INTERIOR-POINT METHODS FOR SEMIDEFINITE PROGRAMMING

by

Rongqin Sheng

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Mathematics in the
Graduate College of The
University of Iowa

August 1997

Thesis supervisor: Professor Florian A. Potra

ABSTRACT

This thesis aims to investigate fundamental issues of primal-dual interior-point methods for solving semidefinite programs (SDP), including superlinear convergence, worst case complexity and numerical implementation.

In this thesis, we will first extend the Mizuno-Todd-Ye predictor-corrector algorithm for linear programming (LP) to SDP. The resulting extension is an infeasible-interior-point algorithm which has the best iteration complexity obtained so far, and more interestingly has superlinear convergence under a sufficient condition. Based on this sufficient condition, superlinear convergence results are derived for algorithms using different search directions. Secondly, we propose a homogeneous formulation of SDP and show that the low iteration complexity of feasible interior-point algorithms can be achieved by corresponding infeasible-interior-point algorithms as long as the homogeneous model is used. Furthermore, the drawback of big M initialization of infeasible-interior-point algorithms is avoided. Finally, we demonstrate numerical implementations of homogeneous algorithms and show that homogeneous algorithms in general outperform their non-homogeneous versions.

Abstract approved: ____________________________
Thesis supervisor

____
Title and department

____
Date

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Rongqin Sheng

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Mathematics at the August 1997 graduation.

Thesis committee: ______________________
Thesis supervisor

____
Member

____
Member

____
Member

____
Member

INTERIOR-POINT METHODS FOR SEMIDEFINITE PROGRAMMING

by

Rongqin Sheng

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Mathematics in the
Graduate College of The
University of Iowa

August 1997

Thesis supervisor: Professor Florian A. Potra

To my wife, Beili
my son, Yuren
and my daughter, Ann

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisor, Professor Florian A. Potra, for introducing me to this topic of investigation, for his years of encouragement, his infinite patience, and his adept guidance.

My sincere thanks are also due to Professor Weimin Han for his constant support.

I am also indebted to Professor Kendall E. Atkinson, Professor Ken O. Kortanek and Professor Yinyu Ye who spend their valuable time serving on my committee.

Furthermore, the financial support of the Department of Mathematics of the University of Iowa, and the National Science Foundation is gratefully acknowledged.

I also owe many thanks to my parents and my other family members for their many years of supports in all possible means.

Finally, my deepest gratitude goes to my wife, Beili, whose love and support have motivated me to persevere with my study.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I
INTRODUCTION

1.1 Overview

In a recent survey paper [14] published in October 1996, Freund and Mizuno state that "Interior point methods in mathematical programming have been the largest and most dramatic area of research in optimization since the development of the simplex method... Interior point methods have permanently changed the landscape of mathematical programming theory, practice and computation...". While most of the research in this area was devoted to linear programming, in the opinion of the authors, "semidefinite programming is the most exciting development in mathematical programming in 1990s."

Semidefinite programming (SDP) consists in minimizing a linear functional of a matrix subject to linear equality and inequality constraints, the inequalities being understood in the sense of the cone of positive semidefinite matrices. It contains as particular cases linear programming (LP), (linearly constrained) quadratic programming (QP), quadratically constrained quadratic programming (QQP) and other optimization problems (see [50] and [65]). Semidefinite programming has numerous applications in such diverse areas as optimal control, combinatorial optimization, structural optimization, pattern recognition, trace factor analysis in statistics, matrix completions, etc. (see the excellent survey paper by Vandenberghe and Boyd [65]). However it was only after the advent of interior point methods that SDP problems could be efficiently solved.

Interior point methods have been known since the 1960s (cf. [11, 22, 12]), but the scientific community realized their potential for solving large LP problems only after the publication of the seminal paper of Karmarkar [28] in 1984. A decade later it became clear that LP solvers based on interior point methods are competitive with simplex based codes for medium size problems, and are outperforming the latter on large-scale problems. The pioneering work of Nesterov and Nemirovskii from the late 1980s showed that interior point methods can be generalized for solving in polynomial time a large class of convex programs that includes SDP (see [50]). The possibility of extending interior point methods to SDP was independently demonstrated by Alizadeh in his Ph.D. thesis [1] (see also [2]).

In the last couple of years an impressive number of interior point methods for SDP have been proposed [2, 4, 6, 7, 13, 20, 23, 33, 35, 38, 37, 39, 41, 42, 44, 43, 46, 49, 52, 51, 59, 62, 64, 70]. Some of them have been successfully implemented for solving important application problems. However the theory and practice of interior point methods for SDP does not have yet the level of maturity reached by interior point methods for LP and QP. The study of interior point methods for SDP is extremely important because while LP and QP can also be solved by other methods (e.g. the simplex method or Lemke's method), interior point methods appear to be the only efficient methods for solving general SDP problems known to date. In this thesis we discuss interior point methods for SDP with emphasis on the following areas: polynomial complexity, superlinear convergence, infeasibility detection, software development. Technical details are given in the following chapters.

1.2 Polynomial convergence

In his landmark paper from 1979 Khachiyan [29] has shown that the ellipsoid method of Yudin and Nemirovskii can solve linear programs with integer data in polynomial time. In spite of the fact that the worst case computational complexity of the ellipsoid method is polynomial while that of the simplex method is exponential, the simplex method outperforms the ellipsoid method on practical problems. This is explained by the fact that the average performance of the ellipsoid method is about the same as its worst case performance while the average performance of the simplex method is much better than its worst case performance. In fact, it has been proven that in certain probabilistic models the average complexity of the simplex method is polynomial (see [8]). Many of the interior point methods proposed after the publication of Karmarkar's paper [28] have both polynomial complexity and very good practical behaviour. Polynomial complexity reduces essentially to showing that the objective function approaches its optimum value at a global linear rate at least $1 - 1/O(n^\omega)$ where n is the dimension of the problem and ω is a positive constant (for the best complexity results proved so far $\omega = 1/2$). This means that the distance to optimum (or the duality gap) is reduced by a factor of $2^{-O(L)}$ in at most $O(n^\omega L)$ iterations. For linear programs with integer data of bit-length L, it is known that if the duality gap is smaller than $2^{-O(L)}$, then by a rounding procedure (consisting in at most $O(n^3)$ arithmetic operations) it is possible to recover the exact solution of the problem. Since the computational cost per iteration is at most $O(n^3)$ arithmetic operations, it follows that an exact solution of the problem can be found in polynomial time in the length L of a binary encoding of the data, and the dimension n of the problem.

When speaking about polynomiality of an SDP problem we mean that there is a positive constant ω such that the distance to optimum (or the duality gap) is reduced by a factor of $2^{-O(L)}$ in at most $O(n^\omega L)$ iterations. In this case we will say that the interior point method has $O(n^\omega L)$ iteration complexity. Nesterov and

Nemirovskii [50] have indicated a general way of constructing polynomial interior point methods for minimizing a linear function over a convex set, provided that a computable self-concordant barrier function for the convex set is known. Since the function $f(X) := \ln\det(X)$ is a self-concordant barrier for the set of all symmetric positive semidefinite matrices X, their result shows that there are polynomial interior point methods for solving SDP.

In what follows we will only consider primal-dual interior point methods that simultaneously solve the standard form SDP problem

$$(P) \quad \min\{C \bullet X : A_i \bullet X = b_i,\ i = 1, \ldots, m,\ X \succeq 0\}, \tag{1.1}$$

and its associated dual problem

$$(D) \quad \max\{b^T y : \sum_{i=1}^m y_i A_i + S = C,\ S \succeq 0\}, \tag{1.2}$$

where $C \in \mathcal{S}^n$, $A_i \in \mathcal{S}^n$, $i = 1, \ldots, m$, $b = (b_1, \ldots, b_m)^T \in \mathbb{R}^m$ are given data, and $X \in \mathcal{S}^n_+$, $(y, S) \in \mathbb{R}^m \times \mathcal{S}^n_+$ are the primal and dual variables, respectively. Here we use $\mathcal{S}^n$ to denote the set of all $n \times n$ symmetric matrices and $\mathcal{S}^n_+$ the set of all $n \times n$ symmetric positive semidefinite matrices. By $G \bullet H$ we denote the trace of $(G^T H)$.

Assuming that the matrices A_i $i = 1, \ldots, m,$ are linearly independent and that both the primal and the dual problems have strictly feasible solutions, it can be shown that for each scalar $\mu > 0$ the following system of equations

$$A_i \bullet X = b_i,\ i = 1, \ldots, m, \tag{1.3a}$$

$$\sum_{i=1}^m y_i A_i + S = C, \tag{1.3b}$$

$$X \bullet S = \mu I, \quad X \succeq 0, \quad S \succeq 0, \quad . \tag{1.3c}$$

has a unique solution $Z_\mu = (X_\mu, y_\mu, S_\mu) \in \mathcal{S}^n_{++} \times \mathbb{R}^m \times \mathcal{S}^n_{++}$. The set of all Z_μ with $\mu > 0$ form the so-called primal-dual central path. For some properties of the primal-dual central path for SDP see [35] and [18]. The primal-dual central path has played an crucial role in the design and analysis of primal-dual interior point methods for LP. We note that the best LP codes to date are based on primal-dual interior point methods (see [67]). The primal-dual central path for LP can be obtained as a particular case of (1.3) if all matrices involved are diagonal matrices. The search direction for all primal-dual interior point methods for LP can be interpreted as Newton directions for (1.3). While in the LP case the Newton direction for (1.3) is uniquely defined, this is no longer the case for SDP. Indeed by linearizing (1.3) and by replacing μ with $\xi\tau$ we obtain the following linear system in the search direction $(U, w, V) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$:

$$A_i \bullet U = (1-\xi) r_i, \quad i = 1, \ldots, m, \tag{1.4a}$$

$$\sum_{i=1}^m w_i A_i + V = (1-\xi) R_d, \tag{1.4b}$$

$$US + XV = \xi\tau I - XS \tag{1.4c}$$

where,

$$r_i = b_i - A_i \bullet X, \; i = 1, \ldots, m, \quad R_d = C - \sum_{i=1}^m y_i A_i - S. \tag{1.5}$$

In the LP case, the above system has a unique solution with matrices U, V diagonal. In the SDP case if a solution exists then clearly $V \in \mathcal{S}^n$, but U is not necessarily symmetric. There have been many attempts to replace equation (1.4c) by a similar equation so that the resulting linear system has a unique solution $(U, w, V) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$. This has given rise to a multitude of search directions, the most notable being the following three:

1) the AHO search direction proposed by Alizadeh, Haeberly and Overton [4],

2) the HRVW/KSH/M search direction independently proposed by Kojima, Shindoh and Hara [35] and Helmberg, Rendl, Vanderbei and Wolkowicz [20], and later rediscovered by Monteiro [41] via a formulation based on a scaling and symmetrization of the Newton equation, and

3) the NT direction introduced in [51].

Several authors [35, 70, 46, 62] have tried to unify these directions. In particular, the Monteiro and Zhang [41, 70, 46] family (MZ) of search directions contains all three search directions mentioned above. Each direction of the MZ family is defined by a linear system of the form (1.4) with equation (1.4c) replaced by

$$H_P(US + XV) = \xi\tau I - H_P(XS), \tag{1.6}$$

where

$$H_P(M) = \frac{1}{2}\left[PMP^{-1} + (PMP^{-1})^T\right], \quad \forall M \in \mathbb{R}^{n\times n},$$

(with a given a nonsingular matrix $P \in \mathbb{R}^{n\times n}$) is the similar symmetrization operator of Zhang [70]. The search directions 1), 2), 3) mentioned above are obtained by taking P equal to I, $S^{1/2}$, and $[S^{1/2}(S^{1/2}XS^{1/2})^{-1/2}S^{1/2}]^{1/2}$ respectively.

In the LP case, the search direction is uniquely defined and the iteration complexity of the interior point method depends only on the choice of some scalar parameters defining the method, and on the neighborhood of the central path that contains the iteration sequence produced by the method. Roughly speaking, most feasible interior point methods that use a "small" neighborhood of the central path have $O(\sqrt{n}L)$ iteration complexity, while those that use a "semi-large" or "large" neighborhood of the central path have $O(nL)$ iteration complexity. In the SDP case the situation is different, and the iteration complexity seems to be dependent on the choice of search direction as well. For SDP the three types of neighborhoods mentioned above are defined in terms of the eigenvalues $\lambda_i(XS), i = 1, \ldots, n$ as:

$$\mathcal{N}_F(\gamma) := \{(X, y, S) \in \mathcal{S}^n_{++} \times \mathbb{R}^m \times \mathcal{S}^n_{++} : \left(\sum_{i=1}^m (\lambda_i(XS) - \tau)^2\right)^{1/2} \le \gamma\tau, \tau > 0\}, \tag{1.7}$$

$$\mathcal{N}_\infty(\gamma) := \{(X, y, S) \in \mathcal{S}^n_{++} \times \mathbb{R}^m \times \mathcal{S}^n_{++} : \max_{i=1,\ldots,m} |\lambda_i(XS) - \tau| \le \gamma\tau, \tau > 0\}. \tag{1.8}$$

$$\mathcal{N}_{-\infty}(\gamma) := \{(X, y, S) \in \mathcal{S}^n_{++} \times \mathbb{R}^m \times \mathcal{S}^n_{++} : \max(0, \tau - \lambda_{\min}(XS)) \le \gamma\tau, \tau > 0\}, \tag{1.9}$$

For feasible interior point methods one usually takes $\tau = \mu = (X \bullet S)/n$ and one considers the intersection of the above neighborhoods with the set $\mathcal{F}^0$ of strictly feasible points.

The best complexity results obtained so far show that some feasible interior point methods based on the small neighborhood have $O(\sqrt{n}L)$ iteration complexity for all three search directions mentioned above. This result was generalized by Monteiro [42] for the whole MZ family, and by Monteiro and Tsuchiya [44] for the KSH family of directions (the KSH family contains the NT direction and the HRVW/KSH/M direction, but not the AHO direction).

Monteiro and Zhang [46] proved that algorithms acting in the large neighborhood have $O(nL)$ iteration complexity if based on the NT direction and $O(n^{3/2}L)$ if based on the HRVW/KSH/M search direction. They also gave iteration complexity bounds (depending on the condition number of matrices J_x and J_s defined by $P^TP = X^{-1/2}J_xX^{1/2} = S^{-1/2}J_sS^{1/2}$) for algorithms acting in the large neighborhood and based on the MZ* family of directions which is a subclass of the MZ family, containing the NT and the HRVW/KSH/M search direction, but not the AHO direction. So far, no complexity results are known for algorithms based on the large neighborhood and the AHO direction.

In a very recent paper Monteiro and Tsuchiya [43] introduced a new class (MT) of search directions (which contains the HRVW/KSH/M and the NT search directions) and proved that feasible interior point methods based on this class and acting in the semi-large neighborhood have $O(nL)$ iteration complexity. We note that the AHO direction does not belong the to the MT class, and nothing is known about the polynomiality of algorithms acting in the semi-large neighborhood based on it.

The polynomiality of infeasible interior point algorithms is considerably more difficult than that of their feasible counterparts. The first complexity result in this respect was obtained by Kojima, Shindoh, and Hara [35] who showed that an infeasible interior point potential reduction method for SDP has $O(n^{5/2}L)$ iteration complexity. Subsequently Zhang [70] analyzed an infeasible interior point method based on the HRVW/KSH/M search direction which has $O(n^2L)$ iteration complexity when acting in the semi-large neighborhood and $O(n^{5/2}L)$ iteration complexity in the large neighborhood. He also stated that the Mizuno-Todd-Ye predictor-corrector method for SDP based on this direction has $O(\sqrt{n}L)$ iteration complexity if started from strictly feasible points. The analysis of the Mizuno-Todd-Ye predictor-corrector method for infeasible starting points was done independently by Kojima, Shida and Shindoh [33], and Potra and Sheng [54]. The analysis in the latter paper is more refined (see the comment on page 3 of [33]) and shows that the iteration complexity depends on the quality of the starting point. If the problem has a solution, then the algorithm is globally convergent. If the starting point is feasible or close to being feasible, the algorithms finds an optimal solution in at most $O(\sqrt{n}L)$ iterations. If the starting point is large enough then the algorithm terminates in at most $O(nL)$ steps either by finding a strictly complementary solution or by determining that the primal-dual problem has no solution of norm less than a given number. This result has been generalized for Mizuno-Todd-Ye predictor-corrector methods based on the MZ* class of directions (a subclass of the MZ class cf. [43]) by Sheng, Potra and Ji [58] and on the T class of directions by Tseng [62] (the T class is introduced in [62]).

The only other papers on infeasible interior point algorithms for SDP of which we are aware are the paper of Lin and Saigal [37] which proves that two infeasible interior point methods acting in the small neighborhood have $O(\sqrt{n}L)$ iteration complexity, the paper of Freund [13] which gives the iteration complexity in terms of the condition number of some matrices depending on the starting point, and the papers of Potra and Sheng [53, 55]. In [53] a simple homogeneous primal-dual feasibility model is proposed for a semidefinite programming (SDP) problems and two infeasible-interior-point algorithms (the Mizuno-Todd-Ye and a short step algorithm based on the HRVW/KSH/M search direction) are applied to the homogeneous formulation. It is shown that when started from the standard (infeasible) starting point $(X^0, y^0, S^0) = (I, 0, I)$ these algorithms have $O(\sqrt{n}L)$ iteration complexity, the same as the lowest complexity attained by feasible algorithms.

1.3 Superlinear convergence

The study of superlinear convergence of interior–point methods for LP was initiated in the early 90s in an effort to explain the fact that interior point methods tend to perform significantly better in practice than indicated by the polynomial complexity bounds. This discrepancy is due to the limitation of the worst case analysis used in deriving polynomial complexity bounds and reflects the inherent conflict between the requirements of global convergence and fast local convergence. Superlinear convergence is especially important for SDP since no finite termination schemes exist for such problems. As predicted by theory and confirmed by numerical experiments the condition number of the linear systems defining the search directions increases as $1/\mu$, so that the respective systems become very ill conditioned as we approach the solution. Therefore an interior point method that is not superlinearly convergent is unlikely to obtain high accuracy in practice in spite of its theoretical "polynomial complexity". On the other hand a superlinearly convergent interior point method will achieve good accuracy (e.g. 10^{-10} or better) in substantially fewer iterations than indicated by its worse case global linear convergence rate that is related to polynomial complexity.

The local convergence analysis for interior point algorithms for SDP is much more challenging than those for LP as shown by a relatively smaller number of papers addressing this subject. The first two papers investigating superlinear convergence of interior point algorithms were written independently by Kojima, Shida and Shindoh [33] and by Potra and Sheng [54]. The algorithm investigated in these papers is an extension of Mizuno-Todd-Ye predictor-corrector algorithm for LP and uses the HRVW/KSH/M search direction. Kojima, Shida and Shindoh [33] established the superlinear convergence under the following three assumptions:

(A) SDP has a strictly complementary solution;

(B) SDP is nondegenerate in the sense that the Jacobian matrix of its KKT system is nonsingular;

(C) the iterates converge tangentially to the central path in the sense that the size of the neighborhood containing the iterates must approach zero, namely,

$$\lim_{k\to\infty} \|(X^k)^{1/2} S^k (X^k)^{1/2} - (X^k \bullet S^k/n) I\|_F / (X^k \bullet S^k/n) = 0.$$

Assumption (B) and (C) are quite restrictive and they are not required for the superlinear convergence of interior point methods for LP or QP. In [54] we have not used assumptions (B) and (C). Instead we proposed a sufficient condition for superlinear convergence that is implied by the above assumptions.

An interesting superlinearly convergent predictor-corrector algorithm based on the NT search direction was proposed by Luo, Sturm and Zhang [39]. The algorithm depends on a parameter $\epsilon > 0$. It produces points $(X^k, y^k, S^k) \in \mathcal{N}_F(\gamma_k)$, where $\gamma_k = 1/4$ if $\mu_k := X^k \bullet S^k/n \geq \epsilon/4$ and $\gamma_k = \mu_k/\epsilon$ if $\mu_k < \epsilon/4$. The algorithm starts from a feasible point $(X^0, y^0, S^0) \in \mathcal{N}_F(1/4)$ and for any given $\tilde{\epsilon} \geq \epsilon/4$ finds a feasible point (X^k, y^k, S^k) with $\mu_k \leq \tilde{\epsilon}$ in at most $O(\sqrt{n} log(\mu_0/\tilde{\epsilon})$ iterations. However this bound on the number of iterations is not proved to hold for $0 < \tilde{\epsilon} < \epsilon/4$, hence the algorithm is not polynomial in the usual sense. The algorithm is superlinearly convergent under assumption (A). Of course (C) is enforced by the algorithm since for sufficiently large k

$$\|(X^k)^1/2S^k(X^k)^{1/2} - (X^k \bullet S^k/n)I\|_F/(X^k \bullet S^k/n) \leq (X^k \bullet S^k)/(4n).$$

It has been proven that if one uses one predictor and r correctors per iteration, then μ_k converges to zero with Q-order $2/(1+2^{-2r})$.

In [56] we improved Our result in ([54]) and obtained superlinear convergence under assumption (A) and the following condition:

$$\text{(D)} \qquad \lim_{k\to\infty} X^k S^k/\sqrt{X^k \bullet S^k} = 0.$$

which is clearly weaker than (C). Of course both (C) and (D) can be enforced by the algorithm, but the practical efficiency of such an approach is questionable. However, from a theoretical point of view it is proved in [56] that the modified algorithm of [33] that uses several corrector steps in order to enforce (C) has polynomial complexity and is superlinearly convergent under assumption (A) only. It is well known that assumption (A) is necessary for superlinear convergence of standard interior point methods even in the QP case (see [45]).

Kojima, Shida and Shindoh [33] also gave an example suggesting that interior point algorithms for SDP based on the HRVW/KSH/M search direction are unlikely to be superlinearly convergent without imposing a condition like (C) or (D). In [34] the same authors showed that a predictor-corrector algorithm using the AHO direction is quadratically convergent under assumptions (A) and (B). They also proved that the algorithm is globally convergent but no polynomial complexity bounds have been found for this algorithm. It is shown that the following condition which is similar, but weaker than (C)) is automatically satisfied by the iteration sequence generated by the algorithm:

$$\lim_{k\to\infty} \|(X^kS^k + S^kX^k)/2 - (X^k \bullet S^k/n)I\|_F/(X^k \bullet S^k/n) = 0.$$

It appears that the use of the AHO direction in the corrector step has a strong effect on centering. We exploited this property in [55] where we showed that a direct extension of Mizuno-Todd-Ye algorithm, based on the HRVW/KSH/M direction in the predictor step and the AHO direction in the corrector step, has polynomial complexity and is superlinearly convergent with Q-order 1.5 under assumptions (A) and (B).

Very recently, Ji, Potra and Sheng [24] gave a comprehensive local convergence analysis of Monteiro's extension of the Mizuno-Todd-Ye predictor-corrector algorithm for LP to SDP which uses the Monteiro-Zhang family of search directions.

Local convergence analysis of interior point methods for SDP algorithms is much more complex than for LP. However, this topic is extremely important since fast local convergence is necessary for efficient practical algorithms, especially since for SDP there is neither a simplex-like algorithm nor a finite termination technique. We believe fast local convergence depends heavily on the search directions employed by the algorithms as well as on the type of neighborhood used for steplength computation.

1.4 Infeasibility detection

As mentioned in Section 1.2, the vast majority of results on interior point methods for SDP assume that the starting point is strictly feasible. The existence of a strictly feasible point implies that both the primal and the dual problem have optimal solutions and that the corresponding solution sets are bounded. While for some SDP problems strictly feasible starting points are readily available, there are many SDP problems for which such starting points are difficult to obtain. Moreover there are SDP problems for which the very existence of feasible and/or optimal solutions are open questions. A typical result, showing the importance of considering infeasible interior point methods, is proved in [54]. The standard infeasible starting point considered there is of the form $(X^0, y^0, S^0) = \rho(I, 0, I)$, where $\rho > 0$ is a constant. According to established terminology this is an ϵ_0-approximate solution, with

$$\epsilon_0 = \max\{X^0 \bullet S^0, \|R_d^0\|, |r_i^0|, i = 1, \ldots, m\}. \tag{1.10}$$

The stopping criterion depends on two positive parameters $\rho > 0$ and ϵ, with $0 < \epsilon < \epsilon_0$. At each iteration the algorithms checks whether an ϵ-approximate solution has been found, that is if,

$$\begin{aligned} (X^k, y^k, S^k) \in \mathcal{F}_\epsilon \quad &= \quad \{Z = (X, y, S) \in \mathcal{S}_+^n \times \mathrm{IR}^m \times \mathcal{S}_+^n : \qquad (1.11) \\ & X \bullet S \le \epsilon, \quad |r_i| \le \epsilon, \ i = 1, \ldots, m, \quad \|R_d\| \le \epsilon\}. \quad (1.12) \end{aligned}$$

or if the step-size has fallen bellow a given threshold,

$$\bar{\theta}_k \le 1/(73.5n). \tag{1.13}$$

It is proven that for any choice of $\rho > 0$, and for any $0 < \epsilon < \epsilon_0$ there is an index $k = O(n \ln(\epsilon_0/\epsilon))$ such that either (1.11) or (1.13) holds. In the latter case the problem has no optimal solution (X^*, y^*, S^*) with $\rho \ge \max\{\|X^*\|, \|S^*\|\}$.

The disadvantage of the above result is that both the starting point and the stopping criterion (1.13) depend on ρ. If we choose a small ρ then the algorithm may stop because (1.13) is satisfied. In this case the information that no optimal solution of norm less than ρ exists is of limited value. On the other hand, if a large value for ρ is chosen and if the problem has an optimal solution of small norm, then the algorithm may need many iterations to find an ϵ-approximate solution.

In order to avoid this inconvenience we proposed in [53] the following primal-dual feasibility model for SDP:

$$A_i \bullet X = \tau b_i, \ i = 1, \ldots, m, \tag{1.14a}$$

$$\sum_{i=1}^{m} y_i A_i + S = \tau C, \tag{1.14b}$$

$$b^T y - C \bullet X = \kappa, \tag{1.14c}$$

$$X \succeq 0, \;\; S \succeq 0, \;\; \tau \geq 0, \kappa \geq 0. \tag{1.14d}$$

The above model is an extension of the homogeneous and self–dual feasibility model for LP developed by Goldman and Tucker [19, 63] and later used by Xu, Hung and Ye [68] in an efficient interior point solver for LP.

It is easily seen that the SDP problem (1.3) has a solution if and only if the homogeneous system (1.14) has a solution $(X^*, y^*, S^*, \tau^*, \kappa^*) \in \mathcal{S}^n_+ \times \mathbb{R}^m \times \mathcal{S}^n_+ \times \mathbb{R}_+ \times \mathbb{R}_+$, such that $\tau^* > 0, \;\; \kappa^* = 0$.

In [53] we considered the Mizuno-Todd-Ye predictor corrector method applied to (1.14) with the standard infeasible starting point $(X^0, y^0, S^0, \tau_0, \kappa_0) = (I, 0, I, 1, 1)$ and we proved that for any choice of $\rho > 0$ there is an index $k = \widehat{K_\epsilon} = O(\sqrt{n}\ln(\rho\epsilon_0/\epsilon))$ such that either

(a) $(X^k/\tau_k, y^k/\tau_k, S^k/\tau_k) \in \mathcal{F}_\epsilon$,

or,

(b) $\tau_k < .75/(\rho + 1)$,

and in the latter case there is no solution $(X^*, y^*, S^*) \in \mathcal{F}^*$ with $\rho \geq Tr(X^* + S^*)$.

Let us note that although the starting point $(X^0, y^0, S^0, \tau_0, \kappa_0) = (I, 0, I, 1, 1)$ is infeasible, the above iteration complexity result is the same as the best complexity results obtained for feasible starting points. Also, while the stopping criterion (b) depends on ρ, the iteration sequence produced by the algorithm does not.

1.5 Software development

As mentioned in the introduction, SDP has many applications in such diverse fields as control system analysis and synthesis [15, 16], optimization of structures

[10], eigenvalue optimization [23, 36], determinant maximization [66], matrix completions [25], combinatorial optimization [2] (including maximum cut and satisfiability problems [17], graph coloring [26], graph equipartitioning [27], and the quadratic assignment problem [71]). For further references and other applications see the excellent review article of Vandenberghe and Boyd [65].

Although some research codes have been developed in the above mentioned application areas, and some experimental MATLAB codes are very promising (see [6] and [60]) the software development for SDP is still in its early stages. One of the first codes, SP, written in 1994 by Vandenberghe and Boyd is available on the Internet at *http://www-ISL.Stanford.EDU/people/boyd/SP.html*. It is based on a potential reduction method and requires feasible starting points. A friendly MATLAB interface for SP was developed by researchers from the Optimization and Control Group at ENSTA and INRIA (*http://www.ensta.fr/~gropco/lmi/lmitool.html*). The "Induced Norm Control Toolbox", a new MATLAB toolbox for analysis and synthesis of linear systems, built on top of the public software LMITOOL and SP has recently been developed by Beran (*http://www.iau.dtu.dk/Staff/ebb/INCT*). In May 1996, Wu, Vandenberghe and Boyd released MAXDET, a C code (with calls to LAPACK) for determinant maximization problems based on a long step path following interior point method (see *http://www-ISL.Stanford.EDU/people/boyd/MAXDET.html*). In June 1996, Wu and S. Boyd released the beta version of SDPSOL, a parser/solver for determinant maximization and SDP problems with matrix structure. This version exploits block-diagonal structure for increased efficiency. However it does not support other sparsity structures. In August 1996, Fujisawa and Kojima released version 2.0 of SDPA, a C++ implementation of some Mehrotra type infeasible interior point method for SDP, that can handle not only block diagonal matrices but also sparse matrices.

Moreover, SDPA provides some infeasibility detection capabilities. The user has the option to choose one of the three search directions mentioned above (HRVW/KSH/M, AHO, and NT). In December 1996, Toh, Todd, and Tütüncü [61] announced their MATLAB code SDPT3 which implements an infeasible Mehrotra type method for SDP using the three different search directions. Very recently, Alizadeh, Haeberly, Nayakkankuppam and Overton [3] released their MATLAB code SDPpack which implements a primal-dual Mehrotra predictor-corrector scheme based on the AHO direction.

Recently we have developed a *Mathematica* package for solving SDP [9]. The software as well as the report [9] are available on the Internet at

http://www.cs.uiowa.ed/~brixius/sdp.html

There are twelve algorithms in our *Mathematica* package version 1.0. Six of them are infeasible-interior-point algorithms for solving the standard primal dual SDP problem (1.3), while the other six algorithms use the homogeneous formulation (1.14) of SDP developed by Potra and Sheng [53]. We have implemented the HRVW/KSH/M, AHO, and NT for both formulations. For each direction and for each formulation we have implemented a Mizuno-Todd-Ye and a Mehrotra type predictor corrector method. The algorithms for (1.3) use either a (close to) feasible starting point or a large infeasible starting point (eventually of the form $(X^0, y^0, S^0) = \rho(I, 0, I)$, where $\rho > 0$ is a reasonably large constant). The algorithms based on the homogeneous system (1.14) instead use the fixed starting point $(X^0, y^0, S^0, \tau_0, \kappa_0) = (I, 0, I, 1, 1)$. The homogeneous versions can also detect infeasibility efficiently by monitoring a changing parameter. Numerical results show that homogeneous algorithms have better performance than their corresponding non-homogeneous versions.

We have just completed our MATLAB code SDPHA which is based on the structure of SDPT3 and implements the homogeneous algorithms for SDP. The code is available in the Internet at

http://www.math.uiowa.edu/~rsheng/SDPHA/sdpha.html

The technique of rank-2 update is employed in our MATLAB code so that the computation of homogeneous directions is only slightly more expensive than that of non-homogeneous ones related to infeasible-interior-point versions. However, the homogeneous algorithms generally take fewer iterations to capture a desired approximate solution. Numerical results show that the homogeneous algorithms outperform their non-homogeneous versions, with improvement of more than 20% in many cases, in terms of total CPU time.

The following notation and terminology are used throughout the thesis:

$\mathbb{R}^p$: the p-dimensional Euclidean space;

$\mathbb{R}^p_+$: the nonnegative orthant of $\mathbb{R}^p$;

$\mathbb{R}^p_{++}$: the positive orthant of $\mathbb{R}^p$;

$\mathbb{R}^{p\times q}$: the set of all $p \times q$ matrices with real entries;

$\mathcal{S}^p$: the set of all $p \times p$ symmetric matrices;

$\mathcal{S}^p_+$: the set of all $p \times p$ symmetric positive semidefinite matrices;

$\mathcal{S}^p_{++}$: the set of all $p \times p$ symmetric positive matrices;

$[M]_{ij}$: the (i,j)-th entry of a matrix M;

$\mathrm{Tr}(M)$: the trace of a $p \times p$ matrix, equals $\sum_{i=1}^p [M]_{ii}$;

$M \succeq 0$: M is positive semidefinite;

$M \succ 0$: M is positive definite;

$\lambda_i(M),\ i = 1, \ldots, n$: the eigenvalues of $M \in \mathcal{S}^n$;

$\lambda_{\max}(M),\ \lambda_{\min}(M)$: the largest, smallest, eigenvalue of $M \in \mathcal{S}^n$;

$G \bullet H \equiv \mathrm{Tr}(G^T H)$;

$\|\cdot\|$: Euclidean norm of a vector and the corresponding norm of a matrix, i.e., $\|y\| \equiv \sqrt{\sum_{i=1}^{p} y_i^2}$, $\quad \|M\| \equiv \max\{\|My\| : \|y\| = 1\}$;

$\|M\|_F \equiv \sqrt{M \bullet M}$, $M \in \mathbb{R}^{p \times q}$: Frobenius norm of a matrix;

$\|(G, H)\|_F \equiv \sqrt{G \bullet G + H \bullet H}$, $G, H \in \mathbb{R}^{p \times q}$;

$M^k = o(1)$: $\|M^k\| \to 0$ as $k \to \infty$;

$M^k = O(1)$: $\|M^k\|$ is bounded;

$M^k = o(\nu_k)$: $M^k/\nu_k = o(1)$;

$M^k = O(\nu_k)$: $M^k/\nu_k = O(1)$.

CHAPTER II
AN INFEASIBLE-INTERIOR-POINT METHOD FOR SDP

2.1 Description of the algorithm

In October 1995, Kojima, Shida and Shindoh [32] and Potra and Sheng [54] independently proposed an infeasible-interior-point algorithm for SDP which is the extension of the Mizuno-Todd-Ye predictor-corrector algorithm for LP [40] and uses the HRVW/KSH/M search direction. The algorithm is described below.

We denote the feasible set of the problem (1.3) by

$$\mathcal{F} = \{(X, y, S) \in \mathcal{S}^n_+ \times \mathrm{I\!R}^m \times \mathcal{S}^n_+ : (X, y, S) \text{ satisfies } (1.3a) \text{ and } (1.3b)\}$$

and its solution set by $\mathcal{F}^*$, i.e.,

$$\mathcal{F}^* = \{(X, y, S) \in \mathcal{F} : X \bullet S = 0\}.$$

The residues of (1.3a) and (1.3b) are denoted by (1.5).

For any given $\epsilon > 0$ we define the set of *ϵ-approximate solutions* of (1.3) as

$$\begin{aligned}\mathcal{F}_\epsilon = \{Z = (X, y, S) \quad \in \quad & \mathcal{S}^n_+ \times \mathrm{I\!R}^m \times \mathcal{S}^n_+ : \\ & X \bullet S \le \epsilon, \quad |r_i| \le \epsilon,\ i = 1, \ldots, m, \quad \|R_d\| \le \epsilon\}.\end{aligned}$$

In what follows we present an algorithm that finds a point in this set in a finite number of steps (provided the problem has a solution). The algorithm will perform in a neighborhood of the infeasible central path starting at

$$(X^0, y^0, S^0) \in \mathcal{S}^n_{++} \times \mathrm{I\!R}^m \times \mathcal{S}^n_{++} :$$

$$\begin{aligned}\mathcal{C}(X^0, y^0, S^0) \quad = \quad & \{Z = (X, y, S) \in \mathcal{S}^n_{++} \times \mathrm{I\!R}^m \times \mathcal{S}^n_{++} : \\ & XS = \tau I, \quad r_i = (\tau/\tau_0) r_i^0,\ i = 1, \ldots, m, \quad R_d = (\tau/\tau_0) R_d^0\},\end{aligned}$$

where R_d^0, r_i^0, $i = 1, \ldots, m$, are residues at (X^0, y^0, S^0) which satisfy $X^0 S^0 = \tau_0 I$.

In our algorithm the positive parameter τ will be driven to zero and therefore the residues will also be driven to zero at the same rate as τ. We use the following neighborhood of the above central path:

$$\begin{aligned}\mathcal{N}(\gamma,\tau) &= \{(X,y,S) \in \mathcal{S}^n_{++} \times \mathrm{I\!R}^m \times \mathcal{S}^n_{++} : \|X^{1/2}SX^{1/2} - \tau I\|_F \leq \gamma\tau\} \\ &= \{(X,y,S) \in \mathcal{S}^n_{++} \times \mathrm{I\!R}^m \times \mathcal{S}^n_{++} : \left(\sum_{i=1}^m (\lambda_i(XS) - \tau)^2\right)^{1/2} \leq \gamma\tau\},\end{aligned}$$

where γ is a constant such that $0 < \gamma < 1$. Throughout the chapter we also use the notation:

$$\mu = (X \bullet S)/n. \tag{2.1}$$

The algorithm depends on two positive parameters α, β satisfying the inequalities

$$\frac{\beta^2}{2(1-\beta)^2} \leq \alpha < \beta < \frac{\beta}{1-\beta} < 1. \tag{2.2}$$

For example, $\alpha = 0.25,\ \beta = 0.41$ verify (2.2).

At a typical step of our algorithm we are given $(X,y,S) \in \mathcal{N}(\alpha,\tau)$ and obtain a predictor direction $(U,w,V) \in \mathcal{S}^n \times \mathrm{I\!R}^m \times \mathcal{S}^n$ by solving the linear system

$$X^{-1/2}(XV + US)X^{1/2} + X^{1/2}(VX + SU)X^{-1/2} = -2X^{1/2}SX^{1/2}, \tag{2.3a}$$

$$A_i \bullet U = r_i, \quad i = 1,\ldots,m, \tag{2.3b}$$

$$\sum_{i=1}^m w_i A_i + V = R_d. \tag{2.3c}$$

We will see later on that the above linear system has a unique solution, which we call the affine scaling direction. If we take a steplength θ along this direction we obtain the points

$$X(\theta) = X + \theta U, \quad y(\theta) = y + \theta w, \quad S(\theta) = S + \theta V.$$

Theoretically we would like to compute the step length

$$\check{\theta} = \max\left\{\tilde{\theta} \in [0,1] : \left(\sum_{i=1}^n (\lambda_i(X(\theta)S(\theta)) - (1-\theta)\tau)^2\right)^{1/2} \leq \beta(1-\theta)\tau, \forall\theta \in [0,\tilde{\theta}]\right\}. \tag{2.4}$$

However this involves computing the root of a complicated nonlinear equation. In

Lemma 2.1.3 we will show that

$$\check{\theta} \geq \widehat{\theta} \tag{2.5}$$

where

$$\widehat{\theta} \equiv \frac{2}{\sqrt{1 + 4\delta/(\beta - \alpha)} + 1}, \tag{2.6}$$

and

$$\delta \equiv \frac{1}{\tau} \| X^{-1/2} U V X^{1/2} \|_F. \tag{2.7}$$

In what follows we assume that a steplength $\bar{\theta}$ satisfying

$$\check{\theta} \geq \bar{\theta} \geq \widehat{\theta} \tag{2.8}$$

is computed, and we consider the predicted points

$$\bar{X} = X + \bar{\theta} U, \quad \bar{y} = y + \bar{\theta} w, \quad \bar{S} = S + \bar{\theta} V. \tag{2.9}$$

In case of $\bar{\theta} = 1$ (which is very unlikely), it is easily seen that a solution $(\bar{X}, \bar{y}, \bar{S}) \in \mathcal{F}^*$ is at hand and the algorithm terminates. Now suppose that $\bar{\theta} < 1$. Then $\bar{X}$ and $\bar{S}$ are symmetric positive definite matrices since $(X(\theta), y(\theta), S(\theta)) \in \mathcal{N}(\beta, (1 - \theta)\tau), \ \ \forall \theta \in [0, \ \bar{\theta}]$. Therefore we can define the corrector direction $(\bar{U}, \bar{w}, \bar{V})$ as the solution of the following linear system

$$\bar{X}^{-1/2}(\bar{X}\bar{V} + \bar{U}\bar{S})\bar{X}^{1/2} + \bar{X}^{1/2}(\bar{V}\bar{X} + \bar{S}\bar{U})\bar{X}^{-1/2} \tag{2.10a}$$

$$= 2[(1 - \bar{\theta})\tau I - \bar{X}^{1/2}\bar{S}\bar{X}^{1/2}],$$

$$A_i \bullet \bar{U} = 0, \ \ i = 1, \ldots, m, \tag{2.10b}$$

$$\sum_{i=1}^{m} \bar{w}_i A_i + \bar{V} = 0. \tag{2.10c}$$

We will prove later on that the above linear system has a unique solution. By taking a unit steplength along the corrector direction we obtain a new point

$$X^+ = \bar{X} + \bar{U}, \quad y^+ = \bar{y} + \bar{w}, \quad S^+ = \bar{S} + \bar{V}. \tag{2.11}$$

Clearly

$$r_i^+ = (1 - \bar{\theta}) r_i, \ i = 1, \ldots, m, \quad R_d^+ = (1 - \bar{\theta}) R_d. \tag{2.12}$$

Correspondingly, we define

$$\tau_+ = (1 - \bar{\theta})\tau. \tag{2.13}$$

Let us note that in setting up the linear systems (2.3) and (2.10) it is not necessary to compute square roots of matrices. Indeed, as pointed out by Monteiro [41], it is easily seen that any equation of the form

$$X^{-1/2}(XV + US)X^{1/2} + X^{1/2}(VX + SU)X^{-1/2} = 2(\sigma I - X^{1/2}SX^{1/2})$$

can be written equivalently under the form

$$2XVX + USX + SXU = 2(\sigma X - XSX).$$

Summarizing, we can formally define our algorithm as follows:

Algorithm 2.1.1 *Choose* $(X^0, y^0, S^0) \in \mathcal{N}(\alpha, \tau_0)$ *with* $\tau_0 = \mu_0 = (X^0 \bullet S^0)/n$ *and set* $\psi_0 = 1$*. For* $k = 0, 1, \cdots$*, do A1 through A5:*

A1 Set $X = X^k$*,* $y = y^k$*,* $S = S^k$*,* $\psi = \psi_k$*,* $\tau = \tau_k$ *and define* $R_d = C - \sum_{i=1}^m y_i A_i - S$, $\quad r_i = b_i - A_i \bullet X, \quad i = 1, \ldots, m.$

A2 If $\max\{X \bullet S, \|R_d\|, \ |r_i|, i = 1, \ldots, m.\} \leq \epsilon$*, then report* $(X, y, S) \in \mathcal{F}_\epsilon$ *and terminate.*

A3 Find the solution U, w, V *of the linear system (2.3), define* $\bar{X}, \bar{y}, \bar{S}$ *as in (2.9), and set* $\psi_+ = (1 - \bar{\theta})\psi$*, for a* $\bar{\theta}$ *satisfying (2.8). If* $\bar{\theta} = 1$*, then report* $(\bar{X}, \bar{y}, \bar{S}) \in \mathcal{F}^*$ *and terminate.*

A4 Find the solution $\bar{U}, \bar{w}, \bar{V}$ *of the linear system (2.10) and define* X^+*,* y^+*,* S^+ *and* τ_+ *as in (2.11) and (2.13).*

A5 Set $X^{k+1} = X^+, \quad S^{k+1} = S^+, \quad \tau_{k+1} = \tau_+, \quad \bar{\theta}_k = \bar{\theta}, \quad \psi_{k+1} = \psi_+, \quad R_d^k = R_d, \quad r_i^k = r_i, \quad i = 1, \ldots, m.$

To analyze algorithm we need the following technical results.

Lemma 2.1.1 *Suppose that* $M \in \mathbb{R}^{p\times p}$ *is a nonsingular matrix and* $E \in \mathbb{R}^{p\times p}$ *has only real eigenvalues. Then,*

$$\lambda_{\max}(E) \le \frac{1}{2}\lambda_{\max}(MEM^{-1} + (MEM^{-1})^T), \tag{2.14}$$

$$\lambda_{\min}(E) \ge \frac{1}{2}\lambda_{\min}(MEM^{-1} + (MEM^{-1})^T), \tag{2.15}$$

$$\|G\|_F \le \frac{1}{2}\|MGM^{-1} + (MGM^{-1})^T\|_F, \quad \forall G \in \mathcal{S}^p. \tag{2.16}$$

Proof. See Lemma 3.2 and Lemma 3.3 Monteiro [41]. ∎

Lemma 2.1.2 *(Monteiro [41], Lemma 4.4) Let* $(X, y, S) \in \mathcal{N}(\gamma, \tau)$ *for some* $\gamma \in [0, 1)$ *and* $\tau > 0$. *Suppose that* $(D_x, \Delta y, D_s) \in \mathbb{R}^{n\times n} \times \mathbb{R}^m \times \mathbb{R}^{n\times n}$ *is a solution of the linear system:*

$$X^{-1/2}(XD_s + D_xS)X^{1/2} + X^{1/2}(D_sX + SD_x)X^{-1/2} = 2H, \tag{2.17a}$$

$$A_i \bullet D_x = 0, \quad i = 1, \ldots, m, \tag{2.17b}$$

$$\sum_{i=1}^m \Delta y_i A_i + D_s = 0, \tag{2.17c}$$

for some $H \in \mathbb{R}^{n\times n}$, *and let* $\delta_x = \tau\|X^{-1/2}D_xX^{-1/2}\|_F$ *and* $\delta_s = \|X^{1/2}D_sX^{1/2}\|_F$. *Then,*

$$\delta_x\delta_s \le \frac{1}{2}\left(\delta_x^2 + \delta_s^2\right) \le \frac{\|H\|_F^2}{2(1-\gamma)^2}.$$

The following corollary is essentially Corollary 2.3 of Monteiro [41].

Corollary 2.1.1 *Let* $(X, y, S) \in \mathcal{N}(\gamma, \tau)$ *for some* $\gamma \in [0, 1)$ *and* $\tau > 0$. *Then, for every* $H \in \mathcal{S}^n$, *the system of linear equations (2.17) has a unique solution* $(D_x, \Delta y, D_s) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$. *Therefore, both linear systems (2.3) and (2.10) have unique solutions.*

Lemma 2.1.3 *If* $(X, y, S) \in \mathcal{N}(\alpha, \tau)$, *then* $\check{\theta} \in [0, 1]$ *defined by (2.4) satisfies (2.5) where* $\widehat{\theta}$ *is given by (2.6) and (2.7).*

Proof. By definition, for any $\theta \in \mathbb{R}$, we have

$$X(\theta)S(\theta) - (1-\theta)\tau I = (X + \theta U)(S + \theta V) - (1-\theta)\tau I$$

$$= \quad (1-\theta)(XS - \tau I) + \theta XS + \theta(XV + US) + \theta^2 UV. \tag{2.18}$$

If we set

$$P(\theta) \equiv X^{-1/2}[X(\theta)S(\theta) - (1-\theta)\tau I]X^{1/2},$$

then, in view of (2.18) and (2.3a), we obtain

$$\begin{aligned}
& P(\theta) + P(\theta)^T \\
= \;& 2(1-\theta)(X^{1/2}SX^{1/2} - \tau I) \\
& +\theta[X^{-1/2}(XV+US)X^{1/2} + X^{1/2}(VX+SU)X^{-1/2} + 2X^{1/2}SX^{1/2}] \\
& +\theta^2(X^{-1/2}UVX^{1/2} + X^{1/2}VUX^{-1/2}) \\
= \;& 2(1-\theta)(X^{1/2}SX^{1/2} - \tau I) + \theta^2(X^{-1/2}UVX^{1/2} + X^{1/2}VUX^{-1/2}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \frac{1}{2}\|P(\theta) + P(\theta)^T\|_F \\
\leq \;& (1-\theta)\|X^{1/2}SX^{1/2} - \tau I\|_F + \theta^2\|X^{-1/2}UVX^{1/2}\|_F \\
\leq \;& \alpha\tau(1-\theta) + \theta^2\delta\tau. \qquad (2.19)
\end{aligned}$$

Hence, if $\widehat{\theta} < 1$ then we must have $X(\theta) \succ 0, \;\; S(\theta) \succ 0$ for all $\theta \in [0, \widehat{\theta}]$. Otherwise, there must exist $0 \leq \theta' \leq \widehat{\theta} < 1$ such that $X(\theta')S(\theta')$ is singular, which means

$$\lambda_{\min}(X(\theta')S(\theta') - (1-\theta')\tau) \leq -(1-\theta')\tau. \qquad (2.20)$$

However, using (2.15) with $M = X^{-1/2}$ and $E = X(\theta')S(\theta') - (1-\theta')\tau$, we have

$$\begin{aligned}
& \lambda_{\min}(X(\theta')S(\theta') - (1-\theta')\tau) \\
\geq \;& \frac{1}{2}\lambda_{\min}(P(\theta') + P(\theta')^T) \\
\geq \;& -\frac{1}{2}\|P(\theta') + P(\theta')^T\|_F \\
\geq \;& -[\alpha\tau(1-\theta') + (\theta')^2\delta\tau] \;\; \text{(from (2.19))} \\
\geq \;& -\beta(1-\theta')\tau,
\end{aligned}$$

which contradicts (2.20). Since $X(\theta) \succ 0$, its square root $X(\theta)^{1/2}$ exists and is uniquely defined. Applying (2.16) of Lemma 2.1.1 with

$$G = X(\theta)^{1/2}S(\theta)X(\theta)^{1/2} - (1-\theta)\tau I, \;\; M = X^{-1/2}X(\theta)^{1/2},$$

and noting that $P(\theta) = MGM^{-1}$, we obtain

$$\begin{aligned} & \left(\sum_{i=1}^{n}(\lambda_i(X(\theta)S(\theta)) - (1-\theta)\tau)^2\right)^{1/2} \\ = \; & \|X(\theta)^{1/2}S(\theta)X(\theta)^{1/2} - (1-\theta)\tau I\|_F \\ \leq \; & \frac{1}{2}\|P(\theta) + P(\theta)^T\|_F \\ \leq \; & \alpha\tau(1-\theta) + \theta^2\delta\tau \;\; \text{(from (2.19))} \\ \leq \; & \beta(1-\theta)\tau. \end{aligned}$$

Therefore, $(X(\theta), y(\theta), S(\theta)) \in \mathcal{N}(\beta, (1-\theta)\tau)$, for all $\theta \in [0, \; \widehat{\theta}]$. Thus, if $\widehat{\theta} < 1$, then (2.5) is satisfied. Similarly, if $\widehat{\theta} = 1$, then we must have $(X(\theta), y(\theta), S(\theta)) \in \mathcal{N}(\beta, (1-\theta)\tau)$, for all $\theta \in [0,1)$, which implies $X(1) \succeq 0, S(1) \succeq 0$ and $X(1)S(1) = 0$, and therefore $\check{\theta} = 1 = \widehat{\theta}$. ∎

Before stating our main result let us note that the standard choice of starting points

$$X^0 = \rho_p I, \; y^0 = 0, \; S^0 = \rho_d I$$

is perfectly centered and satisfies $(X^0, y^0, S^0) \in \mathcal{N}(\alpha, \tau_0)$, as required in the algorithm.

We will see that if the problem has a solution, then for any $\epsilon > 0$ Algorithm 2.1.1 terminates in a finite number (say K_ϵ) of iterations. If $\epsilon = 0$ then the algorithm is likely to generate an infinite sequence. However it may happen that at a certain iteration (let us say at iteration K_0) we have $\bar{\theta} = 1$, which implies that an exact solution is obtained, and therefore the algorithm terminates at iteration K_0. If this (unlikely) phenomenon does not happen we set $K_0 = \infty$.

Theorem 2.1.1 *For any integer $0 \leq k < K_0$, Algorithm 2.1.1 defines a triple*

$$(X^k, y^k, S^k) \in \mathcal{N}(\alpha, \tau_k) \tag{2.21}$$

and the corresponding residuals satisfy

$$R_d^k = \psi_k R_d^0, \quad r_i^k = \psi_k r_i^0, \quad i = 1, \ldots, m, \tag{2.22}$$

$$\tau_k = \psi_k \tau_0, \tag{2.23}$$

$$(1-\alpha)\tau_k \le \mu_k = (X^k \bullet S^k)/n \le (1+\alpha)\tau_k \tag{2.24}$$

where

$$\psi_0 = 1, \quad \psi_k = \prod_{j=0}^{k-1}(1-\bar{\theta}_j),$$

and $\bar{\theta}_j$ is defined by (2.8).

Proof. The proof is by induction. For $k = 0$, (2.21)–(2.24) are clearly satisfied. Suppose they are satisfied for some $k \ge 0$. As in Algorithm 2.1.1 we will omit the index k. Therefore we can write

$$(X, y, S) \in \mathcal{N}(\alpha, \tau),$$

$$R_d = \psi R_d^0, \quad r_i = \psi r_i^0, i = 1, \ldots, m,$$

$$\tau = \psi\tau_0, \quad (1-\alpha)\tau \le \mu \le (1+\alpha)\tau.$$

The fact that (2.22) and (2.23) hold for $k+1$ follows immediately from (2.12) and (2.13). From (2.11) and (2.13) we have

$$\begin{aligned} X^+S^+ - \tau_+ I &= (\bar{X}+\bar{U})(\bar{S}+\bar{V}) - \tau_+ I \\ &= \bar{X}\bar{S} - (1-\bar{\theta})\tau I + \bar{X}\bar{U} + \bar{U}\bar{S} + \bar{U}\bar{V}. \end{aligned} \tag{2.25}$$

Define

$$P = \bar{X}^{-1/2}(X^+S^+ - \tau_+ I)\bar{X}^{1/2}.$$

Then, recalling (2.25) and (2.10a), we obtain

$$\begin{aligned} P + P^T &= 2[\bar{X}^{1/2}\bar{S}\bar{X}^{1/2} - (1-\bar{\theta})\tau I] \\ &\quad + [X^{-1/2}(\bar{X}\bar{U} + \bar{U}\bar{S})X^{1/2} + X^{1/2}(\bar{U}\bar{X} + \bar{S}\bar{U})X^{-1/2}] \\ &\quad + [X^{-1/2}\bar{U}\bar{V}X^{1/2} + X^{1/2}\bar{V}\bar{U}X^{-1/2}] \\ &= X^{-1/2}\bar{U}\bar{V}X^{1/2} + X^{1/2}\bar{V}\bar{U}X^{-1/2}. \end{aligned} \tag{2.26}$$

Since $k < K_0$, we see that $\bar{\theta} < 1$ and that $\bar{X} \succ 0$, $\bar{S} \succ 0$. Hence by applying

Lemma 2.1.2 we deduce

$$\begin{aligned}
& \|\bar{X}^{-1/2}\bar{U}\bar{X}^{-1/2}\|_F\|\bar{X}^{1/2}\bar{V}\bar{X}^{1/2}\|_F \\
\leq\ & \frac{\|\bar{X}^{1/2}\bar{S}\bar{X}^{1/2}-(1-\bar{\theta})\tau I\|_F^2}{2(1-\beta)^2(1-\bar{\theta})\tau} \\
\leq\ & \frac{\beta^2}{2(1-\beta)^2}(1-\bar{\theta})\tau \leq \alpha(1-\bar{\theta})\tau.
\end{aligned} \tag{2.27}$$

Using Lemma 2.1.2 again, we have

$$\begin{aligned}
& \|\bar{X}^{-1/2}\bar{U}\bar{X}^{-1/2}\|_F \\
\leq\ & \frac{\|\bar{X}^{1/2}\bar{S}\bar{X}^{1/2}-(1-\bar{\theta})\tau I\|_F}{(1-\beta)(1-\bar{\theta})\tau} \leq \frac{\beta}{1-\beta} < 1,
\end{aligned}$$

which implies that $I + \bar{X}^{-1/2}\bar{U}\bar{X}^{-1/2} \succ 0$, and therefore, $X^+ = \bar{X} + \bar{U} \succ 0$. Thus $(X^+)^{1/2}$ exists. Using (2.25), (2.26), applying Lemma 2.1.1 with

$$G = (X^+)^{1/2}S^+(X^+)^{1/2} - \tau_+ I, \quad M = \bar{X}^{-1/2}(X^+)^{1/2},$$

and noting that $P = MGM^{-1}$, we have

$$\begin{aligned}
& \|(X^+)^{1/2}S^+(X^+)^{1/2} - \tau_+ I\|_F \leq \frac{1}{2}\|P + P^T\|_F \\
=\ & \frac{1}{2}\|\bar{X}^{-1/2}\bar{U}\bar{V}\bar{X}^{1/2} + \bar{X}^{1/2}\bar{V}\bar{U}\bar{X}^{-1/2}\|_F \quad \text{(from (2.26))} \\
\leq\ & \|\bar{X}^{-1/2}\bar{U}\bar{V}\bar{X}^{1/2}\|_F \leq \|\bar{X}^{-1/2}\bar{U}\bar{X}^{-1/2}\|_F\|\bar{X}^{1/2}\bar{V}\bar{X}^{1/2}\|_F \\
\leq\ & \alpha(1-\bar{\theta})\tau = \alpha\tau_+. \quad \text{(from (2.27))}
\end{aligned} \tag{2.28}$$

The above inequality implies that

$$\lambda_{\min}((X^+)^{1/2}S^+(X^+)^{1/2}) \geq (1-\alpha)\tau_+ > 0.$$

Hence $(X^+)^{1/2}S^+(X^+)^{1/2} \succ 0$, which gives $S^+ \succ 0$. In view of (2.28), this shows that (2.21) holds for $k+1$. Finally, (2.24) is an immediate consequence of (2.21). ∎

2.2 Global convergence and polynomiality

In this section we assume that $\mathcal{F}^*$ is nonempty. Under this assumption we will prove that Algorithm 2.1.1, with $\epsilon = 0$, is globally convergent in the sense that

$$\lim_{k\to\infty}\mu_k = 0, \quad \lim_{k\to\infty} R_d^k = 0, \quad \lim_{k\to\infty} r_i^k = 0, \ i = 1, \ldots, m.$$

In the sequel, we will frequently use the following inequality: for any $M_1, M_2 \in \mathbb{R}^{n\times n}$,

$$\|M_1 M_2\|_F \le \min\{\|M_1\|\|M_2\|_F, \ \ \|M_1\|_F\|M_2\|\}, \tag{2.29}$$

(see exercise 20 in section 5.6 of [21]).

Lemma 2.2.1 *For any* $(\widetilde{X}, \tilde{y}, \tilde{S}) \in \mathcal{F}$*, we have*

$$\begin{aligned} \psi(S \bullet X^0 + S^0 \bullet X) &= S \bullet X + \psi^2 S^0 \bullet X^0 + (1-\psi)^2 \tilde{S} \bullet \widetilde{X} \\ &\quad + \psi(1-\psi)(S^0 \bullet \widetilde{X} + \tilde{S} \bullet X^0) - (1-\psi)(S \bullet \widetilde{X} + \tilde{S} \bullet X). \end{aligned}$$

Proof. Let

$$\begin{aligned} X' &= X - \psi X^0 - (1-\psi)\widetilde{X}, \\ y' &= y - \psi y^0 - (1-\psi)\tilde{y}, \\ S' &= S - \psi S^0 - (1-\psi)\tilde{S}. \end{aligned}$$

From (1.5), (2.22), it is easily seen that (X', y', S') satisfies

$$A_i \bullet X' = 0, i = 1, \ldots, m,$$
$$\sum_{i=1}^{m} y'_i A_i + S' = 0,$$

which implies $X' \bullet S' = 0$, i.e.,

$$[S - \psi S^0 - (1-\psi)\tilde{S}] \bullet [X - \psi X^0 - (1-\psi)\widetilde{X}] = 0. \tag{2.30}$$

By expanding (2.30) we obtain the desired result. ∎

Lemma 2.2.2 *Assume that* $\mathcal{F}^*$ *is nonempty. Then for any* $(X^*, y^*, S^*) \in \mathcal{F}^*$ *and* $(X, y, S) \in \mathcal{N}(\alpha, \tau)$ *we have*

$$X \bullet S^0 + X^0 \bullet S \ \le \ (2 + \alpha + \zeta) n \tau_0 \ , \tag{2.31a}$$

$$X \bullet S^* + X^* \bullet S \ \le \ ((1 + \alpha + \psi)/(1-\psi) + \zeta) n \tau \ , \tag{2.31b}$$

where

$$\zeta = (X^0 \bullet S^* + X^* \bullet S^0)/(X^0 \bullet S^0). \tag{2.32}$$

Proof. The results follow by using Lemma 2.2.1 with $(\widetilde{X}, \tilde{y}, \tilde{S}) = (X^*, y^*, S^*)$, Theorem 2.1.1 and the fact that $S^* \bullet X^* = 0$. ∎

Lemma 2.2.2 shows that the pair (X^k, S^k) generated by Algorithm 2.1.1 is bounded. More precisely, we have the following corollary, which can easily be deduced from Lemma 2.2.2 and Theorem 2.1.1.

Corollary 2.2.1

$$\|X^{1/2}(S^0)^{1/2}\|_F \le \sqrt{(2+\alpha+\zeta)n\tau_0},$$
$$\|S^{1/2}(X^0)^{1/2}\|_F \le \sqrt{(2+\alpha+\zeta)n\tau_0},$$
$$\|X^{1/2}\|_F \le \|(S^0)^{-1/2}\|\sqrt{(2+\alpha+\zeta)n\tau_0},$$
$$\|S^{1/2}\|_F \le \|(X^0)^{-1/2}\|\sqrt{(2+\alpha+\zeta)n\tau_0},$$
$$\|X^{1/2}S^{1/2}\|^2 = \|X^{1/2}SX^{1/2}\| \le (1+\alpha)\tau,$$
$$\|X^{-1/2}S^{-1/2}\|^2 = \|X^{-1/2}S^{-1}X^{-1/2}\| \le \frac{1}{(1-\alpha)\tau}.$$

Lemma 2.2.3 *Suppose $(\widetilde{X}, \tilde{y}, \widetilde{S}) \in \mathcal{F}$, and denote*

$$\begin{aligned}
T &= \psi[2X^{1/2}(S^0-\widetilde{S})X^{1/2} + X^{-1/2}(X^0-\widetilde{X})SX^{1/2} + X^{1/2}S(X^0-\widetilde{X})X^{-1/2}] \\
&\quad -2X^{1/2}SX^{1/2}, \\
T_x &= \psi X^{-1/2}(X^0-\widetilde{X})X^{-1/2}, \\
T_s &= \psi X^{1/2}(S^0-\widetilde{S})X^{1/2}.
\end{aligned}$$

Then the quantity δ defined by (2.7) satisfies the inequality:

$$\delta \le \frac{1}{\tau^2}\left(\tau\|T_x\|_F + \frac{\|T\|_F}{2(1-\alpha)}\right)\left(\|T_s\|_F + \frac{\|T\|_F}{2(1-\alpha)}\right).$$

Proof. It is easily seen that $(U+\psi(X^0-\widetilde{X}),\ w+\psi(y^0-\tilde{y}),\ V+\psi(S^0-\widetilde{S}))$ satisfies (2.17) with $H = T/2$, where (U, w, V) is the predictor direction defined by (2.3). Hence, according to Lemma 2.1.2, we have

$$\tau\|X^{-1/2}(U+\psi(X^0-\widetilde{X}))X^{-1/2}\|_F \le \|T/2\|_F/(1-\alpha),$$
$$\|X^{1/2}(V+\psi(S^0-\widetilde{S}))X^{1/2}\|_F \le \|T/2\|_F/(1-\alpha).$$

Therefore,

$$\tau\|X^{-1/2}UX^{-1/2}\| \le \tau\|T_x\| + \|T/2\|_F/(1-\alpha),$$

$$\|X^{1/2}VX^{1/2}\| \leq \|T_s\| + \|T/2\|_F/(1-\alpha),$$

and the lemma follows by virtue of (2.7). ∎

Lemma 2.2.4 *Under the hypothesis of Lemma 2.2.3 we have*

$$\delta \leq \left(\frac{(3-\alpha)(2+\alpha+\zeta)n}{(1-\alpha)^2}d_0 + \frac{1+\alpha}{1-\alpha}\sqrt{n}\right)^2, \tag{2.33}$$

where

$$d_0 = \max(\|(X^0)^{-1/2}(X^0-\widetilde{X})(X^0)^{-1/2}\|_F,\ \|(S^0)^{-1/2}(S^0-\widetilde{S})(S^0)^{-1/2}\|_F).$$

Proof. Using the notation of Lemma 2.2.3, and Lemma 2.2.1, we can write

$$\begin{aligned}
& \|T\|_F \\
\leq\ & 2\psi\|X^{1/2}(S^0-\widetilde{S})X^{1/2}\|_F + 2\psi\|X^{-1/2}(X^0-\widetilde{X})SX^{1/2}\|_F + 2\|X^{1/2}SX^{1/2}\|_F \\
=\ & 2\psi\|X^{1/2}(S^0)^{1/2}(S^0)^{-1/2}(S^0-\widetilde{S})(S^0)^{-1/2}(S^0)^{1/2}X^{1/2}\|_F \\
& +2\psi\|X^{-1/2}S^{-1/2}S^{1/2}(X^0)^{1/2}(X^0)^{-1/2}(X^0-\widetilde{X})(X^0)^{-1/2}(X^0)^{1/2}S^{1/2}S^{1/2}X^{1/2}\|_F \\
& +2\|X^{1/2}SX^{1/2}\|_F \\
\leq\ & 2\psi\|X^{1/2}(S^0)^{1/2}\|^2\|(S^0)^{-1/2}(S^0-\widetilde{S})(S^0)^{-1/2}\|_F \\
& +2\psi\|X^{-1/2}S^{-1/2}\|\|S^{1/2}(X^0)^{1/2}\|^2\|X^{1/2}S^{1/2}\|\|(X^0)^{-1/2}(X^0-\widetilde{X})(X^0)^{-1/2}\|_F \\
& +2\sqrt{n}\|X^{1/2}SX^{1/2}\| \\
\leq\ & 2\psi(2+\alpha+\zeta)n\tau_0 d_0(1+\frac{1+\alpha}{1-\alpha}) + 2\sqrt{n}(1+\alpha)\tau \\
=\ & 2\tau[2(2+\alpha+\zeta)nd_0/(1-\alpha) + (1+\alpha)\sqrt{n}].
\end{aligned}$$

Also,

$$\begin{aligned}
& \|T_x\|_F \\
\leq\ & \psi\|X^{-1/2}S^{-1/2}S^{1/2}(X^0)^{1/2}(X^0)^{-1/2}(X^0-\widetilde{X})(X^0)^{-1/2}(X^0)^{1/2}S^{1/2}S^{-1/2}X^{-1/2}\|_F \\
\leq\ & \psi\|X^{-1/2}S^{-1/2}\|^2\|S^{1/2}(X^0)^{1/2}\|^2 d_0 \\
\leq\ & \frac{\psi(2+\alpha+\zeta)n\tau_0}{\tau(1-\alpha)}d_0 \\
=\ & \frac{(2+\alpha+\zeta)n}{1-\alpha}d_0,
\end{aligned}$$

and

$$\begin{aligned}
\|T_s\|_F &\le \psi \|X^{1/2}(S^0)^{1/2}(S^0)^{-1/2}(S^0-\widetilde{S})(S^0)^{-1/2}(S^0)^{1/2}X^{1/2}\| \\
&\le \psi \|X^{1/2}(S^0)^{1/2}\|^2 d_0 \\
&\le \psi(2+\alpha+\zeta)n\tau_0 d_0 = (2+\alpha+\zeta)n\tau d_0.
\end{aligned}$$

Then (2.33) follows from Lemma 2.2.3. ∎

According to Lemma 2.1.3 and Lemma 2.2.4, it follows that if $\mathcal{F}^*$ is not empty, then the step length $\bar{\theta}_k$ defined by (2.8) is bounded away from 0. This implies global convergence as shown in the following theorem.

Theorem 2.2.1 *If $\mathcal{F}^*$ is not empty, then Algorithm 2.1.1 is globally convergent at a linear rate. Moreover, the iteration sequence (X^k, y^k, S^k) is bounded and every accumulation point of (X^k, y^k, S^k) belongs to $\mathcal{F}^*$ (i.e., is a primal dual optimal solution of the SDP problem).*

Using Lemma 2.2.4, we can easily deduce the following result.

Theorem 2.2.2 *Suppose that $\mathcal{F}^*$ is nonempty and that the starting point is chosen such that there is a constant κ independent of n satisfying the inequality*

$$\begin{aligned}
&(2+\alpha+\zeta)\max(\|(X^0)^{-1/2}(X^0-\widetilde{X})(X^0)^{-1/2}\|_F,\ \|(S^0)^{-1/2}(S^0-\widetilde{S})(S^0)^{-1/2}\|_F) \\
\le\ & \kappa n^{-1/2}.
\end{aligned}$$

Then Algorithm 2.1.1 terminates in at most $O(\sqrt{n}\ln(\epsilon_0/\epsilon))$ iterations, where ϵ_0 is defined by (1.10).

Theorem 2.2.3 *Suppose $X^0 = S^0 = \rho I$, where $\rho > 0$ is a constant such that $\|X^*\| \le \rho$, $\|S^*\| \le \rho$ for some $(X^*, y^*, S^*) \in \mathcal{F}^*$. Then the step length $\bar{\theta}_k$ defined by (2.8) satisfies the inequality*

$$\bar{\theta}_k \ge 1/(\omega n)$$

with $\omega \equiv 1 + 29/\sqrt{\beta-\alpha}$.

Proof. According to Lemma 2.2.2, we have

$$\rho(\mathrm{Tr}(X) + \mathrm{Tr}(S)) \leq (2 + \alpha + \zeta) n\tau_0 = (2 + \alpha + \zeta) n\rho^2,$$

i.e.,

$$\sum_{i=1}^{n} (\lambda_i(X) + \lambda_i(S)) \leq (2 + \alpha + \zeta)\rho n.$$

In the sequel we will frequently use the fact that $\alpha < 0.5$. Since $X^* \bullet S^* = 0$ we get the relation

$$\begin{aligned} \zeta &= (S^* \bullet X^0 + X^* \bullet S^0)/(X^0 \bullet S^0) \\ &= (\mathrm{Tr}(X^*) + \mathrm{Tr}(S^*))/(n\rho) \\ &\leq 1, \end{aligned}$$

which implies

$$\|X^{1/2}\|_F^2 + \|S^{1/2}\|_F^2 = \sum_{i=1}^{n} (\lambda_i(X) + \lambda_i(S)) \leq (3 + \alpha)\rho n. \tag{2.34}$$

It is easily seen that

$$\|X^0 - X^*\| \leq \rho, \quad \|S^0 - S^*\| \leq \rho. \tag{2.35}$$

Applying (2.34), (2.35), Lemma 2.2.1, and Lemma 2.2.3 with $(\widetilde{X}, \tilde{y}, \tilde{S}) = (X^*, y^*, S^*)$ we have

$$\|X^{1/2}(S^0 - S^*)X^{1/2}\| \leq \|X^{1/2}\|_F^2 \|S^0 - S^*\| \leq 3.5\rho^2 n, \tag{2.36}$$

$$\begin{aligned} &\|X^{-1/2}(X^0 - X^*)SX^{1/2}\| \\ &= \|(X^{-1/2}S^{-1/2})S^{1/2}(X^0 - X^*)S^{1/2}(S^{1/2}X^{1/2})\| \\ &\leq \|X^{-1/2}S^{-1/2}\| \|S^{1/2}X^{1/2}\| \|S^{1/2}\|_F^2 \|X^0 - X^*\| \\ &\leq 6.1\rho^2 n. \end{aligned} \tag{2.37}$$

In view of (2.36), (2.37) and Lemma 2.2.3, we obtain

$$\begin{aligned} \|T\|_F &\leq 2[\psi\|X^{1/2}(S^0 - S^*)X^{1/2}\| + \psi\|X^{-1/2}(X^0 - X^*)SX^{1/2}\| \\ &\quad + \|X^{1/2}SX^{1/2}\|] \\ &< 2\tau[9.6n + 1.5] \leq 23\tau n. \end{aligned} \tag{2.38}$$

$$\begin{aligned}
\tau\|T_x\|_F &= \tau\psi\|X^{-1/2}(X^0 - X^*)X^{-1/2}\| \qquad (2.39)\\
&= \tau\psi\|(X^{-1/2}S^{-1/2})S^{1/2}(X^0 - X^*)S^{1/2}(S^{-1/2}X^{-1/2})\|\\
&\le \tau\psi\|S^{1/2}\|_F^2\|X^{-1/2}S^{-1/2}\|^2\|X^0 - X^*\|\\
&< 7\tau n.
\end{aligned}$$

$$\begin{aligned}
\|T_s\|_F &= \psi\|X^{1/2}(S^0 - S^*)X^{1/2}\| \qquad (2.40)\\
&\le \psi\|X^{1/2}\|_F^2\|S^0 - S^*\|\\
&\le 3.5\tau n.
\end{aligned}$$

Therefore, from Lemma 2.2.3, (2.38), (2.39) and (2.40), we obtain

$$\begin{aligned}
\delta &\le \frac{1}{\tau^2}\left(\tau\|T_x\|_F + \frac{\|T\|_F}{2(1-\alpha)}\right)\left(\|T_s\|_F + \frac{\|T\|_F}{2(1-\alpha)}\right)\\
&\le 795n^2.
\end{aligned}$$

Consequently, $\bar{\theta} > 1/(\omega n)$. ∎

In the following corollary we summarize the complexity results for standard starting points of the form $X^0 = S^0 = \rho I$.

Corollary 2.2.2 *Assume that in Algorithm 2.1.1 we choose a starting point of the form $X^0 = S^0 = \rho I$, where $\rho > 0$ is a constant. Let ϵ_0 be given by (1.10) and let $\epsilon > 0$ be arbitrary. Then the following statements hold:*

(i) If $\mathcal{F}^* \neq \emptyset$, then the algorithm terminates with an ϵ-approximate solution

$$(X^k, y^k, S^k) \in \mathcal{F}_\epsilon$$

in a finite number of steps $k = K_\epsilon < \infty$.

(ii) If $\rho \ge \max\{\|X^*\|, \|S^*\|\}$, for some $(X^*, y^*, S^*) \in \mathcal{F}^*$ then

$$K_\epsilon = O(n\ln(\epsilon_0/\epsilon)).$$

(iii) For any choice of $\rho > 0$ there is an index $k = \widehat{K}_\epsilon = O(n\ln(\epsilon_0/\epsilon))$ such that either

(iiia) $(X^k, y^k, S^k) \in \mathcal{F}_\epsilon$,

or,

(iiib) $\bar{\theta} \leq 1/(\omega n)$,

and in the latter case there is no solution $(X^*, y^*, S^*) \in \mathcal{F}^*$ *with* $\rho \geq \max\{\|X^*\|, \|S^*\|\}$.

CHAPTER III
LOCAL CONVERGENCE ANALYSIS

3.1 A sufficient condition

In this section, we discuss the asymptotic behavior of the infeasible-interior-point algorithm described in Chapter 1. A sufficient condition is derived for the superlinear convergence of the algorithm.

The next two lemmas are well known and can be easily proved.

Lemma 3.1.1 *Let* $A_1, A_2 \in S^p_+$. *Then* $A_1 \bullet A_2 = 0$ *if and only if* $A_1 A_2 = A_2 A_1 = 0$.

Lemma 3.1.2 *Let* $(X^*, y^*, S^*) \in \mathcal{F}^*$. *Then, there exists an orthogonal matrix* $Q = (q_1, q_2, \ldots, q_n)$ *such that* $Q^T X^* Q$ *and* $Q^T S^* Q$ *are diagonal matrices. In other words,* $q_1, q_2, \ldots, q_n$ *are eigenvectors of* X^* *and* S^*.

Definition 3.1.1 *A triple* $(X^*, y^*, S^*) \in \mathcal{F}^*$ *is called a* strictly complementary solution *of (1.3) if* $X^* + S^* \succ 0$.

In this section we investigate the asymptotic behavior of Algorithm 2.1.1. We will propose a sufficient condition for the superlinear convergence of Algorithm 2.1.1.

Assumption 1. The SDP problem has a strictly complementary solution.

Let (X^*, y^*, S^*) be a strictly complementary solution of (1.3) and $Q = (q_1, \ldots, q_n)$ be an orthogonal matrix such that $q_1, \ldots, q_n$ are eigenvectors of X^* and S^*. Define

$$\mathbb{B} = \{i : q_i^T X^* q_i > 0\}, \qquad \mathbb{N} = \{i : q_i^T S^* q_i > 0\}.$$

It is easily seen that $\mathbb{B} \cup \mathbb{N} = \{1, 2, \ldots, n\}$. For simplicity, let us assume without loss of generality that

$$Q^T X^* Q = \begin{pmatrix} \Lambda_B & 0 \\ 0 & 0 \end{pmatrix}, \quad Q^T S^* Q = \begin{pmatrix} 0 & 0 \\ 0 & \Lambda_N \end{pmatrix},$$

where Λ_B and Λ_N are diagonal matrices. Here and in the sequel, if we write a matrix M in the block form

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix},$$

then we assume that the dimensions of M_{11} and M_{22} are $|\mathbb{B}| \times |\mathbb{B}|$ and $|\mathbb{N}| \times |\mathbb{N}|$, respectively. In the next lemma we use the following notations:

$$Q^T X^{1/2} Q = (x_1, x_2, \ldots, x_n),$$

$$Q^T S^{1/2} Q = (s_1, s_2, \ldots, s_n),$$

$$Q^T X^{-1/2} Q = (\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_n),$$

$$Q^T S^{-1/2} Q = (\widehat{s}_1, \widehat{s}_2, \ldots, \widehat{s}_n),$$

where $x_i,\ s_i,\ \widehat{x}_i,\ \widehat{s}_i \in \mathbb{R}^n,\ i = 1, \ldots, n$.

Lemma 3.1.3 *Under Assumption 1,*

$$\begin{aligned}
\|x_i\| &= O(\sqrt{\tau}), \quad \|s_i\| = O(1),\ \forall i \in \mathbb{N};\\
\|x_i\| &= O(1), \qquad \|s_i\| = O(\sqrt{\tau}),\ \forall i \in \mathbb{B};\\
\|\widehat{x}_i\| &= O(1), \qquad \|\widehat{s}_i\| = O(1/\sqrt{\tau}),\ \forall i \in \mathbb{B};\\
\|\widehat{x}_i\| &= O(1/\sqrt{\tau}), \quad \|\widehat{s}_i\| = O(1),\ \forall i \in \mathbb{N}.
\end{aligned}$$

Proof. Because the sequence $\{(X^k, S^k)\}$ is bounded, we have

$$\|x_i\| = O(1), \quad \|s_i\| = O(1) \quad \forall i \in \{1, \ldots, n\}.$$

In view of (2.31b), we get

$$\mathrm{Tr}(X^{1/2} S^* X^{1/2}) + \mathrm{Tr}(S^{1/2} X^* S^{1/2}) \le ((1 + \alpha + \psi)/(1 - \psi) + \zeta) n\tau \le \nu_1 \tau,$$

where

$$\nu_1 = ((2 + \alpha)/(1 - \psi_1) + \zeta) n.$$

Let $\nu = \lambda_{\min}(X^* + S^*)$. Since

$$\begin{aligned}
\mathrm{Tr}(X^{1/2} S^* X^{1/2}) &= \mathrm{Tr}(Q^T X^{1/2} Q Q^T S^* Q Q^T X^{1/2} Q)\\
&= \mathrm{Tr}((x_1, \ldots, x_n) Q^T S^* Q (x_1, \ldots, x_n)^T)
\end{aligned}$$

$$= \sum_{i \in \mathbb{N}} (q_i^T S^* q_i) \|x_i\|^2,$$

and $q_i^T S^* q_i \geq \nu > 0$ for each $i \in \mathbb{N}$, we have

$$\|x_i\|^2 \leq \nu_2 \tau, \forall i \in \mathbb{N},$$

where $\nu_2 = \nu_1/\nu$. Hence $\|x_i\| = O(\sqrt{\tau}), \forall i \in \mathbb{N}$. Similarly, $\|s_i\| = O(\sqrt{\tau}), \forall i \in \mathbb{B}$. From Theorem 2.1.1, we obtain

$$X^{1/2} S X^{1/2} - (1-\alpha)\tau I \succeq 0,$$

which implies

$$S - (1-\alpha)\tau X^{-1} \succeq 0,$$

i.e.,

$$[Q^T S^{1/2} Q]^2 - (1-\alpha)\tau [Q^T X^{-1/2} Q]^2 \succeq 0.$$

Therefore,

$$e_i^T [Q^T S^{1/2} Q]^2 e_i \geq (1-\alpha)\tau e_i^T [Q^T X^{-1/2} Q]^2 e_i, \;\; 1 \leq i \leq n.$$

Hence, for any $i \in \mathbb{B}$ we can write

$$\begin{aligned} \|\widehat{x}_i\|^2 &= e_i^T [Q^T X^{-1/2} Q]^2 e_i \\ &\leq \frac{e_i^T [Q^T S^{1/2} Q]^2 e_i}{(1-\alpha)\tau} \\ &= \frac{\|s_i\|^2}{(1-\alpha)\tau} \\ &\leq \nu_2/(1-\alpha). \end{aligned}$$

Also, for any $i \in \mathbb{N}$,

$$\begin{aligned} \|\widehat{x}_i\|^2 &= e_i^T [Q^T X^{-1/2} Q]^2 e_i \\ &\leq \frac{\|s_i\|^2}{(1-\alpha)\tau} = O(1/\tau). \end{aligned}$$

Therefore, $\|\widehat{x}_i\| = O(1), \forall i \in \mathbb{B}$, $\|\widehat{x}_i\| = O(1/\sqrt{\tau}), \; \forall \, i \in \mathbb{N}$. Similarly, by considering $X - (1-\alpha)\tau S^{-1} \succeq 0$, we can show that $\|\widehat{s}_i\| = O(1), \; \forall \; i \in \mathbb{N}$ and $\|\widehat{x}_i\| = O(1/\sqrt{\tau}), \; \forall \, i \in \mathbb{B}$. ∎

Using Lemma 3.1.3, we can write

$$Q^T(X^k)^{1/2}Q = \begin{pmatrix} O(1) & O(\sqrt{\tau_k}) \\ O(\sqrt{\tau_k}) & O(\sqrt{\tau_k}) \end{pmatrix}, \quad Q^T(X^k)^{-1/2}Q = \begin{pmatrix} O(1) & O(1) \\ O(1) & O(1/\sqrt{\tau_k}) \end{pmatrix}, \tag{3.1}$$

$$Q^T(S^k)^{1/2}Q = \begin{pmatrix} O(\sqrt{\tau_k}) & O(\sqrt{\tau_k}) \\ O(\sqrt{\tau_k}) & O(1) \end{pmatrix}, \quad Q^T(S^k)^{-1/2}Q = \begin{pmatrix} O(1/\sqrt{\tau_k}) & O(1) \\ O(1) & O(1) \end{pmatrix}, \tag{3.2}$$

$$Q^T X^k Q = \begin{pmatrix} O(1) & O(\sqrt{\tau_k}) \\ O(\sqrt{\tau_k}) & O(\tau_k) \end{pmatrix}, \quad Q^T S^k Q = \begin{pmatrix} O(\tau_k) & O(\sqrt{\tau_k}) \\ O(\sqrt{\tau_k}) & O(1) \end{pmatrix}. \tag{3.3}$$

Let us define a linear manifold:

$$\begin{aligned} \mathcal{M} \equiv \{(X', y', S') \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n \quad : \quad & A_i \bullet X' = b_i, \quad i = 1, \ldots, m, \\ & \sum_{i=1}^m y_i' A_i + S' = C, \\ & q_i^T X' q_j = 0 \text{ if } i \text{ or } j \in \mathbb{N}, \\ & q_i^T S' q_j = 0 \text{ if } i \text{ or } j \in \mathbb{B}.\} \end{aligned} \tag{3.4}$$

It is easily seen that if $(X', y', S') \in \mathcal{M}$, then

$$Q^T X' Q = \begin{pmatrix} M_B & 0 \\ 0 & 0 \end{pmatrix}, \quad Q^T S' Q = \begin{pmatrix} 0 & 0 \\ 0 & M_N \end{pmatrix}.$$

Lemma 3.1.4 *Under Assumption 1, $\mathcal{F}^* \subset \mathcal{M}$.*

Proof. For any $(\bar{X}^*, \bar{y}^*, \bar{S}^*) \in \mathcal{F}^*$, we have

$$A_i \bullet (\bar{X}^* - X^*) = 0, i = 1, \ldots, m, \quad \sum_{i=1}^m A_i(\bar{y}_i^* - y_i^*) + (\bar{S}^* - S^*) = 0.$$

Hence,

$$(\bar{S}^* - S^*) \bullet (\bar{X}^* - X^*) = 0,$$

which implies

$$\bar{S}^* \bullet X^* + S^* \bullet \bar{X}^* = \bar{S}^* \bullet \bar{X}^* + S^* \bullet X^* = 0.$$

Then, $\bar{S}^* \bullet X^* = S^* \bullet \bar{X}^* = 0$. By Lemma 3.1.1, we have

$$\bar{S}^* X^* = X^* \bar{S}^* = 0,$$

which implies

$$[Q^T \bar{S}^* Q][Q^T X^* Q] = [Q^T X^* Q][Q^T \bar{S}^* Q] = 0,$$

i.e.,

$$[q_i^T \bar{S}^* q_j][q_j^T X^* q_j] = [q_i^T X^* q_i][q_i^T \bar{S}^* q_j] = 0. \tag{3.5}$$

If i or $j \in \mathbb{B}$, then $q_j^T X^* q_j$ or $q_i^T X^* q_i$ is positive, which implies $q_i^T \bar{S}^* q_j = 0$ according to (3.5). Similarly, we can show that $q_i^T \bar{X}^* q_j = 0$ whenever i or $j \in \mathbb{N}$. Therefore, $(\bar{X}^*, \bar{y}^*, \bar{S}^*) \in \mathcal{M}$, which gives $\mathcal{F}^* \subset \mathcal{M}$. ∎

Lemma 3.1.5 *Under Assumption 1, every accumulation point of* (X^k, y^k, S^k) *is a strictly complementary solution of (1.3).*

Proof. Suppose $(X", y", S") \in \mathcal{F}^*$ is an accumulation point. Let us assume, without loss of generality, that $(X^k, y^k, S^k) \to (X", y", S")$. Then, according to Lemma 3.1.4,

$$Q^T X^k Q \to Q^T X" Q = \begin{pmatrix} M_B & 0 \\ 0 & 0 \end{pmatrix},$$
$$Q^T S^k Q \to Q^T S" Q = \begin{pmatrix} 0 & 0 \\ 0 & M_N \end{pmatrix}.$$

for some symmetric positive semidefinite matrices M_B and M_N. In order to show $X" + S" \succ 0$, it remains to prove that M_B and M_N are nonsingular and therefore positive definite. From Lemma 3.1.3 or (3.1), we have

$$Q^T (X^k)^{1/2} Q = \begin{pmatrix} M_1^k & O(\sqrt{\tau_k}) \\ O(\sqrt{\tau_k}) & O(\sqrt{\tau_k}) \end{pmatrix},$$
$$Q^T (X^k)^{-1/2} Q = \begin{pmatrix} M_2^k & O(1) \\ O(1) & O(1/\sqrt{\tau_k}) \end{pmatrix},$$

where $M_1^k = O(1)$, $M_2^k = O(1)$. Since $((X^k)^{1/2}, (S^k)^{1/2})$ is bounded, it has an accumulation point. Without loss of generality, we may assume

$$Q^T (X^k)^{1/2} Q \to \begin{pmatrix} \widetilde{M}_1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then, $M_1^k \to \widetilde{M}_1$. Since

$$[Q^T (X^k)^{1/2} Q][Q^T (X^k)^{-1/2} Q] = I,$$

we obtain

$$M_1^k M_2^k = I_B - O(\sqrt{\tau_k}) \to I_B,$$

which implies

$$(M_1^k)^{-1} = M_2^k \left(I_B - O(\sqrt{\tau_k})\right)^{-1} = O(1).$$

Hence $\widetilde{M}_1$ must be nonsingular. Obviously $\widetilde{M}_1^2 = M_B$, so that M_B is nonsingular. Similarly, we can show that M_N is nonsingular. ∎

In the next theorem, we propose a sufficient condition for the superlinear convergence of Algorithm 2.1.1. Let us define

$$\eta_k = \eta_k(\Gamma) = \frac{1}{\tau_k}\|(X^k)^{-1/2}(X^k - \breve{X}^k)(S^k - \breve{S}^k)(X^k)^{1/2}\|_F, \tag{3.6}$$

where $(\breve{X}^k, \breve{y}^k, \breve{S}^k)$ is the solution of the following minimization problem:

$$\min\{\|(X^k)^{-1/2}(X^k - X')(S^k - S')(X^k)^{1/2}\|_F : (X', y', S') \in \mathcal{M},\ \|(X', S')\|_F \leq \Gamma\}, \tag{3.7}$$

and Γ is a constant such that $\|(X^k, S^k)\|_F \leq \Gamma$, $\forall k$. For a definition of $\|(X, S)\|_F$ see the notation introduced at the end of Section 1. Note that every accumulation point of (X^k, y^k, S^k) belongs to the feasible set of the above minimization problem and the feasible set is bounded. Therefore $(\breve{X}^k, \breve{S}^k)$ exists for each k.

Theorem 3.1.1 *Under Assumption 1, if $\eta_k \to 0$ as $k \to \infty$, then Algorithm 1.3 is superlinearly convergent. Moreover, if there exists a constant $\sigma > 0$ such that $\eta_k = O(\tau_k^\sigma)$, then the convergence has Q-order of at least $1 + \sigma$ in the sense that $\mu_{k+1} = O(\mu_k^{1+\sigma})$.*

Proof. By Lemma 2.1.3, it remains to prove that $\delta_k \to 0$ as $k \to \infty$. For simplicity, let us omit the index k. It is easily seen that $(U + X - \breve{X}, w + y - \breve{y}, V + S - \breve{S})$ satisfies (3.57) with

$$\begin{aligned} 2H &= X^{-1/2}[X(S - \breve{S}) + (X - \breve{X})S]X^{1/2} \\ &\quad + X^{1/2}[(S - \breve{S})X + S(X - \breve{X})]X^{-1/2} - 2X^{1/2}SX^{1/2} \\ &= \Delta + \Delta^T, \end{aligned}$$

where

$$\Delta = X^{-1/2}(X - \breve{X})(S - \breve{S})X^{1/2}.$$

Here we have used the relation $\breve{X}\breve{S} = \breve{S}\breve{X} = 0$. Then,

$$\|H\|_F \leq \|\Delta\|_F = \eta\tau. \tag{3.8}$$

Denoting

$$\Delta_x = X^{-1/2}(U + X - \breve{X})X^{-1/2},$$

$$\Delta_s = X^{1/2}(V + S - \breve{S})X^{1/2},$$

and applying Lemma 3.5.3, we obtain

$$\tau\|\Delta_x\|_F \leq \|H\|_F/(1-\alpha) \leq \eta\tau/(1-\alpha),$$

which implies

$$\|\Delta_x\|_F \leq \eta/(1-\alpha). \tag{3.9}$$

Similarly,

$$\|\Delta_s\|_F \leq \|H\|_F/(1-\alpha) \leq \eta\tau/(1-\alpha). \tag{3.10}$$

By Lemma 3.1.3 and the fact that $(\breve{X}, \breve{y}, \breve{S}) \in \mathcal{M}$, we have

$$\begin{aligned}
\|X^{-1/2}(X - \breve{X})X^{-1/2}\|_F &= \|I - X^{-1/2}\breve{X}X^{-1/2}\|_F \\
&\leq \|I\|_F + \|X^{-1/2}\breve{X}X^{-1/2}\|_F \\
&= \sqrt{n} + \|Q^T X^{-1/2} Q Q^T \breve{X} Q Q^T X^{-1/2} Q\|_F \\
&= \sqrt{n} + \|(\widehat{x_1}, \ldots, \widehat{x_n}) Q^T \breve{X} Q (\widehat{x_1}, \ldots, \widehat{x_n})^T\|_F \\
&= \sqrt{n} + \|\textstyle\sum_{i,j \in \mathsf{B}} (q_i^T \breve{X} q_j) \widehat{x_i} \widehat{x_j}^T\|_F \\
&= O(1).
\end{aligned} \tag{3.11}$$

Similarly,

$$\|S^{-1/2}(S - \breve{S})S^{-1/2}\|_F = O(1). \tag{3.12}$$

Let us observe that

$$X^{-1/2}UVX^{1/2} = \left(X^{-1/2}UX^{-1/2}\right)\left(X^{1/2}VX^{1/2}\right)$$

$$
\begin{aligned}
&= \left(\Delta_x - X^{-1/2}(X-\check{X})X^{-1/2}\right)\left(\Delta_s - X^{1/2}(S-\check{S})X^{1/2}\right)\\
&= \Delta_x\Delta_s - \left(X^{-1/2}(X-\check{X})X^{-1/2}\right)\Delta_s\\
&\quad -\Delta_x X^{1/2}S^{1/2}\left(S^{-1/2}(S-\check{S})S^{-1/2}\right)S^{1/2}X^{1/2} + \Delta.
\end{aligned}
$$

Then from (3.8), (3.9), (3.10), (3.11), (3.12) and Corollary 2.2.1, we have

$$
\begin{aligned}
\|X^{-1/2}UVX^{1/2}\|_F &\le \|\Delta_x\|_F\|\Delta_s\|_F + \|X^{-1/2}(X-\check{X})X^{-1/2}\|_F\|\Delta_s\|_F\\
&\quad +\|X^{1/2}S^{1/2}\|^2\|\Delta_x\|_F\|S^{-1/2}(S-\check{S})S^{-1/2}\|_F + \|\Delta\|_F\\
&= O(\eta\tau).
\end{aligned}
$$

Hence, $\delta_k = O(\eta_k) \to 0$. Finally, if $\eta_k = O(\tau_k^\sigma)$ for some constant $\sigma > 0$, then we have $\delta_k = O(\tau_k^\sigma)$. From Lemma 2.1.3,

$$
\begin{aligned}
1-\bar{\theta} &\le 1 - \frac{2}{1+\sqrt{1+4\delta/(\beta-\alpha)}}\\
&= \frac{\sqrt{1+4\delta/(\beta-\alpha)}-1}{\sqrt{1+4\delta/(\beta-\alpha)}+1}\\
&= \frac{4\delta/(\beta-\alpha)}{(\sqrt{1+4\delta/(\beta-\alpha)}+1)^2}\\
&\le \delta/(\beta-\alpha) = O(\tau^\sigma).
\end{aligned}
$$

Therefore, $\tau_{k+1} = (1-\bar{\theta}_k)\tau_k = O(\tau_k^{1+\sigma})$. Recalling (2.24), we obtain $\mu_{k+1} = O(\mu_k^{1+\sigma})$.
∎

Kojima, Shida and Shindoh [32] established the superlinear convergence under the following three assumptions:

(A) SDP has a strictly complementary solution;

(B) SDP is nondegenerate in the sense that the Jacobian matrix of its KKT system is nonsingular;

(C) the iterates converge tangentially to the central path in the sense that the size of the neighborhood in wich the iterates reside must approach zero.

We mention that our sufficient condition in the above theorem is satisfied under conditions (A), (B) and (C) of Kojima, Shida and Shindoh [32]. The proof is provided in the next section. As a matter of fact, our sufficient condition is satisfied under conditions (A) and (C) only.

3.2 More specific conditions

The purpose of this section is to establish the superlinear convergence of the algorithm proposed by Kojima, Shida and Shindoh [32] and the present authors [54], under assumptions (A) and (C). In other words, the nondegeneracy condition (B) is not necessary for the superlinear convergence of the algorithm. Our analysis is based on the sufficient condition for the superlinear convergence established in [54]. We show that under assumptions (A) and (C) that sufficient condition is satisfied.

We write (C) as

Assumption 2. $\lim_{k\to\infty} \|(X^k)^{1/2}S^k(X^k)^{1/2} - \tau_k I\|_F/\tau_k = 0.$

Let us write

$$Q^T X^k Q = \begin{pmatrix} \hat{X}_B^k & \hat{X}_J^k \\ (\hat{X}_J^k)^T & \hat{X}_N^k \end{pmatrix}, \quad Q^T S^k Q = \begin{pmatrix} \hat{S}_B^k & \hat{S}_J^k \\ (\hat{S}_J^k)^T & \hat{S}_N^k \end{pmatrix}. \tag{3.13}$$

From Lemma 3.1.3 and Lemma 3.1.4, we can easily get

$$\hat{X}_B^k = \Omega(1), \quad \hat{S}_N^k = \Omega(1). \tag{3.14}$$

The quantities ω_k and ϕ_k defined below are used in the next two lemmas.

$$\omega_k = \max\{\|\hat{X}_J^k\|_F, \ \|\hat{S}_J^k\|_F, \ \tau_k\}. \tag{3.15}$$

$$\phi_k = \max\left\{\|(X^k)^{1/2}S^k(X^k)^{1/2} - \tau_k I\|_F/\tau_k, \quad \sqrt{\tau_k}\right\}. \tag{3.16}$$

Lemma 3.2.1 *Under Assumption 1, we have*

$$\|(X^k - \check{X}^k, \ S^k - \check{S}^k)\|_F = O(\omega_k), \tag{3.17}$$

where $(\check{X}^k, \check{S}^k)$ *is the solution of the minimization problem:*

$$\min_{(X',y',S')\in\mathcal{M}} \|(X^k - X', \ S^k - S')\|_F. \tag{3.18}$$

Proof. Suppose by contradiction that (3.17) does not hold, i.e., there exists a subsequence such that

$$\xi_k = \|(X^k - \check{X}^k,\ S^k - \check{S}^k)\|_F/\omega_k \to \infty \text{ as } k \to \infty. \tag{3.19}$$

Let us define

$$(\Delta X^k, \Delta S^k) = \frac{1}{\omega_k \xi_k}(X^k - \check{X}^k,\ S^k - \check{S}^k). \tag{3.20}$$

It is easily seen that

$$A_i \bullet \Delta X^k = \frac{\tau_k}{\tau_0 \omega_k \xi_k} A_i \bullet (X^0 - X^*),\ i = 1, \ldots, m, \tag{3.21a}$$

$$\sum_{i=1}^{m} (\Delta y_i^k) A_i + \Delta S^k = \frac{\tau_k}{\tau_0 \omega_k \xi_k}[\sum_{i=1}^{m}(y_i^0 - y_i^*)A_i + S^0 - S^*]. \tag{3.21b}$$

Since $\|(\Delta X^k, \Delta S^k)\|_F = 1$ and Δy^k depends linearly on ΔS^k (cf. (3.21)), we deduce that there exists a convergent subsequence of $(\Delta X^k, \Delta y^k, \Delta S^k)$. Without loss of generality we can write

$$(\Delta X^k, \Delta y^k, \Delta S^k) \to (\Delta X', \Delta y', \Delta S').$$

Letting $k \to \infty$ in (3.21), we obtain

$$A_i \bullet \Delta X' = 0,\ i = 1, \ldots, m,\ \ \sum_{i=1}^{m} (\Delta y_i') A_i + \Delta S' = 0. \tag{3.22}$$

From (3.3), we have for any $i \in \mathbb{B}$,

$$|q_i^T(\Delta S^k) q_i| = |q_i^T S^k q_i|/(\xi_k \omega_k) \le \|\hat{S}_B^k\|_F/(\xi_k \omega_k) = O(\tau_k)/(\xi_k \omega_k) = O(1/\xi_k) = o(1),$$

which implies $q_i^T \Delta S' q_i = 0$ for each $i \in \mathbb{B}$. Similarly, we can show that $q_i^T \Delta X' q_i = 0$ for any $i \in \mathbb{N}$. For each pair i, j where $i \in \mathbb{B}, j \in \mathbb{N}$ or $i \in \mathbb{N}, j \in \mathbb{B}$, we have

$$|q_i^T(\Delta S^k) q_j| = |q_i^T S^k q_j|/(\xi_k \omega_k) \le \|\hat{S}_J^k\|_F/(\xi_k \omega_k) \le 1/\xi_k = o(1),$$

which implies $q_i^T \Delta S' q_j = 0$. Similarly, $q_i^T \Delta X' q_j = 0$ for any pair i, j described above. Hence,

$$(\check{X}^k, \check{y}^k, \check{S}^k) + \nu(\Delta X', \Delta y', \Delta S') \in \mathcal{M},\ \text{for all } \nu \in \mathbb{R}.$$

Since

$$\begin{aligned} & \frac{\|(X^k - (\check{X}^k + \omega_k \xi_k \Delta X'),\ \ S^k - (\check{S}^k + \omega_k \xi_k \Delta S'))\|_F}{\|(X^k - \check{X}^k,\ S^k - \check{S}^k))\|_F} \\ = \ & \frac{1}{\omega_k \xi_k}\|(X^k - (\check{X}^k + \omega_k \xi_k \Delta X'),\ \ S^k - (\check{S}^k + \omega_k \xi_k \Delta S'))\|_F \end{aligned}$$

$$= \quad \|(\Delta X^k - \Delta X',\ \Delta S^k - \Delta S')\|_F$$

$$\to \quad 0 \ \text{ as } \ k \to \infty,$$

$(\check{X}^k, \check{y}^k\ \check{S}^k)$ is not the solution of the minimization problem (3.18) for sufficiently large k, which is a contradiction. ∎

Lemma 3.2.2 *Under Assumption 1, the quantities defined by (3.13) and (3.16) satisfy the following relation:*

$$\hat{X}_J^k = O(\phi_k\sqrt{\tau_k}), \quad \hat{S}_J^k = O(\phi_k\sqrt{\tau_k}). \tag{3.23}$$

Proof. The proof is by contradiction. Suppose that (3.23) does not hold. Without loss of generality, we may assume $\hat{X}_J^k \neq O(\phi_k\sqrt{\tau_k})$. Then, there exists a subsequence

$$\chi_k = \frac{\|\hat{X}_J^k\|_F}{\phi_k\sqrt{\tau_k}} \to \infty, \quad \text{as} \quad k \to \infty. \tag{3.24}$$

By (3.16), we have

$$\|(X^k)^{1/2}S^k(X^k)^{1/2} - \tau_k I\|_F \leq \phi_k\tau_k. \tag{3.25}$$

In view of (3.1), we obtain

$$\|X^kS^k - \tau_k I\|_F \leq \|(X^k)^{1/2}\|_F\|(X^k)^{1/2}S(X^k)^{1/2} - \tau_k I\|_F\|(X^k)^{-1/2}\|_F = O(\phi_k\sqrt{\tau_k}). \tag{3.26}$$

Therefore,

$$X^kS^k = O(\phi_k\sqrt{\tau_k}),$$

which implies

$$\begin{pmatrix} \hat{X}_B^k & \hat{X}_J^k \\ (\hat{X}_J^k)^T & \hat{X}_N^k \end{pmatrix}\begin{pmatrix} \hat{S}_B^k & \hat{S}_J^k \\ (\hat{S}_J^k)^T & \hat{S}_N^k \end{pmatrix} = O(\phi_k\sqrt{\tau_k}). \tag{3.27}$$

The above equation and (3.24) yield

$$\hat{X}_B^k\hat{S}_J^k + \hat{X}_J^k\hat{S}_N^k = O(\phi_k\sqrt{\tau_k}) = o(\|\hat{X}_J^k\|_F). \tag{3.28}$$

Hence,

$$\hat{S}_J^k = -(\hat{X}_B^k)^{-1}\hat{X}_J^k\hat{S}_N^k + (\hat{X}_B^k)^{-1}o(\|\hat{X}_J^k\|_F). \tag{3.29}$$

Since according to (3.14) we have $\hat{X}_B^k = \Omega(1),\ \hat{S}_N^k = \Omega(1)$ it follows that

$$\hat{S}_J^k = \Omega(\|\hat{X}_J^k\|_F). \tag{3.30}$$

Now, let us choose a convergent subsequence

$$\lim_{k\to\infty} X^k = X^*, \quad \lim_{k\to\infty} S^k = S^*.$$

By virtue of (3.30), we can also choose a convergent subsequence

$$\lim_{k\to\infty} \frac{\hat{X}_J^k}{\|\hat{X}_J^k\|_F} = X'_J,\ \lim_{k\to\infty} \frac{\hat{S}_J^k}{\|\hat{X}_J^k\|_F} = S'_J$$

Define

$$Q^T X^* Q = \begin{pmatrix} \hat{X}_B^* & 0 \\ 0 & 0 \end{pmatrix},\ \ Q^T S^* Q = \begin{pmatrix} 0 & 0 \\ 0 & \hat{S}_N^* \end{pmatrix}.$$

Upon dividing both sides of (3.28) by $\|\hat{X}_J^k\|_F$ and letting $k \to 0$ we get

$$\hat{X}_B^* S'_J + X'_J \hat{S}_N^* = 0.$$

According to Lemma 3.1.5, $\hat{X}_B^*, \hat{S}_N^*$ are both positive definite matrices. Note that $\|X'_J\|_F = 1$. Hence,

$$\begin{aligned} X'_J \bullet S'_J &= -X'_J \bullet [(X_B^*)^{-1} X'_J S_N^*] \\ &= -\mathrm{Tr}((X_B^*)^{-1} X'_J S_N^* X_J'^T) \\ &= -\mathrm{Tr}((X_B^*)^{-1/2} X'_J S_N^* X_J'^T (X_B^*)^{-1/2}) \\ &= -\mathrm{Tr}([(X_B^*)^{-1/2} X'_J (S_N^*)^{1/2}][(X_B^*)^{-1/2} X'_J (S_N^*)^{1/2}]^T) \\ &< 0. \end{aligned} \tag{3.31}$$

Let $(\check{X}^k,\ \check{y}^k,\ \check{S}^k) \in \mathcal{M}$ be defined as in Lemma 3.2.1 and let

$$X'' = X^k - \check{X}^k + \psi_k(X^* - X^0),$$

$$y'' = y^k - \check{y}^k + \psi_k(y^* - y^0),$$

$$S'' = S^k - \check{S}^k + \psi_k(S^* - S^0),$$

where $\psi_k = \tau_k/\tau_0$. From Theorem 2.1.1, it is easily seen that

$$A_i \bullet X'' = 0,\ i = 1, \ldots, m,\ \sum_{i=1}^m A_i y''_i + S'' = 0.$$

A simple calculation shows that

$$[X^k - \check{X}^k + \psi_k(X^* - X^0)] \bullet [S^k - \check{S}^k + \psi_k(S^* - S^0)] = 0.$$

Then, by using Lemma 3.2.1 we obtain

$$(X^k - \check{X}^k) \bullet (S^k - \check{S}^k) = O(\omega_k \tau_k),$$

which implies

$$[Q^T(X^k - \check{X}^k)Q] \bullet [Q^T(S^k - \check{S}^k)Q] = O(\omega_k \tau_k),$$

i.e.,

$$(\hat{X}_B^k - \tilde{X}_B^k) \bullet \hat{S}_B^k + 2\hat{X}_J^k \bullet \hat{S}_J^k + \hat{X}_N^k \bullet (\hat{S}_N^k - \tilde{S}_N^k) = O(\omega_k \tau_k), \tag{3.32}$$

where

$$Q^T \check{X}^k Q = \begin{pmatrix} \tilde{X}_B^k & 0 \\ 0 & 0 \end{pmatrix}, \quad Q^T \check{S}^k Q = \begin{pmatrix} 0 & 0 \\ 0 & \tilde{S}_N^k \end{pmatrix}.$$

Note that from Lemma 3.2.1 we have

$$\hat{X}_B^k - \tilde{X}_B^k = O(\omega_k), \quad \hat{S}_N^k - \tilde{S}_N^k = O(\omega_k). \tag{3.33}$$

Also, from Lemma 3.1.3 it follows that

$$\hat{S}_B^k = O(\tau_k), \; \hat{X}_N^k = O(\tau_k). \tag{3.34}$$

In view of (3.32), (3.33) and (3.34), we deduce

$$\hat{X}_J^k \bullet \hat{S}_J^k = O(\omega_k \tau_k). \tag{3.35}$$

Let us observe that

$$\begin{aligned} \frac{\omega_k \tau_k}{\|\hat{X}_J^k\|_F^2} &= \frac{\tau_k \max\{\|\hat{X}_J^k\|_F, \; \|\hat{S}_J^k\|_F, \; \tau_k\}}{\|\hat{X}_J^k\|_F^2} \\ &= \frac{\tau_k}{\|\hat{X}_J^k\|_F} \max\left\{1, \; \frac{\|\hat{S}_J^k\|_F}{\|\hat{X}_J^k\|_F}, \; \frac{\tau_k}{\|\hat{X}_J^k\|_F}\right\} \\ &= o(1), \end{aligned} \tag{3.36}$$

since $\dfrac{\|\hat{S}_J^k\|_F}{\|\hat{X}_J^k\|_F}$ is bounded and

$$\frac{\tau_k}{\|\hat{X}_J^k\|_F} = \frac{\sqrt{\tau_k}}{\chi_k \phi_k} \le 1/\chi_k = o(1).$$

Dividing both sides of (3.35) by $\|\hat{X}_J^k\|_F^2$, recalling (3.36), and letting $k \to \infty$, we obtain

$$X'_J \bullet S'_J = 0,$$

which contradicts (3.31).

∎

From Lemma 3.2.2 and (3.15), we get

$$\omega_k = O(\phi_k \sqrt{\tau_k}). \tag{3.37}$$

Theorem 3.2.1 *Under Assumptions 1 and 2, Algorithm 2.1.1 is superlinearly convergent. Moreover, if* $\|X^{1/2}SX^{1/2} - \tau I\|_F = O(\tau^{1+\sigma})$ *for some constant* $\sigma > 0$*, then the convergence has Q-order at least* $1 + \min\{\sigma,\ 0.5\}$*.*

Proof. From Lemma 3.1.3, (3.14) and Lemma 3.2.2, we deduce that

$$Q^T X^k Q = \begin{pmatrix} \Omega(1) & O(\phi_k\sqrt{\tau_k}) \\ O(\phi_k\sqrt{\tau_k}) & O(\tau_k) \end{pmatrix}, \quad Q^T S^k Q = \begin{pmatrix} O(\tau_k) & O(\phi_k\sqrt{\tau_k}) \\ O(\phi_k\sqrt{\tau_k}) & \Omega(1) \end{pmatrix}. \tag{3.38}$$

Let $(\check{X}^k, \check{y}^k, \check{S}^k) \in \mathcal{M}$ be defined as in Lemma 3.2.1. Then, from (3.38), (3.37) and Lemma 3.2.1, we have

$$Q^T(X^k - \check{X}^k)Q = \begin{pmatrix} O(\phi_k\sqrt{\tau_k}) & O(\phi_k\sqrt{\tau_k}) \\ O(\phi_k\sqrt{\tau_k}) & O(\tau_k) \end{pmatrix}, \tag{3.39}$$

$$Q^T(S^k - \check{S}^k)Q = \begin{pmatrix} O(\tau_k) & O(\phi_k\sqrt{\tau_k}) \\ O(\phi_k\sqrt{\tau_k}) & O(\phi_k\sqrt{\tau_k}) \end{pmatrix}. \tag{3.40}$$

Since (X^k, S^k) is bounded, so is $(\check{X}^k,\ \check{S}^k)$. So, we may choose Γ such that

$$\|(X^k, S^k)\|_F \le \Gamma, \quad \|(\check{X}^k, \check{S}^k)\|_F \le \Gamma, \quad \forall\ k.$$

Hence,

$$\begin{aligned} \eta_k &= \eta_k(\Gamma) = \frac{1}{\tau_k}\|(X^k)^{-1/2}(X^k - \check{X}^k)(S^k - \check{S}^k)(X^k)^{1/2}\|_F \\ &\le \frac{1}{\tau_k}\|(X^k)^{-1/2}(X^k - \check{X}^k)(S^k - \check{S}^k)(X^k)^{1/2}\|_F \\ &= \frac{1}{\tau_k}\|[Q^T(X^k)^{-1/2}Q][Q^T(X^k - \check{X}^k)Q][Q^T(S^k - \check{S}^k)Q][Q^T(X^k)^{1/2}Q]\|_F. \end{aligned}$$

In view of (3.1), (3.39) and (3.40), we have

$$[Q^T X^{-1/2} Q][Q^T(X - \check{X})Q][Q^T(S - \check{S})Q][Q^T X^{1/2} Q]$$

$$= \begin{pmatrix} O(1) & O(1) \\ O(1) & O(\tau^{-.5}) \end{pmatrix} \begin{pmatrix} O(\phi\tau^{.5}) & O(\phi\tau^{.5}) \\ O(\phi\tau^{.5}) & O(\tau) \end{pmatrix} \begin{pmatrix} O(\tau) & O(\phi\tau^{.5}) \\ O(\phi\tau^{.5}) & O(\phi\tau^{.5}) \end{pmatrix} \begin{pmatrix} O(1) & O(\tau^{.5}) \\ O(\tau^{.5}) & O(\tau^{.5}) \end{pmatrix}$$
$$= \begin{pmatrix} O(\phi^2\tau) & O(\phi^2\tau^{1.5}) \\ O(\phi\tau) & O(\phi^2\tau) \end{pmatrix},$$

where for simplicity we omit the index k. Therefore $\eta_k = O(\phi_k)$, which proves the theorem by invoking Theorem 3.1.1. ∎

3.3 The modified algorithm

In a recent paper, Kojima, Shida and Shindoh [33] proposed a modification of Algorithm 2.1.1 which enforces Assumption 2. By performing additional corrector steps, the iterates get more and more centered at a quadratic rate.

Algorithm 3.3.1 $k \leftarrow 0$; $(X, y, S) \leftarrow (X^0, y^0, S^0)$;

Repeat until $(X, y, S) \in \mathcal{F}_\epsilon$ *or* $\bar{\theta}$ *is sufficiently small:*

(Predictor step) *Do the same as in Algorithm 2.1.1*

(Corrector step) *Repeat the corrector step in Algorithm 2.1.1 until* $(X, S) \in \mathcal{N}(\min\{\tau^{\bar{\sigma}}, \alpha\},\ \tau)$;

$k \leftarrow k + 1$; $(X^k, y^k, S^k) \leftarrow (X, y, S)$.

In Algorithm 3.3.1, $\bar{\sigma}$ is a positive constant. Kojima, Shida and Shindoh proved in [33] that Algorithm 3.3.1 has polynomial complexity, and is superlinearly convergent under assumptions (A) and (B). Since the local convergence results in the preceding section also apply to Algorithm 3.3.1, we obtain the following theorem.

Theorem 3.3.1 *Under Assumption 1, Algorithm 3.3.1 is superlinearly convergent with Q-order at least* $1 + \min\{\bar{\sigma},\ 0.5\}$.

3.4 A weaker assumption

Assumption 2 is quite strong since even in the LP case (X^k and S^k are diagonal) it is not satisfied in general. This arises an interesting question: is there a weaker assumption than Assumption 2 so that we can expect it is satisfied at least for some cases such as LP? The following weaker assumption works.

Assumption 2'. The iteration sequence satisfies $\lim_{k\to\infty} \dfrac{X^k S^k}{\sqrt{X^k \bullet S^k}} = 0$.

Assumption 2' is equivalent to $\lim_{k\to\infty} X^k S^k/\sqrt{\tau_k} = 0$. It is easily seen from (3.26) that Assumption 2 implies Assumption 2'. In the LP case, Assumption 2' is satisfied as long as Algorithm 2.1.1 is convergent.

Theorem 3.4.1 *Under Assumptions 1 and 2', Algorithm 2.1.1 is superlinearly convergent. Moreover, if* $X^k S^k = O(\tau_k^{0.5+\sigma})$ *for some constant* $\sigma > 0$*, then the convergence has Q-order at least* $1 + \min\{\sigma,\ 0.5\}$.

Proof. Define

$$\phi'_k = \max\left\{\|X^k S^k\|_F/\sqrt{\tau_k}, \quad \sqrt{\tau_k}\right\}.$$

Then, a modification of the proof of Lemma 3.2.2 gives

$$\hat{X}_J^k = O(\phi'_k\sqrt{\tau_k}), \quad \hat{S}_J^k = O(\phi'_k\sqrt{\tau_k}). \tag{3.41}$$

This can be done by replacing ϕ_k by ϕ'_k and removing (3.25) and (3.26) in the proof of Lemma 3.2.2. Therefore,

$$\omega_k = O(\phi'_k\sqrt{\tau_k}). \tag{3.42}$$

In view of (3.41) and (3.42), the theorem follows by an analogous proof to Theorem 3.2.1. ∎

3.5 On Monteiro's algorithm

In this section, we will investigate the local behavior of the predictor-corrector algorithm considered by Monteiro [42] for SDP using the MZ-family of search directions. We show that the sufficient condition of Potra and Sheng [54] (see Theorem 3.1.1) for superlinear convergence applies for this algorithm. The sufficient condition is independent of scaling matrices. In particular we show that the algorithm is superlinearly convergent if (A) and (D) are satisfied. More specifically, we show that under the assumptions (A) and (B), superlinear convergence with Q-order

1.5 is obtained if the scaling matrices in the corrector step have bounded condition number. Finally, we propose a new version of the predictor-corrector algorithm which enjoys quadratic convergence if the scaling matrices in both predictor and corrector steps have bounded condition numbers and (A) and (B) are satisfied.

We consider the symmetrization operator [70]

$$H_P(M) = \frac{1}{2}\left[PMP^{-1} + (PMP^{-1})^T\right], \quad \forall M \in \mathbb{R}^{n\times n}.$$

Since, as observed by Zhang [70],

$$H_P(M) = \tau I \text{ iff } M = \tau I,$$

for any nonsingular matrix P, any matrix M with real spectrum, and any $\tau \in \mathbb{R}$, it follows that for any given nonsingular matrix P, (1.3) is equivalent to

$$A_i \bullet X = b_i, \; i = 1, \ldots, m, \tag{3.43a}$$

$$\sum_{i=1}^m y_i A_i + S = C, \tag{3.43b}$$

$$H_P(XS) = 0, \quad X \succeq 0, \quad S \succeq 0. \tag{3.43c}$$

A perturbed Newton method applied to the system (3.43) leads to the following linear system:

$$H_P(XV + US) = \xi\mu I - H_P(XS), \tag{3.44a}$$

$$A_i \bullet U = 0, \quad i = 1, \ldots, m, \tag{3.44b}$$

$$\sum_{i=1}^m w_i A_i + V = 0, \tag{3.44c}$$

where $(U, w, V) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$ is the unknown search direction, $\xi \in [0, 1]$ is the centering parameter, and $\mu = (X \bullet S)/n$ is the normalized duality gap corresponding to (X, y, S).

The search direction obtained through (3.44) is called the Monteiro-Zhang (MZ) unified direction [70, 46]. The matrix P used in (3.44) is called the scaling matrix for the search direction. It is well known that taking $P = I$ results in the Alizadeh-Haeberly-Overton (AHO) search direction [4], $P = S^{1/2}$ corresponds to the HRVW/KSH/M search direction [35, 20, 41], and the case of

$$P^T P = X^{-1/2}[X^{1/2}SX^{1/2}]^{1/2}X^{-1/2}$$

coincides with the Nesterov-Todd (NT) search direction [51]. Monteiro and Zhang [46] established the polynomiality of a long-step path-following method based on search directions defined by scaling matrices belonging to the class

$$\{W^{1/2} \ : \ W \in \mathcal{S}^n_+ \text{ such that } WXS = SXW\}.$$

Following [46], Sheng et al. [58] proved the polynomiality of a Mizuno-Todd-Ye type predictor-corrector algorithm for SDP by imposing the scaling matrices to be chosen from the class

$$\{P \ : \ P \in \mathbb{R}^{n\times n} \text{ is nonsingular and } \quad PXSP^{-1} \in \mathcal{S}^n\}.$$

Moreover, its superlinear convergence was proved under an additional simple condition. The primal-dual algorithms considered by Monteiro [42] are based on the centrality measure

$$d(X,S) \equiv \|X^{1/2}SX^{1/2} - \mu I\|_F = \left(\sum_{i=1}^{n}(\lambda_i(XS) - \mu)^2\right)^{1/2}, \tag{3.45}$$

where $(X,S) \in \mathcal{S}^n_+ \times \mathcal{S}^n_+$, $\mu = (X \bullet S)/n = (\sum_{i=1}^n \lambda_i(XS))/n$. Given $\gamma \in (0,1)$, we denote by $\mathcal{N}(\gamma)$ the following neighborhood of the central path:

$$\mathcal{N}(\gamma) = \{(X,y,S) \in \mathcal{F} : d(X,S) \le \gamma\mu, \ X \succ 0, \ S \succ 0, \ \mu = (X \bullet S)/n\}. \tag{3.46}$$

Monteiro's generalized predictor-corrector algorithm for semidefinite programming based on the MZ family of directions consists of a predictor step and a corrector step at each iteration. Starting from a strictly feasible pair (X^0, y^0, S^0) in $\mathcal{N}(\alpha)$, it generates a sequence of iterates $\{(X^k, y^k, S^k)\}$ in $\mathcal{N}(\alpha)$. An iteration of Monteiro's generalized predictor-corrector algorithm can be described as follows.

Predictor-Corrector Algorithm Given $(X^k, y^k, S^k) \in \mathcal{N}(\alpha)$, choose nonsingular $n \times n$ matrices P^k and $\bar{P}^k$.

• **Predictor Step.** Solve the system (3.44) with $(X, y, S) = (X^k, y^k, S^k)$, $\xi = 0$ and $P = P^k$. Denote the solution $(U, w, V) \in \mathcal{S}^n \times \mathrm{IR}^m \times \mathcal{S}^n$, and set

$$X^k(\theta) = X^k + \theta U, \quad y^k(\theta) = y^k + \theta w, \quad S^k(\theta) = S^k + \theta V. \tag{3.47}$$

Compute the steplength

$$\bar{\theta}_k = \max\left\{\tilde{\theta} \in [0, 1] : (X^k(\theta), S^k(\theta)) \in \mathcal{N}(\beta) \ \ \forall\, \theta \in [0, \tilde{\theta}]\right\}. \tag{3.48}$$

• **Corrector Step.** Solve the system (3.44) with $(X, y, S) = (X^k(\bar{\theta}_k), y^k(\bar{\theta}_k), S^k(\bar{\theta}_k))$, $\xi = 1$ and $P = \bar{P}^k$. Let $(\bar{U}, \bar{w}, \bar{V})$ be the solution, and set

$$X^{k+1} = X^k(\bar{\theta}_k) + \bar{U}, \quad y^{k+1} = y^k(\bar{\theta}_k) + \bar{w}, \quad S^{k+1} = S^k(\bar{\theta}_k) + \bar{V}. \tag{3.49}$$

End of iteration.

Using an elegant analysis, Monteiro [42] proved that the predictor-corrector algorithm defined above with properly chosen parameters α and β $(0 < \alpha < \beta < 1)$ is well defined and that it needs at most $O(\sqrt{n}\ln(\epsilon_0/\epsilon))$ iterations for producing a pair (X^k, y^k, S^k) such that $X^k \bullet S^k \le \epsilon$, where $\epsilon_0 = X^0 \bullet S^0$ is the initial gap. More precisely, Monteiro showed that

$$(X^k, y^k, S^k) \in \mathcal{N}(\alpha) \ \text{ and } \ (X^k(\bar{\theta}_k), y^k(\bar{\theta}_k), S^k(\bar{\theta}_k)) \in \mathcal{N}(\beta), \tag{3.50}$$

$$X^{k+1} \bullet S^{k+1} = (1 - 1/O(\sqrt{n}))X^k \bullet S^k, \tag{3.51}$$

for all $k \ge 0$.

In analyzing the local behavior of the predictor-corrector algorithm of Monteiro, we need the following technical result proved in [41, Lemma 2.6] and [42, Lemma 2.1(b)].

Lemma 3.5.1 *Suppose that $M \in \mathrm{IR}^{p\times p}$ is a nonsingular matrix and $E \in \mathrm{IR}^{p\times p}$ has at least one real eigenvalue. Then,*

$$\lambda_{\max}(E) \ \le \ \lambda_{\max}(H_M(E)), \tag{3.52}$$

$$\lambda_{\min}(E) \geq \lambda_{\min}(H_M(E)), \tag{3.53}$$

$$\|G\|_F \leq \|H_M(G)\|_F, \quad \forall G \in \mathcal{S}^p, \tag{3.54}$$

$$d(G,J) \leq \|H_M(GJ - \tfrac{G \bullet J}{n} I)\|_F, \quad \forall G, J \in \mathcal{S}^n_{++}. \tag{3.55}$$

The following lemma is part of Lemma 3.5 of Monteiro [42].

Lemma 3.5.2 *Let $W \in \mathbb{R}^{n\times n}$ be such that GWG^{-1} is skew-symmetric for some nonsingular $G \in \mathbb{R}^{n\times n}$. Then,*

$$\|W\|_F \leq \frac{\sqrt{2}}{2}\|W - W^T\|_F. \tag{3.56}$$

The following technical result will play an important role in our analysis.

Lemma 3.5.3 *Let $(X, y, S) \in \mathcal{N}(\gamma)$ for some $\gamma \in [0, \ \sqrt{2} - 1)$. Suppose that $(D_x, \Delta y, D_s) \in \mathcal{S}^{n\times n} \times \mathbb{R}^m \times \mathcal{S}^{n\times n}$ is a solution of the linear system:*

$$H_P(XD_s + D_xS) = H_P(K), \tag{3.57a}$$

$$A_i \bullet D_x = 0, \ \ i = 1, \ldots, m, \tag{3.57b}$$

$$\sum_{i=1}^m \Delta y_i A_i + D_s = 0, \tag{3.57c}$$

for some $K \in \mathbb{R}^{n\times n}$. Then we have

(i) $\|X^{1/2}D_sX^{1/2}\|_F^2 + \|\mu X^{-1/2}D_xX^{-1/2}\|_F^2 \leq \gamma_1^2\|X^{-1/2}KX^{1/2}\|_F^2$,

(ii) $\|X^{-1/2}(XD_s + D_xS - K)X^{1/2}\|_F \leq \gamma_2\|X^{-1/2}KX^{1/2}\|_F$,

where

$$\gamma_1 = \frac{\sqrt{2}+1}{(1 - (\sqrt{2}+1)\gamma)}, \quad \gamma_2 = \sqrt{2} + \sqrt{2}\,\gamma_1\gamma.$$

Proof. By denoting

$$X' = PXP^T, \ \ S' = (P^{-1})^T SP^{-1},$$

$$D_x' = PD_xP^T, \ \ D_s' = (P^{-1})^T D_s P^{-1}, \ \ K' = PKP^{-1},$$

$$\delta_x = \|\mu X^{-1/2}D_xX^{-1/2}\|_F, \quad \delta_s = \|X^{1/2}D_sX^{1/2}\|_F$$

we can write

$$X'D_s' + D_s'X' + D_x'S' + S'D_x' = K' + K'^T, \tag{3.58}$$

and

$$D'_x \bullet D'_s = 0.$$

It is easily seen that $\hat{Q} = (X')^{1/2}(P^{-1})^T X^{-1/2}$ is orthogonal. Then

$$(X')^{1/2} = \hat{Q} X^{1/2} P^T = P X^{1/2} \hat{Q}^T,$$

and

$$(X')^{-1/2} = (P^{-1})^T X^{-1/2} \hat{Q}^T = \hat{Q} X^{-1/2} P^{-1}.$$

Using the notation

$$B = X' D'_s + D'_x S' - K',$$

it follows that

$$\begin{aligned}(X')^{-1/2} B (X')^{1/2} &= (X')^{1/2} D'_s (X')^{1/2} + (X')^{-1/2} D'_x S' (X')^{1/2} - (X')^{-1/2} K' (X')^{1/2} \\ &= \hat{Q}[X^{1/2} D_s X^{1/2} + X^{-1/2} D_x S X^{1/2} - X^{-1/2} K X^{1/2}] \hat{Q}^T. \quad (3.59)\end{aligned}$$

Using (3.59) and Lemma 3.5.2 with $W = (X')^{-1/2} B (X')^{1/2}$ and $G = (X')^{1/2}$, we have

$$\begin{aligned}&\|(X')^{-1/2} B (X')^{1/2}\|_F \\ \le\ & \frac{\sqrt{2}}{2} \|(X')^{-1/2} B (X')^{1/2} - [(X')^{-1/2} B (X')^{1/2}]^T\|_F \\ \le\ & \frac{\sqrt{2}}{2} \|X^{-1/2} D_x S X^{1/2} - [X^{-1/2} D_x S X^{1/2}]^T\|_F \\ & + \frac{\sqrt{2}}{2} \|X^{-1/2} K X^{1/2} - [X^{-1/2} K X^{1/2}]^T\|_F \\ \le\ & \frac{\sqrt{2}}{2} \|X^{-1/2} D_x X^{-1/2} (X^{1/2} S X^{1/2} - \mu I) - [X^{-1/2} D_x X^{-1/2} (X^{1/2} S X^{1/2} - \mu I)]^T\|_F \\ & + \sqrt{2} \|X^{-1/2} K X^{1/2}\|_F \\ \le\ & \sqrt{2} \gamma \delta_x + \sqrt{2} \|X^{-1/2} K X^{1/2}\|_F. \quad (3.60)\end{aligned}$$

On the other hand, using (3.59) again, we obtain

$$\begin{aligned}&\|X^{-1/2} K X^{1/2}\|_F \\ &= \|\hat{Q} X^{-1/2} K X^{1/2} \hat{Q}^T\|_F \\ &\ge \|X^{1/2} D_s X^{1/2} + X^{-1/2} D_x S X^{1/2}\|_F - \|(X')^{-1/2} B (X')^{1/2}\|_F\end{aligned}$$

$$
\begin{aligned}
&= \|X^{1/2}D_sX^{1/2} + \mu X^{-1/2}D_xX^{-1/2} + X^{-1/2}D_xX^{-1/2}(X^{1/2}SX^{1/2} - \mu I)\|_F \\
&\quad - \|(X')^{-1/2}B(X')^{1/2}\|_F \\
&\geq \|X^{1/2}D_sX^{1/2} + \mu X^{-1/2}D_xX^{-1/2}\|_F \\
&\quad - \|X^{-1/2}D_xX^{-1/2}(X^{1/2}SX^{1/2} - \mu I)\|_F - \|(X')^{-1/2}B(X')^{1/2}\|_F \\
&\geq \left(\delta_x^2 + \delta_s^2\right)^{1/2} - \gamma\delta_x - \|(X')^{-1/2}B(X')^{1/2}\|_F \\
&\geq \left(\delta_x^2 + \delta_s^2\right)^{1/2} - (\sqrt{2}+1)\gamma\delta_x - \sqrt{2}\|X^{-1/2}KX^{1/2}\|_F \\
&\geq \left(\delta_x^2 + \delta_s^2\right)^{1/2}(1 - (\sqrt{2}+1)\gamma) - \sqrt{2}\|X^{-1/2}KX^{1/2}\|_F
\end{aligned}
$$

which implies (i). Then (ii) follows from (i), (3.60), and the fact that

$$\|X^{-1/2}(XD_s + D_xS - K)X^{1/2}\|_F = \|(X')^{-1/2}B(X')^{1/2}\|_F.$$

∎

It is interesting to note that the inequalities in the above lemma are independent of the scaling matrix P. In the next lemma we establish a lower bound for the step size $\bar{\theta}_k$, which together with Lemma 3.5.3 enables us to analyze the asymptotic behavior of the predictor-corrector algorithm.

Lemma 3.5.4 *Let (X^k, S^k), (U, V), and $\bar{\theta}_k$ be generated by the predictor-corrector algorithm. Then*

$$\bar{\theta}_k \geq \hat{\theta}_k,$$

where

$$
\begin{aligned}
\omega_k &= \frac{1}{\mu_k}\|(X^k)^{-1/2}(X^kV + US^k + X^kS^k)(X^k)^{1/2}\|_F, \\
\delta_k &= \frac{1}{\mu_k}\|(X^k)^{-1/2}UV(X^k)^{1/2}\|_F, \\
\hat{\theta}_k &= \frac{2}{\sqrt{\left(\frac{\omega_k+\beta-\alpha}{\beta-\alpha}\right)^2 + \frac{4\delta_k}{\beta-\alpha}} + \frac{\omega_k+\beta-\alpha}{\beta-\alpha}}.
\end{aligned}
$$

Proof. For simplicity, let us omit the index k. By (3.47), we have

$$X(\theta)S(\theta) = XS + \theta(XV + US) + \theta^2 UV,$$

which together with the linearity of $H_P(\cdot)$, the fact that $\mathrm{Tr}(H_P(M)) = \mathrm{Tr}(M)$ for $M \in \mathbb{R}^{n\times n}$, and (3.44a) with $\xi = 0$ imply that

$$\begin{aligned} X(\theta) \bullet S(\theta) &= \mathrm{Tr}(X(\theta)S(\theta)) \\ &= \mathrm{Tr}(H_P(X(\theta)S(\theta))) \\ &= \mathrm{Tr}((1-\theta)H_P(XS) + \theta^2 H_P(UV)) \\ &= (1-\theta)X \bullet S + \theta U \bullet V. \end{aligned}$$

Using the fact that $U \bullet V = 0$ we have

$$\mu(\theta) = (X(\theta) \bullet S(\theta))/n = (1-\theta)(X \bullet S)/n = (1-\theta)\mu.$$

Therefore,

$$X(\theta)S(\theta) - \mu(\theta)I = (1-\theta)(XS - \mu I) + \theta(XV + US + XS) + \theta^2 UV,$$

and

$$\begin{aligned} & \|H_{X^{-1/2}}(X(\theta)S(\theta) - \mu(\theta)I)\|_F \\ = \;& (1-\theta)\|X^{1/2}SX^{1/2} - \mu I\|_F + \theta\|X^{-1/2}(XV + US + XS)X^{1/2}\|_F + \\ & \theta^2\|X^{-1/2}UVX^{1/2}\|_F \\ \leq \;& (1-\theta)\alpha\mu + w\mu\theta + \delta\mu\theta^2 \\ \leq \;& (1-\theta)\mu\beta \quad \text{for} \quad 0 \leq \theta \leq \hat{\theta} \\ = \;& \beta\mu(\theta). \end{aligned}$$

Hence, we have $X(\theta) \succ 0$ and $S(\theta) \succ 0$ for all $\theta \in [0, \hat{\theta}]$. Otherwise, there exists a $\theta' \in [0, \hat{\theta}]$ such that $X(\theta')S(\theta')$ is singular, which means

$$\lambda_{\min}(X(\theta')S(\theta') - \mu(\theta')I) \leq -\mu(\theta'). \tag{3.61}$$

On the other hand, (3.53) with $M = X^{-1/2}$ and $E = X(\theta')S(\theta') - \mu(\theta')I$ implies that

$$\begin{aligned} \lambda_{\min}(X(\theta')S(\theta') - \mu(\theta')I) &\geq \lambda_{\min}(H_{X^{-1/2}}(X(\theta')S(\theta') - \mu(\theta')I) \\ &\geq -\|H_{X^{-1/2}}(X(\theta')S(\theta') - \mu(\theta')I)\|_F \\ &\geq -\beta\mu(\theta') > -\mu(\theta'), \end{aligned}$$

which contradicts (3.61). Using (3.55) with $G = X(\theta), J = S(\theta)$, and $M = X^{-1/2}$, we have

$$d(X(\theta), S(\theta)) \le \|H_{X^{-1/2}}(X(\theta)S(\theta) - \mu(\theta)I)\|_F \le \beta\mu(\theta) \quad \text{for} \quad \theta \in [0, \hat{\theta}].$$

Therefore, $(X(\theta), S(\theta)) \in \mathcal{N}(\beta)$ for $0 \le \theta \le \hat{\theta}$. The result follows from the definition of $\bar{\theta}$. ∎

Now let us investigate the asymptotic behavior of the predictor-corrector algorithm and obtain a sufficient condition for superlinear convergence.

Throughout this section Assumption 1 is used.

Let $Q = (q_1, \ldots, q_n)$ be an orthogonal matrix such that $q_1, \ldots, q_n$ are eigenvectors of X^* and S^*, and define

$$\mathbb{B} = \{i : q_i^T X^* q_i > 0\}, \qquad \mathbb{N} = \{i : q_i^T S^* q_i > 0\}.$$

It is easily seen that $\mathbb{B} \cup \mathbb{N} = \{1, 2, \ldots, n\}$. For simplicity, let us assume that

$$Q^T X^* Q = \begin{pmatrix} \Lambda_B & 0 \\ 0 & 0 \end{pmatrix}, \quad Q^T S^* Q = \begin{pmatrix} 0 & 0 \\ 0 & \Lambda_N \end{pmatrix},$$

where Λ_B and Λ_N are diagonal matrices. Here and in the sequel, if we write a matrix M in the block form

$$M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix},$$

then we assume that the dimensions of M_{11} and M_{22} are $|\mathbb{B}| \times |\mathbb{B}|$ and $|\mathbb{N}| \times |\mathbb{N}|$, respectively. In the next lemma we use the following notation:

$$Q^T X^{1/2} Q = (x_1, x_2, \ldots, x_n),$$

$$Q^T S^{1/2} Q = (s_1, s_2, \ldots, s_n),$$

$$Q^T X^{-1/2} Q = (\widehat{x}_1, \widehat{x}_2, \ldots, \widehat{x}_n),$$

$$Q^T S^{-1/2} Q = (\widehat{s}_1, \widehat{s}_2, \ldots, \widehat{s}_n).$$

Lemma 3.5.5 *(Potra-Sheng [54, Lemma 4.4]) Under Assumption 1 we have*

$$\|x_i\| = O(\sqrt{\mu}), \quad \|s_i\| = O(1), \ \forall i \in \mathbb{N};$$

$$\|x_i\| = O(1), \qquad \|s_i\| = O(\sqrt{\mu}), \ \forall i \in \mathbb{B};$$

$$\|\hat{x}_i\| = O(1), \qquad \|\hat{s}_i\| = O(1/\sqrt{\mu}),\ \forall i \in \mathbb{B};$$

$$\|\hat{x}_i\| = O(1/\sqrt{\mu}), \quad \|\hat{s}_i\| = O(1),\ \forall i \in \mathbb{N}.$$

Using Lemma 3.1.3, we can write

$$Q^T(X^k)^{1/2}Q = \begin{pmatrix} O(1) & O(\sqrt{\mu_k}) \\ O(\sqrt{\mu_k}) & O(\sqrt{\mu_k}) \end{pmatrix}, \quad Q^T(X^k)^{-1/2}Q = \begin{pmatrix} O(1) & O(1) \\ O(1) & O(1/\sqrt{\mu_k}) \end{pmatrix},$$
$$Q^T(S^k)^{1/2}Q = \begin{pmatrix} O(\sqrt{\mu_k}) & O(\sqrt{\mu_k}) \\ O(\sqrt{\mu_k}) & O(1) \end{pmatrix}, \quad Q^T(S^k)^{-1/2}Q = \begin{pmatrix} O(1/\sqrt{\mu_k}) & O(1) \\ O(1) & O(1) \end{pmatrix}.$$

Using the same techniques, we obtain a similar result for the predicted pair $(\bar{X}^k, \bar{S}^k)$.

Lemma 3.5.6 *Let $\bar{X}^k = X^k(\bar{\theta}_k)$, $\bar{S}^k = S^k(\bar{\theta}_k)$. If Assumption 1 is satisfied, then we have*

$$Q^T(\bar{X}^k)^{1/2}Q = \begin{pmatrix} O(1) & O(\sqrt{\mu_{k+1}}) \\ O(\sqrt{\mu_{k+1}}) & O(\sqrt{\mu_{k+1}}) \end{pmatrix}, \quad Q^T(\bar{X}^k)^{-1/2}Q = \begin{pmatrix} O(1) & O(1) \\ O(1) & O(1/\sqrt{\mu_{k+1}}) \end{pmatrix},$$
$$Q^T(\bar{S}^k)^{1/2}Q = \begin{pmatrix} O(\sqrt{\mu_{k+1}}) & O(\sqrt{\mu_{k+1}}) \\ O(\sqrt{\mu_{k+1}}) & O(1) \end{pmatrix}, \quad Q^T(\bar{S}^k)^{-1/2}Q = \begin{pmatrix} O(1/\sqrt{\mu_{k+1}}) & O(1) \\ O(1) & O(1) \end{pmatrix}.$$

Theorem 3.5.1 *Under Assumption 1, if $\eta_k \to 0$ as $k \to \infty$, then the predictor-corrector algorithm is superlinearly convergent. Moreover, if there exists a constant $\sigma > 0$ such that $\eta_k = O(\mu_k^\sigma)$, then the convergence has Q-order at least $1+\sigma$ in the sense that $\mu_{k+1} = O(\mu_k^{1+\sigma})$.*

Proof. For simplicity, let us omit the index k. It is easily seen that $(U + X - \breve{X}, w + y - \breve{y}, V + S - \breve{S})$ satisfies (3.57) with

$$K = (X - \breve{X})(S - \breve{S}). \tag{3.62}$$

Here we have used the relation $\breve{X}\breve{S} = \breve{S}\breve{X} = 0$. The matrix

$$\Delta = X^{-1/2}(X - \breve{X})(S - \breve{S})X^{1/2} \tag{3.63}$$

clearly satisfies the equation

$$\|\Delta\|_F = \|X^{-1/2}KX^{1/2}\|_F = \eta. \tag{3.64}$$

Denoting

$$\Delta_x = X^{-1/2}(U + X - \breve{X})X^{-1/2},$$

$$\Delta_s = X^{1/2}(V + S - \breve{S})X^{1/2},$$

and applying (i) of Lemma 3.5.3, we obtain

$$\mu\|\Delta_x\|_F \leq \gamma_1\|\Delta\|_F \leq \gamma_1\eta\mu,$$

which implies

$$\|\Delta_x\|_F \leq \gamma_1\eta. \tag{3.65}$$

Similarly,

$$\|\Delta_s\|_F \leq \gamma_1\|\Delta\|_F \leq \gamma_1\eta\mu. \tag{3.66}$$

By Lemma 3.5.6 and the fact that $(\breve{X}, \breve{y}, \breve{S}) \in \mathcal{M}$, we have

$$\begin{aligned}
\|X^{-1/2}(X-\breve{X})X^{-1/2}\|_F &= \|I - X^{-1/2}\breve{X}X^{-1/2}\|_F \\
&\leq \|I\|_F + \|X^{-1/2}\breve{X}X^{-1/2}\|_F \\
&= \sqrt{n} + \|Q^T X^{-1/2} Q Q^T \breve{X} Q Q^T X^{-1/2} Q\|_F \\
&= \sqrt{n} + \|(\widehat{x_1}, \ldots, \widehat{x_n}) Q^T \breve{X} Q (\widehat{x_1}, \ldots, \widehat{x_n})^T\|_F \\
&= \sqrt{n} + \left\|\textstyle\sum_{i,j\in\mathbb{B}} (q_i^T \breve{X} q_j)\widehat{x_i}\widehat{x_j}^T\right\|_F \\
&= O(1). \qquad (3.67)
\end{aligned}$$

In a similar manner we obtain

$$\|S^{-1/2}(S-\breve{S})S^{-1/2}\|_F = O(1). \tag{3.68}$$

Let us observe that

$$\begin{aligned}
X^{-1/2}UVX^{1/2} &= \left(X^{-1/2}UX^{-1/2}\right)\left(X^{1/2}VX^{1/2}\right) \\
&= \left(\Delta_x - X^{-1/2}(X-\breve{X})X^{-1/2}\right)\left(\Delta_s - X^{1/2}(S-\breve{S})X^{1/2}\right) \\
&= \Delta_x\Delta_s - \left(X^{-1/2}(X-\breve{X})X^{-1/2}\right)\Delta_s \\
&\quad -\Delta_x X^{1/2}S^{1/2}\left(S^{-1/2}(S-\breve{S})S^{-1/2}\right)S^{1/2}X^{1/2} + \Delta.
\end{aligned}$$

Then from (3.64), (3.65), (3.66), (3.67) and (3.68), we get

$$\begin{aligned}
\|X^{-1/2}UVX^{1/2}\|_F &\leq \|\Delta_x\|_F\|\Delta_s\|_F + \|X^{-1/2}(X-\breve{X})X^{-1/2}\|_F\|\Delta_s\|_F \\
&\quad +\|X^{1/2}S^{1/2}\|^2\|\Delta_x\|_F\|S^{-1/2}(S-\breve{S})S^{-1/2}\|_F + \|\Delta\|_F \\
&= O(\eta\mu).
\end{aligned}$$

Hence, $\delta_k = O(\eta_k) \to 0$. Applying (ii) of Lemma 3.5.3, we obtain

$$\begin{aligned} & \|X^{-1/2}[X(V+S-\breve{S})+(U+X-\breve{X})S-K]X^{1/2}\|_F \\ \le\ & \gamma_2\|\Delta\|_F \le \gamma_2\eta\mu. \end{aligned}$$

Noting that

$$X(V+S-\breve{S})+(U+X-\breve{X})S-K = XV+US+XS,$$

we deduce

$$\begin{aligned} \omega &= \frac{1}{\mu}\|X^{-1/2}(XV+US+XS)X^{1/2}\|_F \\ &\le \gamma_2\eta = O(\eta). \end{aligned}$$

Hence, $\omega_k \to 0$. Finally, if $\eta_k = O(\mu_k^\sigma)$ for some constant $\sigma > 0$, then we have $\delta_k = O(\mu_k^\sigma)$ and $\omega_k = O(\mu_k^\sigma)$. From Lemma 3.5.4, it follows that

$$\begin{aligned} 1-\bar{\theta} &\le 1 - \frac{2}{\sqrt{\left(\frac{\omega}{\beta-\alpha}+1\right)^2+\frac{4\delta}{\beta-\alpha}}+\frac{\omega}{\beta-\alpha}+1} \\ &= \frac{\omega/(\beta-\alpha)}{\sqrt{\left(\frac{\omega}{\beta-\alpha}+1\right)^2+\frac{4\delta}{\beta-\alpha}}+\frac{\omega}{\beta-\alpha}+1} + \frac{\sqrt{\left(\frac{\omega}{\beta-\alpha}+1\right)^2+\frac{4\delta}{\beta-\alpha}}-1}{\sqrt{\left(\frac{\omega}{\beta-\alpha}+1\right)^2+\frac{4\delta}{\beta-\alpha}}+\frac{\omega}{\beta-\alpha}+1} \\ &\le \frac{\omega}{\beta-\alpha} + \left(\frac{\omega}{\beta-\alpha}+1\right)^2 + \frac{4\delta}{\beta-\alpha} - 1 \\ &= \frac{3\omega}{\beta-\alpha} + \frac{\omega^2}{(\beta-\alpha)^2} + \frac{4\delta}{\beta-\alpha} \\ &= O(\eta) = O(\mu^\sigma). \end{aligned}$$

Therefore, $\mu_{k+1} = (1-\bar{\theta}_k)\mu_k = O(\mu_k^{1+\sigma})$. ∎

Lemma 3.5.1 was originally obtained by Potra and Sheng [56]. Based on Lemma 3.5.1, we establish the following generalization of the result of Potra and Sheng [56, Theorem 6.1].

Theorem 3.5.2 *Under Assumption 1, if $X^kS^k/\sqrt{X^k \bullet S^k} \to 0$ as $k \to \infty$, then the predictor-corrector algorithm is superlinearly convergent. Moreover, if $X^kS^k = O(\mu_k^{0.5+\sigma})$ for some constant $\sigma > 0$, then the convergence has Q-order at least $1+$*

$\min\{\sigma,\ 0.5\}$.

We will assume that Assumption 1 (strict complementarity) holds. Let $(X^*,$ $y^*,\ S^*)$ be a strictly complementary solution of (1.1) and (1.2). We will also assume the following nondegeneracy condition introduced by Kojima, Shida and Shindoh [33, 34]. First, let us define an affine space $\mathcal{G}_0$ by

$$\begin{aligned}\mathcal{G}_0 = \{(U,V) \in \mathcal{S}^n \times \mathcal{S}^n \quad : \quad & A_i \bullet U = 0,\ i = 1, \cdots, m,\\ & \sum_{i=1}^m w_i A_i + V = 0,\ w_i \in \mathbb{R}^m\}.\end{aligned}$$

Assumption 3. (Nondegeneracy) If $X^*V + US^* = 0$ and $(U,V) \in \mathcal{G}_0$, then $(U,V) = 0$.

As remarked in Section 5 of Kojima, Shida and Shindoh [34], under the strict complementarity assumption, the above nondegeneracy condition is equivalent to the combination of primal and dual nondegeneracy conditions given by Alizadeh, Haeberly and Overton [5].

Under Assumptions 1 and 2, the solution (X^*, S^*) is unique. Therefore the iteration sequence $\left\{(X^k,\ S^k)\right\}$ converges to (X^*, S^*) and so does the sequence of predicted pairs $\left\{(\bar{X}^k,\ \bar{S}^k)\right\}$.

Lemma 3.5.7 *(Kojima-Shida-Shindoh [34], Lemma 5.3) Assume that*

$$H_I(US^* + X^*V) = 0 \quad \textit{and} \quad (U,V) \in \mathcal{G}_0.$$

Then $(U,V) = (0,0)$.

Let R be a nonsingular matrix and

$$\tilde{A}_i = (R^{-1})^T A_i R^{-1},\ i = 1, \ldots, m, \tilde{C} = (R^{-1})^T C R^{-1}, \tilde{b} = b.$$

It is easily seen that the R-scaled SDP

$$\tilde{A}_i \bullet X = \tilde{b}_i,\ i = 1, \ldots, m, \tag{3.69a}$$

$$\sum_{i=1}^m y_i \tilde{A}_i + S = \tilde{C}, \tag{3.69b}$$

$$XS = 0, \quad X \succeq 0, \quad S \succeq 0, \tag{3.69c}$$

also satisfies the strict complementarity and nondegeneracy conditions. Its unique solution is $(RX^*R^T,\ y^*, (R^{-1})^T S^* R^{-1})$.

Using Lemma 3.5.7 and considering the new SDP (3.69), we can easily obtain the following lemma.

Lemma 3.5.8 *Assume that for some nonsingular matrix R,*

$$H_R(US^* + X^*V) = 0 \quad \textit{and} \ \ (U, V) \in \mathcal{G}_0.$$

Then $(U, V) = (0, 0)$.

In the next lemma $\mathrm{cond}_F(B) = \|B\|_F \|B^{-1}\|_F$ denotes the condition number of a matrix B.

Lemma 3.5.9 *If $cond_F(\bar{P}^k) = O(1)$, then*

$$\|(\bar{U}^k, \bar{V}^k)\|_F = O(\sqrt{\mu_{k+1}}). \tag{3.70}$$

Proof. Let $R^k = \bar{P}^k / \|\bar{P}^k\|_F$. Then $\|R^k\|_F = 1$ and $\|(R^k)^{-1}\|_F = \mathrm{cond}_F(\bar{P}^k) = O(1)$. At the corrector step of the algorithm, we have

$$H_{R^k}\left(\bar{U}^k \bar{S}^k + \bar{X}^k \bar{V}^k\right) = H_{R^k}\left(\mu_{k+1} I - \bar{X}^k \bar{S}^k\right), \tag{3.71}$$

where $\bar{X}^k = X^k(\bar{\theta}_k),\ \bar{S}^k = S^k(\bar{\theta}_k)$. From Lemma 3.5.6, it is easily seen that

$$\|\bar{X}^k \bar{S}^k + \bar{S}^k \bar{X}^k\|_F = O(\sqrt{\mu_{k+1}}).$$

Suppose (3.70) is not true, i.e., the sequence $\left\{(\bar{U}^k, \bar{V}^k)/\sqrt{\mu_{k+1}}\right\}$ is unbounded. Then we can choose a subsequence such that

$$\frac{(\bar{U}^k, \bar{V}^k)}{\sqrt{\mu_{k+1}}} \to \infty,$$
$$R^k \to R^*, \ \ (R^k)^{-1} \to (R^*)^{-1},$$

and

$$\frac{(\bar{U}^k, \bar{V}^k)}{\|(\bar{U}^k, \bar{V}^k)\|_F} \to (U', V') \neq 0.$$

Obviously, $(U',\ V') \in \mathcal{G}_0$. The fact that the matrices $A_i,\ i = 1, \cdots, m$ are linearly independent, together with $(U',\ V') \in \mathcal{G}_0$, implies that $(U', V') \neq 0$. Dividing both sides of (3.71) by $\|(\bar{U}^k, \bar{V}^k)\|_F$ and letting $k \to \infty$ along a subsequence, we obtain

$$H_{R^*}(U'S^* + X^*V') = 0,$$

which contradicts Lemma 3.5.8. ∎

Theorem 3.5.3 *Under the strict complementarity and nondegeneracy assumptions, if $cond_F(\bar{P}^k) = O(1)$, then the algorithm is superlinearly convergent with Q-order at least 1.5.*

Proof. At the predictor step, we have

$$(X^k,\ S^k) = (\bar{X}^{k-1},\ \bar{S}^{k-1}) + (\bar{U}^{k-1},\ \bar{V}^{k-1}).$$

Thus,

$$\begin{aligned} & H_{\bar{P}^{k-1}}(X^kS^k) \\ = \ & H_{\bar{P}^{k-1}}\left([\bar{X}^{k-1} + \bar{U}^{k-1}][\bar{S}^{k-1} + \bar{V}^{k-1}]\right) \\ = \ & \mu_k I + H_{\bar{P}^{k-1}}(\bar{U}^{k-1}\bar{V}^{k-1}). \end{aligned}$$

Then, by Lemma 3.5.9, we obtain

$$\|H_{\bar{P}^{k-1}}(X^kS^k)\|_F = O(\mu_k).$$

Note that

$$\begin{aligned} \|H_{\bar{P}^{k-1}}(X^kS^k)\|_F^2 &= \|\bar{P}^{k-1}X^kS^k(\bar{P}^{k-1})^{-1}\|_F^2 + \|(X^k)^{1/2}S^k(X^k)^{1/2}\|_F^2 \\ &\geq \|\bar{P}^{k-1}X^kS^k(\bar{P}^{k-1})^{-1}\|_F^2. \end{aligned} \tag{3.72}$$

Therefore,

$$\|X^kS^k\|_F = \|(\bar{P}^{k-1})^{-1}\bar{P}^{k-1}X^kS^k(\bar{P}^{k-1})^{-1}\bar{P}^{k-1}\|_F \leq O(1)\|H_{\bar{P}^{k-1}}(X^kS^k)\|_F = O(\mu_k), \tag{3.73}$$

which ends the proof by invoking Theorem 3.5.2. ∎

The above result says that the superlinear convergence of the predictor-corrector algorithm is independent of the choice of the scaling matrix P^k in the predictor step of the algorithm, while the scaling matrices used in the corrector step need to be "well-conditioned" for superlinear convergence. Clearly, the family of scaling matrices admissible in the corrector step for superlinear convergence includes the identity matrix defining the AHO as a special case. By imposing the same assumption on the scaling matrices used in the predictor step and a new strategy for the step size, we can improve the order of convergence stated in Theorem 3.5.3.

In order to achieve quadratic convergence we need to slightly modify the choice of the step size. Instead of $\bar{\theta}_k$ given by (3.48), we will use:

$$\bar{\theta}_k = \max\left\{\tilde{\theta} \in [0, \max\{.99, 1-\mu_k^2\}] : (X^k(\theta), S^k(\theta)) \in \mathcal{N}(\beta) \ \ \forall\, \theta \in [0, \tilde{\theta}]\right\}. \quad (3.74)$$

The predictor-corrector algorithm with this new strategy will be called the modified predictor-corrector algorithm. It is easily seen that the modified predictor-corrector algorithm still has polynomial complexity. In what follows we will show that it is also quadratically convergent.

Theorem 3.5.4 *Under the hypothesis of Theorem 3.5.3, if* $cond_F(P^k) = O(1)$*, then the modified predictor-corrector algorithm is quadratically convergent.*

Proof. From the proof of Theorem 3.5.3 (cf. (3.73)), we have

$$X^k S^k = O(\mu_k). \quad (3.75)$$

Using (3.75) and an argument similar to that employed in the proof of Lemma 3.5.9 we get

$$U^k = O(\mu_k), \quad V^k = O(\mu_k). \quad (3.76)$$

Then we can write

$$\begin{aligned} H_{P^k}(\bar{X}^k\bar{S}^k) &= H_{P^k}([X^k + \bar{\theta}_k U^k][S^k + \bar{\theta}_k V^k]) \\ &= (1-\bar{\theta}_k)H_{P^k}(X^kS^k) + \bar{\theta}_k^2 H_{P^k}(U^kV^k) \\ &= (1-\bar{\theta}_k)O(\mu_k) + O(\mu_k^2) = O(\omega_k\mu_k), \end{aligned}$$

where $\omega_k = \max\{\mu_k,\ 1-\bar{\theta}_k\}$. As in (3.72)-(3.73), we can prove that

$$\|\bar{X}^k\bar{S}^k\|_F = O(\omega_k\mu_k). \quad (3.77)$$

Using (3.77) and the same argument as in Lemma 3.5.9, we get

$$\bar{U}^k = O(\omega_k \mu_k) \quad \text{and} \qquad \bar{V}^k = O(\omega_k \mu_k). \tag{3.78}$$

Observing that

$$H_{\bar{P}^k}(X^{k+1}S^{k+1}) - \mu_{k+1} I = H_{\bar{P}^k}(\bar{U}^k \bar{V}^k) = O(\omega_k^2 \mu_k^2),$$

we have

$$\|H_{\bar{P}^k}(X^{k+1}S^{k+1}) - \mu_{k+1} I\|_F / \mu_{k+1} \;\; = \;\; \frac{O(\omega_k^2 \mu_k^2)}{\mu_{k+1}} = \frac{O(\omega_k^2 \mu_k)}{1-\bar{\theta}_k}.$$

Since $\text{cond}_F(P^k) \leq C_1$ and $\text{cond}_F(\bar{P}^k) \leq C_1$ for some constant C_1, we can write

$$\begin{aligned} \|H_{P^k}(X^k S^k) - \mu_k I\|_F / \mu_k &\leq C_1 \|X^k S^k - \mu_k I\|_F / \mu_k && (3.79) \\ &\leq C_1^2 \|\bar{P}^{k-1} X^k S^k (\bar{P}^{k-1})^{-1} - \mu_k I\|_F / \mu_k && (3.80) \\ &\leq C_2 \mu_{k-1} \max\left\{ \frac{\mu_{k-1}^2}{1-\bar{\theta}_{k-1}}, 1 - \bar{\theta}_{k-1} \right\}, && (3.81) \end{aligned}$$

where C_2 is a positive constant. Without loss of generality, we may assume

$$\mu_{k-1} \leq \min\{.1, \alpha / C_2\} \qquad \text{for} \;\; k \geq K,$$

which, together with (3.74) and (3.81), implies that

$$\bar{\theta}_{k-1} \leq 1 - \mu_{k-1}^2 \qquad \text{for} \;\; k \geq K,$$

and

$$\bar{\alpha}_k \equiv \|H_{P^k}(X^k S^k) - \mu_k I\|_F / \mu_k \leq C_2 \mu_{k-1} \leq \alpha \qquad \text{for} \;\; k \geq K.$$

Let

$$\theta_k' = \frac{2}{\sqrt{1 + 4\bar{\delta}_k / (\beta - \bar{\alpha}_k)} + 1}, \;\; \bar{\delta}_k = \|P^k U^k V^k (P^k)^{-1}\|_F / \mu_k.$$

Evidently, $\bar{\delta}_k = O(\mu_k)$. Then for all $\theta \in [0, \theta_k']$, we have

$$\begin{aligned} & \|(X^k(\theta))^{1/2} S^k (S^k(\theta))^{1/2} - \mu_k I\|_F \\ \leq \;\; & \|H_{P^k}(X^k(\theta) S^k(\theta)) - \mu_k I\|_F \\ \leq \;\; & [(1-\theta)\bar{\alpha}_k + \theta^2 \bar{\delta}_k] \mu_k \leq \beta (1-\theta) \mu_k. \end{aligned}$$

This means

$$\bar{\theta}_k \geq \theta_k' \qquad \text{for} \;\; k \geq K.$$

Therefore

$$1 - \bar{\theta}_k \leq 1 - \theta'_k = O(\bar{\delta}_k) = O(\mu_k),$$

and $\mu_{k+1} = (1 - \bar{\theta}_k)\mu_k = O(\mu_k^2)$. ∎

In this section we only consider the feasible version of the predictor-corrector method to keep the presentation simple. However, the analysis used here can be easily extended to the infeasible predictor-corrector algorithms based on the unified direction proposed by Monteiro and Zhang. Under the strict complementarity and nondegeneracy assumptions we have established the superlinear convergence with Q-order 1.5 of the "pure" predictor-corrector algorithm if the scaling matrices for the corrector step satisfy $cond_F(\bar{P}^k) = O(1)$. Whether superlinear convergence can be obtained under a weaker condition is an interesting topic for future research. Finally, we mention that quadratic convergence is established for the predictor-corrector algorithm with a slight modification of the step size selection. It would be interesting to find out whether quadratic convergence can be proved for the "original" predictor-corrector algorithm.

CHAPTER IV
HOMOGENEOUS ALGORITHMS

4.1 A homogeneous feasibility model

In the present chapter, we propose a simple homogeneous primal-dual feasibility model for the problem (1.3). This formulation is an extension of the homogeneous and self–dual feasibility model for linear programming developed by Goldman and Tucker [19, 63]. We mention that Ye, Todd and Mizuno [69] proposed a homogeneous algorithm for linear programming, and their results were simplified by Xu, Hung and Ye [68] who developed an efficient code by using Goldman and Tucker's homogeneous formulation for linear programming and an infeasible–interior–point method to solve it. Using a similar idea, we apply two interior–point algorithms to solve the homogeneous formulation of SDP from infeasible starting points. The algorithms do not need big M initialization. If the original SDP problem has a solution of trace norm ρ^*, then both algorithms find an ϵ-approximate solution in at most $O(\sqrt{n}\ln(\rho^*\epsilon_0/\epsilon))$ steps. We also provide a simple way to monitor possible infeasibility of the original SDP problem such that in at most $O(\sqrt{n}\ln(\rho\epsilon_0/\epsilon))$ steps we can either get a ϵ-approximate solution or determine that there is no solution with trace norm less than or equal to a given number $\rho > 0$. This is better that the complexity of the algorithm considered in [54] where we assumes that a solution exists and proved that the algorithm finds it in $O(n\ln(\epsilon_0/\epsilon)$ iterations if the starting point is large enough (big M initialization) or in $O(\sqrt{n}\ln(\epsilon_0/\epsilon))$ iterations if the starting point is almost feasible, where ϵ_0 denotes residual error at the starting point.

Our homogeneous model is given by the following homogeneous system:

$$A_i \bullet X = \tau b_i,\ i = 1, \ldots, m, \tag{4.1a}$$

$$\sum_{i=1}^{m} y_i A_i + S = \tau C, \tag{4.1b}$$

$$b^T y - C \bullet X = \kappa, \tag{4.1c}$$

$$X \succeq 0,\quad S \succeq 0,\ \tau \geq 0, \kappa \geq 0. \tag{4.1d}$$

It is easily seen that (4.1a), (4.1b) and (4.1c) imply

$$X \bullet S + \tau\kappa = 0. \tag{4.2}$$

The above formulation is an extension of the homogeneous and self–dual feasibility model for linear programming developed by Goldman and Tucker [19, 63]. It is also closely related to the shifted homogeneous primal–dual model considered by Nesterov [47, 48]. The following theorem is easily checked.

Theorem 4.1.1 *The SDP problem (1.3) has a solution if and only if the homogeneous system (4.1) has a solution*

$$(X^*, y^*, S^*, \tau^*, \kappa^*) \in \mathcal{S}_+^n \times \mathbb{R}^m \times \mathcal{S}_+^n \times \mathbb{R}_+ \times \mathbb{R}_+,$$

such that $\tau^* > 0,\ \ \kappa^* = 0.$

We denote the solution set of the problem (1.3) by $\mathcal{F}^*$. Also, let us define the solution set of (1.3) by $\mathcal{H}^*$. Notice that $\mathcal{H}^*$ is not empty since (4.1) has a trivial zero solution.

The residues of (4.1a)–(4.1c) are denoted by:

$$r_i = b_i\tau - A_i \bullet X,\ i = 1, \ldots, m, \tag{4.3a}$$

$$R_d = \tau C - \sum_{i=1}^{m} y_i A_i - S, \tag{4.3b}$$

$$r = \kappa - b^T y + C \bullet X. \tag{4.3c}$$

Throughout the chapter we will use the notation:

$$\mu = \frac{X \bullet S + \tau\kappa}{n+1}. \tag{4.4}$$

Let us define

$$\mathcal{H}_{++} = \mathcal{S}^n_{++} \times \mathbb{R}^m \times \mathcal{S}^n_{++} \times \mathbb{R}_{++} \times \mathbb{R}_{++}.$$

Then we can define the infeasible central path of the homogeneous problem (4.1) by

$$\begin{aligned} \mathcal{C} \;=\; & \{(X, y, S, \tau, \kappa) \in \mathcal{H}_{++} : \\ & XS = \mu I,\ \tau\kappa = \mu,\ r_i = (\mu/\mu_0) r_i^0,\ i = 1, \ldots, m, \\ & R_d = (\mu/\mu_0) R_d^0,\ r = (\mu/\mu_0) r_0\}. \end{aligned}$$

In our algorithms, we will use the following neighborhood of this central path:

$$\begin{aligned} \mathcal{N}(\gamma) &= \{(X, y, S, \tau, \kappa) \in \mathcal{H}_{++} : \left(\|X^{1/2} S X^{1/2} - \mu I\|_F^2 + (\tau\kappa - \mu)^2\right)^{1/2} \le \gamma\mu\} \\ &= \{(X, y, S, \tau, \kappa) \in \mathcal{H}_{++} : \left(\sum_{i=1}^m (\lambda_i(XS) - \tau)^2 + (\tau\kappa - \mu)^2\right)^{1/2} \le \gamma\tau\}, \end{aligned}$$

where γ is a constant such that $0 < \gamma < 1$.

4.2 Search directions

The search direction $(U, w, V, \Delta\tau, \Delta\kappa)$ of our algorithms is defined by the following linear system:

$$X^{-1/2}(XV + US)X^{1/2} + X^{1/2}(VX + SU)X^{-1/2}$$

$$= 2(\xi\mu I - X^{1/2} S X^{1/2}), \tag{4.5a}$$

$$\kappa\Delta\tau + \tau\Delta\kappa = \xi\mu - \tau\kappa, \tag{4.5b}$$

$$A_i \bullet U - b_i \Delta\tau = (1 - \xi) r_i, \quad i = 1, \ldots, m, \tag{4.5c}$$

$$\sum_{i=1}^m w_i A_i + V - \Delta\tau C = (1 - \xi) R_d, \tag{4.5d}$$

$$\Delta\kappa - b^T w + C \bullet U = (1 - \xi) r, \tag{4.5e}$$

where ξ is a parameter such that $\xi \in [0,\ 1]$. We note that equation (4.5a) can be written under the form

$$2XVX + USX + SXU = 2(\xi\mu X - XSX),$$

so that the solution of the linear system (4.5) does not require computation of square roots of matrices (see [41]).

Throughout the chapter we will use the following notation:

$$\widetilde{X} = \begin{pmatrix} X & 0 \\ 0 & \tau \end{pmatrix}, \; \widetilde{S} = \begin{pmatrix} S & 0 \\ 0 & \kappa \end{pmatrix}, \tag{4.6}$$

$$\widetilde{U} = \begin{pmatrix} U & 0 \\ 0 & \Delta\tau \end{pmatrix}, \; \widetilde{V} = \begin{pmatrix} V & 0 \\ 0 & \Delta\kappa \end{pmatrix}. \tag{4.7}$$

Then it is easily seen that

$$\mu = \frac{\widetilde{X} \bullet \widetilde{S}}{n+1}.$$

Lemma 4.2.1 *Suppose $(X, y, S, \tau, \kappa) \in \mathcal{H}_{++}$. Then the linear system (4.5) has a unique solution*

$$(U, w, V, \Delta\tau, \Delta\kappa) \in \mathcal{S}^n \times \mathrm{IR}^m \times \mathcal{S}^n \times \mathrm{IR} \times \mathrm{IR}.$$

Proof. Let us consider the following homogeneous linear system

$$X^{-1/2}(XV + US)X^{1/2} + X^{1/2}(VX + SU)X^{-1/2} = 0 \tag{4.8a}$$

$$\kappa\Delta\tau + \tau\Delta\kappa = 0, \tag{4.8b}$$

$$A_i \bullet U - b_i \Delta\tau = 0, \quad i = 1, \ldots, m, \tag{4.8c}$$

$$\sum_{i=1}^{m} w_i A_i + V - \Delta\tau C = 0, \tag{4.8d}$$

$$\Delta\kappa - b^T w + C \bullet U = 0. \tag{4.8e}$$

In view of (4.8a), (4.8c), (4.8d) and the proof of Lemma 2.1 of Monteiro [41], it follows that U and V are symmetric. From (4.8c)–(4.8e), we have $U \bullet V + \tau\kappa = 0$, i.e., $\widetilde{U} \bullet \widetilde{V} = 0$ with $\widetilde{U}, \widetilde{V}$ given by (4.6) and (4.7). According to (4.8a) and (4.8b), we have

$$\widetilde{X}^{-1/2}(\widetilde{X}\widetilde{V} + \widetilde{U}\widetilde{S})\widetilde{X}^{1/2} + \widetilde{X}^{1/2}(\widetilde{V}\widetilde{X} + \widetilde{S}\widetilde{U})\widetilde{X}^{-1/2} = 0,$$

or equivalently,

$$2\widetilde{X}^{1/2}\widetilde{V}\widetilde{X}^{1/2} + [\widetilde{X}^{-1/2}\widetilde{U}\widetilde{S}\widetilde{X}^{1/2} + \widetilde{X}^{1/2}\widetilde{S}\widetilde{U}\widetilde{X}^{-1/2}] = 0. \tag{4.9}$$

Therefore we obtain

$$[\widetilde{X}^{-1/2}\widetilde{U}\widetilde{X}^{-1/2}] \bullet [\widetilde{X}^{-1/2}\widetilde{U}\widetilde{S}\widetilde{X}^{1/2} + \widetilde{X}^{1/2}\widetilde{S}\widetilde{U}\widetilde{X}^{-1/2}] = 0,$$

which implies

$$[\widetilde{X}^{-1/2}\widetilde{U}\widetilde{X}^{-1/2}] \bullet [\widetilde{X}^{-1/2}\widetilde{U}\widetilde{S}\widetilde{X}^{1/2}] = 0.$$

Hence

$$\mathrm{Tr}(\widetilde{X}^{-1/2}\widetilde{U}\widetilde{S}\widetilde{U}\widetilde{X}^{-1/2}) = 0,$$

i.e.,

$$\mathrm{Tr}([\widetilde{X}^{-1/2}\widetilde{U}\widetilde{S}^{1/2}][\widetilde{X}^{-1/2}\widetilde{U}\widetilde{S}^{1/2}]^T) = 0.$$

It follows from the above relation that $\widetilde{X}^{-1/2}\widetilde{U}\widetilde{S}^{1/2} = 0$ which gives $\widetilde{U} = 0$. Moreover, from (4.9) we obtain $\widetilde{X}^{1/2}\widetilde{V}\widetilde{X}^{1/2} = 0$ which implies $\widetilde{V} = 0$. So, we deduce that $U = V = 0,\ \Delta\tau = \Delta\kappa = 0$. Consequently, the coefficient matrix of the linear system (4.5), which has $2n^2+m+2$ linear equations and $2n^2+m+2$ unknowns, is nonsingular, which implies the existence of a unique solution of (4.5). The symmetry of U and V determined by (4.5) can be deduced as in Lemma 2.1 of Monteiro [41]. ∎

Lemma 4.2.2 *If $(U, w, V, \Delta\tau, \Delta\kappa)$ is a solution of the linear system (4.5), then,*

$$U \bullet V + \Delta\tau\Delta\kappa = 0.$$

Proof. Let us write

$$\begin{aligned} & (U', w', V', \Delta\tau', \Delta\kappa') \\ = \ & ((U + (1-\xi)X,\ w + (1-\xi)y,\ V + (1-\xi)S,\ \tau + (1-\xi)\Delta\tau, \kappa + (1-\xi)\Delta\kappa). \end{aligned}$$

Then from (4.5c)–(4.5e), we have

$$\begin{aligned} & A_i \bullet U' - b_i\Delta\tau' = 0, \quad i = 1, \ldots, m, \\ & \sum_{i=1}^m w_i' A_i + V' - \Delta\tau' C = 0, \\ & \Delta\kappa' - b^T w' + C \bullet U' = 0, \end{aligned}$$

and therefore

$$U' \bullet V' + \Delta\tau'\Delta\kappa' = 0.$$

Hence, using the notation of (4.6) and (4.7), we get

$$(\widetilde{U} + (1-\xi)\widetilde{X}) \bullet (\widetilde{V} + (1-\xi)\widetilde{S})$$

$$= \quad (U + (1-\xi)X) \bullet (V + (1-\xi)S) + (\tau + (1-\xi)\Delta\tau)(\kappa + (1-\xi)\Delta\kappa)$$

$$= \quad 0. \tag{4.10}$$

In view of (4.5a) and (4.5b), we obtain

$$\frac{1}{2}[\widetilde{X}^{-1/2}(\widetilde{X}\widetilde{V} + \widetilde{U}\widetilde{S})\widetilde{X}^{1/2} + \widetilde{X}^{1/2}(\widetilde{V}\widetilde{X} + \widetilde{S}\widetilde{U})\widetilde{X}^{-1/2}] = (\xi\mu I - \widetilde{X}^{1/2}\widetilde{S}\widetilde{X}^{1/2}). \tag{4.11}$$

By taking the trace of both sides of (4.11), we have

$$\widetilde{X} \bullet \widetilde{V} + \widetilde{U} \bullet \widetilde{S} = (\xi - 1)\widetilde{X} \bullet \widetilde{S}. \tag{4.12}$$

Then by expanding (4.10) and using (4.12) we get $\widetilde{U} \bullet \widetilde{V} = 0$, i.e.,

$$U \bullet V + \Delta\tau\Delta\kappa = 0.$$

∎

Lemma 4.2.3 *Let $(X, y, S, \tau, \kappa) \in \mathcal{N}(\gamma)$ for some $\gamma \in [0, 1)$. Suppose that $(U, w, V, \Delta\tau, \Delta\kappa)$ is a solution of the linear system (4.5). Then,*

$$\|\widetilde{X}^{-1/2}\widetilde{U}\widetilde{V}\widetilde{X}^{1/2}\| \leq \frac{\|\xi\mu I - \widetilde{X}^{1/2}\widetilde{S}\widetilde{X}^{1/2}\|_F^2}{2(1-\gamma)^2\mu}.$$

Proof. Using the notation introduced in (4.6), (4.7) and Lemma 4.2.2, we have

$$\widetilde{X}^{1/2}\widetilde{V}\widetilde{X}^{1/2} + \widetilde{X}^{-1/2}\widetilde{U}\widetilde{S}\widetilde{X}^{1/2} + \widetilde{X}^{1/2}\widetilde{S}\widetilde{U}\widetilde{X}^{-1/2} = 2(\xi\mu I - \widetilde{X}^{1/2}\widetilde{S}\widetilde{X}^{1/2}), \tag{4.13}$$

and

$$\widetilde{U} \bullet \widetilde{V} = 0. \tag{4.14}$$

Since $(X, y, S, \tau, \kappa) \in \mathcal{N}(\gamma)$, we obtain

$$\|\widetilde{X}^{1/2}\widetilde{S}\widetilde{X}^{1/2} - \mu I\|_F \leq \gamma\mu. \tag{4.15}$$

Then the desired inequality follows from (4.13), (4.14), (4.15) and a manipulation similar to that used in the proof of Lemma 4.4 of Monteiro [41]. ∎

4.3 A generic homogeneous algorithm

Generic Homogeneous Algorithm

Let $(X^0, y^0, S^0, \tau_0, \kappa_0) = (I, 0, I, 1, 1)$.

Repeat until a stopping criterion is satisfied:

- Choose $\xi \in [0,1]$ and compute the solution $(U, w, V, \Delta\tau, \Delta\kappa)$ of the linear system (4.5).

- Compute steplength $\bar{\theta}$ such that

$$X + \bar{\theta}U \succ 0, \;\; S + \bar{\theta}V \succ 0, \;\; \tau + \bar{\theta}\Delta\tau > 0, \;\; \kappa + \bar{\theta}\Delta\kappa > 0.$$

- Update the iterates

$$(X^+, y^+, S^+, \tau_+, \kappa_+) = (X, y, S, \tau, \kappa) + \bar{\theta}(U, w, V, \Delta\tau, \Delta\kappa).$$

We will define a stopping criterion later in this section.

Theorem 4.3.1 *The iterates generated by the Generic Homogeneous Algorithm satisfy:*

$$\mu_+ = \frac{X^+ \bullet S^+ + \tau_+\kappa_+}{n+1} = (1-(1-\xi)\bar{\theta})\mu, \tag{4.16a}$$

$$r_i^+ = (1-(1-\xi)\bar{\theta})r_i, \;\; i = 1, \ldots, m, \tag{4.16b}$$

$$R_d = (1-(1-\xi)\bar{\theta})R_d, \;\; r^+ = (1-(1-\xi)\bar{\theta})r. \tag{4.16c}$$

Proof. By the definition of the Generic Homogeneous Algorithm, we have

$$\begin{aligned} X^+ \bullet S^+ &= (X + \bar{\theta}U) \bullet (S + \bar{\theta}V) \\ &= X \bullet S + \bar{\theta}(X \bullet V + U \bullet S) + U \bullet V \\ &= X \bullet S + \bar{\theta}(n\xi\mu - X \bullet S) + U \bullet V, \end{aligned}$$

and

$$\begin{aligned} \tau_+\kappa_+ &= (\tau + \bar{\theta}\Delta\tau)(\kappa + \bar{\theta}\Delta\kappa) \\ &= \tau\kappa + \bar{\theta}(\tau\Delta\kappa + \kappa\Delta\tau) + \Delta\tau\Delta\kappa \\ &= \tau\kappa + \bar{\theta}(\xi\mu - \tau\kappa) + \Delta\tau\Delta\kappa. \end{aligned}$$

Therefore, from Lemma 4.2.2, we obtain

$$X^+ \bullet S^+ + \tau_+\kappa_+ = (1-(1-\xi)\bar{\theta})(n+1)\mu,$$

which implies (4.16a). Finally, (4.16b) and (4.16c) can be easily deduced from (4.5c)–(4.5e). ∎

By using the above theorem and the fact that $\mu_0 = 1$, we deduce the following corollary.

Corollary 4.3.1 *Let $(X^k, y^k, S^k, \tau_k, \kappa_k)$ be generated by Generic Homogeneous Algorithm. Then,*

$$r_i^k = \mu_k r_i^0, \; i = 1, \dots, m,$$

$$R_d^k = \mu_k R_d^0, \quad r_k = \mu_k r_0.$$

Let us define a linear manifold

$$\begin{aligned} \mathcal{H}_0 = \quad & \{(X', y', S', \tau', \kappa') \in \mathcal{S}^n \times \mathrm{IR}^m \times \mathcal{S}^n \times \mathrm{IR} \times \mathrm{IR} : \\ & A_i \bullet X' - \tau' b_i = 0, \; i = 1, \dots, m, \\ & \sum_{i=1}^m y_i' A_i + S' - \tau' C = 0, \\ & \kappa' - b^T y' + C \bullet X' = 0\} \end{aligned}$$

Then it is easily seen that if $(X', y', S', \tau', \kappa') \in \mathcal{H}_0$, then

$$X' \bullet S' + \tau'\kappa' = 0.$$

The following lemma shows that the iterates $(X^k, y^k, S^k, \tau_k, \kappa_k)$ are bounded.

Lemma 4.3.1 *If $(X^k, y^k, S^k, \tau_k, \kappa_k)$ are generated by the Generic Homogeneous Algorithm, then*

$$\mathrm{Tr}(X^k + S^k) + \tau_k + \kappa_k = (n+1)(1+\mu). \tag{4.17}$$

Proof. For simplicity, we omit the index k. From Corollary 4.3.1, we have

$$(X - \mu X^0, \; y - \mu y^0, \; S - \mu S^0, \; \tau - \mu\tau_0, \; \kappa - \mu\kappa_0) \in \mathcal{H}_0.$$

Therefore,

$$(S - \mu S^0) \bullet (X - \mu X^0) + (\tau - \mu\tau_0)(\kappa - \mu\kappa_0) = 0. \tag{4.18}$$

The desired result follows by expanding (4.18) and using the fact that $X^0 = S^0 = I$ and $\tau_0 = \kappa_0 = 1$. ∎

Theorem 4.3.2 *Let $(X^k, y^k, S^k, \tau_k, \kappa_k)$ be generated by the Generic Homogeneous Algorithm and let $\rho > 0$ be a given number. Suppose that the following two conditions are satisfied:*

(a) there exists a solution (X^*, y^*, S^*) of the original problem (1.3) such that $\mathrm{Tr}(X^* + S^*) \le \rho$;

(b) there exists a constant $\omega \in (0,1)$ such that $\tau_k \kappa_k \ge \omega \mu_k$, $\forall k$.

Then,

$$\tau_k \ge \frac{\omega}{\rho+1}, \quad \forall\, k.$$

Proof. Let $\tau_* = 1$, $\kappa_* = 0$. Obviously, $(X^*, y^*, S^*, \tau_*, \kappa_*)$ is a solution of (4.1). From the proof of Lemma 4.3.1, for any real number η, we have

$(X - \mu X^0 + \eta X^*,\ y - \mu y^0 + \eta y^*,\ S - \mu S^0 + \eta S^*,\ \tau - \mu\tau_0 + \eta\tau_*,\ \kappa - \mu\kappa_0 + \eta\kappa_*) \in \mathcal{H}_0$.

Therefore, we get

$$(S - \mu S^0 + \eta S^*) \bullet (X - \mu X^0 + \eta X^*) + (\tau - \mu\tau_0 + \eta\tau_*)(\kappa - \mu\kappa_0 + \eta\kappa_*) = 0.$$

By expanding the above equality and noting that η is an arbitrary real number and that $X^* \bullet S^* + \tau_*\kappa_* = 0$, we obtain

$S \bullet X^* + S^* \bullet X + \tau\kappa_* + \tau_*\kappa = \mu(S^0 \bullet X^* + S^* \bullet X^0 + \tau_0\kappa_* + \tau_*\kappa_0) = \mu[\mathrm{Tr}(X^* + S^*) + 1]$.

Then, from condition (a) and $\tau_* = 1$, we deduce that

$$\kappa \le \mu[\mathrm{Tr}(X^* + S^*) + 1] \le \mu(\rho + 1). \tag{4.19}$$

Hence, by (4.19) and condition (b), we have

$$\tau \ge \frac{\omega\mu}{\kappa} \ge \frac{\omega}{\rho+1}.$$

∎

Corollary 4.3.2 *Under condition (b) of Theorem 4.3.2 suppose that* $\mu_k \to 0$ *as* $k \to \infty$*. Then*

(i) the primal–dual problem (1.3) has a solution if and only if $\kappa_k \to 0$ *as* $k \to \infty$ *and there exists a constant* $\tau_* > 0$ *such that* $\tau_k \geq \tau_*$, $\forall k$;

(ii)(1.3) has no solution if and only if $\tau_k \to 0$ *as* $k \to \infty$*.*

Proof. Part (i) is an immediate consequence of Theorem 4.3.2. From Lemma 4.3.1, the sequence $\{\tau_k\}$ is bounded. Let us observe that if (1.3) has no solution then $\{\tau_k\}$ does not have a positive accumulation point (otherwise, there exists a subsequence $\{(X^k/\tau_k, y^k/\tau_k, S^k/\tau_k)\}$ converging to a solution of (1.3)), and therefore $\tau_k \to 0$. Conversely, if $\tau_k \to 0$, then by Theorem 4.3.2 (1.3) has no solution. ∎

Let us define the stopping criterion of the Generic Homogeneous Algorithm:

$$E_k \equiv \max\{\|R_d^k\|/\tau_k,\ |r_i^k|/\tau_k, i = 1, \ldots, m,\ (X^k \bullet S^k + \tau_k\kappa_k)/\tau_k^2\} \leq \epsilon, \tag{4.20}$$

where $\epsilon > 0$ is the tolerance. Let us define the set of ϵ-approximate solutions of (1.3) by

$$\begin{aligned} \mathcal{F}_\epsilon = \{ \quad & (X, y, S) \in \mathcal{S}_+^n \times \mathbb{R}^m \times \mathcal{S}_+^n : \\ & X \bullet S \leq \epsilon, \quad |b_i - A \bullet X| \leq \epsilon,\ i = 1, \ldots, m, \quad \|C - \textstyle\sum_{i=1}^m y_i A_i - S\| \leq \epsilon\}. \end{aligned}$$

Obviously, if (4.20) is satisfied, then $(X^k/\tau_k, y^k/\tau_k, S^k/\tau_k)$ is an ϵ-approximate solution of the original problem (1.3). Let us observe that if our algorithm starts at a point that is a multiple of $(I, 0, I, 1, 1)$, say $(\nu I, 0, \nu, \nu, \nu)$, then we obtain an iteration sequence

$$(\check{X}^k, \check{y}^k, \check{S}^k, \check{\tau}_k, \check{\kappa}_k) = (\nu X^k, \nu y^k, \nu S^k, \nu\tau_k, \nu\kappa_k).$$

However, we note that

$$(\check{X}^k/\check{\tau}_k, \check{y}^k/\check{\tau}_k, \check{S}^k/\check{\tau}_k) = (X^k/\tau_k, y^k/\tau_k, S^k/\tau_k).$$

Therefore, we see that the sequence $(X^k/\tau_k, y^k/\tau_k, S^k/\tau_k)$ does not depend on the magnitude of the starting point. This explains why we do not need a big M initialization. If we define the residual error at the point (X, y, S, τ, κ) by

$$\mathrm{res}((X, y, S, \tau, \kappa)) = \max\{X \bullet S + \tau\kappa,\ \|R_d\|_F,\ |r_i|, i = 1, \ldots, m\}, \tag{4.21}$$

then for our starting point we have

$$\begin{aligned}
\epsilon_0 &\equiv \mathrm{res}((I, 0, I, 1, 1)) \\
&= \max\{n+1, \|C - I\|_F, |b_i - \mathrm{Tr}(A_i)|, i = 1, \ldots, m\},
\end{aligned}$$

while for a $(\rho + 1)$ multiple of our starting point we have

$$\begin{aligned}
\chi(\rho) &\equiv \mathrm{res}((\rho+1)(I, 0, I, 1, 1)) \qquad (4.22)\\
&= \max\{(\rho+1)^2(n+1), (\rho+1)\|C - I\|_F, (\rho+1)|b_i - \mathrm{Tr}(A_i)|, i = 1, \ldots, m\} \\
&\leq \epsilon_0(\rho+1)^2.
\end{aligned}$$

Using the above introduced quantity we obtain the following result.

Lemma 4.3.2 *If $\tau_k \geq \omega/(\rho+1)$, then then the following estimate holds:*

$$E_k \leq \mu_k \chi(\rho)/\omega^2,$$

where $\chi(\rho)$ is the residual error defined by (4.22).

Proof. From (4.20) and Corollary 4.3.1, we get

$$\begin{aligned}
E_k &= \max\{\mu_k\|R_d^0\|_F/\tau_k,\ \mu_k|r_i^0|/\tau_k, i = 1, \ldots, m,\ \mu_k(n+1)/\tau_k^2\} \\
&= \mu_k \max\{\|R_d^0\|_F/\tau_k,\ |r_i^0|/\tau_k, i = 1, \ldots, m,\ (n+1)/\tau_k^2\} \\
&\leq \mu_k \max\{\frac{\|R_d^0\|_F}{\omega/(\rho+1)},\ \frac{|r_i^0|}{\omega/(\rho+1)}, i = 1, \ldots, m,\ \frac{n+1}{(\omega/(\rho+1))^2}\} \\
&\leq (\mu_k/\omega^2)\max\{(\rho+1)\|R_d^0\|_F,\ (\rho+1)|r_i^0|, i = 1, \ldots, m,\ (\rho+1)^2(n+1)\} \\
&= \mu_k\chi(\rho)/\omega^2.
\end{aligned}$$

∎

From Theorem 4.3.2 and Lemma 4.3.2, we see that the quantity τ_k is a useful certificate for monitoring possible infeasibility of the problem (1.3).

Corollary 4.3.3 *Under condition (b) of Theorem 4.3.2 if*

$$\tau_k < \omega/(\rho+1), \text{ for some } k, \tag{4.23}$$

then the primal–dual problem (1.3) has no solution (X^*, y^*, S^*) *such that* $\mathrm{Tr}(X^* + S^*) \leq \rho$.

In the next two sections, we will describe two examples of the Generic Homogeneous Algorithm, which achieves $O(\sqrt{n}\ln(\rho\epsilon_0/\epsilon))$–iteration complexity.

4.4 Two specific algorithms

In this section, we apply the recent infeasible–interior–point predictor–corrector algorithm of Potra and Sheng [54] to the homogeneous system (4.1) and show that it has $O(\sqrt{n}\ln(\rho\epsilon_0/\epsilon))$-iteration complexity. Let α, β be two positive constants satisfying the inequalities

$$\frac{\beta^2}{2(1-\beta)^2} \leq \alpha < \beta < \frac{\beta}{1-\beta} < 1. \tag{4.24}$$

For example, $\alpha = 0.25,\ \beta = 0.41$ verify (4.24).

Algorithm 4.4.1 $(X, y, S, \tau, \kappa) \leftarrow (I, 0, I, 1, 1)$;

Repeat until stopping criterion (4.20) is satisfied or τ *is sufficiently small:*

(Predictor step)

Solve the linear system (4.5) with $\xi = 0$;

Compute

$$\begin{aligned}&\bar\theta = \max\{\tilde\theta \in [0,1] :\\ &\left(\sum_{i=1}^{n}(\lambda_i(X(\theta)S(\theta)) - (1-\theta)\mu)^2 + (\tau(\theta)\kappa(\theta) - \mu)^2\right)^{1/2} \leq \beta(1-\theta)\mu,\\ &\forall\, \theta \in [0, \tilde\theta]\},\end{aligned}$$

where

$$(X(\theta), S(\theta), \tau(\theta), \kappa(\theta)) = (X, S, \tau, \kappa) + \theta(U, V, \Delta\tau, \Delta\kappa);$$

$(X, y, S, \tau, \kappa) \leftarrow X, y, S, \tau, \kappa) + \bar\theta(U, w, V, \Delta\tau, \Delta\kappa)$;

(Corrector step)

Solve the linear system (4.5) with $\xi = 1$;

$(X, y, S, \tau, \kappa) \leftarrow X, y, S, \tau, \kappa) + (U, w, V, \Delta\tau, \Delta\kappa)$.

Using a proof similar to that of Lemma 2.5 of Potra and Sheng [54], we get the following lower bound for the steplength $\bar{\theta}$:

$$\bar{\theta} \geq \frac{2}{\sqrt{1 + 4\delta/(\beta - \alpha)} + 1}, \tag{4.25}$$

where

$$\delta \equiv \frac{1}{\mu} \| \widetilde{X}^{-1/2} \widetilde{U} \widetilde{V} \widetilde{X}^{1/2} \|_F. \tag{4.26}$$

Then, in view of Lemma 4.2.3, we obtain

$$\begin{aligned} \delta &\leq \frac{\|\widetilde{X}^{1/2} \widetilde{S} \widetilde{X}^{1/2}\|_F^2}{2(1-\alpha)^2 \mu^2} \\ &= \frac{\|\mu I\|_F^2 + \|\mu I - \widetilde{X}^{1/2} \widetilde{S} \widetilde{X}^{1/2}\|_F^2}{2(1-\alpha)^2 \mu^2} \\ &\leq \frac{(n+1)\mu^2 + \alpha^2 \mu^2}{2(1-\alpha)^2 \mu^2} = \frac{n + 1 + \alpha^2}{2(1-\alpha)^2} = O(n), \end{aligned}$$

which implies

$$1 - \bar{\theta} \leq 1 - \frac{1}{O(\sqrt{n})}.$$

Here we have used the relation

$$(\mu I) \bullet [\mu I - \widetilde{X}^{1/2} \widetilde{S} \widetilde{X}^{1/2}] = 0.$$

Then, from Theorem 4.3.1, we have

$$\mu_{k+1} = (1 - \bar{\theta}_k)\mu_k \leq (1 - \frac{1}{O(\sqrt{n})})\mu_k. \tag{4.27}$$

Also, we can prove that

$$(X^k, y^k, S^k, \tau_k, \kappa_k) \in \mathcal{N}(\alpha), \ \forall k.$$

Hence, condition (b) of Theorem 4.3.2 is satisfied with $\omega = 1 - \alpha$. Therefore, in view of (4.27), Theorem 4.3.2 and Lemma 4.3.2, we see that in at most $O(\sqrt{n} \ln(\rho\epsilon_0/\epsilon))$ iterations, either the inequality $\tau_k \geq (1-\alpha)/\rho$ holds all the time or it is violated for some $k \leq O(\sqrt{n} \ln(\rho\epsilon_0/\epsilon))$. In the former case, we get an ϵ-approximate solution of (1.3) while in the later case (1.3) has no solution (X^*, y^*, S^*) such that $\mathrm{Tr}(X^* + S^*) \leq$

ρ. To summarize, we obtain the following polynomial complexity result.

Theorem 4.4.1

(i) If $\mathcal{F}^*$ is not empty, then Algorithm 4.4.1 terminates with an ϵ-approximate solution $(X^k/\tau_k, y^k/\tau_k, S^k/\tau_k) \in \mathcal{F}_\epsilon$ in a finite number of steps $k = K_\epsilon < \infty$.

(ii) If $\rho^* = \mathrm{Tr}(X^* + S^*)$ for some $(X^*, y^*, S^*) \in \mathcal{F}^*$, then $K_\epsilon = O(\sqrt{n}\ln(\rho^*\epsilon_0/\epsilon))$.

(iii) For any choice of $\rho > 0$ there is an index $k = \widehat{K}_\epsilon = O(\sqrt{n}\ln(\rho\epsilon_0/\epsilon))$ such that either

(iiia) $(X^k/\tau_k, y^k/\tau_k, S^k/\tau_k) \in \mathcal{F}_\epsilon$,

or,

(iiib) $\tau_k < (1-\alpha)/(\rho+1)$,

and in the latter case there is no solution $(X^*, y^*, S^*) \in \mathcal{F}^*$ with $\rho \geq \mathrm{Tr}(X^* + S^*)$.

Next, let us extend a recent short–step feasible path–following algorithm by Kojima, Shindoh and Hara [35] and simplified by Monteiro [41] for solving the homogeneous system from infeasible starting points.

Algorithm 4.4.2 $\quad (X, y, S, \tau, \kappa) \leftarrow (I, 0, I, 1, 1)$;

Repeat until the stopping criterion (4.20) is satisfied or τ is sufficiently small:

Solve the linear system (4.5) with $\xi = 1 - 0.3/\sqrt{n+1}$;

$(X, y, S, \tau, \kappa) \leftarrow X, y, S, \tau, \kappa) + (U, w, V, \Delta\tau, \Delta\kappa)$.

By an analogous analysis to that used in Theorem 4.1 of Monteiro [41], we can prove that $(X^k, y^k, S^k, \tau_k, \kappa_k) \in \mathcal{N}(0.3),\ \forall k$. From Theorem 4.3.1, we get

$$\mu_{k+1} = (1 - 0.3/\sqrt{n+1})\mu_k.$$

Obviously, condition (b) of Theorem 4.3.2 is satisfied with $\omega = 0.7$. Hence, similar to Theorem 4.4.1, we obtain the following complexity result for Algorithm 4.4.2.

Theorem 4.4.2

(i) If $\mathcal{F}^*$ is not empty, then Algorithm 4.4.2 terminates with an ϵ-approximate solution $(X^k/\tau_k, y^k/\tau_k, S^k/\tau_k) \in \mathcal{F}_\epsilon$ in a finite number of steps $k = K_\epsilon < \infty$.

(ii) If $\rho^* = \mathrm{Tr}(X^* + S^*)$ for some $(X^*, y^*, S^*) \in \mathcal{F}^*$, then $K_\epsilon = O(\sqrt{n}\ln(\rho^*\epsilon_0/\epsilon))$.

(iii) For any choice of $\rho > 0$ there is an index $k = \widehat{K}_\epsilon = O(\sqrt{n}\ln(\rho\epsilon_0/\epsilon))$ such that either

(iiia) $(X^k/\tau_k, y^k/\tau_k, S^k/\tau_k) \in \mathcal{F}_\epsilon$,

or,

(iiib) $\tau_k < 0.7/(\rho + 1)$,

and in the latter case there is no solution $(X^*, y^*, S^*) \in \mathcal{F}^*$ with $\rho \geq \mathrm{Tr}(X^* + S^*)$.

Further remarks

For simplicity of analysis, we have chosen the starting point $(I, 0, I, 1, 1)$. Indeed, we can use any starting point $(X^0, y^0, S^0, \tau_0, \kappa_0) \in \mathcal{H}_{++}$ near the infeasible central path because we are actually interested in the sequence $(X^k/\tau_k, y^k/\tau_k, S^k/\tau_k)$ which does not depend on the magnitude of the starting point.

We have only used the HRVW/KSH/M direction in the exposition of the homogeneous algorithms. Other directions can be used in a similar way.

As a final comment we note that we can also extend the long–step algorithm for linear programming by Kojima Mizuno and Yoshise [31] to the homogeneous system (4.1). By using the proof techniques and the results of Monteiro [41] we can prove that the iteration complexity of this extended algorithm is $O(n^{1.5}\ln(\rho\epsilon_0/\epsilon))$.

CHAPTER V
NUMERICAL IMPLEMENTATION

5.1 Infeasible predictor-corrector algorithms

As described in many other papers (cf. [6, 60, 70]), the linear system (2.17) for the Monteiro-Zhang unified search direction can be written in the following matrix form.

$$\begin{pmatrix} 0 & A & 0 \\ A^T & 0 & I \\ 0 & E & F \end{pmatrix} \begin{pmatrix} \Delta y \\ \mathbf{vec}(\Delta X) \\ \mathbf{vec}(\Delta S) \end{pmatrix} = \begin{pmatrix} r_p \\ \mathbf{vec}(R_d) \\ \mathbf{vec}(R_c) \end{pmatrix}, \tag{5.1}$$

where

$$A^T = [\mathbf{vec}(A_1), \mathbf{vec}(A_2), ..., \mathbf{vec}(A_m)],$$

$$r_p^T = [r_1, r_2, ..., r_m],$$

$E,\ F \in \mathbb{R}^{n^2 \times n^2},\ R_c \in \mathbb{R}^{n \times n}$ are such that (4.5a) has an equivalent vector form:

$$E\mathbf{vec}(\Delta X) + F\mathbf{vec}(\Delta S) = \mathbf{vec}(R_c).$$

For any $n \times n$ matrix M, $\mathbf{vec}(M)$ denotes the operation of stacking the columns of M on top of another so that $\mathbf{vec}(M) = (m_{11}, m_{21}, \ldots, m_{1n}, \ldots, m_{nn})^T$. The above linear system (2.17) can be solved by the following procedure:

Procedure 1

- Compute Δy by solving the linear system

$$[AE^{-1}FA^T]\Delta y = r + AE^{-1}[F\mathbf{vec}(R_d) - \mathbf{vec}(R_c)],$$

- Compute $\Delta S,\ \Delta X$ as follows:

$$\Delta S \ = \ (1-\xi)R_d - \sum_{i=1}^{m} w_i A_i,$$

$$\mathbf{vec}(\Delta X) \;=\; E^{-1}(R_c - F\mathbf{vec}(\Delta S)).$$

Mizuno-Todd-Ye type algorithms using the different search directions described have been investigated in many papers (cf. [32, 39, 42, 51, 54, 53, 56, 55, 58]). Here, we present a general version of the algorithm. Let $0 < \alpha < \beta < 1$ be two parameters measuring the sizes of two neighborhoods of the central path. Then a typical Mizuno-Todd-Ye type algorithm moves the iterates forward to a neighborhood of size β and then back to the smaller one of size α by alternately using the affine and centering directions.

Algorithm 5.1.1 *Choose $\beta \in (0,1)$, and $\rho > 0$;*

$(X, y, S) \leftarrow (\rho I, 0, \rho I)$;

Repeat until $(X,\ y,\ S) \in \mathcal{F}_\epsilon$:

(Predictor step)

Solve the linear system (2.17) with $\sigma = 0$;

Compute

$$\bar{\theta} = \max\{\tilde{\theta} \in [0,1] : \left(\sum_{i=1}^{n}(\lambda_i(X(\theta)S(\theta) - (1-\theta)\mu I)^2\right)^{1/2} \le \beta(1-\theta)\mu, \ \ \forall\, \theta \in [0, \tilde{\theta}]\},$$

where

$$(X(\theta), S(\theta)) = (X, S) + \theta(\Delta X, \Delta S), \ \ \mu = X \bullet S/n;$$

$(X, y, S) \leftarrow (X, y, S) + \bar{\theta}(\Delta X, \Delta y, \Delta S)$;

(Corrector step)

Solve the linear system (2.17) with $\sigma = 1$;

$(X, y, S) \leftarrow (X, y, S) + (\Delta X, \Delta y, \Delta S)$.

The computation of the step size $\bar{\theta}$ involves solving a complicated nonlinear equation. Actually, we can use a lower bound as proposed in [54, 58] for the HKM and NT directions, and in [42] for the AHO directions. Specifically, for the HKM and

NT directions, we have

$$\bar{\theta} \geq \frac{2}{\sqrt{1 + 4\|P\Delta X \Delta S P^{-1}\|_F/(\mu(\beta-\alpha))} + 1}.$$

As described in [42], in the case of AHO direction, we have

$$\bar{\theta} \geq \frac{2}{\sqrt{\left(\frac{\omega+\beta-\alpha}{\beta-\alpha}\right)^2 + \frac{4\delta}{\beta-\alpha}} + \frac{\omega+\beta-\alpha}{\beta-\alpha}},$$

where

$$\begin{aligned}\omega &= \frac{1}{\mu}\|X^{-1/2}(X\Delta S + \Delta X S + XS)X^{1/2}\|_F,\\ \delta &= \frac{1}{\mu}\|X^{-1/2}\Delta X \Delta S X^{1/2}\|_F.\end{aligned}$$

In our *Mathematica* package, we set $\beta = 0.499$ and α to be the size of the current neighborhood:

$$\alpha := \|X^{1/2}SX^{1/2} - \mu I\|_F/\mu.$$

The Mehrotra type algorithms in our *Mathematica* package follow the paper of Todd, Toh and Tütüncü [60]. Let us briefly describe the algorithm.

Algorithm 5.1.2 *(Todd-Toh-Tütüncü[60])*

Choose $\nu \in (0,1)$ *and* $\rho > 0$*;*

$(X, y, S) \leftarrow (\rho I, 0, \rho I)$;

Repeat until $(X,\ y,\ S) \in \mathcal{F}_\epsilon$*:*

(Predictor step)

Compute the predictor direction $(\delta X, \delta y, \delta S)$ *by solving the linear system (2.17) with* $\sigma = 0$*;*

Determine the parameter σ*:*

$$\sigma := \frac{[(X + \psi\delta X) \bullet (S + \phi\delta S)]^2}{[X \bullet S]^2}, \tag{5.2}$$

where

$$\psi := \frac{-\nu}{\min(-\nu,\ \lambda_{\min}(X^{-1}\delta X))}, \quad \phi := \frac{-\nu}{\min(-\nu,\ \lambda_{\min}(S^{-1}\delta S))}. \tag{5.3}$$

(Corrector step)

Compute the corrector direction $(\Delta X, \Delta y, \Delta S)$ *by solving linear system (2.17) with* σ *defined by (5.2) and right side of (4.5a) modified as*

$$\sigma\mu I - H_P(XS + \delta X \delta S).$$

Compute ψ *and* ϕ *from (5.3) with* $\delta X, \delta S$ *replaced by* $\Delta X, \Delta S$.

$(X, y, S) \leftarrow (X, y, S) + (\psi \Delta X, \phi \Delta y, \phi \Delta S)$.

In our *Mathematica* package, we choose $\nu = 0.98$. Following [60], we choose σ as follows:

$$\sigma_{\text{actual}} = \begin{cases} \max(0.05, \sigma) & \text{if } \psi + \phi \geq 1.8, \\ \max(0.1, \sigma) & \text{if } 1.4 \leq \psi + \phi < 1.8 \\ \max(0.2, \sigma) & \text{if } \psi + \phi < 1.4. \end{cases},$$

5.2 Homogeneous predictor-corrector algorithms

Potra and Sheng [53] proposed a homogeneous formulation of SDP:

$$A_i \bullet X = \tau b_i, \ i = 1, \ldots, m, \tag{5.4a}$$

$$\sum_{i=1}^{m} y_i A_i + S = \tau C, \tag{5.4b}$$

$$b^T y - C \bullet X = \kappa, \tag{5.4c}$$

$$X \succeq 0, \quad S \succeq 0, \ \tau \geq 0, \kappa \geq 0. \tag{5.4d}$$

The search direction $(\Delta X, \Delta y, \Delta S, \Delta\tau, \Delta\kappa)$ of homogeneous algorithms is defined by the following linear system:

$$H_P(\Delta X S + X \Delta S) = \xi\mu I - H_P(XS), \tag{5.5a}$$

$$\kappa \Delta\tau + \tau \Delta\kappa = \xi\mu - \tau\kappa, \tag{5.5b}$$

$$A_i \bullet \Delta X - b_i \Delta\tau = (1 - \xi) R_i, \quad i = 1, \ldots, m, \tag{5.5c}$$

$$\sum_{i=1}^{m} \Delta y_i A_i + \Delta S - \Delta\tau C = (1 - \xi) R_d, \tag{5.5d}$$

$$\Delta\kappa - b^T \Delta y + C \bullet \Delta X = (1 - \xi)\gamma, \tag{5.5e}$$

where ξ is a parameter such that $\xi \in [0,\ 1]$,

$$\begin{aligned}
\mu &= \frac{X \bullet S + \tau\kappa}{n+1},\\
R_d &= -\left(\sum_{i=1}^{m} y_i A_i + S - \tau C\right),\\
r_i &= -\left(A_i \bullet X - b_i \tau\right), i = 1, \dots, m,\\
\gamma &= -\left(\kappa - b^T y + C \bullet X\right).
\end{aligned}$$

In what follows we will derive a simple procedure to compute the direction $(\Delta X, \Delta y, \Delta S, \Delta\tau, \Delta\kappa)$ defined by (5.5).

Let

$$\begin{aligned}
\tilde{A}^T &= [\mathbf{vec}(A_1), \mathbf{vec}(A_2), ..., \mathbf{vec}(A_m), -\mathbf{vec}(C)] = [A^T, -\mathbf{vec}(C)],\\
b^T &= [b_1, b_2, ..., b_m],\\
r^T &= [r_1, r_2, ..., r_m],\\
\gamma_c &= \xi\mu - \tau\kappa. \qquad (5.6)
\end{aligned}$$

Further, let $E,\ F \in \mathbb{R}^{n^2 \times n^2}$, $R_c \in \mathbb{R}^{n \times n}$ be such that (5.5a) has an equivalent vector form:

$$E\mathbf{vec}(\Delta X) + F\mathbf{vec}(\Delta S) = \mathbf{vec}(R_c). \qquad (5.7)$$

From (5.5b) and (5.6), we have

$$\Delta\kappa = \gamma_c/\tau - (\kappa/\tau)\Delta\tau.$$

Putting the above expression for $\Delta\kappa$ into (5.5e), we obtain

$$-C \bullet \Delta X = -(\kappa/\tau)\Delta\tau - b^T \Delta y + \gamma_c/\tau - (1-\xi)\gamma. \qquad (5.8)$$

By (5.5d), we get

$$\tilde{A}^T \begin{pmatrix} \Delta y \\ \Delta\tau \end{pmatrix} + \mathbf{vec}(\Delta S) = (1-\xi)\mathbf{vec}(R_d). \qquad (5.9)$$

Therefore, from (5.7), (5.8) and (5.9), we deduce

$$\begin{aligned}
&\tilde{A}E^{-1}F\tilde{A}^T \begin{pmatrix} \Delta y \\ \Delta\tau \end{pmatrix}\\
= \;& -\tilde{A}E^{-1}F\mathbf{vec}(\Delta S) + (1-\xi)\tilde{A}E^{-1}F\mathbf{vec}(R_d)\\
= \;& \tilde{A}(\mathbf{vec}(\Delta X) - E^{-1}\mathbf{vec}(R_c)) + (1-\xi)\tilde{A}E^{-1}F\mathbf{vec}(R_d)
\end{aligned}$$

$$
\begin{aligned}
&= \tilde{A}\mathbf{vec}(\Delta X) + \tilde{A}E^{-1}[(1-\xi)F\mathbf{vec}(R_d) - \mathbf{vec}(R_c)] \\
&= \begin{pmatrix} \Delta\tau b + (1-\xi)r \\ -C \bullet \Delta X \end{pmatrix} + \tilde{A}E^{-1}[(1-\xi)F\mathbf{vec}(R_d) - \mathbf{vec}(R_c)] \\
&= \begin{pmatrix} \Delta\tau b + (1-\xi)r \\ \gamma_c/\tau - (1-\xi)\gamma - (\kappa/\tau)\Delta\tau - b^T\Delta y \end{pmatrix} + \tilde{A}E^{-1}[(1-\xi)F\mathbf{vec}(R_d) - \mathbf{vec}(R_c)] \\
&= \begin{pmatrix} 0 & b \\ -b^T & -\kappa/\tau \end{pmatrix} \begin{pmatrix} \Delta y \\ \Delta\tau \end{pmatrix} + \begin{pmatrix} (1-\xi)r \\ \gamma_c/\tau - (1-\xi)\gamma \end{pmatrix} \\
&\quad + \tilde{A}E^{-1}[(1-\xi)F\mathbf{vec}(R_d) - \mathbf{vec}(R_c)].
\end{aligned}
$$

Then, we have the following procedure.

Procedure 2

- Compute Δy and $\Delta\tau$ by solving the linear system

$$
\begin{aligned}
\left[\tilde{A}E^{-1}F\tilde{A}^T + \begin{pmatrix} 0 & -b \\ b^T & \kappa/\tau \end{pmatrix}\right] \begin{pmatrix} \Delta y \\ \Delta\tau \end{pmatrix} &= \tilde{A}E^{-1}[(1-\xi)F\mathbf{vec}(R_d) - \mathbf{vec}(R_c)] \\
&\quad + \begin{pmatrix} (1-\xi)r \\ \xi\mu/\tau - \kappa - (1-\xi)\gamma \end{pmatrix}. \qquad (5.10)
\end{aligned}
$$

- Compute ΔS, ΔX and $\Delta\kappa$ as follows:

$$
\begin{aligned}
\Delta S &= (1-\xi)R_d - \sum_{i=1}^{m} w_i A_i + \Delta\tau C, \\
\mathbf{vec}(\Delta X) &= E^{-1}(R_c - F\mathbf{vec}(\Delta S)), \\
\Delta\kappa &= (\xi\mu - \kappa\Delta\tau)/\tau - \kappa.
\end{aligned}
$$

The Mizuno-Todd-Ye type homogeneous algorithms are described as follows.

Algorithm 5.2.1

Choose $\beta \in (0,1)$*;*

$(X, y, S, \tau, \kappa) \leftarrow (I, 0, I, 1, 1)$;

Repeat until $(X/\tau,\ y/\tau,\ S/\tau) \in \mathcal{F}_\epsilon$ *or* τ *is sufficiently small:*

(Predictor step)

Solve the linear system (5.5) with $\xi = 0$*;*

Compute

$$\bar\theta = \max\{\tilde\theta \in [0,1] :$$
$$\left(\sum_{i=1}^{n}(\lambda_i(X(\theta)S(\theta)) - (1-\theta)\mu)^2 + (\tau(\theta)\kappa(\theta)-\mu)^2\right)^{1/2} \le \beta(1-\theta)\mu,$$
$$\forall\, \theta \in [0,\tilde\theta]\},$$

where

$$(X(\theta), S(\theta), \tau(\theta), \kappa(\theta)) = (X, S, \tau, \kappa) + \theta(\Delta X, \Delta S, \Delta\tau, \Delta\kappa);$$

$(X, y, S, \tau, \kappa) \leftarrow X, y, S, \tau, \kappa) + \bar\theta(\Delta X, \Delta y, \Delta S, \Delta\tau, \Delta\kappa);$

(Corrector step)

Solve the linear system (5.5) with $\xi = 1$;

$(X, y, S, \tau, \kappa) \leftarrow X, y, S, \tau, \kappa) + (\Delta X, \Delta y, \Delta S, \Delta\tau, \Delta\kappa).$

In accordance with Algorithm 5.1.1, we can use a lower bound for the step size $\bar\theta$. If the HKM or NT direction is used, then we have

$$\bar\theta \ge \frac{2}{\sqrt{1 + 4\zeta/(\beta-\alpha)} + 1},$$

where

$$\zeta = \frac{1}{\mu}\left(\|\Delta X \Delta S\|_F^2 + (\Delta\tau\Delta\kappa)^2\right)^{1/2}.$$

In the case of AHO direction, we have

$$\bar\theta \ge \frac{2}{\sqrt{\left(\frac{\omega+\beta-\alpha}{\beta-\alpha}\right)^2 + \frac{4\delta}{\beta-\alpha}} + \frac{\omega+\beta-\alpha}{\beta-\alpha}},$$

where

$$\begin{aligned} \omega &= \frac{1}{\mu}\left(\|X^{-1/2}(X\Delta S + \Delta X S + XS)X^{1/2}\|_F^2 + (\tau\Delta\kappa + \kappa\Delta\tau + \tau\kappa)^2\right)^{1/2}, \\ \delta &= \frac{1}{\mu}\left(\|X^{-1/2}\Delta X\Delta S X^{1/2}\|_F^2 + (\Delta\tau\Delta\kappa)^2\right)^{1/2}. \end{aligned}$$

In our *Mathematica* package, we set $\beta = 0.499$ and α to be the size of the current neighborhood:

$$\alpha := \frac{1}{\mu}\left(\|X^{1/2}SX^{1/2} - \mu I\|_F^2 + (\tau\kappa-\mu)^2\right)^{1/2}.$$

Next, let use describe the Mehrotra type homogeneous algorithms.

Algorithm 5.2.2

Choose $\nu \in (0,1)$.

$(X, y, S, \tau, \kappa) \leftarrow (I, 0, I, 1, 1)$;

Repeat until $(X/\tau,\ y/\tau,\ S/\tau) \in \mathcal{F}_\epsilon$ *or* τ *is sufficiently small:*

(Predictor step)

Compute the predictor direction $(\delta X, \delta y, \delta S, \delta\tau, \delta\kappa)$ *by solving the linear system (5.5) with* $\xi = 0$;

Determine the parameter ξ:

$$\xi := \frac{[(X + \theta\delta X) \bullet (S + \theta\delta S) + (\tau + \theta\delta\tau)(\kappa + \theta\delta\kappa)]^2}{[X \bullet S + \tau\kappa]^2}, \tag{5.11}$$

where

$$\theta := \frac{-\nu}{\min\left(-\nu,\ \delta\tau/\tau,\ \lambda_{\min}(X^{-1}\delta X),\ \delta\kappa/\kappa,\ \lambda_{\min}(S^{-1}\delta S)\right)}. \tag{5.12}$$

(Corrector step)

Compute the corrector direction $(\Delta X, \Delta y, \Delta S, \Delta\tau, \Delta\kappa)$ *by solving linear system (5.5) with* ξ *defined by (5.11) and right sides of (5.5a)and (5.5b) replaced by*

$$\xi\mu I - H_P(XS + \delta X \delta S),$$

and

$$\xi\mu - \tau\kappa - \delta\tau\delta\kappa$$

respectively;

Compute θ *from (5.12) with* $\delta X, \delta S, \delta\tau, \delta\kappa$ *replaced by* $\Delta X, \Delta S, \Delta\tau, \Delta\kappa$;

$(X, y, S, \tau, \kappa) \leftarrow (X, y, S, \tau, \kappa) + \theta(\Delta X, \Delta y, \Delta S, \Delta\tau, \Delta\kappa)$.

In our *Mathematica* package, we choose $\nu = 0.98$. Similar to [60], we choose ξ as follows:

$$\xi_{\text{actual}} = \begin{cases} \max(0.05, \xi) & \text{if } \theta \geq 0.9, \\ \max(0.1, \xi) & \text{if } 0.7 \leq \theta < 0.9 \\ \max(0.2, \xi) & \text{if } \theta < 0.7. \end{cases},$$

5.3 SDPMATH – a Mathematica package

We have recently developed a Mathematica package for SDP which implements the algorithms discussed the above sections. We implement the following search directions: AHO, HKM (i.e., HRVW/KSH/M), and NT.

For each of these search directions, we implement two Mizuno-Todd-Ye type predictor-corrector algorithms and two Mehrotra (M) type predictor algorithms. Half of the algorithms are infeasible-interior-point algorithms (I), which require starting points of the form $(X, y, S) = (\rho I, 0, \rho I)$ with large ρ, while the other half use the Potra-Sheng homogeneous formulation (H) [53] which uses the starting point $(X, y, S) = (I, 0, I)$.

We consider the following classes of semidefinite programs:

1. **A simple SDP: m=3, n=2**

First we test the algorithms on a semidefinite program of small size:

$$C = \begin{pmatrix} 5 & -1 \\ -1 & 0 \end{pmatrix} \tag{5.13a}$$

$$A = \left[\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \begin{pmatrix} 3 & 1 \\ 1 & -2 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} \right] \tag{5.13b}$$

$$b = \begin{pmatrix} 5 \\ -1 \\ -4 \end{pmatrix} \tag{5.13c}$$

It can be verified that one solution to this SDP is:

$$X = \begin{pmatrix} 1 & -3/7 \\ -3/7 & 11/7 \end{pmatrix} \tag{5.14a}$$

$$y = \begin{pmatrix} 10/7 \\ 9/7 \\ 1/35 \end{pmatrix} \tag{5.14b}$$

$$S = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \tag{5.14c}$$

2. **An infeasible SDP: m=3, n=2**

Permuting (5.13) slightly, we obtain an SDP with no solutions:

$$C = \begin{pmatrix} 5 & -1 \\ -1 & 0 \end{pmatrix} \tag{5.15a}$$

$$A = \left[\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \begin{pmatrix} 3 & 1 \\ 1 & -2 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 3 & 1.01 \end{pmatrix} \right] \tag{5.15b}$$

$$b = \begin{pmatrix} 5 \\ -1 \\ -4 \end{pmatrix} \tag{5.15c}$$

A useful feature of our homogeneous algorithms is that these algorithms can determine the feasibility of a given problem after a small number of steps. If, during the execution of a homogeneous algorithm, the condition $\tau < \epsilon^2$ holds, then the problem is considered to be infeasible and the algorithm halts.

3. **Minimum eigenvalue problem:**

$$min\{\lambda : \lambda \text{ is an eigenvalue of } M\} \tag{5.16}$$

where $M \in \mathcal{S}^N$. It is well-known that this problem can be expressed as an SDP with m = 1, n = N, and

$$C = M \tag{5.17a}$$

$$A = I \tag{5.17b}$$

$$b = (1) \tag{5.17c}$$

To test this problem, we randomly generated a dense random symmetric matrix for $N = 5$, $N = 10$. To ensure the matrix is symmetric, we generate a random $N \times N$ matrix A and compute $(A + A^T)/2$. Note that Mehrotra-type methods require the computation of the minimum eigenvalue of a matrix. For this reason, we do not test the Mehrotra-type algorithms on this type of problem.

our stopping criterion is to cease execution when $E_k < 10^{-6}$, where

$$E_k := \max\{X \bullet S, \ \ |r_i|, \ i = 1, 2, \ldots, m, \ \ \|R_d\|\},$$

for the infeasible algorithms, and

$$E_k := \max\{X \bullet S/\tau^2, \ \ |r_i|/\tau, \ i = 1, 2, \ldots, m, \ \ \|R_d\|/\tau\},$$

for homogeneous algorithms.

We choose $(X, y, S) = (\rho I, 0, \rho I)$ with $\rho = 10^4$ as our starting point for the infeasible-interior-point algorithms, while for the homogeneous algorithms we use the fixed starting point $(X, y, S, \tau, \kappa) = (I, 0, I, 1, 1)$.

Tables 5.1-5.3 show the number of iterations required to solve the above SDP problems, where "PC" the Mizuno-Todd-Ye type algorithms, "M-PC" indicates the Mehrotra type algorithms, and "H" denotes the homogeneous versions.

Our *Mathematica* package is freely available at

http://www.cs.uiowa.edu/~brixius/sdp.html.

Finally, we remark that this package would not have been possible without the help of M. Kojima, who wrote the Mathematica code PINPAL [30], and M. J. Todd, K. C. Toh, and R. H. Tütüncü, who generously provided their MATLAB code, allowing us to effectively implement the Mehrotra-type methods and present many of the examples considered here [60].

PC			*H-PC*			*I-M-PC*			*H-M-PC*		
HKM	AHO	NT	HKM	AHO	NT	HKM	AHO	NT	HKM	AHO	NT
7	7	7	5	5	5	10	11	14	6	6	6

Table 5.1: Iterations to solve (5.13)

H-PC			*H-M-PC*		
AHO	HKM	NT	AHO	HKM	NT
10	11	11	13	13	13

Table 5.2: Iterations to detect infeasibility in (5.15)

PC			*H-PC*		
AHO	HKM	NT	AHO	HKM	NT
20	20	20	14	15	15

Table 5.3: Iterations to solve minimum eigenvalue problem on 10x10 matrix

5.4 SDPHA – a MATLAB package

Very recently, we have developed a MATLAB package – SDPHA for SDP which implements Mehrotra predictor-corrector algorithm and uses the homogeneous formulation [57]. In our MATLAB code SDPHA, we adopted a technique of rank-2 update to solve the linear system (5.10), which is described below.

Case 1. $\tilde{A}E^{-1}F\tilde{A}^T \in \mathbb{R}^{(m+1)\times(m+1)}$ is symmetric.

Note that the HKM and NT directions belong to this case. Let

$$M = \tilde{A}E^{-1}F\tilde{A}^T + \begin{pmatrix} 0 & 0 \\ 0 & \kappa/\tau + \alpha \end{pmatrix},$$

where $\alpha > 0$. Also, let

$$p = (f, \ \ -\alpha f - g), \qquad q = (g, \ \ f), \tag{5.18}$$

where

$$f = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} = e_{m+1}, \qquad g = \begin{pmatrix} b \\ 0 \end{pmatrix}.$$

Note that $p, \ q \in \mathbb{R}^{(m+1)\times 2}$. Then it is easily seen that (5.10) is equivalent to

$$(M + pq^T)\begin{pmatrix} \Delta y \\ \Delta\tau \end{pmatrix} = r_h. \tag{5.19}$$

Let us consider the Cholesky factorization of M: $M = LL^T$. Then by (5.19) we can write

$$\begin{aligned} \begin{pmatrix} \Delta y \\ \Delta\tau \end{pmatrix} &= (M + ab^T)^{-1} r_h \\ &= (LL^T + pq^T)^{-1} r_h \\ &= \{L[I + (L^{-1}p)(L^{-1}q)^T]L^T\}^{-1} r_h \\ &= (L^T)^{-1}[I + (L^{-1}p)(L^{-1}q)^T]^{-1} L^{-1} r_h, \end{aligned} \tag{5.20}$$

By denoting

$$u = L^{-1}p, \quad v = L^{-1}q, \quad w = L^{-1}r_h.$$

and by using (5.20) and the fact that

$$(I + uv^T)^{-1} = I - u(I_2 + v^T u)^{-1} v^T,$$

we obtain

$$\begin{aligned} \begin{pmatrix} \Delta y \\ \Delta\tau \end{pmatrix} &= (L^T)^{-1}[I - u(I_2 + v^T u)^{-1} v^T] w \\ &= (L^T)^{-1}[w - u(I_2 + v^T u)^{-1} v^T w]. \end{aligned} \tag{5.21}$$

Here $I_2 + v^T u \in \mathbb{R}^{2\times 2}$, and u and v can be computed as follows:

$$u = L^{-1}p = (L^{-1}f, \ \ -L^{-1}g - \alpha L^{-1}f),$$

$$v = L^{-1}q = (L^{-1}g, \ \ L^{-1}f).$$

Notice that $L^{-1}f$ is $\begin{pmatrix} \mathbf{0} \\ 1/l_{m+1,m+1} \end{pmatrix}$, where $l_{m+1,m+1}$ is the $(m+1, m+1)$-entry of L. Therefore the computation of (5.21) only involves two forward-iterations $L^{-1}g,\ L^{-1}r_h$ and one backward-iteration $(L^T)^{-1}\hat{w}$ with

$$\hat{w} = w - u(I_2 + v^T u)^{-1} v^T w.$$

Case 2. $\tilde{A}E^{-1}F\tilde{A}^T$ is not symmetric.

Note that the AHO direction belongs to this case. Then we can use the strategy similar to Case 1 by considering the LU decomposition of M instead of the Cholesky decomposition. Then it is easy to see that three forward-iterations and one backward-iteration are needed for the computation of Δy and $\Delta\tau$ in (5.10).

Our computations were performed on an HP-UX9000 workstation. We used the stopping criterion

$$\max\{\frac{\|r_p\|}{\|b\|+1},\ \frac{\|R_d\|_F}{\|C\|_F+1},\ \frac{X \bullet S}{\max\{1,\ (C \bullet X + b^T y)/2\}}\} \le 10^{-9}. \tag{5.22}$$

The left side of the above inequality represents the "relative error". The numerical results are given in Tables 5.4-5.8 for the performance of SDPT3 and SDPHA while using the AHO, HKM and NT directions, respectively. We tested the following problems:

1. the matrix norm minimization problem,
2. the problem of computing the Chebyshev polynomial of a matrix,
3. the Max-Cut problem,
4. the Chebyshev approximation in the complex plane,
5. the logarithmic Chebyshev approximation problem,
6. the problem of "Controller Design in Active Noise Control".

All these problems are taken from [60] and [61] except problem 6 which comes from: Julia A. Olkin, SIAG/OPT Views-and-News, No.8, Fall 1996.

The results in Tables 5.4-5.6 use the starting point

$$(X^0, y^0, S^0) = (1000I,\ \mathbf{0},\ 1000I),$$

for infeasible-interior algorithms in SDPT3.

Table 5.7 compares the performance of SDPT3 and SDPHA by solving a small problem of $m = n = 2$ which has a solution that is large in magnitude. The problem is actually a scaled version of the Kojima-Shida-Shindoh problem [34]:

$$C = \begin{pmatrix} 0 & 0 \\ 0 & 10^6 \end{pmatrix} \tag{5.23a}$$

$$A = \left[\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & -2 \end{pmatrix} \right] \tag{5.23b}$$

$$b = \begin{pmatrix} -2 \cdot 10^6 \\ 0 \end{pmatrix} \tag{5.23c}$$

whose solution is:

$$X^* = \begin{pmatrix} 10^6 & 0 \\ 0 & 0 \end{pmatrix} \tag{5.24a}$$

$$y^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{5.24b}$$

$$S^* = \begin{pmatrix} 0 & 0 \\ 0 & 10^6 \end{pmatrix} \tag{5.24c}$$

We implemented the infeasible-interior-point algorithms in SDPT3 using the following starting point:

$$(X^0, y^0, S^0) = (\rho I,\ \mathbf{0},\ \rho I).$$

Results are given for $\rho = 10, 10^3, \ldots, 10^{11}$, and also for the tune-up starting point suggested in SDPT3. We note that the symbol "$*$" means the algorithm fails while "$+$" indicates the algorithm stops without satisfying the stopping rule (5.22). Finally, Table 5.8 gives the same comparison by solving an educational testing problem (ETP) [60, 61] with $n = 30, m = 15$ which is scaled such that it has a solution with $\|X^*\|_F \geq 10^3,\ \ \|S^*\|_F \geq 10^3$.

Our numerical results indicate that the homogeneous algorithms generally take fewer iterations to capture a desired solution. and that the homogeneous algorithms outperform their non-homogeneous versions, with improvement of more than 20% in many cases, in terms of total CPU time.

The MATLAB package SDPHA can be obtained through URL:

http://www.math.uiowa.edu/~rsheng/SDPHA/sdpha.html.

			SDPT3			SDPHA		
Problem	n	m	its.	cpu sec	gap	its.	cpu sec	gap
1	20	50	13	77.1	3.18e-11	10	58.7	1.66e-10
2	50	11	12	125.3	4.59e-11	8	83.4	8.59e-11
3	20	20	11	17.7	1.55e-9	9	14.9	5.93e-10
4	90	11	13	109.4	4.78e-11	9	78.0	1.45e-10
5	60	10	15	40.3	5.48e-11	11	30.1	1.16e-10
6	20	20	9	13.4	3.48e-10	6	9.8	6.06e-11

Table 5.4: Comparing SDPHA with SDPT3 using the AHO direction.

Problem	n	m	SDPT3			SDPHA		
			its.	cpu sec	gap	its.	cpu sec	gap
1	20	50	13	47.7	2.23e-9	11	37.1	4.56e-10
2	50	11	13	74.5	3.83e-10	9	53.4	3.46e-10
3	20	20	12	12.2	3.49e-8	11	9.9	3.50e-10
4	90	11	14	108.6	1.49e-10	11	86.8	1.80e-11
5	60	10	16	22.9	1.09e-10	12	16.0	3.81e-10
6	20	20	9	8.7	3.81e-10	6	6.0	5.28e-11

Table 5.5: Comparing SDPHA with SDPT3 using the HKM direction.

Problem	n	m	SDPT3			SDPHA		
			its.	cpu sec	gap	its.	cpu sec	gap
1	20	50	13	48.4	2.66e-9	11	39.3	1.64e-10
2	50	11	13	84.2	7.09e-10	10	66.5	3.57e-11
3	20	20	12	13.0	3.25e-8	10	10.2	6.85e-9
4	40	11	14	143.5	6.29e-10	10	106.5	9.39e-11
5	60	6	15	25.0	1.57e-9	12	19.5	1.98e-10
6	20	20	9	9.3	3.81e-10	6	6.7	5.28e-11

Table 5.6: Comparing SDPHA with SDPT3 using the NT direction.

SDPT3	AHO			HKM			NT		
ρ	its.	cpu sec	gap	its.	cpu sec	gap	its.	cpu sec	gap
10	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$
10^3	21+	2.4+	2.72e-6	$*$	$*$	$*$	$*$	$*$	$*$
10^5	16	1.9	1.36e-11	16	1.4	1.19e-10	18	1.9	1.03e-10
10^7	16	1.8	8.10e-10	15	1.3	6.72e-10	16	1.7	2.53e-11
10^9	19	2.2	3.27e-10	20+	1.7+	3.72e-9	20	2.1	1.25e-10
10^{11}	21	2.5	2.04e-10	21	1.8	3.08e-10	21	2.2	1.88e-10
tune-up	$*$	$*$	$*$	12	1.0	9.93e-10	14	1.5	2.52e-10
SDPHA	12	1.9	4.02e-10	12	1.3	6.04e-10	12	1.6	6.21e-10

Table 5.7: Comparing SDPHA with SDPT3 by solving a problem of $n = m = 2$ and using various starting points.

SDPT3	AHO			HKM			NT		
ρ	its.	cpu sec	gap	its.	cpu sec	gap	its.	cpu sec	gap
10^2	18	29.6	3.19e-5	22	21.0	5.46e-5	20	20.9	4.67e-5
10^4	13	22.0	1.91e-5	15	14.8	3.05e-5	15+	15.9+	1.40e-4
10^6	16	26.2	3.07e-6	17	16.8	4.87e-5	17	18.3	5.70e-5
10^8	18	29.4	2.24e-6	19	18.0	6.47e-5	19	19.5	8.31e-5
10^{10}	20	32.5	2.83e-6	21	19.8	7.42e-5	21	21.6	7.11e-5
10^{12}	22	35.6	3.38e-6	23	21.6	8.69e-5	23	23.6	5.94e-5
tune-up	13	21.2	2.62e-6	15	14.2	2.88e-5	15+	15.5+	1.23e-4
SDPHA	12	20.8	2.83e-6	14	12.2	3.03e-5	13	13.3	3.11e-5

Table 5.8: Comparing SDPHA with SDPT3 by solving a ETP problem of $n = 30, m = 15$ and using various starting points.

REFERENCES

[1] F. Alizadeh. *Combinatorial optimization with interior point methods and semi–definite matrices*. PhD thesis, University of Minnesota, Minneapolis, Minnesota, USA, 1991.

[2] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5(1):13–51, 1995.

[3] F. Alizadeh, J.-P. A. Haeberly, M. V. Nayakkankuppam, and M. L. Overton. SDPpack user's guide version 0.8 beta. Technical Report 734, NYU Computer Science Department, March 1997.

[4] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton. Primal-dual interior point methods for semidefinite programming. Technical Report 659, Computer Science Department, Courant Institute of Mathematical Sciences, New York University, 1994.

[5] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton. Complementarity and nondegeneracy in semidefinite programming. Technical Report 681, NYU Computer Science Dept, 1995. To appear in Mathematical Programming (Series B).

[6] F. Alizadeh, J.-P. A. Haeberly, and M. L. Overton. Primal-dual interior-point methods for semidefinite programming: convergence rates, stability and numerical results. Technical Report 721, New York University, New York, NY, May 1996.

[7] K. M. Anstreicher and M. Fampa. A long-step path following algorithm for semidefinite programming problems. Technical report, University of Iowa, Iowa City, IA, January 1996. (Revised January 29, 1996).

[8] K. H. Borgwardt. *The Simplex Method: A Probabilistic Analysis*. Springer–Verlag, Berlin, 1987.

[9] N. Brixius, F. A. Potra, and R. Sheng. Solving semidefinite programs in *mathematica*. Reports on Computational Mathematics 97, Department of Mathematics, The University of Iowa, Iowa City, IA 52242, USA, November 1996.

[10] E. de Klerk, C. Roos, and T. Terlaky. Semi-definite problems in truss topology optimization. Report of the Faculty of Technical Mathematics and Informatics 95-128, Delft University of Technology, November 1995.

[11] I. I. Dikin. Iterative solution of problems of linear and quadratic programming. *Doklady Akademii Nauk SSSR*, 174:747–748, 1967. Translated in : *Soviet Mathematics Doklady*, 8:674–675, 1967.

[12] A. V. Fiacco and G. P. McCormick. *Nonlinear Programming : Sequential Unconstrained Minimization Techniques.* John Wiley & Sons, New York, 1968. Reprint : Volume 4 of *SIAM Classics in Applied Mathematics*, SIAM Publications, Philadelphia, PA 19104–2688, USA, 1990.

[13] R. M. Freund. Complexity of an algorithm for finding an approximate solution of a semi-definite program with no regular assumption. Working Paper OR 302-94, Operations Research Center, Massachusetts Institute of Technology, December 1994.

[14] R. M. Freund and S. Mizuno. Interior point methods: Current status and future directions. *Optima*, 51:1–9, October 1996.

[15] L. El Ghaoui, F. Oustry, and M. AitRami. An lmi-based linearization algorithm for static output-feedback and related problems. Technical report, Ecole Nationale Supérieure de Techniques Avancés, Paris, France, November 1995.

[16] L. El Ghaoui and G. Scorletti. Control of rational systems using linear-fractional representaions and linear matrix inequalities. Preprint, Ecole Nationale Supérieure de Techniques Avancées, Paris, France, April 1996.

[17] M. X. Goemans and D. P. Williamson. Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming. *J. Assoc. Comput. Mach.*, 1994. (To appear). A preliminary version appeared in Proc. 26th Annual ACM Symposium on Theory of Computing.

[18] D. Goldfarb and K. Scheinberg. Interior point trajectories in semidefinite programming. Technical report, Columbia University, New York, NY, November 1996.

[19] A. J. Goldman and A. W. Tucker. Polyhedral convex cones. In H. W. Kuhn and A. W. Tucker, editors, *Linear Inequalities and Related Systems*. Princeton University Press, Princeton, NJ, 1956. p. 19.

[20] C. Helmberg, F. Rendl, R.J. Vanderbei, and H. Wolkowicz. An interior-point method for semidefinite programming. *SIAM J. Optimization*, 6(2):342–361, 1996.

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

[22] P. Huard and B. T. Lieu. La methode des centres dans un espace topologique. *Numerische Mathematik*, 8:56–67, 1968. (In French).

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

[24] J. Ji, F. A. Potra, and R. Sheng. On local convergence of Monteiro's generalized predictor-corrector method for semidefinite programming. Work in progress, 1996.

[25] C. R. Johnson, B. Kroschel, and H. Wolkowicz. An interior-point method for approximate positive semidefinite completions. Technical report, University of Waterloo, August 1996.

[26] D. Karger, R. Motwani, and M. Sudan. Improved graph coloring by semidefinite programming. In *35th Symposium on Foundations of Computer Science*, pages 2–13. IEEE Computer Science Society Press, 1994.

[27] S. E. Karisch and F. Rendl. Semidefinite programming and graph equipartition. Technical report, Technische Universität Graz, December 1995.

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

[29] L. G. Khachiyan. A polynomial algorithm for linear programming. *Soviet Mathematics Doklady*, 20:191–194, 1979.

[30] M. Kojima. A primitive interior point algorithm for semidefinite programs in *mathematica*. Research Reports on Information Sciences B-293, Department of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, December 1994.

[31] M. Kojima, S. Mizuno, and A. Yoshise. A primal–dual interior point algorithm for linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming*, pages 29–47. Springer Verlag, New York, 1989.

[32] M. Kojima, M. Shida, and S. Shindoh. Global and local convergence of predictor–corrector infeasible–interior–point algorithms for semidefinite programs. Research Reports on Information Sciences B-305, Department of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, October 1995.

[33] M. Kojima, M. Shida, and S. Shindoh. Local convergence of predictor–corrector infeasible–interior–point algorithms for semidefinite programs. Research Reports on Information Sciences B-306, Department of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, December 1995.

[34] M. Kojima, M. Shida, and S. Shindoh. A predictor-corrector interior-point algorithm for the semidefinite linear complementarity problem using the Alizadeh-Haeberly-Overton search direction. Research Reports on Information Sciences B-311, Department of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, January 1996.

[35] M. Kojima, S. Shindoh, and S. Hara. Interior-point methods for the monotone linear complementarity problem in symmetric matrices. Research Reports on Information Sciences B-282, Department of Information Sciences, Tokyo Institute of Technology, 2-12-1 Oh-Okayama, Meguro-ku, Tokyo 152, Japan, April 1994.

[36] A. S. Lewis and M. L. Overton. Eigenvalue optimization. In *Acta Numerica 1996*, pages 149–190. Cambridge University Press, 1996.

[37] C-J. Lin and R. Saigal. An infeasible start predictor corrector method for semi-definite linear programming. Technical report, University of Michigan, Ann Arbor, Michigan, December 1995.

[38] C-J. Lin and R. Saigal. A predictor corrector method for semi-definite linear programming. Technical report, University of Michigan, Ann Arbor, Michigan, October 1995. (revised December 10, 1995).

[39] Z-Q. Luo, J. F. Sturm, and S. Zhang. Superlinear convergence of a symmetric primal-dual path following algorithm for semidefinite programming. Report 9607/A, Econometric Institute, Erasmus University Rotterdam, The Netherlands, January 1996.

[40] S. Mizuno, M. J. Todd, and Y. Ye. On adaptive-step primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, 18(4):964–981, 1993.

[41] R. D. C. Monteiro. Primal-dual path following algorithms for semidefinite programming. Working paper, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA, September 1995.

[42] R. D. C. Monteiro. Polynomial convergence of primal-dual algorithms for semidefinite programming based on Monteiro and Zhang family of directions. Working paper, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA, July 1996.

[43] R. D. C. Monteiro and T. Tsuchiya. Polynomial convergence of a new family of primal-dual algorithms for semidefinite programming. Report Memorandum 627, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, JAPAN, November 1996.

[44] R. D. C. Monteiro and T. Tsuchiya. Polynomiality of primal-dual algorithms for semidefinite linear complementarity problems based on the Kojima-Shindoh-Hara family of directions. Technical report, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, JAPAN, August 1996.

[45] R. D. C. Monteiro and S. J. Wright. Local convergence of interior-point algorithms for degenerate monotone LCP. *Computational Optimization and Applications*, 3:131–155, 1994.

[46] R. D. C. Monteiro and Y. Zhang. A unified analysis for a class of path-following primal-dual interior-point algorithms for semidefinite programming. Working paper, School of Industrial and Systems Engineering, Georgia Institute of Technology, June 1996.

[47] Y. E. Nesterov. Long–step strategies in interior point potential–reduction methods. Working paper, Department SES COMIN, Geneva University, 1993. to appear in Mathematical Programming.

[48] Y. E. Nesterov. Infeasible start interior point primal–dual methods in nonlinear programming. Working paper, Center for Operations Research and Econometrics, 1348 Louvain-la-Neuve, Belgium, 1994.

[49] Y. E. Nesterov and A. S. Nemirovsky. Optimization over positive semidefinite matrices: Mathematical background and user's manual. Technical report, Central Economic & Mathematical Institute, USSR Acad. Sci., Moscow, 1990.

[50] Y. E. Nesterov and A. S. Nemirovsky. *Interior Point Polynomial Methods in Convex Programming : Theory and Applications.* SIAM Publications. SIAM, Philadelphia, USA, 1994.

[51] Y. E. Nesterov and M. J. Todd. Primal–dual interior–point methods for self–scaled cones. Technical Report 1125, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York 14853–3801, USA, 1995.

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

[53] F. A. Potra and R. Sheng. Homogeneous interior–point algorithms for semidefinite programming. Reports on Computational Mathematics 82, Department of Mathematics, The University of Iowa, Iowa City, IA 52242, USA, November 1995.

[54] F. A. Potra and R. Sheng. A superlinearly convergent primal-dual infeasible-interior-point algorithm for semidefinite programming. Reports on Computational Mathematics 78, Department of Mathematics, The University of Iowa, Iowa City, IA 52242, USA, October 1995. (to appear in SIAM Journal on Optimization).

[55] F. A. Potra and R. Sheng. Superlinear convergence of a predictor-corrector method for semidefinite programming without shrinking central path neighborhood. Reports on Computational Mathematics 91, Department of Mathematics, The University of Iowa, Iowa City, IA 52242, USA, August 1996.

[56] F. A. Potra and R. Sheng. Superlinear convergence of interior-point algorithms for semidefinite programming. Reports on Computational Mathematics 86, Department of Mathematics, The University of Iowa, Iowa City, IA 52242, USA, April 1996. (Revised May 1996).

[57] F. A. Potra, R. Sheng, and N. Brixius. SDPHA – a MATLAB implementation of homogeneous interior-point algorithms for semidefinite programming. Reports on Computational Mathematics 100, Department of Mathematics, The University of Iowa, Iowa City, IA 52242, USA, April 1997.

[58] R. Sheng, F. A. Potra, and J. Ji. On a general class of interior-point algorithms for semidefinite programming with polynomial complexity and superlinear convergence. Reports on Computational Mathematics 89, Department of Mathematics, The University of Iowa, Iowa City, IA 52242, USA, June 1996.

[59] J. F. Sturm and S. Zhang. Symmetric primal-dual path following algorithms for semidefinite programming. Report 9554/A, Econometric Institute, Erasmus University Rotterdam, The Netherlands, November 1995. revised in February 1996.

[60] M. J. Todd, K. C. Toh, and R. H. Tütüncü. On the Nesterov-Todd direction in semidefinite programming. Technical Report, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853–3801, USA, 1996.

[61] K. C. Toh, M. J. Todd, and R. H. Tütüncü. SDPT3 — a Matlab software package for semidefinite programming. Manuscript, December 1996.

[62] P. Tseng. Search directions and convergence analysis of some infeasible path-following methods for the monotone semi-definite LCP. Technical report, Department of Mathematics, University of Washington, Seattle, WA 98195, USA, June 1996.

[63] A. W. Tucker. Dual systems of homogeneous relations. In H. W. Kuhn and A. W. Tucker, editors, *Linear Inequalities and Related Systems*. Princeton University Press, Princeton, NJ, 1956. p. 3.

[64] L. Vandenberghe and S. Boyd. A primal–dual potential reduction method for problems involving matrix inequalities. *Mathematical Programming*, 69:205–236, 1995.

[65] L. Vandenberghe and S. Boyd. Semidefinite progamming. *SIAM Review*, 38(1):49–95, 1996.

[66] L. Vandenberghe, S. Boyd, and S.-P. Wu. Determinant maximization with linear matrix inequality constraints. Technical Report, Information Systems Laboratory, Electrical Engineering Department, Stanford University, Stanford, CA 94305, USA, July 1996.

[67] S. J. Wright. *Primal-Dual Interior-Point Methods*. SIAM Publications, 1997.

[68] X. Xu, P. Hung, and Y. Ye. A simplification of the homogeneous and self-dual linear programming algorithm and its implementation. Working paper, Department of Management Sciences, The University of Iowa, Iowa City, Iowa 52242, USA, 1993. to appear in Annals of Operations Research.

[69] Y. Ye, M. J. Todd, and S. Mizuno. An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Mathematics of Operations Research*, 19(1):53–67, 1994.

[70] Y. Zhang. On extending primal–dual interior–point algorithms from linear programming to semidefinite programming. TR 95-20, Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, Maryland 21228–5398, USA, October 1995.

[71] Q. Zhao, S. E. Karisch, F. Rendl, and H. Wolkowicz. Semidefinite programming relaxations for the quadratic assignment problem. CORR Report 95-27, University of Waterloo, N2L 3G1 Canada, September 1996.