$

that is, inequality (2.13) holds.

By (2.15), we obtain

$$\begin{aligned}
 & \nabla f(x^k)^T d^k - \alpha(\lambda_1(\mathcal{A}(x^k))_+ - \lambda_1(z_k E_m)_+) \\
 &\leq - \sum_{i=1}^n d_i^k \text{Tr} \left( \Lambda_k \frac{\partial \mathcal{A}(x^k)}{\partial x_i} \right) - (d^k)^T H_k d^k - \alpha(\lambda_1(\mathcal{A}(x^k))_+ - \lambda_1(z_k E_m)_+). \tag{2.17}
 \end{aligned}$$

It follows from (2.10c) that

$$- \text{Tr} \left( \Lambda_k \left( \sum_{i=1}^n d_i^k \frac{\partial \mathcal{A}(x^k)}{\partial x_i} \right) \right) = - \text{Tr}(\Lambda_k (D\mathcal{A}(x^k)d^k)) = \text{Tr}(\Lambda_k (\mathcal{A}(x^k) - z_k E_m)). \tag{2.18}$$

Substituting the above equality into (2.17), one has

$$\begin{aligned}
 & \nabla f(x^k)^T d^k - \alpha(\lambda_1(\mathcal{A}(x^k))_+ - \lambda_1(z_k E_m)_+) \\
 &\leq -(d^k)^T H_k d^k + \text{Tr}(\Lambda_k (\mathcal{A}(x^k) - z_k E_m)) - \alpha(\lambda_1(\mathcal{A}(x^k))_+ - \lambda_1(z_k E_m)_+),
 \end{aligned}$$

that is, inequality (2.14) holds. □

### 3 The algorithm

In this section, we first present our algorithm in detail, and then analyze its feasibility.

#### Algorithm A

- Step 0. Given  $x^0 \in \mathbb{R}^n$ ,  $H_0 = E_n$  (identity matrix),  $t \in (0, 1)$ ,  $\alpha_0 > 0$ ,  $\bar{P} \in (1, 10)$ ,  $\sigma \in (0, 1)$ ,  $\beta \in (0, \frac{1}{2})$ ,  $\eta_1 > 0$ . Let  $k := 0$ .
- Step 1. Solve LSDP( $x^k$ ) (2.7) to get an optimal solution  $((\hat{d}^k)^T, z_k)^T$ . If  $\lambda_1(\mathcal{A}(x^k)) > 0$  and  $z_k = \lambda_1(\mathcal{A}(x^k))$ , then stop.
- Step 2. (Generate a master direction). Solve QSDP( $x^k, H_k$ ) (2.8) to get the optimal solution  $d^k$ . If  $d^k = 0$ , then stop.

Step 3. (Generate a second-order correction direction). Solve subproblem (2.9) and let  $\tilde{d}^k$  be the solution. If there is no solution of subproblem (2.9) or  $\|\tilde{d}^k\| > \|d^k\|$ , then set  $\tilde{d}^k = 0$ .

Step 4. (Update  $\alpha_k$ ) The update rule of  $\alpha_k$  is as follows:

$$\alpha_{k+1} := \begin{cases} \alpha_k, & \text{if } \Delta(x^k, \alpha_k) \leq -(d^k)^T H_k d^k; \\ \frac{\nabla f(x^k)^T d^k + (d^k)^T H_k d^k}{\lambda_1(\mathcal{A}(x^k))_+ - z_k} + \eta_1, & \text{otherwise,} \end{cases} \tag{3.1}$$

where  $\Delta(x^k, \alpha_k)$  is defined by

$$\Delta(x^k, \alpha_k) = \nabla f(x^k)^T d^k - \alpha_k (\lambda_1(\mathcal{A}(x^k))_+ - \lambda_1(z_k E_m)_+), \tag{3.2}$$

that is,  $\Delta(x^k, \alpha_k)$  is the right-hand side of inequality (2.13).

Step 5. (Arc search) Let  $t_k$  be the first number of the sequence  $\{1, \sigma, \sigma^2, \dots\}$  satisfying the following inequalities:

$$\theta_{\alpha_{k+1}}(x^k + t_k d^k + t_k^2 \tilde{d}^k) \leq \theta_{\alpha_{k+1}}(x^k) + \beta t_k \Delta(x^k, \alpha_{k+1}), \quad \text{if } P(x^k) \leq \bar{P}; \tag{3.3}$$

$$\begin{cases} \theta_{\alpha_{k+1}}(x^k + t_k d^k + t_k^2 \tilde{d}^k) \leq \theta_{\alpha_{k+1}}(x^k) + \beta t_k \Delta(x^k, \alpha_{k+1}), \\ P(x^k + t_k d^k) \leq P(x^k), \end{cases} \quad \text{if } P(x^k) > \bar{P}. \tag{3.4}$$

Step 6. Let  $x^{k+1} := x^k + t_k d^k + t_k^2 \tilde{d}^k$ , update  $H_k$  by some method to  $H_{k+1}$  such that  $H_{k+1}$  is positive definite. Set  $k := k + 1$ , and return to Step 1.

*Remark* In Algorithm A, by means of new modified strategies of subproblem, the quadratic semidefinite programming subproblem (2.8) yielding master searching direction is guaranteed to be consistent; further, it is ensured that the solution to subproblem (2.8) exists. This is very different from the ways in [20, 21].

In what follows, we analyze the feasibility of Algorithm A. To this end, it is necessary to extend the definition of infeasible stationary point for nonlinear programming [25, 26] to nonlinear semidefinite programming.

**Definition 3.1** A point  $\tilde{x} \in \mathbb{R}^n$  is called an infeasible stationary point of NLSDP (1.1) if  $\lambda_1(\mathcal{A}(\tilde{x})) > 0$  and

$$\min_{d \in \mathbb{R}^n} \max \{ \lambda_1(\mathcal{A}(\tilde{x}) + D\mathcal{A}(\tilde{x})d), 0 \} = \lambda_1(\mathcal{A}(\tilde{x}))_+ = P(\tilde{x}). \tag{3.5}$$

**Theorem 3.1** Suppose that Assumptions A1–A2 hold, then the following two results are true:

- (1) If Algorithm A stops at Step 1, then  $x^k$  is an infeasible stationary point of NLSDP (1.1).
- (2) If Algorithm A stops at Step 2, then  $x^k$  is a KKT point of NLSDP (1.1).

*Proof* (1) If Algorithm A stops at Step 1, i.e.,  $z_k = \lambda_1(\mathcal{A}(x^k)) = P(x^k) > 0$ , then  $x^k$  is an infeasible solution of NLSDP (1.1). In the following, we will prove  $x^k$  is an infeasible stationary

point of NLSDP (1.1), i.e.,

$$\min_{d \in \mathbb{R}^n} \max \{ \lambda_1(\mathcal{A}(x^k) + D\mathcal{A}(x^k)d), 0 \} = \max \{ \lambda_1(\mathcal{A}(x^k)), 0 \} = P(x^k).$$

By contradiction, suppose that  $x^k$  is not an infeasible stationary point, so there exists  $d^{k,0} \in \mathbb{R}^n$  such that

$$\max \{ \lambda_1(\mathcal{A}(x^k) + D\mathcal{A}(x^k)d^{k,0}), 0 \} < P(x^k). \tag{3.6}$$

If  $\|d^{k,0}\| > 1$ , then  $\frac{1}{\|d^{k,0}\|} < 1$ , so by Lemma 2.4, we have

$$\max \left\{ \lambda_1 \left( \mathcal{A}(x^k) + \frac{1}{\|d^{k,0}\|} D\mathcal{A}(x^k)d^{k,0} \right), 0 \right\} < P(x^k), \tag{3.7}$$

hence, we suppose, without loss of generality, that  $\|d^{k,0}\| \leq 1$ . Let

$$\tilde{z}_k := \max \{ \lambda_1(\mathcal{A}(x^k)) + D\mathcal{A}(x^k)d^{k,0}, 0 \} < P(x^k). \tag{3.8}$$

Obviously,  $((d^{k,0})^T, \tilde{z}_k)^T$  is a feasible point of subproblem LSDP( $x^k$ ) (2.7). Since  $z_k$  is the optimal value of LSDP( $x^k$ ) (2.7), one has

$$z_k \leq \tilde{z}_k < P(x^k), \tag{3.9}$$

which contradicts to  $z_k = P(x^k)$ . Hence, if Algorithm A stops at Step 1, then  $x^k$  is an infeasible stationary point of NLSDP (1.1).

(2) If Algorithm A stops at Step 2, i.e.,  $d^k = 0$ , then by the KKT conditions (2.10a)–(2.10c) of QSDP( $x^k, H_k$ ) (2.8), we obtain

$$\nabla f(x^k) + D\mathcal{A}(x^k)^* \Lambda_k = 0, \tag{3.10a}$$

$$\mathcal{A}(x^k) \preceq z_k E_m, \tag{3.10b}$$

$$\Lambda_k \succeq 0, \tag{3.10c}$$

$$\text{Tr}(\Lambda_k(\mathcal{A}(x^k) - z_k E_m)) = 0. \tag{3.10d}$$

Now we prove  $z_k = 0$ . By contradiction, suppose  $z_k \neq 0$ , then  $z_k > 0$ . So it follows that  $x^k$  is an infeasible point of NLSDP (1.1), i.e.,  $\lambda_1(\mathcal{A}(x^k)) > 0$ . It is obvious that  $(0, \lambda_1(\mathcal{A}(x^k)))$  is a feasible point of LSDP( $x^k$ ) (2.7), so the optimal solution  $z_k \leq \lambda_1(\mathcal{A}(x^k))$ . Since Algorithm A does not stop at Step 1, one has  $z_k < \lambda_1(\mathcal{A}(x^k))$ , which implies that  $d = 0$  is not a feasible point of QSDP( $x^k, H_k$ ) (2.8). This contradicts the fact that 0 is the optimal solution of QSDP( $x^k, H_k$ ) (2.8). Therefore,  $z_k = 0$ . Substituting  $z_k = 0$  into (3.10b) and (3.10d), combining with (3.10a), (3.10c), we can conclude that  $x^k$  is a KKT point of NLSDP (1.1).  $\square$

**Lemma 3.1** *Suppose that Assumptions A1–A2 hold, if Algorithm A does not stop at Step 1 and  $d^k \neq 0$ , then the following conclusions are true:*

- (i) *If  $P(x^k) > 0$ , then the directional derivative  $P'(x^k; d^k) < 0$ ;*
- (ii)  *$\theta'_{\alpha_{k+1}}(x^k; d^k) < 0$ ;*



(iii)  $\Delta(x^k, \alpha_{k+1}) < 0$ , so Algorithm A is well defined.

*Proof* (i) If  $P(x^k) > 0$ , it means that  $x^k$  is an infeasible point of NLSDP (1.1). We can prove  $z_k < \lambda_1(\mathcal{A}(x^k))_+$ . By contradiction, if  $z_k \geq \lambda_1(\mathcal{A}(x^k))_+$ , then  $(0^T, \lambda_1(\mathcal{A}(x^k))_+)$  is the optimal solution of LSDP( $x^k$ ) (2.7), which implies that Algorithm A stops at Step 1. This is a contradiction. So it follows from the definition of directional derivative and (2.16) that

$$P'(x^k; d^k) = [\lambda_1(\cdot)_+]'(\mathcal{A}(x^k); D\mathcal{A}(x^k)d^k) \leq -(\lambda_1(\mathcal{A}(x^k))_+ - \lambda_1(z_k E_m)_+) < 0, \tag{3.11}$$

that is, the result (i) is true.

(ii) The proof is divided into two cases.

*Case A.*  $x^k$  is a feasible point of NLSDP (1.1). We obtain  $z_k = 0$  and  $\lambda_1(\mathcal{A}(x^k))_+ = 0$ , so by (2.14) and Lemma 2.2, we obtain

$$\begin{aligned} \theta'_{\alpha_{k+1}}(x^k; d^k) &\leq -(d^k)^T H_k d^k + \text{Tr}(\Lambda_k(\mathcal{A}(x^k))) \\ &\leq -(d^k)^T H_k d^k + \text{Tr}(\Lambda_k)\lambda_1(\mathcal{A}(x^k))_+ \\ &= -(d^k)^T H_k d^k < 0, \end{aligned}$$

the last inequality above is due to Assumption A2 and  $d^k \neq 0$ .

*Case B.*  $x^k$  is an infeasible solution of NLSDP (1.1). This implies  $\lambda_1(\mathcal{A}(x^k))_+ = \lambda_1(\mathcal{A}(x^k)) > 0$ .

Since Algorithm A does not stop at Step 1, we have  $z_k < \lambda_1(\mathcal{A}(x^k))$ . Therefore, it follows from (2.13), (3.1), and Assumption A2 that

$$\theta'_{\alpha_{k+1}}(x^k; d^k) \leq -(d^k)^T H_k d^k < 0.$$

(iii) If  $x^k$  is a feasible point of NLSDP (1.1), then it is obvious that  $z_k = \lambda_1(\mathcal{A}(x^k))_+ = 0$  is the optimal value of LSDP( $x^k$ ) (2.7). So  $\Delta(x^k, \alpha_{k+1}) = \nabla f(x^k)^T d^k$ . Note that  $d = 0$  is a feasible solution of QSDP( $x^k, H_k$ ) (2.8), so we get

$$\Delta(x^k, \alpha_{k+1}) = \nabla f(x^k)^T d^k \leq -(d^k)^T H_k d^k < 0.$$

If  $x^k$  is an infeasible point of NLSDP (1.1), then, according to the update rule (3.1) of  $\alpha_k$ , it is sufficient to consider the second part of (3.1). It follows from (3.2) and (3.1) that

$$\begin{aligned} \Delta(x^k, \alpha_{k+1}) &\leq \nabla f(x^k)^T d^k - \left( \frac{\nabla f(x^k)^T d^k + (d^k)^T H_k d^k}{\lambda_1(\mathcal{A}(x^k))_+ - z_k} + \eta_1 \right) (\lambda_1(\mathcal{A}(x^k))_+ - \lambda_1(z_k E_m)_+) \\ &\leq -(d^k)^T H_k d^k < 0. \end{aligned} \tag{3.12}$$

Further, the arc search of Algorithm A is valid. Hence, Algorithm A is well defined.  $\square$

### 4 Global convergence

Based on Theorem 3.1, in this section, without loss of generality, we suppose that the sequence  $\{x^k\}$  generated by Algorithm A is infinite. In what follows, we will prove that

$\{x^k\}$  is bounded under some appropriate conditions, and that any accumulation point  $x^*$  of  $\{x^k\}$  is either an infeasible stationary point, or a KKT point of NLSDP (1.1). To this end, the following additional assumptions are necessary.

**A 3** For any  $c > 0$ , the level set  $L_c := \{x \in \mathbb{R}^n \mid P(x) \leq c\}$  is bounded.

**A 4** For any feasible point  $x$  of NLSDP (1.1), MFCQ is satisfied at  $x$ , that is, there exists  $d \in \mathbb{R}^n$  such that

$$A(x) + DA(x)d < 0.$$

**Lemma 4.1** *Suppose that Assumptions A1–A3 hold, then the iterative sequence  $\{x^k\}$  is bounded.*

*Proof* One of the following situations occurs:

- (i) If there exists an integer  $k_1$  such that  $P(x^k) \leq \bar{P}$  for any  $k > k_1$ , then  $x^k \in L_{\bar{P}}$  for any  $k > k_1$ . So  $\{x^k\}$  is also bounded because  $L_{\bar{P}}$  is bounded.
- (ii) If there exists an integer  $k_2$  such that  $P(x^k) > \bar{P}$  for any  $k > k_2$ , then it follows from Step 5 that  $x^k \in L_{P(x^{k_2})}$  for any  $k > k_2$ . So  $\{x^k\}$  is also bounded because  $L_{P(x^{k_2})}$  is bounded.
- (iii) If both (i) and (ii) do not occur, i.e.,  $P(x^k) \leq \bar{P}$  and  $P(x^k) > \bar{P}$  occur infinitely, respectively, then there exists an index set  $\{k_j\}$  satisfying

$$P(x^{k_j}) \leq \bar{P}, \quad P(x^{k_{j+1}}) > \bar{P}, \quad \forall j \in \{1, 2, \dots\}. \tag{4.1}$$

So by arc search strategy, there exists an index set  $\{s_j\}$  associated with  $\{k_j\}$  such that

$$k_j < s_j < k_{j+1}, \quad P(x^{s_j}) > \bar{P}, \quad P(x^{s_{j+1}}) \leq \bar{P}, \quad \forall j \in \{1, 2, \dots\}.$$

For convenience, let  $N := \{1, 2, \dots\}$ ,  $N_j := \{k \mid k_j < k < k_{j+1}\}$ , then we obtain

$$\{k \in N \mid k > k_1\} = \bigcup_j \{N_j \cup \{k_2, k_3, k_4, \dots\}\}.$$

We know from (4.1) that  $\{x^{k_j}\} \subseteq L_{\bar{P}}$ , so  $\{x^{k_j}\}$  is bounded. Hence,  $\{x^k\}$  is bounded as long as we can prove that there exists  $\bar{c} > 0$  such that  $x^k \in L_{\bar{c}}, \forall j \in N, k \in N_j$ . Combining with the boundedness of  $\{x^{k_j}\}$  and Assumption A1, we get  $\{\nabla f(x^{k_j})\}$  is bounded, i.e., there exists  $M > 0$  such that  $\|\nabla f(x^{k_j})\| \leq M$  for any  $j \in N$ . In addition, it follows from LSDP( $x^k$ ) (2.7) and QSDP( $x^k, H_k$ ) (2.8) that  $\{d^{k_j}\}$  is bounded. In view of  $x^{k_{j+1}} = x^{k_j}$ , we get  $\{x^{k_{j+1}}\}$  is bounded. Further, one obtains  $\{P(x^{k_{j+1}})\}$  is bounded due to the continuity of  $P(x)$ . So there exists  $\bar{c} > 0$  such that  $P(x^{k_{j+1}}) \leq \bar{c}$ .

At last, by (4.1) and Step 5 in Algorithm A, one has

$$\bar{c} \geq P(x^{k_{j+1}}) \geq P(x^{k_{j+2}}) \geq \dots \geq P(x^{s_j}) \geq \bar{P}, \tag{4.2}$$

$$\bar{P} \geq P(x^{s_{j+1}}), \bar{P} \geq P(x^{s_{j+2}}), \dots, \bar{P} \geq P(x^{k_{j+1}}). \tag{4.3}$$

We can find  $k_j$  and  $k_{j+1}$  such that  $k \in (k_j, k_{j+1})$  for  $k \in N$ . So it follows from (4.2) and (4.3) that  $x^k \in L_{\bar{c}}$ , i.e.,  $\{x^k\}$  is bounded. □

**Lemma 4.2** *Suppose that Assumptions A1–A4 hold. If  $\alpha_k \rightarrow +\infty$ , then every accumulation point  $x^*$  of  $\{x^k\}$  generated by Algorithm A is an infeasible stationary point of NLSDP (1.1).*

*Proof* If  $\alpha_k \rightarrow +\infty$ , then it follows from (3.1) that the sequence  $\left\{ \frac{\nabla f(x^k)^T d^k + (d^k)^T H_k d^k}{\lambda_1(\mathcal{A}(x^k))_+ - z_k} \right\}$  diverges to  $+\infty$ .

By (2.15) and (2.18), we have

$$\begin{aligned} \frac{\nabla f(x^k)^T d^k + (d^k)^T H_k d^k}{\lambda_1(\mathcal{A}(x^k))_+ - z_k} &\leq \frac{\text{Tr}(\Lambda_k(\mathcal{A}(x^k) - z_k E_m))}{\lambda_1(\mathcal{A}(x^k) - z_k E_m)} \\ &\leq \frac{\text{Tr}(\Lambda_k) \lambda_1(\mathcal{A}(x^k) - z_k E_m)}{\lambda_1(\mathcal{A}(x^k) - z_k E_m)} \\ &= \text{Tr}(\Lambda_k). \end{aligned} \tag{4.4}$$

If  $x^*$  is a feasible point of NLSDP (1.1), then, by Assumption A4, we know that MFCQ is satisfied at  $x^*$ . Similar to the proof of Theorem 5.1 in [27], we obtain that the set  $\Omega$  of the KKT Lagrangian multipliers for QSDP( $x^*, H_*$ ) (2.8) is nonempty and bounded. Note that  $\Lambda_k \xrightarrow{\mathcal{K}} \Lambda_*$  and  $\Lambda_* \in \Omega$ , so  $\{\Lambda_k\}$  is bounded. Therefore it follows from (4.4) that  $\left\{ \frac{\nabla f(x^k)^T d^k + (d^k)^T H_k d^k}{\lambda_1(\mathcal{A}(x^k))_+ - z_k} \right\}$  is bounded. This is a contradiction. Hence,  $x^*$  is an infeasible point, i.e.,  $\lambda_1(\mathcal{A}(x^*)) > 0$ . Further, it is obvious that  $(0, \lambda_1(\mathcal{A}(x^*)))$  is a feasible solution of LSDP( $x^*$ ) (2.7), so  $z_* \leq \lambda_1(\mathcal{A}(x^*))$ . Let  $(d^*, z_*)$  be an optimal solution of LSDP( $x^*$ ) (2.7), then by the constraint of LSDP( $x^*$ ) (2.7), we have

$$\lambda_1(\mathcal{A}(x^*) + D\mathcal{A}(x^*)d^*) \leq \lambda_1(\mathcal{A}(x^*)),$$

further,

$$\max\{\lambda_1(\mathcal{A}(x^*) + D\mathcal{A}(x^*)d^*), 0\} \leq \lambda_1(\mathcal{A}(x^*)).$$

Therefore, we get

$$\min_{d \in \mathbb{R}^n} \max\{\lambda_1(\mathcal{A}(x^*) + D\mathcal{A}(x^*)d), 0\} \leq \max\{\lambda_1(\mathcal{A}(x^*) + D\mathcal{A}(x^*)d^*), 0\} \leq \lambda_1(\mathcal{A}(x^*)).$$

Let  $d = 0$ , then  $\max\{\lambda_1(\mathcal{A}(x^*) + D\mathcal{A}(x^*)d), 0\} = \lambda_1(\mathcal{A}(x^*))$ , which together with the above inequality implies

$$\min_{d \in \mathbb{R}^n} \max\{\lambda_1(\mathcal{A}(x^*) + D\mathcal{A}(x^*)d), 0\} = \lambda_1(\mathcal{A}(x^*)),$$

that is,  $x^*$  is an infeasible stationary point of NLSDP (1.1). □

In the rest of the paper, we assume  $\alpha_k < +\infty$ . According to the update rule (3.1), the following conclusion is shown easily.

**Lemma 4.3** *Suppose that Assumptions A1–A4 hold. Then there exists an integer  $k_0$  such that  $\alpha_k \equiv \alpha_{k_0} \triangleq \alpha > 0$  for any  $k \geq k_0$ .*

Based on Lemma 4.3, in the rest of the paper, without loss of generality, we assume that  $\alpha_k \equiv \alpha, k = 1, 2, \dots$

**Lemma 4.4** *Suppose that Assumptions A1–A2 hold,  $x^k \rightarrow x^*, H_k \rightarrow H_*$ . Then  $z_k \rightarrow z_*, d^k \rightarrow d^*$ , where  $z_k, z_*$  are the optimal solutions of LSDP( $x^k$ ) (2.7) and LSDP( $x^*$ ) (2.7), respectively, and  $d^k, d^*$  are the optimal solutions of QSDP( $x^k, H_k$ ) (2.8) and QSDP( $x^*, H_*$ ) (2.8), respectively.*

*Proof* Since  $z_k$  is the optimal value of LSDP( $x^k$ ) (2.7), we obtain  $z_k < \lambda_1(\mathcal{A}(x^k))_+$  due to the fact that  $(0, \lambda_1(\mathcal{A}(x^k))_+)$  is a feasible solution of LSDP( $x^k$ ) (2.7). By the boundedness of  $\{\lambda_1(\mathcal{A}(x^k))\}$  and  $z_k > 0$ , it is true that  $\{z_k\}$  is bounded. According to the sensitivity theory of semidefinite programming in [28], we know that the first part of the conclusions is true.

Now consider the second part of the conclusions. We first prove  $\{d^k\}$  is bounded. It follows from LSDP( $x^k$ ) (2.7) that  $\|\widehat{d}^k\| \leq 1$ . And obviously,  $\widehat{d}^k$  is a feasible solution of QSDP( $x^k, H_k$ ) (2.8), so one obtains

$$\nabla f(x^k)^T d^k + \frac{1}{2}(d^k)^T H_k d^k \leq \nabla f(x^k)^T \widehat{d}^k + \frac{1}{2}(\widehat{d}^k)^T H_k \widehat{d}^k,$$

further, the above inequality gives rise to

$$\nabla f(x^k)^T d^k + \frac{1}{2}(d^k)^T H_k d^k \leq \|\nabla f(x^k)\| \|\widehat{d}^k\| + \frac{1}{2}\|\widehat{d}^k\|^2 \bar{a} \leq M_1 + \frac{1}{2}\bar{a}.$$

On the other hand, one has

$$\nabla f(x^k)^T d^k + \frac{1}{2}(d^k)^T H_k d^k \geq -\|\nabla f(x^k)\| \|d^k\| + a \|d^k\|^2 \geq -M_1 \|d^k\| + a \|d^k\|^2.$$

The two inequalities above indicate that  $\{d^k\}$  is bounded.

Suppose that  $d^k \not\rightarrow d^*$ , then there exists a subsequence  $\{d^s\}_{K_1} \subseteq \{d^k\}$  converging to  $\bar{d}$  ( $\neq d^*$ ). For any feasible solution  $d$  of QSDP( $x^*, H_*$ ) (2.8), since  $z_k \rightarrow z_*$ , there exists a feasible solution  $d^m$  of QSDP( $x^s, H_s$ ) (2.8) such that

$$d^m \xrightarrow{K_1} d.$$

Since  $d^s$  is the solution of QSDP( $x^s, H_s$ ) (2.8), one has

$$\nabla f(x^s)^T d^s + \frac{1}{2}(d^s)^T H_s d^s \leq \nabla f(x^s)^T d^m + \frac{1}{2}(d^m)^T H_s d^m.$$

Let  $s \xrightarrow{K_1} \infty, m \xrightarrow{K_1} \infty$ , one gets

$$\nabla f(x^*)^T \bar{d} + \frac{1}{2}\bar{d}^T H_* \bar{d} \leq \nabla f(x^*)^T d + \frac{1}{2}d^T H_* d,$$

which means that  $\bar{d}$  is a solution of QSDP( $x^*, H_*$ ) (2.8). This contradicts the uniqueness of the solution of QSDP( $x^*, B_*$ ) (2.8). □

**Lemma 4.5** *Suppose that Assumptions A1–A4 hold,  $x^*$  is an accumulation point of the sequence  $\{x^k\}$  generated by Algorithm A, i.e.,  $x^k \xrightarrow{K} x^*$ . If  $x^*$  is not an infeasible stationary point of NLSDP (1.1), then  $d^k \xrightarrow{K} 0$ .*

*Proof* By contradiction, suppose that  $d^k \not\xrightarrow{K} 0$ , then there exist a constant  $b > 0$  and an index subset  $K' \subseteq K$  such that

$$\|d^k\| \geq b > 0 \tag{4.5}$$

for any  $k \in K'$ . The following proof is divided into two steps.

*Step A.* We first prove  $\underline{t} := \inf\{t_k, k \in K'\} > 0$ .

By Taylor expansion and the boundedness of the sequence  $\{d^k\}$ , one has

$$\begin{aligned} f(x^k + td^k + t^2\tilde{d}^k) &= f(x^k) + t\nabla f(x^k)^T d^k + o(t), \\ \mathcal{A}(x^k + td^k + t^2\tilde{d}^k) &= \mathcal{A}(x^k) + t \sum_{i=1}^n d_i^k \frac{\partial \mathcal{A}(x^k)}{\partial x_i} + o(t). \end{aligned} \tag{4.6}$$

In view of  $t \leq 1$ , combining with the convexity of  $\lambda_1(\cdot)_+$  and QSDP( $x^k, H_k$ ) (2.8), one obtains

$$\begin{aligned} &\lambda_1(\mathcal{A}(x^k + td^k + t^2\tilde{d}^k))_+ \\ &\leq (1-t)\lambda_1(\mathcal{A}(x^k))_+ + t\lambda_1\left(\mathcal{A}(x^k) + \sum_{i=1}^n d_{k_i} \frac{\partial \mathcal{A}(x^k)}{\partial x_i}\right)_+ + o(t) \\ &\leq (1-t)\lambda_1(\mathcal{A}(x^k))_+ + t\lambda_1(z_k E_m)_+ + o(t), \end{aligned} \tag{4.7}$$

which together with (2.12), (4.6), and (4.7) gives

$$\begin{aligned} &\theta_\alpha(x^k + td^k + t^2\tilde{d}^k) \\ &\leq f(x^k) + t\nabla f(x^k)^T d^k + o(t) + \alpha[(1-t)\lambda_1(\mathcal{A}(x^k))_+ + t\lambda_1(z_k E_m)_+ + o(t)] \\ &= f(x^k) + \alpha\lambda_1(\mathcal{A}(x^k))_+ + t(\nabla f(x^k)^T d^k - \alpha(\lambda_1(\mathcal{A}(x^k))_+ - \lambda_1(z_k E_m)_+)) + o(t) \\ &= \theta_\alpha(x^k) + t\Delta(x^k, \alpha) + o(t), \end{aligned}$$

so we obtain

$$\theta_\alpha(x^k + td^k + t^2\tilde{d}^k) - \theta_\alpha(x^k) - \beta t\Delta(x^k, \alpha) \leq (1-\beta)t\Delta(x^k, \alpha) + o(t). \tag{4.8}$$

By Lemma 4.4 and (3.12), we get

$$\Delta(x^k, \alpha) \leq -(d^k)^T H_k d^k \longrightarrow -(d^*)^T H_* d^* < 0 \quad \text{as } k (\in K') \longrightarrow \infty,$$

so we have

$$\Delta(x^k, \alpha) \leq -0.5(d^*)^T H_* d^*$$

for  $k \in K'$  sufficiently large. Substituting the above inequality into (4.8) gives

$$\theta_\alpha(x^k + td^k + t^2\tilde{d}^k) - \theta_\alpha(x^k) - \beta t \Delta(x^k, \alpha) \leq -0.5(1 - \beta)t(d^*)^T H_* d^* + o(t), \tag{4.9}$$

which means that, for  $k \in K'$  sufficiently large and  $t$  sufficiently small, inequality (3.3) or the first inequality of (3.4) holds.

In what follows, we consider the second inequality of (3.4).

Note that  $P(x^k) > \bar{P} > 0$ , so  $P(x^*) = \lim_{K'} P(x^k) \geq \bar{P} > 0$ , which means  $x^*$  is an infeasible solution of NLSDP (1.1). Since  $x^*$  is not an infeasible stationary point of NLSDP (1.1), it follows that  $z_* - \lambda_1(\mathcal{A}(x^*))_+ < 0$ . Further, we have

$$\lambda_1(z_k E_m)_+ - \lambda_1(\mathcal{A}(x^k))_+ \rightarrow z_* - \lambda_1(\mathcal{A}(x^*))_+ < 0 \quad \text{as } k \in K' \rightarrow \infty,$$

so it follows that, for  $k \in K'$  sufficiently large,

$$\lambda_1(z_k E_m)_+ - \lambda_1(\mathcal{A}(x^k))_+ < 0.5(z_* - \lambda_1(\mathcal{A}(x^*))_+).$$

By (4.7), (2.11), and the above inequality, one has

$$\begin{aligned} P(x^k + td^k + t^2\tilde{d}^k) &\leq P(x^k) + t(\lambda_1(z_k E_m)_+ - \lambda_1(\mathcal{A}(x^k))_+) + o(t) \\ &\leq P(x^k) + 0.5t(z_* - \lambda_1(\mathcal{A}(x^*))_+) + o(t), \end{aligned}$$

equivalently,

$$P(x^k + td^k + t^2\tilde{d}^k) - P(x^k) \leq 0.5t(z_* - \lambda_1(\mathcal{A}(x^*))_+) + o(t), \tag{4.10}$$

which implies that, for  $k \in K'$  sufficiently large and  $t$  sufficiently small, the second inequality in (3.4) holds.

Summarizing the analysis above, we can conclude  $\underline{t} := \inf\{t_k, k \in K'\} > 0$ .

*Step B.* Based on  $\underline{t} = \inf\{t_k, k \in K'\} > 0$ , we prove a contradiction will occur.

It follows from (3.3) or (3.4) that  $\{\theta_\alpha(x^k)\}$  is nonincreasing and

$$\theta_\alpha(x^{k+1}) \leq \theta_\alpha(x^k) - 0.5ab^2\beta\underline{t} \tag{4.11}$$

for any  $k \in K'$ , where  $b$  is defined in (4.5). And one obtains from (2.12) that

$$\theta_\alpha(x^k) = f(x^k) + \alpha(\lambda_1(\mathcal{A}(x^k))_+) \geq f(x^k),$$

combining with the boundedness of  $\{f(x^k)\}$ , we conclude that  $\{\theta_\alpha(x^k)\}$  is convergent. Taking  $k \xrightarrow{K'} \infty$  in (4.11), we obtain  $-0.5ab^2\beta\underline{t} \geq 0$ . This is a contradiction. So  $\lim_K d^k = 0$ .  $\square$

Based on the above results, we are now in a position to present the global convergence of Algorithm A.

**Theorem 4.1** *Suppose that Assumptions A1–A4 hold,  $x^*$  is an accumulation point of the sequence  $\{x^k\}$  generated by Algorithm A. Then either  $x^*$  is an infeasible stationary point, or a KKT point of NLSDP (1.1).*

*Proof* Without loss of generality, we suppose that  $x^*$  is not an infeasible stationary point of NLSDP (1.1). In what follows, we show that  $x^*$  is a KKT point of NLSDP (1.1).

By Lemmas 4.4–4.5, we know that  $d^* = 0$  is an optimal solution of QSDP( $x^*, H_*$ ) (2.8), so it follows from Lemma 2.5 that there exists  $\Lambda_* \in \mathbb{S}_+^m$  such that

$$\nabla f(x^*) + D\mathcal{A}(x^*)^* \Lambda_* = 0, \tag{4.12a}$$

$$\mathcal{A}(x^*) \preceq z_* E_m, \tag{4.12b}$$

$$\text{Tr}(\Lambda_* (\mathcal{A}(x^*) - z_* E_m)) = 0. \tag{4.12c}$$

Now we prove  $z_* = 0$ . By contradiction, if  $z_* \neq 0$ , then  $z_* > 0$ . Since  $(0, \lambda_1(\mathcal{A}(x^*))_+)$  is a feasible solution of LSDP( $x^*$ ) (2.7), we get  $\lambda_1(\mathcal{A}(x^*))_+ \geq z_* > 0$ , which implies  $x^*$  is an infeasible point of NLSDP (1.1). On the other hand, we get  $z_* \geq \lambda_1(\mathcal{A}(x^*))_+ > 0$  by (4.12b), so  $z_* = \lambda_1(\mathcal{A}(x^*))_+$ . Obviously,  $(0, z_* = \lambda_1(\mathcal{A}(x^*))_+)$  is an optimal solution of LSDP( $x^*$ ) (2.7). In a manner similar to the proof of Theorem 3.1, we can conclude that  $x^*$  is an infeasible stationary point of NLSDP (1.1), this is a contradiction.

Substituting  $z_* = 0$  into (4.12a)–(4.12c), one obtains

$$\begin{aligned} \nabla f(x^*) + D\mathcal{A}(x^*)^* \Lambda_* &= 0, \\ \mathcal{A}(x^*) &\leq 0, \quad \text{Tr}(\Lambda_* \mathcal{A}(x^*)) = 0, \end{aligned}$$

which means that  $x^*$  is a KKT point of NLSDP (1.1). □

### 5 Superlinear convergence

In this section, we analyze the superlinear convergence of Algorithm A. At first, we prove that a full step can be accepted for  $k$  sufficiently large, and then we present the superlinear convergence. To this end, Assumption A1 should be strengthened as the following one:

**A 5** The functions  $f(x)$  and  $\mathcal{A}(x)$  are twice continuously differentiable.

Besides, the following assumptions are necessary:

**A 6** ([20]) The sequence  $\{x^k\}$  generated by Algorithm A is an infinite sequence, and  $\lim_{k \rightarrow \infty} x^k = x^*$ , where  $x^*$  is a KKT point of NLSDP (1.1). In addition, let  $\Lambda_*$  be the corresponding Lagrangian multiplier.

**A 7** ([29]) The constrained nondegeneracy condition holds at  $(x^*, \Lambda_*)$ .

Denote

$$Y_{ij}^k = \left( (p_i^k)^T \frac{\partial \mathcal{A}(x^k)}{\partial x_1} p_j^k, (p_i^k)^T \frac{\partial \mathcal{A}(x^k)}{\partial x_2} p_j^k, \dots, (p_i^k)^T \frac{\partial \mathcal{A}(x^k)}{\partial x_n} p_j^k \right)^T,$$

where  $1 \leq i \leq j \leq m-r$ ,  $\{p_1^k, \dots, p_{m-r}^k\}$  is an orthogonal basis, which is introduced in Sect. 2. And let  $\lim_{k \rightarrow \infty} Y_{ij}^k = Y_{ij}^*$ . The constrained nondegeneracy condition (i.e., Assumption A7) is equivalent to the fact that the matrix  $(Y_{11}^*, Y_{12}^*, Y_{22}^*, Y_{13}^*, Y_{23}^*, Y_{33}^*, \dots, Y_{m-r, m-r}^*)$  is full of column rank, which implies that  $\Lambda_*$  is unique. Based on this result, we know that subproblem (2.9) has a unique solution.

**A 8** ([29]) The strong second-order sufficient condition holds at  $x^*$ , i.e.,

$$d^T \nabla_{xx} L(x^*, \Lambda_*) d + \Gamma_{\mathcal{A}(x^*)}(\Lambda_*, D\mathcal{A}(x^*)d) > 0$$

for any  $d \in \text{app}(\Lambda_*) \setminus \{0\}$ , where  $\text{app}(\Lambda_*) := \{d \mid D\mathcal{A}(x^*)d \in \text{aff}(\mathcal{C}(\mathcal{A}(x^*) + \Lambda_*; \mathbb{S}_-^m))\}$ ,  $\Gamma_A(B, C) := -2\langle B, CA^\dagger C \rangle$ ,  $A^\dagger$  is the Moore–Penrose pseudoinverse of  $A$ .

**A 9** The strict complementarity condition is satisfied at  $(x^*, \Lambda_*)$ , i.e.,

$$\text{rank}(\mathcal{A}(x^*)) = r, \quad \text{rank}(\Lambda_*) = m - r.$$

**A 10**  $\|(W_k - H_k)d^k\| = o(\|d^k\|)$ , where

$$W_k = \nabla_{xx} L(x^k, \Lambda_k) = \nabla^2 f(x^k) + \nabla_{xx}^2 \langle \Lambda_k, \mathcal{A}(x^k) \rangle,$$

$$\nabla_{xx}^2 \langle \Lambda_k, \mathcal{A}(x^k) \rangle = \begin{pmatrix} \langle \Lambda_k, \frac{\partial^2 \mathcal{A}(x^k)}{\partial x_1 \partial x_1} \rangle & \dots & \langle \Lambda_k, \frac{\partial^2 \mathcal{A}(x^k)}{\partial x_1 \partial x_n} \rangle \\ \dots & \dots & \dots \\ \langle \Lambda_k, \frac{\partial^2 \mathcal{A}(x^k)}{\partial x_n \partial x_1} \rangle & \dots & \langle \Lambda_k, \frac{\partial^2 \mathcal{A}(x^k)}{\partial x_n \partial x_n} \rangle \end{pmatrix}.$$

Based on Assumptions A5–A6, we know that  $z_k = 0$  for sufficiently large  $k$ , so for sufficiently large  $k$ , the two subproblems LSDP (2.7) and QSDP (2.8) can be replaced by the following subproblem:

$$\begin{aligned} \min \quad & \nabla f(x^k)^T d + \frac{1}{2} d^T H_k d \\ \text{s.t.} \quad & \mathcal{A}(x^k) + D\mathcal{A}(x^k)d \leq 0. \end{aligned} \tag{5.1}$$

The following conclusions are the results in [20], which are important for the analysis of the acceptance of full step size (i.e., Lemma 5.2).

**Lemma 5.1** *Suppose that Assumptions A2–A10 hold, then the following conclusions are true:*

- (i)  $\lim_{k \rightarrow \infty} \Lambda_k = \Lambda_*$ .
- (ii)  $\text{rank}(\mathcal{A}(x^k) + D\mathcal{A}(x^k)d^k) = \text{rank}(\mathcal{A}(x^*)) = r$  for all  $k$  sufficiently large.
- (iii) If  $\tilde{d}^k$  is a solution to subproblem (2.9), then there exists  $\widehat{\Phi}_k \in \mathbb{S}^q$  such that

$$\nabla f(x^k) + H_k d_k + H_k \tilde{d}^k + D\mathcal{A}(x^k)^* \tilde{N}_k^T \widehat{\Phi}_k \tilde{N}_k = 0. \tag{5.2}$$

- (iv)  $\|\tilde{d}^k\| = O(\|d^k\|^2)$ ,  $\|\Phi_k - \widehat{\Phi}_k\| = O(\|d^k\|^2)$  for all  $k$  sufficiently large, where  $\Phi_k$  satisfies  $\Lambda_k = \tilde{N}_k \Phi_k \tilde{N}_k^T$ .
- (v)  $\lambda_1(\mathcal{A}(x^k) + d^k + \tilde{d}^k) \leq 0$  holds for all  $k$  sufficiently large.

Based on Lemma 5.1, we are now in a position to show that the full size step can be accepted for  $k$  sufficiently large, which plays a key role for the superlinear convergence.

**Lemma 5.2** *Suppose that Assumptions A2–A10 hold, then the full step size can be accepted, i.e.,  $t_k \equiv 1$  for  $k$  sufficiently large.*



*Proof* By Lemma 5.1(v), we know  $\lambda_1(\mathcal{A}(x^k + d^k + \tilde{d}^k)) \leq 0$  for all  $k$  sufficiently large, this implies  $P(x^k) \leq \bar{P}$  for all  $k$  sufficiently large. So according to the arc search strategy (see Step 5 in Algorithm A), it is sufficient to prove

$$\theta_\alpha(x^k + d^k + \tilde{d}^k) \leq \theta_\alpha(x^k) + \beta \Delta_k$$

for  $k$  sufficiently large, or equivalently,

$$T_k := \theta_\alpha(x^k + d^k + \tilde{d}^k) - \theta_\alpha(x^k) - \beta \Delta_k \leq 0 \tag{5.3}$$

for  $k$  sufficiently large.

By (2.12), Taylor expansion, and Lemma 5.1(iv), (v), we obtain

$$\begin{aligned} &\theta_\alpha(x^k + d^k + \tilde{d}^k) - \theta_\alpha(x^k) \\ &= \nabla f(x^k)^\top d^k + \nabla f(x^k)^\top \tilde{d}^k + \frac{1}{2}(d^k)^\top \nabla^2 f(x^k) d^k - \alpha P(x^k) + o(\|d^k\|^2). \end{aligned} \tag{5.4}$$

It follows from the constraints of subproblem (2.9) that

$$\bar{N}_k^\top (\mathcal{A}(x^k + d^k)) \bar{N}_k = -\bar{N}_k^\top (D\mathcal{A}(x^k) \tilde{d}^k) \bar{N}_k + o(\|d^k\|^2), \tag{5.5}$$

which gives rise to

$$-\langle D\mathcal{A}(x^k) \tilde{d}^k, \bar{N}_k \hat{\Phi}_k \bar{N}_k^\top \rangle = \langle \mathcal{A}(x^k + d^k), \bar{N}_k \hat{\Phi}_k \bar{N}_k^\top \rangle + o(\|d^k\|^2).$$

By (5.2), (2.2), and the above equality, one has

$$\begin{aligned} \nabla f(x^k)^\top \tilde{d}^k &= -\langle D\mathcal{A}(x^k) \tilde{d}^k, \bar{N}_k \hat{\Phi}_k \bar{N}_k^\top \rangle - (d^k + \tilde{d}^k)^\top B_k \tilde{d}^k \\ &= \langle \mathcal{A}(x^k + d^k), \bar{N}_k \hat{\Phi}_k \bar{N}_k^\top \rangle + o(\|d^k\|^2) \\ &= \langle \bar{N}_k^\top \mathcal{A}(x^k + d^k) \bar{N}_k, \hat{\Phi}_k \rangle + o(\|d^k\|^2) \\ &= \langle \bar{N}_k^\top \mathcal{A}(x^k + d^k) \bar{N}_k, \hat{\Phi}_k - \Phi_k \rangle \\ &\quad + \langle \bar{N}_k^\top \mathcal{A}(x^k + d^k) \bar{N}_k, \Phi_k \rangle + o(\|d^k\|^2) \\ &= \langle \mathcal{A}(x^k + d^k), \bar{N}_k \Phi_k \bar{N}_k^\top \rangle + o(\|d^k\|^2), \end{aligned}$$

where the last equality is due to Lemma 5.1(iv) and (5.5).

Note that  $\bar{N}_k \Phi_k \bar{N}_k^\top = \Lambda_k$ , by Taylor expansion and (2.10c), we get

$$\begin{aligned} \nabla f(x^k)^\top \tilde{d}^k &= \left\langle \mathcal{A}(x^k) + D\mathcal{A}(x^k) d^k + \frac{1}{2}(d^k)^\top D^2 \mathcal{A}(x^k) d^k, \Lambda_k \right\rangle + o(\|d^k\|^2) \\ &= \left\langle \frac{1}{2}(d^k)^\top D^2 \mathcal{A}(x^k) d^k, \Lambda_k \right\rangle + o(\|d^k\|^2) \\ &= \frac{1}{2}(d^k)^\top \nabla_{xx}^2 \langle \Lambda_k, \mathcal{A}(x^k) \rangle d^k + o(\|d^k\|^2). \end{aligned} \tag{5.6}$$

By (5.3), (5.4), and (5.6), we have

$$\begin{aligned}
 T_k &= \nabla f(x^k)^T d^k - \alpha P(x^k) + \nabla f(x^k)^T \tilde{d}^k + \frac{1}{2} (d^k)^T \nabla^2 f(x^k) d^k \\
 &\quad + o(\|d^k\|^2) - \beta \Delta_k \\
 &= \nabla f(x^k)^T d^k - \alpha P(x^k) + \frac{1}{2} (d^k)^T \nabla_{xx}^2 \langle \Lambda_k, \mathcal{A}(x^k) \rangle d^k \\
 &\quad + \frac{1}{2} (d^k)^T \nabla^2 f(x^k) d^k - \beta \Delta_k + o(\|d^k\|^2).
 \end{aligned}
 \tag{5.7}$$

It follows from (3.2) and (3.12) that

$$\nabla f(x^k)^T d^k - \alpha P(x^k) = \Delta_k \leq - (d^k)^T H_k d^k.
 \tag{5.8}$$

Noting that  $\frac{1}{2} - \beta > 0$ , it follows from (5.7), (5.8), and Assumption A2 that

$$\begin{aligned}
 T_k &\leq \left(\frac{1}{2} - \beta\right) \Delta_k + \frac{1}{2} (d^k)^T (W_k - H_k) d^k + o(\|d^k\|^2) \\
 &\leq -\left(\frac{1}{2} - \beta\right) a \|d^k\|^2 + \frac{1}{2} (d^k)^T (W_k - H_k) d^k + o(\|d^k\|^2),
 \end{aligned}$$

which together with Assumption A10 gives

$$T_k \leq -\left(\frac{1}{2} - \beta\right) a \|d^k\|^2 + o(\|d^k\|^2) \leq 0$$

for  $k$  sufficiently large. So we get the conclusion. □

Based on Lemma 5.2, we now present the superlinear convergence of Algorithm A. The proof is similar to that of Theorem 3.3 in [17].

**Theorem 5.1** *Suppose that Assumptions A2–A10 hold, the sequence  $\{x^k\}$  generated by Algorithm A is superlinearly convergent, that is,  $\|x^{k+1} - x^*\| = o(\|x^k - x^*\|)$ .*

### 6 Numerical experiments

In this section, preliminary numerical experiments of Algorithm A are implemented. The tested problems are chosen from [13, 30]. Algorithm A was coded by Matlab (2017a) and run on a computer with 3.60 GHz CPU with Windows 7 (64 bite) system.

The parameters are chosen as follows:

$$\alpha_0 = 80.1, \quad \bar{P} = 5, \quad \eta_1 = 0.1, \quad \eta_2 = 0.2, \quad \beta = 0.4;$$

The stop criterion is:  $\|d^k\| \leq 10^{-6}$ .

**Problem 1 ([30])**

$$\begin{aligned}
 &\min f(x) = \sin x_1 + \cos x_2 \\
 &\text{s.t. } \begin{pmatrix} x_1 & 1 \\ 1 & x_2 \end{pmatrix} \preceq 0, \\
 &x = (x_1, x_2)^T \in \mathbb{R}^2.
 \end{aligned} \tag{6.1}$$

**Problem 2 ([30])**

$$\begin{aligned}
 &\min f(x) = e^{-x_1-x_2} \\
 &\text{s.t. } \begin{pmatrix} x_1 & 1 \\ 1 & x_2 \end{pmatrix} \preceq 0, \\
 &x = (x_1, x_2)^T \in \mathbb{R}^2.
 \end{aligned} \tag{6.2}$$

For the above two tested problems, we compare Algorithm A with the algorithm in [30]. The numerical results are listed in Table 1. The meaning of the notations in Table 1 is described as follows:

$x^0$  : the initial point,      Iter : the number of iterations,  
 time : the CPU time,       $f_{\text{final}}$  : the final objective value.

The numerical results in Table 1 indicate that Algorithm A is much more robust than Algorithm [30], although the CPU time of Algorithm A is more than that of Algorithm [30]. The less time of Algorithm [30] is due to the fact that Algorithm [30] only solves a subproblem at each iteration.

**Table 1** Numerical results

Problem	$n$	$m$	$x^0$	Algorithm A			Algorithm [30]		
				Iter	time (s)	$f_{\text{final}}$	Iter	time (s)	$f_{\text{final}}$
Problem 1	2	2	(1, 2)	9	1.43	-2.000	10	1.12	-2.000
	2	2	(-5, 4)	9	1.24	-2.000	11	1.08	-2.000
	2	2	(-4, -2)	10	1.04	-2.000	10	1.01	-2.000
	2	2	(-4, 5)	9	1.16	-2.000	9	1.26	-2.000
	2	2	(10, -10)	7	1.36	-2.000	4	0.32	-1.102
	2	2	(2, 1)	11	1.35	-2.000	4	0.30	-2.244
	2	2	(1, -10)	8	0.91	-2.000	3	0.45	-1.891
Problem 2	2	2	(1, 1)	2	0.68	7.389	3	0.35	7.389
	2	2	(1, -7)	5	0.93	7.389	5	0.79	7.389
	2	2	(9, -11)	6	0.91	7.389	6	0.81	7.389
	2	2	(1, 5)	6	1.43	7.389	9	1.12	7.389
	2	2	(1, 4)	5	0.72	7.389	4	1.04	5.431
	2	2	(0, 4)	6	0.78	7.389	4	0.58	8.175
	2	2	(0, 2)	5	0.61	7.389	3	0.43	7.842

**Table 2** Numerical results of NCM

$n$	$m$	$x^0$	Algorithm	A-Iter
10	5	$(0.5, \dots, 0.5)^T$	Algorithm A	5
			Algorithm in [31]	8
			Algorithm in [13]	9
45	10	$(0.5, \dots, 0.5)^T$	Algorithm A	15
			Algorithm in [31]	8
			Algorithm in [13]	10
105	15	$(0.5, \dots, 0.5)^T$	Algorithm A	4
			Algorithm in [31]	10
			Algorithm in [13]	11
190	20	$(0.5, \dots, 0.5)^T$	Algorithm A	2
			Algorithm in [31]	11
			Algorithm in [13]	12

**Problem 3** ([13]) The nearest correlation matrix (NCM) problem:

$$\begin{aligned}
 \min \quad & f(X) = \frac{1}{2} \|X - C\|_F^2 \\
 \text{s.t.} \quad & X \preceq \epsilon I, \\
 & X_{ii} = 1, \quad i = 1, 2, \dots, m,
 \end{aligned} \tag{6.3}$$

where  $C \in \mathbb{S}^m$  is a given matrix,  $X \in \mathbb{S}^m$ ,  $\epsilon$  is a scalar.

In the implementation,  $\epsilon = 10^{-3}$ , the matrix  $C$  is generated randomly, and its diagonal elements are 1. We test ten times for every fixed dimensionality.

The numerical results are listed in Table 2. The meaning of the notations in Table 2 is described as follows:

$n$  : the dimensionality of  $x$ ;                       $m$  : the order of the matrix  $\mathcal{A}(x)$ ;  
 A-Iter : the average iterative number.

The numerical results in Table 2 indicate that the average iterative number of Algorithm A is less than that of the other two algorithms. Hence, Algorithm A is comparable.

### 7 Concluding remarks

In this paper, we have presented a new SSDP algorithm for nonlinear semidefinite programming. Two subproblems, which are constructed skillfully, are solved to generate the master search directions. In order to avoid the Maratos effect, a second-order correction direction is introduced by solving a new quadratic programming. A penalty function is used as a merit function for arc search. The global convergence and superlinearly convergence of the proposed algorithm are shown under some mild conditions. The preliminary numerical results indicate that the proposed algorithm is effective and comparable.

#### Funding

The work is supported by the National Natural Science Foundation (No.11561005) and the Science Foundation of Guangxi Province (No. 2016GXNSFAA380248).

#### Availability of data and materials

Not applicable

**Competing interests**

The authors declare that they have no competing interests.

**Authors' contributions**

The authors completed the paper. The authors read and approved the final manuscript.

**Authors' information**

Jianling Li, Email: jianlingli@126.com; Hui Zhang, Email: 605007820@qq.com

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 19 March 2019 Accepted: 7 August 2019 Published online: 20 August 2019

**References**

1. Tuyen, N.V., Yao, J.C., Wen, C.F.: A note on approximate Karush–Kuhn–Tucker conditions in locally Lipschitz multiobjective optimization. *Optim. Lett.* **13**, 163–174 (2019)
2. Chang, S.S., Wen, C.F., Yao, J.C.: Common zero point for a finite family of inclusion problems of accretive mappings in Banach spaces. *Optimization* **67**, 1183–1196 (2018)
3. Takahashi, W., Wen, C.F., Yao, J.C.: The shrinking projection method for a finite family of demimetric mappings with variational inequality problems in a Hilbert space. *Fixed Point Theory* **19**, 407–419 (2018)
4. Kanno, Y., Takewaki, I.: Sequential semidefinite program for maximum robustness design of structures under load uncertainty. *J. Optim. Theory Appl.* **130**(2), 265–287 (2006)
5. Jarre, F.: An interior point method for semidefinite programming. *Optim. Eng.* **1**, 347–372 (2000)
6. Ben, T.A., Jarre, F., Kovara, M., Nemirovski, A., Zowe, J.: Optimization design of trusses under a nonconvex global buckling constraint. *Optim. Eng.* **1**, 189–213 (2000)
7. Noll, D., Torki, M., Apkarian, P.: Partially augmented Lagrangian method for matrix inequality constraints. *SIAM J. Optim.* **15**(1), 161–184 (2001)
8. Sun, J., Zhang, L.W., Wu, Y.: Properties of the augmented Lagrangian in nonlinear semidefinite optimization. *J. Optim. Theory Appl.* **129**(3), 437–456 (2006)
9. Sun, D.F., Sun, J., Zhang, L.W.: The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. *Math. Program.* **114**(2), 349–391 (2008)
10. 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)
11. Wu, H.X., Luo, H.Z., Ding, X.D.: Global convergence of modified augmented Lagrangian methods for nonlinear semidefinite programming. *Comput. Optim. Appl.* **56**(3), 531–558 (2013)
12. Wu, H.X., Luo, H.Z., Yang, J.F.: Nonlinear separation approach for the augmented Lagrangian in nonlinear semidefinite programming. *J. Glob. Optim.* **59**(4), 695–727 (2014)
13. Yamashita, H., Yabe, H., Harada, K.: A primal-dual interior point method for nonlinear semidefinite programming. *Math. Program.* **135**, 89–121 (2012)
14. Yamashita, H., Yabe, H.: Local and superlinear convergence of a primal-dual interior point method for nonlinear semidefinite programming. *Math. Program.* **132**, 1–30 (2012)
15. Fares, B., Noll, D., Apkarian, P.: Robust control via sequential semidefinite programming. *SIAM J. Control Optim.* **40**, 1791–1820 (2002)
16. Correa, R., Ramirez, H.C.: A global algorithm for nonlinear semidefinite programming. *SIAM J. Optim.* **15**(1), 303–318 (2004)
17. Wang, Y., Zhang, S.W., Zhang, L.W.: A note on convergence analysis of an SQP-type method for nonlinear semidefinite programming. *J. Inequal. Appl.* **2008**, 218345 (2007)
18. Gomez, W., Ramirez, H.: A filter algorithm for nonlinear semidefinite programming. *Appl. Math. Comput.* **29**(29), 297–328 (2010)
19. Zhu, Z.B., Zhu, H.L.: A filter method for nonlinear semidefinite programming with global convergence. *Acta Math. Sin.* **30**(10), 1810–1826 (2014)
20. Zhao, Q., Chen, Z.W.: On the superlinear local convergence of a penalty-free method for nonlinear semidefinite programming. *J. Comput. Appl. Math.* **308**, 1–19 (2016)
21. Zhao, Q., Chen, Z.W.: An SQP-type method with superlinear convergence for nonlinear semidefinite programming. *Asia-Pac. J. Oper. Res.* **35**, 1850009 (2018)
22. Zhang, J.L., Zhang, X.S.: A robust SQP method for optimization with inequality constraints. *J. Comput. Appl. Math.* **21**(2), 247–256 (2003)
23. 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)
24. So, W.: Commutativity and spectra of Hermitian matrices. *Linear Algebra Appl.* **212–213**(15), 121–129 (1994)
25. Dunn, J.C.: On the convergence of projected gradient processes to singular critical points. *J. Optim. Theory Appl.* **55**(2), 203–216 (1987)
26. Burke, J.V., Moré, J.J.: On the identification of active constraints. *SIAM J. Numer. Anal.* **25**, 1197–1211 (1988)
27. Burke, J.V., Han, S.P.: A robust sequential quadratic programming method. *Math. Program.* **43**(1–3), 277–303 (1989)
28. Bonnans, J.F., Shapiro, A.: *Perturbation Analysis of Optimization Problems*, vol. 13, pp. 63–83. Springer, Berlin (2000)
29. Sun, D.F.: The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Math. Oper. Res.* **31**(4), 761–776 (2006)
30. Li, Z.Y.: *Differential-Algebraic Algorithm and Sequential Penalty Algorithm for Semidefinite Programming*. Fujian Normal University, Fujian (2006) (in Chinese)
31. Yamakawa, Y., Yamashita, N.: A differentiable merit function for the shifted perturbed Karush–Kuhn–Tucker conditions of the nonlinear semidefinite programming. *Pac. J. Optim.* **11**(3), 557–579 (2015)