

# A new class of large neighborhood path-following interior point algorithms for semidefinite optimization with $O(\sqrt{n} \log \frac{\text{Tr}(X^0 S^0)}{\epsilon})$ iteration complexity

Yang Li\*      Tamás Terlaky†

July 27, 2009

## Abstract

In this paper, we extend the Ai-Zhang direction to the class of semidefinite optimization problems. We define a new wide neighborhood  $\mathcal{N}(\tau_1, \tau_2, \eta)$  and, as usual but with a small change, we make use of the scaled Newton equations for symmetric search directions. After defining the “positive part” and the “negative part” of a symmetric matrix, we recommend to solve the Newton equation with its right hand side replaced first by its positive part and then by its negative part, respectively. In such a way, we obtain a decomposition of the classical Newton direction and use different step lengths for each of them.

Starting with a feasible point  $(X^0, y^0, S^0)$  in  $\mathcal{N}(\tau_1, \tau_2, \eta)$ , the algorithm terminates in at most  $O(\eta\sqrt{\kappa_\infty n} \log \frac{\text{Tr}(X^0 S^0)}{\epsilon})$  iterations, where  $\kappa_\infty$  is a parameter associated with the scaling matrix  $P$  and  $\epsilon$  is the required precision. To our best knowledge, when the parameter  $\eta$  is a constant, this is the first large neighborhood path-following Interior Point Method (IPM) with the same complexity as small neighborhood path-following IPMs for semidefinite optimization that use the Nesterov-Todd direction. In the case when  $\eta$  is chosen to be in the order of  $\sqrt{n}$ , our result coincides with the results for the classical large neighborhood IPMs. Some preliminary numerical results are provided as well.

**Keywords:** interior point methods, large neighborhood, path-following algorithm, semidefinite optimization.

---

\*Department of Mathematics and Statistics, McMaster University, Hamilton, ON, L8S 4K1, Canada (liy99@mcmaster.ca).

†Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA 18015-1582, USA (terlaky@lehigh.edu).

# 1 Introduction

Semidefinite optimization (SDO) problems yield a generalization of linear optimization (LO) problems. Since Alizadeh [3] explored various applications of SDO in combinatorial optimization, SDO has been applied in many areas, including control theory, probability and signal processing [31].

Due to the success of Interior Point Methods (IPMs) in solving LO, most IPM variants were extended to SDO. The first IPMs for SDO were independently developed by Alizadeh [3] and Nesterov and Nemirovskii [21]. Alizadeh [3] applied Ye's potential reduction idea to SDO and showed how variants of dual IPMs could be extended to SDO. Almost at the same time, in their milestone book [21], Nesterov and Nemirovskii proved that IPMs are able to solve general conic optimization problems, in particular SDO problems, in polynomial time.

The difficulty to extend primal-dual path-following IPMs from LO to SDO lies in acquiring a symmetric search direction. The Newton method applied to the central path equation  $XS = \tau\mu I$  leads to the linear system

$$X\Delta S + \Delta XS = \tau\mu I - XS, \quad (1.1)$$

which generally results in non-symmetric search directions. Over the years, people suggested many strategies to deal with this problem. Alizadeh, Haeberly and Overton (AHO) [5] proposed to symmetrize both sides of (1.1). Alternatively, a similarity transformation  $P(\cdot)P^{-1}$  could be applied to both sides of (1.1). This strategy was first investigated by Monteiro [17] for  $P = X^{-1/2}$  and  $P = S^{1/2}$ . It turned out that the resulting directions by this approach could be seen as two special cases of the class of directions introduced earlier by Kojima, Shindoh and Hara [15]. At the same time, another motivation led Helmberg, Rendl, Vanderbei and Wolkowicz [11] to the direction given by  $P = S^{1/2}$ . The search directions given by  $P = X^{-1/2}$  and  $P = S^{1/2}$  are usually referred to as the H.K.M directions. Nesterov and Todd [22, 23] introduced the so-called NT direction in their attempt to generalize primal-dual IPMs beyond SDO. In [35], based on Monteiro's idea, Zhang generalized all these aforementioned approaches to a unified scheme parameterized by a nonsingular scaling matrix  $P$ . This family of search directions is referred to as the Monterio-Zhang (MZ) family of search directions.

As in the case of LO, there is an intriguing fact about IPMs for SDO. Although the theoretical complexity is worse, large neighborhood algorithms perform better in practice than small neighborhood algorithms. Many efforts were spent to bridge this gap. In [25], Peng, Roos and Terlaky established a new paradigm based on the class of the so-called self-regular functions. Under their new paradigm, large neighborhood IPMs can come arbitrarily close to the best known iteration bounds of small neighborhood IPMs. Later, based on Ai's original paper [1], a result of interest was given by Ai and Zhang [2] for linear complementarity problems (LCP). Their algorithm decomposes the classical Newton direction into two orthogonal ones and proceeds in a wide neighborhood. It is proved that their algorithm stops after at most  $O(\sqrt{n}L)$  iterations, where  $n$  is the number of variables and  $L$  is the input data length. This result yields the first large neighborhood path-following algorithm having the same theoretical complexity as a small neighborhood path-following algorithm for monotone LCPs, which include LO as a special case.

In this paper, we extend the Ai-Zhang scheme to SDO. We first define a new neighborhood  $\mathcal{N}(\tau_1, \tau_2, \eta)$ , where  $0 < \tau_2 < \tau_1 < 1$  and  $\eta \geq 1$  are given parameters. This new

neighborhood is proved to be a wide neighborhood itself. Not surprisingly, the neighborhood defined by Ai and Zhang [2] is a simple case of our wide neighborhood for LO. Another important ingredient of our algorithm is the decomposition of the classical Newton direction into two separate ones. To make this point clear, let us only consider LO, where (1.1) becomes

$$x\Delta s + s\Delta x = \tau\mu e - xs. \quad (1.2)$$

Moving from  $(x, s)$  to  $(x + \alpha\Delta x, s + \alpha\Delta s)$ , at each iteration we expect the duality gap to decrease by

$$x^T s - (x + \alpha\Delta x)^T (s + \alpha\Delta s) = -\alpha(x^T \Delta s + s^T \Delta x) = -\alpha e^T (\tau\mu e - xs) = \alpha(1 - \tau)x^T s > 0,$$

where the relation  $\mu = \frac{x^T s}{n}$  is used. Further, we may also write

$$\begin{aligned} -\alpha(x^T \Delta s + s^T \Delta x) &= -\alpha \sum \min([\tau\mu e - xs]_i, 0) - \alpha \sum \max([\tau\mu e - xs]_i, 0) \\ &= -\alpha \sum [\tau\mu e - xs]_i^- - \alpha \sum [\tau\mu e - xs]_i^+, \end{aligned}$$

from which we see that the negative components of  $\tau\mu e - xs$  are responsible for reducing the duality gap. On the other hand, the small components of  $x_i s_i$ , i.e.,  $x_i s_i < \tau \frac{x^T s}{n}$  indicate that the iterate is “close” to the boundary of the positive orthant. For these coordinates, using (1.2) we have

$$\begin{aligned} x_i(\alpha)s_i(\alpha) &= (x_i + \alpha\Delta x_i)(s_i + \alpha\Delta s_i) \\ &= x_i s_i + \alpha(x_i \Delta s_i + s_i \Delta x_i) + \alpha^2 \Delta x_i \Delta s_i \\ &= x_i s_i + \alpha\left(\tau \frac{x^T s}{n} - x_i s_i\right) + \alpha^2 \Delta x_i \Delta s_i. \end{aligned}$$

Because the coefficient of the first order term of  $\alpha$ ,  $\tau \frac{x^T s}{n} - x_i s_i$ , is bigger than zero, then  $x_i(\alpha)s_i(\alpha)$  increases locally and pushes the iterate to the interior of the first orthant, i.e., keeping the centrality. Ai and Zhang [2] suggested to treat negative and positive components of  $\tau\mu e - xs$  separately to obtain a better iteration complexity bound for large neighborhood IPMs. We generalize this idea to SDO and show that, given a feasible starting point  $(X^0, y^0, S^0)$  in  $\mathcal{N}(\tau_1, \tau_2, \eta)$ , our algorithm terminates in at most  $O(\eta\sqrt{\kappa_\infty n} \log \frac{\text{Tr}(X^0 S^0)}{\epsilon})$  iterations for SDO, where  $n$  is the dimension of the problem,  $\kappa_\infty$  is a parameter associated with the scaling matrix  $P$ , and  $\epsilon$  is the required precision. In particular, when the parameter  $\eta$  is a fixed constant, our large neighborhood path following algorithm has the same theoretical complexity as a small neighborhood algorithm that uses NT scaling. Likewise, when  $\eta$  is chosen to be in the order of  $\sqrt{n}$ , this complexity coincides with the known results for the classical large neighborhood algorithms.

We organize our paper as follows. In Section 2, we introduce the primal-dual pair of SDO problems and briefly explain how path-following interior point methods work. In Section 3, we define the positive and negative parts of a symmetric matrix, and prove some of their intriguing properties. By using these new definitions, we introduce a new neighborhood which is proved to be a large neighborhood. In Section 4, we explain the way to decompose the classical Newton direction and present the framework of our algorithm. The convergency analysis and the theoretical complexity bound are presented in Section 5 and some numerical results are provided in Section 6. Finally, we finish the paper with some conclusions and considerations about future work.

## 1.1 Notations

Throughout the paper, we use the following notations:

- $\mathcal{R}^n$ : the  $n$ -dimensional Euclidean space.
- $\mathcal{R}^{m \times n}$ : the set of all  $m \times n$  matrices.
- $\mathcal{S}^n$ : the set of all  $n \times n$  symmetric matrices.
- $\mathcal{S}_+^n$ : the set of all  $n \times n$  symmetric positive semidefinite matrices.
- $\mathcal{S}_{++}^n$ : the set of all  $n \times n$  symmetric positive definite matrices.
- $Q \succeq 0$ :  $Q$  is positive semidefinite, where  $Q \in \mathcal{S}^n$ .
- $Q \succ 0$ :  $Q$  is positive definite, where  $Q \in \mathcal{S}^n$ .
- $\text{Tr}(Q)$ : the trace of a matrix  $Q \in \mathcal{R}^{n \times n}$ , i.e.,  $\text{Tr}(Q) := \sum_{i=1}^n Q_{ii}$ .
- $\lambda_i(Q)$ : the eigenvalues of  $Q \in \mathcal{S}^n$ ,  $i = 1, 2, \dots, n$ .
- $\lambda_{\min}(Q)$ : the smallest eigenvalue of  $Q \in \mathcal{S}^n$ .
- $\lambda_{\max}(Q)$ : the largest eigenvalue of  $Q \in \mathcal{S}^n$ .
- $\Lambda(Q)$ : the diagonal matrix with all the eigenvalues of  $Q$  as diagonal elements.
- $\text{cond}(Q)$ : the condition number of  $Q$ , defined as  $\text{cond}(Q) = \lambda_{\max}(Q)/\lambda_{\min}(Q)$ .
- $\|Q\|$ : the Euclidean norm for  $Q \in \mathcal{R}^{n \times n}$ , i.e.,  $\|Q\| = \max_{\|u\|=1} \|Qu\|$ .
- $\|Q\|_F$ : the Frobenius norm of  $Q \in \mathcal{R}^{n \times n}$ , i.e.,  $\|Q\|_F = \sqrt{\text{Tr}(Q^T Q)}$ .
- $\text{vec}(Q)$ : the vector obtained by stacking  $Q$ 's columns one by one. See Appendix A.

## 2 Semidefinite Optimization Problem

We consider the following semidefinite optimization (SDO) problem

$$\begin{aligned}
 (\mathcal{P}) \quad & \min \quad \text{Tr}(CX) \\
 & \text{s.t.} \quad \text{Tr}(A_i X) = b_i, \quad i = 1, \dots, m, \\
 & \quad \quad X \succeq 0,
 \end{aligned}$$

where  $C, X \in \mathcal{S}^n$ ,  $A_i \in \mathcal{S}^n$ ,  $i = 1, \dots, m$  are linearly independent and  $b = (b_1, \dots, b_m)^T \in \mathcal{R}^m$ . We call problem  $(\mathcal{P})$  the primal form of SDO, and  $X$  is the primal matrix variable.

Corresponding to every primal problem  $(\mathcal{P})$ , there exists a dual problem  $(\mathcal{D})$

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

where  $y \in \mathcal{R}^m$ ,  $S \in \mathcal{S}^n$  and  $(y, S)$  is the dual variable.

The primal-dual feasible set is defined as

$$\mathcal{F} = \left\{ (X, y, S) \in \mathcal{S}_+^n \times \mathcal{R}^m \times \mathcal{S}_+^n \left| \begin{array}{l} \text{Tr}(A_i X) = b_i, \quad i = 1, \dots, m \\ \sum_{i=1}^m y_i A_i + S = C \end{array} \right. \right\},$$

and the relative interior of the primal-dual feasible set is

$$\mathcal{F}^0 = \left\{ (X, y, S) \in \mathcal{S}_{++}^n \times \mathcal{R}^m \times \mathcal{S}_{++}^n \left| \begin{array}{l} \text{Tr}(A_i X) = b_i, \quad i = 1, \dots, m \\ \sum_{i=1}^m y_i A_i + S = C \end{array} \right. \right\}.$$

Under the assumptions that  $\mathcal{F}^0$  is nonempty and the matrices  $A_i$ ,  $i = 1, 2, \dots, m$ , are linearly independent, then  $X^*$  and  $(y^*, S^*)$  are optimal if and only if they satisfy the optimality conditions [8],

$$\begin{aligned} \text{Tr}(A_i X) &= b_i, \quad X \succeq 0, \quad i = 1, \dots, m \\ \sum_{i=1}^m y_i A_i + S &= C, \quad S \succeq 0, \\ X S &= 0. \end{aligned} \tag{2.1}$$

Path-following Interior Point Methods (IPMs) follow the central path that is given as the set of solutions of the perturbed optimality conditions

$$\begin{aligned} \text{Tr}(A_i X) &= b_i, \quad X \succ 0, \quad i = 1, \dots, m \\ \sum_{i=1}^m y_i A_i + S &= C, \quad S \succ 0, \\ X S &= \mu I, \end{aligned} \tag{2.2}$$

rather than (2.1). It is proved in [15, 21] that there is a unique solution  $(X(\mu), y(\mu), S(\mu))$  to the central path equations (2.2) for any barrier parameter  $\mu > 0$ , assuming that the  $\mathcal{F}^0$  is nonempty, and the coefficient matrices  $A_i$ ,  $i = 1, \dots, m$  are linearly independent. Moreover, the limit point  $(X^*, y^*, S^*)$  as  $\mu$  goes to 0 is a primal-dual optimal solution of the corresponding SDO problem.

### 3 Neighborhood

Although path-following interior point algorithms follow the central path while the barrier parameter  $\mu$  is decreasing to 0, they do not stay on the central path exactly. All the iterates are merely required to reside in a neighborhood of the central path, while steadily approaching the optimal set.

One of the popular neighborhoods is the so-called *small neighborhood*, defined as

$$\mathcal{N}_F(\theta) := \left\{ (X, y, S) \in \mathcal{F}^0 \left| \left\| X^{1/2} S X^{1/2} - \mu_g I \right\|_F = \left[ \sum_{i=1}^n (\lambda_i(XS) - \mu_g)^2 \right]^{1/2} \leq \theta \mu_g \right. \right\},$$

where  $\theta \in (0, 1)$  and  $\mu_g := \text{Tr}(XS)/n$  is associated with the actual duality gap. Another popular alternative is called *negative infinity neighborhood* that is a *large neighborhood*, defined as

$$\mathcal{N}_\infty^-(1 - \gamma) := \{(X, y, S) \in \mathcal{F}^0 \mid \lambda_{\min}(XS) \geq \gamma \mu_g\},$$

where  $\gamma \in (0, 1)$ .

In theory, IPMs based on the small neighborhood  $\mathcal{N}_F(\theta)$ , e.g., short step algorithms, have a better iteration complexity bound than algorithms based on the large neighborhood, e.g., large update algorithms. However, computational experience reveals that large update IPMs usually perform better in practice than short step algorithms.

In this paper we explore a variant of large neighborhood path-following IPMs and prove that its iteration complexity is  $O(\eta \sqrt{\kappa_\infty n} \log \frac{\text{Tr}(X^0 S^0)}{\epsilon})$ , where  $n$  is the dimension of

the problem,  $\kappa_\infty$  is a parameter associated with the scaling matrix  $P$  and  $\epsilon$  is the required precision. The new parameter  $\eta$  is used to define our new neighborhood  $\mathcal{N}(\tau_1, \tau_2, \eta)$ . In particular, when  $\eta$  is chosen to be a constant, our new algorithm has the best complexity result  $O(\sqrt{n} \log \frac{\text{Tr}(X^0 S^0)}{\epsilon})$ , which coincides with the complexity of short step IPMs, when the Nesterov-Todd (NT) scaling is used.

In order to introduce the algorithm, we need to investigate a new neighborhood which combines the classical small and large neighborhoods. Before doing so, we need to present some notations.

Let  $M$  be a symmetric real matrix, i.e.,  $M \in \mathcal{S}^n$ , with the *Eigenvalue Decomposition*  $M = Q\Lambda Q^T = \sum_{i=1}^n \lambda_i q_i q_i^T$ , where  $\Lambda$  is a diagonal matrix with all the eigenvalues of  $M$  along its diagonal,  $Q$  is an orthonormal matrix, i.e.,  $QQ^T = I$ , and each column  $q_i$  of  $Q$  is an eigenvector of  $M$  corresponding to the eigenvalue  $\lambda_i$ . Then, we define the positive part  $M^+$  and the negative part  $M^-$  of  $M$  as

$$M^+ := \sum_{\lambda_i \geq 0} \lambda_i q_i q_i^T, \quad M^- := \sum_{\lambda_i \leq 0} \lambda_i q_i q_i^T. \quad (3.1)$$

In particular, for a real number  $M \in \mathcal{S}^1$ ,  $M^+ = \max\{M, 0\}$  and  $M^- = \min\{M, 0\}$ . Likewise, if  $M \in \mathcal{S}^n$  is a diagonal matrix,  $M^+$  and  $M^-$  could be constructed by simply separating the nonnegative and nonpositive entries. Apparently,  $M = M^+ + M^-$ , where  $M^+, -M^- \succeq 0$ .

It turns out that the positive and negative parts of a symmetric matrix have many interesting properties. We present and verify some of them which play a crucial role in the complexity analysis.

First, we show that the triangle inequality holds for the positive part.

**Proposition 3.1** *Assume  $U, V \in \mathcal{S}^n$ , then we have*

$$\|(U + V)^+\|_F \leq \|U^+ + V^+\|_F \leq \|U^+\|_F + \|V^+\|_F.$$

**Proof.** As we see,

$$U = U^+ + U^- = U^+ + \sum_{\lambda_i(U) \leq 0} \lambda_i(U) q_i(U) q_i(U)^T$$

and

$$V = V^+ + V^- = V^+ + \sum_{\lambda_i(V) \leq 0} \lambda_i(V) q_i(V) q_i(V)^T.$$

According to Lemma B.2, we obtain

$$\lambda_i(U + V) \leq \lambda_i(U^+ + V^+)$$

for  $i = 1, \dots, n$ .

Let  $\mathcal{I}$  denote the index set satisfying

$$\mathcal{I} := \{ i \mid \lambda_i(U + V) \geq 0 \}.$$

Then,

$$\begin{aligned}
\|(U + V)^+\|_F &= \left[ \sum_{i \in \mathcal{I}} \lambda_i^2(U + V) \right]^{1/2} \\
&\leq \left[ \sum_{i \in \mathcal{I}} \lambda_i^2(U^+ + V^+) \right]^{1/2} \\
&\leq \|U^+ + V^+\|_F \\
&\leq \|U^+\|_F + \|V^+\|_F,
\end{aligned}$$

which completes the proof. ■

The next lemma reveals that a similarity transformation preserves the Frobenius norm over the positive part of a symmetric matrix.

**Lemma 3.2** *Let  $M \in \mathcal{S}^n$  and  $W$  be a nonsingular matrix. Then we have*

$$\|M^+\|_F = \|(WMW^{-1})^+\|_F.$$

**Proof.** Because  $M$  is similar to  $WMW^{-1}$ , they have the same eigenvalues. In particular, they have the same nonnegative eigenvalues. Then the result follows easily. ■

The next lemma exhibits that the positive part of a symmetric matrix does not exceed, in the sense of Frobenius norm, its positive part after a similarity transformation.

**Lemma 3.3** *Suppose that  $W \in \mathcal{R}^{n \times n}$  is a nonsingular matrix. Then, for any  $M \in \mathcal{S}^n$ , we have*

$$\|M^+\|_F \leq \frac{1}{2} \left\| [WMW^{-1} + (WMW^{-1})^T]^+ \right\|_F.$$

To prove this result, we need to first demonstrate an interesting fact about symmetric matrices.

**Lemma 3.4** *Let  $M \in \mathcal{S}^n$  and  $\lambda_i$  and  $m_{ii}$  denote the  $i^{\text{th}}$  eigenvalue and the  $i^{\text{th}}$  diagonal element of  $M$ , respectively. Then we have*

$$\sum_{\lambda_i \geq 0} \lambda_i^2 \geq \sum_{m_{ii} \geq 0} m_{ii}^2.$$

**Proof.** If  $M$  is positive semidefinite, then for any eigenvalue of  $M$ , we have  $\lambda_i \geq 0$  and  $m_{ii} \geq 0$ . In this case,

$$\sum_{\lambda_i(M) \geq 0} \lambda_i^2 = \sum_{i=1}^n \lambda_i^2 = \|M\|_F^2 \geq \sum_{i=1}^n m_{ii}^2 = \sum_{m_{ii} \geq 0} m_{ii}^2.$$

Let us consider the general case. For any symmetric matrix  $M \in \mathcal{S}^n$ , there exists an eigenvalue decomposition such that  $M = Q\Lambda Q^T = \sum_{i=1}^n \lambda_i q_i q_i^T$ , where  $\Lambda$  is the diagonal matrix with all of the eigenvalues of  $M$  and  $Q$  is an orthogonal matrix, i.e.,  $QQ^T = I$ , and  $q_i$  is the eigenvector of  $M$  corresponding to the eigenvalue  $\lambda_i$ .

Recall the definitions of  $M^+$  and  $M^-$  as in (3.1), and let  $m_{ij}^+$  and  $m_{ij}^-$  denote the  $(i, j)$  element for  $M^+$  and  $M^-$ , respectively. Note the fact that for any  $i$ ,  $m_{ii}^+ \geq 0$  and  $m_{ii}^- \leq 0$ , then we can define the set  $\mathcal{I}$  as

$$\mathcal{I} = \{i \mid m_{ii} = m_{ii}^+ + m_{ii}^- \geq 0\}.$$

For any  $i \in \mathcal{I}$ , we have  $m_{ii}^+ \geq m_{ii}^+ + m_{ii}^- \geq 0$ , since  $m_{ii}^- \leq 0$ . Further, we obtain  $(m_{ii}^+)^2 \geq (m_{ii}^+ + m_{ii}^-)^2$ , for all  $i \in \mathcal{I}$ .

The proof of the lemma follows by

$$\sum_{\lambda_i \geq 0} \lambda_i^2 = \|M^+\|_F^2 \geq \sum_{i=1}^n (m_{ii}^+)^2 \geq \sum_{i \in \mathcal{I}} (m_{ii}^+)^2 \geq \sum_{i \in \mathcal{I}} (m_{ii}^+ + m_{ii}^-)^2 = \sum_{m_{ii} \geq 0} m_{ii}^2. \quad \blacksquare$$

Now, we are ready to prove Lemma 3.3.

**Proof of Lemma 3.3.** It is easy to see that  $\|M^+\|_F^2 = \|[\Lambda(M)]^+\|_F^2 = \sum_{\lambda_i(M) \geq 0} \lambda_i^2(M)$ .

Let us consider the right hand side. According to Theorem B.1, there exists a unitary matrix  $U$  such that  $U(WMW^{-1})U^* = \Lambda(WMW^{-1}) + N = \Lambda(M) + N$ , where  $N$  is a strictly upper triangular matrix and  $U^*$  is the Hermitian adjoint of  $U$ . The last equality is due to the similarity of  $WMW^{-1}$  and  $M$ . From Lemma 3.2, we know that

$$\begin{aligned} \frac{1}{2} \left\| [WMW^{-1} + (WMW^{-1})^T]^+ \right\|_F &= \frac{1}{2} \left\| [U(WMW^{-1} + (WMW^{-1})^T)U^*]^+ \right\|_F \\ &= \frac{1}{2} \left\| [\Lambda(M) + N + \Lambda(M) + N^T]^+ \right\|_F \\ &= \left\| \left[ \Lambda(M) + \frac{N+N^T}{2} \right]^+ \right\|_F. \end{aligned}$$

On the other hand, from Lemma 3.4 we conclude that

$$\|[\Lambda(M)]^+\|_F^2 \leq \left\| \left[ \Lambda(M) + \frac{N+N^T}{2} \right]^+ \right\|_F^2,$$

which implies  $\|M^+\|_F \leq \frac{1}{2} \left\| [WMW^{-1} + (WMW^{-1})^T]^+ \right\|_F$ . \blacksquare

We have adequate results on the positive and negative parts of symmetric matrices so far. This will assist us in the analysis of convergence and complexity analysis. Now, we move to the definition of a new large neighborhood. Analogous to the neighborhood introduced by Ai and Zhang [2], we define our neighborhood, using the positive part in (3.1), as

$$\mathcal{N}(\tau_1, \tau_2, \eta) := \mathcal{N}_\infty^-(1 - \tau_2) \cap \left\{ (X, y, S) \in \mathcal{F}^0 : \left\| [\tau_1 \mu_g I - X^{1/2} S X^{1/2}]^+ \right\|_F \leq \eta(\tau_1 - \tau_2) \mu_g \right\}, \quad (3.2)$$

where  $\eta \geq 1$  and  $0 < \tau_2 < \tau_1 < 1$ .

The next proposition indicates that the neighborhood  $\mathcal{N}(\tau_1, \tau_2, \eta)$  is indeed a large neighborhood.

**Proposition 3.5** *If  $\eta \geq 1$  and  $0 < \tau_2 < \tau_1 < 1$ , then we have*

$$\mathcal{N}_\infty^-(1 - \tau_1) \subseteq \mathcal{N}(\tau_1, \tau_2, \eta) \subseteq \mathcal{N}_\infty^-(1 - \tau_2).$$



**Proof.** From the definition of  $\mathcal{N}(\tau_1, \tau_2, \eta)$  it is obvious that  $\mathcal{N}(\tau_1, \tau_2, \eta) \subseteq \mathcal{N}_\infty^-(1 - \tau_2)$ . For the first inclusion, we need to prove that

$$\mathcal{N}_\infty^-(1 - \tau_1) \subseteq \{(X, y, S) \in \mathcal{F}^0 : \|\tau_1 \mu_g I - X^{1/2} S X^{1/2}\|_F \leq \eta(\tau_1 - \tau_2) \mu_g\}.$$

Given that for  $(X, y, S) \in \mathcal{N}_\infty^-(1 - \tau_1)$ , one has

$$\tau_1 \mu_g I - X^{1/2} S X^{1/2} \preceq 0, \quad (3.3)$$

which implies  $[\tau_1 \mu_g I - X^{1/2} S X^{1/2}]^+ = 0$ , leading to the claimed relationship.  $\blacksquare$

Moreover, if the parameter  $\eta \geq \sqrt{n}$ , then the neighborhood  $\mathcal{N}(\tau_1, \tau_2, \eta)$  is exactly the negative infinity neighborhood  $\mathcal{N}_\infty^-(1 - \tau_2)$ .

**Proposition 3.6** *If  $\eta \geq \sqrt{n}$  and  $0 < \tau_2 < \tau_1 < 1$ , then we have*

$$\mathcal{N}(\tau_1, \tau_2, \eta) = \mathcal{N}_\infty^-(1 - \tau_2).$$

**Proof.** To complete the proof, it is sufficient to show that for any  $(X, y, S) \in \mathcal{N}_\infty^-(1 - \tau_2)$ , we have

$$\mathcal{N}_\infty^-(1 - \tau_2) \subseteq \{(X, y, S) \in \mathcal{F}^0 : \|\tau_1 \mu_g I - X^{1/2} S X^{1/2}\|_F \leq \eta(\tau_1 - \tau_2) \mu_g\}. \quad (3.4)$$

Because  $(X, y, S) \in \mathcal{N}_\infty^-(1 - \tau_2)$ , it follows that

$$\lambda_{\min}(X^{1/2} S X^{1/2}) = \lambda_{\min}(XS) \geq \tau_2 \mu_g.$$

Therefore,

$$\lambda_{\max}([\tau_1 \mu_g I - X^{1/2} S X^{1/2}]^+) \leq (\tau_1 - \tau_2) \mu_g.$$

This implies

$$\|\tau_1 \mu_g I - X^{1/2} S X^{1/2}\|_F \leq \sqrt{n}(\tau_1 - \tau_2) \mu_g,$$

which shows that (3.4) holds when  $\eta \geq \sqrt{n}$ .  $\blacksquare$

## 4 Search Direction

Given an iterate  $(X, y, S)$ , path-following IPMs generate the next iterate by taking a Newton step to system (2.2). Let the target be a point on the central path corresponding to  $\mu = \tau \mu_g$ , where  $\tau \in [0, 1]$  is called centering parameter and  $\mu_g = \text{Tr}(XS)/n$  corresponds to the actual duality gap. To move from the current point  $(X, y, S)$  towards the target on the central path, we wish we could obtain a symmetric search direction from the following linear system

$$\begin{aligned} \text{Tr}(A_i \Delta X) &= 0, \\ \sum_{i=1}^m \Delta y_i A_i + \Delta S &= 0, \\ \Delta X S + X \Delta S &= \tau \mu_g I - X S. \end{aligned} \quad (4.1)$$

Although the second equality guarantees us a symmetric  $\Delta S$ , system (4.1) does not allow a symmetric solution matrix  $\Delta X$ . Various remedies have been proposed since the middle of 1990's. The readers who are interested are referred to the paper [28] for a comprehensive discussion. We use the approach proposed by Zhang [35], who suggested to replace the last equation in system (2.2) by

$$H_P(XS) = \mu I, \quad (4.2)$$

where  $H_P(\cdot)$  is a symmetrization transformation defined as

$$H_P(M) = \frac{1}{2} [PMP^{-1} + (PMP^{-1})^T]$$

for a given matrix  $M$  and a given nonsingular matrix  $P$ . In particular, if  $P = I$  then for any symmetric matrix  $M$ ,  $H_I(M) = M$ . In [35], Zhang observed that if  $P$  is nonsingular, then

$$H_P(M) = \mu I \Leftrightarrow M = \mu I.$$

Thus, the search direction is well defined by the following system

$$\text{Tr}(A_i \Delta X) = 0, \quad (4.3a)$$

$$\sum_{i=1}^m \Delta y_i A_i + \Delta S = 0, \quad (4.3b)$$

$$H_P(\Delta XS + X \Delta S) = \tau \mu_g I - H_P(XS). \quad (4.3c)$$

We refer to the directions derived from (4.3) as the Monteiro-Zhang (MZ) family. In particular, when  $P = I$ , the direction obtained from (4.3) coincides with the AHO direction [5]. If  $P = X^{-1/2}$  or  $S^{1/2}$ , then (4.3) gives the H..K..M directions [11, 15, 17, 18]. Further, we obtain the NT direction when  $P = W_{NT}^{-1/2}$ , where  $W_{NT}$  is the solution of the system  $W_{NT}^{-1} X W_{NT}^{-1} = S$ . Nesterov and Todd [22, 23] proved the existence and uniqueness of such a solution as  $W_{NT} = S^{-1/2} (S^{1/2} X S^{1/2})^{1/2} S^{-1/2}$ .

In terms of Kronecker product<sup>1</sup>, equation (4.3c) can be expressed as

$$E \text{vec}(\Delta X) + F \text{vec}(\Delta S) = \text{vec}(\tau \mu_g I - H_P(XS)),$$

where

$$E = \frac{1}{2} (P^{-T} S \otimes P + P \otimes P^{-T} S), \quad F = \frac{1}{2} (PX \otimes P^{-T} + P^{-T} \otimes PX). \quad (4.4)$$

Todd, Toh and Tütüncü [29] proved that system (4.3) has a unique solution for any  $(X, y, S) \in \mathcal{S}_{++}^n \times \mathcal{R}^m \times \mathcal{S}_{++}^n$  and for the scaling matrix  $P$  which satisfies  $PXSP^{-1} \in \mathcal{S}^n$ . Actually, this still holds under some weaker conditions, as the authors pointed out in [19]. Throughout this paper, we restrict the scaling matrix  $P$  to the aforementioned specific class

$$\mathcal{P}(X, S) := \{P \in \mathcal{S}_{++}^n \mid PXSP^{-1} \in \mathcal{S}^n\}, \quad (4.5)$$

where  $X, S \in \mathcal{S}_{++}^n$ . Apparently,  $P = X^{-1/2}$ ,  $S^{1/2}$  and  $W_{NT}^{1/2}$  belong to this specific class. Unfortunately,  $P = I$  does not. We should mention that this restriction on  $P$  is

---

<sup>1</sup>For the definition and properties of Kronecker product, please refer to APPENDIX A.

common for large neighborhood path-following algorithms proposed in [20]. Moreover, this restriction on  $P$  does not lose any generality, in terms of the solution set of system (4.3), as Monteiro indicates in [18].

After obtaining the search direction, most classic large neighborhood path-following algorithms do a linear search to decide how far they move along the search direction in attempt to minimize the duality gap as much as possible within the neighborhood  $\mathcal{N}_\infty^-(1 - \tau_2)$ .

In our new algorithm, we decompose the Newton direction into two separate parts according to the positive and negative parts of  $\tau\mu_g I - H_P(XS)$ . Thus, we need to solve the following two systems:

$$\text{Tr}(A\Delta X_-) = 0, \quad (4.6a)$$

$$\sum_{i=1}^m (\Delta y_i)_- A_i + \Delta S_- = 0, \quad (4.6b)$$

$$H_P(\Delta X_- S + X \Delta S_-) = [\tau\mu_g I - H_P(XS)]^-, \quad (4.6c)$$

and

$$\text{Tr}(A\Delta X_+) = 0, \quad (4.7a)$$

$$\sum_{i=1}^m (\Delta y_i)_+ A_i + \Delta S_+ = 0, \quad (4.7b)$$

$$H_P(\Delta X_+ S + X \Delta S_+) = [\tau\mu_g I - H_P(XS)]^+, \quad (4.7c)$$

where  $P \in \mathcal{P}(X, S)$  and  $(\Delta X_-, \Delta y_-, \Delta S_-)$  denotes the negative part of the search direction, while  $(\Delta X_+, \Delta y_+, \Delta S_+)$  analogously denotes the positive part of the search direction. Again, equations (4.6c) and (4.7c) could be written in Kronecker product form as

$$E\mathbf{vec}(\Delta X_-) + F\mathbf{vec}(\Delta S_-) = \mathbf{vec}([\tau\mu_g I - H_P(XS)]^-) \quad (4.8)$$

and

$$E\mathbf{vec}(\Delta X_+) + F\mathbf{vec}(\Delta S_+) = \mathbf{vec}([\tau\mu_g I - H_P(XS)]^+), \quad (4.9)$$

respectively.

It is easy to see that systems (4.6) and (4.7) are also well-defined and have a unique solution because  $P \in \mathcal{P}(X, S)$ . To get the best step lengths for both of the directions, we expect to solve the following subproblem

$$\begin{aligned} \min \quad & \text{Tr}(X(\alpha)S(\alpha)) \\ \text{s.t.} \quad & (X(\alpha), y(\alpha), S(\alpha)) \in \mathcal{N}(\tau_1, \tau_2, \eta) \\ & 0 \leq \alpha_- \leq 1, 0 \leq \alpha_+ \leq 1, \end{aligned} \quad (4.10)$$

where  $\alpha = (\alpha_-, \alpha_+)$  denotes the step lengths along the direction  $(\Delta X_-, \Delta y_-, \Delta S_-)$  and  $(\Delta X_+, \Delta y_+, \Delta S_+)$ , respectively. Finally, the new iterate is given by

$$\begin{aligned} (X(\alpha), y(\alpha), S(\alpha)) & := (X, y, S) + (\Delta X(\alpha), \Delta y(\alpha), \Delta S(\alpha)) \\ & := (X, y, S) + \alpha_-(\Delta X_-, \Delta y_-, \Delta S_-) + \alpha_+(\Delta X_+, \Delta y_+, \Delta S_+). \end{aligned} \quad (4.11)$$

So far, we have already introduced the most important ingredients of the new algorithm: the newly-defined neighborhood  $\mathcal{N}(\tau_1, \tau_2, \eta)$  given by (3.2) and the new search

directions based on systems (4.6) and (4.7). We are ready to present a generic frame for our algorithm.

---

**Algorithm 1** Path-following IPM based on the  $\mathcal{N}(\tau_1, \tau_2, \eta)$  neighborhood

---

**Input:**

- required precision  $\epsilon > 0$ ;
- neighborhood parameters  $\eta \geq 1, 0 < \tau_2 < \tau_1 < 1$ ;
- reference parameter  $0 \leq \tau \leq 1$ ;
- an initial point  $(X^0, y^0, S^0) \in \mathcal{N}(\tau_1, \tau_2, \eta)$  with  $\mu_g^0 = \text{Tr}(X^0 S^0)/n$ ;

**while**  $\mu_g^k > \epsilon$  **do**

- (1) Compute the scaling matrix  $P^k \in \mathcal{P}(X^k, S^k)$ .
- (2) Compute the directions  $(\Delta X_-^k, \Delta y_-^k, \Delta S_-^k)$  by (4.6) and  $(\Delta X_+^k, \Delta y_+^k, \Delta S_+^k)$  by (4.7).
- (3) Find a step length vector  $\alpha^k = (\alpha_-^k, \alpha_+^k) > 0$  giving a sufficient reduction of the duality gap and assuring  $(X(\alpha^k), y(\alpha^k), S(\alpha^k)) \in \mathcal{N}(\tau_1, \tau_2, \eta)$ .
- (4) Set  $(X^{k+1}, y^{k+1}, S^{k+1}) = (X(\alpha^k), y(\alpha^k), S(\alpha^k))$ .
- (5) Set  $\mu_g^{k+1} := \text{Tr}(X^{k+1} S^{k+1})/n$  and  $k := k + 1$ .

**end while**

---

There are three comments we would like to address about the presented algorithm. First of all, although we suggest to solve problem (4.10) to decide the best step lengths, solving this problem could be expensive. Hence, a “sufficient” duality gap decrease obtained for low computational cost is preferred against the “maximal possible” duality gap decrease for high computational cost. Moreover, it is also not necessary to solve problem (4.10). Even if we do not use the optimal solution of problem (4.10) as the step lengths, we are still able to achieve polynomial convergence as it is discussed in Section 5. Second, in spite of the fact that two linear systems (4.6) and (4.7) have to be solved, however, the additional cost is very marginal, since both of (4.6) and (4.7) have the same coefficient matrix. At each iteration, the algorithm only needs to form and factorize the Schur matrix once which together yield the majority of the total running time. Then backsolving needs to be executed once, simultaneously for two right-hand-sides specified in (4.6) and (4.7), respectively. Third, it might appear computationally expensive to obtain the negative and positive parts in (4.6) and (4.7). However, if the NT scaling is used, the strategy proposed by Todd, Toh and Tütüncü [29] for computing the NT scaling matrix and the NT direction, can help us. The basic idea is to simultaneously scale  $X$  and  $S$  to diagonal matrices. In this case, after scaling, the right hand sides of (4.6c) and (4.7c) also become diagonal. Then, it is straightforward to decide the negative and positive parts. The preliminary numerical tests of Section 6 confirm that (4.6) and (4.7) can be simultaneously solved without significant increase of computational time.

## 5 Complexity Analysis

In this part, we present the convergence and complexity proofs for Algorithm 1. Recall that our algorithm is based on the MZ family, we scale problems  $(\mathcal{P})$  and  $(\mathcal{D})$  as Monteiro and Todd proposed in [19] in order to analyze the algorithm in a unified way for the class of matrices  $P \in \mathcal{P}(X, S)$ . In addition, this scaling procedure simplifies the proofs of the main results. At the end of this section, after demonstrating several technical lemmas, we present the most important result of polynomial convergence.

### 5.1 Scaling Procedure

We scale the primal and dual variables of problems  $(\mathcal{P})$  and  $(\mathcal{D})$  in the form of,

$$\tilde{X} := PXP, \quad (\tilde{y}, \tilde{S}) := (y, P^{-1}SP^{-1}). \quad (5.1)$$

To keep consistency, we also have to apply the same scaling to the other data in  $(\mathcal{P})$  and  $(\mathcal{D})$  as well, i.e.,

$$\tilde{C} := P^{-1}CP^{-1}, \quad (\tilde{A}_i, \tilde{b}_i) := (P^{-1}A_iP^{-1}, b_i), \text{ for } i = 1, \dots, m.$$

As mentioned, we restrict the scaling matrix  $P \in \mathcal{P}(X, S)$  as defined by (4.5). It is easy to see that for  $X, S \in \mathcal{S}_{++}^n$  one has

$$\mathcal{P}(X, S) := \{P \in \mathcal{S}_{++}^n \mid PXS P^{-1} \in \mathcal{S}^n\} = \{P \in \mathcal{S}_{++}^n : \tilde{X}\tilde{S} = \tilde{S}\tilde{X}\}, \quad (5.2)$$

i.e.,  $\tilde{X}$  and  $\tilde{S}$  become commutable after scaling under  $P$ . The commutativity of  $\tilde{X}\tilde{S}$  also implies that  $\tilde{X}\tilde{S}$  is symmetric if  $\tilde{X}$  and  $\tilde{S}$  are both symmetric. Further, the requirement on  $P$  also guarantees that  $\tilde{X}$  and  $\tilde{S}$  can be simultaneously diagonalized (i.e., they have eigenvalue decompositions with the same  $Q$ ) according to Proposition B.6.

From now on, we use  $\Lambda$  to denote the diagonal matrix  $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ , where  $\lambda_i$  for  $i = 1, \dots, n$  are the eigenvalues of  $\tilde{X}\tilde{S}$  with increasing order  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ . We should emphasize that the matrices  $\tilde{X}\tilde{S}$ ,  $\tilde{S}\tilde{X}$ ,  $X S$ ,  $S X$ ,  $X^{1/2}S X^{1/2}$  and  $S^{1/2}X S^{1/2}$  have the same eigenvalues, since they are all similar to each other.

In the scaled space problems  $(\mathcal{P})$  and  $(\mathcal{D})$  are equivalent to the following pair of problems:

$$\begin{aligned} (\tilde{\mathcal{P}}) \quad & \min \quad \text{Tr}(\tilde{C}\tilde{X}) \\ & \text{s.t.} \quad \text{Tr}(\tilde{A}_i\tilde{X}) = \tilde{b}_i, \quad i = 1, \dots, m, \\ & \quad \quad \tilde{X} \succeq 0, \end{aligned}$$

and

$$\begin{aligned} (\tilde{\mathcal{D}}) \quad & \max \quad \tilde{b}^T \tilde{y} \\ & \text{s.t.} \quad \sum_{i=1}^m \tilde{y}_i \tilde{A}_i + \tilde{S} = \tilde{C}, \\ & \quad \quad \tilde{S} \succeq 0. \end{aligned}$$

The search direction  $(\Delta X, \Delta y, \Delta S)$  based on system (4.6) and (4.7) corresponds to the scaled direction  $(\widetilde{\Delta X}, \widetilde{\Delta y}, \widetilde{\Delta S})$  defined as

$$\widetilde{\Delta X}_- = P\Delta X_-P, \quad \widetilde{\Delta y}_- = \Delta y_-, \quad \widetilde{\Delta S}_- = P^{-1}\Delta S_-P^{-1}, \quad (5.3)$$

$$\widetilde{\Delta X}_+ = P\Delta X_+P, \quad \widetilde{\Delta y}_+ = \Delta y_+, \quad \widetilde{\Delta S}_+ = P^{-1}\Delta S_+P^{-1}. \quad (5.4)$$

The directions  $(\widetilde{\Delta X}_-, \widetilde{\Delta y}_-, \widetilde{\Delta S}_-)$  and  $(\widetilde{\Delta X}_+, \widetilde{\Delta y}_+, \widetilde{\Delta S}_+)$  are readily verified to be solutions of the scaled Newton systems

$$\text{Tr}(\widetilde{A}_i \widetilde{\Delta X}_-) = 0, \quad (5.5a)$$

$$\sum_{i=1}^m (\widetilde{\Delta y}_i)_- \widetilde{A}_i + \widetilde{\Delta S}_- = 0, \quad (5.5b)$$

$$H_I(\widetilde{\Delta X}_- \widetilde{S} + \widetilde{X} \widetilde{\Delta S}_-) = [\tau \widetilde{\mu}_g I - \widetilde{X} \widetilde{S}]^-, \quad (5.5c)$$

and

$$\text{Tr}(\widetilde{A}_i \widetilde{\Delta X}_+) = 0, \quad (5.6a)$$

$$\sum_{i=1}^m (\widetilde{\Delta y}_i)_+ \widetilde{A}_i + \widetilde{\Delta S}_+ = 0, \quad (5.6b)$$

$$H_I(\widetilde{\Delta X}_+ \widetilde{S} + \widetilde{X} \widetilde{\Delta S}_+) = [\tau \widetilde{\mu}_g I - \widetilde{X} \widetilde{S}]^+, \quad (5.6c)$$

respectively. To simplify the notation, we use  $\widetilde{X} \widetilde{S}$  rather than  $H_I(\widetilde{X} \widetilde{S})$ , since  $\widetilde{X} \widetilde{S} = H_I(\widetilde{X} \widetilde{S})$  when the scaling matrix  $P \in \mathcal{P}(X, S)$ . In terms of the Kronecker product, equations (5.5c) and (5.6c) become

$$\widetilde{E} \text{vec}(\widetilde{\Delta X}_-) + \widetilde{F} \text{vec}(\widetilde{\Delta S}_-) = \text{vec}([\tau \widetilde{\mu}_g I - \widetilde{X} \widetilde{S}]^-), \quad (5.7a)$$

$$\widetilde{E} \text{vec}(\widetilde{\Delta X}_+) + \widetilde{F} \text{vec}(\widetilde{\Delta S}_+) = \text{vec}([\tau \widetilde{\mu}_g I - \widetilde{X} \widetilde{S}]^+), \quad (5.7b)$$

respectively, where

$$\widetilde{E} = \frac{1}{2}(\widetilde{S} \otimes I + I \otimes \widetilde{S}), \quad \widetilde{F} = \frac{1}{2}(\widetilde{X} \otimes I + I \otimes \widetilde{X}). \quad (5.8)$$

After deciding the step lengths, the iterates are updated as follows:

$$\begin{aligned} (\widetilde{X}(\alpha), \widetilde{y}(\alpha), \widetilde{S}(\alpha)) &= (\widetilde{X}, \widetilde{y}, \widetilde{S}) + (\widetilde{\Delta X}(\alpha), \widetilde{\Delta y}(\alpha), \widetilde{\Delta S}(\alpha)) \\ &= (\widetilde{X}, \widetilde{y}, \widetilde{S}) + \alpha_- (\widetilde{\Delta X}_-, \widetilde{\Delta y}_-, \widetilde{\Delta S}_-) + \alpha_+ (\widetilde{\Delta X}_+, \widetilde{\Delta y}_+, \widetilde{\Delta S}_+). \end{aligned} \quad (5.9)$$

The next proposition formalizes the equivalence between the original and the scaled problems.

**Proposition 5.1** *If  $(X, y, S)$  and  $(\widetilde{X}, \widetilde{y}, \widetilde{S})$  are related to each other as specified by (5.1),  $(X(\alpha), y(\alpha), S(\alpha))$  and  $(\widetilde{X}(\alpha), \widetilde{y}(\alpha), \widetilde{S}(\alpha))$  are defined by (4.11) and (5.9), respectively, then we have*

1.  $(X, y, S) \in \mathcal{F}$  if and only if  $(\widetilde{X}, \widetilde{y}, \widetilde{S})$  is feasible for  $(\widetilde{\mathcal{P}})$  and  $(\widetilde{\mathcal{D}})$ ;
2.  $(X, y, S) \in \mathcal{N}(\tau_1, \tau_2, \eta)$  if and only if  $(\widetilde{X}, \widetilde{y}, \widetilde{S}) \in \widetilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$ , where  $\widetilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$  is the neighborhood corresponding to  $(\widetilde{\mathcal{P}})$  and  $(\widetilde{\mathcal{D}})$ ;
3.  $\widetilde{X}(\alpha) = PX(\alpha)P$ ,  $\widetilde{y}(\alpha) = y(\alpha)$ ,  $\widetilde{S}(\alpha) = P^{-1}S(\alpha)P^{-1}$  and  $\mu(\alpha) = \widetilde{\mu}_g(\alpha)$ , where  $\widetilde{\mu}_g(\alpha) = \text{Tr}(\widetilde{X}(\alpha)\widetilde{S}(\alpha))/n$ .

## 5.2 Technical Results

Before proceeding to the complexity result, we have to prove some technical lemmas. Throughout this section we fix the reference parameter to  $\tau = \tau_1$  and let:

- A.1  $(\widetilde{\Delta X}_-, \widetilde{\Delta y}_-, \widetilde{\Delta S}_-)$  and  $(\widetilde{\Delta X}_+, \widetilde{\Delta y}_+, \widetilde{\Delta S}_+)$  be the solutions of (5.5) and (5.6), respectively;
- A.2  $\widetilde{\Delta X}(\alpha) := \alpha_- \widetilde{\Delta X}_- + \alpha_+ \widetilde{\Delta X}_+$  and  $\widetilde{\Delta S}(\alpha) := \alpha_- \widetilde{\Delta S}_- + \alpha_+ \widetilde{\Delta S}_+$ .

From the following lemma, we see that if the current iterate is feasible, then the search directions are orthogonal.

**Lemma 5.2** *Under A.1 and A.2, we have*

$$\text{Tr}(\widetilde{\Delta X}(\alpha) \widetilde{\Delta S}(\alpha)) = 0.$$

**Proof.** The proof is straightforward by using (5.5a), (5.5b), (5.6a), and (5.6b).  $\blacksquare$

**Lemma 5.3** *If  $P \in \mathcal{P}(X, S)$ , then we have*

$$\text{Tr}(\widetilde{X} \widetilde{\Delta S}_-) + \text{Tr}(\widetilde{\Delta X}_- \widetilde{S}) = \text{Tr}([\tau_1 \widetilde{\mu}_g I - \widetilde{X} \widetilde{S}]^-), \quad (5.10)$$

and

$$\text{Tr}(\widetilde{X} \widetilde{\Delta S}_+) + \text{Tr}(\widetilde{\Delta X}_+ \widetilde{S}) = \text{Tr}([\tau_1 \widetilde{\mu}_g I - \widetilde{X} \widetilde{S}]^+). \quad (5.11)$$

**Proof.** Using the fact that  $\text{Tr}(M) = \text{Tr}(H_I(M))$  for any matrix  $M \in \mathcal{R}^{n \times n}$ , it is easy to see that

$$\begin{aligned} \text{Tr}(\widetilde{X} \widetilde{\Delta S}_-) + \text{Tr}(\widetilde{\Delta X}_- \widetilde{S}) &= \text{Tr}(\widetilde{X} \widetilde{\Delta S}_- + \widetilde{\Delta X}_- \widetilde{S}) \\ &= \text{Tr}(H_I(\widetilde{X} \widetilde{\Delta S}_- + \widetilde{\Delta X}_- \widetilde{S})) \\ &= \text{Tr}([\tau_1 \widetilde{\mu}_g I - \widetilde{X} \widetilde{S}]^-). \end{aligned}$$

One can show (5.11) analogously.  $\blacksquare$

Intuitively, we wish to reduce the duality gap as much as possible in every iteration. The next result, however, shows that Algorithm 1 holds a lower bound for duality gap reduction. This bound derives from feasibility considerations as we will see in later discussions.

**Lemma 5.4** *Let  $(\widetilde{X}, \widetilde{y}, \widetilde{S}) \in \mathcal{F}^0$ , then for every  $\alpha := (\alpha_-, \alpha_+) \in [0, 1]$ , we have*

$$\text{Tr}(\widetilde{X}(\alpha) \widetilde{S}(\alpha)) = \text{Tr}(\widetilde{X} \widetilde{S}) + \alpha_- \text{Tr}([\tau_1 \widetilde{\mu}_g I - \widetilde{X} \widetilde{S}]^-) + \alpha_+ \text{Tr}([\tau_1 \widetilde{\mu}_g I - \widetilde{X} \widetilde{S}]^+).$$

Furthermore,

$$\widetilde{\mu}_g(\alpha) = \widetilde{\mu}_g + \alpha_- \frac{\text{Tr}([\tau_1 \widetilde{\mu}_g I - \widetilde{X} \widetilde{S}]^-)}{n} + \alpha_+ \frac{\text{Tr}([\tau_1 \widetilde{\mu}_g I - \widetilde{X} \widetilde{S}]^+)}{n} \geq (1 - \alpha_-) \widetilde{\mu}_g.$$

**Proof.** Using Lemma 5.2 and Lemma 5.3, we have

$$\begin{aligned}
\text{Tr}(\tilde{X}(\alpha)\tilde{S}(\alpha)) &= \text{Tr}((\tilde{X} + \alpha_- \widetilde{\Delta X}_- + \alpha_+ \widetilde{\Delta X}_+)(\tilde{S} + \alpha_- \widetilde{\Delta S}_- + \alpha_+ \widetilde{\Delta S}_+)) \\
&= \text{Tr}(\tilde{X}\tilde{S}) + \alpha_- (\text{Tr}(\widetilde{\Delta X}_- \tilde{S}) + \text{Tr}(\tilde{X} \widetilde{\Delta S}_-)) + \alpha_+ (\text{Tr}(\widetilde{\Delta X}_+ \tilde{S}) + \text{Tr}(\tilde{X} \widetilde{\Delta S}_+)) \\
&\quad + \text{Tr}(\widetilde{\Delta X}(\alpha) \widetilde{\Delta S}(\alpha)) \\
&= \text{Tr}(\tilde{X}\tilde{S}) + \alpha_- \text{Tr}([\tau_1 \tilde{\mu}_g I - \tilde{X}\tilde{S}]^-) + \alpha_+ \text{Tr}([\tau_1 \tilde{\mu}_g I - \tilde{X}\tilde{S}]^+).
\end{aligned}$$

Then, we have

$$\begin{aligned}
\tilde{\mu}_g(\alpha) &= \frac{\text{Tr}(\tilde{X}(\alpha)\tilde{S}(\alpha))}{n} \\
&= \frac{\text{Tr}(\tilde{X}\tilde{S})}{n} + \alpha_- \frac{\text{Tr}([\tau_1 \tilde{\mu}_g I - \tilde{X}\tilde{S}]^-)}{n} + \alpha_+ \frac{\text{Tr}([\tau_1 \tilde{\mu}_g I - \tilde{X}\tilde{S}]^+)}{n} \\
&\geq \tilde{\mu}_g - \alpha_- \frac{\text{Tr}(\tilde{X}\tilde{S})}{n} \\
&= (1 - \alpha_-) \tilde{\mu}_g,
\end{aligned}$$

where the inequality is due to the fact that  $\tilde{X}, \tilde{S} \in \mathcal{S}_+^n$  implies

$$\text{Tr}([\tau_1 \tilde{\mu}_g I - \tilde{X}\tilde{S}]^-) \geq \text{Tr}(-\tilde{X}\tilde{S}). \quad \blacksquare$$

The next lemma shows that the negative part of  $\tau_1 \tilde{\mu}_g I - \tilde{X}\tilde{S}$  is also bounded in terms of the duality gap at the current iteration.

**Lemma 5.5** *Let  $(\tilde{X}, \tilde{y}, \tilde{S}) \in \mathcal{F}^0$ , then*

$$\text{Tr}([\tau_1 \tilde{\mu}_g I - \tilde{X}\tilde{S}]^-) \leq -(1 - \tau_1) \text{Tr}(\tilde{X}\tilde{S}). \quad (5.12)$$

**Proof.** Recall that

$$[\tau_1 \tilde{\mu}_g I - \tilde{X}\tilde{S}]^- + [\tau_1 \tilde{\mu}_g I - \tilde{X}\tilde{S}]^+ = \tau_1 \tilde{\mu}_g I - \tilde{X}\tilde{S}.$$

Taking the trace of both sides, we have

$$\begin{aligned}
\text{Tr}([\tau_1 \tilde{\mu}_g I - \tilde{X}\tilde{S}]^-) &= (\tau_1 - 1) \text{Tr}(\tilde{X}\tilde{S}) - \text{Tr}([\tau_1 \tilde{\mu}_g I - \tilde{X}\tilde{S}]^+) \\
&\leq -(1 - \tau_1) \text{Tr}(\tilde{X}\tilde{S}),
\end{aligned}$$

which completes the proof. \blacksquare

The next results, Proposition 5.6 and Corollary 5.7, imply that Algorithm 1 reduces the duality gap steadily if the feasibility of the iterates can be preserved. From now on, we introduce the notation  $\beta = (\tau_1 - \tau_2)/\tau_1$ , then we have  $\beta \in (0, 1)$  and  $\tau_2 = (1 - \beta)\tau_1$ . Further let us denote

$$\hat{\eta} = \max \left\{ \frac{\left\| [\tau_1 \tilde{\mu}_g I - \tilde{X}\tilde{S}]^+ \right\|_F}{\beta \tau_1 \tilde{\mu}_g}, 1 \right\}.$$

It follows that if  $(\tilde{X}, \tilde{S}) \in \tilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$ , then  $1 \leq \hat{\eta} \leq \eta$ .



**Proposition 5.6** Let  $(\tilde{X}, \tilde{y}, \tilde{S}) \in \tilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$ . Then we have

$$\tilde{\mu}_g(\alpha) \leq \tilde{\mu}_g - \alpha_-(1 - \tau_1)\tilde{\mu}_g + \alpha_+ \frac{\hat{\eta}\beta\tau_1\tilde{\mu}_g}{\sqrt{n}}.$$

**Proof.** Using Lemmas 5.3 and 5.4, we see that

$$\begin{aligned} \tilde{\mu}_g(\alpha) &= \tilde{\mu}_g + \alpha_- \frac{\text{Tr}([\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^-)}{n} + \alpha_+ \frac{\text{Tr}([\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^+)}{n} \\ &\leq \tilde{\mu}_g - \alpha_-(1 - \tau_1) \frac{\text{Tr}(\tilde{X}\tilde{S})}{n} + \alpha_+ \sqrt{n} \frac{\|[\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^+\|_F}{n} \\ &\leq \tilde{\mu}_g - \alpha_-(1 - \tau_1)\tilde{\mu}_g + \alpha_+ \frac{\hat{\eta}\beta\tau_1\tilde{\mu}_g}{\sqrt{n}}, \end{aligned}$$

where the first inequality is due to Lemma 5.5 and the Cauchy-Schwarz inequality and the last inequality derives from the assumption that  $(\tilde{X}, \tilde{y}, \tilde{S}) \in \tilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$ .  $\blacksquare$

When the parameters  $\tau_1$  and  $\beta$  are chosen appropriately and all the iterates reside in the neighborhood  $\tilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$ , we claim that the duality gap is decreasing at the rate of  $O(1/\sqrt{n})$ .

**Corollary 5.7** Let  $\tau_1 \leq \frac{4}{9}$ ,  $\beta \leq \frac{1}{4}$  and  $(\tilde{X}, \tilde{y}, \tilde{S}) \in \tilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$ . If  $\alpha_- = \alpha_+ \hat{\eta} \sqrt{\frac{\beta\tau_1}{n}}$ , then we have

$$\tilde{\mu}_g(\alpha) \leq \left(1 - \alpha_+ \frac{2\hat{\eta}\sqrt{\beta\tau_1}}{9\sqrt{n}}\right) \tilde{\mu}_g.$$

**Proof.** From Proposition 5.6, it follows that

$$\begin{aligned} \tilde{\mu}_g(\alpha) &\leq \tilde{\mu}_g - \alpha_-(1 - \tau_1)\tilde{\mu}_g + \alpha_+ \frac{\hat{\eta}\beta\tau_1\tilde{\mu}_g}{\sqrt{n}} \\ &\leq \tilde{\mu}_g - \frac{5}{9}\alpha_-\tilde{\mu}_g + \alpha_+ \frac{\hat{\eta}\beta\tau_1\tilde{\mu}_g}{\sqrt{n}} \\ &= \tilde{\mu}_g - \alpha_+ \left(\frac{5}{9} - \sqrt{\beta\tau_1}\right) \frac{\hat{\eta}\sqrt{\beta\tau_1}}{\sqrt{n}} \tilde{\mu}_g \\ &\leq \left(1 - \alpha_+ \frac{2\hat{\eta}\sqrt{\beta\tau_1}}{9\sqrt{n}}\right) \tilde{\mu}_g. \end{aligned}$$

Here the second inequality holds because  $\tau_1 \leq \frac{4}{9}$  and the last inequality is due to the fact that  $\beta\tau_1 \leq \frac{1}{9}$ .  $\blacksquare$

Subsequently, we show how to ensure all the iterates in the neighborhood  $\tilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$ . Although we wish to decrease the duality gap as much as possible, we still need to control the smallest eigenvalue of  $\tilde{X}(\alpha)\tilde{S}(\alpha)$  in order to stay in the neighborhood  $\tilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$ .

**Lemma 5.8** Suppose  $P \in \mathcal{P}(X, S)$  and  $\chi(\alpha) = \tilde{X}\tilde{S} + \alpha_- [\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^- + \alpha_+ [\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^+$ . If  $(\tilde{X}, \tilde{y}, \tilde{S}) \in \tilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$ , then we have

$$\lambda_{\min}(\chi(\alpha)) \geq \tau_2\tilde{\mu}_g + \alpha_+(\tau_1 - \tau_2)\tilde{\mu}_g. \quad (5.13)$$

**Proof.** We first consider the situation when  $\lambda_{\min}(\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}) \geq 0$  and note that  $\lambda_{\min}(\cdot)$  is a homogeneous concave function. Then, we have

$$\begin{aligned}
\lambda_{\min}(\chi(\alpha)) &= \lambda_{\min}(\tilde{X}\tilde{S} + \alpha_+[\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^+) \\
&= \lambda_{\min}(\tilde{X}\tilde{S} + \alpha_+(\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S})) \\
&= \lambda_{\min}((1 - \alpha_+)\tilde{X}\tilde{S} + \alpha_+\tau_1\tilde{\mu}_g I) \\
&\geq (1 - \alpha_+)\lambda_{\min}(\tilde{X}\tilde{S}) + \alpha_+\tau_1\tilde{\mu}_g \\
&\geq (1 - \alpha_+)\tau_2\tilde{\mu}_g + \alpha_+\tau_1\tilde{\mu}_g \\
&= \tau_2\tilde{\mu}_g + \alpha_+(\tau_1 - \tau_2)\tilde{\mu}_g.
\end{aligned}$$

The second inequality holds due to  $(\tilde{X}, \tilde{y}, \tilde{S}) \in \tilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$ .

When  $\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}$  is negative semidefinite, i.e.,  $[\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^- = \tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}$  and  $[\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^+ = 0$ , we have

$$\begin{aligned}
\lambda_{\min}(\chi(\alpha)) &= \lambda_{\min}(\tilde{X}\tilde{S} + \alpha_-[\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^-) \\
&= \lambda_{\min}(\Lambda + \alpha_-(\tau_1\tilde{\mu}_g I - \Lambda)) \\
&\geq \lambda_{\min}(\Lambda + (\tau_1\tilde{\mu}_g I - \Lambda)) \\
&= \tau_1\tilde{\mu}_g \\
&= \tau_2\tilde{\mu}_g + (\tau_1 - \tau_2)\tilde{\mu}_g \\
&\geq \tau_2\tilde{\mu}_g + \alpha_+(\tau_1 - \tau_2)\tilde{\mu}_g.
\end{aligned}$$

Now, let us consider the last case where  $\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}$  is indefinite. Recall that the eigenvalues of  $\tilde{X}\tilde{S}$  are ordered increasingly, i.e.,  $\lambda_1 \leq \dots \leq \lambda_n$ . Assume  $\lambda_k$  is the first eigenvalue of  $\tilde{X}\tilde{S}$  such that  $\tau_1\tilde{\mu}_g - \lambda_k \leq 0$ , i.e.,  $\tau_1\tilde{\mu}_g - \lambda_1 \geq \dots \geq \tau_1\tilde{\mu}_g - \lambda_{k-1} > 0 \geq \tau_1\tilde{\mu}_g - \lambda_k \geq \dots \geq \tau_1\tilde{\mu}_g - \lambda_n$ . It is easy to see that

$$\begin{aligned}
\lambda_{\min}(\chi(\alpha)) &= \lambda_{\min}(\tilde{X}\tilde{S} + \alpha_-[\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^- + \alpha_+[\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^+) \\
&= \lambda_{\min}(Q(\Lambda + \alpha_-[\tau_1\tilde{\mu}_g I - \Lambda]^- + \alpha_+[\tau_1\tilde{\mu}_g I - \Lambda]^+)Q^T) \\
&= \min\{\lambda_1 + \alpha_+(\tau_1\tilde{\mu}_g - \lambda_1), \lambda_k + \alpha_-(\tau_1\tilde{\mu}_g - \lambda_k)\} \\
&\geq \min\{\tau_2\tilde{\mu}_g + \alpha_+(\tau_1 - \tau_2)\tilde{\mu}_g, \tau_1\tilde{\mu}_g\} \\
&\geq \tau_2\tilde{\mu}_g + \alpha_+(\tau_1 - \tau_2)\tilde{\mu}_g.
\end{aligned}$$

Taking all of the possible cases into account, we conclude that (5.13) is true.  $\blacksquare$

To follow the central path, we also need to make sure that the iterates remain in the prescribed neighborhood of the central path.

**Lemma 5.9** *Suppose  $P \in \mathcal{P}(X, S)$  and  $\chi(\alpha) = \tilde{X}\tilde{S} + \alpha_-[\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^- + \alpha_+[\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^+$ . If  $(\tilde{X}, \tilde{y}, \tilde{S}) \in \tilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$ , then we have*

$$\|[\tau_1\tilde{\mu}_g(\alpha)I - \chi(\alpha)]^+\|_F \leq (1 - \alpha_+)\hat{\eta}\beta\tau_1\tilde{\mu}_g(\alpha). \quad (5.14)$$

**Proof.** Assume that the eigenvalues of  $\tilde{X}\tilde{S}$  are ordered so that

$$\tau_1\tilde{\mu}_g - \lambda_1 \geq \tau_1\tilde{\mu}_g - \lambda_2 \geq \dots \geq \tau_1\tilde{\mu}_g - \lambda_{k-1} \geq 0 \geq \tau_1\tilde{\mu}_g - \lambda_k \geq \dots \geq \tau_1\tilde{\mu}_g - \lambda_n.$$

Now, let us consider the diagonal elements of  $\Lambda + \alpha_-[\tau_1\tilde{\mu}_g I - \Lambda]^- + \alpha_+[\tau_1\tilde{\mu}_g I - \Lambda]^+$ . For  $i = 1, \dots, k-1$ ,  $\lambda_i + \alpha_+(\tau_1\tilde{\mu}_g - \lambda_i) = (1 - \alpha_+)\lambda_i + \alpha_+\tau_1\tilde{\mu}_g$ , then

$$\begin{aligned} \tau_1\tilde{\mu}_g(\alpha) - (\lambda_i + \alpha_+(\tau_1\tilde{\mu}_g - \lambda_i)) &\leq \tau_1\tilde{\mu}_g(\alpha) - \frac{\tilde{\mu}_g(\alpha)}{\tilde{\mu}_g}(\lambda_i + \alpha_+(\tau_1\tilde{\mu}_g - \lambda_i)) \\ &= \frac{\tilde{\mu}_g(\alpha)}{\tilde{\mu}_g}(\tau_1\tilde{\mu}_g - (1 - \alpha_+)\lambda_i - \alpha_+\tau_1\tilde{\mu}_g) \\ &= \frac{\tilde{\mu}_g(\alpha)}{\tilde{\mu}_g}(1 - \alpha_+)(\tau_1\tilde{\mu}_g - \lambda_i). \end{aligned}$$

For  $i = k, \dots, n$ ,  $\lambda_i + \alpha_-(\tau_1\tilde{\mu}_g - \lambda_i) \geq \lambda_i + \tau_1\tilde{\mu}_g - \lambda_i = \tau_1\tilde{\mu}_g \geq 0$ , then

$$\tau_1\tilde{\mu}_g(\alpha) - (\lambda_i + \alpha_-(\tau_1\tilde{\mu}_g - \lambda_i)) \leq \tau_1\tilde{\mu}_g - \tau_1\tilde{\mu}_g = 0.$$

For convenience, let  $\varphi(\alpha) = [\tau_1\tilde{\mu}_g(\alpha)I - (\Lambda + \alpha_-[\tau_1\tilde{\mu}_g I - \Lambda]^- + \alpha_+[\tau_1\tilde{\mu}_g I - \Lambda]^+)]^+$ . Therefore, together with Lemma 3.2, we have

$$\begin{aligned} \|\varphi(\alpha)\|_F &\leq \frac{\tilde{\mu}_g(\alpha)}{\tilde{\mu}_g}(1 - \alpha_+) \left\| [\tau_1\tilde{\mu}_g I - \Lambda]^+ \right\|_F \\ &= \frac{\tilde{\mu}_g(\alpha)}{\tilde{\mu}_g}(1 - \alpha_+) \left\| [\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^+ \right\|_F \\ &\leq (1 - \alpha_+)\hat{\eta}\beta\tau_1\tilde{\mu}_g(\alpha). \end{aligned} \tag{5.15}$$

On the other hand, let  $\phi(\alpha) = [\tau_1\tilde{\mu}_g(\alpha)I - \chi(\alpha)]^+$ , then we have

$$\begin{aligned} \|\phi(\alpha)\|_F &= \left\| [\tau_1\tilde{\mu}_g(\alpha)I - (\tilde{X}\tilde{S} + \alpha_-[\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^- + \alpha_+[\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^+)]^+ \right\|_F \\ &= \left\| [\tau_1\tilde{\mu}_g(\alpha)I - (\Lambda + \alpha_-[\tau_1\tilde{\mu}_g I - \Lambda]^- + \alpha_+[\tau_1\tilde{\mu}_g I - \Lambda]^+)]^+ \right\|_F \\ &\leq (1 - \alpha_+)\hat{\eta}\beta\tau_1\tilde{\mu}_g(\alpha). \end{aligned}$$

The proof is completed. ■

The next two lemmas together bound the distance between the current iterate and our reference point  $\tau_1\tilde{\mu}_g I$  on the central path.

**Lemma 5.10** *Let  $X, S \in \mathcal{S}_{++}^n$ ,  $P \in \mathcal{P}(X, S)$ ,  $\tilde{X}$  and  $\tilde{S}$  are defined by (5.1), and  $\tilde{E}$  and  $\tilde{F}$  are defined by (5.8). Then,*

$$\left\| (\tilde{F}\tilde{E})^{-1/2} \mathbf{vec}([\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^-) \right\|^2 \leq \text{Tr}(\tilde{X}\tilde{S}). \tag{5.16}$$

**Proof.** Using Equation (5.8) and Proposition B.6, we find the spectral decompositions of  $\tilde{E}$  and  $\tilde{F}$  to be

$$\begin{aligned} \tilde{E} &= \frac{1}{2}(\tilde{S} \otimes I + I \otimes \tilde{S}) = \frac{1}{2}Q_K(\Lambda(\tilde{S}) \otimes I + I \otimes \Lambda(\tilde{S}))Q_K^T, \\ \tilde{F} &= \frac{1}{2}(\tilde{X} \otimes I + I \otimes \tilde{X}) = \frac{1}{2}Q_K(\Lambda(\tilde{X}) \otimes I + I \otimes \Lambda(\tilde{X}))Q_K^T, \end{aligned}$$

where  $Q_K = Q \otimes Q$  is an  $n^2 \times n^2$  orthogonal matrix. Furthermore, because  $\tilde{X}$  and  $\tilde{S}$  commute, from Proposition B.5, we have  $\tilde{F}\tilde{E} \in \mathcal{S}_{++}^{n^2}$ . Then, we have

$$(\tilde{F}\tilde{E})^{-1} = 4Q_K(\Lambda \otimes I + I \otimes \Lambda + \Lambda(\tilde{X}) \otimes \Lambda(\tilde{S}) + \Lambda(\tilde{S}) \otimes \Lambda(\tilde{X}))^{-1}Q_K^T,$$

where the matrix in the middle is diagonal with the properties that the  $((i-1)n+i)^{\text{th}}$  component is  $1/(4\lambda_i)$  and the largest component is  $1/(4\lambda_1)$ . On the other hand,

$$\begin{aligned} \mathbf{vec}(\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}) &= \mathbf{vec}(\tau_1\tilde{\mu}_g I - Q\Lambda Q^T) \\ &= (Q \otimes Q)\mathbf{vec}(\tau_1\tilde{\mu}_g I - \Lambda) \\ &= Q_K\mathbf{vec}(\tau_1\tilde{\mu}_g I - \Lambda), \end{aligned}$$

where  $\mathbf{vec}(\tau_1\tilde{\mu}_g I - \Lambda)$  is an  $n^2$ -vector with at most  $n$  nonzeros at the  $((i-1)n+i)^{\text{th}}$  entries which are equal to  $\tau_1\tilde{\mu}_g - \lambda_i$ . Finally, we have

$$\begin{aligned} \left\| (\tilde{F}\tilde{E})^{-1/2}\mathbf{vec}([\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^-) \right\|^2 &= (\mathbf{vec}([\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^-))^T (\tilde{E}\tilde{F})^{-1}\mathbf{vec}([\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^-) \\ &= \sum_{i=1}^n ([\tau_1\tilde{\mu}_g - \lambda_i]^-)^2 / \lambda_i \\ &= \sum_{i=1}^n ([\sqrt{\lambda_i} - \tau_1\tilde{\mu}_g / \sqrt{\lambda_i}]^+)^2 \\ &\leq \sum_{i=1}^n \lambda_i \\ &= \text{Tr}(\tilde{X}\tilde{S}), \end{aligned}$$

which leads to inequality (5.16). ■

**Lemma 5.11** *Let  $P \in \mathcal{P}(X, S)$ ,  $\tilde{X}$  and  $\tilde{S}$  be defined by (5.1), and  $\tilde{E}$  and  $\tilde{F}$  be defined by (5.8). If  $(\tilde{X}, \tilde{y}, \tilde{S}) \in \tilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$  and  $\beta \leq 1/4$ , then*

$$\left\| (\tilde{F}\tilde{E})^{-1/2}\mathbf{vec}([\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^+) \right\|^2 \leq \hat{\eta}^2 \beta \tau_1 \tilde{\mu}_g / 3.$$

**Proof.** Note the fact that  $\lambda_{\min}(\tilde{F}\tilde{E}) = \lambda_1 \geq \tau_2 \tilde{\mu}_g$ . Then one has

$$\begin{aligned} \left\| (\tilde{F}\tilde{E})^{-1/2}\mathbf{vec}([\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^+) \right\|^2 &\leq \left\| (\tilde{F}\tilde{E})^{-1/2} \right\|^2 \left\| \mathbf{vec}([\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^+) \right\|^2 \\ &= \left\| (\tilde{F}\tilde{E})^{-1/2} \right\|^2 \left\| [\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^+ \right\|_F^2 \\ &\leq \hat{\eta}^2 \beta^2 \tau_1^2 \tilde{\mu}_g^2 / (\tau_2 \tilde{\mu}_g) \\ &\leq \hat{\eta}^2 \beta \tau_1 \tilde{\mu}_g / 3. \end{aligned}$$

The last inequality follows from the fact that  $\beta \leq 1/4$  implies  $\beta\tau_1/\tau_2 \leq 1/3$ . ■

We now apply Lemma 5.10 and Lemma 5.11, together with Lemma B.4, to conclude the following result.

**Lemma 5.12** *Let  $P \in \mathcal{P}(X, S)$  and  $G = \tilde{E}^{-1}\tilde{F}$ . If  $(\tilde{X}, \tilde{y}, \tilde{S}) \in \tilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$  and  $\beta \leq 1/4$ , then*

$$\left\| G^{-1/2}\mathbf{vec}(\tilde{\Delta}\tilde{X}(\alpha)) \right\|^2 + \left\| G^{1/2}\mathbf{vec}(\tilde{\Delta}\tilde{S}(\alpha)) \right\|^2 + 2\text{Tr}(\tilde{\Delta}\tilde{X}(\alpha)\tilde{\Delta}\tilde{S}(\alpha)) \leq \alpha_-^2 \text{Tr}(\tilde{X}\tilde{S}) + \alpha_+^2 \hat{\eta}^2 \beta \tau_1 \tilde{\mu}_g / 3.$$

**Proof.** From (5.7), we have

$$\tilde{E}\mathbf{vec}(\widetilde{\Delta X}(\alpha)) + \tilde{F}\mathbf{vec}(\widetilde{\Delta S}(\alpha)) = \alpha_- \mathbf{vec}([\tau_1 \tilde{\mu}_g I - \tilde{X} \tilde{S}]^-) + \alpha_+ \mathbf{vec}([\tau_1 \tilde{\mu}_g I - \tilde{X} \tilde{S}]^+).$$

Applying Lemma B.4 to this equality, we obtain

$$\begin{aligned} & \left\| (\tilde{F}\tilde{E})^{-1/2} \tilde{E} \mathbf{vec}(\widetilde{\Delta X}(\alpha)) \right\|^2 + \left\| (\tilde{F}\tilde{E})^{-1/2} \tilde{F} \mathbf{vec}(\widetilde{\Delta S}(\alpha)) \right\|^2 + 2\mathrm{Tr}(\widetilde{\Delta X}(\alpha) \widetilde{\Delta S}(\alpha)) \\ &= \left\| (\tilde{F}\tilde{E})^{-1/2} [\alpha_- \mathbf{vec}([\tau_1 \tilde{\mu}_g I - \tilde{X} \tilde{S}]^-) + \alpha_+ \mathbf{vec}([\tau_1 \tilde{\mu}_g I - \tilde{X} \tilde{S}]^+)] \right\|^2. \end{aligned}$$

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

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

Hence, to complete the proof, it is sufficient to show that

$$\begin{aligned} & \left\| (\tilde{F}\tilde{E})^{-1/2} [\alpha_- \mathbf{vec}([\tau_1 \tilde{\mu}_g I - \tilde{X} \tilde{S}]^-) + \alpha_+ \mathbf{vec}([\tau_1 \tilde{\mu}_g I - \tilde{X} \tilde{S}]^+)] \right\|^2 \\ & \leq \alpha_-^2 \left\| (\tilde{F}\tilde{E})^{-1/2} \mathbf{vec}([\tau_1 \tilde{\mu}_g I - \tilde{X} \tilde{S}]^-) \right\|^2 + \alpha_+^2 \left\| (\tilde{F}\tilde{E})^{-1/2} \mathbf{vec}([\tau_1 \tilde{\mu}_g I - \tilde{X} \tilde{S}]^+) \right\|^2 \\ & \leq \alpha_-^2 \mathrm{Tr}(\tilde{X} \tilde{S}) + \alpha_+^2 \hat{\eta}^2 \beta \tau_1 \tilde{\mu}_g / 3, \end{aligned}$$

where the last inequality can be derived from Lemma 5.10 and 5.11.  $\blacksquare$

Using Lemma B.7, we can explore a bound for the second order term  $\widetilde{\Delta X}(\alpha) \widetilde{\Delta S}(\alpha)$ .

**Lemma 5.13** *Let  $P \in \mathcal{P}(X, S)$  and  $G = \tilde{E}^{-1} \tilde{F}$ . If  $\beta \leq 1/4$ ,  $\alpha_- = \alpha_+ \hat{\eta} \sqrt{\frac{\beta \tau_1}{n}}$  and  $(\tilde{X}, \tilde{y}, \tilde{S}) \in \tilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$ , then we have*

$$\left\| H_I(\widetilde{\Delta X}(\alpha) \widetilde{\Delta S}(\alpha)) \right\|_F \leq \left\| \mathbf{vec}(\widetilde{\Delta X}(\alpha)) \right\| \left\| \mathbf{vec}(\widetilde{\Delta S}(\alpha)) \right\| \leq \frac{2}{3} \sqrt{\mathrm{cond}(G)} \alpha_+^2 \hat{\eta}^2 \beta \tau_1 \tilde{\mu}_g. \quad (5.17)$$

**Proof.** From the last inequality in Lemma B.9, we have

$$\begin{aligned} \left\| H_I(\widetilde{\Delta X}(\alpha) \widetilde{\Delta S}(\alpha)) \right\|_F & \leq \left\| \widetilde{\Delta X}(\alpha) \widetilde{\Delta S}(\alpha) \right\|_F \\ & \leq \left\| \widetilde{\Delta X}(\alpha) \right\|_F \left\| \widetilde{\Delta S}(\alpha) \right\|_F \\ & = \left\| \mathbf{vec}(\widetilde{\Delta X}(\alpha)) \right\| \left\| \mathbf{vec}(\widetilde{\Delta S}(\alpha)) \right\|. \end{aligned}$$

Further because of Lemma 5.2 and Lemma B.7, it follows that

$$\begin{aligned} \left\| H_I(\widetilde{\Delta X}(\alpha) \widetilde{\Delta S}(\alpha)) \right\|_F & \leq \left\| \mathbf{vec}(\widetilde{\Delta X}(\alpha)) \right\| \left\| \mathbf{vec}(\widetilde{\Delta S}(\alpha)) \right\| \\ & \leq \frac{\sqrt{\mathrm{cond}(G)}}{2} \left( \left\| G^{-1/2} \mathbf{vec}(\widetilde{\Delta X}(\alpha)) \right\|^2 + \left\| G^{1/2} \mathbf{vec}(\widetilde{\Delta S}(\alpha)) \right\|^2 \right) \\ & \leq \frac{\sqrt{\mathrm{cond}(G)}}{2} (\alpha_-^2 \mathrm{Tr}(\tilde{X} \tilde{S}) + \alpha_+^2 \hat{\eta}^2 \beta \tau_1 \tilde{\mu}_g / 3). \end{aligned}$$

Substitute  $\alpha_-$  with  $\alpha_+ \hat{\eta} \sqrt{\frac{\beta \tau_1}{n}}$  and apply Lemma 5.12, then we eventually obtain

$$\begin{aligned} \left\| H_I(\widetilde{\Delta X}(\alpha) \widetilde{\Delta S}(\alpha)) \right\|_F & \leq \frac{\sqrt{\mathrm{cond}(G)}}{2} (\alpha_+^2 \hat{\eta}^2 \beta \tau_1 n \tilde{\mu}_g / n + \alpha_+^2 \hat{\eta}^2 \beta \tau_1 \tilde{\mu}_g / 3) \\ & = \frac{2}{3} \sqrt{\mathrm{cond}(G)} \alpha_+^2 \hat{\eta}^2 \beta \tau_1 \tilde{\mu}_g, \end{aligned}$$

observing that  $\text{Tr}(\tilde{X}\tilde{S}) = n\tilde{\mu}_g$ . ■

The next proposition gives out a sufficient condition which guarantees all the iterates in the neighborhood  $\mathcal{N}(\tau_1, \tau_2, \eta)$ .

**Proposition 5.14** *Let  $(X, y, S) \in \mathcal{N}(\tau_1, \tau_2, \eta)$ ,  $\tau_1 < 4/9$ ,  $\beta \leq 1/4$ ,  $P \in \mathcal{P}(X, S)$  and  $G = \tilde{E}^{-1}\tilde{F}$ . If  $\alpha_- = \alpha_+\hat{\eta}\sqrt{\beta\tau_1/n}$  and  $\alpha_+ \leq 1/(\sqrt{\text{cond}(G)}\hat{\eta}^2)$ , then*

$$(X(\alpha), y(\alpha), S(\alpha)) \in \mathcal{N}(\tau_1, \tau_2, \eta).$$

**Proof.** By Corollary 5.7 we have  $\tilde{\mu}_g(\alpha) \leq \tilde{\mu}_g$ . Further, using Lemma 5.8 and the fact that  $\lambda_{\min}(\cdot)$  is a homogeneous concave function on the space of symmetric matrices, one has

$$\begin{aligned} \lambda_{\min}(H_I(\tilde{X}(\alpha)\tilde{S}(\alpha))) &\geq \lambda_{\min}(H_I(\tilde{X}\tilde{S} + \alpha_-[\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^- + \alpha_+[\tau_1\tilde{\mu}_g I - \tilde{X}\tilde{S}]^+)) \\ &\quad + \lambda_{\min}(H_I(\tilde{\Delta}\tilde{X}(\alpha)\tilde{\Delta}\tilde{S}(\alpha))) \\ &\geq \lambda_{\min}(\chi(\alpha)) - \left\| H_I(\tilde{\Delta}\tilde{X}(\alpha)\tilde{\Delta}\tilde{S}(\alpha)) \right\| \\ &\geq \tau_2\tilde{\mu}_g + \alpha_+(\tau_1 - \tau_2)\tilde{\mu}_g - \left\| H_I(\tilde{\Delta}\tilde{X}(\alpha)\tilde{\Delta}\tilde{S}(\alpha)) \right\|_F, \end{aligned}$$

where the second inequality follows from the fact that the absolute value of the smallest eigenvalue of a symmetric matrix is less than or equal to its Euclidean norm, and the third inequality is due to the fact that the Euclidean norm of a symmetric matrix is less than or equal to its Frobenius norm. These two results follow directly from Lemma B.8.

From Lemma 5.13 one can further derive that

$$\begin{aligned} \lambda_{\min}(H_I(\tilde{X}(\alpha)\tilde{S}(\alpha))) &\geq \tau_2\tilde{\mu}_g + \alpha_+(\tau_1 - \tau_2)\tilde{\mu}_g - \frac{2}{3}\sqrt{\text{cond}(G)}\alpha_+^2\hat{\eta}^2\beta\tau_1\tilde{\mu}_g \\ &\geq \tau_2\tilde{\mu}_g + \alpha_+\beta\tau_1\tilde{\mu}_g - \alpha_+\beta\tau_1\tilde{\mu}_g \\ &= \tau_2\tilde{\mu}_g \\ &\geq \tau_2\tilde{\mu}_g(\alpha) \\ &> 0. \end{aligned}$$

This reveals that  $\tilde{X}(\alpha)\tilde{S}(\alpha)$  is nonsingular, and further implies that each of the factors  $\tilde{X}(\alpha)$  and  $\tilde{S}(\alpha)$  are nonsingular as well. By using continuity, it follows that  $\tilde{X}(\alpha)$  and  $\tilde{S}(\alpha)$  are also in  $\mathcal{S}_{++}^n$ , since  $\tilde{X}, \tilde{S} \succ 0$ . Then, we claim that

$$\lambda_{\min}(\tilde{X}(\alpha)\tilde{S}(\alpha)) \geq \lambda_{\min}(H_I(\tilde{X}(\alpha)\tilde{S}(\alpha))) \geq \tau_2\tilde{\mu}_g(\alpha). \quad (5.18)$$

Since  $\beta \leq 1/4$  and  $\tau_1 \leq 4/9$ , from Lemma 5.4, we have

$$\tilde{\mu}_g(\alpha) \geq (1 - \alpha_-)\tilde{\mu}_g = (1 - \alpha_+\hat{\eta}\sqrt{\beta\tau_1/n})\tilde{\mu}_g \geq (1 - \sqrt{\beta\tau_1})\tilde{\mu}_g \geq \frac{2}{3}\tilde{\mu}_g. \quad (5.19)$$

From Lemma 3.3, we have

$$\begin{aligned} \psi(\alpha) &:= \left\| [\tau_1\tilde{\mu}_g(\alpha)I - \tilde{X}^{1/2}(\alpha)\tilde{S}(\alpha)\tilde{X}^{1/2}(\alpha)]^+ \right\|_F \\ &\leq \left\| [H_{\tilde{X}^{1/2}(\alpha)}(\tau_1\tilde{\mu}_g(\alpha)I - \tilde{X}^{1/2}(\alpha)\tilde{S}(\alpha)\tilde{X}^{1/2}(\alpha))]^+ \right\|_F \\ &= \left\| [H_I(\tau_1\tilde{\mu}_g(\alpha)I - \tilde{X}(\alpha)\tilde{S}(\alpha))]^+ \right\|_F. \end{aligned}$$

Because  $\tilde{X}(\alpha)\tilde{S}(\alpha) = (\tilde{X} + \alpha_-\widetilde{\Delta X}_- + \alpha_+\widetilde{\Delta X}_+)(\tilde{X} + \alpha_-\widetilde{\Delta X}_- + \alpha_+\widetilde{\Delta X}_+)$ , we have

$$\begin{aligned}\psi(\alpha) &\leq \left\| [H_I(\tau_1\tilde{\mu}_g(\alpha)I - \tilde{X}\tilde{S} - \alpha_-[\tau\tilde{\mu}_gI - \tilde{X}\tilde{S}]^- - \alpha_+[\tau\tilde{\mu}_gI - \tilde{X}\tilde{S}]^+ - \widetilde{\Delta X}(\alpha)\widetilde{\Delta S}(\alpha))]^+ \right\|_F \\ &\leq \left\| [H_I(\tau_1\tilde{\mu}_g(\alpha)I - \tilde{X}\tilde{S} - \alpha_-[\tau\tilde{\mu}_gI - \tilde{X}\tilde{S}]^- - \alpha_+[\tau\tilde{\mu}_gI - \tilde{X}\tilde{S}]^+)]^+ \right\|_F + \\ &\quad \left\| [-H_I(\widetilde{\Delta X}(\alpha)\widetilde{\Delta S}(\alpha))]^+ \right\|_F \\ &= \left\| [\tau_1\tilde{\mu}_g(\alpha)I - \tilde{X}\tilde{S} - \alpha_-[\tau\tilde{\mu}_gI - \tilde{X}\tilde{S}]^- - \alpha_+[\tau\tilde{\mu}_gI - \tilde{X}\tilde{S}]^+]^+ \right\|_F + \\ &\quad \left\| [H_I(\widetilde{\Delta X}(\alpha)\widetilde{\Delta S}(\alpha))]^- \right\|_F,\end{aligned}$$

where the second inequality is from Proposition 3.1.

Using the fact that  $\left\| [H_I(\widetilde{\Delta X}(\alpha)\widetilde{\Delta S}(\alpha))]^- \right\|_F \leq \left\| H_I(\widetilde{\Delta X}(\alpha)\widetilde{\Delta S}(\alpha)) \right\|_F$  and Lemma 5.9, we can prove

$$\psi(\alpha) \leq (1 - \alpha_+)\hat{\eta}\beta\tau_1\tilde{\mu}_g(\alpha) + \left\| H_I(\widetilde{\Delta X}(\alpha)\widetilde{\Delta S}(\alpha)) \right\|_F.$$

Further, from Lemma 5.13 and inequality (5.19), one has

$$\begin{aligned}\psi(\alpha) &\leq (1 - \alpha_+)\hat{\eta}\beta\tau_1\tilde{\mu}_g(\alpha) + \frac{2}{3}\sqrt{\text{cond}(G)}\alpha_+^2\hat{\eta}\beta\tau_1\tilde{\mu}_g \\ &\leq (1 - \alpha_+)\hat{\eta}\beta\tau_1\tilde{\mu}_g(\alpha) + \sqrt{\text{cond}(G)}\alpha_+^2\hat{\eta}\beta\tau_1\tilde{\mu}_g(\alpha).\end{aligned}$$

Since  $\alpha_+ \leq 1/(\sqrt{\text{cond}(G)}\hat{\eta}^2)$  and  $\tilde{\eta} \geq 1$ , we have  $\sqrt{\text{cond}(G)}\alpha_+^2\hat{\eta}\beta\tau_1\tilde{\mu}_g(\alpha) \leq \alpha_+\hat{\eta}\beta\tau_1\tilde{\mu}_g(\alpha)$ . Thus,

$$\begin{aligned}\psi(\alpha) &\leq (1 - \alpha_+)\hat{\eta}\beta\tau_1\tilde{\mu}_g(\alpha) + \alpha_+\hat{\eta}\beta\tau_1\tilde{\mu}_g(\alpha) \\ &= \hat{\eta}\beta\tau_1\tilde{\mu}_g(\alpha) \\ &\leq \eta\beta\tau_1\tilde{\mu}_g(\alpha) \\ &= \eta(\tau_1 - \tau_2)\tilde{\mu}_g(\alpha).\end{aligned}$$

This, together with (5.18), implies that

$$(\tilde{X}(\alpha), \tilde{y}(\alpha), \tilde{S}(\alpha)) \in \tilde{\mathcal{N}}(\tau_1, \tau_2, \eta).$$

Consequently, according to Proposition (5.1), one has

$$(X(\alpha), y(\alpha), S(\alpha)) \in \mathcal{N}(\tau_1, \tau_2, \eta). \quad \blacksquare$$

### 5.3 Polynomial Complexity

In this section we present our main complexity result. The next theorem gives an iteration-complexity bound for Algorithm 1 in terms of a parameter  $\kappa_\infty$  defined as

$$\kappa_\infty = \sup \left\{ \text{cond}((\tilde{E}^k)^{-1}\tilde{F}^k) : k = 0, 1, \dots \right\}. \quad (5.20)$$

Obviously,  $\kappa_\infty \geq 1$ .

**Theorem 5.15** *Suppose that  $\kappa_\infty \leq \infty$ ,  $\eta \geq 1$ ,  $0 < \tau_2 < \tau_1 \leq 4/9$ , and  $\beta \leq 1/4$  are fixed parameters. At each iteration, let  $P^k \in \mathcal{P}(X^k, S^k)$ . Then Algorithm 1 will terminate in  $O(\eta\sqrt{\kappa_\infty n} \log \frac{\text{Tr}(X^0 S^0)}{\epsilon})$  iterations with a solution  $\text{Tr}(XS) \leq \epsilon$ .*

**Proof.** In every iteration, let  $\hat{\alpha} = (\sqrt{\beta\tau_1}/(\kappa_\infty n)/\hat{\eta}, 1/(\sqrt{\kappa_\infty}\hat{\eta}^2))$ . By Proposition 5.14, we have

$$(X(\hat{\alpha}), y(\hat{\alpha}), S(\hat{\alpha})) \in \mathcal{N}(\tau_1, \tau_2, \eta).$$

Furthermore, from Lemma 5.7, we also conclude

$$\tilde{\mu}_g(\alpha) \leq \left(1 - \frac{2\sqrt{\beta\tau_1}}{9\hat{\eta}\sqrt{\text{cond}(G)n}}\right) \tilde{\mu}_g \leq \left(1 - \frac{2\sqrt{\beta\tau_1}}{9\eta\sqrt{\text{cond}(G)n}}\right) \tilde{\mu}_g \leq \left(1 - \frac{2\sqrt{\beta\tau_1}}{9\eta\sqrt{\kappa_\infty n}}\right) \tilde{\mu}_g,$$

from which the statement of the theorem follows. ■

From Theorem 5.15, it is easy to present various iteration complexities of Algorithm 1 in terms of some specific aforementioned scaling matrices  $P$ .

**Corollary 5.16** *If the parameter  $\eta$  is a constant, then for Algorithm 1, when it is based on the NT direction, the iteration-complexity bound is  $O(\sqrt{n} \log \frac{\text{Tr}(X^0 S^0)}{\epsilon})$ . When the H..K..M scaling is used, then Algorithm 1 terminates in at most  $O(n \log \frac{\text{Tr}(X^0 S^0)}{\epsilon})$  iterations.*

**Corollary 5.17** *If the parameter  $\eta$  is in the order of  $\sqrt{n}$ , then for Algorithm 1, when it is based on the NT direction, the iteration-complexity bound is  $O(n \log \frac{\text{Tr}(X^0 S^0)}{\epsilon})$ . When the H..K..M scaling is used, then Algorithm 1 terminates in at most  $O(n^{3/2} \log \frac{\text{Tr}(X^0 S^0)}{\epsilon})$  iterations.*

From Lemma B.3, Corollaries 5.16 and 5.17 are readily achieved.

As we see, when  $\eta$  is a constant and the NT scaling is used, Algorithm 1 achieves its best complexity bound which coincides with the best known complexity of IPMs for SDO. When  $\eta$  is in the order of  $\sqrt{n}$ , our complexity result is the same as the one for classical large neighborhood IPMs, since we have shown in Proposition 3.6 that in that case our neighborhood  $\mathcal{N}(\tau_1, \tau_2, \eta)$  is exactly the large neighborhood  $\mathcal{N}(1 - \tau_2)$ .

## 6 Preliminary Numerical Experiments

We compare our algorithm, Algorithm 1, with classical large neighborhood IPMs without using Mehrotra's predictor-corrector strategy. The comparison is carried out by using SDPT3 [30], which is capable to simultaneously scale the primal variable  $X$  and the dual variable  $S$  to diagonal matrices when NT scaling is used. Moreover, SDPT3 does not require too many modifications to implement our algorithm.

We have to point out that our implementation differs from Algorithm 1 in the same ways that typical implementations of IPMs differ from theory. First, we use an infeasible starting point and reduce infeasibility together with the duality gap. Second, we do not explicitly force the iterates to stay in the neighborhood  $\mathcal{N}(\tau_1, \tau_2, \eta)$ , because it is computationally expensive. However, we do check positive semidefiniteness of the primal and dual variables at every iteration.

For this computational experience we implement Algorithm 1 using NT scaling. In this case, the computation of the positive and negative parts  $[\tau\mu_g I - H_P(XS)]^\pm$  can be



Table 1: Classical IPM in SDPT3 vs Algorithm 1 ( $\alpha_- = 1.2\alpha_+$ )

| problem | Classical IPM |      |                               |       |        |          |          |          |       |        | ALG 1    |          |          |  |
|---------|---------------|------|-------------------------------|-------|--------|----------|----------|----------|-------|--------|----------|----------|----------|--|
|         | $m$           | $n$  | iter                          | time  | per    | gap      | relgap   | iter     | time  | per    | gap      | relgap   |          |  |
| DIMACS  | bm1           | 883  | 882                           | 34    | 173.75 | 5.11     | 1.02e-02 | 2.13e-04 | 29    | 171.93 | 5.93     | 1.69e-01 | 3.53e-03 |  |
|         | copo14        | 1275 | $14 \times 14^a$              | 23    | 2.76   | 0.12     | 3.69e-09 | 3.69e-09 | 55    | 8.00   | 0.15     | 2.86e-12 | 2.86e-12 |  |
|         | hamming_7_5_6 | 1793 | 128                           | 16    | 8.40   | 0.52     | 2.63e-07 | 3.05e-09 | 16    | 9.32   | 0.58     | 1.17e-07 | 1.35e-09 |  |
|         | hamming_9_8   | 2305 | 512                           | 18    | 42.44  | 2.36     | 6.92e-07 | 1.54e-09 | 19    | 48.41  | 2.55     | 9.51e-07 | 2.12e-09 |  |
|         | minphase      | 48   | 48                            | 22    | 0.53   | 0.02     | 4.99e-02 | 3.84e-03 | 24    | 0.61   | 0.03     | 8.49e-03 | 6.55e-04 |  |
|         | torusg3-8     | 512  | 512                           | 26    | 53.77  | 2.07     | 3.55e-01 | 3.68e-09 | 20    | 42.38  | 2.12     | 9.57e-01 | 9.90e-09 |  |
|         | toruspm3-8-50 | 512  | 512                           | 19    | 38.70  | 2.04     | 5.59e-06 | 5.29e-09 | 19    | 39.45  | 2.08     | 5.70e-06 | 5.39e-09 |  |
|         | truss8        | 496  | $33 \times 19 + 1 \times 1^b$ | 27    | 4.75   | 0.18     | 9.82e-07 | 3.67e-09 | 26    | 4.68   | 0.18     | 3.46e-07 | 1.29e-09 |  |
|         | arch8         | 174  | 335                           | 56    | 8.71   | 0.16     | 9.10e-08 | 6.02e-09 | 50    | 8.56   | 0.17     | 5.40e-08 | 3.58e-09 |  |
|         | equalG11      | 801  | 801                           | 22    | 85.20  | 3.87     | 1.04e-05 | 8.26e-09 | 22    | 99.97  | 4.54     | 1.23e-05 | 9.78e-09 |  |
| SDPLIB  | gpp250-4      | 250  | 250                           | 18    | 3.04   | 0.17     | 7.93e-06 | 5.30e-09 | 17    | 3.23   | 0.19     | 1.41e-05 | 9.46e-09 |  |
|         | gpp500-4      | 501  | 500                           | 22    | 23.75  | 1.08     | 4.75e-04 | 1.52e-07 | 20    | 24.13  | 1.21     | 5.89e-06 | 1.88e-09 |  |
|         | maxG11        | 800  | 800                           | 22    | 64.73  | 2.94     | 9.81e-06 | 7.79e-09 | 21    | 66.38  | 3.16     | 3.14e-06 | 2.49e-09 |  |
|         | maxG51        | 1000 | 1000                          | 28    | 202.17 | 7.22     | 5.17e-05 | 6.45e-09 | 27    | 208.59 | 7.73     | 5.49e-05 | 6.85e-09 |  |
|         | mcp250-1      | 250  | 250                           | 24    | 2.51   | 0.10     | 7.58e-07 | 1.19e-09 | 21    | 2.63   | 0.13     | 3.22e-06 | 5.06e-09 |  |
|         | mcp500-1      | 500  | 500                           | 24    | 13.67  | 0.57     | 4.39e-06 | 3.66e-09 | 25    | 15.88  | 0.64     | 1.45e-06 | 1.21e-09 |  |
|         | qpG11         | 800  | 1600                          | 23    | 68.31  | 2.97     | 8.59e-06 | 1.75e-09 | 22    | 70.68  | 3.21     | 1.90e-05 | 3.87e-09 |  |
|         | ss30          | 132  | 426                           | 51    | 45.88  | 0.90     | 4.88e-08 | 1.18e-09 | 47    | 44.25  | 0.94     | 3.28e-07 | 7.91e-09 |  |
|         | theta3        | 1106 | 150                           | 20    | 4.02   | 0.20     | 5.54e-07 | 6.49e-09 | 20    | 4.30   | 0.22     | 1.65e-07 | 1.94e-09 |  |
|         | theta4        | 1949 | 200                           | 21    | 14.00  | 0.67     | 1.80e-07 | 1.77e-09 | 20    | 14.87  | 0.74     | 2.42e-07 | 2.39e-09 |  |
| theta5  | 3028          | 250  | 20                            | 40.98 | 2.05   | 6.35e-07 | 5.50e-09 | 20       | 44.28 | 2.21   | 7.84e-07 | 6.79e-09 |          |  |

<sup>a</sup> The problem has 14 SDP blocks whose size is 14. In addition, the problem has 364 linear constraints as well.

<sup>b</sup> The problem has 33 SDP blocks whose size is 19 and 1 SDP block whose size is 1.

done cheaply as mentioned on page 12. The numerical results confirm that the average CPU time per iteration increases only marginally. Furthermore, we do not solve problem (4.10) to decide the step lengths. In our implementation, we fix the ratio between  $\alpha_-$  and  $\alpha_+$ . According to our experiments, a ratio slightly bigger than one, say, 1.20, gives the smallest number of iterations for most of the problems in the benchmark sets of DIMACS [24] and SDPLIB [7]. Nevertheless, there does exist problems, e.g., *hinf12* from DIMACS, for which the best result is obtained for a ratio of  $\alpha_-$  and  $\alpha_+$  that is smaller than one, say, 0.9.

All of our tests are run on an Intel Core i7 PC (2.67GHz) with 4G RAM under Windows SP3 and MATLAB R2007(a). Table 1 lists the names of the test problems, the number of iterations, the total CPU time, the average CPU time per iteration, the duality gap when the algorithm terminates and the relative duality gap defined as

$$\text{relgap} = \frac{\text{duality gap}}{1 + |\text{Tr}(\text{CX})| + |\mathbf{b}^T \mathbf{y}|}.$$

Overall, in term of the number of iterations, Algorithm 1 appears to be as good as classical large neighborhood IPMs without using predictor-corrector strategy, even a little better for all of the tested problems except *copo14*. The computational experiments reported in Table 1 confirm that Algorithm 1, even with a relatively naive implementation, is efficient and accurate, regardless of the peripheral increased computational cost of a single iteration.

## 7 Conclusions and Further Works

As discussed previously, this paper provides a new large neighborhood path-following algorithm with the same theoretical complexity bound as the best short-step path-following algorithm. Our preliminary implementation also provides us an encouraging evidence that our new algorithm may also perform well in practice. However, there are still some unsettled issues for implementation. For example, sophisticated and efficient strategies to choose step lengths deserve more work for real applications of the algorithm. Another issue of interest is how to efficiently compute the positive and negative part of the right-hand-side in the Newton equation when we do not use the NT scaling. It is apparent that, due to the high computational cost, explicitly computing them with eigenvalue decomposition is not desirable. To develop computationally efficient implementation strategies for our large neighborhood algorithm remains the subject of further research.

Because we were successful in combining Ai and Zhang's work [2] with LO and our work with SDO, a natural question to ask is whether this applies to second order cone optimization (SOCO) problems and further to general conic optimization. At last, we are also curious about the relationship among the classical IPMs, the self-regular IPMs and our algorithm.

## Acknowledgements

We would like to thank the anonymous referees and associate editor for their constructive comments that help us improve the presentation of the paper. We also acknowledge the

support of the Natural Sciences and Engineering Research Council of Canada (NSERC), Canada Research Chairs (CRC), Lehigh University, and Advanced Optimization Laboratory (AdvOL), McMaster University.

# APPENDIX

## A Some Properties of the Kronecker product

The Kronecker product of two matrices  $G \in \mathcal{R}^{m \times n}$  and  $K \in \mathcal{R}^{p \times q}$  is denoted by  $G \otimes K$  and is defined to be the block matrix

$$G \otimes K = \begin{bmatrix} g_{11}K & \cdots & g_{1n}K \\ \vdots & \ddots & \vdots \\ g_{m1}K & \cdots & g_{mn}K \end{bmatrix} \in \mathcal{R}^{mp \times nq}.$$

With each matrix  $Q \in \mathcal{R}^{m \times n}$ , we associate the vector  $\mathbf{vec}(Q) \in \mathcal{R}^{m \times n}$  defined by

$$\mathbf{vec}(Q) = [q_{11}, \cdots, q_{m1}, q_{12}, \cdots, q_{m2}, \cdots, q_{1n}, \cdots, q_{mn}]^T.$$

We present some useful properties of the Kronecker products.

1.  $(G \otimes K)\mathbf{vec}(H) = \mathbf{vec}(KHG^T)$ .
2.  $(G \otimes K)^T = G^T \otimes K^T$ .
3.  $G \otimes I$  is symmetric if and only if  $G$  is symmetric.
4.  $(G \otimes K)^{-1} = G^{-1} \otimes K^{-1}$ .
5.  $(G \otimes K)(H \otimes L) = GH \otimes KL$ .
6. If  $\Lambda(G) = \text{diag}(\lambda_i)$  and  $\Lambda(K) = \text{diag}(\mu_j)$ , then  $\Lambda(G \otimes K) = \text{diag}(\lambda_i \mu_j)$ . If  $q_i$  and  $r_j$  are the eigenvectors corresponding to the eigenvalues  $\lambda_i$  and  $\mu_j$  of  $G$  and  $K$ , then  $\mathbf{vec}(r_j q_i^T)$  is the eigenvector corresponding to the eigenvalue  $\lambda_i \mu_j$  of  $G \otimes K$ .
7.  $\mathbf{vec}(G)^T \mathbf{vec}(K) = \text{Tr}(GK)$ .

## B Some Properties of Square and Symmetric Matrices

**Theorem B.1 (Schur Triangulation)** *Given  $Q \in \mathcal{R}^{n \times n}$ , there is a unitary matrix  $U \in \mathcal{R}^{n \times n}$  such that*

$$UQU^* = \Lambda(Q) + N,$$

where  $N$  is a strictly upper triangular matrix.

**Proof.** For the proof, see Horn and Johnson [12], page 79. ■

**Lemma B.2** Suppose  $B = A + \tau cc^T$ , where  $A \in \mathcal{S}^n$  and  $c \in \mathcal{R}^n$  is a unit vector. Let  $\lambda_i(A)$  and  $\lambda_i(B)$  denote the  $i^{\text{th}}$  largest eigenvalues of  $A$  and  $B$ , respectively, i.e.,

$$\begin{aligned}\lambda_1(A) &\leq \lambda_2(A) \leq \cdots \leq \lambda_{n-1}(A) \leq \lambda_n(A), \\ \lambda_1(B) &\leq \lambda_2(B) \leq \cdots \leq \lambda_{n-1}(B) \leq \lambda_n(B).\end{aligned}$$

Then there exists nonnegative numbers  $\delta_1, \dots, \delta_n$  such that

$$\lambda_i(B) = \lambda_i(A) + \delta_i \tau, \quad i = 1, \dots, n$$

with  $\delta_1 + \cdots + \delta_n = 1$ .

**Proof.** For the proof, see Golub and Van Loan [10], page 412. ■

**Lemma B.3** Let  $\kappa_\infty$  be defined by (5.20), then

- if for all  $k$  the scaling matrix  $P^k = (W_{NT}^k)^{1/2}$ , then  $\kappa_\infty = 1$ ;
- if for all  $k$  the scaling matrix  $P^k = (S^k)^{1/2}$ , then  $\kappa_\infty \leq \frac{n}{r_2}$ ;
- if for all  $k$  the scaling matrix  $P^k = (X^k)^{-1/2}$ , then  $\kappa_\infty \leq \frac{n}{r_2}$ .

**Proof.** For the proof, we refer to Lemma 3.1 in Monteiro and Zhang's paper [20], page 289. ■

The following technical lemma was first introduced and proved in Zhang [35].

**Lemma B.4** Let  $u, v, r \in \mathcal{R}^n$  and  $Q, R \in \mathcal{R}^{n \times n}$  satisfying  $Qu + Rv = r$ . If  $RQ^T \in \mathcal{S}_{++}^n$  then

$$\|(RQ^T)^{-1/2}Qu\|^2 + \|(RQ^T)^{-1/2}Rv\|^2 + 2u^T v = \|(RQ^T)^{-1/2}r\|^2. \quad (\text{B.1})$$

**Proof.** For the proof, we refer to Proposition 1.1 in Zhang's paper [35], page 367. ■

To utilize Lemma B.4, we need to explore the conditions under which  $\tilde{F}\tilde{E}^T \in \mathcal{S}_{++}^n$ , where  $\tilde{F}$  and  $\tilde{E}$  is defined by (5.8). In [29] and [20], the authors state the same necessary and sufficient condition for  $\tilde{F}\tilde{E}^T \in \mathcal{S}_{++}^n$  but in different formats. In our paper, we utilize the proposition stated in [20]. For those who are interested in the proof, they are advised to consult Proposition 4.1 of Monteiro and Zhang [20].

**Proposition B.5** Let  $X, S \in \mathcal{S}_{++}^n$ ,  $\tilde{X}$  and  $\tilde{S}$  be defined by (5.1), and  $\tilde{E}$  and  $\tilde{F}$  be defined by (5.8). Then

- (i)  $\tilde{E}, \tilde{F} \in \mathcal{S}_{++}^{n^2}$ , and thus  $\tilde{F}\tilde{E}^T = \tilde{F}\tilde{E}$ ;
- (ii)  $\tilde{F}\tilde{E} \in \mathcal{S}^{n^2}$  if and only if  $\tilde{X}\tilde{S} \in \mathcal{S}^n$ ;
- (iii)  $\tilde{F}\tilde{E} \in \mathcal{S}^{n^2}$  implies  $\tilde{F}\tilde{E} \in \mathcal{S}_{++}^{n^2}$ .

We also use the next two results which are presented by Proposition 4.2 and Lemma 4.6 in Monteiro and Zhang [20].

**Proposition B.6** *For any  $P \in \mathcal{P}(X, S)$ , there exists an orthogonal matrix  $Q$  and diagonal matrices  $\Lambda(\tilde{X})$  and  $\Lambda(\tilde{S})$  such that:*

- (i)  $\tilde{X} = PXP = Q\Lambda(\tilde{X})Q^T$ ;
- (ii)  $\tilde{S} = P^{-1}SP^{-1} = Q\Lambda(\tilde{S})Q^T$ ;
- (iii)  $\Lambda = \Lambda(\tilde{X})\Lambda(\tilde{S})$ , and hence  $\tilde{X}\tilde{S} = \tilde{S}\tilde{X} = Q\Lambda Q^T$ .

**Lemma B.7** *For any  $u, v \in \mathcal{R}^n$  and  $G \in \mathcal{S}_{++}^n$ , we have*

$$\|u\| \|v\| \leq \frac{\sqrt{\text{cond}(G)}}{2} \left( \|G^{-1/2}u\|^2 + \|G^{1/2}v\|^2 \right). \quad (\text{B.2})$$

The remaining properties are from Lemmas 3.1, 3.2, and 3.3 of Monteiro [17], respectively. We use these results throughout this paper.

**Lemma B.8** *For any  $Q \in \mathcal{S}^n$ , we have*

$$\lambda_{\max}(Q) = \max_{\|u\|=1} u^T Q u, \quad (\text{B.3})$$

$$\lambda_{\min}(Q) = \min_{\|u\|=1} u^T Q u, \quad (\text{B.4})$$

$$\|Q\| = \max_{i=1, \dots, n} |\lambda_i(Q)|, \quad (\text{B.5})$$

$$\|Q\|_F^2 = \sum_{i=1}^n |\lambda_i(Q)|^2. \quad (\text{B.6})$$

**Lemma B.9** *For any  $Q \in \mathcal{R}^{n \times n}$  the following relations hold:*

$$\max_{i=1, \dots, n} \text{Re}[(\lambda_i(Q))] \leq \frac{1}{2} \lambda_{\max}(Q + Q^T), \quad (\text{B.7})$$

$$\min_{i=1, \dots, n} \text{Re}[(\lambda_i(Q))] \geq \frac{1}{2} \lambda_{\min}(Q + Q^T), \quad (\text{B.8})$$

$$\sum_{i=1}^n |\lambda_i(Q)|^2 \leq \|Q\|_F^2 = \|Q^T\|_F^2, \quad (\text{B.9})$$

$$\lambda_{\max}(Q^T Q) = \|Q^T Q\| = \|Q\|^2 = \|Q^T\|^2, \quad (\text{B.10})$$

$$\|Q\|_F \geq \|(Q + Q^T)/2\|_F. \quad (\text{B.11})$$

**Lemma B.10** *Let  $W \in \mathcal{R}^{n \times n}$  be a nonsingular matrix. Then, for any  $Q \in \mathcal{S}^n$ , we have*

$$\lambda_{\max}(Q) \leq \frac{1}{2} \lambda_{\max}(WQW^{-1} + (WQW^{-1})^T), \quad (\text{B.12})$$

$$\lambda_{\min}(Q) \geq \frac{1}{2} \lambda_{\min}(WQW^{-1} + (WQW^{-1})^T), \quad (\text{B.13})$$

$$\|Q\| \leq \frac{1}{2} \|(WQW^{-1} + (WQW^{-1})^T)\|, \quad (\text{B.14})$$

$$\|Q\|_F \leq \frac{1}{2} \|(WQW^{-1} + (WQW^{-1})^T)\|_F. \quad (\text{B.15})$$

## References

- [1] W. Ai. Neighborhood-following algorithms for linear programming. *Sci. China Ser. A*, 47:812–820, 2004.
- [2] W. Ai and S. Zhang. An  $O(\sqrt{n}L)$  iteration primal-dual path-following method, based on wide neighborhoods and large updates, for monotone LCP. *SIAM J. on Optimization*, 16:400–417, 2005.
- [3] F. Alizadeh. *Combinatorial Optimization with Interior-Point Methods and Semi-Definite Matrices*. PhD thesis, Computer Science Department, University of Minnesota, Minneapolis, MN, USA, 1991.
- [4] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM J. on Optimization*, 5:13–51, 1995.
- [5] F. Alizadeh, J.A. Haeberly, and M.L. Overton. Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results. *SIAM J. on Optimization*, 8:746–768, 1998.
- [6] D.S. Bernstein. *Matrix Mathematics*. Princeton University Press, Princeton, New Jersey, 2005.
- [7] B. Borchers. SDPLIB 1.2, library of semidefinite programming test problems. *Optimization Methods and Software*, 11:683–690, 1999.
- [8] E. de Klerk. *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.
- [9] E. de Klerk, C. Roos, and T. Terlaky. *Self-dual embeddings*, pages 111–138. In H. Wolkowicz, R. Saigal, and L. Vandenberghe (eds.), *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*. Kluwer Academic Publishers, Boston, 2000.
- [10] G.H. Golub and C.F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Maryland, 2nd edition, 1989.
- [11] C. Helmberg, F. Rendl, R.J. Vanderbei, and H. Wolkowicz. An interior-point method for semidefinite programming. *SIAM J. on Optimization*, 6:342–361, 1996.
- [12] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, 1985.
- [13] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, New York, 1990.
- [14] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

- [15] M. Kojima, S. Shindoh, and S. Hara. Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM J. on Optimization*, 7:86–125, 1997.
- [16] M.H. Koulaei and T. Terlaky. On the extension of a Mehrotra-type algorithm for semidefinite optimization. Technical Report 2007/4, Advanced Optimization Lab., Department of Computing and Software, McMaster University, Hamilton, Ontario, Canada. To appear in *Optimization Methods and Softwares*.
- [17] R.D.C. Monteiro. Primal–dual path-following algorithms for semidefinite programming. *SIAM J. on Optimization*, 7:663–678, 1997.
- [18] R.D.C. Monteiro. Polynomial convergence of primal-dual algorithms for semidefinite programming based on the Monteiro and Zhang family of directions. *SIAM J. on Optimization*, 8:797–812, 1998.
- [19] R.D.C. Monteiro and M.J. Todd. *Path-following methods*, pages 267–306. In H. Wolkowicz, R. Saigal, and L. Vandenberghe (eds.), *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*. Kluwer Academic Publishers, Boston, 2000.
- [20] R.D.C. Monteiro and Y. Zhang. A unified analysis for a class of long-step primal-dual path-following interior-point algorithms for semidefinite programming. *Math. Program.*, 81:281–299, 1998.
- [21] Y.E. Nesterov and A.S. Nemirovsky. *Interior Point Polynomial Algorithms in Convex Programming: Theory and Applications*. SIAM, Philadelphia, PA, 1994.
- [22] Y.E. Nesterov and M.J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.*, 22:1–42, 1997.
- [23] Y.E. Nesterov and M.J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM J. on Optimization*, 8:324–364, 1998.
- [24] G. Pataki and S. Schmieta. The DIMACS Library of Semidefinite-Quadratic-Linear Programs. Technical report, Rutgers, 1999. See <http://dimacs.rutgers.edu/Challenges/Seventh/Instances/>.
- [25] J. Peng, C. Roos, and T. Terlaky. *Self-Regularity: A New Paradigm for Primal-Dual Interior-Point Algorithms*. Princeton University Press, Princeton, New Jersey, 2002.
- [26] C. Roos, T. Terlaky, and J.-Ph. Vial. *Interior Point Methods for Linear Optimization*. Springer Science, New York, 2nd edition, 2006.
- [27] J.F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *SIAM J. on Optimization*, 7:86–125, 1997.
- [28] M.J. Todd. A study of search directions in primal-dual interior-point methods for semidefinite programming. *Optimization Methods and Software*, 11:1–46, 1999.



- [29] M.J. Todd, K.C. Toh, and R.H. Tütüncü. On the Nesterov–Todd direction in semidefinite programming. *SIAM J. on Optimization*, 8:769–796, 1998.
- [30] R.H. Tütüncü, K.C. Toh, and M.J. Todd. Solving semidefinite-quadratic-linear programs using SDPT3. *Math. Program.*, 95:189–217, 2003.
- [31] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Rev.*, 38(1):49–95, 1996.
- [32] H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*. International Series in Operations Research & Management Science. Kluwer Academic Publishers, Boston, 2000.
- [33] S. Wright. *Primal-Dual Interior-point Methods*. SIAM, Philadelphia, 1997.
- [34] Y. Ye. A class of projective transformations for linear programming. *SIAM J. on Computing*, 19:457–466, 1990.
- [35] Y. Zhang. On extending some primal–dual interior-point algorithms from linear programming to semidefinite programming. *SIAM J. on Optimization*, 8:365–386, 1998.