Mathematical Programming 81 (1998) 281–299

A unified analysis for a class of long-step primal-dual path-following interior-point algorithms for semidefinite programming

Renato D.C. Monteiro [a,1], Yin Zhang [b,2,*]

[a] *School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA*

[b] *Department of Computational and Applied Mathematics, Rice University, MS 134 Houston, TX 77005-1892, USA*

Received 30 August 1996; revised manuscript received 20 May 1997

Abstract

We present a unified analysis for a class of long-step primal-dual path-following algorithms for semidefinite programming whose search directions are obtained through linearization of the symmetrized equation of the central path $H_P(XS) \equiv [PXSP^{-1} + (PXSP^{-1})^T]/2 = \mu I$, introduced by Zhang. At an iterate (X, S), we choose a scaling matrix P from the class of nonsingular matrices P such that $PXSP^{-1}$ is symmetric. This class of matrices includes the three well-known choices, namely: $P = S^{1/2}$ and $P = X^{-1/2}$ proposed by Monteiro, and the matrix P corresponding to the Nesterov–Todd direction. We show that within the class of algorithms studied in this paper, the one based on the Nesterov–Todd direction has the lowest possible iteration-complexity bound that can provably be derived from our analysis. More specifically, its iteration-complexity bound is of the same order as that of the corresponding long-step primal-dual path-following algorithm for linear programming introduced by Kojima, Mizuno and Yoshise. © 1998 The Mathematical Programming Society, Inc. Published by Elsevier Science B.V.

Keywords: Semidefinite programming; Primal-dual; Path-following; Interior-point methods

1. Introduction

Several authors have discussed generalizations of interior-point algorithms for linear programming (LP) to the context of semidefinite programming (SDP). The landmark work in this direction is due to Nesterov and Nemirovskii [14,15] where a

* Corresponding author.

[1] This author's research is supported in part by the National Science Foundation under grants INT-9600343 and CCR-9700448 and the Office of Naval Research under grant N00014-94-1-0340.

[2] This author's research was supported in part by DOE DE-FG02-93ER25171-A001.

PII S0025-5610(97)00070-1

general approach for using interior-point methods to solve convex programs is proposed based on the notion of self-concordant functions. (See their book [17] for a comprehensive treatment of this subject.) They show that the problem of minimizing a linear function over a convex set can be solved in "polynomial time" as long as a self-concordant barrier function for the convex set is known. In particular, Nesterov and Nemirovskii show that linear programs, convex quadratic programs with convex quadratic constraints, and semidefinite programs all have explicit and easily computable self-concordant barrier functions, and hence can be solved in "polynomial time". On the other hand, Alizadeh [1] extends Ye's projective potential reduction algorithm [26] for LP to SDP and argues that many known interior-point algorithms for LP can also be transformed into algorithms for SDP in a mechanical way. Since then many authors have proposed interior-point algorithms for solving SDP problems, including Alizadeh et al. [2], Helmberg et al. [3], Jarre [5], Kojima et al. [7,9], Lin and Saigal [10], Luo et al. [11], Monteiro [12], Nesterov and Nemirovskii [16], Nesterov and Todd [18,19], Potra and Sheng [20], Sturm and Zhang [22], Vandenberghe and Boyd [25], and Zhang [27]. Most of these more recent works are concentrated on primal-dual methods.

This paper studies long-step primal-dual path-following interior-point algorithms for SDP. Each search direction is the solution of a linear system containing a symmetrization of the equation $X\Delta S + \Delta XS = R$, where R is an appropriate matrix. The first suggested and most natural symmetrization is to take the symmetric part of both sides of this equation. This approach results in the direction introduced by Alizadeh et al. [2].

Another way of symmetrizing is to first apply a similarity transformation $P(\cdot)P^{-1}$ to both sides of the equation and then symmetrize. Such an approach was first introduced by Monteiro [12] for the cases of $P = X^{-1/2}$ and $P = S^{1/2}$. The resulting directions were found to be equivalent to two special directions in the class of directions introduced earlier by Kojima et al. [9] through a different formulation. The second direction (with $P = S^{1/2}$) was also proposed by Helmberg et al. [3] independently from [9]. To unify the above directions, including the one by Alizadeh et al. [2], Zhang [27] formally introduced a general symmetrization scheme using an arbitrary nonsingular scaling matrix P, which leads to a class of search directions corresponding to different P matrices. In a recent paper, Todd et al. [24] study conditions for the existence and uniqueness of search directions in the above class, and show that the Nesterov–Todd direction [19] is a member of this class corresponding to any scaling matrix P such that $P^{\mathrm{T}}P = S^{1/2}(S^{1/2}XS^{1/2})^{-1/2}S^{1/2}$. More recently, Kojima et al. [8] demonstrate that the Nesterov–Todd direction also belongs to the class of directions introduced by Kojima et al. [9].

The goal of this paper is to establish iteration-complexity bounds for a class of long-step primal-dual path-following algorithms based on a certain "commutative" subset of the set of search directions obtained by a scaling and symmetrization scheme, which will be described in detail in Section 2. This scheme was originally introduced by Monteiro [12] for two special cases of scaling matrices and later generalized by Zhang [27] to general scaling matrices. For convenience, we will call the set of search directions generated by this scheme the Monteiro–Zhang family of search directions. As a result of our unified analysis, we are able to derive polynomial convergence for several long-step path-following algorithms based on search directions from the Monteiro–Zhang family, including all the directions discussed above but the Alizadeh–Haeberly–Overton (AHO) direction. Our derived iteration-complexity bound is $O(\sqrt{\kappa_\infty} nL)$, where κ_∞ is the supremum of the spectral condition numbers of certain matrices G^k, with each G^k determined by the current iterate (X^k, S^k) and scaling matrix P^k. We show that the method based on the Nesterov–Todd direction has the lowest possible iteration-complexity bound derivable from our analysis because the corresponding G-matrices are always equal to the identity matrix. More specifically, its iteration-complexity is of the same order, namely $\mathcal{O}(nL)$, as that of the corresponding long-step primal-dual path-following algorithm for LP introduced by Kojima et al. [6]. In contrast, the scaling matrices $P = X^{-1/2}$ and $P = S^{1/2}$ lead to G-matrices that have an $\mathcal{O}(n)$ upper bound on their condition numbers. Consequently, the algorithms based on these directions have $\mathcal{O}(n^{1.5}L)$ iteration-complexity bounds. These latter iteration-complexity bounds were first proved in Monteiro [12], and in this paper follow as byproducts from our unified analysis.

This paper is organized as follows. In Section 2, we introduce the SDP problem and a general symmetrization scheme that motivates the Monteiro–Zhang family of search directions. In Section 3, we introduce the so-called commutative class, a subset of the Monteiro–Zhang family, of scaling matrices, describe the long-step primal-dual path-following algorithms corresponding to this class of scaling matrices, and present the main convergence results. The key result, Theorem 3.1, is stated in Section 3 without a proof. Section 4 contains technical results and a proof for Theorem 3.1. We present several other relevant results in Section 5. Finally, concluding remarks are given in Section 6.

1.1. Notation and terminology

The following notation is used throughout the paper. The superscript T denotes transpose. $\mathbb{R}^p$ denotes the p-dimensional Euclidean space. The set of all $p \times q$ matrices with real entries is denoted by $\mathbb{R}^{p\times q}$. The set of all symmetric $p \times p$ matrices is denoted by $\mathcal{S}^p$. For $Q \in \mathcal{S}^p$, $Q \succeq 0$ means Q is positive semidefinite and $Q \succ 0$ means Q is positive definite. The trace of a matrix $Q \in \mathbb{R}^{p\times p}$ is denoted by $\operatorname{tr} Q \equiv \sum_{i=1}^{p} Q_{ii}$. For a matrix $Q \in \mathbb{R}^{p\times p}$ with all real eigenvalues, we denote its eigenvalues by $\lambda_i[Q]$, $i = 1, \dots, p$, and its smallest and largest eigenvalues by $\lambda_{\min}[Q]$ and $\lambda_{\max}[Q]$, respectively; moreover, the spectral condition number of a symmetric matrix Q is denoted by $\operatorname{cond}(Q) \equiv \lambda_{\max}[Q]/\lambda_{\min}[Q]$. Given P and Q in $\mathbb{R}^{p\times q}$, the inner product between them in the vector space $\mathbb{R}^{p\times q}$ is defined as $P \cdot Q \equiv \operatorname{tr} P^{\mathrm{T}} Q$. The Euclidean norm and its associated operator norm are both denoted by $\|\cdot\|$; hence, $\|Q\| \equiv \max_{\|u\|=1} \|Qu\|$ for any $Q \in \mathbb{R}^{p\times p}$. The Frobenius norm of $Q \in \mathbb{R}^{p\times p}$ is $\|Q\|_F \equiv (Q \cdot Q)^{1/2}$. $\mathcal{S}^p_+$ and $\mathcal{S}^p_{++}$ denote the set of all matrices in $\mathcal{S}^p$ which are positive semidefinite and positive definite, respectively. For any $p \times q$ matrix A, *vec* A denotes the pq-vector obtained by stacking the columns of A one by one from the first to the last column. The Kronecker product of two matrices A and B is denoted by $A \otimes B$ (see [4] for a comprehensive treatment on Kronecker products and related topics).

2. The SDP problem and a symmetrization scheme

In this section we first introduce the SDP problem and the corresponding assumptions that will be used in our presentation. We also motivate the Monteiro–Zhang family of search directions used by our class of algorithms via a scaled symmetric equation of the central path which was introduced by Zhang [27].

Consider the following SDP problem which we call the primal SDP problem

$$\begin{aligned} \min \quad & C \cdot X \\ \text{s.t.} \quad & A_i \cdot X = b_i, \quad i = 1, 2, \ldots, m, \\ & X \succeq 0, \end{aligned} \tag{2.1}$$

where $C \in \mathscr{S}^n$, $A_i \in \mathscr{S}^n$, $i = 1, \ldots, m$ and $b = (b_1, \ldots, b_m) \in \mathbb{R}^m$ are the data, and $X \in \mathscr{S}^n$ is the primal variable.

The corresponding dual SDP problem is

$$\begin{aligned} \max \quad & b^{\mathrm{T}} y \\ \text{s.t.} \quad & \sum_{i=1}^{m} y_i A_i + S = C, \\ & S \succeq 0, \end{aligned} \tag{2.2}$$

where $y \in \mathbb{R}^m$ and $S \in \mathscr{S}^n$ are the dual variables.

The sets of *feasible interior points* for (2.1) and (2.2) are, respectively,

$$\mathscr{F}^0(P) \equiv \{X \in \mathscr{S}^n_{++}\colon\ A_i \cdot X = b_i, i = 1, \ldots, m\},$$

$$\mathscr{F}^0(D) \equiv \{(S, y) \in \mathscr{S}^n_{++} \times \mathbb{R}^m\colon\ \sum_{i=1}^{m} y_i A_i + S = C\}.$$

Assumption 2.1. We make the following assumptions throughout our presentation:
(a) $\mathscr{F}^0(P) \times \mathscr{F}^0(D) \neq \emptyset$;
(b) the matrices A_i, $i = 1, \ldots, m$, are linearly independent.

Under Assumption 2.1(a), it is well known that both (2.1) and (2.2) have optimal solutions and that their optimal values are the same. Hence, if X^* and (S^*, y^*) are optimal solutions of (2.1) and (2.2), respectively, we have $C \cdot X^* = b^{\mathrm{T}} y^*$. This last condition can be alternatively expressed as $X^* \cdot S^* = 0$, or equivalently $X^* S^* = 0$. Thus, the set of primal and dual optimal solutions of (2.1) and (2.2) consist of all solutions $(X, S, y) \in \mathscr{S}^n_+ \times \mathscr{S}^n_+ \times \mathbb{R}^m$ of the following optimality system:

$$\begin{aligned} XS &= 0, \\ \sum_{i=1}^{m} y_i A_i + S - C &= 0, \\ A_i \cdot X - b_i &= 0, \quad i = 1, \ldots, m, \end{aligned} \tag{2.3}$$

where the first equation is called the complementarity condition. Observe that for $X, S \in \mathcal{S}^n$, the product XS is generally not in $\mathcal{S}^n$. Hence, the left-hand side of Eq. (2.3) is a map from $\mathcal{S}^n \times \mathcal{S}^n \times \mathbb{R}^m$ to $\mathbb{R}^{n\times n} \times \mathcal{S}^n \times \mathbb{R}^m$. Thus, the optimality system (2.3) is *not* square when X and S are restricted to $\mathcal{S}^n$.

The methods discussed in this paper are all based on applying Newton-like methods to Eq. (2.3). One possible approach is to apply a (damped) Newton-type method to Eq. (2.3) viewing its left-hand side as a map from $\mathbb{R}^{n\times n} \times \mathcal{S}^n \times \mathbb{R}^m$ into itself. This approach leads to a well-defined direction $(\widetilde{\Delta X}, \Delta S, \Delta y)$ in $\mathbb{R}^{n\times n} \times \mathcal{S}^n \times \mathbb{R}^m$. Since we want all iterates to be in $\mathcal{S}^n \times \mathcal{S}^n \times \mathbb{R}^m$, we must then symmetrize $(\widetilde{\Delta X}, \Delta S, \Delta y)$ to obtain a search direction in $\mathcal{S}^n \times \mathcal{S}^n \times \mathbb{R}^m$. The approach suggested by Helmberg et al. [3] and Kojima et al. [9] is to take the search direction $(\Delta X, \Delta S, \Delta y)$ where $\Delta X \equiv (\widetilde{\Delta X} + \widetilde{\Delta X}^{\mathrm{T}})/2$, i.e., to project $\widetilde{\Delta X}$ onto $\mathcal{S}^n$. An independent and subsequent derivation of this direction was given in [12]. This paper describes two search directions generated by two novel symmetrization schemes that are special cases of the general symmetrization scheme studied in this paper (see the following paragraphs). In the sequel, we refer to the direction proposed by Helmberg et al. [3], Kojima et al. [9] and Monteiro [12] as the HRVW/KSH/M direction. The second direction introduced by Monteiro in [12], a dual counterpart of the HRVW/KSH/M direction, has also been found to be a member of the class proposed by Kojima et al. [9]. We will refer to this second direction as the KSH/M direction. A more detailed description of these directions will appear below in the context of the general symmetrization scheme.

Another approach that enables the application of Newton-like methods is to make the optimality system (2.3) square by modifying the left-hand side of Eq. (2.3) to a map from $\mathcal{S}^n \times \mathcal{S}^n \times \mathbb{R}^m$ into itself. To achieve this, it is necessary to replace the first equation in (2.3) by an equivalent equation of the form $\Phi(X,S) = 0$, where $\Phi: \mathcal{S}^n_+ \times \mathcal{S}^n_+ \to \mathcal{S}^n$ is a map such that, for every $(X,S) \in \mathcal{S}^n_+ \times \mathcal{S}^n_+$, $\Phi(X,S) = 0$ if and only if $XS = 0$. Approaches along this line were first proposed by Alizadeh et al. [2] (without scaling) and later by Monteiro [12] (with scaling). Motivated by these works, especially by the latter one, Zhang [27] introduced a general symmetrization scheme based on the so-called *similar symmetrization* operator $H_{\mathrm{P}}: \mathbb{R}^{n\times n} \to \mathcal{S}^n$ defined as

$$H_{\mathrm{P}}(M) \equiv \tfrac{1}{2}\left[PMP^{-1} + (PMP^{-1})^{\mathrm{T}}\right] \quad \forall M \in \mathbb{R}^{n\times n},$$

where $P \in \mathbb{R}^{n\times n}$ is some nonsingular matrix. Zhang [27] also observes that

$$H_{\mathrm{P}}(M) = \tau I \iff M = \tau I,$$

for any nonsingular matrix P, any matrix M with real spectrum (e.g., $M = XS$ with $X, S \in \mathcal{S}^n_+$) and any $\tau \in \mathbb{R}$. Consequently, for any given nonsingular matrix P, Eq. (2.3) is equivalent to the following square system from $\mathscr{S}^n \times \mathscr{S}^n \times \mathbb{R}^m$ into itself:

$$H_P(XS) = 0, \tag{2.4}$$

$$\sum_{i=1}^{m} y_i A_i + S - C = 0, \tag{2.5}$$

$$A_i \cdot X - b_i = 0, \quad i = 1, \ldots, m \tag{2.6}$$

to which Newton-like methods can be applied. A perturbed Newton method applied to this system leads to the following linear system for direction $(\Delta X, \Delta S, \Delta y) \in \mathscr{S}^n \times \mathscr{S}^n \times \mathbb{R}^m$:

$$H_P(\Delta X\, S + X\, \Delta S) = \sigma\mu I - H_P(XS), \tag{2.7}$$

$$\sum_{i=1}^{m} \Delta\, y_i A_i + \Delta S = C - S - \sum_{i=1}^{m} y_i A_i, \tag{2.8}$$

$$A_i \cdot \Delta X = b_i - A_i \cdot X, \quad i = 1, \ldots, m, \tag{2.9}$$

where $\sigma \in (0, 1)$ is the centering parameter and $\mu = \mu(X, S) \equiv (X \cdot S)/n$ the normalized duality gap corresponding to (X, S, y). The solution of this linear system is the Newton step at the point (X, S, y) with respect to a system of equations that defines the unique point on the central path with duality gap $\sigma\mu$, namely the system consisting of Eqs. (2.5) and (2.6) and the equation $H_P(XS) = \sigma\mu I$. However, we observe that if the scaling matrix P varies from iteration to iteration, the directions determined by Eqs. (2.7)–(2.9) are *not* Newton steps with respect to the same system, even when $\sigma\mu$ remains constant.

The following simple result shows that there is no loss of generality in restricting our attention to those scaling matrices P that are in $\mathscr{S}^n_{++}$, since they yield all the possible directions that can be generated by system (2.7)–(2.9) as P varies over the set of nonsingular matrices.

Proposition 2.1. *The set of solutions to system (2.7)–(2.9) remains invariant as long as the matrix $W = P^{\mathrm{T}}P$ does not change.*

Proof. Note that Eq. (2.7) is equivalent to the equation

$$W(X\,\Delta S + \Delta X\, S) + (\Delta S X + S\,\Delta X)W = W(\sigma\mu I - XS) + (\sigma\mu I - SX)W$$

obtained by multiplying Eq. (2.7) on the left by P^{T} and on the right by P. Since this last equation depends on W only, the result follows. □

Hence, for fixed $W \in \mathscr{S}^n_{++}$, there is no loss of generality in considering only the matrix $W^{1/2}$ among all those scaling matrices P such that $P^{\mathrm{T}}P = W$, since their corresponding system (2.7)–(2.9) all have the same solution set. Thus we only need to consider scaling matrices that are in $\mathscr{S}^n_{++}$.

Eq. (2.7) can also be written as

$$H(\hat{X}\,\Delta\hat{S} + \Delta\hat{X}\,\hat{S}) = \sigma\mu I - H(\hat{X}\hat{S}), \tag{2.10}$$

where $H \equiv H_I$ is the plain symmetrization operator and

$$\hat{X} \equiv PXP, \quad \Delta\hat{X} \equiv P\Delta X P, \quad \hat{S} \equiv P^{-1}SP^{-1}, \quad \Delta\hat{S} \equiv P^{-1}\Delta S P^{-1}. \tag{2.11}$$

Moreover, in terms of the Kronecker product, Eq. (2.10) becomes

$$\hat{E}\mathrm{vec}\ \Delta\hat{X} + \hat{F}\mathrm{vec}\ \Delta\hat{S} = \mathrm{vec}(\sigma\mu I - H(\hat{X}\hat{S})), \tag{2.12}$$

where

$$\hat{E} \equiv \tfrac{1}{2}(\hat{S}\otimes I + I\otimes\hat{S}), \qquad \hat{F} \equiv \tfrac{1}{2}(\hat{X}\otimes I + I\otimes\hat{X}). \tag{2.13}$$

The following result due to Todd et al. (see Theorem 3.1 of [24]) gives a sufficient condition for system (2.7)–(2.9) to have a unique solution. It is a generalization of a result of Shida et al. [21] corresponding to the case in which $P = I$.

Proposition 2.2. *Let $X, S, P \in \mathcal{S}^n_{++}$ be given and suppose Assumption 2.1(b) holds. Then, a sufficient condition for system (2.7)–(2.9) to have a unique solution is that $\hat{E}\hat{F} + \hat{F}\hat{E} \succ 0$. Moreover, the latter condition holds if $H(\hat{X}\hat{S}) = H_P(XS) \succeq 0$.*

The choices of $P = S^{1/2}$ and $P = X^{-1/2}$ in Eq. (2.7) lead to the two directions described by Monteiro in [12], namely, the HRVW/KSH/M direction and the KSH/M direction, respectively. These directions are in the class of directions proposed earlier by Kojima et al. [9] using a different approach. The HRVW/KSH/M direction, corresponding to $P = S^{1/2}$, was also proposed by Helmberg et al. [3] independently from [9]. As was shown by Todd et al. [24], the choice of $P = W_{\mathrm{nt}}^{1/2}$ in Eq. (2.7), where

$$W_{\mathrm{nt}} \equiv S^{1/2}(S^{1/2}XS^{1/2})^{-1/2}S^{1/2} = X^{-1/2}(X^{1/2}SX^{1/2})^{1/2}X^{-1/2} \tag{2.14}$$

leads to the Nesterov–Todd direction [19], which was originally derived via a different approach (see also [22] for another derivation). The defining property of the matrix $W = W_{\mathrm{nt}}$ (or $P = W_{\mathrm{nt}}^{1/2}$) is that it is the unique solution W (or P) in $\mathcal{S}^n_{++}$ of the equation

$$WXW = S \quad (\text{or } PXP = P^{-1}SP^{-1}) \tag{2.15}$$

(see [19,22] for more details). Observe that using the notation (2.11), identity (2.15) corresponds to the following symmetric property of the Nesterov–Todd direction

$$P = W_{\mathrm{nt}}^{1/2} \;\Rightarrow\; \hat{X} = \hat{S} \quad \text{and} \quad \hat{E} = \hat{F}. \tag{2.16}$$

Finally, we mention that the choice of $P = I$ in Eq. (2.7) leads to the AHO direction proposed in [2].

3. The algorithm and main results

In this section we describe the class of long-step primal-dual path-following algorithms that we will consider in this paper. We also present the main convergence results for this class of algorithms. The key result is Theorem 3.1 whose proof will be given in Section 4.

The class of algorithms we consider are based on the following so-called "large neighborhood" of the central path:

$$\mathscr{N}(\gamma) \equiv \{(X,S,y) \in \mathscr{F}^0(P) \times \mathscr{F}^0(D)\colon\ \lambda_{\min}[XS] \geqslant \gamma\mu(X,S)\}, \tag{3.1}$$

where $\gamma \in (0,1)$ is a given constant, and

$$\mu(X,S) \equiv \frac{X \cdot S}{n} \quad \text{for every } (X,S) \in \mathscr{S}^n \times \mathscr{S}^n.$$

Recall from Proposition 2.2 that system (2.7)–(2.9) has a unique solution whenever the scaling matrix $P \in \mathscr{S}^n_{++}$ satisfies $H_{\mathrm{P}}(XS) = H(\hat{X}\hat{S}) \succeq 0$ (also recall that $\hat{X}$ and $\hat{S}$ are defined in Eq. (2.11)). In our class of algorithms, we impose a stronger condition on P; that is, P belongs to the following class of scaling matrices:

$$\mathscr{P}(X,S) \equiv \{P \in \mathscr{S}^n_{++}\colon\ P^2XS = SXP^2\} = \{P \in \mathscr{S}^n_{++}\colon\ \hat{X}\hat{S} = \hat{S}\hat{X}\}. \tag{3.2}$$

Since the scaling matrices in $\mathscr{P}(X,S)$ make $\hat{X}$ and $\hat{S}$ commute, we will refer to $\mathscr{P}(X,S)$ as the *commutative class* of scaling matrices at (X,S). Observe that the scaling matrices $P = W_{\mathrm{nt}}^{1/2}$, $P = S^{1/2}$ and $P = X^{-1/2}$, corresponding to the NT direction, the HRVW/KSH/M direction and the KSH/M direction, are contained in $\mathscr{P}(X,S)$, and hence our analysis applies to these directions. However, $P = I$ does not belong to $\mathscr{P}(X,S)$, except for the unlikely case when X and S commute; hence, our analysis does not apply to the AHO direction.

The long-step primal-dual path-following algorithm stated below uses search directions exclusively from the commutative class. We will call it Algorithm-CC, where "CC" stands for the "commutative class". It depends on three parameters γ, σ and L which are assumed to be independent of the dimension n.

Algorithm-CC

Let $\gamma, \sigma \in (0,1)$, $L > 1$ and $(X^0, S^0, y^0) \in \mathscr{N}(\gamma)$ be given.
Set $k = 0$ and $\mu_0 = (X^0 \cdot S^0)/n$.
Repeat until $\mu_k \leqslant 2^{-L}\mu_0$, **do**
(1) Choose a scaling matrix $P^k \in \mathscr{P}(X^k, S^k)$.
(2) Compute a solution $(\Delta X^k, \Delta S^k, \Delta y^k)$ of system (2.7)–(2.9) with $(X,S,y) = (X^k, S^k, y^k)$, $P = P^k$ and $\mu = \mu_k$.
(3) Let $\alpha_k = \max\{\bar{\alpha} \in [0,1] : (X^k, S^k, y^k) + \alpha(\Delta X^k, \Delta S^k, \Delta y^k) \in \mathscr{N}(\gamma),$ $\alpha \in [0, \bar{\alpha}]\}$.
(4) Set $(X^{k+1}, S^{k+1}, y^{k+1}) = (X^k, S^k, y^k) + \alpha_k(\Delta X^k, \Delta S^k, \Delta y^k)$.
(5) Set $\mu_{k+1} = (X^{k+1} \cdot S^{k+1})/n$ and increment k by 1.

End

We emphasize that the scaling matrix $P = P^k$ in Algorithm-CC is allowed to change from iteration to iteration. Consequently, the optimality system (2.4)–(2.6) changes from iteration to iteration, although its solution set remains invariant. Because of this, these algorithms are not Newton method in the usual sense.

Recall that the matrices $\hat{E}$ and $\hat{F}$ depend on X, S and P. Let $\hat{E}^k$ and $\hat{F}^k$ denote the matrices $\hat{E}$ and $\hat{F}$ evaluated at X^k, S^k and P^k, respectively, and define

$$\kappa_\infty \equiv \sup\left\{\text{cond}\left[(\hat{E}^k)^{-1}\hat{F}^k\right]:\ k = 0, 1, 2, \ldots\right\}. \tag{3.3}$$

Clearly, $\kappa_\infty \geqslant 1$. The following result gives an iteration-complexity bound for Algorithm-CC in terms of κ_∞.

Theorem 3.1. *Assume that $\kappa_\infty < \infty$. Then the sequence $\{\mu_k\}$ generated by Algorithm-CC satisfies*

$$\mu_{k+1} = (1 - (1 - \sigma)\alpha_k)\mu_k,$$

where

$$\alpha_k \geqslant \min\left(1, \frac{2\sigma(1-\gamma)}{1 - 2\sigma + \sigma^2/\gamma}\,\frac{1}{\sqrt{\kappa_\infty}\ n}\right).$$

Consequently, Algorithm-CC terminates in at most $\mathrm{O}(\sqrt{\kappa_\infty} nL)$ iterations.

We will leave the proof of Theorem 3.1 to Section 4 after developing some necessary technical lemmas. We now specialize Theorem 3.1 to the three special choices of the sequence $\{P^k\}$ that lead to the NT direction, the HRVW/KSH/M direction and the KSH/M direction.

Lemma 3.1. *The number κ_∞ defined by Eq. (3.3) satisfies:*

$$\kappa_\infty \begin{cases} = 1 & \text{if } P^k = (W^k_{\text{nt}})^{1/2} \quad \forall k, \\ \leqslant n/\gamma & \text{if } P^k = (X^k)^{-1/2} \quad \forall k, \\ \leqslant n/\gamma & \text{if } P^k = (S^k)^{1/2} \quad \forall k. \end{cases}$$

Proof. For simplicity, let us drop the superscript k. Assume first that $P = (W_{\text{nt}})^{1/2}$. By (2.16), we have $\hat{X} = \hat{S}$, and hence $\hat{E} = \hat{F}$ due to (2.13). Thus it follows from (3.3) that $\kappa_\infty = 1$.

Consider now the case in which $P = X^{-1/2}$. By (2.11), we have $\hat{X} = I$ and $\hat{S} = X^{1/2}SX^{1/2}$, which in view of (2.13) imply that $\hat{F} = I$ and

$$\hat{E} = \tfrac{1}{2}\left[X^{1/2}SX^{1/2} \otimes I + I \otimes X^{1/2}SX^{1/2}\right].$$

By well-known properties of the Kronecker product (see Ch. 4 of [4]), the spectrum of $\hat{E}$ is $\{(\lambda_i + \lambda_j)/2:\ 1 \leqslant i,\ j \leqslant n\}$, where $\{\lambda_i:\ i = 1, \ldots, n\}$ is the spectrum of $X^{1/2}SX^{1/2}$. Since the latter matrix is contained in the interval $[\gamma\mu, n\mu]$ due to the fact that $(X^k, S^k, y^k) \in \mathcal{N}(\gamma)$, so is the former one, that is

$$\gamma\mu \leqslant \lambda_{\min}[\hat{E}] \leqslant \lambda_{\max}[\hat{E}] \leqslant n\mu.$$

Hence, we have

$$\mathrm{cond}(\hat{E}^{-1}\hat{F}) = \mathrm{cond}(\hat{E}^{-1}) = \mathrm{cond}(\hat{E}) \leqslant n/\gamma,$$

which, in view of Eq. (3.3), implies that $\kappa_\infty \leqslant n/\gamma$. The proof for the case $P = S^{1/2}$ is similar. □

As a consequence of Theorem 3.1 and Lemma 3.1, we immediately obtain the following result.

Theorem 3.2. *Algorithm-CC based on the NT direction, the HRVW/KSH/M direction and the KSH/M direction has iteration-complexity bounds equal to* $\mathcal{O}(nL)$, $\mathcal{O}(n^{3/2}L)$ *and* $\mathcal{O}(n^{3/2}L)$, *respectively.*

We note that the iteration-complexity bounds obtained in Theorem 3.2 for the HRVW/KSH/M and KSH/M directions were obtained earlier by Monteiro [12] within a more specialized context. We also note that, in the context of Algorithm-CC, Theorem 3.2 guarantees that the Nesterov–Todd direction achieves the best possible iteration-complexity bound of $\mathrm{O}(nL)$ that can provably be derived from our analysis. This is due to the fact that by Lemma 3.1 the condition number κ_∞ achieves its lowest possible value for the sequence $\{P^k\}$ corresponding to the NT direction.

4. Technical results and the proof of Theorem 3.1

In this section, we first develop a number of technical results and then use these results to prove Theorem 3.1.

We will use the following notation throughout this section:

$$X(\alpha) \equiv X + \alpha\Delta X, \qquad S(\alpha) \equiv S + \alpha\Delta S, \tag{4.1}$$

$$\mu(\alpha) \equiv \frac{X(\alpha)\cdot S(\alpha)}{n}, \qquad \mu \equiv \frac{X\cdot S}{n}. \tag{4.2}$$

The following simple identity introduced in [27] is useful for deriving bounds on the direction $(\Delta X, \Delta S, \Delta y)$.

Lemma 4.1. *Let* $u, v, r \in \mathbb{R}^p$ *and* $E, F \in \mathbb{R}^{p\times p}$ *satisfy* $Eu + Fv = r$. *If* $FE^{\mathrm{T}} \in \mathcal{S}^p_{++}$ *then*

$$\|(FE^{\mathrm{T}})^{-1/2}Eu\|^2 + \|(FE^{\mathrm{T}})^{-1/2}Fv\|^2 + 2u^{\mathrm{T}}v = \|(FE^{\mathrm{T}})^{-1/2}r\|^2. \tag{4.3}$$

Proof. Pre-multiply both sides of $Eu + Fv = r$ by $(FE^{\mathrm{T}})^{-1/2}$ and take 2-norm squared. □

Later we will apply Lemma 4.1 to the linear equation (2.12). With this goal in mind, we first need to derive conditions under which the matrices $\hat{E}$ and $\hat{F}$ of Eq. (2.13) satisfy the assumption $\hat{F}\hat{E}^{\mathrm{T}} \in \mathcal{S}^{n^2}_{++}$ of Lemma 4.1. In the following proposition, we show that a necessary and sufficient condition for $\hat{F}\hat{E}^{\mathrm{T}} \in \mathscr{S}^{n^2}_{++}$ is that $\hat{X}\hat{S} \in \mathscr{S}^n$, or equivalently $\hat{X}$ and $\hat{S}$ commute. This result is in essence the same as Proposition 3.1 of [24], which is formulated in terms of different matrices involving symmetrized Kronecker products. We give its (short) proof for completeness.

Proposition 4.1. *Let $X, S \in \mathscr{S}^n_{++}$, $\hat{X}$ and $\hat{S}$ be defined as in Eq. (2.11), and $\hat{E}$ and $\hat{F}$ as in Eq. (2.13). Then:*
(i) *$\hat{E}, \hat{F} \in \mathscr{S}^{n^2}_{++}$, and thus $\hat{F}\hat{E}^{\mathrm{T}} = \hat{F}\hat{E}$;*
(ii) *$\hat{F}\hat{E} \in \mathscr{S}^{n^2}$ if and only if $\hat{X}\hat{S} \in \mathscr{S}^n$;*
(iii) *$\hat{F}\hat{E} \in \mathscr{S}^{n^2}$ implies $\hat{F}\hat{E} \in \mathscr{S}^{n^2}_{++}$.*

Proof. (i) is obvious. For (ii), note that $\hat{F}\hat{E} \in \mathscr{S}^{n^2}$ if and only if $\hat{E}$ and $\hat{F}$ commute. By direct calculation,

$$4\left(\hat{F}\hat{E} - \hat{E}\hat{F}\right) = (\hat{X}\hat{S} - \hat{S}\hat{X}) \otimes I + I \otimes (\hat{X}\hat{S} - \hat{S}\hat{X}).$$

Therefore, $\hat{F}\hat{E} = \hat{E}\hat{F}$ if and only if $\hat{X}\hat{S} = \hat{S}\hat{X}$. Since $\hat{E}, \hat{F} \in \mathscr{S}^{n^2}_{++}$ and $\hat{F}\hat{E}$ is similar to $\hat{F}^{1/2}\hat{E}\hat{F}^{1/2}$, clearly the symmetry of $\hat{F}\hat{E}$ implies its positive definiteness. This proves (iii). □

In the sequel, we will denote the eigenvalues of the matrix XS as $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$. Observe that since the matrices XS, SX, $S^{1/2}XS^{1/2}$, $X^{1/2}SX^{1/2}$, $\hat{X}\hat{S}$, $\hat{S}\hat{X}$ are similar, they all have the same eigenvalues. In addition, we let Λ denote the diagonal matrix

$$\Lambda \equiv \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

Proposition 4.2. *For any $P \in \mathscr{P}(X, S)$, there exist an orthogonal matrix Q_{P} and diagonal matrices $\Lambda(\hat{X})$ and $\Lambda(\hat{S})$ such that:*
(i) *$\hat{X} \equiv PXP = Q_{\mathrm{P}}[\Lambda(\hat{X})]Q_{\mathrm{P}}^{\mathrm{T}}$;*
(ii) *$\hat{S} \equiv P^{-1}SP^{-1} = Q_{\mathrm{P}}[\Lambda(\hat{S})]Q_{\mathrm{P}}^{\mathrm{T}}$;*
(iii) *$\Lambda = \Lambda(\hat{X})\Lambda(\hat{S})$, and hence $\hat{X}\hat{S} = \hat{S}\hat{X} = Q_{\mathrm{P}}\Lambda Q_{\mathrm{P}}^{\mathrm{T}}$.*

Proof. The commutativity of $\hat{X}$ and $\hat{S}$ ensures that the two matrices share a common set of orthonormal eigenvectors, from which (i) and (ii) follow. Moreover, by (i) and (ii), we have $\hat{X}\hat{S} = Q_{\mathrm{P}}[\Lambda(\hat{X})\Lambda(\hat{S})]Q_{\mathrm{P}}^{\mathrm{T}}$. Since the spectrum of XS and $\hat{X}\hat{S}$ is the same, by permuting the columns of Q_{P} if necessary, we have $\Lambda = \Lambda(\hat{X})\Lambda(\hat{S})$, that is (iii) holds. □

The result below follows from applying Lemma 4.1. to Eq. (2.12).

Lemma 4.2. *Let $P \in \mathscr{P}(X, S)$ and $G \equiv \hat{E}^{-1}\hat{F}$. Then,*

$$\|G^{-1/2} \text{ vec } \Delta\hat{X}\|^2 + \|G^{1/2} \text{ vec } \Delta\hat{S}\|^2 + 2\Delta\hat{X} \cdot \Delta\hat{S} = \sum_{i=1}^{n} \frac{(\sigma\mu - \lambda_i)^2}{\lambda_i}.$$

Moreover, if $\lambda_{\min}[XS] \geqslant \gamma\mu$ *for some* $\gamma \in (0,1)$, *then*

$$\sum_{i=1}^{n} \frac{(\sigma\mu - \lambda_i)^2}{\lambda_i} \leqslant (1 - 2\sigma + \sigma^2/\gamma)n\mu. \tag{4.4}$$

Proof. Applying Lemma 4.1 to relation (2.12) we obtain

$$\|(\hat{F}\hat{E})^{-1/2}\hat{E} \text{ vec } \Delta\hat{X}\|^2 + \|(\hat{F}\hat{E})^{-1/2}\hat{F} \text{ vec } \Delta\hat{S}\|^2 + 2\Delta\hat{X}\cdot\Delta\hat{S} = \|(\hat{F}\hat{E})^{-1/2}\text{vec}(\sigma\mu I - H(\hat{X}\hat{S}))\|^2.$$

The commutativity of $\hat{E}$ and $\hat{F}$ implies that

$$(\hat{F}\hat{E})^{-1/2}\hat{E} = (\hat{E}^{-1}\hat{F})^{-1/2} = G^{-1/2}, \qquad (\hat{F}\hat{E})^{-1/2}\hat{F} = (\hat{E}^{-1}\hat{F})^{1/2} = G^{1/2}.$$

Hence, for the proof of the first statement it remains to show that

$$\|(\hat{F}\hat{E})^{-1/2} \text{ vec}(\sigma\mu I - \hat{X}\hat{S})\|^2 = \sum_{i=1}^{n} \frac{(\sigma\mu - \lambda_i)^2}{\lambda_i}. \tag{4.5}$$

Using Eq. (2.13) and Proposition 4.2(ii), we find the spectral decomposition of $\hat{E}$ to be

$$\hat{E} = \tfrac{1}{2}(\hat{S}\otimes I + I \otimes \hat{S}) = \tfrac{1}{2}\hat{Q}[\Lambda(\hat{S})\otimes I + I\otimes\Lambda(\hat{S})]\hat{Q}^{\mathrm{T}},$$

where $\hat{Q} \equiv Q_{\mathrm{P}} \otimes Q_{\mathrm{P}}$ is an orthogonal matrix of dimension n^2. Similarly, by Eq. (2.13) and Proposition 4.2(i), we have

$$\hat{F} = \tfrac{1}{2}(\hat{X}\otimes I + I \otimes \hat{X}) = \tfrac{1}{2}\hat{Q}[\Lambda(\hat{X})\otimes I + I\otimes\Lambda(\hat{X})]\hat{Q}^{\mathrm{T}}.$$

Therefore, using Proposition 4.2(iii) we obtain

$$(\hat{F}\hat{E})^{-1} = 4\hat{Q}[\Lambda\otimes I + I\otimes\Lambda + \Lambda(\hat{X})\otimes\Lambda(\hat{S}) + \Lambda(\hat{S})\otimes\Lambda(\hat{X})]^{-1}\hat{Q}^{\mathrm{T}},$$

where the matrix in the middle is diagonal with the property that its $((i-1)n+i)$th diagonal element is equal to $1/(4\lambda_i)$ for $i = 1,\ldots,n$. On the other hand, observe that

$$\text{vec}(\sigma\mu I - \hat{X}\hat{S}) = \hat{Q} \text{ vec}(\sigma\mu I - \Lambda),$$

where $\text{vec}(\sigma\mu I - \Lambda)$ is an n^2-vector having at most n nonzero components, namely: its $((i-1)n+i)$th component is equal to $\sigma\mu - \lambda_i$ for $i = 1,\ldots,n$. The above two relations and a straightforward verification finally yield

$$\text{vec}(\sigma\mu I - \hat{X}\hat{S})^{\mathrm{T}}(\hat{E}\hat{F})^{-1} \text{ vec}(\sigma\mu I - \hat{X}\hat{S}) = \sum_{i=1}^{n}\frac{(\sigma\mu - \lambda_i)^2}{\lambda_i},$$

which proves Eq. (4.5) and hence the first part of the lemma.

To prove (4.4), we use the fact that $n\mu = \text{tr } XS = \sum_{i=1}^{n}\lambda_i$ to obtain

$$\sum_{i=1}^{n}\frac{(\sigma\mu - \lambda_i)^2}{\lambda_i} \leqslant \sigma^2\mu^2\frac{n}{\lambda_1} - 2\sigma n\mu + \sum_{i=1}^{n}\lambda_i \leqslant \frac{\sigma^2 n\mu}{\gamma} - 2\sigma n\mu + n\mu,$$

which completes the proof of the lemma. □

Lemma 4.3. *Let* $(X,S,y) \in \mathcal{S}^n_{++} \times \mathcal{S}^n_{++} \times \mathbb{R}^m$ *and* $(\Delta X, \Delta S, \Delta y)$ *satisfy Eq. (2.7). Then*

$$H_P(X(\alpha)S(\alpha)) = (1-\alpha)H_P(XS) + \alpha\sigma\mu I + \alpha^2 H_P(\Delta X\,\Delta S), \tag{4.6}$$

$$\mu(\alpha) = (1-\alpha+\alpha\sigma)\mu + \alpha^2 \Delta X \cdot \Delta S/n, \tag{4.7}$$

Proof: By Eq. (4.1), we have

$$X(\alpha)S(\alpha) = (X+\alpha\Delta X)(S+\alpha\Delta S) = XS + \alpha(X\,\Delta S + \Delta X\,S) + \alpha^2\Delta X\,\Delta S.$$

This expression, together with the linearity of $H_P(\cdot)$ and Eq. (2.7), implies that

$$\begin{aligned} H_P(X(\alpha)S(\alpha)) &= H_P(XS) + \alpha H_P(X\,\Delta S + \Delta X\,S) + \alpha^2 H_P(\Delta X\,\Delta S) \\ &= H_P(XS) + \alpha[\sigma\mu I - H_P(XS)] + \alpha^2 H_P(\Delta X\,\Delta S), \end{aligned}$$

and hence Eq. (4.6) holds. Using Eq. (4.6) and the identity $\operatorname{tr} H_P(M) = \operatorname{tr} M$, we obtain

$$\begin{aligned} X(\alpha)\cdot S(\alpha) &= \operatorname{tr}[(1-\alpha)H_P(XS) + \alpha\sigma\mu I + \alpha^2 H_P(\Delta X\,\Delta S)] \\ &= (1-\alpha)\operatorname{tr} H_P(XS) + \alpha\sigma\mu n + \alpha^2 \operatorname{tr} H_P(\Delta X\,\Delta S) \\ &= (1-\alpha)X\cdot S + \alpha\sigma\mu n + \alpha^2\Delta X\cdot\Delta S. \end{aligned}$$

Dividing this expression by n and noting Eq. (4.2), we obtain Eq. (4.7). □

Lemma 4.4. *Suppose that* $(X,S,y) \in \mathcal{S}^n_{++} \times \mathcal{S}^n_{++} \times \mathbb{R}^m$, $P \in \mathcal{S}^n_{++}$ *and* $Q \in \mathcal{P}(X,S)$. *Then,*

$$\lambda_{\min}[H_P(XS)] \leqslant \lambda_{\min}[XS] = \lambda_{\min}[H_Q(XS)]. \tag{4.8}$$

Proof. Since $Q \in \mathcal{P}(X,S)$, we have $H_Q(XS) = QXSQ^{-1}$. By similarity,

$$\lambda_{\min}[XS] = \lambda_{\min}[QXSQ^{-1}] = \lambda_{\min}[H_Q(XS)].$$

Moreover,

$$\lambda_{\min}[XS] = \lambda_{\min}[PXSP^{-1}] \geqslant \lambda_{\min}[H(PXSP^{-1})] = \lambda_{\min}[H_P(XS)],$$

where the inequality follows from the fact that the real part of the spectrum of a real matrix is contained between the largest and the smallest eigenvalues of its Hermitian part (see p. 187 of [4], for example). We have thus shown that (4.8) holds. □

Lemma 4.5. *Suppose that* $(X,S,y) \in \mathcal{S}^n_{++} \times \mathcal{S}^n_{++} \times \mathbb{R}^m$, $(\Delta X, \Delta S, \Delta y) \in \mathcal{S}^n \times \mathcal{S}^n \times \mathbb{R}^m$ and $P \in \mathcal{P}(X,S)$ *satisfy Eq. (2.7) and that* $\lambda_{\min}[XS] \geqslant \gamma\mu$ *for some constant* $\gamma \in (0,1)$. *Then*

$$\lambda_{\min}[X(\alpha)S(\alpha)] \geqslant \gamma\mu(\alpha)$$

for every $\alpha \in [0,\tilde{\alpha}]$, *where*

$$\tilde{\alpha} \equiv \min\left(1, \frac{\sigma(1-\gamma)\mu}{\omega}\right) \quad \textit{and} \quad \omega \equiv \|H_P(\Delta X\,\Delta S)\| + \frac{\gamma}{n}|\Delta X\cdot\Delta S|. \tag{4.9}$$

Proof. Using the assumptions that $\lambda_{\min}[XS] \geqslant \gamma\mu$ and $P \in \mathcal{P}(X,S)$ together with Lemma 4.4, we obtain

$$\lambda_{\min}[H_P(XS) - \gamma\mu I] = \lambda_{\min}[H_P(XS)] - \gamma\mu = \lambda_{\min}[XS] - \gamma\mu \geqslant 0. \tag{4.10}$$

Combining Eqs. (4.6) and (4.7) and simplifying, we obtain

$$H_P(X(\alpha)S(\alpha)) - \gamma\mu(\alpha)I = (1-\alpha)(H_P(XS) - \gamma\mu I) + \alpha\sigma(1-\gamma)\mu I + \alpha^2\Omega, \tag{4.11}$$

where

$$\Omega \equiv H_P(\Delta X \Delta S) - \frac{\gamma}{n}(\Delta X \cdot \Delta S)I.$$

It follows from Lemma 4.4, (4.9)–(4.11) and the fact that $\lambda_{\min}[\cdot]$ is a homogeneous concave function on the space of symmetric matrices that

$$\begin{aligned}\lambda_{\min}[X(\alpha)S(\alpha)] - \gamma\mu(\alpha) &\geqslant \lambda_{\min}[H_P(X(\alpha)S(\alpha))] - \gamma\mu(\alpha)\\ &= \lambda_{\min}[H_P(X(\alpha)S(\alpha)) - \gamma\mu(\alpha)I]\\ &\geqslant (1-\alpha)\lambda_{\min}[H_P(XS) - \gamma\mu I] + \alpha\sigma(1-\gamma)\mu + \alpha^2\lambda_{\min}[\Omega]\\ &\geqslant \alpha\sigma(1-\gamma)\mu - \alpha^2\|\Omega\|\\ &\geqslant \alpha[\sigma(1-\gamma)\mu - \alpha\omega] \geqslant 0\end{aligned}$$

for every $\alpha \in [0, \tilde{\alpha}]$. □

Lemma 4.6. *For any $u, v \in \mathbb{R}^n$ and $G \in \mathcal{S}^n_{++}$, we have*

$$\|u\|\|v\| \leqslant \frac{\sqrt{\operatorname{cond}(G)}}{2}\left(\|G^{-1/2}u\|^2 + \|G^{1/2}v\|^2\right). \tag{4.12}$$

Proof. We have:

$$\|u\|^2 \leqslant \frac{u^T G^{-1} u}{\lambda_{\min}(G^{-1})} = \lambda_{\max}(G)\|G^{-1/2}u\|^2,$$

$$\|v\|^2 \leqslant \frac{v^T G v}{\lambda_{\min}(G)} = \frac{\|G^{1/2}v\|^2}{\lambda_{\min}(G)}.$$

Hence

$$\begin{aligned}\|u\|\|v\| &\leqslant \sqrt{\operatorname{cond}(G)}\|G^{-1/2}u\|\|G^{1/2}v\|\\ &\leqslant \frac{\sqrt{\operatorname{cond}(G)}}{2}\left(\|G^{-1/2}u\|^2 + \|G^{1/2}v\|^2\right).\end{aligned}$$ □

Lemma 4.7. *Let a point $(X,S,y) \in \mathcal{N}(\gamma)$ and a scaling matrix $P \in \mathcal{P}(X,S)$ be given, and define $G \equiv \hat{E}^{-1}\hat{F}$ where $\hat{E}$ and $\hat{F}$ are given by Eq. (2.13). Then, the solution $(\Delta X, \Delta S, \Delta y)$ of Eqs. (2.7)–(2.9) satisfies*

$$\|H_P(\Delta X \Delta S)\|_F \leqslant \frac{\sqrt{\operatorname{cond}(G)}}{2}(1 - 2\sigma + \sigma^2/\gamma)n\mu. \tag{4.13}$$

Proof. Let $\Delta\hat{X}$ and $\Delta\hat{S}$ be defined as in Eq. (2.11). Using relations (2.8) and (2.9) and the fact that (X,S,y) is a feasible point, we easily see that

$$\Delta X \cdot \Delta S = \Delta\hat{X} \cdot \Delta\hat{S} = 0. \tag{4.14}$$

We have

$$\begin{aligned}\|H_P(\Delta X\,\Delta S)\|_F &= \|H(\Delta\hat{X}\,\Delta\hat{S})\|_F \leqslant \|\Delta\hat{X}\,\Delta\hat{S}\|_F \leqslant \|\Delta\hat{X}\|_F\|\Delta\hat{S}\|_F \\ &= \|\mathrm{vec}\,\Delta\hat{X}\|\,\|\mathrm{vec}\,\Delta\hat{S}\|.\end{aligned}$$

Using Lemmas 4.6, 4.2 and relation (4.14), we obtain

$$\begin{aligned}\|\mathrm{vec}\ \Delta\hat{X}\|\|\mathrm{vec}\ \Delta\hat{S}\| &\leqslant \frac{\sqrt{\mathrm{cond}(G)}}{2}\left(\|G^{-1/2}\mathrm{vec}\ \Delta\hat{X}\|^2+\|G^{1/2}\mathrm{vec}\ \Delta\hat{S}\|^2\right) \\ &= \frac{\sqrt{\mathrm{cond}(G)}}{2}\left(\sum_{i=1}^{n}\frac{(\sigma\mu-\lambda_i)^2}{\lambda_i} - 2\,\Delta\hat{X}\cdot\Delta\hat{S}\right) \\ &\leqslant \frac{\sqrt{\mathrm{cond}(G)}}{2}(1-2\sigma+\sigma^2/\gamma)n\mu.\end{aligned}$$

The result now follows by combining the above two inequalities. □

Finally, we are in a position to prove Theorem 3.1.

Proof of Theorem 3.1. Due to Eq. (4.14), relation (4.7) implies that $\mu_{k+1} = (1-(1-\sigma)\alpha_k)\mu_k$. By Lemma 4.5 and Eq. (4.14), it follows that $(X^k(\alpha), S^k(\alpha), y^k(\alpha)) \in \mathcal{N}(\gamma)$ for every $\alpha \in [0, \tilde{\alpha}_k]$, where

$$\tilde{\alpha}_k \equiv \min\left(1, \frac{\sigma(1-\gamma)\mu}{\|H_P(\Delta X^k\,\Delta S^k)\|}\right).$$

Hence, by the definition of α_k, Lemma 4.7 and relation (3.3), we have

$$\alpha_k \geqslant \min(1, \tilde{\alpha}_k) \geqslant \min\left(1, \frac{2\sigma(1-\gamma)}{1-2\sigma+\sigma^2/\gamma}\,\frac{1}{\sqrt{\kappa_\infty}\;n}\right),$$

which proves the first part of the theorem. The second part about the iteration-complexity bound follows from a now standard argument. □

5. Other results

Besides the three special cases considered in Lemma 3.1 and Theorem 3.2, there are many other choices of $\{P^k\}$ that lead to polynomial algorithms. In this section, we derive a representation for the commutative class $\mathcal{P}(X,S)$ and establish an alternative iteration-complexity bound for Algorithm-CC based on this representation.

Without loss of generality, we assume that there are p ($p \leqslant n$) distinct eigenvalues in the spectrum of XS and we group them into p groups. Hence,

$$\Lambda = \mathrm{diag}(\lambda^{(1)}I^{(1)}, \ldots, \lambda^{(p)}I^{(p)}), \tag{5.1}$$

where $\lambda^{(i)}$ is the ith distinct eigenvalue, $I^{(i)}$ the identity matrix of dimension $n^{(i)}$, $n^{(i)} \geqslant 1$ the multiplicity of $\lambda^{(i)}$, and $\sum_{i=1}^{p} n^{(i)} = n$.

Consider the following class of block diagonal matrices associated with the pair (X, S):

$$\mathscr{T}(X,S) \equiv \{\mathrm{diag}(T^{(1)},\dots,T^{(p)}):\ T^{(i)} \in \mathscr{S}^{n^{(i)}}_{++},\ for\ i=1,\dots,p\}.$$

The following result gives a representation for the class of matrices $W \in \mathscr{S}^n_{++}$ such that $W = P^2$ for some $P \in \mathscr{P}(X,S)$, namely

$$\mathscr{W}(X,S) \equiv \{W \in \mathscr{S}^n_{++}:\ WXS = XSW\},$$

in terms of the set $\mathscr{T}(X,S)$.

Theorem 5.1. *Let $(X,S) \in \mathscr{S}^n_{++} \times \mathscr{S}^n_{++}$ and a fixed $P \in \mathscr{P}(X,S)$ be given. Then,*

$$\mathscr{W}(X,S) = \{PQ_\mathrm{P}TQ_\mathrm{P}^\mathrm{T}P:\ T \in \mathscr{T}(X,S)\}, \tag{5.2}$$

where Q_P is as in Proposition 4.2. Moreover, $W(X,S)$ is a convex cone of dimension at least n.

Proof. Let $W \in \mathscr{S}^n_{++}$ be given. The equation $WXS = SXW$ is equivalent to

$$(P^{-1}WP^{-1})(\hat{X}\hat{S}) = (\hat{S}\hat{X})(P^{-1}WP^{-1}),$$

which in turn, by Proposition 4.2.(iii), is equivalent to $T\Lambda = \Lambda T$, where $T \equiv Q_\mathrm{P}^\mathrm{T}P^{-1}WP^{-1}Q_\mathrm{P} \succ 0$. Given the structure of Λ in Eq. (5.1), one can easily see that the last relation is equivalent to $T \in \mathscr{T}(X,S)$. We have thus proved that Eq. (5.2) holds. Obviously, the set $\{W \in \mathscr{S}^n_{++}:\ WXS = SXW\}$ is a convex cone. Since, for fixed $P \in \mathscr{P}(X,S)$, $PQ_\mathrm{P}TQ_\mathrm{P}^\mathrm{T}P$ is a one-to-one and linear function of T, it follows from the characterization of T that this set has dimension at least n (equal to n when $p = n$). □

We now give a complexity result in terms of the representation (5.2) in Theorem 5.1. If we fix P at $(S^k)^{1/2}$ in Theorem 5.1, then by the representation (5.2) any $P^k \in \mathscr{P}(X^k,S^k)$ can be written as

$$P^k = \left[(S^k)^{1/2}(Q^k)T^k(Q^k)^\mathrm{T}(S^k)^{1/2}\right]^{1/2}, \tag{5.3}$$

where Q^k is the matrix Q_P in Theorem 5.1 corresponding to $P = (S^k)^{1/2}$ and $T^k \in \mathscr{T}(X^k,S^k)$. We stress that similar results corresponding to other choices of P in the representations can also be similarly established. Recall that κ_∞ is defined in Eq. (3.3).

Theorem 5.2. *For any sequences $\{(X^k,S^k,y^k)\} \subset \mathscr{N}(\gamma)$ and $\{P^k:\ P^k \in \mathscr{P}(X^k,S^k)\}$, there holds*

$$\kappa_\infty \leqslant \frac{n}{\gamma}\left(\sup_k \mathrm{cond}(T^k)\right)^2,$$

where $T^k \in \mathscr{T}(X^k,S^k)$ is associated with P^k through Eq. (5.3).

Proof. Again let us drop the superscript k. Recall that

$$G \equiv \hat{E}^{-1}\hat{F} = (\hat{S}\otimes I + I\otimes\hat{S})^{-1}(\hat{X}\otimes I + I\otimes\hat{X}).$$

It follows from commutativity of $\hat{X}$ and $\hat{S}$ that the lth eigenvalue of G for $l=(i-1)n+j$, $1\leqslant i,j\leqslant n$, is

$$\lambda_l[G] = \frac{\lambda_i[\hat{X}]+\lambda_j[\hat{X}]}{\lambda_i[\hat{S}]+\lambda_j[\hat{S}]}. \tag{5.4}$$

Observe that $\hat{X}=PXP$ is similar to $X^{1/2}WX^{1/2}$ and $\hat{S}=P^{-1}SP^{-1}$ to $S^{1/2}W^{-1}S^{1/2}$, where $W=P^2$. Substituting these relations and $W=S^{1/2}QTQ^{\mathrm{T}}S^{1/2}$, we obtain:

$$\lambda_i[\hat{X}] = \lambda_i[X^{1/2}WX^{1/2}] = \lambda_i[X^{1/2}S^{1/2}QTQ^{\mathrm{T}}S^{1/2}X^{1/2}],$$

$$\lambda_i[\hat{S}] = \lambda_i[S^{1/2}W^{-1}S^{1/2}] = \lambda_i[T^{-1}].$$

Observe that for any $A\in\mathscr{S}^n$, $B\in\mathbb{R}^{n\times n}$ and $v\in\mathbb{R}^n$

$$\frac{v^{\mathrm{T}}B^{\mathrm{T}}ABv}{v^{\mathrm{T}}v} = \frac{(Bv)^{\mathrm{T}}A(Bv)}{(Bv)^{\mathrm{T}}(Bv)}\frac{v^{\mathrm{T}}B^{\mathrm{T}}Bv}{v^{\mathrm{T}}v},$$

which implies that

$$\lambda_{\min}[A]\lambda_{\min}[B^{\mathrm{T}}B] \leqslant \lambda_i[B^{\mathrm{T}}AB] \leqslant \lambda_{\max}[A]\lambda_{\max}[B^{\mathrm{T}}B].$$

Using these inequalities, we have

$$\lambda_{\min}[X^{1/2}SX^{1/2}]\lambda_{\min}[T] \leqslant \lambda_i[\hat{X}] \leqslant \lambda_{\max}[X^{1/2}SX^{1/2}]\lambda_{\max}[T].$$

Consequently, from Eq. (5.4) we obtain

$$\lambda_{\max}[G] \leqslant \frac{\lambda_{\max}[X^{1/2}SX^{1/2}]\lambda_{\max}[T]}{\lambda_{\min}[T^{-1}]} = \lambda_{\max}[X^{1/2}SX^{1/2}]\lambda^2_{\max}[T].$$

Similarly,

$$\lambda_{\min}[G] \geqslant \frac{\lambda_{\min}[X^{1/2}SX^{1/2}]\lambda_{\min}[T]}{\lambda_{\max}[T^{-1}]} = \lambda_{\min}[X^{1/2}SX^{1/2}]\lambda^2_{\min}[T].$$

Therefore,

$$\mathrm{cond}(G) \leqslant \mathrm{cond}(S^{1/2}XS^{1/2})\mathrm{cond}(T)^2,$$

where we used the fact that $S^{1/2}XS^{1/2}$ is similar to $X^{1/2}SX^{1/2}$. Finally, it is known (see the proof of Theorem 3.1) that $(X,S,y)\in\mathscr{N}(\gamma)$ implies that $\mathrm{cond}(S^{1/2}XS^{1/2}) \leqslant n/\gamma$. □

This theorem guarantees that Algorithm-CC will have a polynomial iteration-complexity bound if the matrices $\{T^k\}$ are chosen so that their spectral condition numbers are uniformly bounded above by a polynomial of n. In particular, it will have an $\mathrm{O}(n^{1.5}L)$-iteration complexity bound if $\mathrm{cond}(T^k)$ is bounded above by a constant independent of n. Furthermore, for any number τ, the choice $T^k=(\Lambda^k)^\tau$ would give an $\mathrm{O}(n^{1.5+|\tau|}L)$-iteration polynomial algorithm since $\mathrm{cond}(T^k)\leqslant(n/\gamma)^{|\tau|}$.

6. Concluding remarks

After the release of the first version of the current paper, several new developments have occurred. Sturm and Zhang [23] independently provided another proof for the $\mathcal{O}(nL)$-iteration convergence of the long-step path-following algorithm using the Nesterov–Todd direction. Monteiro [13] proved that short-step primal-dual path-following algorithms based on any direction of the Monteiro–Zhang family, including the AHO direction, have iteration-complexity bound equal to $\mathcal{O}(\sqrt{n}L)$. Monteiro's result for the AHO direction is of particular interest since the AHO direction corresponds to a fixed scaling matrix (the identity matrix) and has good local convergence properties and interesting numerical behavior (see [2,7]).

It is worth noting that for long-step path-following algorithms, namely those that use the "large neighborhood" defined in Eq. (3.1), no polynomial complexity result has been obtained so far for a search direction outside the commutative class studied in this paper.

Acknowledgements

We would like to thank the anonymous referees and associate editor for their constructive comments that have helped improve the presentation of the paper.

References

[1] F. Alizadeh, Combinatorial Optimization with Interior Point Methods and Semi-Definite Matrices, Ph.D. Thesis, Computer Science Department, University of Minnesota, Minneapolis, 1991.

[2] F. Alizadeh, J.A. Haeberly, M. Overton, Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results, Technical report, Computer Science Department, New York University, New York, 1996.

[3] C. Helmberg, F. Rendl, R.J. Vanderbei, H. Wolkowicz, An interior-point method for semidefinite programming, SIAM Journal on Optimization 6 (1996) 342–361.

[4] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, New York, 1991.

[5] F. Jarre, An interior-point method for minimizing the maximum eigenvalue of a linear combination of matrices, SIAM Journal on Control and Optimization 31 (1993) 1360–1377.

[6] M. Kojima, S. Mizuno, A. Yoshise, A primal-dual interior point method for linear programming, in: Nimrod Megiddo (Ed.), Progress in Mathematical Programming, Interior-Point and Related Methods, Springer, New York, 1989, pp. 29–47.

[7] M. Kojima, M. Shida, S. Shindoh, Local convergence of predictor-corrector infeasible-interior-point algorithms for SDPs and SDLCPs, Mathematical Programming 80 (1998) 129–160.

[8] M. Kojima, M. Shida, S. Shindoh, A note on the Nesterov-Todd and the Kojima-Shindoh-Hara search directions in semidefinite programming, Research Report #B-313, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, 1996.

[9] M. Kojima, S. Shindoh, S. Hara, Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices, SIAM Journal on Optimization 7 (1997) 86–125.

[10] C-J. Lin, R. Saigal, A predictor-corrector method for semi-definite programming, Working paper, Department of Industrial and Operations Engineering, The University of Michigan, Ann Arbor, Michigan 48109-2177, 1995.

[11] Z-Q. Luo, J.F. Sturm, S. Zhang, Superlinear convergence of a symmetric primal-dual path-following algorithms for semidefinite programming, Working paper, 1996 (to appear in SIAM Journal on Optimization).

[12] R.D.C. Monteiro, Primal-dual path-following algorithms for semidefinite programming, SIAM Journal on Optimization 7 (1997) 663–678.

[13] R.D.C. Monteiro, Polynomial convergence of primal-dual algorithms for semidefinite programming based on Monteiro and Zhang family of directions, Manuscript, School of ISyE, Georgia Institute of Technology, Atlanta, GA 30332, USA, 1996.

[14] Y.E. Nesterov, A.S. Nemirovskii, A general approach to the design of optimal methods for smooth convex functions minimization, Ekonomika i Matem, Metody 24 (1988) 509–517 (In Russian; English transl. Matekon: Translations of Russian and East European Math. Economics.).

[15] Y.E. Nesterov, A.S. Nemirovskii, Self-concordant functions and polynomial time methods in convex programming, Preprint, Central Economic & Mathematical Institute, USSR Acadamy of Science, Moscow, USSR, 1989.

[16] Y.E. Nesterov, A.S. Nemirovskii, Optimization over positive semidefinite matrices: Mathematical background and user's manual, Technical report, Central Economic & Mathematical Institute, USSR Academy of Science Moscow, USSR, 1990.

[17] Y.E. Nesterov, A.S. Nemirovskii, Interior Point Methods in Convex Programming: Theory and Applications, Society for Industrial and Applied Mathematics, Philadelphia, 1994.

[18] Y.E. Nesterov, M.J. Todd, Self-scaled barriers and interior-point methods for convex programming, Technical Report 1091, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York, 14853-3801, 1995.

[19] Y.E. Nesterov, M.J. Todd, Primal-dual interior-point methods for self-scaled cones, Mathematics of Operations Research 22 (1997) 1–42.

[20] F.A. Potra, R. Sheng, A superlinearly convergent primal-dual infeasible-interior-point algorithm for semidefinite programming, Reports on Computational Mathematics, No. 78, Department of Mathematics, The University of Iowa, 1995.

[21] M. Shida, S. Shindoh, M. Kojima, Existence of search directions in interior-point algorithms for the SDP and monotone SDLCP, Technical Report B-310, Department of Mathematical and Computer Science, Tokyo Institute of Technology, 1996 (to appear in SIAM Journal on Optimization).

[22] J.F. Sturm, S. Zhang, Symmetric primal-dual path-following algorithms for semidefinite programming, Report 9554/A, Ecometric Institute, Erasmus University Rotterdam, The Netherlands, 1995.

[23] J.F. Sturm, S. Zhang, On the long-step path-following method for semidefinite programming, Report 9638/A, Ecometric Institute, Erasmus University Rotterdam, The Netherlands, 1996.

[24] M.J. Todd, K.C. Toh, R.H. Tütüncü, On the Nesterov-Todd direction in semidefinite programming, Technical Report, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853, USA, 1996.

[25] L. Vandenberghe, S. Boyd, A primal-dual potential reduction method for problems involving matrix inequalities, Mathematical Programming 69 (1995) 205–236.

[26] Y. Ye, A class of projective transformations for linear programming, SIAM Journal on Computing 19 (1990) 457–466.

[27] Y. Zhang, On extending some primal-dual interior-Point algorithms from linear programming to semidefinite programming, Technical Report, 1995 (to appear in SIAM Journal on Optimization).