Primal-dual path following method for nonlinear semi-infinite programs with semi-definite constraints

Takayuki Okuno\textsuperscript{1,2} · Masao Fukushima\textsuperscript{3}

Received: 1 October 2018 / Accepted: 30 April 2022 / Published online: 4 June 2022
© Springer-Verlag GmbH Germany, part of Springer Nature and Mathematical Optimization Society 2022

Abstract
In this paper, we propose a primal-dual path following method for nonlinear semi-infinite semi-definite programs with infinitely many convex inequality constraints, called SISDP for short. A straightforward approach to the SISDP is to use classical methods for semi-infinite programs such as discretization and exchange methods and solve a sequence of (nonlinear) semi-definite programs (SDPs). However, it is often too demanding to find exact solutions of SDPs. In contrast, our approach does not rely on solving SDPs accurately but approximately following a path leading to a solution, which is formed on the intersection of the semi-infinite feasible region and the interior of the semi-definite feasible region. Specifically, we first present a prototype path-following method and show its global weak* convergence to a Karush-Kuhn-Tucker point of the SISDP under some mild assumptions. Next, to achieve fast local convergence, we integrate a two-step sequential quadratic programming method equipped with the Monteiro-Zhang scaling technique into the prototype method. We prove two-step superlinear convergence of the resulting algorithm using Alizadeh-Hareberly-Overton-like, Nesterov-Todd, and Helmberg-Rendl-Vanderbei-Wolkowicz/Kojima-Shindoh-Hara/Monteiro scaling directions. Finally, we conduct some numerical experiments to demonstrate the efficiency of the proposed method through comparison with a discretization method that solves SDPs obtained by finite relaxation of the SISDP.

The work was supported by JSPS KAKENHI Grant Number [15K15943, 20K19748, 20H04145].

\begin{itemize}
\item Takayuki Okuno
takayuki-okuno@st.seikei.ac.jp
\item Faculty of Science and Technology Department of Science and Technology, Seikei University, 3-3-1, Kita, Kichijoji, Musashino, Tokyo 180-8633, Japan
\item Center for Advanced Intelligence Project, RIKEN, Nihonbashi 1-chome Mitsui Building, 15th floor, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan
\item The Kyoto College of Graduate Studies for Informatics, Kyoto 606-8225, Japan
\end{itemize}
Keywords  Semi-infinite program · Nonlinear semi-definite program · Path-following method · Superlinear convergence · Global convergence

Mathematics Subject Classification  90C22 · 90C26 · 90C34

1 Introduction

In this paper, we consider the following nonlinear semi-infinite semi-definite program with an infinite number of convex inequality constraints and one linear matrix inequality constraint, SISDP for short:

\[
\begin{align*}
\text{Minimize } f(x) \\
\text{subject to } g(x, \tau) &\leq 0 \quad (\tau \in T), \\
F(x) &\in S^m_+,
\end{align*}
\]

where \( f : \mathcal{R}^n \to \mathcal{R} \) is a continuously differentiable function and \( T \) is a compact metric space. In addition, \( g : \mathcal{R}^n \times T \to \mathcal{R} \) is a continuous function, and \( g(\cdot, \tau) : \mathcal{R}^n \to \mathcal{R} \) is convex and continuously differentiable for each \( \tau \in T \). Moreover, \( S^m \) and \( S^m_+(S^m_+) \) denote the sets of \( m \times m \) real symmetric matrices and real symmetric positive (semi-)definite matrices, respectively, and \( F(\cdot) : \mathcal{R}^n \to S^m \) is an affine function, i.e.,

\[ F(x) := F_0 + \sum_{i=1}^{n} x_i F_i \]

with \( F_i \in S^m \) for \( i = 0, 1, \ldots, n \) and \( x = (x_1, x_2, \ldots, x_n)^\top \). We assume that SISDP (1.1) has a nonempty solution set. We may allow SISDP (1.1) to include linear equality constraints. However, for simplicity of expression, we omit them, since the algorithms and analysis given in the subsequent sections can be extended straightforwardly.

When \( T \) comprises a finite number of elements, the SISDP reduces to a nonlinear semi-definite program (nonlinear SDP or NSDP). Particularly when all the functions are affine with respect to \( x \), it further reduces to a linear SDP (LSDP). LSDPs have been studied extensively in the aspects of theory, algorithms, and applications [35]. Meanwhile, studies on NSDPs are still scarce, although important applications are found in various fields in the real world [7, 13, 14]. Shapiro [27] expanded an elaborate theory on the first and second order optimality conditions of NSDPs. See [2] for a comprehensive description of the optimality conditions and duality theory of NSDPs. Yamashita et al. [43] proposed a primal-dual interior point-type method using the Monteiro-Zhang (MZ) family of directions [35, Chapter 10] and showed its global convergence property. They further made a local convergence analysis in [41] by exploiting some scaling directions. A sequential quadratic programming (SQP) method for nonlinear programs was also extended to NSDPs by Correa and Ramirez [4] and Freund et al. [7]. See the survey article [42] for more algorithms designed to solve NSDPs.
In the absence of the semi-definite constraint, SISDP (1.1) becomes a nonlinear semi-infinite program (SIP) with an infinite number of convex constraints. For solving nonlinear SIPs, many researchers proposed various kinds of algorithms, for example discretization based methods [25, 30], local reduction based methods [9, 22, 23, 31], Newton-type methods [15, 24], smoothing projection methods [40], convexification based methods [6, 28, 29, 34], and so on. In this paper, we make use of the local reduction method to design the proposed algorithm. The fundamental idea of the method is to represent the infinitely many inequality constraints as locally equivalent finitely many inequality constraints by means of implicit functions. Such a reduction approach is often costly and requires some strong theoretical assumptions at each point. But, it is a useful technique for designing an algorithm with fast local convergence to a solution. For an overview of SIPs, see [8, 11, 26] and the references therein.

Most closely related to SISDP (1.1) are SIPs involving (possibly infinitely many) conic constraints. Li et al. [18] considered a linear SIP with semi-definite constraints and proposed a discretization based method. Subsequently, Li et al. [17] tackled the same problem and developed a relaxed cutting plane method. Hayashi and Wu [10] focused on a linear SIP involving second-order cone (SOC) constraints and proposed an exchange-type method. It is worth mentioning that SISDP (1.1) can be viewed as a generalization of those problems in the sense that SOC constraints are representable as semi-definite constraints (e.g., see [1]) and the functions \( f(x, \tau) \) and \( g(x, \tau) \) are allowed to be nonlinear. More recently, Okuno et al. [21] considered an SIP with a convex objective function and an infinite number of linear conic constraints, and proposed an exchange-type method combined with Tikhonov’s regularization technique. Okuno and Fukushima [19] restricted themselves to a nonlinear SIP with infinitely many SOC constraints, and constructed a quadratically convergent SQP-type method based on the local reduction method. One of the common features of the algorithms mentioned in this paragraph is that they sequentially solve a sequence of certain conic constrained problems to generate iterates.

We find some important applications of SISDP (1.1). For example, semi-infinite eigenvalue optimization problems [18], finite impulse response filter design [38], and robust envelope-constrained filter design with orthonormal bases [16] can be formulated as linear SISDPs, namely, SISDP (1.1) whose functions are all affine with respect to \( x \). Secondly, we briefly mention two possible applications of SISDP (1.1) with a nonlinear objective function. Both applications are related to filter design problems. The first one arises in minimax design of infinite impulse response (IIR) filter design using the SDP relaxation technique [12]. Let \( D(\omega) \) be an ideal frequency response over \([0, \pi]\) and \( H(z) \) be a transfer function of the IIR digital filter with real coefficients. Then, the design problem is to minimize the maximum of a complex approximation error \( W(\omega)(D(\omega) - H(e^{\sqrt{-1}\omega})) \) over a frequency band of interest \( \Omega \subseteq [0, \pi] \), where \( e \) denotes Napier’s constant and \( W(\omega) : [0, \pi] \to \mathbb{R}^+ \) is a prefixed weighting function. In a manner similar to Jiang et al. [12], this problem can be reformulated as a linear SISDP with a semi-definite constraint, say \( X \in S^m_+ \), and the constraint rank \( X = 1 \) via a problem having nonconvex quadratic constraints. To mitigate the difficulty of this problem, Jiang et al. [12] considered a linear SISDP relaxation problem obtained by removing the rank 1 constraint, and furthermore, to gain a low-rank matrix solution.
by inducing as many zero eigenvalues as possible, they incorporated the $\ell_1$ norm of a vector of the eigenvalues $\lambda_1(X), \lambda_2(X), \ldots, \lambda_m(X)$ ($\sum_{i=1}^{m} |\lambda_i(X)| = \text{Tr}(X)$ for $X \in S_m^n$) into the objective function of the SISDP as a penalty, which results in a linear SISDP.\footnote{Actually, Jiang et al. \cite{12} solved linear SDPs with finitely many inequality constraints, which are obtained by discretizing the SISDP.} Note that such an $\ell_1$-type function can be interpreted as a so-called sparse regularizer in eigenvalues. Inspired by this observation, we may utilize other nonconvex sparse regularizers \cite{37} in eigenvalues, e.g., $\sum_{i=1}^{m} (\lambda_i(X)^2 + \epsilon)^{\frac{q}{2}}$ with $0 < q < 1$ and $\epsilon \geq 0$, although it is non-smooth when $\epsilon = 0$. Alternatively, letting $I$ be the $m \times m$ identity matrix, $\log \det(X + \epsilon I) = \sum_{i=1}^{m} \log (\lambda_i(X) + \epsilon)$ ($\epsilon > 0$) \cite{5} may also serve as a regularizer to obtain a rank 1 solution. In both cases, the resulting problem is nothing but SISDP (1.1) with a nonlinear objective function. The second application comes from spectral factorization associated with finite impulse response (FIR) filter design \cite{38, 39}. A detailed explanation of the problem as well as some numerical results of the algorithm proposed in this paper are given in Sect. 4.

In spite of the important applications, to the best of our knowledge, there is no existing work that deals with SISDP (1.1) itself. In this paper, we propose a path-following algorithm to solve the SISDP. To begin with, we present a prototype algorithm for generating a sequence approaching a Karush-Kuhn-Tucker (KKT) point of the SISDP. In this prototype algorithm, we approximately follow a central path formed by barrier KKT (BKKT) points of the SISDP. A BKKT point, which is a perturbed KKT point of the SISDP equipped with a so-called barrier parameter, can be computed efficiently using the interior-point SQP-type method proposed in the authors’ recent work \cite{20}. However, this algorithm actually generates a sequence of Lagrange multipliers in a measure space, and thus is conceptual in that sense. Subsequently, we present a more practical version of the prototype algorithm coupled with the local reduction method. Thanks to this local reduction technique, we are able to handle Lagrange multipliers in a finite dimensional space by expressing the infinitely many inequality constraints as finitely many ones locally. To attain fast local convergence, we further integrate a two-step SQP method into the prototype method. Specifically, using the Monteiro-Zhang (MZ) scaling technique and the local reduction method, we derive a certain scaled equation system of the SISDP, called a scaled BKKT system, whose solution set consists of BKKT points. We then apply a two-step SQP method to this system, while decreasing a barrier parameter to zero superlinearly. In each step of the two-step SQP method, we produce a search direction by solving a mixed linear complementarity system or a linear system approximating the scaled BKKT system. We then adjust a step-size along the obtained search direction so that the next iteration point remains in the interior of the semi-definite region.

To clarify the limiting behavior of the proposed algorithm, we first consider a local version of the algorithm that works only around a KKT point that satisfies a certain nondegeneracy condition. We will show that a step-size of unity is eventually adopted and, as a result, two-step superlinear convergence to the KKT point is achieved. Next, we give a global version of the algorithm. Global and local convergence properties are inherited from the prototype and local versions of the algorithm, respectively.
Although it is possible to design a convergent algorithm that solves NSDPs iteratively like the existing algorithms mentioned in the preceding paragraph, it is often too demanding to get an accurate solution of an NSDP at each iteration. In contrast, the proposed path-following algorithm only requires solving quadratic programs when combined with the interior-point SQP-type method [20].

The proposed algorithm is similar in spirit to the primal-dual interior-point method proposed by Yamashita et al. [41, 43] for NSDPs. Actually, they share the same idea of approaching a KKT point by tracking BKKT points. Nonetheless, our algorithm is distinguished from those of Yamashita et al. [41, 43] in the following points:

1. The algorithms proposed in [41, 43] are not ready to handle infinitely many inequality constraints, which we call semi-infinite constraints. On the other hand, to handle both semi-infinite constraints and semi-definite constraints adequately, our algorithm utilizes the local reduction method and the MZ scaling technique jointly, which come from the fields of SIP and SDP, respectively.
2. Due to the semi-infinite constraints, we have to examine convergence properties of a sequence of dual variables in an infinite-dimensional space.
3. Our algorithm solves a scaled BKKT system that contains complementarity conditions associated with the semi-infinite constraints. To solve this system for each barrier parameter value, we propose to solve a certain quadratic program repeatedly like SQP methods in nonlinear optimization, while Yamashita et al. [41, 43] solve linear Newton equations. Accordingly, the resulting analysis of our algorithms is quite different from that of Yamashita et al. [41, 43].

The paper is organized as follows: In Sect. 2, we propose the prototype primal-dual path-following method for the SISDP. We prove that any weak*-accumulation point of the generated sequence is a KKT point of the SISDP under some mild assumptions. In Sect. 3, we further combine the local-reduction based SQP method with the prototype method. We first give the local version of the algorithm and analyze its local convergence speed towards a KKT point. We next provide the global version of the algorithm, which is the final form of the proposed method. In Sect. 4, we conduct some numerical experiments to exhibit the efficiency of the proposed method. Finally, we conclude this paper with some remarks.

Notations

Throughout this paper, we use the following notations: Let $\mathcal{R}^n_+$ and $\mathcal{R}_n^{++}$ be the $n$-dimensional nonnegative and positive orthants, respectively. The $m \times m$ identity matrix is denoted by $I$. For any $P \in \mathcal{R}^{m \times m}$, $\text{Tr}(P)$ denotes the trace of $P$. For any symmetric matrices $X, Y \in \mathcal{S}^m$, we denote the Jordan product of $X$ and $Y$ by $X \circ Y := (XY + YX)/2$ and the inner product of $X$ and $Y$ by $X \bullet Y = \text{Tr}(XY)$. Also, we denote the Frobenius norm $\|X\|_F := \sqrt{X \bullet X}$ and

$$svec(X) := (X_{11}, \sqrt{2}X_{21}, \ldots, \sqrt{2}X_{m1}, X_{22}, \sqrt{2}X_{32}, \ldots, \sqrt{2}X_{m2}, X_{33}, \ldots, X_{mm})^{\top} \in \mathcal{R}_{\frac{m(m+1)}{2}}$$
for $X \in S^m$. We write $(F_i \bullet V)_{i=1}^n := (F_1 \bullet V, F_2 \bullet V, \ldots, F_n \bullet V)^\top \in \mathcal{R}^n$ for $V, F_1, F_2, \ldots, F_n \in S^m$. We also denote $(a)_+ := \max(a, 0)$ for any $a \in \mathcal{R}$. For sequences $\{y_k\}$ and $\{z_k\}$, if $\|y_k\| \leq L\|z_k\|$ for any $k$ with some $L > 0$, we write $\|y_k\| = O(\|z_k\|)$. If $L_1\|z_k\| \leq \|y_k\| \leq L_2\|z_k\|$ for any $k$ with some $L_1, L_2 > 0$, we represent $\|y_k\| = \Theta(\|z_k\|)$. Moreover, if there exists a sequence $\{\alpha_k\}$ with $\lim_{k \to \infty} \alpha_k = 0$ and $\|y_k\| \leq \alpha_k\|z_k\|$ for any $k$, we write $\|y_k\| = \alpha(\|z_k\|)$. Finally, we denote by $\perp$ the perpendicularity with respect to the canonical inner product in the Euclidean space.

**Terminologies from functional analysis**

Let us review some terminologies from functional analysis briefly. For more details, refer to the basic material [2, Section 2] or suitable textbooks of functional analysis.

Let $\mathcal{C}(T)$ be the set of real-valued continuous functions defined on $T$ endowed with the supremum norm $\|h\| := \max_{\tau \in T} |h(\tau)|$. Let $\mathcal{M}(T)$ be the dual space of $\mathcal{C}(T)$, which can be identified with the space of (finite signed) regular Borel measures with the Borel sigma algebra $\mathcal{B}$ on $T$ equipped with the total variation norm, i.e., $\|\mu\| := \sup_{A \in \mathcal{B}} \int_A f(y) \, d\mu(y)$ for $\mu \in \mathcal{M}(T)$. Denote by $\mathcal{M}_+(T)$ the set of all the nonnegative Borel measures of $\mathcal{M}(T)$. Especially if $\mu \in \mathcal{M}_+(T)$, $\|\mu\| = \mu(T)$ since $\inf_{A \in \mathcal{B}} \int_A f(y) \, d\mu(y) = 0$ and $\sup_{A \in \mathcal{B}} \int_A f(y) \, d\mu(y) = \mu(T)$. For $\mu \in \mathcal{M}(T)$, we define the support of $\mu$ by $\text{supp}(\mu) := \{ \tau \in T \mid \mu(N_\tau) > 0 \}$, where $N_\tau$ is an open neighborhood of $\tau$ in $T$, then $\mu(N_\tau) > 0$. We say that $\mu \in \mathcal{M}(T)$ is a finite discrete measure if there exist a finite number of indices $\tau_1, \tau_2, \ldots, \tau_q \in T$ and scalars $\alpha_1, \alpha_2, \ldots, \alpha_q \in \mathcal{R}$ such that $\mu(A) = \sum_{i=1}^q \alpha_i I_A(\tau_i)$ for any Borel set $A \in \mathcal{B}$, where $I_A : T \to \mathcal{R}$ is the indicator function satisfying $I_A(\tau) = 1$ if $\tau \in A$ and $I_A(\tau) = 0$ otherwise.

Let $\langle \cdot, \cdot \rangle : \mathcal{M}(T) \times \mathcal{C}(T) \to \mathcal{R}$ be the bilinear form defined by $\langle \cdot, h \rangle := \int_T \int_T f(\tau) \, d\mu(\tau)$ for $\mu \in \mathcal{M}(T)$ and $h \in \mathcal{C}(T)$. We then endow $\mathcal{M}(T)$ with the weak-*-topology, which is the minimum topology such that any semi-norm $p_A$ on $\mathcal{M}(T)$ is continuous for any finite subset $A \subseteq \mathcal{C}(T)$, where $p_A : \mathcal{M}(T) \to \mathcal{R}$ is defined by $p_A(\mu) := \max_{h \in A} |\langle \mu, h \rangle|$. Let us here specify the concept of accumulation points and limit points in the sense of the weak-*-topology. Let $\{y_k\}$ be a sequence in $\mathcal{M}(T)$ and $y^* \in \mathcal{M}(T)$.

1. We call $y^*$ the weak-* limit point of $\{y_k\}$ if for any neighborhood $\mathcal{N}(y^*)$ of $y^*$ with respect to the weak-*-topology there exists an integer $\overline{K} \geq 0$ such that $y_k^* \in \mathcal{N}(y^*)$ for any $k \geq \overline{K}$. We then say $\{y_k\}$ weakly converges to $y^*$ and often write it as $\text{w}^*\lim_{k \to \infty} y_k = y^*$.

2. We call $y^*$ a weak-* accumulation point of $\{y_k\}$ if for any integer $\overline{K} \geq 0$ and neighborhood $\mathcal{N}(y^*)$ of $y^*$ with respect to the weak-*-topology there exists an integer $k \geq \overline{K}$ such that $y_k^* \in \mathcal{N}(y^*)$. 

\(\square\) Springer
2 Primal-dual path following method (Prototype)

In this section, we give a prototype algorithm for solving SISDP (1.1). A more specific version of the algorithm equipped with superlinear convergence will be presented in the subsequent section.

2.1 KKT conditions for the SISDP

First, we present the Karush-Kuhn-Tucker (KKT) conditions for the SISDP together with Slater’s constraint qualification, abbreviated as SCQ. Here, SCQ for the SISDP is defined precisely as below:

**Definition 2.1** We say that the Slater constraint qualification (SCQ) holds for SISDP (1.1) if there exists some \( \bar{x} \in \mathbb{R}^n \) such that \( F(\bar{x}) \in S^+_m \) and \( g(\bar{x}, \tau) < 0 \) (\( \tau \in T \)). Such a point \( \bar{x} \) is called a Slater point.

**Theorem 2.1** Let \( x^* \in \mathbb{R}^n \) be a local optimal solution of SISDP (1.1). Then, under the SCQ, there exists some finite Borel-measure \( y \in \mathcal{M}(T) \) such that

\[
\nabla f(x^*) + \int_T \nabla x g(x^*, \tau)dy(\tau) - (F_i \cdot V)_{i=1}^n = 0, \tag{2.1}
\]

\[
F(x^*) \circ V = O, \ F(x^*) \in S^+_m, \ V \in S^+_m, \tag{2.2}
\]

\[
\int_T g(x^*, \tau)dy(\tau) = 0, \ g(x^*, \tau) \leq 0 (\tau \in T), \ y \in \mathcal{M}_+(T), \tag{2.3}
\]

where \( V \in S^+_m \) is a Lagrange multiplier matrix associated with the constraint \( F(x) \in S^+_m \). In particular, there exists some discrete measure \( y \in \mathcal{M}_+(T) \) satisfying the above conditions and \( |\text{supp}(y)| \leq n \), where \( \text{supp}(y) = \{ \tau \in T \mid y(\{\tau\}) > 0 \} \). Conversely, when \( f \) is convex, if the above conditions (2.1)–(2.3) hold, then \( x^* \) is an optimum of SISDP (1.1).

**Proof** This can be proved by applying [21, Theorem 2.4]. We leave the detailed proof to Appendix. \( \square \)

**Definition 2.2** We call conditions (2.1)–(2.3) the Karush-Kuhn-Tucker (KKT) conditions for SISDP (1.1). We also call \( (x, y, V) \) satisfying (2.1)–(2.3) a KKT point of SISDP (1.1).

2.2 Description of the prototype algorithm

Let us define the function \( R_\mu : \mathbb{R}^n \times \mathcal{M}(T) \times S^+_m \to \mathbb{R} \) with a parameter \( \mu \geq 0 \) by

\[
R_\mu(x, y, V) := \sqrt{\psi(x)^2 + \|\varphi_1(x, y, V)\|^2 + \varphi_2(x, y)^2 + \|\varphi_3(x, V, \mu)\|^2_F},
\]

\( \odot \) Springer
where

\[ \psi(x) := \max_{\tau \in T} (g(x, \tau) + \varphi_1(x, y, V) := \nabla f(x) + \int_T \nabla_x g(x, \tau)dy(\tau) - (F_i \cdot V)^n_{i=1}, \]

\[ \varphi_2(x, y) := \int_T g(x, \tau)dy(\tau), \]

\[ \varphi_3(x, V, \mu) := F(x) \circ V - \mu I. \]

Notice that a point satisfying \( R_0(x, y, V) = 0 \) with \( F(x) \in S^m_+ \) and \( V \in S^m_+ \) is nothing but a KKT point of SISDP (1.1). In terms of the function \( R_\mu \), we define a barrier KKT (BKKT) point by perturbing the semi-definite complementarity condition in the KKT conditions (2.1)–(2.3).

**Definition 2.3** Let \( \mu > 0 \). We call \((x, y, V) \in \mathcal{R}^n \times \mathcal{M}(T) \times S^m \) a barrier Karush-Kuhn-Tucker (BKKT) point of SISDP (1.1) if \( R_\mu(x, y, V) = 0, y \in \mathcal{M}(T), F(x) \in S^m_+, \) and \( V \in S^m_+ \). In particular, we refer to \( \mu \) as a barrier parameter.

Note that a BKKT point actually satisfies \( F(x) \in S^m_+ \) and \( V \in S^m_+ \) since \( R_\mu(x, y, V) = 0 \) implies \( F(x) \circ V = \mu I \). Given a parameter \( \varepsilon \geq 0 \) together with a barrier parameter \( \mu > 0 \), we define the following set that contains BKKT points with barrier parameter \( \mu \):

\[ \mathcal{N}_\varepsilon^\mu := \{ w := (x, y, V) \in \mathcal{R}^n \times \mathcal{M}(T) \times S^m_+ | R_\mu(w) \leq \varepsilon, F(x) \in S^m_+ \}. \]

Notice that \( \mathcal{N}_\varepsilon^\mu \) with \( \varepsilon = 0 \) coincides with the set of BKKT points with \( \mu \). Hence, the set \( \mathcal{N}_\varepsilon^\mu \) can be regarded as a set of approximations of BKKT points, whose approximation degree is controlled by the parameter \( \varepsilon \). Hereafter, we often call a point in \( \mathcal{N}_\varepsilon^\mu \) an approximate BKKT point. Moreover, by definition, we expect that a point in \( \mathcal{N}_\varepsilon^\mu \) approaches the set of KKT points as \( \varepsilon \) and \( \mu \) get closer to zero. Inspired by this observation, we propose the following prototype algorithm for computing a KKT point of the SISDP, in which we generate a sequence of approximate BKKT points \( \{w^k\} \) for SISDP (1.1) such that \( w^k \in \mathcal{N}_{\varepsilon \mu_k}^\mu \) for each \( k \) with a positive parameter sequence \( \{(\mu_k, \varepsilon_k)\} \) converging to \((0, 0)\).

**Algorithm 1** (Primal-dual path following method (Prototype))

**Step 0** (Initial setting): Choose an initial iteration point \( w^0 := (x^0, y^0, V_0) \in \mathcal{R}^n \times \mathcal{M}(T) \times S^m \) such that \( F(x^0) \in S^m_+ \) and \( V_0 \in S^m_+ \). Choose the initial parameters \( \mu_0 > 0, \varepsilon_0 > 0 \) and \( \beta \in (0, 1) \). Let \( k := 0 \).

**Step 1** (Initial setting): (Stopping rule): Stop if

\[ R_0(w^k) = 0, F(x^k) \in S^m_+, V_k \in S^m_+, y^k \in \mathcal{M}(T). \]

Otherwise, go to Step 2.
Step 2 (Computing an approximate BKKT point): Find an approximate BKKT point \( w_{k+1} = (x_{k+1}, y_{k+1}, V_{k+1}) \) such that
\[
w_{k+1} \in N \mu_k. \tag{2.4}
\]

Step 3 (Update): Set \( \mu_{k+1} := \beta \mu_k \) and \( \epsilon_{k+1} := \beta \epsilon_k \). Let \( k := k + 1 \). Return to Step 1.

In the recent work [20], the authors proposed an interior-point SQP method for computing a BKKT point and showed its global convergence property. This method can be used in Step 2.

2.3 Convergence analysis

In this section, we establish weak\(^*\) convergence of Algorithm 1 to KKT points of SISDP (1.1). Furthermore, we will characterize weak\(^*\) accumulation points of the generated sequence more precisely for some special cases. For the sake of analysis, we assume that Algorithm 1 produces an infinite sequence and further make the following assumptions:

Assumption A

1. The feasible set of SISDP (1.1) is nonempty and compact.
2. Later’s constraint qualification holds for SISDP (1.1).

Let \( S^* \subseteq \mathbb{R}^n \) be the optimal solution set of SISDP (1.1) and \( \bar{v} \in \mathcal{R} \) be a constant larger than the optimal value of the SISDP. If \( f \) is convex, Assumption A-1 can be replaced with the milder assumption that \( S^* \) is compact by adding a convex constraint \( f(x) \leq \bar{v} \) to the SISDP without changing the shape of \( S^* \). Under the above assumptions, we first show that the generated sequences \( \{x_k\} \) and \( \{(y_k, V_k)\} \) are bounded.

Proposition 2.1 Suppose that Assumption A-1 holds. Then, any sequence \( \{x^k\} \) produced by Algorithm 1 is bounded.

Proof Denote the feasible set of SISDP (1.1) by \( \mathcal{F} \) and define a proper closed convex function \( \varphi : \mathcal{R}^n \rightarrow \mathcal{R} \) by
\[
\varphi(x) := \max \left( -\lambda_{\min}(F(x)), \max_{\tau \in T} g(x, \tau) \right),
\]
where \( \lambda_{\min}(F(x)) \) denotes the least eigenvalue of \( F(x) \). Note that, from (2.4) and \( \epsilon_k \leq \epsilon_0 \) for all \( k \) by the update rule of \( \epsilon_k \) in Step 3, it holds that
\[
\{x^k\} \subseteq \{x \in \mathcal{R}^n \mid \varphi(x) \leq \epsilon_0\}. \tag{2.5}
\]
where \( \epsilon_0 \) is an algorithmic parameter given in Step 0. This is because, by Step 2 of the algorithm, we have \( F(x^k) \in S^*_{\mu_k} \) for all \( k \geq 0 \), which implies \( -\lambda_{\min}(F(x^k)) < 0 < \epsilon_0 \) for each \( k \). In addition, by (2.4), we have \( \max_{\tau \in T} g(x^k, \tau) \leq \psi(x^k) \leq R_{\mu_k}(x^k, y^k, V_k) \leq \epsilon_{k-1} \leq \epsilon_0 \) for any \( k \geq 1 \). Hence, we obtain (2.5).

Recall that if a given proper closed convex function has a nonempty and compact level set, then any
level set is compact \[3, \text{Proposition 2.3.1}\]. Hence, any level set \( \{ x \in \mathbb{R}^n \mid \varphi(x) \leq \eta \} \) with \( \eta > 0 \) is compact because the level set \( \{ x \in \mathbb{R}^n \mid \varphi(x) \leq 0 \} (= \mathcal{F}) \) is nonempty and compact. Combining this fact with (2.5) implies that \( \{ x^k \} \) is bounded. The proof is complete. \( \square \)

**Proposition 2.2.** Suppose that Assumption A holds. Then, the generated Lagrange multiplier sequences \( \{ V_k \} \subseteq S^m_{++} \) and \( \{ y^k \} \subseteq \mathcal{M}_+(T) \) are bounded.

**Proof.** For simplicity of expression, denote \( \tilde{w}^k := (V_k, y^k) \in S^m_{++} \times \mathcal{M}_+(T) \) and

\[
W_k := \frac{V_k}{\| \tilde{w}^k \|}, \quad q^k := \frac{y^k}{\| \tilde{w}^k \|}
\]

where \( \| \cdot \| \) is a suitable norm such that \( \| \tilde{w}^k \|^2 = \| V_k \|^2 + \| y_k \|^2 \) on \( S^m \times \mathcal{M}(T) \). For contradiction, suppose that there exists a subsequence \( \{ \tilde{w}^k \}_{k \in \mathcal{K}} \subseteq \{ \tilde{w}^k \} \) such that \( \tilde{w}^k \rightarrow \infty (k \in \mathcal{K} \rightarrow \infty) \), where \( \mathcal{K} \subseteq \{ 1, 2, \ldots, \} \) indicates a subset of indices. Note that \( \{ (W_k, q_k) \} \) is bounded. Notice also that the corresponding sequence \( \{ x^k \}_{k \in \mathcal{K}} \) is bounded from Proposition 2.1. Recall that any bounded sequence in \( \mathcal{M}(T) \) has at least one weak* accumulation point and one can extract a subsequence weakly* converging to that point. Thanks to this property, without loss of generality we can assume that there exists a point \( (x^*, W_*, q^*) \in \mathbb{R}^n \times S^m_{++} \times \mathcal{M}_+(T) \) such that

\[
\lim_{k \in \mathcal{K} \rightarrow \infty} (x^k, W_k) = (x^*, W_*), \quad \lim_{k \in \mathcal{K} \rightarrow \infty} q^k = q^*.
\]

Note, in particular, that \( \|(W_*, q^*)\| = 1 \), since \( w^* - \lim_{k \in \mathcal{K} \rightarrow \infty} q^k = q^* \) entails the relation that

\[
\lim_{k \in \mathcal{K} \rightarrow \infty} \| q^k \| = \lim_{k \in \mathcal{K} \rightarrow \infty} \int_T dq^k(\tau) = \int_T dq^*(\tau) = \| q^* \|
\]

and therefore

\[
\| (W_*, q^*) \|^2 = \| W_* \|^2 + \| q^* \|^2 = \lim_{k \in \mathcal{K} \rightarrow \infty} \left( \| W_k \|^2 + \| q^k \|^2 \right) = 1.
\]

From (2.4), \( \max (\psi(x^k), \| \varphi_1(x^k, y^k, V_k) \|, \| \varphi_2(x^k, y^k) \|, \| \varphi_3(x^k, V_k, \mu_k) \| \) \) \( \leq \varepsilon_{k-1} \) follows for each \( k \geq 1 \). Dividing both sides of this inequality by \( \| \tilde{w}^k \| \) yields

\[
\left\| \frac{\nabla f(x^k)}{\| \tilde{w}^k \|} - (F_i \cdot W_k)_{i=1}^n + \int_T \nabla x \mathcal{G}(x^k, \tau) dq^k(\tau) \right\| \leq \frac{\varepsilon_{k-1}}{\| \tilde{w}^k \|},
\]

\[
\left| \int_T g(x^k, \tau) dq^k(\tau) \right| \leq \frac{\varepsilon_{k-1}}{\| \tilde{w}^k \|}, \quad q^k \in \mathcal{M}_+(T),
\]

\[
\left\| F(x^k) \circ W_k - \frac{\mu_{k-1}}{\| \tilde{w}^k \|} I \right\| \leq \frac{\varepsilon_{k-1}}{\| \tilde{w}^k \|}, \quad F(x^k) \in S^m_{++}, \quad W_k \in S^m_{++}.
\]
By letting $k(\in K) \to \infty$, we obtain
\[
(F_i \bullet W_*)_{i=1}^n - \int_T \nabla_x g(x^*, \tau) dq^*(\tau) = 0, \tag{2.6}
\]
\[
\int_T g(x^*, \tau) dq^*(\tau) = 0, \quad q^* \in \mathcal{M}_+(T), \tag{2.7}
\]
\[
F(x^*) \circ W_* = O, \quad F(x^*) \in S^m_+, \quad W_* \in S^m_+. \tag{2.8}
\]

Now, choose a Slater point $\tilde{x} \in \mathcal{R}^n$ arbitrarily and let $\tilde{d} := \tilde{x} - x^*$. Notice here that
\[
F(\tilde{x}) \bullet W_* \geq 0, \quad \int_T g(\tilde{x}, \tau) dq^*(\tau) \leq 0, \tag{2.9}
\]
since
\[
F(\tilde{x}) \in S^m_+ \quad \text{and} \quad W_* \in S^m_+. \tag{2.10}
\]
On the other hand, it holds that
\[
F(\tilde{x}) \bullet W_* - \int_T g(\tilde{x}, \tau) dq^*(\tau)
= F(x^* + \tilde{d}) \bullet W_* - \int_T g(x^* + \tilde{d}, \tau) dq^*(\tau)
\leq F(x^* + \tilde{d}) \bullet W_* - \int_T (g(x^*, \tau) + \nabla_x g(x^*, \tau)^\top \tilde{d}) dq^*(\tau)
= \tilde{d}^\top (F_i \bullet W^*)_i = -\int_T \nabla_x g(x^*, \tau) dq^*(\tau)
= 0, \tag{2.11}
\]
where the first inequality holds because $g(x^*, \tau) + \nabla_x g(x^*, \tau)^\top \tilde{d} \leq g(x^* + \tilde{d}, \tau)$ ($\tau \in T$) by the convexity of $g(\cdot, \tau)$. Moreover, the third equality is obtained from (2.7) and the fact that $F(x^*) \bullet W_* = 0$ by (2.8). The last equality is due to (2.6). Combining (2.9) and (2.11) implies that $F(\tilde{x}) \bullet W_* = 0$ and $\int_T g(\tilde{x}) dq^*(\tau) = 0$, from which we can conclude $W_* = O$ and $q^* = 0$ by using (2.10) again. This contradicts $\|W_*, q^*\| = 1$. The proof is complete. \hfill \Box

Now, we are ready to establish the global convergence property of Algorithm 1. Recall that the definition of KKT points is given in Definition 2.2.

**Theorem 2.2** Suppose that Assumption A holds. Then, the sequence $\{(x^k, y^k, V_k)\}$ produced by Algorithm 1 is bounded. Let $(x^*, y^*, V_*) \in \mathcal{R}^n \times \mathcal{M}_+(T) \times S^m$ be a weak* accumulation point of $\{(x^k, y^k, V_k)\}$. Then, $(x^*, y^*, V_*)$ is a KKT point of SISDP (1.1). In particular, if $f$ is convex, $x^*$ is an optimum.
Proof The boundedness of \((x^k, y^k, V_k)\) follows from Propositions 2.1 and 2.2. It remains to show the second half of the theorem. We can assume \(\lim_{k \to \infty} (x^k, V_k) = (x^*, V_*)\) and \(w^* = \lim_{k \to \infty} y^k = y^*\) without loss of generality. Then, by letting \(k \to \infty\) in (2.4), we see that the KKT conditions (2.1)–(2.3) hold with some \(M\). For each \(2\) Here, the metric topology in \(R^n \times S^m \times R^M \times T^M\) is the one which is naturally induced from the norm topology in \(R^n \times S^m \times R^m\) and the metric topology in \(T^M\).

Up to now, we have imposed no particular conditions on \(y\) in the KKT conditions and \(y^k\) that is computed in Step 2. In the proposed algorithm presented in the next section, in order to compute \(y^k\), we will make use of the interior-point SQP-type method [20].

In the following theorem, assuming that \(y^k\) is discrete, thus \(\text{supp}(y) = \{\tau \in T | y(\tau) > 0\}\), and \(|\text{supp}(y^k)| \leq M\) with some \(M > 0\) for all \(k\), we describe a more precise form of a weak* accumulation point of \(\{y^k\}\). The boundedness assumption on \(|\text{supp}(y^k)|\) is not restrictive in practice. In fact, for the numerical examples shown in Sect. 4, we observe that \(|\text{supp}(y^k)|\) is bounded by the dimension of \(x\), at least for all test examples solved in the numerical experiments in Sect. 4.

**Theorem 2.3** Assume that \(y^k\) is discrete and \(|\text{supp}(y^k)| \leq M\) for any \(k \geq 0\) with \(M\) times some \(M > 0\), and consider a sequence \(\{t^k\} \subset T^M := T \times \cdots \times T\) with \(t^k := (\tau_1^k, \tau_2^k, \ldots, \tau_M^k)\) such that \(\text{supp}(y^k) \subset \{\tau_1^k, \tau_2^k, \ldots, \tau_M^k\}\) and \(y^k(\{\tau_i^k\}) = 0\) if \(\tau_i^k \notin \text{supp}(y^k)\) for \(i = 1, 2, \ldots, M\). Furthermore, denote

\[
\xi^k := (y^k(\{\tau_1^k\}), y^k(\{\tau_2^k\}), \ldots, y^k(\{\tau_M^k\}))^T \in R^M_+
\]

for \(k = 1, 2, \ldots\). Then,

(i) with regard to the metric topology in \(R^n \times S^m \times R^M \times T^M\), the sequence \(\{(x^k, V_k, \xi^k, t^k)\}\) is bounded and has accumulation points.\(^2\)

(ii) Let \((x^*, V_*, \xi^*, t^*)\) be an arbitrary accumulation point of \(\{(x^k, V_k, \xi^k, t^k)\}\) in the sense of (i), where \((x^*, V_*) \subset R^n \times S^m_+\), \(t^* := (\tau_1^*, \tau_2^*, \ldots, \tau_M^*) \in T^M\), and \(\xi^* = (\xi_1^*, \xi_2^*, \ldots, \xi_M^*) \in R^M_+\), denote the distinct elements of \(\{\tau_1^*, \tau_2^*, \ldots, \tau_M^*\}\) by \(s_1, s_2, \ldots, s_p \in T\) with \(p \leq M\).\(^3\) Moreover, define the finite discrete measure \(y^* : B \to R_+\) by

\[
y^*(A) := \sum_{j=1}^p \xi_j^* I_A(s_j) \quad (A \in B),\tag{2.12}
\]

where \(\xi_j^* := \sum_{i: \tau_i^* = s_j} \xi_i^*\) for \(j = 1, 2, \ldots, p\). Then, \(y^*\) is a weak* accumulation point of \(\{y^k\}\) and \((x^*, y^*, V_*)\) is a KKT point of SISDP (1.1).

\(^2\) Here, the metric topology in \(R^n \times S^m \times R^M \times T^M\) is the one which is naturally induced from the norm topology in \(R^n \times S^m \times R^m\) and the metric topology in \(T^M\).

\(^3\) For each \(k\), by definition, \(\tau_i^k \neq \tau_j^k\) if \(i \neq j\). But, at the limit, \(\tau_i^* = \tau_j^*\) possibly occurs, since two distinct sequences may converge to the identical point.
(iii) Especially, if $|\text{supp}(y^*)| = M$, we have
\[ y^*(A) = \sum_{i=1}^{M} \xi_i^* I_A(\tau_i^*) \quad (A \in \mathcal{B}). \]

**Proof** The first statement (i) is proved in a manner similar to Proposition 2.2. Let us show the second statement (ii). Without loss of generality, by taking a subsequence if necessary, we may assume that $\{ (x^k, V_k, \xi^k, t^k) \}$ converges to $(x^*, V_*, \xi^*, t^*)$. Then, for an arbitrary $h \in \mathcal{C}(T)$, it holds that
\[ \lim_{k \to \infty} \langle y^k, h \rangle = \langle y^*, h \rangle, \]
since, by the definitions of $t^*$, $\xi^*$, and $\zeta^*$ together with the continuity of $h$ over $T$,
\[ \langle y^k, h \rangle = \int_T h(\tau) dy^k(\tau) = \sum_{i=1}^{M} h(\tau_i^k) y^k(\{\tau_i^k\}) = \sum_{i=1}^{M} h(\tau_i^k) \xi_i^k \]
\[ \overset{k \to \infty}{\longrightarrow} \sum_{i=1}^{M} h(\tau_i^*) \xi_i^* = \sum_{j=1}^{p} h(s_j) \zeta_j^* = \int_T h(\tau) dy^*(\tau) = \langle y^*, h \rangle. \]

Finally, by Theorem 2.2, $(x^*, y^*, V_*)$ is a KKT point of the SISDP. The statement (iii) follows from (2.12) together with $\text{supp}(y^*) = \{ \tau_1^*, \tau_2^*, \ldots, \tau_M^* \}$ and $\tau_i^* \neq \tau_j^* (i \neq j)$.

---

3 Proposed algorithm

Algorithm 1 generates a sequence which is merely weak* convergent. In this section, we further integrate an SQP method coupled with the local reduction method [11, 19, 23, 31] into Algorithm 1 so as to achieve fast convergence. Then, the semi-infinite constraints are treated locally as a finite number of inequality constraints, and as a result, the Lagrange multiplier $y$ in the measure space $\mathcal{M}_+(T)$ reduces to a finite dimensional vector of Lagrange multipliers. This manipulation will enable us to argue the convergence rate of a sequence generated by the algorithm. For simplicity of explanation, we first describe a local version of the proposed algorithm as Algorithm 2 that works locally around a KKT point of SISDP (1.1) and study its limiting behavior. A global version will be presented as Algorithm 3 at the end of this section.

Throughout this section, we assume that the compact metric space $T$ is a bounded closed set in $\mathbb{R}^q$ formed by finitely many sufficiently smooth inequality constraints and $g(x, \cdot)$ is twice continuously differentiable for any $x \in \mathbb{R}^n$. Moreover, we often identify $X \in S^m$ with $\text{svec}(X) \in \mathbb{R}^{m(m+1)/2}$.
3.1 Local reduction technique

We explain the local reduction method to SISDP (1.1) briefly. Suppose that we are standing at a point \( \bar{x} \in \mathcal{R}^n \). The local reduction method represents the semi-infinite region \( D := \{ x \in \mathcal{R}^n \mid g(x, \tau) \leq 0 (\tau \in T) \} \) with finitely many inequality constraints locally around \( \bar{x} \). Specifically, in some open neighborhood of \( \bar{x} \), say \( U(\bar{x}) \), it expresses the region \( D \cap U(\bar{x}) \) as

\[
D \cap U(\bar{x}) = \{ x \in U(\bar{x}) \mid g(x, \tau^i(\bar{x})) \leq 0 (i = 1, 2, \ldots, p(\bar{x})) \}
\]

using distinct smooth implicit functions \( \tau^i(\bar{x}) : U(\bar{x}) \rightarrow T (i = 1, 2, \ldots, p(\bar{x})) \) with some positive integer \( p(\bar{x}) \). Then, SISDP (1.1) is locally equivalent to the problem with finitely many inequality constraints in \( U(\bar{x}) \), namely,

\[
\begin{align*}
\text{Minimize} & \quad f(x) \\
\text{subject to} & \quad \hat{g}_i(x) := g(x, \tau^i(\bar{x})) \leq 0 (i = 1, 2, \ldots, p(\bar{x})), \\
& \quad F(x) \in S^m_+,
\end{align*}
\]

(3.1)

to which standard nonlinear optimization algorithms such as the SQP method are conceptually applicable. In what follows, we clarify the condition under which the functions \( \tau^i(\cdot) (i = 1, 2, \ldots, p(\bar{x})) \) and the open neighborhood \( U(\bar{x}) \) exist. Let \( S(x) \) denote the set of all local maximizers of \( \max_{\tau \in T} g(x, \tau) \) and let

\[
S_\delta(x) := \{ \tau \in S(x) \mid g(x, \tau) > \max_{\tau \in T} g(x, \tau) - \delta \}
\]

(3.2)

for a given constant \( \delta > 0 \). We introduce the concept of \( \delta \)-nondegeneracy, which will play an important role in the subsequent analysis.

**Definition 3.1** Given \( \delta > 0 \), we say that \( \bar{x} \) is \( \delta \)-nondegenerate for \( \max_{\tau \in T} g(\bar{x}, \tau) \) if \( |S_\delta(\bar{x})| < \infty \) and the linear independence constraint qualification, the second-order sufficient conditions, and the strict complementarity condition regarding \( \max_{\tau \in T} g(\bar{x}, \tau) \) hold at any \( \tau \in S_\delta(\bar{x}) \).

As \( \delta \) gets larger, the \( \delta \)-nondegeneracy becomes restrictive. Indeed, if \( \delta \) is sufficiently large, this concept implies that all local maxima of \( g(\bar{x}, \tau) \) in \( \tau \in T \) are strict local maximizers. If \( \bar{x} \) is \( \delta \)-nondegenerate, there exist an open neighborhood \( U(\bar{x}) \subseteq \mathcal{R}^n \), a nonnegative integer \( p(\bar{x}) := |S_\delta(\bar{x})| \), and twice continuously differentiable implicit functions \( \tau^i(\cdot) : U(\bar{x}) \rightarrow T \) such that \( S_\delta(x) = \{ \tau^i(x) \}_{i=1}^{p(\bar{x})} \) with \( \tau^i(x) (i = 1, \ldots, p(\bar{x})) \) being strict local maximizers of \( \max_{\tau \in T} g(x, \tau) \) for any \( x \in U(\bar{x}) \).

With those implicit functions, it holds that \( \max_{\tau \in T} g(x, \tau) = \max_{1 \leq i \leq p(\bar{x})} \hat{g}_i(x) \) in \( U(\bar{x}) \) and thus SISDP (1.1) and nonlinear SDP (3.1) are equivalent locally.

**Remark 3.1** If \( g(x, \cdot) \) is strongly concave for any \( x \in \mathcal{R}^n \) and \( T \) is a box set, then \( |S_\delta(x)| = 1 \) for any \( \delta \geq 0 \) and \( x \in \mathcal{R}^n \), and the linear independence constraint qualification and the second-order sufficient conditions hold at the unique maximum in \( \max_{\tau \in T} g(x, \cdot) \). If, in addition, the strict complementarity holds, \( \delta \)-nondegeneracy holds at each \( x \).
The functions $\hat{g}_i(\cdot)$ ($i = 1, 2, \ldots, p(\bar{x}))$ are convex in $U(\bar{x})$ when the functions $g(\cdot, \tau)$ ($\tau \in T$) are convex as supposed in SISDP (1.1). Indeed, for each $i = 1, 2, \ldots, p(\bar{x})$, there exists some neighborhood $T_i \subseteq T$ of $\tau_i(\bar{x})$ such that $\max_{\tau \in T_i} g(x, \tau) = \hat{g}_i(x)$ holds for any $x \in U(\bar{x})$. By noting that $\max_{\tau \in T_i} g(\cdot, \tau)$ is convex for each $i = 1, 2, \ldots, p(\bar{x})$, we then ensure the convexity of $\hat{g}_i(\cdot)$ ($i = 1, 2, \ldots, p(\bar{x})$) in $U(\bar{x})$.

We can compute the values of $\nabla \tau_i(\bar{x})$ for $i = 1, 2, \ldots, p(\bar{x})$ by solving a certain linear system derived from the implicit function theorem, from which we further obtain the values of $\nabla \hat{g}_i(\bar{x})$ and $\nabla^2 \hat{g}_i(\bar{x})$ for each $i$. Thanks to this result, we acquire the concrete forms of the quadratic programs (QPs) that arise in the SQP iterations for (3.1), although it is difficult in general to know explicit forms of the functions $\tau_i^j(\cdot)$ ($i = 1, 2, \ldots, p(\bar{x}))$. See (3.17) and (3.19)-(3.22), which will be shown later.

The following proposition is concerned with the relationship between KKT points of SISDP (1.1) and those of the locally equivalent NSDP (3.1) under the $\delta$-nondegeneracy assumption. The proof is deferred to Appendix.

**Proposition 3.1** Let $u^* := (x^*, y^*, V_u) \in \mathcal{R}^n \times \mathcal{M}_+(T) \times S_+^m$ be a KKT point of SISDP (1.1) and suppose that $x^*$ is $\delta$-nondegenerate for some $\delta > 0$ so that SISDP (1.1) is represented as NSDP (3.1) with $\bar{x} = x^*$ equivalently around $x^*$. Then, $y^*$ is discrete and

$$\{\tau_1^1(x^*), \tau_2^2(x^*), \ldots, \tau_{p(x^*)}(x^*)\} \supseteq \text{supp}(y^*).$$

In addition, $z^* := (x^*, z^*, V_u) \in \mathcal{R}^n \times \mathcal{R}_+^{p(x^*)} \times S_+^m$ with $z^* := y^*([\tau_1^i(x^*)])$ ($i = 1, 2, \ldots, p(x^*))$ is a KKT point of NSDP (3.1) with $\bar{x} = x^*$.  

**Remark 3.2** (About connection of Proposition 3.1 and Theorem 2.3) Consider a sequence $\{w^k\} = \{(x^k, y^k, V_k)\}$ generated by Algorithm 1. Suppose that $\{y^k\} \subseteq \mathcal{M}_+(T)$ is a discrete measure sequence and $|\text{supp}(y^k)| \leq M$ holds for all $k$ with some $M > 0$. Then, by Theorem 2.3, there exists a weak* accumulation point $w^* := (x^*, y^*, V_u) \in \mathcal{R}^n \times \mathcal{M}_+(T) \times S_+^m$ such that $y^*$ satisfies (2.12). Additionally, if $x^*$ is $\delta$-nondegenerate for some $\delta > 0$, then $y^*$ satisfies condition (3.3) from Proposition 3.1. Let $\{t^k\} \subseteq T^M$ with $t^k := (\tau^k_1, \tau^k_2, \ldots, \tau^k_M)$ and $\tau^*_1, \tau^*_2, \ldots, \tau^*_M \in T$ as in item (ii) of Theorem 2.3. Since $\tau^*_1, \tau^*_2, \ldots, \tau^*_M \supseteq \text{supp}(y^*)$ by (2.12), (3.3) implies

$$\{\tau^*_1, \tau^*_2, \ldots, \tau^*_M\} \supseteq \left(\text{supp}(y^*) \cap \{\tau_1^1(x^*), \tau_2^2(x^*), \ldots, \tau_{p(x^*)}(x^*)\}\right) =: A_*.$$

From this fact, we see that, for any element of the set $A_*$, say $\hat{\tau}^*_i(x^*)$, there exists a subsequence $\{\tau^k_i\}_{k \in \mathcal{K}} \subseteq T$ with some $i \in \{1, 2, \ldots, M\}$ such that $\tau^k_i \in \text{supp}(y^k)$ for each $k \in \mathcal{K}$ and $\lim_{k \to \infty} \tau^k_i = \hat{\tau}^*_i(x^*)$.

---

4 We say that $z^*$ is a KKT point of NSDP (3.1) with $\bar{x} = x^*$ if $\nabla f(x^*) + \sum_{i=1}^{p(x^*)} \nabla \hat{g}_i(x^*) \zeta^* = -(F_j \bullet V_u)_{j=1}^q = 0, 0 \leq \zeta^* \perp \hat{g}(x^*) \leq 0$, and condition (2.2) holds with $V = V_u$. 

[Springer]
3.2 Local version of the proposed algorithm

Throughout this section, we let $w^* = (x^*, y^*, V_*) \in \mathcal{R}^n \times \mathcal{M}_+ (T) \times S^n_{++}$ be a KKT point of SISDP (1.1) and assume that $x^*$ is $\delta$-nondegenerate for some $\delta > 0$. Consider NSDP (3.1) with $\bar{x} = x^*$ whose semi-infinite constraints are represented as $g(x, \tau^i_{x^*}(x)) \leq 0 \ (i = 1, 2, \ldots, p(x^*))$. Then, $z^* := (x^*, \zeta^*, V_*) \in \mathcal{R}^n \times \mathcal{R}^{p(x^*)} \times S^m_+$ defined as in Proposition 3.1 is a KKT point of this NSDP. The local algorithm is supposed to solve this NSDP, which is denoted by NSDP*.

3.2.1 The structure of the local algorithm

The local algorithm works in the neighborhood $U(x^*)$, where $U(x^*)$ is the set $U(\bar{x})$ with $\bar{x} = x^*$ defined in Sect. 3.1. This implies that each generated point $x^k$ is contained by $U(x^*)$. Thus, $x^k$ is $\delta$-nondegenerate and the implicit functions defined around $x^k$ are identical to those around $x^*$, namely, $\{\tau^i_{x^k}()\}_{i=1}^{p(x^k)} = \{\tau^i_{x^*}()\}_{j=1}^{p(x^*)}$ and thus $p(x^k) = p(x^*)$ for each $k \geq 0$. This means that NSDP (3.1) with $\bar{x} = x^k$ coincides with NSDP*, that is, the structure of NSDP* is eventually identified. In what follows, we suppose that

$$\tau^i_{x^k}(\cdot) = \tau^i_{x^*}(\cdot) \ (i = 1, 2, \ldots, p(x^*)).$$  \hspace{1cm} (3.4)

Moreover, we suppose that other related points such as $x^{k+\frac{1}{2}}$ and $x^k_+$ shown below are $\delta$-nondegenerate and the set of implicit functions at any such point is common to those at $x^*$.

Remark 3.3 To establish locally fast convergence, essentially the same conditions as above are usually assumed in local reduction-type methods, for example, [31, Assumption 2(b)]. If condition (3.4) is judged invalid, we need to execute a safeguard so as to maintain convergence to a KKT point of the SISDP. Actually, Step 2.4 of Algorithm 2 given later is intended to play such a role.

The algorithm produces a sequence $\{z^k\} := \{(x^k, \zeta^k, V_k)\} \subseteq \mathcal{R}^n \times \mathcal{R}^{p(x^*)} \times S^m_+$, where $\zeta^k_i \geq 0$ is a Lagrange multiplier estimate of the inequality constraint

$$\hat{g}_i(x) := g(x, \tau^i_{x^k}(x)) \leq 0$$

for $i = 1, 2, \ldots, p(x^k)(= p(x^*))$. Note that $g(x, \tau^i_{x^k}(x)) = g(x, \tau^i_{x^*}(x))$ for each $i$ by (3.4). Similarly to Algorithm 1, $\{z^k\}$ is a sequence of approximate BKKT points of NSDP*. Specifically, for a given sequence $\{(\varepsilon_k, \mu_k)\}$ converging to $(0, 0)$,

$$z^k \in \tilde{N}^{\varepsilon_k}_{\mu_k}$$

holds for each $k$, where for $\varepsilon > 0$ and $\mu > 0$,

$$\tilde{N}^{\varepsilon}_{\mu} := \{z = (x, \zeta, V) \in \mathcal{R}^n \times \mathcal{R}^{p(x^*)} \times S^m_+ \mid \hat{R}_\mu(z) \leq \varepsilon, \ F(x) \in S^m_+ \}$$  \hspace{1cm} (3.5)
with
\[ \hat{R}_\mu(z) := \sqrt{\psi(x)^2 + \| \hat{\varphi}_1(x, \zeta, V)^2 + \| \varphi_2(x, \zeta) \|^2 + \| \varphi_3(x, V, \mu) \|^2}, \]
\[ \hat{\varphi}_1(x, \zeta, V) := \nabla f(x) + \sum_{i=1}^{p(x^*)} \nabla \hat{g}_i(x) \xi_i - (F_j \cdot V)^n_{j=1}, \]
\[ \hat{\varphi}_2(x, \zeta) := \sum_{i=1}^{p(x^*)} \hat{g}_i(x) \xi_i, \]
and \( \psi, \varphi_3 \) being the functions defined in Sect. 2.2.

Actually, \( z^k \) is identified with \( w^k := (x^k, y^k, V_k) \in \mathbb{R}^n \times \mathcal{M}_+(T) \times S^m_+ \) where \( y^k(A) := \sum_{i=1}^{p(x^*)} \zeta_i \mathcal{A}(\{ \tau^i_x(x^k) \}) \) \( (A \in \mathcal{B}) \). In particular, we have
\[ w^k \in N_{\mu_k}^{e_k} \iff z^k \in \hat{N}_{\mu_k}^{e_k}, \quad (3.6) \]
which indicates that Algorithm 1 underlies the proposed algorithm.

In every iteration, we produce two kinds of search directions
\[ \Delta_1 z^k := (\Delta_1 x^k, \Delta_1 \zeta^k, \Delta_1 V_k), \quad \Delta_1 x^k := \left( \Delta_1 x^k, \Delta_1 \zeta^k, \Delta_1 V_k \right) \]
in \( \mathbb{R}^n \times \mathbb{R}^{p(x^*)} \times S^m \), along which we determine step sizes \( s^k_j \left( j = \frac{1}{2}, 1 \right) \) by
\[ s^k_j = \min(t^k_j, u^k_j), \quad (3.7) \]
where
\[ t^k_j := \begin{cases} \frac{-\sigma}{\lambda_{\min}(F(x^{k+\frac{j}{2}}-\frac{1}{2}) - \sum_{i=1}^{n} \Delta_j x^k F_i)} & \text{if } \lambda_{\min} \left( F(x^{k+\frac{j}{2}}-\frac{1}{2}) - \sum_{i=1}^{n} \Delta_j x^k F_i \right) \leq -1 \\ 1 & \text{otherwise,} \end{cases} \]
\[ u^k_j := \begin{cases} \frac{-\sigma}{\lambda_{\min}(V_{k+\frac{j}{2}} - \sum_{i=1}^{n} \Delta_j V_{k+\frac{j}{2}})} & \text{if } \lambda_{\min} \left( V_{k+\frac{j}{2}} - \sum_{i=1}^{n} \Delta_j V_{k+\frac{j}{2}} \right) \leq -1 \\ 1 & \text{otherwise,} \end{cases} \]
for \( j = \frac{1}{2}, 1 \). Here, \( \sigma \in (0, 1) \) is a prescribed algorithmic constant, \( \lambda_{\min}(\cdot) \) denotes the minimum eigenvalue, and \( (x^{k+\frac{1}{2}}, V_{k+\frac{1}{2}}) \) is defined as
\[ (x^{k+\frac{1}{2}}, V_{k+\frac{1}{2}}) := \left( x^k + s^k_{\frac{1}{2}} \Delta_1 x^k, V_k + s^k_{\frac{1}{2}} \Delta_1 V_k \right). \]
In fact, by (3.7), $s^k_1, s^k_2 \in (0, 1]$ hold. Moreover, since
\[
\bar{s} := \sup \{s \mid \lambda_{\min}(X + s \Delta X) \geq 0, s \geq 0\} = \begin{cases} 
\frac{1}{\lambda_{\min}(X^{-1} \Delta X)} & \text{if } \lambda_{\min}(X^{-1} \Delta X) < 0 \\
\infty & \text{otherwise}
\end{cases}
\] (3.8)
holds for any $X \in S^m_{++}$ and $\Delta X \in S^m$. \(^5\) the following interior point conditions are fulfilled:
\[
F(x^k + s^k_1 \Delta_1 x^k) \in S^m_{++}, \ F(x^k + s^k_2 \Delta_1 x^k + s^k_1 \Delta_1 x^k) \in S^m_{++},
\]
\[
V_k + s^k_1 \Delta_1 V_k \in S^m_{++}, \ V_k + s^k_2 \Delta_1 V_k + s^k_1 \Delta_1 V_k \in S^m_{++}.
\]
We remark here that if $X + \Delta X \in S^m_{++}$, then $\bar{s} > 1$ and hence (3.8) necessarily implies
\[
\lambda_{\min}(X^{-1} \Delta X) > -1.
\] (3.9)

To attain fast convergence of $\{z^k\}$, we try to follow the central path closely by updating the barrier parameter $\mu_k$ in such a way that $\mu_k + 1 = o(\mu_k)$ and solving certain systems to have the search directions $\Delta_1 z^k$ and $\Delta_2 z^k$. The systems are actually related to the BKKT conditions for NSDP (3.1) with $\bar{x} = x^k$. When those directions turn out to be unsuccessful, a point near the central path is computed (cf. Step 2-4 of Algorithm 2 shown below). Before describing the details, we first show the overall structure of the proposed algorithm:

**Algorithm 2** (Local primal-dual path following method)

**Step 0** (Initial setting): Choose parameters
\[
0 < \alpha < 1, \ 0 < \beta < 1, \ \gamma_1, \gamma_2 > 0, \ \delta > 0, \ \sigma \in (0, 1),
\mu_0 > 0, \ 0 < c \leq \frac{1}{\alpha + 2}.
\]
Set $\varepsilon_0 := \gamma_1 \mu_0^{1+\alpha}$. Moreover, choose an initial iteration point $z^0 := (x^0, \xi^0, 0) \in R^n \times R^p(x^*) \times S^m_{++}$ such that $F(x^0) \in S^m_{++}$. Let $k := 0$.

**Step 1** (Stopping rule): Stop if
\[
\hat{R}_0(z^k) = 0, \ F(x^k) \in S^m_+, \ V_k \in S^m_+, \ \xi^k \in R^p(x^*).
\]
Otherwise, go to Step 2.

**Step 2** (Computing an approximate BKKT point): Perform the following procedure:
\(^5\) For $X \in S^m_{++}, Y \in S^m$, the eigenvalues of $X^{-1} Y$ are all real numbers.
Step 2-1: Compute $S_{\delta}(x^k)$ (see (3.2)). Choose a scaling matrix $P_k$ and obtain $\Delta_{\frac{1}{2}}z^k := (\Delta_{\frac{1}{2}}x^k, \Delta_{\frac{1}{2}}\bar{z}^k, \Delta_{\frac{1}{2}}V_k)$ by solving the mixed linear complementarity system (3.13), (3.15), and (3.16), which amounts to solving QP (3.17) (see Sect. 3.2.2) with $\mu = \mu_k$, $P = P_k$ and $\bar{z} = z^k$. Compute $s_{\frac{1}{2}}^k$ by (3.7) with $j = \frac{1}{2}$ and define $z^k + \frac{1}{2} := (x^k + \frac{1}{2}, \bar{z}^k + \frac{1}{2}, V_{k+\frac{1}{2}})$ by setting $(x^k + \frac{1}{2}, V_{k+\frac{1}{2}}) := (x_k + s_{\frac{1}{2}}^k \Delta_{\frac{1}{2}}x_k, V_k + s_{\frac{1}{2}}^k \Delta_{\frac{1}{2}}V_k)$ and

$$\zeta^k + \frac{1}{2} := \zeta^k + \Delta_{\frac{1}{2}}\zeta^k.$$  \hspace{1cm} (3.10)

**Step 2-2:** Compute $S_{\delta}(x^k + \frac{1}{2})$ (see (3.2)). Choose a scaling matrix $P_{k+\frac{1}{2}}$. If the system of linear equations (3.19)–(3.22) (see Sect. 3.2.2) with $\mu = \mu_k$, $P = P_{k+\frac{1}{2}}$ and $\bar{z} = z^k + \frac{1}{2}$ is unsolvable, then go to Step 2-4. Otherwise, compute the solution $\Delta_{\frac{1}{2}}\tilde{z}^k := (\Delta_{\frac{1}{2}}x^k, \Delta_{\frac{1}{2}}\bar{z}^k, \Delta_{\frac{1}{2}}\tilde{V}_k)$, from which we derive $\Delta_{\frac{1}{2}}z^k := (\Delta_{\frac{1}{2}}x^k, \Delta_{\frac{1}{2}}\bar{z}^k, \Delta_{\frac{1}{2}}V_k)$. Moreover, compute $s_{\frac{1}{2}}^k$ by (3.7) with $j = 1$, and define $z^k_+ := (x^k_+, \bar{z}^k_+, V^k_+)$ by setting $(x^k_+, V^k_+) := (x_k^k + s_{\frac{1}{2}}^k \Delta_{\frac{1}{2}}x_k, V_k + s_{\frac{1}{2}}^k \Delta_{\frac{1}{2}}V_k)$ and

$$\zeta^k_+ := \zeta^k + \Delta_{\frac{1}{2}}\zeta^k.$$  \hspace{1cm} (3.11)

Step 2-3: If $z^k_+ \in \mathcal{N}_{\mu_k}^\epsilon$, set $z^{k+1} := z^k_+$, and go to Step 3. Otherwise, go to Step 2-4.

**Step 2-4:** Compute $z^{k+1} \in \mathcal{N}_{\mu_k}^\epsilon$.

**Step 3** (Update): Update the parameters $\mu_k$ and $\epsilon_k$ as

$$\mu_{k+1} := \min \left( \beta \mu_k, \gamma_2 \mu_k^{1+c\alpha} \right), \epsilon_{k+1} := \gamma_1 \mu_{k+1}^{1+c\alpha}.$$  \hspace{1cm} (3.12)

Set $k := k + 1$ and return to Step 1.

In (3.10) of Step 2-1, we update $\zeta^k$ with the unit step size along $\Delta_{\frac{1}{2}}\zeta^k$ like the standard SQP method. In (3.11) of Step 2-2, $\zeta^k + \frac{1}{2}$ is updated in a similar way. If Steps 2-1–2-3 generate $z^k$ successfully, $\{z^k\}$ is expected to converge to a KKT point of NSDP superlinearly. If not, Step 2-4 plays a role of safeguarding the global convergence of $\{z^k\}$. We will specify how Step 2-4 is executed in the global version (Algorithm 3) of the algorithm.

### 3.2.2 Computing the directions $\Delta_{\frac{1}{2}}z^k$ and $\Delta_{\frac{1}{2}}\bar{z}^k$

**First direction $\Delta_{\frac{1}{2}}z$**

Let $\bar{z} = (\bar{x}, \bar{\zeta}, \bar{V}) \in \mathcal{R}^n \times \mathcal{R}^{n(\bar{x})} \times S^{m}_{++}$ be a current point such that $F(\bar{x}) \in S^{m}_{++}$ and $\bar{x}$ is $\delta$-nondegenerate. We show that a first search direction $\Delta_{\frac{1}{2}}z = (\Delta_{\frac{1}{2}}x, \Delta_{\frac{1}{2}}\bar{z}, \Delta_{\frac{1}{2}}V)$
$\in \mathcal{R}^n \times \mathcal{R}^{p(\bar{x})} \times S^m$ can be computed through the local reduction method in a manner similar to the interior-point SQP method proposed in [20].

To start with, we apply the Monteiro-Zhang scaling to $F(x)$ and $V$, in which we select a nonsingular matrix $P \in \mathcal{R}^{m \times m}$ and scale the matrices $F(x)$ and $V$ as

$$F_p(x) := PF(x)P^T = F_0 + \sum_{i=1}^{n} x_i F_p^i, \quad V_p := P^{-T}VP^{-1},$$

where $F_p^i := PF_iP^T$ for $i = 0, 1, 2, \ldots, n$. Let us consider NSDP (3.1) with $F(x) \in S^m_+$ replaced by $F_p(x) \in S^m_+$, called a scaled NSDP. Since $F(x) \circ V = \mu I, \ F(x) \in S^m_+, V \in S^m_+$ if and only if $F_p(x) \circ V_p = \mu I, \ F_p(x) \in S^m_+, \ V_p \in S^m_+$ for any $\mu \geq 0$, the KKT and BKKT conditions of the reduced NSDP (3.1) are equivalent to those of the scaled NSDP (3.1). Therefore, to produce a search direction, it is natural to solve the following mixed linear complementarity system approximating the BKKT system of the scaled NSDP:

\[
\nabla f(\bar{x}) + \nabla^2_{xx}L(\bar{x}, \bar{\zeta}, \bar{V})\Delta_\frac{1}{2}x + \nabla \hat{g}(\bar{x})(\bar{\zeta} + \Delta_\frac{1}{2}\xi) - \left( F_p \bullet \left( V_p + \Delta_\frac{1}{2}V_p \right) \right)_{i=1}^{n} = 0, \tag{3.13}\nabla \hat{g}(\bar{x}) \perp \bar{\zeta} + \Delta_\frac{1}{2}\xi = 0, \tag{3.15}
\]

where $L : \mathcal{R}^n \times \mathcal{R}^{p(\bar{x})} \times S^m \rightarrow \mathcal{R}$ is the Lagrangian of the NSDP defined by $L(x, \xi, V) := f(x) + \sum_{i=1}^{p(\bar{x})} \hat{g}_i(x)\xi_i - F(x) \bullet V$. Moreover, for any $X \in S^m$, the linear operator $\mathcal{L}_X : S^m \rightarrow S^m$ is defined by $\mathcal{L}_X(Z) := X \circ Z$.

In our method, we make a slight modification to the above system. Specifically, we replace the second equation (3.14) with the following equation:

$$F_p(\bar{x}) \circ (V_p + \Delta_\frac{1}{2}V_p) + \frac{1}{2} \left( \mathcal{L}_{V_p} + \mathcal{L}_{F_p(\bar{x})}\mathcal{L}_{V_p}^{-1} \mathcal{L}_{F_p(\bar{x})}^{-1} \right) \sum_{i=1}^{n} \Delta_{\frac{1}{2}}x_i F_p^i = \mu I. \tag{3.16}$$

The second term of the left hand side approximates $\mathcal{L}_{V_p} \sum_{i=1}^{n} \Delta_{\frac{1}{2}}x_i F_p^i$ around a BKKT point. Actually, at any BKKT point, those two expressions are identical to each other since $\mathcal{L}_{F_p(\bar{x})}$ and $\mathcal{L}_{V_p}$ commute there. Particularly when choosing a scaling matrix $P$ so that $F_p(\bar{x})$ and $V_p$ commute, (3.14) and (3.16) become identical to each other.
The reason for using the system (3.13), (3.15), and (3.16) is that it can be solved via the KKT system of the following quadratic program (QP):

\[
\begin{align*}
\text{Minimize } & \nabla f(\bar{x})^\top \Delta x + \frac{1}{2} \Delta x^\top B_P(\bar{x}, \xi, \bar{V}) \Delta x - \mu \xi_P(\bar{x})^\top \Delta x \\
\text{subject to } & \hat{g}(\bar{x}) + \nabla \hat{g}(\bar{x})^\top \Delta x \leq 0,
\end{align*}
\]

\[\text{(3.17)}\]

where \(\xi_P(\cdot) := \nabla \log \det F_P(\cdot) = (F_P^i \cdot F_P(\cdot)^{-1})_{i=1}^n, \hat{g}(\cdot) := (\hat{g}_1(\cdot), \hat{g}_2(\cdot), \ldots, \hat{g}_p(\bar{x})(\cdot))^\top,\]

\[B_P(x, \xi, V) := \nabla_{xx}^2 L(x, \xi, V) + H_P(x, V),\]

and \(H_P(x, V)\) is the symmetric matrix whose elements are defined by

\[\left(H_P(x, V)\right)_{i,j} := \frac{1}{2} F_P^i \cdot \left(\mathcal{L}_{F_P(x)}^{-1} \mathcal{L}_{V_P} + \mathcal{L}_{V_P} \mathcal{L}_{F_P(x)}^{-1}\right) F_P^j\]

\[\text{(3.18)}\]

for \(i, j = 1, 2, \ldots, n\). Note that the linear operator \(\mathcal{L}_{F_P(x)}\) is invertible when \(F(x) \in S^m_{++}\). Denote a KKT pair of QP (3.17) by \(\left(\Delta_1 x, \xi, \Delta_1 V\right)\) and define

\[
\Delta_1 V := \mu F(\bar{x})^{-1} - \bar{V} - \sum_{i=1}^n \Delta_1 x_i P^\top \frac{1}{2} \left(\mathcal{L}_{F_P(\bar{x})}^{-1} \mathcal{L}_{V_P} + \mathcal{L}_{V_P} \mathcal{L}_{F_P(\bar{x})}^{-1}\right) F_P^i P,
\]

\[
\Delta_1 V_P := P^{-\top} \Delta_1 V P^{-1}.
\]

Then, we can verify that the triple \(\left(\Delta_1 x, \Delta_1 \xi, \Delta_1 V\right)\) solves the system (3.13), (3.15), and (3.16).

QP (3.17) is necessarily feasible if the original problem (1.1) is feasible, since the functions \(\hat{g}_i(\cdot) (i = 1, 2, \ldots, p(\bar{x}))\) are convex as mentioned in Sect. 3.1. Furthermore, we have the following property concerning the strong convexity of the objective function of QP (3.17).

**Proposition 3.2** Suppose that \(F_P(x) \in S^m_{++}, V_P \in S^m_{++}, \) and \(F_1, F_2, \ldots, F_n\) are linearly independent in \(S^m\). Also, suppose that either of the following is true:

(i) \(\|F_P(x) \circ V_P - \mu I\| \leq \theta \mu\) with \(0 \leq \theta < 1\);

(ii) \(F_P(x)\) and \(V_P\) commute.

Then, \(H_P(x, V)\) is positive definite. Especially, if \(f\) is convex and \(\bar{\xi} \in \mathcal{R}^{p(\bar{x})}_+,\) the objective function of QP (3.17) is strongly convex. Therefore, it has a unique optimum.

**Proof** Note that \(F_P(x) \circ V_P \in S^m_{++}\) holds if either of the assumptions (i) and (ii) holds. Then, the operators \(\mathcal{L}_{F_P(x)} \mathcal{L}_{V_P}\) and \(\mathcal{L}_{V_P} \mathcal{L}_{F_P(x)}\) are positive definite. Actually, for any \(D \in S^m \setminus \{0\}, D \cdot \mathcal{L}_{F_P(x)} \mathcal{L}_{V_P} D = D \cdot \mathcal{L}_{V_P} \mathcal{L}_{F_P(x)} D = \text{Tr}(D(F_P(x) \circ V_P)D) > 0\). Then, letting \(\Delta F := \sum_{i=1}^n \Delta x_i F_P^i\) and noting the linear independence of \(F_1, F_2, \ldots, F_n\) in \(S^m\), we obtain
\[2\Delta x^\top H_P(x, V)\Delta x = \Delta F \cdot \left(\mathcal{L}_{F_P(x)}^{-1}\mathcal{L}_{V_P} + \mathcal{L}_{V_P}\mathcal{L}_{F_P(x)}^{-1}\right) \Delta F = \mathcal{L}_{F_P(x)}^{-1}(\Delta F) \cdot \left(\mathcal{L}_{V_P}\mathcal{L}_{F_P(x)} + \mathcal{L}_{F_P(x)}\mathcal{L}_{V_P}\right)\mathcal{L}_{F_P(x)}^{-1}(\Delta F) > 0\]

for any \(\Delta x \neq 0\). We omit the proof for the latter claim. \(\square\)

Below, we list some particular choices for the scaling matrix \(P\) and the corresponding direction \(\Delta_{\frac{1}{2}} V\):

(i) \(P = I\): In this case, \(F_P(\bar{x}) = F(\bar{x})\) and

\[\Delta_{\frac{1}{2}} V = \mu F(\bar{x})^{-1} - \bar{V} - \frac{1}{2} \left(\mathcal{L}_{F(\bar{x})}^{-1}\mathcal{L}_{V} + \mathcal{L}_{V}\mathcal{L}_{F(\bar{x})}^{-1}\right) \sum_{i=1}^{n} \Delta_{\frac{1}{2}} x_i F_i.\]

(ii) \(P = F(\bar{x})^{-\frac{1}{2}}\): In this case, \(F_P(\bar{x}) = I\) and

\[\Delta_{\frac{1}{2}} V = \mu F(\bar{x})^{-1} - \bar{V} - \frac{1}{2} F(\bar{x})^{-1} \left(\sum_{i=1}^{n} \Delta_{\frac{1}{2}} x_i F_i\right) \bar{V}\]

\[+ \bar{V} \left(\sum_{i=1}^{n} \Delta_{\frac{1}{2}} x_i F_i\right) F(\bar{x})^{-1}.\]

(iii) \(P = W^{-\frac{1}{2}}, W := F(\bar{x})^\frac{1}{2}(F(\bar{x})^\frac{1}{2}V F(\bar{x})^\frac{1}{2})^{-\frac{1}{2}} F(\bar{x})^\frac{1}{2}\): In this case, \(F_P(\bar{x}) = V_P\) and

\[\Delta_{\frac{1}{2}} V = \mu F(\bar{x})^{-1} - \bar{V} - W^{-1} \left(\sum_{i=1}^{n} \Delta_{\frac{1}{2}} x_i F_i\right) W^{-1}.\]

The direction \(\Delta_{\frac{1}{2}} V\) obtained as above can be related to the family of Monteiro-Zhang (MZ) directions [35]. Actually, as for (i), if \(F_P(\bar{x})\) and \(V_P\) commute, the generated direction can be cast as the Alizadeh-Hareberly-Overton (AHO) direction. On the other hand, the generated directions in (ii) and (iii) are nothing but the Helmberg-Rendle-Vanderbei-Wolkowicz/Kojima-Shindoh-Hara/Monteiro (HRVW/KSH/M) and Nesterov-Todd (NT) directions, respectively, by themselves.

**Second direction \(\Delta_1 z\)**

We next show how to compute the second direction \(\Delta_1 z = (\Delta_1 x, \Delta_1 \zeta, \Delta_1 V)\) at \(\bar{z} + s \Delta_{\frac{1}{2}} z\), where \(s\) is a step size. In a manner similar to \(\Delta_{\frac{1}{2}} z\), we may compute the second direction \(\Delta_1 z\) by solving \(QP(3.17)\) with \(\bar{z}\) and \(P\) replaced by \(\bar{z} + s \Delta_{\frac{1}{2}} z\) and another scaling matrix \(\hat{P} \in \mathcal{R}^{m \times m}\), respectively. However, by exploiting information associated to \(\Delta_{\frac{1}{2}} x\), we can replace the QP with certain linear equations as follows: Let

\[J_d(\bar{x}) := \{i \in \{1, 2, \ldots, p(\bar{x})\} \mid \hat{g}_i(\bar{x}) + \nabla \hat{g}_i(\bar{x})^\top \Delta_{\frac{1}{2}} x = 0\}.\]
If the current point \( \bar{x} \) is sufficiently close to \( x^* \) that forms the KKT point \( \bar{z}^* \), we can expect that the inequality constraints \( \hat{g}_i(x) \leq 0 \) \((i \in J_a(\bar{x}))\) are also active at \( x^* \).

Motivated by this observation, we propose to solve the following linear equations for \( \Delta \hat{\bar{z}} := (\Delta_1 x, \Delta_1 \xi, \Delta_1 \hat{V}) \):

\[
\nabla f(\hat{x}) + \nabla^2 f(\hat{x}) L(\hat{x}, \hat{\xi}, \hat{V}) \Delta_1 x + \nabla \hat{g}(\hat{x})(\hat{\xi} + \Delta_1 \xi)
= - \left( F_p^i \bullet \left( \hat{V}_p + \Delta_1 \hat{V} \right) \right)_{i=1}^n = 0,
\]

\[
F_p(\hat{x}) \circ (\hat{V}_p + \Delta_1 \hat{V}) + \frac{1}{2} \left( \mathcal{L}_{\hat{V}_p} + \mathcal{L}_{F_p(\hat{x})} \mathcal{L}_{\hat{V}_p} \mathcal{L}_{F_p(\hat{x})}^{-1} \right) \sum_{i=1}^n \Delta_1 x_i F_p^i = \mu I, \tag{3.19}
\]

\[
\hat{g}(\hat{x}) + \nabla \hat{g}(\hat{x})^\top \Delta_1 x = 0 \quad (i \in J_a(\bar{x})), \tag{3.20}
\]

\[
\hat{\xi}_i + \Delta_1 \xi_i = 0 \quad (i \notin J_a(\bar{x})), \tag{3.21}
\]

\[
\hat{\xi} + \Delta_1 \xi = 0 \quad (i \notin J_a(\bar{x})), \tag{3.22}
\]

where \((\hat{x}, \hat{V}) := (\bar{x} + s\Delta_1 x, \bar{V} + s\Delta_1 V), \hat{V}_p := \hat{P}^{-1} \hat{V} \hat{P}^{-1}, \) and \( \hat{\xi} := \bar{\xi} + \Delta_1 \zeta \).

Once we obtain \( \Delta_1 \hat{V} \), we set \( \Delta_1 V := \hat{P}^\top \Delta_1 \hat{V} \hat{P}. \) If the above linear equations are not solvable or not well-defined because \( \{ \tau_i(\cdot) \}_{i \in J_a(\bar{x})} \subseteq \{ \tau_i(\cdot) \}_{1 \leq i \leq p(\bar{x})} \) does not hold, i.e., the family of functions \( \hat{g}_i(\cdot) \ (i \in J_a(\bar{x})) \) defined at \( \bar{x} \) is not valid at \( \hat{x} \), then we skip the above procedure and proceed to the next step.

### 3.2.3 Local convergence analysis of Algorithm 2

In this section, we focus on the case where the identity matrix is selected as a scaling matrix, i.e., \( P = I \). Accordingly, \( F_p(x) = F(x) \) and \( V_p = V \) hold throughout the section.

For the other cases where scaling matrices corresponding to HRVW/KSH/M and NT directions are used (recall (ii) and (iii) in the subsection 'First direction \( \Delta_1 \hat{z}^* \), in Sect. 3.2.2), we can also show results similar to the ones given below in a manner analogous to [41, Theorems 3.4].

In what follows, we will make two sets of assumptions. In order to describe the assumptions, we define additional functions and notations. Denote by \( I_a(x^*) \) the set of indices corresponding to the active inequality constraints at \( x^* \) among \( \hat{g}_i(x) \leq 0, \ldots, \hat{g}_{p(x^*)}(x) \leq 0 \), i.e., \( I_a(x^*) := \{ i \in \{ 1, 2, \ldots, p(x^*) \} \mid \hat{g}_i(x^*) = 0 \} \). For a fixed barrier parameter \( \mu > 0 \), we consider the function \( \Phi_{\mu} : \mathcal{R}^n \times \mathcal{R}^{\vert I_a(x^*) \vert} \times \mathcal{R}^{m(m+1)/2} \rightarrow \mathcal{R}^d \) with \( d := n + \vert I_a(x^*) \vert + m(m+1)/2 \) defined by

\[
\Phi_{\mu}(\bar{z}) := \left( \nabla f(x) + \sum_{i \in I_a(x^*)} \xi_i \nabla \hat{g}_i(x) - (F_j \bullet V)_{j=1}^n \right)_{svec(F(x) \circ V - \mu I)}, \tag{3.23}
\]

where we write

\[
\bar{z} := (x, \bar{\xi}, svec(V)) \in \mathcal{R}^n \times \mathcal{R}^{\vert I_a(x^*) \vert} \times \mathcal{R}^{m(m+1)/2},
\]

\[
\hat{\zeta} := (\xi_i)_{i \in I_a(x^*)}.
\]
Next, we denote the Jacobian of $\Phi_\mu(\cdot)$ by $\mathcal{J} \Phi_\mu(\cdot)$, that is,

$$
\mathcal{J} \Phi_\mu(\tilde{z}) := \begin{pmatrix}
\nabla^2 f(x) + \sum_{i \in I_a(x^*)} \zeta_i \nabla^2 \hat{g}_i(x) & (\nabla \hat{g}_i(x))_{i \in I_a(x^*)} - (\text{svec}(F_1)^T) \\
(\nabla \hat{g}_i(x))_{i \in I_a(x^*)} & 0 \\
\text{svec}(L_V(F_1)) \cdots \text{svec}(L_V(F_n)) & 0 \\
\end{pmatrix},
$$

where $T_X \in \mathcal{R}^{m(m+1)/2 \times m(m+1)/2}$ is defined as the matrix such that $T_X \text{svec}(Y) := \text{svec}(L_X(Y))$ for any $X, Y \in \mathcal{S}^m$. Note that $\mathcal{J} \Phi_\mu = \mathcal{J} \Phi_0$ holds for any $\mu \geq 0$. Then, $\tilde{z}^* := (x^*, \tilde{\zeta}^*, \text{svec}(V_*))$ solves $\Phi_0(\tilde{z}) = 0$.

Now, the first set of assumptions is imposed on the functions $f$ and $g$, and the KKT point $\tilde{z}^*$ as follows:

**Assumption B:**

1. The functions $f$ and $g(\cdot, \tau)$ ($\tau \in T$) are three times continuously differentiable.
2. The strict complementarity condition holds for the semi-definite constraint and the inequality constraints, i.e.,

   $$
   F(x^*) + V_* \in \mathcal{S}^m_{++}, \quad -\hat{g}_i(x^*) + \zeta_i^* > 0 \ (i = 1, 2, \ldots, p(x^*)).
   $$

3. The Jacobian $\mathcal{J} \Phi_0(\tilde{z}^*)$ is nonsingular.

In a manner analogous to the proof of [41, Theorem 1], it is not difficult to verify Assumption B-3 under B-2 together with the second-order sufficiency and nondegeneracy conditions regarding NSDP*, where NSDP* was defined in the beginning of Sect. 3.1. We will observe that Assumptions B-2 and B-3 hold in most test problems solved in the numerical experiments. By the implicit function theorem, we can ensure the existence of a central path converging to $\tilde{z}^*$ under Assumption B-3. Namely, there exist some $\tilde{\mu} > 0$ and a unique smooth curve $\tilde{z}(\cdot) : (0, \tilde{\mu}] \to \mathcal{R}^{m} \times \mathcal{R}^{\|I_a(x^*)\|} \times \mathcal{R}^{m(m+1)/2}$ such that $\lim_{\mu \to 0} \tilde{z}(\mu) = \tilde{z}^*$ and $\tilde{z}(\mu)$ represents a BKKT point with barrier parameter $\mu \in (0, \tilde{\mu}]$.

Next, we give the second set of assumptions, which is imposed on the generated sequence.

**Assumption C:**

1. $\lim_{k \to \infty} z^k = z^*$.
2. For any $k$ large enough, $\tau_{x^*}^i(\cdot) = \tau_{x^*}^i(\cdot) = \tau_{x^*}^i(\cdot) (i = 1, 2, \ldots, p(x^*))$ holds.
3. The active inequality constraints at $x^*$ are eventually identified in the sense that $J_a^k = I_a(x^*)$ holds for any $k$ large enough, where

   $$
   J_a^k := \{ i \in \{1, 2, \ldots, p(x^*) \} \mid \hat{g}_i(x^*) + \nabla \hat{g}_i(x^*)^T \Delta_{+} x^k = 0 \}
   $$

   for each $k$. 

\( \Box \) Springer
Assumption C-1 is set in order to focus on the limiting behavior and convergence rate of \( \{\tilde{z}^k\} \). Nonetheless, it can be regarded as reasonable by taking into consideration that the central path \( \tilde{z}^* \) converges to \( \tilde{z}^\bullet \) as mentioned above and the algorithm generates points close to the path.

Under Assumptions C-2 and C-3, SISDP (1.1) can be regarded as an NSDP with the equality-constraints \( \hat{g}_i(x) = 0 \) \((i \in I_a(x^*)) \). C-2 is along the same line as (3.4), meaning that the structure of the implicit functions around \( x^* \) is identified at \( x^k \) and \( x^{k+\frac{1}{2}} \). Under Assumptions B and C-1, we may expect that C-2 and C-3 hold. We explain the reason only for C-2 below. C-3 holds by an argument similar to C-2. Recall that since \( x^* \) was assumed to be \( \delta \)-nondegenerate, the sets of implicit functions are identical for all \( x \in U(x^*) \). From C-1, we have \( x^k \in U(x^*) \) for all \( k \) large enough, implying that \( \tau_i^k(\cdot) = \tau_i^{k+\frac{1}{2}}(\cdot) \) \((i \in I_a(x^*)) \). From construction, \( x^{k+\frac{1}{2}} \) will be generated nearby \( x^k \) and thus eventually lie in \( U(x^*) \) as \( x^k \) gets closer to \( x^* \). This leads us to \( \tau_i^{k+\frac{1}{2}}(\cdot) = \tau_i^{k+\frac{1}{2}}(\cdot) \) \((i \in I_a(x^*)) \). In the numerical experiments, we will demonstrate that Assumptions C-2 and C-3 are reasonable in practice.

Before moving onto the main analysis, we derive important equations. Assumption C-3 along with the complementarity condition (3.15) with \( (\tilde{x}, \tilde{z}) = (x^k, \xi^k) \) yields that, for \( i \in \{1, 2, \ldots, p(x^*)\} \setminus I_a(x^*) \) and all \( k \) sufficiently large, \( \hat{g}_i(x^k) + \nabla \hat{g}_i(x^k)^\top \Delta_1 x^k < 0 \) and \( \xi_i^k + \Delta_1 \xi_i^k = 0 \), which together with (3.10) implies \( \xi_i^{k+\frac{1}{2}} = 0 \). Similarly, by (3.22) with \( \tilde{x} = x^k \) and \( J_a(\tilde{x}) \) replaced by \( J_a^{(k)} \), i.e., \( I_a(x^*) \) under Assumption C-3, we have \( \xi_i^{k+\frac{1}{2}} + \Delta_1 \xi_i^k = 0 \) \((i \in \{1, 2, \ldots, p(x^*)\} \setminus I_a(x^*)) \), and thus \( (\xi_i^{k+\frac{1}{2}}) = 0 \) by (3.11). Therefore, we can reduce the system (3.13), (3.15), and (3.16) with \( P = I \) and \( (\mu, \tilde{z}) = (\mu_k, z^k) \) and the linear equations (3.19)–(3.22) with \( \hat{P} = I \) and \( (\mu, \tilde{z}) = (\mu_k, z^k+\frac{1}{2}) \) to the following equations for \( j = 0 \) and \( j = \frac{1}{2} \), respectively:

\[
\Phi_{\mu_k}(\tilde{z}^{k+j}) + Q(\tilde{z}^{k+j}) \Delta_{j+\frac{1}{2}} \tilde{z} = 0, \tag{3.24}
\]

where \( \Delta_{j+\frac{1}{2}} \tilde{z} := \left( \Delta_{j+\frac{1}{2}} x, \Delta_{j+\frac{1}{2}} \tilde{z}, \text{svec} \left( \Delta_{j+\frac{1}{2}} V \right) \right)^\top \) and the function \( Q : \mathbb{R}^n \times \mathbb{R}^{I_a(x^*)} \times \mathbb{R}^{m(m+1)/2} \to \mathbb{R}^{d \times d} \) is defined by replacing the block consisting of \( \text{svec}(\mathcal{L}_V(F_i)) \), \( i = 1, 2, \ldots, n \) in \( \mathcal{J} \Phi_{\mu}(\tilde{z}) \) with the matrix with columns \( v_1, v_2, \ldots, v_n \) defined by

\[
v_i := \frac{1}{2} \text{svec} \left( \mathcal{L}_V F_i + \mathcal{L}_{F(x)} \mathcal{L}_V \mathcal{L}_{F(x)}^{-1} F_i \right) \quad (i = 1, 2, \ldots, n).
\]

**Remark 3.4** If \( f \) is strongly convex, then C-1 holds under Assumptions A and B-3 by virtue of Step 2.4 as safeguard. Moreover, under B-2, B-3, and C-1, we actually obtain C-3 by exploiting the fact that, for any \( k \) large enough, the \( x \)-component of the unique solution of the linear equation (3.24) coincides with \( \Delta_{\frac{1}{2}} x^k \). Then, since
\[ \lim_{k \to \infty} \Delta \frac{1}{2} x^k = 0 \text{ from B-3, both } x^k \text{ and } x^{k + \frac{1}{2}} \text{ lie in } U(x^*) \text{ for } k \text{ large enough, which yields C-2.} \]

**Technical results**

In this section, we provide two useful propositions before entering the essential part of the convergence analysis. See the Appendix for the proofs. The first proposition is associated with the limiting behavior of the operator \( L_{F(x^k)} \mathcal{L}_{V_k} L_{F(x^k)}^{-1} \) as \( k \) tends to \( \infty \).

**Proposition 3.3** Let \( (X^*, Y^*) \in S^m_+ \times S^m_+ \) satisfy the strict complementarity condition that \( X^* \circ Y^* = O \) and \( X^* + Y^* \in S^m_{++} \). Let \( \{\mu_r\} \subseteq R_+ \) and \( \{(X_r, Y_r)\} \subseteq S^m_+ \times S^m_+ \) be sequences such that \( \lim_{r \to \infty} \mu_r = 0 \), \( \lim_{r \to \infty} (X_r, Y_r) = (X^*, Y^*) \), and \( \|X_r \circ Y_r - \mu_r I\|_F = O(\mu_r^{1+\theta}) \) with \( \theta > 0 \). Then, \( \|L_{X_r} \mathcal{L}_{Y_r} L_{X_r}^{-1} - \mathcal{L}_{Y_r}\|_2 = O(\mu_r^\theta) \) and thus \( \lim_{r \to \infty} L_{X_r} \mathcal{L}_{Y_r} L_{X_r}^{-1} = \mathcal{L}_{Y^*}, \) where \( \|\cdot\|_2 \) denotes the operator norm, namely, for any linear operator \( T: S^m \to S^m \),
\[
\|T\|_2 := \sup \|X\|_F = 1 \|T(X)\|_F.
\]

The next proposition will be useful in proving that the interior point conditions \( F(x^{k+j-\frac{1}{2}} + s_j \Delta_j x^k) \in S^m_{+++} \) and \( V_{k+j-\frac{1}{2}} + s_j \Delta_j V_k \in S^m_{+++} \) are eventually satisfied by \( s_j \) for \( j = \frac{1}{2}, 1 \).

**Proposition 3.4** Let \( 0 < \theta < 1 \) and \( \{(X_r, Y_r)\} \subseteq S^m_{+++} \times S^m_{+++} \), \( \{\Delta X_r, \Delta Y_r\} \subseteq S^m \times S^m \), and \( \{\mu_r\} \subseteq R_+ \) be sequences such that \( \lim_{r \to \infty} \mu_r = 0 \),
\[
\|\Delta X_r \circ \Delta Y_r\|_F = O(\mu_r^2), \tag{3.25}
\]
\[
\|X_r \circ Y_r - \mu_r I\|_F = O(\mu_r^{1+\theta}). \tag{3.26}
\]

Moreover, let \( 0 < \hat{\theta} < 1 \) and \( \{\hat{\mu}_r\} \subseteq R_+ \) be a sequence such that \( \lim_{r \to \infty} \hat{\mu}_r = 0 \),
\[
\|Z_r - \hat{\mu}_r I\|_F = O(\hat{\mu}_r^{1+\hat{\theta}}), \tag{3.27}
\]
\[
\mu_r^2 = o(\hat{\mu}_r), \tag{3.28}
\]

where \( Z_r := X_r \circ Y_r + X_r \circ \Delta Y_r + Y_r \circ \Delta X_r. \) Then, we have \( X_r + \Delta X_r \in S^m_{+++} \) and \( Y_r + \Delta Y_r \in S^m_{+++} \) for any sufficiently large \( r \).

**Main convergence results**

In this section, we provide the main convergence results for the proposed local algorithm. For the sake of analysis, we choose a parameter \( \tilde{c} \) such that
\[
\frac{1}{2} < \tilde{c} < \frac{1 - c}{1 + c\alpha} < 1, \tag{3.29}
\]

\( \tilde{c} \) Springer
where \( c \) and \( \alpha \) are algorithmic parameters selected in Step 0. Such \( \tilde{c} \) exists, since \( 0 < \alpha < 1 \) and \( 0 \leq c \leq \frac{1}{\alpha + 2} \). In terms of \( \tilde{c} \), let us define the parameter sequence \( \{ \tilde{\epsilon}_k \} \) by

\[
\tilde{\epsilon}_k := \gamma_1 \mu_{k-1}^{1+c\alpha}
\]

for each \( k \). Note that the second inequality in (3.29) implies \((1+c\alpha)(1+\tilde{c}\alpha) < 1 + \alpha\). Then, from (3.12), we have

\[
\tilde{\epsilon}_k > \epsilon_k
\]

for all \( k \) large enough. Furthermore, the update rule (3.12) of \( \{ \mu_k \} \) and \( \{ \epsilon_k \} \) yields

\[
\mu_k = \gamma_2 \mu_{k-1}^{1+c\alpha}, \quad \epsilon_k = \gamma_1 \gamma_2 \mu_k^{1+c\alpha}(1+\alpha)
\]

for all \( k \) sufficiently large. Hereafter, we assume that the iteration number \( k \) is so large that (3.31) and (3.32) hold.

To show the final theorem concerning two-step superlinear convergence (see Theorem 3.1), we prove the following two propositions:

**Proposition 3.5** Suppose that Assumptions B and C hold. We have

1. the full step size \( s_{k+1}^k = 1 \) is eventually adopted in Step 2-1,
2. \( z_{k+1}^k \in \hat{N}_{\tilde{\epsilon}_k}^\mu, \) i.e., \( \hat{R}_{\mu_k}(z_{k+1}^k) \leq \tilde{\epsilon}_k \), \( F(x_{k+1}^k) \in S_{++}^m \), and \( V_{k+1}^k \in S_{++}^m \) for all \( k \) sufficiently large,
3. \( \| \Phi_0(z_{k+1}^k) \| = O(\mu_{k}^{1+c\alpha}) \), and
4. \( \| z_{k+1}^k - z^* \| = O(\| z_k - z^* \|^{1+c\alpha}) \).

**Proposition 3.6** Suppose that Assumptions B and C hold. We have

1. the full step size \( s_1^k = 1 \) is eventually adopted in Step 2-2,
2. \( z_{k+1}^k \in \hat{N}_{\epsilon_k}^\mu, \) i.e., \( \hat{R}_{\mu_k}(z_{k+1}^k) \leq \epsilon_k = \gamma_1 \mu_{k}^{1+c\alpha}, \) \( F(x_{k+1}^k + \Delta_1 x_k) \in S_{++}^m \), and \( V_{k+1}^k + \Delta_1 V_k \in S_{++}^m \) for all \( k \) sufficiently large, and
3. \( \| z_{k+1}^k + \Delta_1 z - z^* \| = O(\| z_{k+1}^k - z^* \|) \).

Items 1 and 2 of Proposition 3.5 mean that \( z_{k+1}^k = z_k + \Delta_1 z_k \) eventually holds and \( z_{k+1}^k \) is accommodated by \( \hat{N}_{\epsilon_k}^\mu \), which is larger than \( \hat{N}_{\tilde{\epsilon}_k}^\mu \) since \( \tilde{\epsilon}_k > \epsilon_k = \gamma_1 \mu_{k}^{1+c\alpha} \) for \( k \) sufficiently large by (3.31). On the other hand, items 1 and 2 of Proposition 3.6 indicate that \( z_{k+1}^k \) is necessarily accepted by the targeted neighborhood \( \hat{N}_{\epsilon_k}^\mu \) for all \( k \) sufficiently large and thus \( z_{k+1}^k \in \hat{N}_{\epsilon_k}^\mu \) holds. Hence, the condition in Step 2-3 is eventually satisfied with \( s_{k+1}^k = 1 \).

In what follows, we devote ourselves to prove the above two propositions. To begin with, we give some lemmas that help to show Proposition 3.5. The following lemma is concerned with the convergence speed of \( \mu_{k-1} \), \( \| \Phi_0(z_k^k) \| \), and \( \| z_k^k - z^* \| \).
Lemma 3.1 Suppose that Assumptions B and C hold. Then, we have $\|z^k - z^*\| = \|\tilde{z}^k - \tilde{z}^*\|$ for sufficiently large $k$ and $\mu_{k-1} = \Theta(||\Phi_0(\tilde{z}^k)||) = \Theta(||\tilde{z}^k - \tilde{z}^*||) = \Theta(||z^k - z^*||)$.

Proof See Appendix A.5.

Lemma 3.2 Suppose that Assumptions B and C hold. Then,

1. We have

$$\|\mathcal{J}\Phi_0(\tilde{z}^k) - Q(\tilde{z}^k)\|_F = \sqrt{\sum_{i=1}^{n} \frac{1}{2}(\mathcal{L}_{F(\alpha_i)}^\mathcal{L}_{V_k} - \mathcal{L}_{V_k}) F_i} = O(\mu_{k-1}^\alpha)$$

and hence $\lim_{k \to \infty} Q(\tilde{z}^k) = \mathcal{J}\Phi_0(\tilde{z}^*)$, and

2. $Q(\tilde{z}^k)$ is nonsingular for sufficiently large $k$ and $\{Q(\tilde{z}^k)^{-1}\}$ is bounded.

Proof Notice that $|F(x^k) \circ V_k - \mu_{k-1}I| \leq \varepsilon_{k-1} = \gamma_1 \mu_{k-1}^{1+\alpha}$ by $z^k \in \tilde{N}_{\mu_{k-1}}^{-1}$ and (3.12). In addition, note that $\lim_{k \to \infty} F(x^k) + V_k = F(x^*) + V_* \in S_+^m \times S_+^m$ and $\lim_{k \to \infty} F(x^k) + V_k = F(x^*) + V_* \in S_+^m$ by Assumptions B-2 and C-1. Then, Proposition 3.3 with $\{X_r\}, \{V_r\}, \{\mu_r\}$, and $\theta$ replaced by $\{F(x^k)\}, \{V_k\}, \{\mu_{k-1}\}$, and $\alpha$, respectively, yields

$$\|\mathcal{L}_{F(\alpha_i)}^\mathcal{L}_{V_k} - \mathcal{L}_{V_k}\|_2 = O(\mu_{k-1}^\alpha),$$

which further implies

$$\|\mathcal{J}\Phi_0(\tilde{z}^k) - Q(\tilde{z}^k)\|_F = \sqrt{\sum_{i=1}^{n} \frac{1}{2}(\mathcal{L}_{F(\alpha_i)}^\mathcal{L}_{V_k} - \mathcal{L}_{V_k}) F_i} = O(\mu_{k-1}^\alpha),$$

where the first equality is a direct consequence of the fact $\|svec(X)\|^2 = \|X\|^2_{F^2}$ ($X \in S^m$) and the forms of $\mathcal{J}\Phi_0(\tilde{z}^k)$ and $Q(\tilde{z}^k)$. To prove item 2, recall that $\mathcal{J}\Phi_0(\tilde{z}^*)$ is nonsingular from Assumption B-3. Then, since $\lim_{k \to \infty} Q(\tilde{z}^k) = \mathcal{J}\Phi_0(\tilde{z}^*)$ from item 1, $Q(\tilde{z}^k)$ is nonsingular for all $k$ sufficiently large. In addition, we obtain the boundedness of $\{Q(\tilde{z}^k)^{-1}\}$ because $\lim_{k \to \infty} Q(\tilde{z}^k)^{-1} = \mathcal{J}\Phi_0(\tilde{z}^*)^{-1}$ holds. The proof is complete. □

We are now ready to prove Proposition 3.5 using Lemmas 3.1 and 3.2.

Proof of Proposition 3.5: Recall that $\tilde{z}^k = (x^k, \tilde{x}^k, svec(V_k))$ with $\tilde{x}^k = (\tilde{x}^k_i)_{i \in I_a(x^*)}$ and $\Delta_1^k \tilde{z}^k = (\Delta_1^k x^k, \Delta_1^k \tilde{x}^k, svec(\frac{1}{2} V_k))$. For simplicity of expression, we identify $svec(V) \in \mathcal{N}_+^{m^{(m+1)/2}}$ with $V \in S^m$ and suppose $I_a(x^*) = \{1, 2, \ldots, p(x^*)\}$, which implies $\tilde{z}^* = z^*$ and, for sufficiently large $k$,

$$\tilde{z}^k = z^k, \Delta_1^k \tilde{z}^k = \Delta_1^k z^k.$$
It is not difficult to extend the subsequent analysis to the more general case of \( I_a(x^*) \subseteq \{1, 2, \ldots, p(x^*)\} \). Hereafter, we write

\[
Q_k := Q(z^k), \quad J_k := J \Phi_0(z^k)
\]

(3.33)

for each \( k \).

From item 2 of Lemma 3.2 together with (3.24), we readily see that there exists some \( L > 0 \) such that

\[
\|Q_k^{-1}\|_F \leq L \tag{3.34}
\]

and

\[
\Delta_{1/2} z^k = -Q_k^{-1} \Phi_{\mu_k}(z^k),
\]

(3.35)

\[
\|\Delta_{1/2} z^k\| \leq \|Q_k^{-1}\|_F \|\Phi_{\mu_k}(z^k)\| = O(\|\Phi_{\mu_k}(z^k)\|),
\]

(3.36)

where the equality in (3.36) follows from (3.34). Especially, combining the above with Lemma 3.1 implies that

\[
\max\left(\|\Delta_{1/2} x^k\|, \|\Delta_{1/2} V_k\|\right) \leq \|\Delta_{1/2} z^k\| = O(\mu_k^{-1}).
\]

(3.37)

Notice that \( \|\Phi_{\mu}(z^*)\| = \|\mu I\|_F = \mu \sqrt{m} \) for any \( \mu \geq 0 \) and

\[
\|\Phi_{\mu_k}(z^k)\| \leq \|\Phi_0(z^k)\| + \mu_k \sqrt{m}
\]

\[
= \|\Phi_0(z^k)\| + o(\mu_k^{-1})
\]

\[
= O(\|z^k - z^*\|)
\]

\[
= O(\mu_k^{-1}),
\]

(3.38)

where the first equality follows from (3.32) and the last two equalities are derived from Lemma 3.1. Additionally, by item 1 of Lemma 3.2, we have

\[
\|J_k - Q_k\|_F = \sqrt{\sum_{i=1}^{n} \frac{1}{2} \left( \mathcal{L}_{F(x^k)} \mathcal{L}_{V_k} \mathcal{L}_{F(x^k)}^{-1} - \mathcal{L}_{V_k} \right) F_i}^2_F = O(\mu_k^{-1}). \tag{3.39}
\]

1. It suffices to show that

\[
F(x^k + \Delta_{1/2} x^k) = F(x^k) + \Delta_{1/2} F_k \in S_{n,+}^n, \quad V_k + \Delta_{1/2} V_k \in S_{n,+}^n
\]

(3.40)

for any \( k \) sufficiently large, where \( \Delta_{1/2} F_k := \sum_{i=1}^{n} \Delta_{1/2} x_i^k F_i \). In fact, if these conditions hold, by (3.9) with \((X, \Delta X) = (F(x^k), \Delta_{1/2} F_k)\) and \((X, \Delta X) = (V_k, \Delta_{1/2} V_k)\),
we see that \( t_k^\frac{1}{2} = u_k^\frac{1}{2} = 1 \) and thus \( s_k^\frac{1}{2} = 1 \) from (3.7). From \( z_k \in \widetilde{\mathcal{N}}_{\mu_k-1} \) and
\( \varepsilon_{k-1} = \gamma_1 \mu_{k-1}^{1+\alpha} \), we have
\[
\| F(x^k) \circ V_k - \mu_{k-1} I \|_F = O(\mu_{k-1}^{1+\alpha}).
\] (3.41)

Denote
\[
\Gamma_k := \frac{1}{2} \sum_{i=1}^n \Delta_x^\frac{1}{2} x_i^k \left( \mathcal{L}_{F(x^k)} \mathcal{L}_{V_k} \mathcal{L}_{F(x^k)}^{-1} - \mathcal{L}_{V_k} \right) F_i.
\]

In view of (3.37) and (3.39), we have
\[
\| \Gamma_k \|_F \leq \| J_k - Q_k \|_F \| \Delta_x^\frac{1}{2} x^k \| = O(\mu_{k-1}^{1+\alpha}),
\] (3.42)
\[
\| \Delta_x^\frac{1}{2} F_k \circ \Delta_x^\frac{1}{2} V_k \|_F = O(\| \Delta_x^\frac{1}{2} z^k \|_2) = O(\mu_k^2).
\] (3.43)

By rearranging (3.16) with \( \tilde{z} = z^k \), \( \mu = \mu_k \), and \( P = I \) in terms of \( \Gamma_k \), we obtain
\[
F(x^k) \circ V_k + \Delta_x^\frac{1}{2} F_k \circ V_k + \Delta_x^\frac{1}{2} V_k \circ F(x^k) - \mu_k I = -\Gamma_k,
\]
which together with (3.42) and (3.32) implies
\[
\| F(x^k) \circ V_k + \Delta_x^\frac{1}{2} F_k \circ V_k + \Delta_x^\frac{1}{2} V_k \circ F(x^k) - \mu_k I \|_F = O(\mu_{k-1}^{1+\alpha})
\]
\[
= O(\mu_k^2 + (1-c_1)\alpha^{\frac{1}{1+c_1}}).
\] (3.44)

Since \( \alpha \in (0, 1) \) and \( c \in (0, 1) \), using (3.32) again, we obtain
\[
\mu_k^2 = o(\mu_k).
\] (3.45)

In Proposition 3.4, replace \( \{X_r\}, \{Y_r\}, \{\Delta X_r\}, \{\Delta Y_r\}, \{\mu_r, \tilde{\mu}_r\} \), and \( (\theta, \tilde{\theta}) \)
by \( \{F(x^k)\}, \{V_k\}, \{\Delta_x^\frac{1}{2} F_k\}, \{\Delta_x^\frac{1}{2} V_k\}, \{\mu_{k-1}, \mu_k\} \), and \( (\alpha, \frac{(1-c_1)\alpha^{\frac{1}{1+c_1}}}{1+c_1}) \), respectively. Then, the relations (3.41), (3.43), (3.44), and (3.45) correspond to conditions (3.25)–(3.28). We thus have (3.40) by Proposition 3.4.

2. We have only to show \( \tilde{\rho}_{\mu_k}(z^{k+\frac{1}{2}}) \leq \tilde{\varepsilon}_k \). We begin with noting that from item 1,
\( s_k^\frac{1}{2} = 1 \) for all \( k \) sufficiently large. Then, \( z^{k+\frac{1}{2}} = z^k + \Delta_x^\frac{1}{2} z^k \) and \( \| \Phi_{\mu_k}(z^{k+\frac{1}{2}}) \| \) is evaluated as follows:
\[
\| \Phi_{\mu_k}(z^{k+\frac{1}{2}}) \| \leq \| \Phi_{\mu_k}(z^k) + J_k \Delta_x^\frac{1}{2} z^k \| + O(\| \Delta_x^\frac{1}{2} z^k \|^2)
\]
\[
= \| \Phi_{\mu_k}(z^k) - J_k Q_k^{-1} \Phi_{\mu_k}(z^k) \| + O(\mu_{k-1}^2)
\]
\[
= \| (Q_k - J_k) Q_k^{-1} \Phi_{\mu_k}(z^k) \| + O(\mu_{k-1}^2)
\]
\[
\leq \| J_k - Q_k \|_F \| Q_k^{-1} \|_F \| \Phi_{\mu_k}(z^k) \| + O(\mu_{k-1}^2)
\]
where the first inequality follows from (3.33) and $J_k = J_{\Phi_k}(z^k)$, the first equality comes from (3.35) and (3.37), and the third equality is derived from (3.38), (3.39), and (3.34). From (3.46), we further obtain

$$\sum_{i=1}^{\mu} (\xi_i^* + \Delta_1 \xi_i^*) \tilde{g}_i(x^k + \Delta_1 x^k) \leq p(x^*) \|z^k + \Delta_1 z^k\| \|\Phi_{\mu_k}(z^k + \Delta_1 z^k)\|$$

$$= O(\|\Phi_{\mu_k}(z^k + \Delta_1 z^k)\|)$$

$$= O(\mu_{k-1}^{1+c\alpha}).$$

and

$$\max_{1 \leq i \leq p(x^*)} (\tilde{g}_i(x^k + \Delta_1 x^k))_+ = \max_{i \in I_0(x^*)} (\tilde{g}_i(x^k + \Delta_1 x^k))_+$$

$$= O(\|\Phi_{\mu_k}(z^k + \Delta_1 z^k)\|)$$

$$= O(\mu_{k-1}^{1+c\alpha}).$$

Combining these facts, we obtain $R_{\mu_k}(z^{k+1/2}) = O(\mu_{k-1}^{1+c\alpha})$. Since (3.30) and (3.29) together yield $\tilde{\epsilon}_k = \gamma_1 \gamma_2 \mu_{k-1}^{1+c\alpha}$ and $(1+c\alpha)(1+c\alpha) < 1+c\alpha$, we have $\mu_{k-1}^{1+c\alpha} = O(\mu_{k-1}^{1+c\alpha})$ and hence conclude $R_{\mu_k}(z^{k+1/2}) \leq \tilde{\epsilon}_k$ for any $k$ sufficiently large.

3. By $\mu_{k-1}^{1+c\alpha} = O(\mu_{k-1}^{1+c\alpha})$ shown above and (3.46), we obtain the desired result.

4. Recall that $\|Q_k^{-1}\|$ is bounded by (3.34) and $\mu_{k-1} = \Theta(\|z^k - z^*\|)$ from Lemma 3.1. It follows that

$$\|z^k + \Delta_2 z^k - z^*\|$$

$$= \|z^k - Q_k^{-1} \Phi_{\mu_k}(z^k) - z^*\|$$

$$\leq \|z^k - Q_k^{-1} \Phi_0(z^k) - z^*\| + \mu_k \|Q_k^{-1}\| F \sqrt{m}$$

$$\leq \|Q_k^{-1}\| F \|Q_k(z^k - z^*) - \Phi_0(z^k) + \Phi_0(z^*)\| + \gamma_2 \mu_{k-1}^{1+c\alpha} \|Q_k^{-1}\| F \sqrt{m}$$

$$= \|Q_k^{-1}\| F \|J_k(z^k - z^*) + (Q_k - J_k)(z^k - z^*) - \Phi_0(z^k) + \Phi_0(z^*)\|$$

$$+ O(\mu_{k-1}^{1+c\alpha})$$

$$\leq \|Q_k^{-1}\| F \|J_k(z^k - z^*) - \Phi_0(z^k) + \Phi_0(z^*)\|$$

$$+ \|Q_k^{-1}\| F \|J_k - Q_k\| F \|z^k - z^*\| + O(\|z^k - z^*\|^{1+c\alpha})$$

$$= O(\|z^k - z^*\|^2) + O(\|z^k - z^*\|^{1+c\alpha}) + O(\|z^k - z^*\|^{1+c\alpha})$$

$$= O(\|z^k - z^*\|^{1+c\alpha}),$$
where the first equality follows from (3.35), the first inequality comes from
\( \Phi_0(z^k) - \Phi_{\mu_k}(z^k) = (0, 0, \text{svec}(\bar{x} I)) \) (see (3.23)), the second inequality is derived
from \( \Phi_0(z^*) = 0 \) and (3.32), and the third equality is due to (3.39) and Lemma 3.1.
Thus, the desired conclusion is obtained. \( \square \)

We next enter the phase of proving Proposition 3.6. First, let us observe several
properties obtained from Proposition 3.5. Note that item 4 of Proposition 3.5 implies
\( \lim_{k \to \infty} z^k + \frac{1}{2} = z^* \), and thus
\( \lim_{k \to \infty} \tilde{z}^k + \frac{1}{2} = \tilde{z}^* \).

(3.47)

Noting \( z^k + \frac{1}{2} \in \mathcal{N}_{\mu_k} \) by Proposition 3.5, we can show that
\( \|J \Phi_0(\tilde{z}^k + \frac{1}{2}) - Q(\tilde{z}^k + \frac{1}{2})\|_F = O(\mu_k) \) in a manner similar to item 1 of Lemma 3.2. Thus, it holds
that \( \lim_{k \to \infty} Q(\tilde{z}^k + \frac{1}{2}) = J \Phi_0(\tilde{z}^*) \) from (3.47), which together with Assumption B-3
implies the nonsingularity of \( Q(\tilde{z}^k + \frac{1}{2}) \) for all \( k \) sufficiently large and the boundedness
of \( \{Q(\tilde{z}^k + \frac{1}{2})^{-1}\} \).

The above observations are summarized in the following lemma:

**Lemma 3.3** Suppose that Assumptions B and C hold. Then, we have

1. \( Q(\tilde{z}^k + \frac{1}{2}) \) is nonsingular for all \( k \) sufficiently large,
2. \( \|J \Phi_0(\tilde{z}^k + \frac{1}{2}) - Q(\tilde{z}^k + \frac{1}{2})\|_F = O(\mu_k) \), and
3. \( \{Q(\tilde{z}^k + \frac{1}{2})^{-1}\} \) is bounded.

Furthermore, in a manner similar to Lemma 3.1, we can derive the following result by
using \( z^k + \frac{1}{2} \in \mathcal{N}_{\mu_k} \):

**Lemma 3.4** Suppose that Assumptions B and C hold. Then, we have
\( \|z^k + \frac{1}{2} - z^*\| = \|z^k + \frac{1}{2} - \tilde{z}^*\| \) for sufficiently large \( k \) and \( \mu_k = \Theta(\|\Phi_0(z^k + \frac{1}{2})\|) = \Theta(\|z^k + \frac{1}{2} - z^*\|) = \Theta(\|z^k + \frac{1}{2} - \tilde{z}^*\|) \).

**Proof** The proof is obtained in a manner analogous to Lemma 3.1. \( \square \)

We are now ready to prove Proposition 3.6. Its proof seems quite similar to Propo-
sitions 3.5. However, we do not omit it since there are some significant differences.
For example, the proof of item 2 of Proposition 3.6 relies on the condition \( \tilde{c} > \frac{1}{2} \) in
(3.29).

**Proof of Proposition 3.6** Like the proof of Proposition 3.5, for simplicity of expres-
sion, we identify \( \text{svec}(V) \in \mathcal{R}^{m(m+1)/2} \) with \( V \in S^m \) and suppose \( I_a(x^*) = \{1, 2, \ldots, p(x^*)\} \), which implies \( \tilde{z}^* = z^* \) and, for sufficiently large \( k \),
\[ z^k + \frac{1}{2} = z^k + 1, \quad \Delta_1 z^k = \Delta_1 z^k. \]

We also write
\[ Q_{k + \frac{1}{2}} := Q(z^k + \frac{1}{2}), \quad J_{k + \frac{1}{2}} := J \Phi_{\mu_k}(z^k + \frac{1}{2}) = J \Phi_0(z^k + \frac{1}{2}). \]
By $z^{k+\frac{1}{2}} = z^{k+\frac{1}{2}}$ and item 3 of Proposition 3.5, we have, for any $k$ sufficiently large,

$$
\| \Phi_{\mu_k}(z^{k+\frac{1}{2}}) \| = O(\mu^{1+\tilde{c}\alpha}). \tag{3.48}
$$

By (3.24) with $j = \frac{1}{2}$ and item 1 of Lemma 3.3, we have

$$
\Delta_1 z^k = -Q_{k+\frac{1}{2}}^{-1}\Phi_{\mu_k}(z^{k+\frac{1}{2}}), \tag{3.49}
$$

which together with (3.48), item 3 of Lemma 3.3, and Lemma 3.4 implies

$$
\| \Delta_1 z^k \| = O(\| \Phi_{\mu_k}(z^{k+\frac{1}{2}}) \|) = O(\mu^{1+\tilde{c}\alpha}). \tag{3.50}
$$

Combining this with Lemma 3.4 yields

$$
\| \Delta_1 z^k \| = O(\| z^{k+\frac{1}{2}} - z^* \|^{1+\tilde{c}\alpha}). \tag{3.51}
$$

Note that, from Lemma 3.3, we have

$$
\| J_{k+\frac{1}{2}} - Q_{k+\frac{1}{2}} \| = O(\mu^{\tilde{c}\alpha}). \tag{3.52}
$$

1. As in the proof of item 1 of Proposition 3.5, it suffices to show that

$$
F(x^{k+\frac{1}{2}} + \Delta_1 x^k) = F(x^{k+\frac{1}{2}}) + \Delta_1 F_k \in S^m_+, \quad V_{k+\frac{1}{2}} + \Delta_1 V_k \in S^m_+ \tag{3.53}
$$

for all $k$ sufficiently large, where $\Delta_1 F_k := \sum_{i=1}^n \Delta_1 x^k_i F_i$. Since $z^{k+\frac{1}{2}} \in N_{\tilde{\mu}_k}$ by Proposition 3.5 and $\tilde{\epsilon}_k = \gamma_1 \mu^{1+\tilde{c}\alpha}$, we obtain

$$
\| F(x^{k+\frac{1}{2}}) \circ V_{k+\frac{1}{2}} - \mu_k I \|_F = O(\mu^{1+\tilde{c}\alpha}). \tag{3.54}
$$

On the other hand, (3.50) implies

$$
\| \Delta_1 F_k \circ \Delta_1 V_k \|_F = O(\| \Delta_1 z^k \|^2) = O(\mu^{2\tilde{c}\alpha}). \tag{3.55}
$$

Moreover, by (3.20) with $\hat{P} = I$, $\hat{z} = z^{k+\frac{1}{2}}$, and $\mu = \mu_k$, it holds that

$$
\| F(x^{k+\frac{1}{2}}) \circ V_{k+\frac{1}{2}} + \Delta_1 F_k \circ V_{k+\frac{1}{2}} + F(x^{k+\frac{1}{2}}) \circ \Delta_1 V_k - \mu_k I \|_F = 0. \tag{3.56}
$$

In Proposition 3.4, replace $\{ X_r \}, \{ Y_r \}, \{ \Delta X_r \}, \{ \Delta Y_r \}, \{ (\mu_r, \hat{\mu}_r) \}, \{ (\theta, \hat{\theta}) \}$ by $\{ F(x^{k+\frac{1}{2}}) \}, \{ V_{k+\frac{1}{2}} \}, \{ \Delta_1 F_k \}, \{ \Delta_1 V_k \}, \{ (\mu_k, \mu_k) \}$, and $(\tilde{c}\alpha, \alpha)$, respectively. Then, in view of (3.54), (3.55), and (3.56), we can verify conditions (3.25)–(3.28). We thus obtain (3.53) and conclude the desired result.
We only need to show \( \hat{R}_{\mu_k}(z_k^+) \leq \varepsilon_k \) for \( k \) sufficiently large. The remaining part is obvious from (3.53). We first note that

\[
\left\| \Phi_{\mu_k}(z^{k+\frac{1}{2}} + \Delta_1 z^k) \right\| \leq \left\| \Phi_{\mu_k}(z^{k+\frac{1}{2}}) + J_{k+\frac{1}{2}} \Delta_1 z^k \right\| + O(\|\Delta_1 z^k\|^2)
\]

\[
\leq \left\| \Phi_{\mu_k}(z^{k+\frac{1}{2}}) + Q_{k+\frac{1}{2}} \Delta_1 z^k \right\|
\quad + \|J_{k+\frac{1}{2}} - Q_{k+\frac{1}{2}}\| \|\Delta_1 z^k\| + O(\|\Delta_1 z^k\|^2)
\]

\[
= O\left( \mu_k \bar{\alpha} \|\Delta_1 z^k\| \right) + O(\|\Delta_1 z^k\|^2)
\]

\[
= O(\mu_k^{1+2\bar{\alpha}}),
\]

where the first equality follows from (3.24), (3.49), and (3.52) and the second equality is obtained from (3.50). Then, in a manner similar to Proposition 3.5, we can show

\[
\hat{R}_{\mu_k}(z_k^+) = O(\mu_k^{1+2\bar{\alpha}}),
\]

which together with \( \bar{\alpha} > \frac{1}{2} \) from (3.29) implies \( \hat{R}_{\mu_k}(z_k^+) \leq \varepsilon_k = \gamma_1 \mu_k^{1+\alpha} \) for all \( k \) sufficiently large.

3. We have

\[
\left\| z^{k+\frac{1}{2}} + \Delta_1 z^k - z^* \right\| \leq \left\| z^{k+\frac{1}{2}} - z^* \right\| + \|\Delta_1 z^k\|
\]

\[
= \left\| z^{k+\frac{1}{2}} - z^* \right\| + O(\left\| z^{k+\frac{1}{2}} - z^* \right\|^{1+\bar{\alpha}})
\]

\[
= O(\left\| z^{k+\frac{1}{2}} - z^* \right\|),
\]

where the first equality follows from (3.51). Combining Propositions 3.5 and 3.6, we get the following two-step superlinear convergence result.

**Theorem 3.1** Suppose that Assumptions B and C hold. Then, \( z^{k+\frac{1}{2}} = z^k + \Delta_1 z^k \) holds for any \( k \) sufficiently large and the update in Step 2-3 is eventually adopted. Moreover, we have

\[
\left\| z^{k+1} - z^* \right\| = O\left( \left\| z^k - z^* \right\|^{1+\alpha} \right). \quad (3.57)
\]

Hence, \( \{z^k\} \) converges to \( z^* \) two-step superlinearly with the order of \( 1 + \alpha \in (1, \frac{4}{3}) \).

**Proof** From item 1 of Proposition 3.5, \( z^{k+\frac{1}{2}} = z^k + \Delta_1 z^k \) holds for any \( k \) sufficiently large. From item 2 of Proposition 3.6, \( z^{k+1} = z^{k+\frac{1}{2}} + \Delta_1 z^k \) holds for any \( k \) sufficiently large, that is to say, the update in Step 2-3 is eventually accepted. Using Propositions 3.5 and 3.6 again yields

\[
\left\| z^{k+\frac{1}{2}} + \Delta_1 z^k - z^* \right\| = O(\left\| z^{k+\frac{1}{2}} - z^* \right\|) = O(\left\| z^k - z^* \right\|^{1+\alpha}).
\]
We thus confirm (3.57). Finally, since the parameter $c$ is chosen so that $0 < c \leq \frac{1}{\alpha + 2}$, we have

$$1 < 1 + c\alpha \leq 1 + \frac{\alpha}{\alpha + 2} = 2 - \frac{2}{\alpha + 2} < \frac{4}{3}.$$  

The proof is complete. \hfill \Box

### 3.3 Global version of the proposed algorithm

Algorithm 2 only works locally around a KKT point. Since the set of implicit functions defined around each generated point may in general vary, we have to be careful in updating Lagrange multipliers $\zeta_k$ and $\zeta_{k+1}$. Moreover, Algorithm 2 does not specify how $z_{k+1}^+ \in \hat{N}_{\mu_k}^\xi$ is computed in Step 2-4. In this section, we present a global form of the algorithm (Algorithm 3), where we modify definition (3.5) of $\hat{N}_{\mu_k}^\xi$ by replacing the implicit functions defined around $x^*$ with those around each generated point. As in Algorithm 2, we assume that each generated point is $\delta$-nondegenerate for a predetermined $\delta > 0$. Note that we actually have $\zeta_k \in \mathbb{R}^p(x_k)$ for each $k \geq 0$, while Algorithm 2 supposes $\zeta_k \in \mathbb{R}^p(x^*)$ for all $k$. Similarly, we have $\zeta_{k+1}^+ \in \mathbb{R}^p(x_{k+1}^+)$ and $\zeta_k^+ \in \mathbb{R}^p(x_k^+)$ for each $k \geq 0$. To clarify the crucial difference from Algorithm 2, the precise description of the procedures common to Algorithm 2 is omitted.

**Algorithm 3** (Global primal-dual path following method)

**Step 0** (Initial setting): Choose $\kappa > 0$. The other settings are the same as in Step 0 of Algorithm 2.

**Step 1** (Stopping rule): Same as in Step 1 of Algorithm 2.

**Step 2** (Computing an approximate BKKT point): Perform the following procedure:

**Step 2-1**: Compute $\Delta_{1/2} z^k := (\Delta_{1/2} x^k, \Delta_{1/2} \zeta_k, \Delta_{1/2} V_k)$ and $s_{1/2}^k \in (0, 1]$ as in Step 2-1 of Algorithm 2. Define $z_{k+1/2}^+ := (x_{k+1}^+, \zeta_{k+1}^+, V_{k+1}^+)$ by setting $(x_{k+1}^+, V_{k+1}^+) := \left( x^k + s_k^k \Delta_{1/2} x^k, V_k + s_k^k \Delta_{1/2} V_k \right)$ and, for $i = 1, 2, \ldots, p(x_{k+1}^+)$,

$$\zeta_i^+ := \begin{cases} \zeta_i^{+j} + \Delta_{1/2} \zeta_j^k & \text{if there exists } j \text{ s.t. } \tau_i^j(x_{k+1}^+) = \tau_i^{+j}(x_{k+1}^+) \\ 0 & \text{otherwise.} \end{cases}$$  

(3.58)

**Step 2-2**: Compute $\Delta_1 z^k := (\Delta_1 x^k, \Delta_1 \zeta_k, \Delta_1 V_k)$ and $s_1^k \in (0, 1]$ as in Step 2-2 of Algorithm 2. Define $z_+^k := (x_+, \zeta_+^k, V_+^k)$ by setting $(x_+, V_+^k) := $  

Springer
\[
(x^{k+\frac{1}{2}} + s^{k}_{1} \Delta_1 x^{k}, \ V_{k+\frac{1}{2}}^{1} + s^{k}_{1} \Delta_1 V_{k}) \text{ and, for } i = 1, 2, \ldots, \ p(x^{k}),
\]

\[
(\xi^{k+\frac{1}{2}}_{i}) := \begin{cases} 
\xi^{k+\frac{1}{2}}_{j} + \Delta_1 \xi^{k}_{j} & \text{if there exists } j \text{ such that } \tau_{x^{k+\frac{1}{2}}}^{j}(x^{k}) = \tau_{x^{k+\frac{1}{2}}}^{j}(x^{k}) \\
0 & \text{otherwise.}
\end{cases}
\tag{3.59}
\]

**Step 2-3:** Same as in Step 2-3 of Algorithm 2.

**Step 2-4:** Find \( w^{k+1} = (x^{k+1}, y^{k+1}, V^{k+1}) \in N_{\mu_{k}}^{\epsilon} \) with discrete measure \( y^{k+1} \in \mathcal{M}_{(+)}(T) \) by using, e.g., the interior-point SQP method [20]. If \( w^{k+1} \) is a KKT point of SISDP (1.1), then terminate the algorithm. If it holds that

\[
\text{supp}(y^{k+1}) \subseteq \{ \tau_{x^{k+1}}^{i}(x^{k+1}) \}_{i=1}^{p(x^{k+1})},
\tag{3.60}
\]

set \( z^{k+1} := (x^{k+1}, \xi^{k+1}, V^{k+1}) \) with

\[
\xi^{k+1}_{i} := y^{k+1}(\{ \tau_{x^{k+1}}^{i}(x^{k+1}) \}) \ (i = 1, 2, \ldots, p(x^{k+1})).
\]

Otherwise, set \( z^{k+1} := (x^{k+1}, \xi^{k+1}, V^{k+1}) \) with

\[
\xi^{k+1}_{i} := \sum_{\tau \in \text{supp}(y^{k}) : \| \tau - \tau_{x^{k+1}}^{i}(x^{k+1}) \| \leq \kappa} y^{k+1}(\{ \tau \})
\tag{3.61}
\]

for \( i = 1, 2, \ldots, p(x^{k+1}) \).

**Step 3:** Same as in Step 3 of Algorithm 2.

**Remark 3.5** (About Steps 2-1 and 2-2) We explain update (3.58). To determine \( \zeta^{k+\frac{1}{2}}_{i} \) properly, we need to examine which constraint of \( g(x, \tau_{x^{k+\frac{1}{2}}}^{i}(x)) \leq 0 \ (i = 1, 2, \ldots, p(x^{k+\frac{1}{2}})) \) corresponds to each component of \( \zeta^{k} + s^{k}_{1} \Delta_1 \xi^{k} \). Taking into consideration that for each \( j = 1, 2, \ldots, p(x^{k}), \) \( (\zeta^{k} + s^{k}_{1} \Delta_1 \xi^{k})_{j} \) is a Lagrange multiplier estimate of \( g(x, \tau_{x^{k}}^{j}(x)) \leq 0 \) at \( x^{k+\frac{1}{2}} \), we may set \( (\zeta^{k+\frac{1}{2}})_{i} := (\zeta^{k} + s^{k}_{1} \Delta_1 \xi^{k})_{j} \) with \( j \) such that \( \tau_{x^{k}}^{j}(x^{k+\frac{1}{2}}) = \tau_{x^{k+\frac{1}{2}}}^{i}(x^{k+\frac{1}{2}}) \).\footnote{If \( x^{k} \) and \( x^{k+\frac{1}{2}} \) are sufficiently close to \( \delta \)-nondegenerate \( x^{\ast} \) as supposed in Algorithm 2, such \( j \) exists for each \( i \).}

\( \zeta^{k} + s^{k}_{1} \Delta_1 \xi^{k} \)

\( \tau_{x^{k}}^{j}(x^{k+\frac{1}{2}}) = \tau_{x^{k+\frac{1}{2}}}^{i}(x^{k+\frac{1}{2}}) \)

\( \delta \)-nondegenerate

\( x^{\ast} \)
\[ \| \tau_{x_{k+\frac{1}{2}}}^{i} (x_{k}) - \tau_{x_{k}}^{i} (x_{k}) - \delta_{k}^{i} \nabla_{x_{k}}^{i} (x_{k})^{\top} \Delta_{1} x_{k} \| \left( \approx \| \tau_{x_{k+\frac{1}{2}}}^{i} (x_{k}) - \tau_{x_{k}}^{i} (x_{k}) \| \right) \]

is smaller than a prescribed threshold value. If it is found, we regard \( \tilde{j} \) as the desired \( j \). If not, we set \( (\zeta_{k+\frac{1}{2}})^{j} := 0 \). Update (3.59) is executed in a similar manner.

**Remark 3.6** (About Step 2-4) If condition (3.60) holds, we can derive \( z_{k+1}^{k+1} \in \hat{N}_{\mu_{k}}^{\varepsilon_{k}} \). On the other hand, even if (3.60) does not hold, \( z_{k+1}^{k+1} \in \hat{N}_{\mu_{k}}^{\varepsilon_{k}} \) may hold for the following reason: As described in Remark 3.2, for any point in \( \text{supp}(y_{k}^{*}) \cap \{ \tau_{x_{k}}^{i} (x_{k})^{p(x_{k})} \}_{i=1}^{p} \), say \( \bar{\tau} \), there exists a sequence consisting of elements in \( \text{supp}(y_{k}^{*}) \) such that it converges to \( \bar{\tau} \). From this fact, by setting \( z_{k}^{k} \) as in (3.61) with sufficiently small \( \kappa \), we have \( \hat{R}_{\mu_{k}}(z_{k}^{k+1}) \approx R_{\mu_{k}}(w_{k}^{k+1}) \) for \( k \) large enough and thus expect the condition \( w_{k+1}^{k+1} \in N_{\mu_{k}}^{\varepsilon_{k}} \) to induce \( z_{k+1}^{k+1} \in \hat{N}_{\mu_{k}}^{\varepsilon_{k}} \).

**Remark 3.7** (About convergence property of Algorithm 3) Step 2-4 plays a role of safeguarding the global weak ∗ convergence of \( \{w_{k}\} \). Indeed, like Algorithm 2, Algorithm 1 underlies Algorithm 3. Hence, Theorem 2.2 ensures global convergence of Algorithm 3 to a KKT point of the SISDP.

4 Numerical experiments

In this section, we conduct some numerical experiments to demonstrate the efficiency of the primal-dual path following method (Algorithm 3) by solving three kinds of SISDPs with a one-dimensional index set of the form \( T = [T_{\min}, T_{\max}] \subseteq \mathcal{R} \) with \( T_{\min}, T_{\max} \in \mathcal{R} \): The first one is a linear SISDP where all functions are affine with respect to \( x \); the second one is an SISDP with a nonlinear objective function. The third one is a nonlinear SISDP arising from real applications, particularly, FIR design problems. Throughout the section, we identify a symmetric matrix variable \( X \in S^{m} \) with a vector variable \( x := (x_{11}, x_{12}, \ldots, x_{1m}, x_{22}, x_{23}, \ldots, x_{mm})^{\top} \in \mathcal{R}^{m(m+1)/2} \) through

\[
X = \begin{pmatrix}
  x_{11} & x_{12} & \ldots & x_{1m} \\
  x_{12} & x_{22} & \ldots & x_{2m} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{1m} & x_{2m} & \ldots & x_{mm}
\end{pmatrix}
\]

The program is coded in MATLAB R2019a and run on a machine with Intel(R) Core(TM) i7-8565U CPU@1.80GHz and 16.0 GB RAM. We compute the scaling matrices for the NT direction according to [32, Sect. 4.1]. As for SISDPs with a
nonlinear objective function, the matrix $B_P$ in the quadratic program (3.17) is not necessarily positive-definite. So as to assure its positive definiteness, we modify $B_P$ by lifting its negative eigenvalues to 1. Let $\bar{x}$ be a current point and $\{\tau^i_{\bar{x}}(\cdot)\}_{i=1}^{p(\bar{x})}$ be the set of implicit functions defined in (3.1). As for the set $S_\delta(\bar{x})$ defined by (3.2), we set $\delta := 1$ and put $N + 1$ grids $\{s_1, s_2, \ldots, s_{N+1}\}$ on $T$ uniformly with $N := 100$. To specify the set $S_\delta(\bar{x})$, we apply Newton’s method combined with the projection onto $T$ for the problem $\max_{\tau \in T} g(\bar{x}, \tau)$ starting from each of the local maximizers $\bar{s}$ of $\max\{ g(\bar{x}, s) \mid s = s_1, s_2, \ldots, s_{N+1}\}$ such that $g(\bar{x}, \bar{s}) > \max_{1 \leq i \leq N+1} g(\bar{x}, s_i) - \delta$.

Next, we explain how each step of Algorithm 3 is implemented. In what follows, for the sake of clarity, we denote each step of Algorithm 3 by A3-Step. For example, Step 2-4 of Algorithm 3 is expressed as A3-Step 2-4. In A3-Step 0, we set

$$\gamma_1 = \sqrt{\frac{m(m+1)}{2}}, \quad \gamma_2 = 5, \quad c = \frac{1}{2.99}, \quad \alpha = 0.99, \quad \beta = 0.8, \quad \delta = 1, \quad \sigma = 0.95, \quad \kappa = 0.01.$$ 

In A3-Step 1, we terminate the algorithm if $\mu_{k+1} < 10^{-10}$ or $\tilde{R}_0(\varepsilon^k) \leq 10^{-8}$. Similarly, in A3-Step 2-4, we stop the algorithm if $R_0(w^{k+1}) \leq 10^{-8}$. In the same step, we implement the interior-point SQP-type method proposed in [20] by using the implementation details described therein. In A3-Step 3, for the sake of numerical stability, we set $\varepsilon_{k+1} := \max(10^{-7}, \gamma_1 \mu_{k+1}^{1+\alpha})$. For $X \in S^m_{+}$ and $Y \in S^m$, we compute $L^1_X Z$ by solving the linear equation $L^1_X Z = Y$ for $Z \in S^m$ with the Matlab built-in solver $\text{lyap2}$. We moreover use $\text{quadprog}$ to solve quadratic programs in A3-Step 2-1. Moreover, in A3-Steps 2-1 and 2-2, let $10^{-1}$ be the threshold value for the left-hand side of (3.62) in Remark 3.5, which is used to check the correspondence between implicit functions defined on two distinct points.

For the sake of comparison, we also implement a discretization method that solves finitely relaxed SISDPs sequentially until an approximate feasible solution is obtained. More precisely, for solving SISDP (1.1), we use the following discretization algorithm:

**Discretization method for SISDP**

**Step 0:** Choose an initial finite set $T_0 \subset T$. Choose $\theta > 0$. Set $r := 0$.

**Step 1:** Get a KKT point $x^r$ of the finitely relaxed SISDP with $T$ replaced by $T_r$.

**Step 2:** Find $\bar{\tau} \in T$ such that $g(x^r, \bar{\tau}) > \theta$ and set $T_{r+1} := T_r \cup \{\bar{\tau}\}$. If such a point does not exist in $T$, terminate the algorithm.

**Step 3:** Increment $r$ by one and return to Step 1.

In Step 0 of the discretization method, we choose $T_0 = \{T_{\min}, T_{\max}\} \subset \mathcal{R}$. In Step 2, to find such a $\bar{\tau} \in T$ we solve $\max_{\tau \in T} g(x^r, \tau)$ by applying Newton’s method with a starting point $s \in \text{argmax}\{ g(x^r, s) \mid s = s_1, s_2, \ldots, s_{N+1}\}$, where $\{s_1, s_2, \ldots, s_{N+1}\}$ is the set of grids defined earlier in this section.\footnote{There is no theoretical guarantee for global optimality of $\tau$ thus found. In practice, however, we may expect to have a global optimum by setting $N$ large enough.} We set $\theta := 10^{-6}$.
4.1 Linear SISDPs

In this section, we consider the linear SISDP (1.1), called LSISDP for short. Specifically, we solve the following problem taken from [18, Sect. 4.2]:

Maximize \( A_0 \bullet X \)
subject to
\[
A(\tau) \bullet X \geq 0 \quad (\tau \in T) \\
I \bullet X = 1 \\
X \in S^m_+, \tag{4.1}
\]

where \( A_0 \in S^m \) and \( A : T \to S^m \) is a symmetric matrix valued function whose elements are \( q \)-th order polynomials in \( \tau \), i.e., \((A(\tau))_{i,j} = \sum_{l=0}^{q} a_{i,j,l} \tau^l \) for \( 1 \leq i, j \leq m \).

In this experiment, we deal with the cases where \( q = 9 \), \( m = 10, 20 \), and \( T = [0, 1] \), i.e., \( T_{\min} = 0 \) and \( T_{\max} = 1 \). We generate 10 test problems for each of \( m = 10, 20 \) as follows: We choose all entries of \( A_0 \) and the coefficients \( a_{i,j,l} \) in \( A(\tau) \) from the interval \([-1, 1]\) randomly. Among those generated data sets, we use only data such that the semi-infinite constraints include at least one active constraint at an optimum of (4.1). Specifically, for each generated data, we compute an optimum, say \( \tilde{X} \), of the SDP obtained by removing the semi-infinite constraints. If \( \min_{1 \leq i \leq 21} A \left( T_{\min} + (i-1)(T_{\max} - T_{\min}) \right) \bullet \tilde{X} \leq -10^{-3} \), which implies that \( \tilde{X} \) does not satisfy the semi-infinite constraints, we adopt it as a valid data set.

We examine the performance of Algorithm 3 by comparing it with the discretization method that uses SDPT3 [33] with the default setting to solve linear SDPs sequentially. In A3-Step 0, we choose \( x^0 \) so that \( X^0 = m^{-1}I \) and set \( \xi^0 = (1, 1, \ldots, 1)^\top \in \mathcal{R}^{p(x^0)} \), \( V_0 = mI \), and \( \mu_0 = 1 \). An initial Lagrange multiplier for the linear equality constraint is set to be zero. The obtained results are shown in Tables 1, 2, in which “ave.time(s)” and “\( \tilde{R}_0^\infty \)” stand for the average running time in seconds and the average value of \( \tilde{R}_0 \) at the solution output by the algorithm, respectively, and “Disc.” stands for the discretization method. Moreover, “AHO-like”, “NT”, and “H.K.M” stand for Algorithm 3 combined with the scaling matrices \( P = I \), \( F(x^k)^{-\frac{1}{2}} \), and \( W^{-\frac{1}{2}} \), respectively.

From the tables, we observe that the computational time for “AHO-like” is largest among all. Actually, it spends around 1.1 seconds for \( m = 10 \) and 24 seconds for \( m = 20 \), while the others spend about 1 second in all cases. This is mainly due to high computational costs for calculating the matrix \( H_P \) defined by (3.18), in which \( L_{F(x)}^{-1} \) must be dealt with. However, in the cases of “NT” and “H.K.M”, \( H_P \) can be

| Table 1 Results for the linear SISDP with \( m = 10 \) |
|-----------------|------------------|
|                | Ave.time(s)      | \( \tilde{R}_0^\infty \) |
| AHO-like       | 1.01             | 1.39e-09         |
| H.K.M.         | 0.31             | 1.39e-09         |
| NT             | 0.28             | 1.39e-09         |
| Disc.          | 0.19             | 7.25e-06         |
handled more efficiently. Second, we observe that “Disc.” solves problems faster than Algorithm 3. This is because an SDP is solved very quickly with SDPT3 at each iteration of “Disc.”, and the number of SDPs solved is very small. In fact, only three or four SDPs are solved on average per run. However, we can see that our methods gain KKT points with higher accuracy than the discretization method. More specifically, the values of $\hat{R}_0^*$ for Algorithm 3 lie between $1.0 \times 10^{-9}$ and $1.5 \times 10^{-9}$, while those for the discretization method are around $10^{-6}$. We also observe that Algorithm 3 skips A3-Step 2-4 in most iterations, namely, $z_{k+1}$ is determined by the directions $\Delta_1 z_{k+1}$ and $\Delta_1 z_{k+1}$. Actually, A3-Step 2-4 is skipped in more than 90% of iterations. Skipping A3-Step 2-4 is desirable since the interior point SQP method employed in A3-Step 2-4 is likely to solve multiple QPs and results in more computational cost than A3-Steps 2-1 and 2-2. Also, in most cases, the full step is accepted eventually and the value of $\hat{R}_{\mu-1}$ converges to 0 superlinearly.

We now turn our attention to the boundedness of $|\text{supp}(y^k)|$ assumed in Theorem 2.3. To this end, we monitor $|\text{supp}(y^k)|$ in Step 2.4 together with Step 2.3, where $y^k$ is computed from $z^k$ through the relation (3.6). From Table 3 showing the maximum of $|\text{supp}(y^k)|$ throughout all runs of the algorithm, we observe that it is bounded for these instances. Second, we investigate C-2 and C-3. For all the instances, we observed that the computed optimum, written $x^*$, is $\delta$-nondegenerate. For example, in one instance, $S_0(x^*)$ for $\max_{\tau \in T} g(x^*, \tau)$ comprised $\tau = 1$ and $\tau = 0.4563205$, of which the latter one is the global maximizer. Moreover, we confirmed that these local maximizers satisfied LICQ and strict complementarity, implying $\delta$-nondegeneracy of $x^*$. Since the sequence $\{x^k\}$ and $\{x_{k+1/2}\}$ converged to $x^*$, we confirmed that the implicit functions at $x^k$ and $x_{k+1/2}$ were identical to those at $x^*$ in the last several iterations, that is, Assumption C-2 held. Next, we check C-3 using the same problem-instance. While the active set at $x^*$, that is, the index set satisfying $g(x^*, \tau) = 0$ was $\{0.4563205\}$, the index set corresponding to $J_0^k$, that is, the set of $\tau \in S_0(x^k)$ such that $g(x^k, \tau) + \nabla g(x^k, \tau)^T \Delta_1 x^k = 0$ consisted of a single index within $10^{-6}$ from 0.4563205 in the last 3 iterations. From this, we can see that the implicit functions at $x^*$ were already captured at $x_{k+1/2}$ prior to convergence.

### Table 2

Results for the linear SISDP with $m = 20$

| Method   | Ave.time(s) | $\hat{R}_0^*$ |
|----------|-------------|---------------|
| AHO-like | 24.53       | 1.97e-09      |
| H.K.M.   | 1.01        | 1.97e-09      |
| NT       | 0.98        | 1.97e-09      |
| Disc.    | 0.47        | 4.83e-06      |

### Table 3

The maximum of $|\text{supp}(y^k)|$ throughout all runs for each $m$

| $m$   | AHO-like | H.K.M. | NT |
|-------|----------|--------|----|
| 10    | 3        | 3      | 3  |
| 20    | 3        | 3      | 3  |
4.2 Nonlinear SISDPs

Next, we solve the following SISDP whose objective function is nonlinear:

$$\begin{align*}
\text{Minimize} & \quad \frac{1}{2}x^\top B x + c^\top x + a\|x\|^4 \\
\text{subject to} & \quad \sum_{i=1}^{n} \tau^{i-1} x_i \leq \sum_{i=1}^{n} \tau^{2i} + \sin(9\pi \tau) + 2 \quad (\tau \in T) 
\end{align*}$$

(4.2)

with $a > 0$, $b > 0$, and $n := m(m + 1)/2$. We deal with the cases of $m = 10, 20$. For each of $m = 10, 20$, all the elements of $B \in S^m$ and $c \in \mathbb{R}^n$ are randomly generated from the interval $[-1, 1]$. We set $T = [0, 1]$ and $a = b = 0.01$. Problem (4.2) is always feasible since $x = 0$ satisfies the constraints. The objective function is not convex in general but coercive in the sense that $f(x) \to \infty$ as $\|x\| \to \infty$, and thus the considered problem is guaranteed to have at least one global optimum. We use the primal-dual interior point method [43] to solve finitely relaxed SISDPs. In A3-Step 0, $x^0 = 0$ is chosen so that $X_0 = 0$, and thus $X_0 + bI = bI \in S^m_{++}$. We set $y^0$, $V_0$, and $\mu_0$ as in the previous experiment for linear SISDPs.

We show the results in Tables 4, 5, where each column and row have the same meaning as in Tables 1, 2. From the tables, “AHO-like” spends the largest CPU-time like in linear SISDPs. We observe that Algorithm 3 (AHO-like, NT, H.K.M.) successfully obtains KKT points with higher accuracy than the discretization method. Actually, the values of $\hat{R}_0^*$ obtained by Algorithm 3 lie between $10^{-9}$ and $2 \times 10^{-9}$, while those for the discretization method are around $10^{-7}$. As in the case of the results for linear SISDPs, the discretization method is likely to spend less time than Algorithm 3. However, as the problem-size grows, solving a single NSDP in the discretization method becomes more costly, and as a result the whole cost of the discretization method may be higher than Algorithm 3. To confirm this prediction, we conduct additional experiments with $m = 40$ and show the obtained comparison results of the discretization method and Algorithm 3 using NT and HKM directions.

| Table 4 Results for the nonlinear SISDP with $m = 10$ |
|------------------------------------------------------|
| Ave.time(s)   | $\hat{R}_0^*$ |
|---------------|----------------|
| AHO-like      | 2.30           | 1.39e-09      |
| H.K.M.        | 0.82           | 1.39e-09      |
| NT            | 0.81           | 1.39e-09      |
| Disc.         | 0.14           | 3.76e-07      |

| Table 5 Results for the nonlinear SISDP with $m = 20$ |
|------------------------------------------------------|
| Ave.time(s)   | $\hat{R}_0^*$ |
|---------------|----------------|
| AHO-like      | 18.00          | 1.97e-09      |
| H.K.M.        | 2.23           | 1.97e-09      |
| NT            | 2.69           | 1.97e-09      |
| Disc.         | 1.11           | 4.95e-07      |
Table 6  Results for the nonlinear SISDP with \( m = 40 \)

|          | Ave.time(s) | \( \tilde{R}_0^n \) |
|----------|-------------|---------------------|
| H.K.M    | 66.56       | 2.78e-09            |
| NT       | 73.03       | 2.78e-09            |
| Disc.    | 94.19       | 6.65e-07            |

Table 7  The maximum of \( |\text{supp}(y^k)| \) throughout all runs for each \( m \)

| \( m \) | AHO-like | H.K.M. | NT |
|---------|----------|--------|----|
| 10      | 5        | 5      | 5  |
| 20      | 9        | 6      | 7  |
| 40      | —        | 25     | 16 |

in Table 6. We observe that Algorithms 3 finds more accurate solutions within much less than the time spent by the discretization method.

We now turn our attention to the boundedness of \( |\text{supp}(y^k)| \) assumed in Theorem 2.3. From Table 7 showing the maximum number of \( |\text{supp}(y^k)| \) throughout all runs of the algorithm, we observe that \( |\text{supp}(y^k)| \) is bounded for these instances.

Second, we considered C-2 and C-3. For all the instances, we observed that the computed optimum, written \( x^\ast \), is \( \delta \)-nondegenerate. For example, in one instance, we had \( S_\delta(x^\ast) = \{0, 0.16589683, 0.38669535, 0.60602213, 1\} \) for \( \max_{\tau \in T} g(x^\ast, \tau) \), among which 0.16589683 and 1 were the global maximizers. In a way similar to the experiments for SISDP (4.1), we confirmed that C-2 held. Moreover, C-3 was also confirmed by observing the indices corresponding to \( J_k^a \) in the last several iterations were within \( 10^{-6} \) from the active indices at \( x^\ast \), that is, \( \tau = 0.16589683 \) and \( \tau = 1 \).

4.3 Application from FIR design problems

In this section, we solve SISDP (1.1) that arises in FIR filter design with upper and lower bounds on its frequency response magnitude. Given the filter length \( n \), the frequency range \( T \subseteq [0, \pi] \), and the lower and upper frequency response bounds functions \( \text{Low}, \text{Upp} : T \rightarrow \mathbb{R}_+ \), we consider to find filter tap coefficients \( x := (x_1, \ldots, x_n)^\top \in \mathbb{R}^n \) such that the frequency response \( X(x, \tau) := x_1 + \sum_{i=2}^n x_i e^{-\sqrt{-1}(i-1)\tau} \), where \( e \) denotes Napier’s constant, satisfies the magnitude bounds: \( \text{Low}(\tau) \leq |X(x, \tau)| \leq \text{Upp}(\tau) \ (\tau \in T) \). This problem is formulated as

Find \( x \in \mathbb{R}^n \) such that \( \text{Low}(\tau) \leq |X(x, \tau)| \leq \text{Upp}(\tau) \ (\tau \in T) \).

Associated with this problem, Wu et al. [38] considered the following problem of finding a vector \( r := (r_1, r_2, \ldots, r_n)^\top \in \mathbb{R}^n \) together with a symmetric matrix \( Y \in \mathbb{S}^{n-1} \):
Find 
\[ (r, \mathcal{Y}) \in \mathbb{R}^n \times S^{n-1} \]
such that 
\[ \text{Low}^2(\tau) \leq \mathcal{F}(r, \tau) \leq \text{Upp}^2(\tau) \quad (\tau \in T) \]
\[ Z(\mathcal{Y}, r) := \begin{pmatrix} \mathcal{Y} - A^\top \mathcal{Y} A & \bar{r} - A^\top \mathcal{Y} b \\ \bar{r}^\top - b^\top \mathcal{Y} A & r_1 - b^\top \mathcal{Y} b \end{pmatrix} \in S^n_+, \]
\[ (4.3) \]

where
\[ \mathcal{F}(r, \tau) := r_1 + 2 \sum_{i=2}^n r_k \cos \tau (i - 1), \]

which is the Fourier transform of \( r \), and
\[ A := \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{(n-1) \times (n-1)}, \quad b := \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{n-1}, \quad \bar{r} := \begin{pmatrix} r_2 \\ r_3 \\ \vdots \\ r_n \end{pmatrix} \in \mathbb{R}^{n-1}. \]

Notice that problem (4.3) is cast as a linear SISDP by setting an arbitrary linear objective function. Wu et al. [38] solved this problem via discretization.

The components of \( r \) are called autocorrelation coefficients relevant to the filter of our interest, and are related to the filter tap vector \( x \) via the following relation:
\[ r_j = \sum_{i=1}^{n-j-1} x_i x_{i+j-1} \quad (j = 1, 2, \ldots, n), \]

which are simply expressed as \( r_j = T_j \cdot xx^\top \quad (j = 1, \ldots, n) \), where \( T_j \in \mathbb{R}^{n \times n} \) is the matrix whose \((i, i + j - 1)\)-element is 1 for \( i = 1, 2, \ldots, n - j + 1 \) and all the others are zeros. By replacing \( xx^\top \) with \( \mathcal{X} \in S^n_+ \), the above conditions are further transformed as
\[ r_j = T_j \cdot \mathcal{X} \quad (j = 1, \ldots, n), \quad \mathcal{X} \in S^n_+, \quad \text{rank}(\mathcal{X}) = 1. \quad (4.4) \]

In this experiment, we compute \( r \) and \( x \) simultaneously. This purpose is achieved by solving problem (4.3) with (4.4) as additional constraints. Since this problem is still hard owing to the rank 1 constraint, we remove it as in the SDP-relaxation technique to have the following problem:
\[ \text{Find} \quad \mathcal{X}, \mathcal{Y}, r \in S^n \times S^{n-1} \times \mathbb{R}^n \]
such that \( \text{Low}^2(\tau) \leq \mathcal{F}(r, \tau) \leq \text{Upp}^2(\tau) \quad (\tau \in T) \]
\[ Z(\mathcal{Y}, r) \in S^n_+, \quad \mathcal{X} \in S^n_+ \]
\[ r_j = T_j \cdot \mathcal{X} \quad (j = 1, \ldots, n). \quad (4.5) \]
Though this problem seems tractable, we do not necessarily gain a rank 1 solution $X$ by merely solving (4.5). One promising way for treating this matter is to minimize a function which can induce low-rankness of $X$. One choice for such a function is

$$f_\varepsilon(X) := \log \det (X + \varepsilon I)$$

with $\varepsilon > 0$. This function was utilized by Fazel et al. [5] in a heuristic method for rank minimization problems.

Now, the SISDP that we actually solve is of the following form:

$$\begin{aligned}
\text{Minimize} \quad & f_\varepsilon(X) + \operatorname{Tr}(X) \\
\text{subject to} \quad & \text{the constraints of (4.5)}.
\end{aligned}$$

(4.6)

Since $\operatorname{Tr}(X) = \sum_{i=1}^{m} \lambda_i(X) = \sum_{i=1}^{m} |\lambda_i(X)|$ on $X \in S^m_+$ and thus it can be regarded as an $\ell_1$-sparse regularizer, the second term $\operatorname{Tr}(X)$ in the objective function is expected to help us to obtain a rank 1 solution. In this experiment, we set $\varepsilon = 1$, $T = [0, 1]$, $\operatorname{Low}(\tau) := |\cos(\cos(\tau))|$, $\operatorname{Upp}(\tau) := 1.1 |\cos(\cos(\tau))|$.

In A3-Step 0, a starting point $(X_0, Y_0, r^0) \in S^n \times S^{n-1} \times \mathcal{R}^n$ is set as

$$X_0 := I_n, \quad Y_0 := \text{diag}(n-1, n-2, \ldots, 1), \quad (r^0)_j := T_j \cdot X_0 \quad (1 \leq j \leq n).$$

We set $\xi^0 := (1, 1, \ldots, 1)^\top$. Two initial Lagrange multiplier matrices corresponding to $Z(Y, r) \in S^n_+$ and $X \in S^m_+$ are set as $Z(Y_0, r^0)^{-1}, X_0^{-1}$, respectively. An initial Lagrange multiplier vector for the linear equality constraints to be a zero vector. The conditions in the other steps of the algorithm are the same as in the previous experiments.

We apply Algorithm 3 using H.K.M directions to this problem with $n = 5, 15, 25$. The obtained results are shown in Table 8, in which, for the case of $n = 15, 25$, we show the cpu-time spent until $\hat{R}_0^* \leq 10^{-6}$ instead of $\hat{R}_0^* \leq 10^{-8}$ is attained. Indeed, Algorithm 3 fails to find a solution with $\hat{R}_0^* \leq 10^{-8}$ owing to numerical instability in the last stage of iterations. When $n = 25$, the cpu-time grows explosively. This is mainly because Algorithm 3 fails to compute $z_{k+1}^*$ successfully in Step 2-3, and almost all of the cpu-time is occupied by computing $w_{k+1}^* \in \mathcal{N}_{\mu_k}^{\rho_k}$ in Step 2-4. It is a strength of the proposed algorithm that a KKT point can be gained even in such a undesirable situation by virtue of Step 2-4 as a safeguard system.

| $n$  | Time(s) | $\hat{R}_0^*$ |
|------|---------|---------------|
| 5    | 2.62    | 6.98e-09      |
| 15   | 44.81   | 9.87e-07      |
| 25   | 1760.53 | 9.70e-07      |
It is worth mentioning that the obtained solutions for \( n = 5, 15, 25 \) satisfy \( \text{rank}(\mathcal{X}) = 1 \) by regarding eigenvalues no greater than \( 10^{-6} \) as zeros. For example, the maximum eigenvalue of the output \( \mathcal{X} \) for \( n = 5 \) is 0.34 and the remaining four eigenvalues are less than \( 10^{-8} \). This is considered an effect of the nonlinear term \( f_\varepsilon \) in the objective function. In fact, supplementary numerical experiments reveal that, when the objective function consists of \( \text{Tr}(\mathcal{X}) \) only, the problem usually yields a solution with \( \text{rank}(\mathcal{X}) > 3 \). This fact indicates that the nonlinear SISDP (4.6) provides a useful model in FIR filter design.

5 Conclusion

In this paper, we proposed a primal-dual path following method for solving SISDP (1.1). First, we provided a prototype algorithm (Algorithm 1) that attempts to find a KKT point of the SISDP by following a path formed by BKKT points in an infinite dimensional space. We showed that a sequence generated by the algorithm weakly\(^*\) converges to a KKT point under some mild assumptions. Then, based on Algorithm 1, we gave two types of algorithms, Algorithms 2 and 3, using the local reduction method and two-step SQP method. Algorithm 2 is a local algorithm, for which we established two-step superlinear convergence. Algorithm 3 is a global version of Algorithm 2. We showed that Algorithm 3 inherits the global convergence property from Algorithm 1 and the fast local convergence property from Algorithm 2.

In the numerical experiments, we compared Algorithm 3 with the discretization method which solves (nonlinear) SDPs obtained by finite relaxation of SISDP (1.1). We observed that Algorithm 3 exhibited the numerical efficiency comparable to the discretization method. In particular, it worked more effectively in finding highly accurate solutions than the discretization method.

Acknowledgements We would like to thank two anonymous reviewers and the associated editor for their valuable comments and suggestions.

Appendix

A.1 Proof of Theorem 2.1

Firstly, note that \( F(x^*) \bullet V = 0, \ F(x^*) \in S^m_+ \) and \( V \in S^m_+ \) hold if and only if \( F(x^*) \circ V = O, \ F(x^*) \in S^m_+, \) and \( V \in S^m_+ \).

To show the theorem, we utilize [21, Theorem 2.4] related to the KKT conditions for semi-infinite programs with infinitely many conic constraints. To apply this theorem, we let \( \mathcal{W} := S^m \times \mathcal{R}, \ C := S^m_+ \times \mathcal{R}_+, \) and \( H(x, \tau) := (F(x), g(x, \tau)) \in \mathcal{W}. \) Furthermore, we define the inner product \( \langle \cdot, \cdot \rangle \) by \( \langle (Y_1, y_1), (Y_2, y_2) \rangle := Y_1 \bullet Y_2 + y_1 y_2 \) for any \( (Y_1, y_1), (Y_2, y_2) \in \mathcal{W}. \) Then, SISDP (1.1) is rewritten as the following semi-infinite program with infinitely many conic constrains:

\[
\min f(x) \ \text{s.t.} \ H(x, \tau) \in C \ (\tau \in T).
\]
In fact, the SCQ for the SISDP implies the RCQ [21, Definition 2.2], that is, for a given feasible point \( x \), there exists \( d \in \mathbb{R}^n \) such that \( H(x, \tau) + \nabla_x H(x, \tau) \top d \in \text{int} \ C \), which is rewritten as \( F(x + d) \in S^m_{++} \) and \( g(x, \tau) + \nabla_x g(x, \tau) \top d < 0 \) (\( \tau \in T \)). This can be ensured by letting \( d := \bar{x} - x \) with a Slater point \( \bar{x} \) and using the fact that \( F(\bar{x}) \in S^m_{++} \) and \( g(x, \tau) + \nabla_x g(x, \tau) \top d \leq g(\bar{x}, \tau) < 0 \) (\( \tau \in T \)) from the convexity of \( g(\cdot, \tau) \) for any \( \tau \in T \). Hence, by applying [21, Theorem 2.4] to problem (A.1), under the presence of the SCQ, there exist \( p \leq n \) Lagrange multipliers \( Z_j := (V_j, y_j) \in \mathbb{W}(1 \leq j \leq p) \) and \( \tau_1, \tau_2, \ldots, \tau_p \in T \) such that

\[
\nabla f(x^*) - \sum_{j=1}^p \nabla_x H(x^*, \tau_j) Z_j = 0, \\
\langle H(x^*, \tau_j), Z_j \rangle = 0, \quad H(x^*, \tau_j) \in C, \quad Z_j \in C^*, \quad (j = 1, 2, \ldots, p) 
\] (A.2)

where \( C^* \) stands for the dual cone of \( C \) and

\[
\nabla_x H(x^*, \tau_j) Z_j = \left( \frac{(F_i \cdot V_j)_{i=1}^n}{\nabla_x g(x, \tau_j) y_j} \right). 
\]

Then, by letting \( V := \sum_{j=1}^p V_j \in S^m_+ \), we have

\[
\sum_{j=1}^p \nabla_x H(x^*, \tau_j) Z_j = \left( \frac{\sum_{j=1}^p (F_i \cdot V_j)_{i=1}^n}{\sum_{j=1}^p \nabla_x g(x, \tau_j) y_j} \right) = \left( \frac{\sum_{j=1}^p (F_i \cdot V)_{i=1}^n}{\sum_{j=1}^p \nabla_x g(x, \tau_j) y_j} \right). 
\]

By noting this fact together with \( C^* = C \), the conditions (A.2)-(A.3) yields

\[
\nabla f(x^*) + \sum_{j=1}^p \nabla_x g(x^*, \tau_j) y_j - (F_i \cdot V)_{i=1}^n = 0, \\
F(x^*) \cdot V = 0, \quad F(x^*) \in S^m_+, \quad V \in S^m_+, \\
\sum_{j=1}^p g(x^*, \tau_j) y_j = 0, \quad g(x^*, \tau_j) \leq 0, \quad y_j \geq 0 \quad (j = 1, 2, \ldots, p). 
\] (A.4)

(A.5)

(A.6)

Finally, define \( y \in \mathcal{M}_+(T) \) as follows: For any \( \tau \in T \), \( y(\tau) := y_j \) if \( \tau = \tau_j \) with \( 1 \leq j \leq p \) and, otherwise, \( y(\tau) := 0 \). Then, by \( \sum_{j=1}^p g(x^*, \tau_j) y_j = \int_T g(x^*, \tau) dy(\tau) \) and \( \sum_{j=1}^p \nabla_x g(x^*, \tau_j) y_j = \int_T \nabla_x g(x^*, \tau) dy(\tau) \), conditions (A.4)-(A.6) are rewritten as the desired KKT conditions (2.1)-(2.3). By definition, \( |\text{supp}(y)| \leq p \leq n \) holds.

The last assertion can be shown in a manner similar to the the one in the standard convex optimization theory. The proof is complete.

**Proof of Proposition 3.1**

Since \( x^* \) is \( \delta \)-nondegenerate and feasible for the SISDP, \( \max_{\tau \in T} g(x^*, \tau) \leq 0 \) and the number of global maximizers of \( \max_{\tau \in T} g(x^*, \tau) \) is finite. Denote the set of these
global maximizers by $\mathcal{S}$. As $w^*$ is a KKT point of the SISDP, the KKT conditions (2.1)-(2.3) hold. In particular, by the complementarity condition (2.3), for each $A \in \mathcal{B}$, we have the implication that $y^*(A) > 0 \Rightarrow g(x^*, \tau) = 0$ ($\tau \in A$). Hereafter, suppose $\text{supp}(y^*) \neq \emptyset$. Then, by definition, for arbitrarily chosen $s \in \text{supp}(y^*)$, $y^*(N_s) > 0$ holds for any open neighborhood $N_s$ of $s$, which together with the above implication implies that $g(x^*, \tau) = 0$ ($\tau \in N_s$). Then, notice that $\max_{\tau \in T} g(x^*, \tau) \leq 0$ since $x^*$ is feasible to the SISDP. Thus, $s \in N_s \subseteq \mathcal{S}$ for any $s \in \text{supp}(y^*)$. This means $\text{supp}(y^*) \subseteq \mathcal{S}$, which together with the fact that $\mathcal{S}$ is a finite discrete set yields that $\text{supp}(y^*)$ must be a finite discrete set. From this fact, we conclude that $y^*$ is a discrete measure with finite support. Then, (3.3) readily follows from $\text{supp}(y^*) \subseteq \mathcal{S} \subseteq \{\tau_{1s}^*(x^*), \tau_{2s}^*(x^*), \ldots, \tau_{ps}^*(x^*)\}$.

Lastly, we show that $\varepsilon^*$ is a KKT point of the NSDP. Noting (3.3) and the fact that $y^*$ is a discrete measure, we have $\int_T \nabla_x g(x^*, \tau) dy^*(\tau) = \sum_{i=1}^{p(x^*)} \nabla_x g(x^*, \tau_s^*(x^*)) \xi_i^*$ and $\int_T g(x^*, \tau) dy^*(\tau) = \sum_{i=1}^{p(x^*)} g(x^*, \tau_s^*(x^*)) \xi_i^*$.

Using this fact together with the KKT conditions (2.1)–(2.3) with $w^*$ replaced by $\bar{w}$, we obtain

$$
\nabla f(x^*) + \sum_{i=1}^{p(x^*)} \nabla_x g(x^*, \tau_s^*(x^*)) \xi_i^* - (F_j \bullet V_*)_{j=1} = 0,
$$

$$
F(x^*) \circ V_* = O, \quad F(x^*) \in S^m_+, \quad V_* \in S^m_+, \quad \sum_{i=1}^{p(x^*)} g(x^*, \tau_s^*(x^*)) \xi_i^* = 0, \quad g(x^*, \tau_s^*(x^*)) \leq 0, \quad \xi_i^* \geq 0 (i = 1, 2, \ldots, p(x^*)).
$$

This means that $\varepsilon^*$ is a KKT point of the NSDP.

\hfill $\Box$

### A.2 Proof of Proposition 3.3

We begin with giving some lemmas that help to show Proposition 3.3.

\textbf{Lemma A.1} Let $X \in S_+^m$, $Y \in S^m$ and $\mu \geq 0$. Then,

1. $\|XY - YX\|_F \leq 2\|X \circ Y - \mu I\|_F$ and
2. $\|L_X L_Y - L_Y L_X\|_F \leq \|X \circ Y - \mu I\|_F$.

\textbf{Proof} Using some orthogonal matrix $O \in \mathbb{R}^{m \times m}$, we make an eigenvalue decomposition of $X$: $O^T X O = D$ with $D \in \mathbb{R}^{m \times m}$ being a diagonal matrix. Denote the $i$-th diagonal entry of $D$ by $d_i \geq 0$ for $i = 1, 2, \ldots, m$. Let $\tilde{Y} := O^T Y O$ with the $(i, j)$-th entry $\tilde{y}_{ij}$ for $1 \leq i, j \leq m$.

1. We have the desired result from

$$
\|XY - YX\|_F^2 = \|O^T X O \circ O^T Y O - O^T Y O \circ O^T X O\|_F^2
\quad = \|D \tilde{Y} - \tilde{Y} D\|_F^2
\quad = \sum_{1 \leq i, j \leq m} (d_i - d_j)^2 \tilde{y}_{ij}^2
$$

$\square$ Springer
\[ \sum_{1 \leq i \neq j \leq m} (d_i + d_j)^2 \tilde{y}_{ij} \leq \sum_{1 \leq i \neq j \leq m} (d_i + d_j)^2 \tilde{y}_{ij} + \sum_{i=1}^{m} (2d_i \tilde{y}_{ii} - 2\mu)^2 \]
\[ = \| D\tilde{Y} + \tilde{Y}D - 2\mu I \|_F^2 \]
\[ = \| XY + YX - 2\mu I \|_F^2 \]
\[ = 4\| X \circ Y - \mu I \|_F^2, \]

where the first inequality follows from \( d_i \geq 0 \) for \( i = 1, 2, \ldots, m \).

2. By direct calculation, we have
\[ \| LXL_Y - LYL_X \|_2^2 = \max_{\| Z \|_F = 1} \| LXL_Z - LYL_Z \|_F \]
\[ = \max_{\| Z \|_F = 1} \frac{\| (XY - YX)Z - Z(XY - YX) \|_F}{4} \]
\[ \leq \frac{\| XY - YX \|_F^2}{2} \]
\[ \leq \| X \circ Y - \mu I \|_F^2, \]

where the second inequality follows from item 1.

\[ \square \]

Lemma A.2 Let \((X_\ast, Y_\ast) \in S_+^m \times S_+^m\) satisfy the strict complementarity condition that \(X_\ast \circ Y_\ast = O\) and \(X_\ast + Y_\ast \in S_+^m\). Let \(\{\mu_r\} \subseteq \mathcal{R}_{++}\) and \(\{(X_r, Y_r)\} \subseteq S_+^m \times S_+^m\) be sequences such that \(\lim_{r \to \infty} \mu_r = 0\) and \(\lim_{r \to \infty} (X_r, Y_r) = (X_\ast, Y_\ast)\). Let spectral decompositions of \(X_\ast\) and \(Y_\ast\) be
\[ O_\ast^\top X_\ast O_\ast = \begin{pmatrix} D_X & O \\ O & O \end{pmatrix}, \quad O_\ast^\top Y_\ast O_\ast = \begin{pmatrix} O & O \\ O & D_Y \end{pmatrix} \]
using some orthogonal matrix \(O_\ast \in \mathcal{R}^{m \times m}\) and positive diagonal matrices \(D_X \in S_+^p\) and \(D_Y \in S_+^q\) with \(p + q = m\). Furthermore, suppose \(p, q > 0\) and choose a sequence of orthogonal matrices \(\{O_r\} \subseteq \mathcal{R}^{m \times m}\) such that
\[ O_r^\top X_r O_r = \begin{pmatrix} D_X & O \\ O & E_X \end{pmatrix}, \quad \lim_{r \to \infty} O_r = O_\ast \]

with \(D_X \in \mathcal{R}^{p \times p}\) and \(E_X \in \mathcal{R}^{q \times q}\) being positive diagonal matrices for \(r \geq 1\). (Notice that \(\lim_{r \to \infty} E_X = O_\ast\).) If \(\| X_r \circ Y_r - \mu_r I \| = o(\mu_r)\), then
\[ \lim_{r \to \infty} \frac{1}{\mu_r} E_X = D_Y^{-1}. \]
Proof Let $\tilde{Y}_r := \mathcal{O}_r^\top Y_r \mathcal{O}_r$ and $\tilde{y}_r^i$ and $e_r^i$ be the $i$-th diagonal entry of $\tilde{Y}_r$ and $E_{X_r}$, respectively for any $i = p + 1, p + 2, \ldots, m$. Since $\|X_r \circ Y_r - \mu_r I\|_F = o(\mu_r)$ and

$$\|X_r \circ Y_r - \mu_r I\|_F = \left\| \begin{pmatrix} D_{X_r} & 0 \\ 0 & E_{X_r} \end{pmatrix} \right\|_F \geq \sqrt{\sum_{i=p+1}^{m} (e_r^i \tilde{y}_r^i - \mu_r)^2},$$

we have

$$0 = \lim_{r \to \infty} \sqrt{\sum_{i=p+1}^{m} \left( \frac{e_r^i \tilde{y}_r^i - \mu_r}{\mu_r} \right)^2} = \lim_{r \to \infty} \sqrt{\sum_{i=p+1}^{m} \left( \frac{e_r^i \tilde{y}_r^i - 1}{\mu_r} \right)^2},$$

which yields $\lim_{r \to \infty} \frac{e_r^i \tilde{y}_r^i}{\mu_r} = 1$ for any $i = p + 1, \ldots, m$. Notice that, for $i \geq p + 1$, $\{\tilde{y}_r^i\}$ converges to the $i$-th positive diagonal entry of $D_{Y_*}$. In view of these facts, we obtain (A.7).

Now, we are ready to show Proposition 3.3. For the case where $X_* \in S^m_{++}$, it is easy to prove the desired result. So, we consider the case of $X_* \in S^m \setminus S^m_{++}$. Let $\lambda_r > 0$ be the smallest eigenvalue of $X_r$. Notice that $\lambda_r \to 0$ ($r \to \infty$) and, by Lemma A.2, $\lim_{r \to \infty} \frac{\lambda_r}{\mu_r}$ exists and is positive. Thus, we also have

$$\lim_{r \to \infty} \frac{\mu_r}{\lambda_r} > 0. \tag{A.8}$$

Note that, for any $X \in S^m$ having $m$ eigenvalues $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_m$, the corresponding symmetric linear operator $L_X$ has $m(m + 1)/2$ eigenvalues $\alpha_1, \alpha_2, \ldots, \alpha_m, \{(\alpha_i + \alpha_j)/2\}_{i \neq j}$. This fact yields that the maximum eigenvalue of the operator $L_{X_r}^{-1}$ is $\lambda_r^{-1}$. Therefore, we have $\|L_{X_r}^{-1}\|_2 = \lambda_r^{-1}$ for any $r \geq 0$. It then follows that

$$\|L_{X_r}^{-1} - L_{Y_r}^{-1}\|_2 \leq \|L_{X_r} - L_{Y_r}\|_2 \|L_{X_r}^{-1}\|_2 \leq \mu_r \left\| \frac{X_r \circ Y_r - \mu_r I}{\mu_r} \right\|_F = \frac{\mu_r}{\lambda_r} \left\| \frac{X_r \circ Y_r - \mu_r I}{\mu_r} \right\|_F.$$

where the second inequality follows from Lemma A.1. This relation together with (A.8) and $\|X_r \circ Y_r - \mu_r I\|_F = O(\mu_r^{1+\theta})$ implies $\|L_{X_r}^{-1} - L_{Y_r}^{-1}\|_2 = O(\mu_r^\theta)$.
A.4 Proof of Proposition 3.4

Define $U_r(s) := (X_r + s \Delta X_r) \circ (Y_r + s \Delta Y_r)$ for $s \in [0, 1]$ and each $r$. By using the fact that $\|X\|_F \geq |\lambda_{\min}(X)|$ for any $X \in S^m$, conditions (3.25)–(3.27) yield that there exists some $K > 0$ such that

\[
\begin{align*}
\lambda_{\min}(\Delta X_r \circ \Delta Y_r) &\geq -K \mu_r^2, \\
\lambda_{\min}(X_r \circ Y_r) &\geq \mu_r - K \mu_r^{1+\theta}, \\
\lambda_{\min}(Z_r - \hat{\mu}_r I) &\geq -K \hat{\mu}_r^{1+\hat{\theta}}.
\end{align*}
\]

(A.9) \hspace{1cm} (A.10) \hspace{1cm} (A.11)

Then, it holds that

\[
\lambda_{\min}(U_r(s)) = \lambda_{\min}\left((X_r \circ Y_r + s X_r \circ \Delta Y_r + s Y_r \circ \Delta X_r + s^2 \Delta X_r \circ \Delta Y_r)\right)
\]

\[
= \lambda_{\min}\left((1 - s)X_r \circ Y_r + s(Z_r - \hat{\mu}_r I) + s \hat{\mu}_r I + s^2 \Delta X_r \circ \Delta Y_r\right)
\]

\[
\geq (1 - s)\lambda_{\min}(X_r \circ Y_r) + s\lambda_{\min}(Z_r - \hat{\mu}_r I)
\]

\[
+ s\lambda_{\min}(\hat{\mu}_r I) + s^2 \lambda_{\min}(\Delta X_r \circ \Delta Y_r)
\]

\[
\geq (1 - s)\left(\mu_r - K \mu_r^{1+\theta}\right) - sK \hat{\mu}_r^{1+\hat{\theta}} + s \hat{\mu}_r - s^2 K \mu_r^2
\]

\[
=: u_r(s)
\]

for any $r$ sufficiently large and $s \in [0, 1]$, where the first inequality follows from the fact that $\lambda_{\min}(A + B) \geq \lambda_{\min}(A) + \lambda_{\min}(B)$ for $A, B \in S^m$ and the second inequality is due to (A.9)–(A.11) and $s \in [0, 1]$. Notice that $u_r(s)$ is concave and quadratic. Then, for any $r$ sufficiently large, we have $u_r(s) > 0$ ($s \in [0, 1]$) since $0 < \theta, \hat{\theta} < 1$, $\lim_{r \to \infty}(\mu_r, \hat{\mu}_r) = (0, 0)$, and (3.28) imply that $u_r(0) = \mu_r - K \mu_r^{1+\theta} > 0$ and $u_r(1) = \hat{\mu}_r - K \hat{\mu}_r^{1+\hat{\theta}} - K \mu_r^2 > 0$ for sufficiently large $r$. This means that $\lambda_{\min}(U_r(s)) \geq u_r(s) > 0$ ($s \in [0, 1]$) and therefore

\[
U_r(s) \in S^m_{++} \quad (s \in [0, 1]),
\]

(A.12)

from which we can derive $X_r + \Delta X_r \in S^m_{++}$ and $Y_r + \Delta Y_r \in S^m_{++}$. Actually, for contradiction, suppose that either one of these two conditions is not true. We can assume $X_r + \Delta X_r \notin S^m_{++}$ without loss of generality. Recall that $X_r \in S^m_{++}$. Then, there exists some $\tilde{s} \in (0, 1]$ such that $X_r + \tilde{s}\Delta X_r \in S^m_+ \setminus S^m_{++}$. Therefore, we can find some nonzero vector $d \in \mathbb{R}^n$ such that $(X_r + \tilde{s}\Delta X_r)d = 0$. From this fact, we readily have

\[
d^T U_r(\tilde{s})d = \frac{d^T (X_r + \tilde{s}\Delta X_r)(Y_r + \tilde{s}\Delta Y_r)d + d^T (Y_r + \tilde{s}\Delta Y_r)(X_r + \tilde{s}\Delta X_r)d}{2}
\]

\[
= 0,
\]

which contradicts (A.12). Hence, we conclude that $X_r + \Delta X_r \in S^m_{++}$ and $Y_r + \Delta Y_r \in S^m_{++}$ for all $r$ sufficiently large. The proof is complete. \qedsymbol
A.5 Proof of Lemma 3.1

To begin with, notice that Assumptions B-2, C-1, and C-3 yield that for sufficiently large $k$,

$$0 < \zeta_i^k = \Theta(1) \ (i \in I_\alpha(x^*)) \quad \text{(A.13)}$$

Furthermore, by (3.10), we have

$$\zeta_i^k = 0 \ (i \in \{1, 2, \ldots, p(x^*)\} \setminus I_\alpha(x^*)) \quad \text{(A.14)}$$

which together with $\zeta_i^* = 0 \ (i \in \{1, 2, \ldots, p(x^*)\} \setminus I_\alpha(x^*))$ implies $\|\tilde{z}^k - \tilde{z}^*\| = \|z^k - z^*\|$ for all $k$ sufficiently large. Next, by (A.14), $z^k \in N_{\varepsilon_{k-1}}$, and $\varepsilon_{k-1} = \gamma_1 \mu_{k-1}^{1+\alpha}$, it follows that

$$\left\| \nabla f(x^k) + \sum_{i \in I_\alpha(x^*)} \nabla \hat{g}_i(x^k) \zeta_i^k - (F_j \cdot V_k)_{j=1}^n \right\| = o(\mu_{k-1}),$$

$$\|F(x^k) \circ V_k\| = \Theta(\mu_{k-1}), \quad \text{(A.15)}$$

$$\left| \sum_{i \in I_\alpha(x^*)} \zeta_i^k \hat{g}_i(x^k) \right| = o(\mu_{k-1}), \quad \max_{1 \leq i \leq p(x^*)} (\hat{g}_i(x^k))_+ = o(\mu_{k-1}). \quad \text{(A.16)}$$

Noting $a = 2(a)_+ - |a|$ for $a \in \mathcal{R}$, we have

$$\left| \sum_{i \in I_\alpha(x^*)} \zeta_i^k \hat{g}_i(x^k) \right| = \left| 2 \sum_{i \in I_\alpha(x^*)} \zeta_i^k (\hat{g}_i(x^k))_+ - \sum_{i \in I_\alpha(x^*)} \zeta_i^k \hat{g}_i(x^k) \right|,$$

which together with (A.13) and (A.16) yields $|\hat{g}_i(x^k)| = o(\mu_{k-1}) \ (i \in I_\alpha(x^*)).$ From this fact and (A.15), we obtain $\|\Phi_0(\tilde{z}^k)\| = \Theta(\mu_{k-1})$ and thus $\mu_{k-1} = \Theta(\|\Phi_0(\tilde{z}^k)\|)$.

We next prove $\mu_{k-1} = \Theta(\|z^k - z^*\|)$ by showing $\|\Phi_0(\tilde{z}^k)\| = \Theta(\|z^k - z^*\|)$. It suffices to show that the sequence $\{\eta_k\} := \{\|\Phi_0(\tilde{z}^k)\|/\|z^k - z^*\|\}$ is bounded above and away from zero. Note that by $\Phi_0(\tilde{z}^*) = 0$ and Assumption B-1,

$$\eta_k = \frac{\|\Phi_0(\tilde{z}^k) - \Phi_0(\tilde{z}^*)\|}{\|z^k - z^*\|} = \left| \mathcal{J} \Phi_0(\tilde{z}^*) \frac{\tilde{z}^k - \tilde{z}^*}{\|z^k - z^*\|} + \frac{O(\|\tilde{z}^k - \tilde{z}^*\|^2)}{\|z^k - z^*\|} \right|.$$

Obviously, $\{\eta_k\}$ is bounded from above. To show $\{\eta_k\}$ is bounded away from zero, suppose to the contrary. Then, without loss of generality, we can assume that $\lim_{k \to \infty} \eta_k = 0$, and hence there exists some $d^*$ with $\|d^*\| = 1$ such that $\lim_{k \to \infty} \frac{\tilde{z}^k - \tilde{z}^*}{\|z^k - z^*\|} = d^*$ and $\mathcal{J} \Phi_0(\tilde{z}^*)d^* = 0$. However, this contradicts the nonsingularity of $\mathcal{J} \Phi_0(\tilde{z}^*)$ from Assumption B-3. We have the desired conclusion. \qed
References

1. Alizadeh, F., Goldfarb, D.: Second-order cone programming. Math. Program. 95, 3–51 (2003)
2. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer-Verlag, New York (2000)
3. Bertsekas, D.P., Nedic, A., Ozdaglar, A.E.: Convex Analysis and Optimization. Athena Scientific, Cambridge (2003)
4. Correa, R., Ramírez, C.H.: A global algorithm for nonlinear semidefinite programming. SIAM J. Optim. 15, 303–318 (2004)
5. Fazel, M., Hindi, H., and Boyd, S. P.: Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices. In: American Control Conference, 2003. Proceedings of the 2003, 3, 2156-2162. IEEE (2003)
6. Floudas, C.A., Stein, O.: The adaptive convexification algorithm: a feasible point method for semi-infinite programming. SIAM J. Optim. 18, 1187–1208 (2007)
7. 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)
8. Goberna, M.A., López, M.A.: Semi-Infinite Programming: Recent Advances. Kluwer Academic Publishers, Dordrecht (2001)
9. Gramlich, G., Hettich, R., Sachs, E.W.: Local convergence of SQP methods in semi-infinite programming. SIAM J. Optim. 5, 641–658 (1995)
10. Hayashi, S. and S.-Y. Wu, S.-Y.: An explicit exchange algorithm for linear semi-infinite programming problems with second-order cone constraint, SIAM J. Optim. 20, 1527–1546 (2009)
11. Hettich, R., Kortanek, K.O.: Semi-infinite programming: Theory, methods, and applications. SIAM Rev. 35, 380–429 (1993)
12. Jiang, A. and Kwan, H. K.: Minimax design of IIR digital filters using SDP relaxation technique, IEEE Trans. Circuits Syst. I, Reg. Papers, 57, 378–390 (2009)
13. Konno, H., Kawadai, N., Tuy, H.: Cutting plane algorithms for nonlinear semi-definite programming problems with applications. J. Global Optim. 25, 141–155 (2003)
14. Leibfritz, F., Maruhn, J.H.: A successive SDP-NSDP approach to a robust optimization problem in finance. Comput. Optim. and Appl. 44, 443–466 (2009)
15. Li, D., Qi, L., Tam, J., Wu, S.-Y.: A smoothing Newton method for semi-infinite programming. J. Global Optim. 30, 169–194 (2004)
16. Li, S., Teo, K.L., Yang, X., Wu, S.-Y.: Robust envelope-constrained filter with orthonormal bases and semi-definite and semi-infinite programming. Optim. Eng. 8, 299–319 (2007)
17. Li, S., Wu, S.-Y., Yang, X., Teo, K.L.: A relaxed cutting plane method for semi-infinite semi-definite programming. Comput. Optim. and Appl. 196, 459–473 (2006)
18. Li, S., Yang, X., Teo, K.L., Wu, S.-Y.: A solution method for combined semi-infinite and semi-definite programming. ANZIAM J. 45, 477–494 (2004)
19. Okuno, T., Fukushima, M.: Local reduction based SQP-type method for semi-infinite programs with an infinite number of second-order cone constraints. J. Global Optim. 60, 25–48 (2014)
20. Okuno, T., Fukushima, M.: An interior point sequential quadratic programming-type method for log-determinant semi-infinite programs. J. Comput. Appl. Math. 376, 112784 (2020)
21. Okuno, T., Hayashi, S., Fukushima, M.: A regularized explicit exchange method for semi-infinite programs with an infinite number of conic constraints. SIAM J. Optim. 22, 1009–1028 (2012)
22. Pereira, A., Costa, M., Fernandes, E.: Interior point filter method for semi-infinite programming problems. Optimization 60, 1309–1338 (2011)
23. Pereira, A., Fernandes, E.: A reduction method for semi-infinite programming by means of a global stochastic approach. Optimization 58, 713–726 (2009)
24. Qi, L., Wu, S.-Y., Zhou, G.: Semismooth Newton methods for solving semi-infinite programming problems. J. Global Optim. 27, 215–232 (2003)
25. Reemtsen, R.: Discretization methods for the solution of semi-infinite programming problems. J. Optim. Theory Appl. 71, 85–103 (1991)
26. Reemtsen, R. and Rückmann, J.: Semi-infinite Programming, Kluwer Academic Publishers, Boston (1998)
27. Shapiro, A.: First and second order analysis of nonlinear semidefinite programs. Math. Program. 77, 301–320 (1997)
Primal-dual path following...

28. Shiu, T., Wu, S.-Y.: Relaxed cutting plane method with convexification for solving nonlinear semi-infinite programming problems. Comput. Optim. and Appl. 53, 1–23 (2012)

29. Stein, O., Steuermann, P.: The adaptive convexification algorithm for semi-infinite programming with arbitrary index sets. Math. Program. 18, 1–25 (2012)

30. Still, G.: Discretization in semi-infinite programming: the rate of convergence. Math. Program. 91, 53–69 (2001)

31. Tanaka, Y., Fukushima, M., Ibaraki, T.: A globally convergent SQP method for semi-infinite nonlinear optimization. J. Comput. Appl. Math. 23, 141–153 (1988)

32. Todd, M.J., Toh, K.C., Tütüncü, R.H.: On the Nesterov-Todd direction in semidefinite programming. SIAM J. Optim. 8, 769–796 (1998)

33. Toh, K. C., Todd, M. J., and Tütüncü, R. H.: SDPT3—a MATLAB software package for semidefinite programming, version 2.1. Optim. Methods Softw. 11, 545–581 (1999)

34. Wang, S., Yuan, Y.: Feasible method for semi-infinite programs. SIAM J. Optim. 25, 2537–2560 (2015)

35. Wolkowicz, H., Saigal, R., Vandenberghe, L.: Handbook of Semidefinite Programming: Theory, Algorithms, and Applications. Springer Science & Business Media, Germany (2012)

36. Wei, D., Sestok, C.K., Oppenheim, A.: V: Sparse filter design under a quadratic constraint: low-complexity algorithms. IEEE Trans. Sig. Process 857–870, 61 (2012)

37. Wen, F., Chu, L., Liu, P., Qiu, R.: C: A survey on nonconvex regularization-based sparse and low-rank recovery in signal processing, statistics, and machine learning. IEEE Access 6983–69906, 6 (2018)

38. Wu, S.-P., Boyd, S., and Vandenberghe, L.: FIR filter design via semidefinite programming and spectral factorization, In: Proc. IEEE Conf. on Decision and Control, pp. 271–276 (1996)

39. Wu, S.-P., Boyd, S., and Vandenberghe, L.: FIR filter design via spectral factorization and convex optimization. Appl. Comput. Control, Signals and Circuits, pp. 215–245, (1999)

40. Xu, M., Wu, S.-Y., Ye, J.J.: Solving semi-infinite programs by smoothing projected gradient method. Comput. Optim. and Appl. 59, 591–616 (2014)

41. 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)

42. Yamashita, H., Yabe, H.: A survey of numerical methods for nonlinear semidefinite programming. J. Oper. Res. Soc. Japan. 58, 24–60 (2015)

43. Yamashita, H., Yabe, H., Harada, K.: A primal-dual interior point method for nonlinear semidefinite programming. Math. Program. 135, 89–121 (2012)

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