NEW LARGE NEIGHBORHOOD IPMS FOR SDO

NEW LARGE NEIGHBORHOOD INTERIOR POINT METHODS FOR SEMIDEFINITE OPTIMIZATION

by

YANG LI, B.S.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Master of Applied Science

McMaster University

MASTER OF APPLIED SCIENCE (2008) McMaster University
(School of Computational Engineering and Science) Hamilton, Ontario

TITLE: New Large Neighborhood Interior Point Methods for Semidefinite Optimization
AUTHOR: Yang Li, B.S.
SUPERVISOR: Dr. Tamás Terlaky
NUMBER OF PAGES: xvi, 67

Abstract

In this thesis, 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 $O(\eta\sqrt{\kappa_\infty n}\log(1/\epsilon))$ iterations, where κ_∞ is a parameter associated with the scaling matrix 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 complexity bound coincides with the known bound for classical large neighborhood IPMs.

To make this thesis more accessible to readers who are new in this area, we start with a brief introduction to IPMs and SDO. The basic concepts and principles of IPMs and SDO are presented in Chapter 2 and 3.

Acknowledgements

I would like to express my sincere appreciations to many people, without whose generous support this thesis would not have been possible.

I gratefully acknowledge to my supervisor, Dr. Tamás Terlaky, who kindly accepted me as a master student at McMaster University. Under his guidance, I started my research work in Interior Point Methods. When I feel confused and lose my ideas, he is always the one helping me out. Without his support, this thesis would not have been completed.

I also thank the other members of my examination committee Dr. Tim Davidson (chair), Dr. Antoine Deza, and Dr. Yuriy Zinchenko, for their careful review and helpful comments.

I am indebted to Dr. Wenbao Ai, who has been my undergraduate advisor and introduced me into optimization, specially Interior Point Methods. He also answered my many questions regarding his Ai-Zhang algorithm, which had significant impact on my work in this thesis.

All the help from the former and current members of the AdvOL are memorable. Special thanks go to Jiaping Zhu, who cooked me my first dinner at Hamilton, and Voicu Chis, with whom I discussed a lot during the process of my thesis.

Finally, I dedicate this work to my parents, Zhenfu Li and Jing Zhang as a tiny present for their constant support and encouragement at any time in my life. They are the real motivation of all my work.

Contents

List of Figures

List of Algorithms

Abbreviations

IPM	Interior Point Method
LCP	Linear Complementarity Problem
LO	Linear Optimization
MZ	Monteiro-Zhang
NT	Nesterov-Todd
SDO	Semidefinite Optimization
SOCO	Second-Order Conic Optimization
SVD	Singular Value Decomposition

Notations and Symbols

e:	the all one vector with an appropriate dimension.
$\mathcal{R}^n$:	the n-dimensional Euclidean space.
$\mathcal{R}^n_+$:	the nonnegative orthant defined as $\mathcal{R}^n_+ = \{x \in \mathcal{R}^n \mid x \geq 0\}$.
$\mathcal{R}^n_{++}$:	the positive orthant defined as $\mathcal{R}^n_{++} = \{x \in \mathcal{R}^n \mid x > 0\}$.
$\mathcal{R}^{m\times n}$:	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.
$\mathrm{cond}(Q)$:	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., $\|Q\|_F = \sqrt{\mathrm{Tr}(Q^T Q)}$.
$\|q\|$:	the 2-norm of $q \in \mathcal{R}^n$, i.e., $\|q\| = (q_1^2 + \cdots + q_n^2)^{1/2}$.
$\mathbf{vec}(Q)$:	the vector obtained by stacking Q's columns one by one. See Appendix A for more details.
$\mathrm{Diag}(q)$:	the diagonal matrix in $\mathcal{R}^{n\times n}$ with elements of $q \in \mathcal{R}^n$, $q_1, \ldots, q_n$, distributed along the diagonal.
$\langle x, y\rangle$:	the inner product of variables x and y. If $x, y \in \mathcal{R}^n$, then $\langle x, y\rangle = x^T y = \sum_{i=1}^{n} x_i y_i$. If $x, y \in \mathcal{R}^{n\times n}$, then $\langle x, y\rangle = \mathrm{Tr}(x^T y) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ij} y_{ij}$.

Chapter 1

History of Semidefinite Optimization

For almost two decades, Semidefinite Optimization (SDO), which yields a generalization of Linear Optimization (LO) problems, has been one of the most active research areas in mathematical programming. More and more difficult problems arising from practice could be modeled as SDO problems. Nevertheless, when we carefully review the literature, the first paper which discusses an SDO problem can be traced back to the 1960s, almost four decades ago [?].

One of the reasons that SDO was out of interest for such a long time was the lack of robust and efficient algorithms for solving SDO problem. This situation has changed since the 1990s. Alizadeh [3] and Nesterov and Nemirovskii [24] independently developed the first IPMs for SDO. 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 [24], Nesterov and Nemirovskii proved that IPMs are able to solve general conic optimization problems, in particular SDO problems, in polynomial time. After that, notable successes are achieved in the theory and practice of SDO. Nowadays, algorithms for SDO has been already matured and one can solve large scale problems even on a desktop PC.

Another reason that SDO was out of interest for a long time is the limited realization of its importance in various applications. In [4], Alizadeh showed how SDO could be used in combinatorial optimization. One of the landmark work was done by Goemans and Williamson [10]. Using SDO they proposed a randomized 0.878-approximation algorithm for the famous NP hard MAX-CUT problem. In the past two decades, more and more applications of SDO were

derived in e.g., control theory, probability theory, statistics, signal processing and structural design [34].

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

$$X\Delta S + \Delta X S = \tau\mu I - XS, \tag{1.0.1}$$

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 (1.0.1). Another possible alternative is to employ a similarity transformation $P(\cdot)P^{-1}$ on both sides of (1.0.1). This strategy was first investigated by Monteiro [20] 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 [16]. At the same time, another motivation led Helmberg, Rendl, Vanderbei and Wolkowicz [12] 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 [25, 26] in their attempt to generalize primal-dual IPMs beyond SDO. In [39], 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.

Thanks to two decades of efforts of numerous researchers and scientists from mathematics and computer science, nowadays SDO turns out to be a very sophisticated technique. Several efficient and accurate solvers are available, such as SeDuMi [30] and SDPT3 [33]. More and more engineers utilize SDO in their own research projects. Nevertheless, there are still some challenging questions in SDO.

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 [27], 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 complexity bounds of small neighborhood IPMs. Their results hold for LO, SOCO, SDO and for monotone nonlinear complementarity problems. Later, based on Ai's original idea [1], an important

result was given by Ai and Zhang [2] for monotone 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 $O(\sqrt{n}\log(1/\epsilon))$ iterations, where ϵ is the required precision. 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 thesis, we extend the Ai-Zhang technique to SDO. We 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 direction into two individual directions: one of them reduces the duality gap and the other one moves 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 $O(\eta\sqrt{\kappa_\infty n}\log(1/\epsilon))$ 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.

Chapter 2

Semidefinite Optimization

In this chapter, we present the Semidefinite Optimization (SDO) problem and some of its characteristics. After reviewing the Linear Optimization (LO) problem in Section 2.1, we introduce the so-called Semidefinite Optimization problem, which could be considered as a generalized LO problem over the space of positive semidefinite matrices. Although *weak duality* still holds for SDO problems, the *strong duality* theorem might not be true in some cases, even if a feasible solution exists for both the primal and the dual problems. In Section 2.2.2, we illustrate by some examples that strong duality may not hold, and present a sufficient condition which guarantee strong duality. At the end of this chapter, we present the optimality condition for SDO.

2.1 Review of Linear Optimization

Consider the Linear Optimization problem in a standard primal form

$$(LP) \quad \begin{array}{lll} \min & \langle c, x \rangle & \\ \text{s.t.} & \langle a_i, x \rangle = b_i, & i = 1, \ldots, m, \\ & x \in \mathcal{R}^n_+, & \end{array}$$

where $c \in \mathcal{R}^n$ and the vectors a_i, $i = 1, \ldots, m$ are linearly independent in $\mathcal{R}^n$ and $x \in \mathcal{R}^n_+ = \{x | x \geq 0\}$. The way we present the LO problem might be a little different from what you see in the LO literature. However, this way might be more straightforward to see the similarity between LO problems and SDO problems.

The notation $\langle\ ,\ \rangle$ denotes the inner product of two variables over a certain

space. In LO, this space is the common vector space $\mathcal{R}^n$ and the associated inner product coincides with the dot product for two vectors, i.e., $\langle c, x\rangle = c^T x = \sum_{i=1}^{n} c_i x_i$. Besides satisfying m equalities, x also has to reside in the nonnegative orthant, i.e., $\mathcal{R}^n_+ = \{x \in \mathcal{R}^n | x \geq 0\}$. A solution x of (LP) is called *primal feasible* if it belongs to the set

$$\mathcal{F}_{LP} := \left\{ x \in \mathcal{R}^n_+ \mid \langle a_i, x\rangle = b_i,\ i = 1, \ldots, m \right\}.$$

A primal feasible solution x^* is called a primal optimal solution if $\langle c, x^*\rangle \leq \langle c, x\rangle$ for any $x \in \mathcal{F}_{LP}$.

In practice, as long as the problem is minimizing or maximizing a linear function over several linear constraints, we are able to refer to it as LO, since we are able to transfer it to the standard primal form [35].

We could also give the dual of the standard primal LO as follows

$$(LD) \quad \begin{array}{ll} \max & \langle b, y\rangle \\ \text{s.t.} & \sum_{i=1}^{m} y_i a_i + s = c, \\ & s \in \mathcal{R}^n_+, \end{array}$$

where $y_i \in \mathcal{R}$ and $a_i \in R^n$, $i = 1, \ldots, m$. A solution (y, s) of (LD) is called *dual feasible* if it belongs to the set

$$\mathcal{F}_{LD} := \left\{ (y, s) \in \mathcal{R}^m \times \mathcal{R}^n_+ \mid \sum_{i=1}^{n} y_i a_i + s = c \right\}.$$

A dual feasible solution x is called a dual optimal solution if $\langle b, y\rangle \leq \langle b, y^*\rangle$ for any $(y, s) \in \mathcal{F}_{LD}$.

Given a primal-dual feasible solution $(x, y, s) \in \mathcal{F}_L := \mathcal{F}_{LP} \times \mathcal{F}_{LD}$, the *weak duality* property gives that the *duality gap* is simply

$$\langle c, x\rangle - \langle b, y\rangle = \langle c - \sum_{i=1}^{n} y_i a_i, x\rangle = \langle s, x\rangle \geq 0.$$

From the *strong duality* theory [29], we know that if primal problem (LP) has an optimal solution x^*, then dual problem (LD) also has an optimal solution (y^*, s^*) and the duality gap vanishes at (x^*, y^*, s^*).

2.2 Introduction to Semidefinite Optimization

2.2.1 Primal and Dual Problems

As mentioned before, Semidefinite Optimization (SDO) could be roughly considered as generalized LO over the space of positive semidefinite matrices. Because the variables we care about in SDO are matrices, we, therefore, replace the nonnegativity requirement of LO by the requirement that the matrix of variables is symmetric positive semidefinite, i.e., $X \in \mathcal{S}^n_+$, where $\mathcal{S}^n$ is the set of symmetric matrices and further $\mathcal{S}^n_+ = \{x \in \mathcal{S}^n \mid u^T X u \geq 0,\ \forall u \in \mathcal{R}^n\}$. Now, we introduce the primal SDO in the form of

$$\begin{array}{cll} & \min & \langle C, X\rangle \\ (SP) & \text{s.t.} & \langle A_i, X\rangle = b_i, \quad i = 1, \ldots, m, \\ & & X \in \mathcal{S}^n_+, \end{array}$$

where $C \in \mathcal{S}^n$, $b = (b_1, \ldots, b_m)^T \in \mathcal{R}^m$, and $A_i \in \mathcal{S}^n$, $i = 1, \ldots, m$, are linearly independent.

In the linear space of $n \times n$ matrices, we define the inner product in $\mathcal{R}^{n\times n}$ as

$$\langle C, X\rangle = \mathrm{Tr}(C^T X) = \sum_{i=1}^{n}\sum_{j=1}^{n} C_{ij}X_{ij}.$$

Since in our case C and X are symmetric, we may simply denote $\langle C, X\rangle = \mathrm{Tr}(CX)$. If we think of X as an array of n^2 components in the form $\mathbf{vec}(X) = (x_{11}, \cdots, x_{m1}, x_{12}, \cdots, x_{m2}, \cdots, x_{n1}, \cdots, x_{nn})^T$, then the inner product of matrices coincides with the dot product of vectors. We also use $X \succeq 0$ instead of $X \in \mathcal{S}^n_+$ when X is symmetric positive semidefinite. With these notations, it is easy to see that (SP) is equivalent to

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

Again, $C,\ X \in \mathcal{S}^n$, and we assume that $A_i \in \mathcal{S}^n$, $i = 1, \ldots, m$, are linearly independent and $b = (b_1, \ldots, b_m)^T \in \mathcal{R}^m$. We call problem (P) in the given form the primal problem, and X is the primal matrix variable. A primal variable X of (P) is called *primal feasible* if

$$X \in \mathcal{F}_P := \{X \succeq 0 \mid \mathrm{Tr}(A_i X) = b_i,\ i = 1, \ldots, m\}.$$

In particular, we say that X is a strictly feasible primal solution if

$$X \in \mathcal{F}_P^0 := \{X \succ 0 \mid \mathrm{Tr}(A_i X) = b_i,\ i = 1, \ldots, m\},$$

where $X \succ 0$ denotes that X is positive definite. A primal feasible solution X^* is called a primal optimal solution if $\mathrm{Tr}(CX^*) \leq \mathrm{Tr}(CX)$ for all $X \in \mathcal{F}_P$.

Note that we are working in the space of symmetric positive semidefinite matrices. The dual of problem (P) [8] can be written in the form

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

where $y \in \mathcal{R}^n$ and $S \in \mathcal{S}^n$. A solution (y, S) is called *dual feasible* if

$$(y, S) \in \mathcal{F}_D := \left\{ S \succeq 0 \;\middle|\; \sum_{i=1}^{m} y_i A_i + S = C \right\}.$$

In particular, we say that (y, S) is a strictly dual feasible solution if

$$(y, S) \in \mathcal{F}_D^0 := \left\{ S \succ 0 \;\middle|\; \sum_{i=1}^{m} y_i A_i + S = C \right\}.$$

A dual feasible solution (y^*, S^*) is called a dual optimal solution if $b^T y^* \geq b^T y$ for all $(y, S) \in \mathcal{F}_D$.

Strictly speaking, we should write "inf" and "sup" instead of "min" and "max". Not only because the problems might be unbounded, but also because the optimal values might not be attained even if they are finite. We will present some examples in the next subsection where positive duality gap or non-attainment of the optimal value is demonstrated. We, however, still choose to stick with the notations "min" and "max" in this thesis due to the fact that we shall impose conditions that ensure that the optimal values are actually attained when they are finite.

Consider the LO problem (LP) and let $X = \mathrm{Diag}(x)$, $S = \mathrm{Diag}(s)$, $C = \mathrm{Diag}(c)$ and $A_i = \mathrm{Diag}(a_i)$, $i = 1, \ldots, m$, then SDO problems (P) and (D) are exactly LO problems (LP) and (LD), respectively. This representation also shows that SDO is a generalization of LO to the space of positive semidefinite matrices.

2.2.2 Duality Theory

Analogous to LO, we are interested in the relation between (P) and (D) as well. In this section, we investigate how a feasible solution of the primal problem (P) implies a bound of the optimal value for the dual problem (D), and vice verse.

First, we note the following trivial but key fact that we call it *weak duality* theorem.

Theorem 2.2.1 (Weak Duality). *If $X \in \mathcal{F}_P$ in (P) and $(y, S) \in \mathcal{F}_D$ in (D), then*

$$\mathrm{Tr}(CX) - b^T y = \mathrm{Tr}(XS) \geq 0.$$

Proof. From the feasibility of X and (y, S), we have

$$\begin{aligned} \mathrm{Tr}(CX) - b^T y &= \mathrm{Tr}\left(\left(\sum_{i=1}^{m} y_i A_i + S\right) X\right) - b^T y \\ &= \mathrm{Tr}\left(\sum_{i=1}^{m} y_i A_i X\right) + \mathrm{Tr}(SX) - b^T y \\ &= \sum_{i=1}^{m} y_i \mathrm{Tr}(A_i X) + \mathrm{Tr}(XS) - b^T y \\ &= \mathrm{Tr}(XS) \\ &\geq 0. \end{aligned}$$

Here we used the fact that $\mathrm{Tr}(SX) = \mathrm{Tr}(XS)$. ■

The difference between the primal and dual objective values at feasible solutions of (P) and (D), which is always nonnegative by Proposition 2.2.1, is called the *duality gap*. *Strong duality* is the assertion that the duality gap is zero and both of (P) and (D) attain their optimal value whenever both of them are feasible. Although this is always true for LO, it does fail for SDO occasionally.

We illustrate two examples from Luo, Sturm, and Zhang [19] and Vandenberghe and Boyd [34] respectively, to show how strong duality fails.

Example 2.2.2. Consider the following problem

$$\max \ -y_1, \quad \text{s.t.} \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} y_1 + \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} y_2 \preceq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The constraint is equivalent to $\begin{pmatrix} y_1 & 1 \\ 1 & y_2 \end{pmatrix} \succeq 0$. Then, we have the feasible region is $\{(y_1, y_2) \mid y_1 > 0,\ y_2 > 0,\ y_1 y_2 \geq 1\}$. To maximize $-y_1$, it is possible to choose y_1 arbitrarily small such that the optimal value is 0, but is never attained. The dual form of this problem is

$$\begin{array}{lrcl} \min & \mathrm{Tr}\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} X\right) & & \\ \text{s.t.} & \mathrm{Tr}\left(\begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} X\right) & = & -1, \\ & \mathrm{Tr}\left(\begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} X\right) & = & 0, \\ & X & \succeq & 0, \end{array}$$

for which the only feasible, hence optimal, solution is $X = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ with optimal value 0. Hence, the duality gap for this pair of primal and dual problems could be arbitrarily close to 0, but it never vanishes.

Example 2.2.3. Consider the problem

$$\begin{array}{lrcl} \min & \mathrm{Tr}\left(\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} X\right) & & \\ \text{s.t.} & \mathrm{Tr}\left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} X\right) & = & 0, \\ & \mathrm{Tr}\left(\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} X\right) & = & 2, \\ & X & \succeq & 0. \end{array}$$

Note the fact that if a diagonal element of a positive semidefinite matrix is 0 the elements in the corresponding column and row are all 0, then we acquire any feasible solution X is of the form $\begin{pmatrix} 0 & 0 & 0 \\ 0 & \omega_1 & \omega_2 \\ 0 & \omega_2 & 1 \end{pmatrix}$. It follows that the

optimal value is 1 with $\omega_1 = 0$ and $\omega_2 = 0$. The dual form is

$$\max\ 2y_2, \quad \text{s.t.} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} y_1 + \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} y_2 \preceq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Equivalently, it requires that

$$S = \begin{pmatrix} -y_1 & -y_2 & 0 \\ -y_2 & 0 & 0 \\ 0 & 0 & 1 - 2y_2 \end{pmatrix} \succeq 0,$$

i.e., $y_1 \geq 0$ and $y_2 = 0$. Therefore, it follows that the optimal solution is $(0, 0)^T$ with an optimal value 0. Here both of the primal and dual problems attain their optimal values, but with a positive duality gap.

From these examples, we see that unlike in the case of LO, we cannot assert that the duality gap vanishes if either the primal or the dual problem has an optimal solution, unless certain conditions are imposed. One such condition is summarized in the following theorem. Before doing that, it is necessary to make some assumptions which is assumed throughout the entire thesis.

Assumption 2.2.4. *The matrices A_i, $i = 1, \ldots, m$, are linearly independent.*

Assumption 2.2.5. $\mathcal{F}^0 := \mathcal{F}_D^0 \times \mathcal{F}_P^0$ *is nonempty.*

Theorem 2.2.6 (Strong Duality)[1]. *Let X^* and (y^*, S^*) denote the optimal solutions of (P) and (D), respectively. Under Assumptions 2.2.4 and 2.2.5, then both (P) and (D) attain their optimal values with zero duality gap, i.e.,*

$$\mathrm{Tr}(CX^*) - b^T y^* = \mathrm{Tr}(X^* S^*) = 0.$$

Eventually, if we assume that the strong duality theorem holds, then X^* and (y^*, S^*) are optimal if and only if they satisfy the optimality conditions [8],

$$\begin{aligned} \mathrm{Tr}(A_i X^*) &= b_i,\ X^* \succeq 0,\ i = 1, \ldots, m \\ \sum_{i=1}^{m} y_i^* A_i + S^* &= C,\ S^* \succeq 0, \\ X^* S^* &= 0, \end{aligned} \tag{2.2.1}$$

[1]For a proof, we refer to Todd [31] and de Klerk [8].

where we replace the strong duality requirement $\mathrm{Tr}(X^*S^*) = 0$ by $X^*S^* = 0$, since they are essentially equivalent. Note that the first two equations request the primal and dual feasibility, respectively. The last equation is called *complementarity condition*, which is a nonlinear function. The complementarity condition guarantees optimality and the solution of (2.2.1) is a challenging problem.

Chapter 3

Interior Point Methods

To start this chapter, we introduce the barrier method which serves to exhibit the basic idea of path-following IPMs and many relevant concepts, such as centrality and neighborhood. We consider the Newton method to solve the optimality condition equations which are derived from the barrier method. However, unlike in the case of LO, the Newton direction usually does not exist for SDO, since a symmetric search direction is not guaranteed to exist for the Newton system. One general remedy, originally proposed by Zhang [39], is stated in Section 3.3. At last, we present two frameworks of primal-dual path-following IPMs based on the small and large neighborhoods.

3.1 Barrier Method

We define the *logarithmic barrier function* over the cone of positive definite matrices $\mathcal{S}^n_{++}$ by

$$f(X) := \begin{cases} -\ln\det X & \text{if } X \succ 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{3.1.1}$$

We say $f(X)$ has the barrier property over $\mathcal{S}^n_{++}$. For the simple case $n = 1$, we get a smooth function $-\ln x$ which is defined on the positive axis and whose function value tends to $+\infty$ as x approaches 0. In general, $f(X)$ goes to $+\infty$ when X approaches the boundary of the cone of positive definite matrices.

Let $X \succ 0$, $\Delta X \in \mathcal{S}^n$ and $\|\Delta X\| \leq \delta$, where δ is an arbitrarily small positive

number. Then, one has

$$\begin{aligned} f(X+\Delta X) &= -\ln\det\left[X(I+X^{-1}\Delta X)\right] \\ &= -\ln\det X - \ln\left(1+\mathrm{Tr}(X^{-1}\Delta X)\right) \\ &= f(X) - \mathrm{Tr}(X^{-1}\Delta X), \end{aligned}$$

from which we derive that

$$f'(X) = -X^{-1}.$$

To utilize the barrier property of $f(X)$, we implicitly enforce the positive semidefiniteness in (P) to be embedded in the objective function. Then, it follows a sequence of parameterized primal and dual problems

$$BP(\mu)\quad \begin{array}{ll} \min & \mathrm{Tr}(CX)+\mu f(X) \\ \text{s.t.} & \mathrm{Tr}(A_iX)=b_i, \quad i=1,\ldots,m, \\ & X \succ 0, \end{array}$$

and

$$BD(\mu)\quad \begin{array}{ll} \max & b^Ty - \mu f(S) \\ \text{s.t.} & \sum_{i=1}^{m} y_iA_i + S = C, \\ & S \succ 0, \end{array}$$

where $\mu > 0$ is referred to as the *barrier parameter*.

Note that $BP(\mu)$ is to minimize a convex function, $\mathrm{Tr}(CX)+\mu f(X)$, with linear constraints, $\mathrm{Tr}(A_iX)=b_i,\ i=1,\ldots,m$, over a convex set, $X \succ 0$. The *Lagrange conditions* are necessary and sufficient. Thus, we are able to derive the optimality conditions from *Lagrange function*

$$\mathcal{L}(X,y) = \mathrm{Tr}(CX)+\mu f(X) - \sum_{i=1}^{m} y_i(\mathrm{Tr}(A_iX)-b_i),$$

where $y = (y_1,\ldots,y_m)^T$ is the Lagrange multipliers. From the Lagrange's theorem [28], X is an optimal solution of $BF(\mu)$, if and only if there exists some $y \in \mathcal{R}^m$ such that

$$\begin{aligned} \mathrm{Tr}(A_iX) &= b_i,\ i=1,\ldots,m \\ \nabla_X\mathcal{L}(X,y) &= C+\mu f'(X) - \sum_{i=1}^{m} y_iA_i \\ &= C-\mu X^{-1} - \sum_{i=1}^{m} y_iA_i \\ &= 0. \end{aligned}$$

Let us denote $S := \mu X^{-1}$. Then we have the following optimality conditions for $BP(\mu)$

$$\begin{aligned} \mathrm{Tr}(A_i X) &= b_i,\ X \succ 0,\ i = 1, \ldots, m \\ \sum_{i=1}^{m} y_i A_i + S &= C,\ S \succ 0, \\ XS &= \mu I. \end{aligned} \tag{3.1.2}$$

The set of equations (3.1.2) is also the optimality conditions for the dual barrier problem $BD(\mu)$. In fact, the first two equations in equations (3.1.2) keep primal and dual feasibility and the last equation in equations (3.1.2) can be considered as a perturbation of the complementarity condition in the equations (2.2.1).

3.2 Central Path and Neighborhood

We show the existence and uniqueness of solutions to equations (3.1.2) for every positive μ. For all $\mu > 0$, the solutions $(X(\mu), y(\mu), S(\mu)) \in \mathcal{S}^n_{++} \times \mathcal{R}^m \times \mathcal{S}^n_{++}$ to equations (3.1.2) form a smooth curve, which is called the *central path*. Note that equations (3.1.2) look almost the same as equations (2.2.1), except the right-hand-side of the last equation. Equations (3.1.2) approximate (2.2.1) more and more closely as μ tends to zero. If the central path converges to any point as $\mu \to 0$, it must be an optimal solution to (P) and (D) because the central path is differentiable, as it is showed later in this subsection.

The next theorem proves the existence and uniqueness of the central path.

Theorem 3.2.1[1] *Suppose that Assumptions 2.2.4 and 2.2.5 hold. Then for every $\mu > 0$, there is a unique solution $(X(\mu), y(\mu), S(\mu)) \in \mathcal{S}^n_{++} \times \mathcal{R}^m \times \mathcal{S}^n_{++}$ to the cental path equations (3.1.2). Further, $X(\mu)$ and $(y(\mu), S(\mu))$ are the unique solutions to $BP(\mu)$ and $BD(\mu)$, respectively.*

So far, it is clear that for any $\mu > 0$, the set of equations (3.1.2) has a unique solution and the set of solutions for all $\mu > 0$ forms the central path. The following theorem reveals that the central path is also differentiable and it makes explicit how the duality gap is associated with the barrier parameter μ.

[1]For a proof, we refer to Todd [31] and de Klerk [8].

Theorem 3.2.2.[2] *Suppose that Assumptions 2.2.4 and 2.2.5 hold. Then the set of solutions to* (3.1.2) *for all* $\mu > 0$ *forms a nonempty infinitely many times differentiable path, called central path. If* $(X(\mu), y(\mu), S(\mu))$ *solves equations* (3.1.2) *for a specific positive* μ*, then* $X(\mu)$ *is a strictly feasible solution to* (P)*, and* $(y(\mu), S(\mu))$ *is a strictly feasible solution to* (D) *with duality gap*

$$\mathrm{Tr}(CX(\mu)) - b^T y(\mu) = \mathrm{Tr}(X(\mu)S(\mu)) = n\mu.$$

Motivated by Theorem 3.2.2, path-following algorithms attempt to track the points on the central path which leads to an optimal solution when μ is steadily decreasing to zero. Usually, we use a primal-dual algorithm, i.e., maintaining both the X and the (y, S) iterates simultaneously.

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

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

where $\theta \in (0, 1)$ and $\mu_g := \mathrm{Tr}(XS)/n$ is associated to 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 $\gamma \in (0, 1)$.

For the special case of SDO, i.e., LO, the small neighborhood is equivalent to

$$\mathcal{LN}_2(\theta) := \left\{ (x, y, s) \in \mathcal{F}_L^0 \,|\, \|xs - \mu_g e\| \leq \theta \mu_g \right\},$$

where $\theta \in (0, 1)$ and $\mu_g := x^T s/n$; the counterpart of the large neighborhood is given as

$$\mathcal{LN}_\infty^-(1-\gamma) := \left\{ (x, y, s) \in \mathcal{F}_L^0 \mid xs \geq \gamma \mu_g e \right\},$$

where $\gamma \in (0, 1)$ and $\mathcal{F}_L^0$ denotes the interior of $\mathcal{F}_L$.

[2] For a proof, we refer to Todd [31] and de Klerk [8].

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 large neighborhoods, e.g., large update algorithms. However, computational experience [28, 37] shows that large neighborhood IPMs typically perform much better in practice than small neighborhood algorithms. In Chapter 4 of this thesis, we will explore a variant of large neighborhood path-following IPMs and prove its polynomial iteration complexity which will coincide with the complexity of small neighborhood IPMs.

3.3 Symmetrization and Search Direction

One of the crucial problem of the path-following IPMs is to choose how to solve system (3.1.2). Due to the last equation, system (3.1.2) is nonlinear. Therefore, it is not trivial to get a solution to system (3.1.2). Usually, people would like to use (damped) Newton methods to solve it.

Given an iterate (X, y, S), path-following IPMs generate the next iterate by taking a Newton step to system (3.1.2). Let us target the point on the central path corresponding to $\mu = \tau\mu_g$, where $\tau \in [0, 1]$ is called the centering parameter and $\mu_g = \mathrm{Tr}(XS)/n$ corresponds to the actual duality gap. To move from the current point (X, y, S) towards the target on the central path, we wish a symmetric search direction $(\Delta X, \Delta y, \Delta S) \in \mathcal{S}^n \times \mathcal{R}^n \times \mathcal{S}^n$ is available from the following linear system

$$\begin{aligned} \mathrm{Tr}(A_i\Delta X) &= 0, \\ \sum_{i=1}^{m} \Delta y_i A_i + \Delta S &= 0, \\ \Delta XS + X\Delta S &= \tau\mu_g I - XS. \end{aligned} \tag{3.3.1}$$

From the second equality we have a symmetric ΔS, however, system (3.3.1) do not allow a symmetric solution matrix ΔX. Various remedies are proposed since the middle of 1990's.

One of the ideas was proposed by Alizadeh, Haeberly, and Overton [5]. They suggested to replace the last equation in system (3.3.1) by

$$\frac{1}{2}(\Delta XS + X\Delta S + X\Delta S + \Delta XS) = \tau\mu_g I - \frac{1}{2}(XS + SX).$$

In [32], Todd, Toh, and Tütüncü pointed out that the resulting search direction, called the AHO direction, gives a unique symmetric ΔX and ΔS only

if $XS + SX$ is positive semidefinite. Otherwise, it might be not well-defined. Another difficulty with the AHO direction is that there is still no polynomial complexity proof for large neighborhood, i.e., long step, IPMs based on the AHO direction.

Another possible alternative is to employ a similarity transformation $P(\cdot)P^{-1}$ on both sides of the third equation of system (3.3.1). This strategy was first investigated by Monteiro [20] 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 [16]. At the same time, another motivation led Helmberg, Rendl, Vanderbei and Wolkowicz [12] 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.

The last popular direction we would like to introduce, the NT direction, was first proposed by Nesterov and Todd [25, 26] in their attempt to generalize primal-dual IPMs beyond SDO. The NT direction not only has many nice properties in theory but is also robust and accurate in practice. It is widely used in state-of-the-art SDO software, such as SeDuMi and SDPT3.

In [39], 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, which turns out to involve all the aforementioned directions, i.e., AHO, H..K..M and NT directions. Zhang suggested to replace the last equation in system (3.1.2) by

$$H_P(XS) \;=\; \mu I, \tag{3.3.2}$$

where $H_P(\cdot)$ is a symmetrization transformation that is defined for a given matrix M and a given nonsingular matrix P as

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

In particular, if $P = I$ then for any symmetric matrix M, $H_I(M) = M$. In [39], 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

$$\mathrm{Tr}(A_i \Delta X) \;=\; 0, \tag{3.3.3a}$$

$$\sum_{i=1}^{m} \Delta y_i A_i + \Delta S = 0, \tag{3.3.3b}$$

$$H_P(\Delta X S + X \Delta S) = \tau \mu_g I - H_P(XS). \tag{3.3.3c}$$

For the choices of P, when $P = I$, the direction obtained from (3.3.3) coincides with the AHO direction [5]. If $P = X^{-1/2}$ or $S^{1/2}$, then (3.3.3) gives the H..K..M directions [12, 16, 20, 21], 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 [25, 26] 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 by (3.3.3) as the Monteiro-Zhang (MZ) family.

In terms of Kronecker product[3], equation (3.3.3c) can be expressed as

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

where

$$E = \frac{1}{2}(S \otimes I + I \otimes S), \quad F = \frac{1}{2}(X \otimes I + I \otimes X). \tag{3.3.4}$$

In [32], Todd, Toh and Tütüncü proved that system (3.3.3) 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 $PXSP^{-1} \in \mathcal{S}^n$. Apparently, $P = X^{-1/2}$, $S^{1/2}$ and $W_{NT}^{-1/2}$ belong to this specific class. However, $P = I$ does not. In [32], the authors proved that the solution to system (3.3.3) is not uniquely defined.

The next result shows that the solution set to system (3.3.3) keep the same in terms of $V = P^T P$ not changing.

Theorem 3.3.1[4] *The set of solutions to system* (3.3.3) *remains invariant as long as the matrix* $V = P^T P$ *does not change.*

Thus, for fixed $V \in \mathcal{S}_{++}^n$, there is no loss of generality in considering only the matrix $V^{1/2}$ among all those scaling matrices P such that $P^T P = V$, since their corresponding system (3.3.3) all have the same solution set. Hence, we will assume $P \in \mathcal{S}_{++}^n$ in the remaining of this thesis, with the exception of Section 4.3.2.

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

[4]For a proof, we refer to Todd, Toh and Tütüncü [32].

3.4 Primal-Dual Path-following Algorithms

From the previous discussions, the motivation of path-following IPMs should be clear: starting with a relatively big but fixed μ, e.g., $\mu = 1$, take a (damped) Newton step along the directions derived from system (3.3.3). After updating the iterates, a decrease in duality gap is obtained and one repeats the process until the required precision is obtained. In practice, we usually use primal-dual methods, i.e., maintain the iterates X and (y, S) in the same time.

Algorithm 1 presents a generic framework of primal-dual path-following IPMs.

Algorithm 1 Generic Primal-Dual Path-following IPMs

Input:
required precision $\epsilon > 0$; neighborhood $\mathcal{N}$
an initial point $(X^0, y^0, S^0) \in \mathcal{F}_P^0 \times \mathcal{F}_D^0$ with $\mu_g^0 = \mathrm{Tr}(X^0 S^0)/n$;

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

(1) Compute the scaling matrix P^k and choose a centering parameter $\tau_k \in [0, 1]$.

(2) Compute the directions $(\Delta X^k, \Delta y^k, \Delta S^k)$ by (3.3.3).

(3) Find a step length $\alpha^k > 0$ giving a sufficient reduction of the duality gap and assuring $(X^k + \alpha^k \Delta X^k, y^k + \alpha^k \Delta y^k, S^k + \alpha^k \Delta S^k) \in \mathcal{N}$.

(4) Set $(X^{k+1}, y^{k+1}, S^{k+1}) = (X^k + \alpha^k \Delta X^k, y^k + \alpha^k \Delta y^k, S^k + \alpha^k \Delta S^k)$.

(5) Set $\mu_g^{k+1} := \mathrm{Tr}(X^{k+1} S^{k+1})/n$ and $k := k + 1$.

end while

In this framework, the algorithms starts with a strictly feasible point. However, people usually utilize either infeasible algorithms or self-dual embedding models. The first strategy is reducing the infeasibility and the duality gap simultaneously; the later one is able to find a trivial strictly feasible point, but works with a larger problem in terms of dimension.

In Algorithm 1, we use the notation $\mathcal{N}$ to denote a certain neighborhood of the central path. According to various algorithms, it might be the small neighborhood or the large neighborhood, i.e. the negative infinity neighborhood, or even our new neighborhood method in the next chapter. In the subsequent subsections of this chapter, we show the main complexity results for both of small and large neighborhood algorithms. Theoretically, the algorithms

based on the small neighborhood are proved to be better than those based on the large neighborhood, in terms of the number of iterations. Nevertheless, in practice the situation seems to be the opposite: large neighborhood algorithms perform much better than small neighborhood ones. The main work in this thesis is motivated by this contradiction. We propose a class of large neighborhood algorithms for SDO whose complexity bound coincide with the one for small neighborhood algorithms. This new algorithm is presented in Chapter 4.

We will state two classical primal-dual path-following algorithms based on the small and large neighborhoods, respectively. The choices of the centering parameter τ also depend on the algorithms used. The details are discussed subsequently.

3.4.1 Small Neighborhood Algorithm

In this subsection, we state a small neighborhood algorithm based on the MZ family, see e.g. Monteiro [20]. The algorithm generates iterates in the small neighborhood $\mathcal{N}_F(\beta)$ of the central path and selects step sizes $\alpha = 1$, i.e., it takes a full Newton step, and centrality parameters $\tau_k = 1 - \delta/\sqrt{n}$, where δ is a constant specified in Theorem 3.4.1.

Theorem 3.4.1[5] *Let $\beta \in (0, 1/2)$ and $\delta \in [0, \sqrt{n})$ be constants satisfying*

$$\frac{\beta^2 + \delta^2}{2(1-\beta)^2(1-\delta/\sqrt{n})} \leq \beta.$$

Suppose that $(X, y, S) \in \mathcal{N}_F(\beta)$ and let $(\Delta X, \Delta y, \Delta S)$ denote the solution of (3.3.3) and $\tau = 1 - \delta/\sqrt{n}$. Then, every iterate (X^k, y^k, S^k) generated by the small neighborhood algorithm is in the neighborhood $\mathcal{N}_F(\beta)$. Moreover, the algorithm terminates in at most $O(\sqrt{n}\log(1/\epsilon))$, where ϵ is the required precision.

A pair β, δ satisfying the conditions stated in Theorem 3.4.1 is $\beta = 0.3$ and $\delta = 0.3$. Using these parameters, we provide a variant of the algorithm.

[5] For the proof, we refer to Monteiro and Todd [22].

Algorithm 2 Path-following IPMs based on small neighborhood $\mathcal{N}_F(\beta)$

Input:
required precision $\epsilon > 0$;
neighborhood parameter $\beta = 0.3$;
parameter associated with the centrality $\delta = 0.3$;
an initial point $(X^0, y^0, S^0) \in \mathcal{N}_F(\beta)$ with $\mu_g^0 = \mathrm{Tr}(X^0 S^0)/n$;

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

(1) Compute the scaling matrix $P^k \in \mathcal{P}(X^k, S^k)$ and set the centering parameter $\tau_k = 1 - \delta/\sqrt{n}$.

(2) Compute the directions $(\Delta X^k, \Delta y^k, \Delta S^k)$ by (3.3.3).

(3) Set $(X^{k+1}, y^{k+1}, S^{k+1}) = (X^k + \Delta X^k, y^k + \Delta y^k, S^k + \Delta S^k)$.

(4) Set $\mu_g^{k+1} := \mathrm{Tr}(X^{k+1} S^{k+1})/n$ and $k := k + 1$.

end while

The main difference between Algorithm 1 and 2 is that line search for largest step length is not required, since a full Newton step guarantees the feasibility of the iterates and polynomial iteration complexity bound. From Theorem 3.4.1, we know that the iteration complexity of small neighborhood algorithms do not depend on the scaling matrix P, i.e., no matter which member of the MZ family is used, the iteration complexity does not change theoretically. Nevertheless, this fact is not true for large neighborhood algorithms.

3.4.2 Large Neighborhood Algorithm

Although small neighborhood algorithms possess the best known iteration complexity bound, they are less closely related to practical algorithms than algorithms based on a large neighborhood. Because the small neighborhood $\mathcal{N}_F(\beta)$ contains only a small fraction of the points in the strictly feasible set $\mathcal{F}^0$, so algorithms based on this neighborhood do not have much room to maneuver, and the amount of progress they can achieve at each iteration is limited. The large neighborhood $\mathcal{N}_\infty^-(1-\gamma)$, on the other hand, is much more expansive. When γ is small, it might take up almost the entire strictly feasible set $\mathcal{F}^0$.

In this subsection, we summarize Monteiro and Zhang's work [23], in which they proposed a unified analysis for large neighborhood algorithms. Different

from small neighborhood algorithms, large neighborhood path-following algorithms do not allow all members of the MZ family. Thus, we further restrict the scaling matrices P to the specific class

$$\mathcal{P}(X,S) := \{P \in \mathcal{S}^n_{++} \mid PXSP^{-1} \in \mathcal{S}^n\}, \tag{3.4.1}$$

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. In other words, the polynomial complexity results for large neighborhood algorithms are only valid for the H..K..M and NT directions, but not the AHO direction. Furthermore, this restriction on P does not lose any generality, in terms of the solution set of system (3.3.3), as Monteiro proves in [21].

Let τ_k be a constant in every iteration, say $\tau_k = \tau$. We present a framework of large neighborhood IPMs in Algorithm 3.

Algorithm 3 Path-following IPMs based on large neighborhood $\mathcal{N}_\infty^-(1-\gamma)$

Input:
required precision $\epsilon > 0$;
neighborhood parameter $\gamma \in (0,1)$;
centrality parameter $\tau_k = \tau$;
an initial point $(X^0, y^0, S^0) \in \mathcal{N}_\infty^-(1-\gamma)$ with $\mu_g^0 = \mathrm{Tr}(X^0S^0)/n$;

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 (3.3.3).

(3) Find a the largest step length $\alpha^k > 0$ such that

$$(X^k + \alpha^k \Delta X^k, y^k + \alpha^k \Delta y^k, S^k + \alpha^k \Delta S^k) \in \mathcal{N}_\infty^-(1-\gamma).$$

(4) Set $(X^{k+1}, y^{k+1}, S^{k+1}) = (X^k + \alpha^k \Delta X^k, y^k + \alpha^k \Delta y^k, S^k + \alpha^k \Delta S^k)$.

(5) Set $\mu_g^{k+1} := \mathrm{Tr}(X^{k+1}S^{k+1})/n$ and $k := k+1$.

end while

The next theorem gives an iteration-complexity bound for Algorithm 3 in terms of a parameter κ_∞ defined as

$$\kappa_\infty = \sup\left\{\mathrm{cond}((E^k)^{-1}F^k) : \ k = 0, 1, \dots\right\}, \tag{3.4.2}$$

where E and F are defined by (3.3.4). Obviously, $\kappa_\infty \geq 1$.

Theorem 3.4.2[6] *Assume that $\kappa_\infty < \infty$. Then the sequence $\{\mu_k\}$ generated by the large neighborhood algorithm satisfies*

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

where

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

Consequently, Algorithm 3 terminates in at most $O(\sqrt{\kappa_\infty} n \log(1/\epsilon))$ iterations.

From Lemma B.0.20, it is easy to obtain the following corollary.

Corollary 3.4.3[7] *Algorithm 3 based on the NT direction and the H..K..M directions have iteration complexity bounds equal to $O(n \log(1/\epsilon))$ and $O(n^{3/2} \log(1/\epsilon))$, respectively.*

Evidently, the iteration complexity of large neighborhood algorithms depends on the scaling matrix P. A sufficient condition is that $P \in \mathcal{P}(X, S)$ is required. Among all of the choices for P, the NT scaling achieves the best iteration result, since for the NT direction $\kappa_\infty = 1$. Nevertheless, although the AHO direction is proved to be convergent, but no polynomial complexity is known so far.

[6] For the proof, we refer to Monteiro and Todd [22].
[7] For the proof, we refer to Monteiro and Todd [22].

Chapter 4

New Interior Point Methods

In this chapter, we intend to propose our new algorithm, which is based on a large neighborhood but with an $O(\sqrt{n}\log(1/\epsilon))$ iteration complexity when the NT scaling is used.

We organize this chapter as follows. We first review the Ai-Zhang algorithm in Section 4.1. Then, from Section 4.2, we extend their algorithm to SDO. We first 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 4.3, we explain the way to decompose the classical Newton direction and present the framework of our algorithm. We also suggest a computationally cheap methodology to apply our algorithm to practical implementation in the same section. In Section 4.4, the theoretical complexity bound and the convergence analysis are presented. Following by the technical lemmas in Section 4.4.2, we present the most important polynomial complexity result in Section 4.4.3.

4.1 Ai-Zhang Algorithm

In this section, we summarize Ai-Zhang algorithm for LO. In [2], Ai and Zhang proved the convergence and polynomial iteration complexity for linear complementarity problems (LCP). To make their idea more accessible, in this thesis we only concentrate on LO, which is a special case of LCP. On the other hand, it is more straightforward to extend their algorithm to SDO from LO.

Recall that if $X = \text{Diag}(x)$, $S = \text{Diag}(s)$, and $A_i = \text{Diag}(a_i)$, $i = 1, \ldots, m$, then SDO problems (P) and (D) are LO problems (LP) and (LD), respectively. In this case, system (3.3.1) is identical to

$$\begin{aligned} \langle a_i, \Delta x \rangle &= 0, \ i = 1, \ldots, m \\ \sum_{i=1}^{m} \Delta y_i a_i + \Delta s &= 0, \\ S\Delta x + X\Delta s &= \tau \mu_g e - xs. \end{aligned} \tag{4.1.1}$$

Note that both X and S are diagonal, hence symmetric, then there is no trouble with symmetrization as general SDO.

In the following part, we would like to explain Ai-Zhang's idea in [2] that motivated our work. Before proceeding, we need to introduce a new notation. For any vector $q \in \mathcal{R}^n$, q^+ denotes its positive part, i.e., $(q^+)_i = \max\{q_i, 0\}$ and q^- denotes its negative part, i.e., $(q^-)_i = \min\{q_i, 0\}$. Later in Section 4.2.1, we will apply this notation to any symmetric matrix. Nevertheless, we stick this notation with vector in this section.

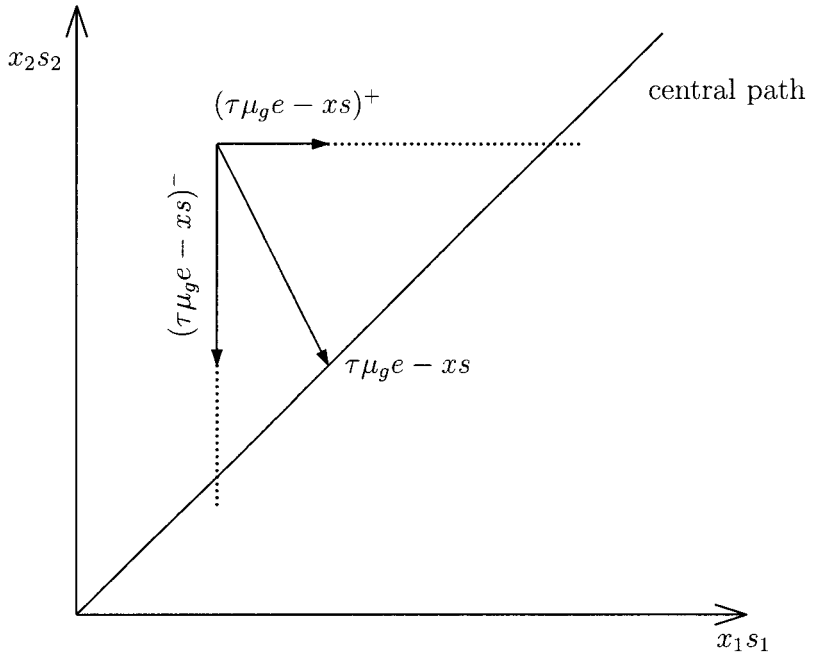

Figure 4.1: Decomposition of the Newton direction

Suppose we have an iterate (x, y, s) in a certain neighborhood which is able to warrantee the convergence, namely, (x, y, s) is feasible and reasonably away from the central path. In the next iterations, IPMs aim at a target point on the central path, say, $\tau\mu_g e$. Figure 4.1 illustrates the classical Newton direction, from xs to $\tau\mu_g e$. Note that $\tau\mu_g e = xs + (\tau\mu_g e - xs)^- + (\tau\mu_g e - xs)^+$ and the fact that $x^T s$ indicates the current duality gap, hence $(\tau\mu_g e - xs)^-$, which is component-wisely non-positive and shooting towards another point on the central path with a smaller duality gap than $\tau\mu_g e$ in this special two dimensional case as illustrated in Figure 4.1, plays the role to reduce the duality gap. However, for higher dimensions, i.e., $n \geq 3$, $(\tau\mu_g e - xs)^-$ might not exactly aim at a point on the central path associated with a smaller duality gap, but provides us the possibility to head to it. In the other hand, $(\tau\mu_g e - xs)^+$ is trying to drag the iterate back close to the central path, namely, protecting the duality gap from decreasing too fast. From this geometrical intuition, Ai and Zhang [2] first proposed to decompose the classical Newton direction and treat the resulting components individually: one for optimality and the other one for centrality.

Another important ingredient of the new algorithm is to introduce a new neighborhood for the central path,

$$\mathcal{LN}(\tau_1, \tau_2, \eta) := \mathcal{LN}_\infty^-(1-\tau_2) \cap \left\{(x, y, s) \in \mathcal{F}_L^0 : \|\tau_1\mu_g e - xs\| \leq \eta(\tau_1 - \tau_2)\mu_g\right\},$$

where $\eta \geq 1$ and $0 < \tau_2 < \tau_1 < 1$. One can easily verify that

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

i.e., the new defined neighborhood is itself a large neighborhood.

To decompose the Newton direction, Ai and Zhang [2] suggested to solve the following two systems:

$$\begin{aligned} \langle a_i, \Delta x_- \rangle &= 0, \ i = 1, \ldots, m \\ \sum_{i=1}^m \Delta(y_i)_- a_i + \Delta s_- &= 0, \\ S\Delta x_- + X\Delta s_- &= (\tau\mu_g e - xs)^- \end{aligned} \tag{4.1.2}$$

and

$$\begin{aligned} \langle a_i, \Delta x_+ \rangle &= 0, \ i = 1, \ldots, m \\ \sum_{i=1}^m \Delta(y_i)_+ a_i + \Delta s_+ &= 0, \\ S\Delta x_+ + X\Delta s_+ &= (\tau\mu_g e - xs)^+ \end{aligned} \tag{4.1.3}$$

rather than system (4.1.1) in every iteration.

Let $\alpha := (\alpha_1, \alpha_2)^T \in \mathcal{R}^2_+$ be the step sizes taken along $(\Delta x_-, \Delta y_-, \Delta z_-)$ and $(\Delta x_+, \Delta y_+, \Delta z_+)$. Then, the new iterate is denoted by

$$(x(\alpha), y(\alpha), s(\alpha)) \;:=\; (x, y, s) + \alpha_-(\Delta x_-, \Delta y_-, \Delta s_-) + \alpha_+(\Delta x_+, \Delta y_+, \Delta s_+).$$

The best choice of the step sizes is the solution of the following optimization subproblem:

$$\begin{array}{ll} \min & \langle x(\alpha)s(\alpha)\rangle \\ \text{s.t.} & (x(\alpha), y(\alpha), s(\alpha)) \in \mathcal{LN}(\tau_1, \tau_2, \eta) \\ & 0 \le \alpha_- \le 1,\ 0 \le \alpha_+ \le 1. \end{array} \tag{4.1.4}$$

But in practice, it is not necessary and also expensive to exactly solve the above problem for step sizes. For details, we will discuss in Section 4.3.1.

So far, we are clear how to decide the search directions in Ai-Zhang's new algorithm. We describe a generic framework for their algorithm which is referred to [2].

Algorithm 4 The Ai-Zhang Large Neighborhood Algorithm

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 $(x^0, y^0, s^0) \in \mathcal{LN}(\tau_1, \tau_2, \eta)$ with $\mu_g^0 = \langle x^0, s^0\rangle/n$;

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

(1) Compute the directions $(\Delta x_-^k, \Delta y_-^k, \Delta s_-^k)$ by (4.1.2) and $(\Delta x_+^k, \Delta y_+^k, \Delta s_+^k)$ by (4.1.2).

(2) Find a step length vector $\alpha^k = (\alpha_-^k, \alpha_+^k) > 0$ giving a sufficient reduction of duality gap and assuring $(x(\alpha^k), y(\alpha^k), s(\alpha^k)) \in \mathcal{LN}(\tau_1, \tau_2, \eta)$.

(3) Set $(x^{k+1}, y^{k+1}, s^{k+1}) = (x(\alpha^k), y(\alpha^k), s(\alpha^k))$.

(4) Set $\mu_g^{k+1} := \langle x^{k+1}, s^{k+1}\rangle/n$ and $k := k + 1$.

end while

Following this framework, Ai and Zhang [2] proved the iteration bound stated as follows.

Theorem 4.1.1. *Suppose that* $\eta \geq 1$, $\tau = \tau_1 \leq 1/4$, *and* $0 \leq \frac{1}{2}\tau_1 \leq \tau_2 < \tau_1 < 1$ *are fixed for all iterations. Then, Algorithm 4 terminates in* $O(\sqrt{n}\log(1/\epsilon))$ *iterations, where* ϵ *is the required precision.*

After Peng, Roos, and Terlaky [27] proposed the self-regular based large neighborhood IPMs which could be arbitrarily closed to the small neighborhood IPMs, it is the Ai-Zhang algorithm that first achieved the same iteration bound for the large neighborhood as the result for the small neighborhood.

4.2 A New Neighborhood

In the remaining of this thesis, we will extend the Ai-Zhang algorithm to a more general class of SDO. In order to do so, we need to first redefine their neighborhood $\mathcal{LN}(\tau_1, \tau_2, \eta)$ for the SDO case.

4.2.1 Separation of Positive and Negative Parts

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

$$M^+ := \sum_{\lambda_i \geq 0} \lambda_i q_i q_i^T \ , \quad M^- := \sum_{\lambda_i \leq 0} \lambda_i q_i q_i^T . \tag{4.2.1}$$

In particular, for a real number $M \in \mathcal{S}^1$, M^+ denotes its positive part, i.e., $M^+ = \max\{M, 0\}$, and M^- denotes its negative 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 play a crucial role throughout the paper.

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

Proposition 4.2.1. *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.0.19, 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 4.2.2. *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 does not exceed, in the sense of Frobenius norm, its positive part after a similarity transformation.

Lemma 4.2.3. *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 4.2.4. *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 (4.2.1), 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 4.2.3.

Proof of Lemma 4.2.3. 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.0.18, 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 4.2.2, we know that

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

From Lemma 4.2.4, we claim

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

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

Proposition 4.2.1 and Lemmas 4.2.2 and 4.2.3 will play a crucial role in proving convergence and complexity of our new large neighborhood IPM.

4.2.2 Neighborhood $\mathcal{N}(\tau_1, \tau_2, \eta)$

With the notations above, we are ready to define a new neighborhood, using the positive part in (4.2.1), as

$$\begin{aligned} \mathcal{N}(\tau_1, \tau_2, \eta) \;:=\; & \mathcal{N}_\infty^-(1-\tau_2) \cap \\ & \left\{(X, y, S) \in \mathcal{F}^0 : \left\|[\tau_1 \mu_g I - X^{1/2} S X^{1/2}]^+\right\|_F \leq \eta(\tau_1 - \tau_2)\mu_g\right\}, \end{aligned} \tag{4.2.2}$$

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 4.2.5. *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

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

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

$$\tau_1\mu_g I - X^{1/2}SX^{1/2} \preceq 0, \tag{4.2.3}$$

which implies $[\tau_1\mu_g I - X^{1/2}SX^{1/2}]^+ = 0$, 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 4.2.6. *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

$$\mathcal{N}_\infty^-(1-\tau_2) \subseteq \left\{(X,y,S) \in \mathcal{F}^0 : \left\|[\tau_1\mu_g I - X^{1/2}SX^{1/2}]^+\right\|_F \leq \eta(\tau_1-\tau_2)\mu_g\right\}. \tag{4.2.4}$$

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

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

Therefore,

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

That implies

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

which proves that (4.2.4) holds when $\eta \geq \sqrt{n}$. ■

From Propositions 4.2.3 and 4.2.4, it is clear that this new neighborhood $\mathcal{N}(\tau_1, \tau_2, \eta)$ could include any given large neighborhood $\mathcal{N}(1-\gamma)$. Hence, $\mathcal{N}(\tau_1, \tau_2, \eta)$ is a large neighborhood itself.

4.3 Search Direction

In this section, we aim to present our strategy to decompose the classical Newton direction. The original idea was proposed by Ai and Zhang [2]. They suggested to treat the Newton direction as a combination of two separate directions for monotone linear complementarity problem (LCP), which is a class of more general problems including LO. Using different step sizes to each of the decomposed Newton direction, they proved their new algorithm terminates in $O(\sqrt{n}\log 1/\epsilon)$ iterations, where n is the problem size and ϵ is the required precision. In this section, we extend their idea to SDO.

4.3.1 Decomposition of the Newton Direction

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:

$$\begin{aligned} \mathrm{Tr}(A\Delta X_-) &= 0, && (4.3.1a)\\ \sum_{i=1}^{m}(\Delta y_i)_- A_i + \Delta S_- &= 0, && (4.3.1b)\\ H_P(\Delta X_- S + X\Delta S_-) &= [\tau\mu_g I - H_P(XS)]^-, && (4.3.1c) \end{aligned}$$

and

$$\begin{aligned} \mathrm{Tr}(A\Delta X_+) &= 0, && (4.3.2a)\\ \sum_{i=1}^{m}(\Delta y_i)_+ A_i + \Delta S_+ &= 0, && (4.3.2b)\\ H_P(\Delta X_+ S + X\Delta S_+) &= [\tau\mu_g I - H_P(XS)]^+, && (4.3.2c) \end{aligned}$$

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 (4.3.1c) and (4.3.2c) could be written, in Kronecker product form, as

$$E\mathbf{vec}(\Delta X_-) + F\mathbf{vec}(\Delta S_-) = \mathbf{vec}([\tau\mu_g I - H_P(XS)]^-) \qquad (4.3.3)$$

and

$$E\mathbf{vec}(\Delta X_+) + F\mathbf{vec}(\Delta S_+) = \mathbf{vec}([\tau\mu_g I - H_P(XS)]^+), \qquad (4.3.4)$$

respectively.

Obviously, systems (4.3.1) and (4.3.2) 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

$$\begin{array}{ll} \min & \mathrm{Tr}(X(\alpha)S(\alpha)) \\ \text{s.t.} & (X(\alpha), y(\alpha), S(\alpha)) \in \mathcal{N}(\tau_1, \tau_2, \eta) \\ & 0 \leq \alpha_- \leq 1,\ 0 \leq \alpha_+ \leq 1, \end{array} \tag{4.3.5}$$

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

$$\begin{aligned} (X(\alpha), y(\alpha), S(\alpha)) &:= (X, y, S) + (\Delta X(\alpha), \Delta y(\alpha), \Delta S(\alpha)) \\ &:= (X, y, S) + \alpha_-(\Delta X_-, \Delta y_-, \Delta S_-) + \alpha_+(\Delta X_+, \Delta y_+, \Delta S_+). \end{aligned} \tag{4.3.6}$$

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 (4.2.2) and the new search directions based on systems (4.3.1) and (4.3.2). Now, we are ready to present a generic framework for our algorithm.

Algorithm 5 Path-following IPMs 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 $\mu_g^0 = \mathrm{Tr}(X^0 S^0)/n$;

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 (4.3.1) and $(\Delta X_+^k, \Delta y_+^k, \Delta S_+^k)$ by (4.3.2).

(3) Find a step length vector $\alpha^k = (\alpha_-^k, \alpha_+^k) > 0$ giving a sufficient reduction of 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 (4.3.5) 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 (4.3.5) is also not a must. Even if we do not use the optimal solution of problem (4.3.5) as the step lengths, we are still able to achieve the polynomial convergence, as it is discussed later. Second, in spite of the fact that two linear systems (4.3.1) and (4.3.2) have to be solved, however, the additional cost is very marginal, since both of (4.3.1) and (4.3.2) 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 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 (4.3.1) and (4.3.2). However, we can utilize the strategy, scaling X and S to the same diagonal matrix, proposed by Todd, Toh and Tütüncü in [32] to obtain the negative and positive parts cheaply as a byproduct when computing the NT scaling matrix. We summarize the procedure in the coming subsection.

4.3.2 Computing Positive and Negative Parts

Obviously, computing the positive and negative parts explicitly is computation-averse, since the eigenvalue decomposition is very expensive. In this subsection, we will suggest a marginal way to compute the positive and negative parts of $\tau\mu_g I - H_P(XS)$ w.r.t NT scaling. It turns out that the positive and negative parts can be obtained together with the scaling matrix $P = W_{NT}^{-1/2}$ when NT scaling is employed.

Recall that when $P = W_{NT}^{-1/2}$, where $W_{NT} = X^{1/2}(X^{1/2}SX^{1/2})^{-1/2}X^{1/2}$ is the unique solution to the system $W_{NT}^{-1}XW_{NT}^{-1} = S$. We called the direction arising from (3.3.3) as NT direction and the scaling used as NT scaling. From Theorem 3.3.1, we know that the solutions of (3.3.3) are equivalent as long as P^TP does not change. For simplicity but without loss of generality, people always assume that $P = V^{1/2}$, e.g., $P = W_{NT}^{-1/2}$ in the convergence and complexity analysis. From the implementation side, however, this is not necessary. In [32], based on Theorem 3.3.1, Todd, Toh and Tütüncü proposed a computationally cheap way to calculate the NT scaling matrix. We first summarize their strategy, from which how to compute the positive and negative parts are very straightforward.

Let the *Cholesky factorization* of positive definite matrices X and S be

$$X = LL^T, \qquad S = RR^T,$$

and let $UDV^T = R^TL$ be the *Singular Value Decomposition* (SVD) of R^TL. Define $Q := L^{-1}X^{1/2}$. It is easy to see that Q is orthogonal, since

$$QQ^T = L^{-1}XL^{-T} = L^{-1}LL^TL^{-T} = I.$$

Then, one has

$$X^{1/2}SX^{1/2} = Q^T(L^TR)(R^TL)Q = (Q^TV)D^2(V^TQ).$$

Note Q^TV is also orthogonal, then we have

$$(X^{1/2}SX^{1/2})^{-1/2} = (Q^TV)D^{-1}(V^TQ).$$

Finally, W_{NT} can be computed easily by

$$W_{NT} = X^{1/2}(X^{1/2}SX^{1/2})^{-1/2}X^{1/2} = LVD^{-1}V^TL^T = \mathcal{G}\mathcal{G}^T, \tag{4.3.7}$$

where

$$\mathcal{G} := LVD^{-1/2}.$$

From (4.3.7), we remark that $\mathcal{G}^{-T}\mathcal{G}^{-1} = W^{-1}$. According to Theorem 3.3.1, the choice $P := \mathcal{G}^{-1}$ yields the same direction to system (3.3.3), hence to systems (4.3.1) and (4.3.2), as $P = W_{NT}^{-1/2}$. Therefore, we have

$$\begin{aligned}\mathcal{G}^TS\mathcal{G} &= D^{-1/2}V^TL^TSLVD^{-1/2}\\ &= D^{-1/2}V^TL^TRR^TLVD^{-1/2}\\ &= D^{-1/2}V^TVD^2V^TVD^{-1/2}\\ &= D.\end{aligned}$$

In a similar way, we also have $\mathcal{G}^{-1}X\mathcal{G}^{-T} = D$. In other words, $\mathcal{G}$ scales X and S to the same diagonal matrix D. In this case, the second term in the right hand side of system (3.3.3) becomes

$$H_{\mathcal{G}^{-1}}(XS) = (\mathcal{G}^{-1}XS\mathcal{G} + \mathcal{G}^TSX\mathcal{G}^{-T})/2 = D^2.$$

Then, according to Theorem 3.3.1, we can obtain the NT direction as the solution of

$$\begin{aligned}\mathrm{Tr}(A_i\Delta X) &= 0, &\text{(4.3.8a)}\\ \sum_{i=1}^{m}\Delta y_iA_i + \Delta S &= 0, &\text{(4.3.8b)}\\ H_{\mathcal{G}^{-1}}(\Delta XS + X\Delta S) &= \tau\mu_gI - D^2. &\text{(4.3.8c)}\end{aligned}$$

Because $\tau\mu_gI - D^2$ is a diagonal matrix, only $O(n)$ operations are needed to compute the positive and negative parts w.r.t. $\tau\mu_gI - D^2$, namely, our new algorithm is also viable from the computational point of view.

4.4 Complexity Analysis

In this section, we present the convergence and complexity proofs for Algorithm 5. Recall that our algorithm is based on the MZ family, we scale problems $(\mathcal{P})$ and $(\mathcal{D})$ as Monteiro and Todd proposed in [22] 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 polynomial convergence result.

4.4.1 Scaling Procedure

Scale the primal and dual variables in the following way,

$$\widetilde{X} := PXP, \qquad (\widetilde{y}, \widetilde{S}) := (y, P^{-1}SP^{-1}). \tag{4.4.1}$$

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

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

As mentioned, to investigate the new algorithm, we restrict the scaling matrix to $P \in \mathcal{P}(X, S)$ as defined by (3.4.1). It is easy to see that for $X, S \in \mathcal{S}^n_{++}$ one has

$$\mathcal{P}(X, S) := \{P \in \mathcal{S}^n_{++} \mid PXSP^{-1} \in \mathcal{S}^n\} = \{P \in \mathcal{S}^n_{++} : \widetilde{X}\widetilde{S} = \widetilde{S}\widetilde{X}\}, \tag{4.4.2}$$

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.0.23.

From now on, we use Λ to denote the diagonal matrix $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, where λ_i for $i = 1, \ldots, n$ are the eigenvalues of $\widetilde{X}\widetilde{S}$ with increasing order $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. 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:

$$(\widetilde{\mathcal{P}}) \quad \begin{array}{lll} \min & \mathrm{Tr}(\widetilde{C}\widetilde{X}) & \\ \text{s.t.} & \mathrm{Tr}(\widetilde{A}_i\widetilde{X}) & = \widetilde{b}_i,\ i=1,\ldots,m, \\ & \widetilde{X} & \succeq 0, \end{array}$$

and

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

The search direction $(\Delta X, \Delta y, \Delta S)$ based on system (4.3.1) and (4.3.2) corresponds to the scaled direction $(\widetilde{\Delta X}, \widetilde{\Delta y}, \widetilde{\Delta S})$ defined as

$$\widetilde{\Delta X}_- = P\Delta X_- P, \quad \widetilde{\Delta y}_- = \Delta y_-, \quad \widetilde{\Delta S}_- = P\Delta S_- P, \tag{4.4.3}$$

$$\widetilde{\Delta X}_+ = P\Delta X_+ P, \quad \widetilde{\Delta y}_+ = \Delta y_+, \quad \widetilde{\Delta S}_+ = P\Delta S_+ P. \tag{4.4.4}$$

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

$$\mathrm{Tr}(\widetilde{A}_i\widetilde{\Delta X}_-) = 0, \tag{4.4.5a}$$

$$\sum_{i=1}^{m}(\widetilde{\Delta y_i})_-\widetilde{A}_i + \widetilde{\Delta S}_- = 0, \tag{4.4.5b}$$

$$H_I(\widetilde{\Delta X}_-\widetilde{S} + \widetilde{X}\widetilde{\Delta S}_-) = [\tau\widetilde{\mu}_g I - \widetilde{X}\widetilde{S}]^-, \tag{4.4.5c}$$

and

$$\mathrm{Tr}(\widetilde{A}_i\widetilde{\Delta X}_+) = 0, \tag{4.4.6a}$$

$$\sum_{i=1}^{m}(\widetilde{\Delta y_i})_+\widetilde{A}_i + \widetilde{\Delta S}_+ = 0, \tag{4.4.6b}$$

$$H_I(\widetilde{\Delta X}_+\widetilde{S} + \widetilde{X}\widetilde{\Delta S}_+) = [\tau\widetilde{\mu}_g I - \widetilde{X}\widetilde{S}]^+, \tag{4.4.6c}$$

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 (4.4.5c) and (4.4.6c) become

$$\widetilde{E}\mathbf{vec}(\widetilde{\Delta X}_-) + \widetilde{F}\mathbf{vec}(\widetilde{\Delta S}_-) = \mathbf{vec}([\tau\mu_g I - \widetilde{X}\widetilde{S}]^-), \tag{4.4.7a}$$

$$\widetilde{E}\mathbf{vec}(\widetilde{\Delta X}_+) + \widetilde{F}\mathbf{vec}(\widetilde{\Delta S}_+) = \mathbf{vec}([\tau\mu_g I - \widetilde{X}\widetilde{S}]^+), \qquad (4.4.7b)$$

respectively, where

$$\widetilde{E} = \frac{1}{2}(\widetilde{S} \otimes I + I \otimes \widetilde{S}), \quad \widetilde{F} = \frac{1}{2}(\widetilde{X} \otimes I + I \otimes \widetilde{X}). \qquad (4.4.8)$$

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

$$\begin{aligned}(\widetilde{X}(\alpha), \widetilde{y}(\alpha), \widetilde{S}(\alpha)) &= (\widetilde{X}, \widetilde{y}, \widetilde{S}) + (\widetilde{\Delta X}(\alpha), \widetilde{\Delta y}(\alpha), \widetilde{\Delta S}(\alpha)) \qquad (4.4.9)\\ &= (\widetilde{X}, \widetilde{y}, \widetilde{S}) + \alpha_-(\widetilde{\Delta X}_-, \Delta y_-, \widetilde{\Delta S}_-) + \alpha_+(\widetilde{\Delta X}_+, \widetilde{\Delta y}_+, \widetilde{\Delta S}_+).\end{aligned}$$

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

Proposition 4.4.1. *If (X, y, S) and $(\widetilde{X}, \widetilde{y}, \widetilde{S})$ are related to each other as specified by (4.4.1), $(X(\alpha), y(\alpha), S(\alpha))$ and $(\widetilde{X}(\alpha), \widetilde{y}(\alpha), \widetilde{S}(\alpha))$ are defined by* (4.3.6) *and* (4.4.9)*, 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 $\widetilde{\mu}_g(\alpha) = \frac{\mathrm{Tr}(\widetilde{X}(\alpha)\widetilde{S}(\alpha))}{n}$.*

4.4.2 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 (4.4.5) and (4.4.6), 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 a feasible point, the search directions are orthogonal.

Lemma 4.4.2. *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 (4.4.5a), (4.4.5b), (4.4.6a), and (4.4.6b). ■

Lemma 4.4.3. *If $P \in \mathcal{P}(X, S)$, then we have*

$$\mathrm{Tr}(\widetilde{X}\widetilde{\Delta S}_-) + \mathrm{Tr}(\widetilde{\Delta X}_-\widetilde{S}) = \mathrm{Tr}([\tau_1\widetilde{\mu}_g I - \widetilde{X}\widetilde{S}]^-) \tag{4.4.10}$$

and

$$\mathrm{Tr}(\widetilde{X}\widetilde{\Delta S}_+) + \mathrm{Tr}(\widetilde{\Delta X}_+\widetilde{S}) = \mathrm{Tr}([\tau_1\widetilde{\mu}_g I - \widetilde{X}\widetilde{S}]^+). \tag{4.4.11}$$

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 (4.4.11) analogously. ■

Intuitively, we wish to reduce the duality gap as much as possible in every iteration. The next result, however, shows that Algorithm 5 holds a lower bound for duality gap reduction. In the later discussion, it will be seen that this bound derives from feasibility considerations.

Lemma 4.4.4. *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 4.4.2 and Lemma 4.4.3, we have

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

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

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

■

In fact, 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 as the next lemma shows.

Lemma 4.4.5. *Let* $(\widetilde{X}, \widetilde{y}, \widetilde{S}) \in \mathcal{F}^0$, *then*

$$\mathrm{Tr}([\tau_1\widetilde{\mu}_g I - \widetilde{X}\widetilde{S}]^-) \leq -(1-\tau_1)\mathrm{Tr}(\widetilde{X}\widetilde{S}). \tag{4.4.12}$$

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 4.4.6 and Corollary 4.4.7, imply that Algorithm 5 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 4.4.6. *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 4.4.3, 4.4.4 and 4.4.5, 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 4.4.7. *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 4.4.6, 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 is 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 4.4.8. *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*

$$\lambda_{\min}(\chi(\alpha)) \geq \tau_2\widetilde{\mu}_g + \alpha_+(\tau_1 - \tau_2)\widetilde{\mu}_g. \tag{4.4.13}$$

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 cases into account, we conclude that (4.4.13) 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 4.4.9. *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*

$$\left\|[\tau_1\widetilde{\mu}_g(\alpha)I - \chi(\alpha)]^+\right\|_F \leq (1 - \alpha_+)\hat{\eta}\beta\tau_1\widetilde{\mu}_g(\alpha). \tag{4.4.14}$$

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 4.2.2, we have

$$\begin{aligned}
\|\varphi(\alpha)\|_F &\leq \frac{\widetilde{\mu}_g(\alpha)}{\widetilde{\mu}_g}(1-\alpha_+)\left\|[\tau_1\widetilde{\mu}_g I - \Lambda]^+\right\|_F \\
&= \frac{\widetilde{\mu}_g(\alpha)}{\widetilde{\mu}_g}(1-\alpha_+)\left\|Q[\tau_1\widetilde{\mu}_g I - \Lambda]^+Q^T\right\|_F \qquad (4.4.15)\\
&= \frac{\widetilde{\mu}_g(\alpha)}{\widetilde{\mu}_g}(1-\alpha_+)\left\|[\tau_1\widetilde{\mu}_g I - \widetilde{X}\widetilde{S}]^+\right\|_F \\
&\leq (1-\alpha_+)\hat{\eta}\beta\tau_1\widetilde{\mu}_g(\alpha).
\end{aligned}$$

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

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

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 4.4.10. *Let $X, S \in \mathcal{S}^n_{++}$, $P \in \mathcal{P}(X,S)$, $\widetilde{X}$ and $\widetilde{S}$ are defined by (4.4.1), and $\widetilde{E}$ and $\widetilde{F}$ are defined by (4.4.8). Then,*

$$\left\|(\widetilde{F}\widetilde{E})^{-1/2}\mathbf{vec}([\tau_1\widetilde{\mu}_g I - \widetilde{X}\widetilde{S}]^-)\right\|^2 \leq \mathrm{Tr}(\widetilde{X}\widetilde{S}). \qquad (4.4.16)$$

Proof. Using Equation (4.4.8) and Proposition B.0.23, we find the spectral decompositions of $\widetilde{E}$ and $\widetilde{F}$ to be

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

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.0.22, 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

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

which leads to inequality (4.4.16). ■

Lemma 4.4.11. *Let $P \in \mathcal{P}(X,S)$, $\widetilde{X}$ and $\widetilde{S}$ be defined by (4.4.1), and $\widetilde{E}$ and $\widetilde{F}$ be defined by (4.4.8). 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 4.4.10 and 4.4.11, together with Lemma B.0.21, to conclude the following result.

Lemma 4.4.12. *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*

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

Proof. From (4.4.7), 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.0.21 to this equality, we obtain

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

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

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

where the last inequality can be derived from Lemma 4.4.10 and 4.4.11. ■

Using Lemma B.0.27, we can explore a bound for the second order term $\widetilde{\Delta X}(\alpha)\widetilde{\Delta S}(\alpha)$.

Lemma 4.4.13. *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*

$$\left\|H_I(\widetilde{\Delta X}(\alpha)\widetilde{\Delta S}(\alpha))\right\|_F \leq \left\|\mathbf{vec}(\widetilde{\Delta X}(\alpha))\right\|\left\|\mathbf{vec}(\widetilde{\Delta S}(\alpha))\right\| \leq \frac{2}{3}\sqrt{\mathrm{cond}(G)}\alpha_+^2\hat{\eta}^2\beta\tau_1\widetilde{\mu}_g. \tag{4.4.17}$$

Proof. Noticing the last inequality in Lemma B.0.26, 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 4.4.2 and B.0.27, it follows that

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

Substitute α_- with $\alpha_+\hat{\eta}\sqrt{\frac{\beta\tau_1}{n}}$ and apply Lemma 4.4.12, then we finally obtain

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

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

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

Proposition 4.4.14. *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 $G = \widetilde{E}^{-1}\widetilde{F}$. If $\alpha_- = \alpha_+\hat{\eta}\sqrt{\beta\tau_1/n}$ and $\alpha_+ \leq 1/(\sqrt{\mathrm{cond}(G)}\hat{\eta}^2)$, then*

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

Proof. By Corollary 4.4.7 we have $\widetilde{\mu}_g(\alpha) \leq \widetilde{\mu}_g$. Further, using Lemmas 4.4.8, 4.4.13 and the fact that $\lambda_{\min}(\cdot)$ is a homogeneous concave function on the space of symmetric matrices, one has

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

One can derive from Lemma 4.4.13 that

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

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

$$\lambda_{\min}(\widetilde{X}(\alpha)\widetilde{S}(\alpha)) \geq \lambda_{\min}(H_I(\widetilde{X}(\alpha)\widetilde{S}(\alpha))) \geq \tau_2\widetilde{\mu}_g(\alpha). \tag{4.4.18}$$

Since $\beta \leq 1/4$ and $\tau_1 \leq 4/9$, from Lemma 4.4.4, we have

$$\widetilde{\mu}_g(\alpha) \geq (1-\alpha_-)\widetilde{\mu}_g \geq (1-\hat{\eta}\sqrt{\beta\tau_1}/\sqrt{n})\widetilde{\mu}_g \geq (1-\sqrt{\beta\tau_1})\widetilde{\mu}_g \geq \frac{2}{3}\widetilde{\mu}_g. \tag{4.4.19}$$

From Proposition 4.2.1, we have

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

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

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

Using the fact that $\left\|[H_I(\widetilde{\Delta X}(\alpha)\widetilde{\Delta S}(\alpha))]^-\right\|_F \leq \left\|H_I(\widetilde{\Delta X}(\alpha)\widetilde{\Delta S}(\alpha))\right\|_F$ and Lemma 4.4.9, we can prove

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

Further, from Lemma 4.4.13 and inequality (4.4.19), one has

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

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

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

This, together with (4.4.18), implies that

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

Consequently, according to Proposition (4.4.1), one has

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

4.4.3 Polynomial Complexity

In this subsection we present our main complexity result. The next theorem gives an iteration-complexity bound for Algorithm 5 in terms of a parameter κ_∞ defined as

$$\kappa_\infty = \sup\left\{\mathrm{cond}((\widetilde{E}^k)^{-1}\widetilde{F}^k) : \ k = 0, 1, \ldots\right\}. \tag{4.4.20}$$

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

Theorem 4.4.15. *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 5 will terminate in $O(\eta\sqrt{\kappa_\infty n}\log(1/\epsilon))$ iterations with a solution $\mathrm{Tr}(XS) \leq \epsilon$.*

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 4.4.14, we have

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

Furthermore, from Lemma 4.4.7, we also conclude

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

from which the statement of the theorem follows. ■

From Theorem 4.4.15, it is easy to present various iteration complexities of Algorithm 5 in terms of some specific aforementioned scaling matrices P.

Corollary 4.4.16. *If the parameter η is a constant, then for Algorithm 5, when it is based on the NT direction, the iteration-complexity bound is $O(\sqrt{n}\log(1/\epsilon))$. When the H..K..M scaling is used, then Algorithm 5 terminates in at most $O(n\log(1/\epsilon))$ iterations.*

Corollary 4.4.17. *If the parameter η is in the order of $\sqrt{n}$, then for Algorithm 5, when it is based on the NT direction, the iteration-complexity bound is $O(n\log(1/\epsilon))$. When the H..K..M scaling is used, then Algorithm 5 terminates in at most $O(n^{3/2}\log(1/\epsilon))$ iterations.*

From Lemma B.0.20, Corollaries 4.4.16 and 4.4.17 are readily achieved.

As we see, when η is a constant and the NT scaling is used, Algorithm 5 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 4.2.6 that in that case our neighborhood $\mathcal{N}(\tau_1, \tau_2, \eta)$ is exactly the large neighborhood $\mathcal{N}(1 - \tau_2)$.

Chapter 5

Conclusions and Future Work

As stated by Renegar [28], "It is one of the ironies of the IPM literature that algorithms which are more efficient in practice often have somewhat worse complexity bounds." After Peng, Roos and Terlaky [27] established the self-regular paradigm, under which the complexity of large neighborhood IPMs for both LO and SDO can come arbitrarily close to the best known iteration bounds of IPMs, Ai and Zhang [2] proposed the so-called Ai-Zhang direction and proved $O(\sqrt{n}\log(1/\epsilon))$ iteration bound for LCP. In this thesis, we have extended Ai-Zhang's algorithm to SDO and successfully proved that a new large neighborhood IPM has $O(\sqrt{n}\log(1/\epsilon))$ iteration complexity, with marginal additional computational cost, when the Nesterov-Todd scaling is used, where n is the measure of the problem size and ϵ is the required precision. We would like to emphasize that, although the generalization of an IPM to SDP may be seen as "expected", or "routine exercise", this was certainly not the case here. The generalization was far from routine, it required several innovative ideas, novel inequalities and new nontrivial mathematical results that may be of general interest on their own.

Now, the most important theoretical work has been completed, as usual, we would like to implement this algorithm to see its performance in practice. Some issues related to the implementation deserve further discussion and exploration.

First, as we see from Chapter 4, our analysis assumes that the algorithm starts from a strictly feasible point and proceeds gradually to an optimal solution. In practice, however, it is not trivial to find such a strictly feasible point. Sometimes, a strictly feasible point even does not exist. To avoid this difficulty, one popular way is to embed the given SDO in a larger problem for which a

strictly feasible point is easily available. This embedding technique not only allows a strictly interior starting point, but also provides infeasibility identification. However, because we have to solve a larger problem, it will not only cost more physical memory, which sometimes causes memory overflow given the fact that SDO problems in practice might be of large dimension, but also need to solve a dense linear system at every iteration. Another alternative is to use an infeasible interior algorithm, namely, an algorithm starting with an infeasible point. This requires more analysis of a global and polynomial convergence. Generally, infeasible algorithms require a new neighborhood similar to $\mathcal{N}(\tau_1, \tau_2, \eta)$ which is able to measure the infeasibility together with distance to the central path. Hence, when the iterates are approaching the optimal set, then infeasibility and duality are reduced simultaneously until a feasible and optimal solution is obtained.

Second, efficient heuristics are needed for calculating step sizes. Ideally, we wish to solve subproblem (4.3.5) with the goal to find the best step sizes that decrease the duality gap as much as possible. Nevertheless, subproblem (4.3.5) itself is a two dimensional linear search, namely, a plane search, which might be computationally expensive. In Theorem 4.4.15, we propose a possible choice of the step length which leads to a polynomial iteration complexity. However, this choice might be too conservative to achieve efficiency in computational practice. Some other efficient heuristics might exist. For example, to make a big reduction of duality gap, we might fix $\alpha_+ = 1$ in every iteration, and do a linear search on α_-. Further research and benchmarking are needed to explore these possible step length heuristics .

Third, we are interested in knowing how to compute efficiently the positive and negative parts of the right-hand-side in the Newton equation when a scaling different from the NT scaling is used. In Subsection 4.3.2, we provide a way to compute positive and negative parts w.r.t $\tau\mu_g I - H_P(XS)$. Although the NT scaling is widely used in SDO solvers, and it turns out to be robust and accurate, it is still valuable to investigate a relatively cheap way to calculate the positive and negative parts when other scaling is employed, e.g., the H..K..M scaling.

Besides these topics about implementation, it might be very interesting to extend our algorithm to second order conic optimization (SOCO). Although we could consider SOCO as a special case of SDO and solve it by using an SDO approach, IPMs that solve SOCO problems directly usually have much better complexity than an IPM applied to the semidefinite formulation of a SOCO problem. Further, at each iteration much less work is needed for IPMs

directly applied to SOCO than those using the SDO approach. We suspect that when the notion of positive and negative parts of symmetric matrices is employed in the Jordan algebra context, the analysis will be more challenging than the one we see in this thesis.

To summarize, SDO and IPMs have matured during the past two decades. Nevertheless, our work in developing a new $O(\sqrt{n}\log(1/\epsilon))$ large neighborhood algorithm for SDO demonstrates that the development of IPMs has not been exhausted yet. There are still many open areas in IPMs including theoretical problems and computational methods. Among others, the solution of large scale SDO problems remains a challenging problem.

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. $(G \otimes K)^{-1} = G^{-1} \otimes K^{-1}$.
4. $(G \otimes K)(H \otimes L) = GH \otimes KL$.
5. If $\Lambda(G) = \mathrm{diag}(\lambda_i)$ and $\Lambda(K) = \mathrm{diag}(\mu_j)$, then $\Lambda(G \otimes K) = \mathrm{diag}(\lambda_i\mu_j)$. 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 $G \otimes K$.
6. $\mathbf{vec}(G)^T\mathbf{vec}(K) = \mathrm{Tr}(GK)$.

Appendix B

Some Properties of Square and Symmetric Matrices

Theorem B.0.18 (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 [13], page 79. ■

Lemma B.0.19. *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 [11], page 412. ■

Lemma B.0.20. *Let κ_∞ be defined by (4.4.20), 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 [23]. ■

The following technical lemma was first introduced and proved in Zhang [39].

Lemma B.0.21. *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*

$$\left\|(RQ^T)^{-1/2}Qu\right\|^2 + \left\|(RQ^T)^{-1/2}Rv\right\|^2 + 2u^Tv = \left\|(RQ^T)^{-1/2}r\right\|^2. \quad \text{(B.0.1)}$$

Proof. For the proof, we refer to Zhang's paper [20]. ■

To utilize Lemma B.0.21, 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 (4.4.8). In [32] and [20], 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 [20]. For those who are interested in the proof, they are advised to consult the paper by Monteiro [20].

Proposition B.0.22. *Let $X, S \in \mathcal{S}_{++}^n$, $\widetilde{X}$ and $\widetilde{S}$ be defined by (4.4.1), and $\widetilde{E}$ and $\widetilde{F}$ be defined by (4.4.8). 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 $\widetilde{F}\widetilde{E} \in \mathcal{S}_{++}^{n^2}$.*

For all the remaining results, we give all our credits to Monteiro and his paper [20]. We use his results throughout this paper from time to time.

Proposition B.0.23. *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.0.24. *For any $Q \in \mathcal{S}^n$, we have*

$$\lambda_{\max}(Q) = \max_{\|u\|=1} u^T Q u, \tag{B.0.2}$$

$$\lambda_{\min}(Q) = \min_{\|u\|=1} u^T Q u, \tag{B.0.3}$$

$$\|Q\| = \max_{i=1,\ldots,n} |\lambda_i(Q)|, \tag{B.0.4}$$

$$\|Q\|_F^2 = \sum_{i=1}^{n} |\lambda_i(Q)|^2. \tag{B.0.5}$$

Lemma B.0.25. *For any $Q \in \mathcal{R}^{n\times n}$ the following relations hold:*

$$\max_{i=1,\ldots,n} \mathrm{Re}[(\lambda_i(Q))] \leq \frac{1}{2}\lambda_{\max}(Q+Q^T), \tag{B.0.6}$$

$$\min_{i=1,\ldots,n} \mathrm{Re}[(\lambda_i(Q))] \geq \frac{1}{2}\lambda_{\min}(Q+Q^T), \tag{B.0.7}$$

$$\sum_{i=1}^{n} |\lambda_i(Q)|^2 \leq \|Q\|_F^2 = \left\|Q^T\right\|_F^2, \tag{B.0.8}$$

$$\lambda_{\max}(Q^TQ) = \left\|Q^TQ\right\| = \|Q\|^2 = \left\|Q^T\right\|^2, \tag{B.0.9}$$

$$\|Q\|_F \geq \left\|(Q+Q^T)/2\right\|_F. \tag{B.0.10}$$

Lemma B.0.26. *Let $W \in \mathcal{R}^{n\times n}$ be a nonsingular matrix. Then, for any $Q \in \mathcal{S}^n$, we have*

$$\lambda_{\max}(Q) \leq \frac{1}{2}\lambda_{\max}(WQW^{-1} + (WQW^{-1})^T), \tag{B.0.11}$$

$$\lambda_{\min}(Q) \geq \frac{1}{2}\lambda_{\min}(WQW^{-1} + (WQW^{-1})^T), \tag{B.0.12}$$

$$\|Q\| \leq \frac{1}{2}\left\|(WQW^{-1} + (WQW^{-1})^T\right\|, \tag{B.0.13}$$

$$\|Q\|_F \leq \frac{1}{2}\left\|(WQW^{-1} + (WQW^{-1})^T\right\|_F. \tag{B.0.14}$$

Lemma B.0.27. *For any $u, v \in \mathcal{R}^n$ and $G \in \mathcal{S}^n_{++}$, we have*

$$\|u\|\,\|v\| \leq \frac{\sqrt{\mathrm{cond}(G)}}{2}\left(\left\|G^{-1/2}u\right\|^2 + \left\|G^{1/2}v\right\|\right). \tag{B.0.15}$$

Bibliography

[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] R. Bellman and K. Fan. On systems of linear inequalities in hermitian matrix variables. In V.L. Klee, editor, *Convexity*, volume 7 of *Proceedings of Symposia in Pure Mathematics*, pages 1–11, Providence, RI, 1963. American Mathematical Society.

[7] D.S. Bernstein. *Matrix Mathematics.* Princeton University Press, Princeton, New Jersey, 2005.

[8] E. de Klerk. *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications.* Kluwer Academic Publishers, Dordrecht, 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] M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, 1995.

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

[12] 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.

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

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

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

[16] 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.

[17] 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.

[18] Y. Li and T. Terlaky. 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. Technical Report 2008/4, Advanced Optimization Lab., Department of Computing and Software, McMaster University, Hamilton, Ontario, Canada.

[19] Z.-Q. Luo, J. F. Sturm, and S. Zhang. Conic convex programming and self-dual embedding. *Optim. Methods. Softw.*, 14:169–218, 2000.

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

[21] 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.

[22] 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.

[23] 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.

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

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

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

[27] 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.

[28] J. Renegar. *A Mathematical View of Interior-Point Methods in Convex Optimization.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2001.

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

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

[31] M.J. Todd. Semidefinite optimization. *Acta Numerica*, 10:515–560, 2001.

[32] 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.

[33] K.C. Toh, M.J. Todd, and R.H. Tutuncu. Sdpt3 c a matlab software package for semidefinite programming. *Optimization Methods and Software*, 11:545–581, 1999.

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

[35] R.J. Vanderbei. *Linear Programming: Foundations and Extensions.* Kluwer Academic Publishers, Boston, 1996.

[36] 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.

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

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

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