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

Yang Li* Tamás Terlaky†

July 2, 2008

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, we utilize symmetric directions by scaling the Newton equation with special matrices. After defining the "positive part" and the "negative part" of a symmetric matrix, we solve the Newton equation with its right hand side replaced first by its positive part and then by its negative part, respectively. In this way, we obtain a decomposition of the usual 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 $O(\eta\sqrt{\kappa_\infty n}\log\frac{\mathrm{Tr}(X^0S^0)}{\epsilon})$ iterations, where κ_∞ is a parameter associated with the scaling matrix P and ϵ is the required precision. To our best knowledge, when the parameter η 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 η is chosen to be in the order of $\sqrt{n}$, our result coincides with the results for the classical large neighborhood IPMs.

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

*School of Computational Engineering and Science, McMaster University, Hamilton, ON, L8S 4K1, Canada (liy99@mcmaster.ca).

†Department of Computing and Software, McMaster University, Hamilton, ON, L8S 4K1, Canada (terlaky@mcmaster.ca; http://www.cas.mcmaster.ca/~terlaky).

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 is applied in various areas, including control theory, probability and signal processing [27].

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 [20]. 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 [20], 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

$$X\Delta S + \Delta X S = \tau\mu I - XS, \tag{1.1}$$

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] suggested to symmetrize both sides of (1.1). Another possible alternative is to employ a similarity transformation $P(\cdot)P^{-1}$ on both sides of (1.1). This strategy was first investigated by Monteiro [16] 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 [14]. At the same time, another motivation led Helmberg, Rendl, Vanderbei and Wolkowicz [10] 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, respectively. Another very popular direction was introduced by Nesterov and Todd [21, 22] in their attempt to generalize primal-dual IPMs beyond SDO. In [31], based on Monteiro's idea, Zhang generalized all the 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 their theoretical complexity is worse, large neighborhood algorithms perform better in practice than small neighborhood algorithms. Many efforts were spent to bridge this gap. In [23], 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 idea [1], an important result was given by Ai and Zhang [2] for linear complementarity problems (LCP). Their algorithm uses a wide neighborhood and decomposes the classical Newton direction into two orthogonal directions using different step-length for each of them. They proved that the 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.

In this paper, we extend the Ai-Zhang technique 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 special case of our wide neighborhood. Another important ingredient of our algorithm is the decomposition of the classical Newton method into two individual directions: one of them reduces the duality gap and the other one keeps the iterates away from the boundary of the positive semidefinite cone. We use different step lengths for each of the directions. Further, we derive a symmetric direction by using scaling matrices P such that $PXSP^{-1}$ is symmetric for any iterate (X, y, S). Such directions are referred to in the literature as the Monteiro-Zhang (MZ) family. We prove 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 $O(\eta\sqrt{\kappa_\infty n}\log\frac{\mathrm{Tr}(X^0S^0)}{\epsilon})$ iterations. Here n is the dimension of the problem, κ_∞ is a parameter associated with the scaling matrix P, and ϵ is the required precision. In other words, when the parameter is a fixed constant, our large neighborhood path following algorithm has the same theoretical complexity as a small neighborhood algorithm that uses NT scaling, and when η 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 part 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. In Section 5, the convergency analysis and the theoretical complexity bound are presented. Finally, Section 6 consists of 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 $Q \in \mathcal{S}^n$.
Tr(Q):	the trace of a matrix $Q \in \mathcal{R}^{n\times n}$, i.e., $\mathrm{Tr}(Q) := \sum_{i=1}^{n} Q_{ii}$.
$\lambda_i(Q)$:	the eigenvalues of $Q \in \mathcal{S}^n$, $i = 1, 2, \ldots, n$.
$\lambda_{\min}(Q)$:	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.
cond(Q):	the condition number of Q, defined as $\mathrm{cond}(Q) = \lambda_{\max}(Q)/\lambda_{\min}(Q)$.
$\|Q\|$:	the Euclidean norm for $Q \in \mathcal{R}^{n\times n}$, i.e., $\|Q\| = \max_{\|\mu\|=1} \|Qu\|$.
$\|Q\|_F$:	the Frobenius norm of $Q \in \mathcal{R}^{n\times n}$, i.e., $\|Q\|_F = \sqrt{\mathrm{Tr}(Q^TQ)}$.
$\mathbf{vec}(Q)$:	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

$$(\mathcal{P}) \quad \begin{array}{lrcl} \min & \mathrm{Tr}(CX) & & \\ \text{s.t.} & \mathrm{Tr}(A_iX) & = & b_i,\ i = 1, \ldots, m, \\ & X & \succeq & 0, \end{array}$$

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

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

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

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

The primal-dual feasible set is defined as

$$\mathcal{F} = \left\{ (X, y, S) \in \mathcal{S}^n_+ \times \mathcal{R}^m \times \mathcal{S}^n_+ \;\middle|\; \begin{array}{c} \mathrm{Tr}(A_iX) = b_i,\ X \succeq 0,\ i = 1, \ldots, m \\ \sum_{i=1}^{m} y_i A_i + S = C, S \succeq 0 \end{array} \right\},$$

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

$$\mathcal{F}^0 = \left\{ (X, y, S) \in \mathcal{S}^n_{++} \times \mathcal{R}^m \times \mathcal{S}^n_{++} \;\middle|\; \begin{array}{c} \mathrm{Tr}(A_iX) = b_i,\ X \succ 0,\ i = 1, \ldots, m \\ \sum_{i=1}^{m} y_i A_i + S = C, S \succ 0 \end{array} \right\}.$$

Under the assumptions that $\mathcal{F}^0$ is nonempty and the matrices A_i, $i = 1, 2, \ldots, m$, are linearly independent, then X^* and (y^*, S^*) are optimal if and only if they satisfy the optimality conditions [7],

$$\begin{array}{rcl} \mathrm{Tr}(A_iX) & = & b_i,\ X \succeq 0,\ i = 1, \ldots, m \\ \sum_{i=1}^{m} y_i A_i + S & = & C,\ S \succeq 0, \\ XS & = & 0. \end{array} \tag{2.1}$$

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

$$\begin{array}{rcl} \mathrm{Tr}(A_iX) & = & b_i,\ X \succ 0,\ i = 1, \ldots, m \\ \sum_{i=1}^{m} y_i A_i + S & = & C,\ S \succ 0, \\ XS & = & \mu I. \end{array} \tag{2.2}$$

It is proved in [14, 18, 20] 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, $i = 1, \ldots, m$ are linearly independent. Moreover, the limit point (X^*, y^*, S^*) as μ goes to 0 is a primal-dual optimal solution for the SDO problem.

3 Neighborhood

Although path-following interior point algorithms follow the central path while the barrier parameter μ is decreasing to 0, they do not stay on the central path exactly. All the iterates are 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

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

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

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

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

In theory, IPMs based on the small neighborhood $\mathcal{N}_F(\beta)$, e.g., short step algorithms, have a better iteration complexity bound than algorithms based on large neighborhoods, e.g., large update algorithms. However, computational experiences shows that large update IPMs perform much 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 $O(\eta\sqrt{\kappa_\infty n} \log \frac{\mathrm{Tr}(X^0 S^0)}{\epsilon})$, where n is the dimension of the problem, κ_∞ is a parameter associated with the scaling matrix P and ϵ is the required precision. The new parameter η is used to defined our new neighborhood $\mathcal{N}(\tau_1, \tau_2, \eta)$. In particular, when η is chosen to be a constant, our new algorithm will have the best complexity result $O(\sqrt{n} \log \frac{\mathrm{Tr}(X^0 S^0)}{\epsilon})$, which coincides with the complexity of short step IPMs, when we use the Nesterov-Todd (NT) scaling.

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 introduce 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 Λ is a diagonal matrix with all the eigenvalues of M in its diagonal, and 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 λ_i. Then, we define the positive part M^+ and the negative part M^- of M as

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

In particular, for a real number $M \in \mathcal{S}^1$, M^+ denotes its nonnegative part, i.e., $M^+ = \max\{M, 0\}$, and M^- denotes its nonpositive part, i.e., $M^- = \min\{M, 0\}$. Furthermore, if $M \in \mathcal{S}^n$ is a diagonal matrix, M^+ and M^- could be easily constructed by taking the positive and negative elements separately along the diagonal and leaving the zeros where they are. Apparently, $M = M^+ + M^-$.

In the next, we investigate some algebraic properties of M^+ and M^-. These properties paly a crucial role throughout the paper.

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*

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

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 $i = 1, \ldots, n$.

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 unitary transformation preserves the Frobenius norm over the positive part of a symmetric matrix.

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

$$\left\|M^+\right\|_F = \left\|(QMQ^T)^+\right\|_F.$$

Proof. Because M is similar to QMQ^T, they have the same eigenvalues. Specially, they have the same nonnegative eigenvalues. Then the result follows easily. ∎

The next paramount lemma reveals that the positive part of a symmetric matrix doesn't exceed, in the sense of Frobenius norm, its positive part after a similar transformation.

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

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

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

Lemma 3.4 *Let $M \in \mathcal{S}^n$ and λ_i and m_{ii} denote the i^{th} eigenvalue and the i^{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, there exists a spectral decomposition such that $M = Q\Lambda Q^T = \sum_{i=1}^{n} \lambda_i q_i q_i^T$, where Λ is the diagonal matrix with all of the eigenvalues of M and Q is a unitary matrix, i.e., $QQ^T = I$, where q_i is the eigenvector of M corresponding to the eigenvalue λ_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. By definition,

$$M = M^+ + M^- = \sum_{\lambda_i \geq 0} \lambda_i q_i q_i^T + \sum_{\lambda_i \leq 0} \lambda_i q_i q_i^T,$$

and M^+ and M^- are positive and negative semidefinite, 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

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

For any $i \in \mathcal{I}$, we have $m_{ii}^+ \geq m_{ii}^+ + m_{ii}^- \geq 0$, since $m_{ii}^- < 0$. 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 $\|M^+\|_F^2 = \left\|[\Lambda(M)]^+\right\|_F^2 = \sum_{\lambda_i(M) \geq 0} \lambda_i^2(M)$. Let us consider the right hand side. According to Theorem B.1, there exists a unitary matrix U such that $U(WMW^{-1})U^T = \Lambda(WMW^{-1}) + N = \Lambda(M) + N$, where N is a strictly upper triangular matrix. The last equality is due to the similarity of WMW^{-1} and M. From Lemma 3.2, we know that

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

From Lemma 3.4, we claim

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

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

Proposition 3.1 and Lemmas 3.2 and 3.3 will play a crucial role in proving convergence and complexity of our new large neighborhood IPM. Analogous to the neighborhood introduced by Ai and Zhang [2], we define our neighborhood, using the positive part in (3.1), as

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

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

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

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

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

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

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

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

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,

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

That implies

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

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

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 target the point on the central path corresponding to $\mu = \tau\mu_g$, where $\tau \in [0, 1]$ is called centering parameter and $\mu_g = X \bullet S/n$ 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 compute a symmetric search direction by solving the following linear system

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

Although from the second equality we have that ΔS is symmetric, system (4.1) do not allow a symmetric solution matrix ΔX. Various remedies are proposed since the middle of 1990's. The interested readers are referred to the papers [5, 10, 14, 21, 22, 16, 17] for comprehensive discussions. We use the approach proposed by Zhang [31], who suggested to replace the last equation in system (2.2) by

$$H_P(XS) = \mu I, \tag{4.2}$$

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

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

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 [31], Zhang observed that if P is nonsingular, then

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

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

$$\mathrm{Tr}(A_i \Delta X) = 0, \tag{4.3a}$$

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

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

For the choices of P, 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 [10, 14, 16, 17], respectively. 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 [21, 22] prove the existence and uniqueness of such a solution as $W_{NT} = X^{1/2}(X^{1/2} S X^{1/2})^{-1/2} X^{1/2}$. We refer to the directions derived from (4.3) as the Monteiro-Zhang (MZ) family.

In terms of Kronecker product[1], equation (4.3c) can be expressed as

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

[1]For the definition and properties of Kronecker product, please refer to APPENDIX A.

where

$$E = \frac{1}{2}(S \otimes I + I \otimes S), \quad F = \frac{1}{2}(X \otimes I + I \otimes X). \tag{4.4}$$

In [26], Todd, Toh and Tütüncü 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 for which $PXSP^{-1} \in \mathcal{S}^n$. Actually, this still holds under weaker conditions, as the authors show in their papers [18]. However, to investigate our new large neighborhood path-following algorithm, throughout the paper, we restrict the scaling matrix P to the aforementioned specific class

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

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. However, $P = I$ does not. We should mention that this restriction on P is common for large neighborhood path-following algorithms proposed in [19]. Furthermore, this restriction on P does not lose any generality, in terms of the solution set of system (4.3), as Monteiro indicates in [17].

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 while they reduce 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:

$$\mathrm{Tr}(A\Delta X_-) \;=\; 0, \tag{4.6a}$$

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

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

and

$$\mathrm{Tr}(A\Delta X_+) \;=\; 0, \tag{4.7a}$$

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

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

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

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

and

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

respectively.

Obviously, 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

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

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

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

So far, we have already introduced the most important ingredients of our 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). Now, 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 $\eta \ge 1$, $0 < \tau_2 < \tau_1 < 1$;
reference parameter $0 \le \tau \le 1$;
an initial point $(X^0, y^0, S^0) \in \mathcal{N}(\tau_1, \tau_2, \eta)$ with $\mu_g^0 = \mathrm{Tr}(X^0 S^0)/n$;
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 $\mu_g^{k+1} := \mathrm{Tr}(X^{k+1}S^{k+1})/n$ and $k := k+1$.
end while

We have to remark three important facts about the presented algorithm. First of all, although we suggest to solve problem (4.10) to decide the best step lengths, to solve this problem is very expensive in general, and thus a "sufficient" duality gap decrease obtained for low computational cost is preferred against the "maximal possible" duality gap decrease for high computational cost. Furthermore, solving problem (4.10) is also unnecessary. 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 later. 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 decompose the Schur matrix once, both of which together usually take up 90% of the total running time, then does the backsolve once for the two right-hand-sides simultaneously. Third, it seems that it might be expensive to obtain the negative and positive parts in (4.6) and (4.7). However, we can utilize the strategy, scaling X and S to the same diagonal matrix, as proposed by Todd, Toh and Tütüncü in [26] to obtain the negative and positive parts cheaply as a byproduct when computing the NT scaling matrix.

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 [18] in order to analyze the algorithm in a unified way for the class of matrices $P \in \mathcal{P}(X,S)$. Furthermore, this scaling procedure simplifies the proofs of the main results. At the end of this section, after proving some technical lemmas, we present the most important result w.r.t. polynomial convergence.

5.1 Scaling Procedure

Scale the primal and dual variables in the following way,

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

To keep consistency, we have to apply the same scaling to the other data as well, i.e.,

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

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

$$\mathcal{P}(X,S) := \{P \in \mathcal{S}^n_{++} \mid PXSP^{-1} \in \mathcal{S}^n\} = \{P \in \mathcal{S}^n_{++} : \widetilde{X}\widetilde{S} = \widetilde{S}\widetilde{X}\}, \tag{5.2}$$

i.e., we require P to make $\widetilde{X}$ and $\widetilde{S}$ to commute after scaling, implying that $\widetilde{X}\widetilde{S}$ is symmetric, as long as $\widetilde{X}$ and $\widetilde{S}$ are both symmetric. This requirement on P also guarantees that $\widetilde{X}$ and $\widetilde{S}$ can be simultaneously diagonalised (i.e., they have eigenvalue decompositions with the same Q) according to Proposition B.6.

From now on, we use Λ to denote the diagonal matrix $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, where λ_i for $i = 1, \dots, n$ are the eigenvalues of $\widetilde{X}\widetilde{S}$ with increasing order $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. We should emphasize that the matrices $\widetilde{X}\widetilde{S}$, $\widetilde{S}\widetilde{X}$, XS, SX, $X^{1/2}S^{1/2}X^{1/2}$ and $S^{1/2}X^{1/2}S^{1/2}$ have the same eigenvalues, since they are similar.

In the scaled space the primal and dual problems are equivalent to the following pair of problems:

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

and

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

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

$$\widetilde{\Delta X}_- = P\Delta X_- P, \quad \widetilde{\Delta y}_- = \Delta y_-, \quad \widetilde{\Delta S}_- = P\Delta S_- P, \tag{5.3}$$

$$\widetilde{\Delta X}_+ = P\Delta X_+ P, \quad \widetilde{\Delta y}_+ = \Delta y_+, \quad \widetilde{\Delta S}_+ = P\Delta S_+ P. \tag{5.4}$$

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

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

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

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

and

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

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

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

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

$$\widetilde{E}\mathbf{vec}(\widetilde{\Delta X}_-) + \widetilde{F}\mathbf{vec}(\widetilde{\Delta S}_-) = \mathbf{vec}([\tau\mu_g I - \widetilde{X}\widetilde{S}]^-), \tag{5.7a}$$

$$\widetilde{E}\mathbf{vec}(\widetilde{\Delta X}_+) + \widetilde{F}\mathbf{vec}(\widetilde{\Delta S}_+) = \mathbf{vec}([\tau\mu_g I - \widetilde{X}\widetilde{S}]^+), \tag{5.7b}$$

respectively, where

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

Having the search directions, and after deciding about the step lengths, the iterates are updated as follows:

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

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 $\widetilde{\mu}_g(\alpha) = \frac{\mathrm{Tr}(\widetilde{X}(\alpha)\widetilde{S}(\alpha))}{n}$.*

5.2 Technical Results

Before proving the complexity of our algorithm, 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 current iterate is feasible, then the search directions are orthogonal.

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

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

Proof. The proof is straightforward by using (5.5a), (5.5b), (5.6a), and (5.6b). ■

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

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

and

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

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

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

One can show (5.11) analogously. ■

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. It will be seen in later discussions that this bound derives from feasibility considerations.

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*

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

Furthermore,

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

Proof. Using Lemma 5.2 and Lemma 5.3, we have

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

Then, we have

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

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

$$\mathrm{Tr}([\tau_1 \widetilde{\mu}_g I - \widetilde{X}\widetilde{S}]^-) \geq \mathrm{Tr}(-\widetilde{X}\widetilde{S}).$$ ■

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

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

$$\mathrm{Tr}([\tau_1 \widetilde{\mu}_g I - \widetilde{X}\widetilde{S}]^-) \leq -(1 - \tau_1)\mathrm{Tr}(\widetilde{X}\widetilde{S}). \tag{5.12}$$

Proof. It is easy to see that

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

Taking the trace of both sides, we have

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

which completes the proof. ■

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

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

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

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

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

Proof. Using Lemmas 5.3, 5.4 and 5.5, we see that

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

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

When the parameters τ_1 and β are chosen appropriately and all the iterates reside in the neighborhood $\widetilde{\mathcal{N}}(\tau_1, \tau_2, \eta)$, we claim that the duality gap is decreasing in $O(1-1/\sqrt{n})$.

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

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

Proof. From Proposition 5.6, it follows that

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

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

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

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

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

Proof. To prove this lemma, we first consider the situation when $\lambda_{\min}(\tau_1\widetilde{\mu}_g - \widetilde{X}\widetilde{S}) \geq 0$. In this case, $[\tau_1\widetilde{\mu}_g I - \widetilde{X}\widetilde{S}]^- = 0$. Then,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

The proof is completed. ■

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

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

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

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

$$\widetilde{E} = \frac{1}{2}(\widetilde{S} \otimes I + I \otimes \widetilde{S}) = \frac{1}{2}Q_K(\Lambda(\widetilde{S}) \otimes I + I \otimes \Lambda(\widetilde{S}))Q_K^T,$$
$$\widetilde{F} = \frac{1}{2}(\widetilde{X} \otimes I + I \otimes \widetilde{X}) = \frac{1}{2}Q_K(\Lambda(\widetilde{X}) \otimes I + I \otimes \Lambda(\widetilde{X}))Q_K^T,$$

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

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

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,

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

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

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

which leads to inequality (5.16). ■

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

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

Proof. Noticing that $\lambda_{\min}(\widetilde{F}\widetilde{E}) = \lambda_1 \geq \tau_2\widetilde{\mu}_g$, it is easy to see that

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

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

Now, we apply Lemmas 5.10 and 5.11, together with Lemma B.4, to conclude the following result.

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

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

Proof. From (5.7), we have

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

Applying Lemma B.4 to this equality, we obtain

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

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

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

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

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

where the last inequality can be derived from Lemma 5.10 and 5.11. ■

Using Lemma B.10, 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 $G = \widetilde{E}^{-1}\widetilde{F}$. If $\beta \leq 1/4$, $\alpha_- = \alpha_+\hat{\eta}\sqrt{\frac{\beta\tau_1}{n}}$ and $(\widetilde{X},\widetilde{y},\widetilde{S}) \in \widetilde{\mathcal{N}}(\tau_1,\tau_2,\eta)$, then we have*

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

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

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

From Lemmas 5.2 and B.10, it follows that

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

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

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

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

By the next proposition, we get one of the most important results in this paper, a sufficient condition to keep all the iterates in the neighborhood $\mathcal{N}(\tau_1,\tau_2,\eta)$.

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

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

Proof. By Corollary 5.7 we have $\widetilde{\mu}_g(\alpha)\le\widetilde{\mu}_g$. Further, using Lemmas 5.8, 5.13 and the fact that $\lambda_{\min}(\cdot)$ is a homogeneous concave function on the space of symmetric matrices, one has

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

One can derive from Lemma 5.13 that

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

This implies that $\widetilde{X}(\alpha)\widetilde{S}(\alpha)$ is nonsingular, implying that each of the factors $\widetilde{X}(\alpha)$ and $\widetilde{S}(\alpha)$ are nonsingular as well. By using continuity, it follows that $\widetilde{X}(\alpha)$ and $\widetilde{S}(\alpha)$ are also in $\mathcal{S}^n_{++}$, since $\widetilde{X}$ and $\widetilde{S}$ are. Then, we may claim that

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

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

$$\widetilde{\mu}_g(\alpha) \ge (1-\alpha_-)\widetilde{\mu}_g \ge (1-\hat{\eta}\sqrt{\beta\tau_1}/\sqrt{n})\widetilde{\mu}_g \ge (1-\sqrt{\beta\tau_1})\widetilde{\mu}_g \ge \frac{2}{3}\widetilde{\mu}_g. \tag{5.19}$$

From Proposition 3.1, we have

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

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

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

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

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

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

$$\begin{aligned}
\psi(\alpha) &\le (1-\alpha_+)\hat{\eta}\beta\tau_1\widetilde{\mu}_g(\alpha) + \tfrac{2}{3}\sqrt{\mathrm{cond}(G)}\alpha_+^2\hat{\eta}\beta\tau_1\widetilde{\mu}_g \\
&\le (1-\alpha_+)\hat{\eta}\beta\tau_1\widetilde{\mu}_g(\alpha) + \sqrt{\mathrm{cond}(G)}\alpha_+^2\hat{\eta}\beta\tau_1\widetilde{\mu}_g(\alpha).
\end{aligned}$$

Since $\alpha_+ \le 1/(\sqrt{\mathrm{cond}(G)}\hat{\eta}^2)$ and $\widetilde{\eta} \ge 1$, we have $\sqrt{\mathrm{cond}(G)}\alpha_+^2\hat{\eta}\beta\tau_1\widetilde{\mu}_g(\alpha) \le \alpha_+\hat{\eta}\beta\tau_1\widetilde{\mu}_g(\alpha)$. Thus,

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

This, together with (5.18), implies that

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

Consequently, according to Proposition (5.1), one has

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

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 κ_∞ defined as

$$\kappa_\infty = \sup\left\{\operatorname{cond}((\widetilde{E}^k)^{-1}\widetilde{F}^k) : \; k = 0, 1, \dots\right\}. \tag{5.20}$$

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 $O(\eta\sqrt{\kappa_\infty n}\log\frac{\operatorname{Tr}(X^0S^0)}{\epsilon})$ iterations with a solution $\operatorname{Tr}(XS) \leq \epsilon$.*

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

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

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

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

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

As we see, when η 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 η 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 Conclusions and Further Works

As discussed previously, the contribution of this paper is a new large neighborhood path-following algorithm having the same theoretical complexity bond as the best short-step path-following algorithm. As usual, the next step would be to discuss issues related to the implementation of the new algorithm. Notice that finding the step lengths in (4.10) is relatively expensive. Although we proved the theoretical complexity for fixed step lengths, looking for efficient heuristics to compute better step lengths in practice deserves special attention. Another issue is how to efficiently compute the positive and negative part of the right-hand-side in the Newton equation when not the NT scaling is used. It is apparent that, due to the high computational cost, explicitly computing them with eigenvalue decomposition is not desired. To develop computationally efficient implementation strategies for our large neighborhood algorithm remains the subject of further research.

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 denote 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

$$\mathbf{vec}(Q) = [q_{11}, \cdots, q_{m1}, q_{12}, \cdots, q_{m2}, \cdots, q_{n1}, \cdots, q_{nn}]^T.$$

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.
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 λ_i and μ_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 $U \in \mathcal{R}^n$ such that*

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

where N is a strictly upper triangular matrix.

Proof. For the proof, see Horn and Johnson [11], 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^{th} *largest eigenvalues of* A *and* B, *respectively, i.e.,*

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

Then there exist nonnegative numbers $\delta_1, \ldots, \delta_n$ *such that*

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

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

Proof. For the proof, see Golub and Van Loan [9], page 412. ■

Lemma B.3 *Let* κ_∞ *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* $\kappa_\infty \leq \frac{n}{\tau_2}$;
- *if for all* k *the scaling matrix* $P^k = (X^k)^{-1/2}$, *then* $\kappa_\infty \leq \frac{n}{\tau_2}$.

Proof. For the proof of this lemma, we refer to Monteiro's paper [19]. ■

The following technical lemma was first introduced and proved in Zhang [31].

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*

$$\left\|(RQ^T)^{-1/2}Qu\right\|^2 + \left\|(RQ^T)^{-1/2}Rv\right\|^2 + 2u^Tv = \left\|(RQ^T)^{-1/2}r\right\|^2. \tag{B.1}$$

Proof. For the proof, we refer to Zhang's paper [16]. ■

To utilize Lemma B.4, we need to explore the conditions under which $\widetilde{F}\widetilde{E}^T \in \mathcal{S}^n_{++}$, where $\widetilde{F}$ and $\widetilde{E}$ is defined by (5.8). In [26] and [16], the authors state the same necessary and sufficient condition for $\widetilde{F}\widetilde{E}^T \in \mathcal{S}^n_{++}$ but in different formats. In our paper, we utilize the proposition stated in [16]. For those who are interested in the proof, they are advised to consult the paper by Monteiro [16].

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

(i) $\widetilde{E}, \widetilde{F} \in \mathcal{S}^{n^2}_{++}$, *and thus* $\widetilde{F}\widetilde{E}^T = \widetilde{F}\widetilde{E}$;

(ii) $\widetilde{F}\widetilde{E} \in \mathcal{S}^{n^2}$ *if and only if* $\widetilde{X}\widetilde{S} \in \mathcal{S}^n$;

(iii) $\widetilde{F}\widetilde{E} \in \mathcal{S}^{n^2}$ *implies* $\widetilde{F}\widetilde{E} \in \mathcal{S}^{n^2}_{++}$.

The following results and their proofs can be found in Monteiro [16]. We use his results throughout this paper.

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

(i) $\widetilde{X} = PXP = Q\Lambda(\widetilde{X})Q^T$;

(ii) $\widetilde{S} = P^{-1}SP^{-1} = Q\Lambda(\widetilde{S})Q^T$;

(iii) $\Lambda = \Lambda(\widetilde{X})\Lambda(\widetilde{S})$, *and hence* $\widetilde{X}\widetilde{S} = \widetilde{S}\widetilde{X} = Q\Lambda Q^T$.

Lemma B.7 *For any $Q \in \mathcal{S}^n$, we have*

$$\begin{aligned} \lambda_{\max}(Q) &= \max_{\|u\|=1} u^T Q u, & \text{(B.2)}\\ \lambda_{\min}(Q) &= \min_{\|u\|=1} u^T Q u, & \text{(B.3)}\\ \|Q\| &= \max_{i=1,\ldots,n} |\lambda_i(Q)|, & \text{(B.4)}\\ \|Q\|_F^2 &= \sum_{i=1}^n |\lambda_i(Q)|^2. & \text{(B.5)} \end{aligned}$$

Lemma B.8 *For any $Q \in \mathcal{R}^{n\times n}$ the following relations hold:*

$$\begin{aligned} \max_{i=1,\ldots,n} \mathrm{Re}[(\lambda_i(Q))] &\leq \frac{1}{2}\lambda_{\max}(Q+Q^T), & \text{(B.6)}\\ \min_{i=1,\ldots,n} \mathrm{Re}[(\lambda_i(Q))] &\geq \frac{1}{2}\lambda_{\min}(Q+Q^T), & \text{(B.7)}\\ \sum_{i=1}^n |\lambda_i(Q)|^2 &\leq \|Q\|_F^2 = \left\|Q^T\right\|_F^2, & \text{(B.8)}\\ \lambda_{\max}(Q^TQ) &= \left\|Q^TQ\right\| = \|Q\|^2 = \left\|Q^T\right\|^2, & \text{(B.9)}\\ \|Q\|_F &\geq \left\|(Q+Q^T)/2\right\|_F. & \text{(B.10)} \end{aligned}$$

Lemma B.9 *Let $W \in \mathcal{R}^{n\times n}$ be a nonsingular matrix. Then, for any $Q \in \mathcal{S}^n$, we have*

$$\begin{aligned} \lambda_{\max}(Q) &\leq \frac{1}{2}\lambda_{\max}(WQW^{-1} + (WQW^{-1})^T), & \text{(B.11)}\\ \lambda_{\min}(Q) &\geq \frac{1}{2}\lambda_{\min}(WQW^{-1} + (WQW^{-1})^T), & \text{(B.12)}\\ \|Q\| &\leq \frac{1}{2}\left\|(WQW^{-1} + (WQW^{-1})^T\right\|, & \text{(B.13)}\\ \|Q\|_F &\leq \frac{1}{2}\left\|(WQW^{-1} + (WQW^{-1})^T\right\|_F. & \text{(B.14)} \end{aligned}$$

Lemma B.10 *For any $u, v \in \mathcal{R}^n$ and $G \in \mathcal{S}^n_{++}$, we have*

$$\|u\|\,\|v\| \leq \frac{\sqrt{\mathrm{cond}(G)}}{2}\left(\left\|G^{-1/2}u\right\|^2 + \left\|G^{1/2}v\right\|\right). \tag{B.15}$$

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] E. de Klerk. *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*. Kluwer Academic Publishers, Dordrecht,Netherlands, 2002.

[8] 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.

[9] G.H. Golub and C.F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Maryland, 2 edition, 1989.

[10] 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.

[11] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, 1985.

[12] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, New York, 1990.

[13] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[14] 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.

[15] 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.

[16] R.D.C. Monteiro. Primal–dual path-following algorithms for semidefinite programming. *SIAM J. on Optimization*, 7:663–678, 1997.

[17] 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.

[18] 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.

[19] 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.

[20] Y.E. Nesterov and A.S. Nemirovsky. *Interior Point Polynomial Algorithms in Convex Programming: Theory and Applications*. SIAM, Philadelphia, PA, 1994.

[21] Y.E. Nesterov and M.J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.*, 22:1–42, 1997.

[22] Y.E. Nesterov and M.J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM J. on Optimization*, 8:324–364, 1998.

[23] 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.

[24] C. Roos, T. Terlaky, and J.-Ph. Vial. *Interior Point Methods for Linear Optimization*. Springer Science, New York, 2nd edition, 2006.

[25] J.F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *SIAM J. on Optimization*, 7:86–125, 1997.

[26] 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.

[27] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Rev.*, 38(1):49–95, 1996.

[28] 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.

[29] S. Wright. *Primal-Dual Interior-point Methods*. SIAM, Philadelphia, 1997.

[30] Y. Ye. A class of projective transformations for linear programming. *SIAM J. Computing*, 19:457–466, 1990.

[31] Y. Zhang. On extending some primal–dual interior-point algorithms from linear programming to semidefinite programming. *SIAM J. on Optimization*, 8:365–386, 1998.