

On the Extension of a Mehrotra-Type Algorithm for Semidefinite Optimization

Mohammad H. Koulaei*, Tamás Terlaky†

March 10, 2007

Abstract

It has been shown in various papers that most interior-point algorithms and their analysis can be generalized to semidefinite optimization. This paper presents an extension of the recent variant of Mehrotra's predictor-corrector algorithm that was proposed by Salahi et al. (2005) for linear optimization problems. Based on the NT (Nesterov and Todd 1997) direction as Newton search direction it is shown that the iteration-complexity bound of the algorithm is of the same order as that of the corresponding algorithm for linear optimization.

Keywords: Semidefinite optimization, predictor-corrector algorithms, Mehrotra-Type Algorithm, Polynomial complexity.

1 Introduction

Semidefinite Optimization (SDO) is a generalization of linear optimization (LO), it has various applications in diverse areas, such as system and control theory [4] and combinatorial optimization [2]. Moreover, it turns out that Interior Point Methods (IPMs) can solve SDO problems efficiently, in a comparable way as they solve LO problems. Generalization of IPMs of LO to the context of SDO started in the early of 1990s. The first algorithms in this direction were introduced independently by Alizadeh [1] and Nesterov and Nemirovskii [17]. Alizadeh [1] extends Ye's projective potential reduction algorithm [29] from LO to SDO and argues that many known interior-point algorithms for LO can be transformed into algorithms for SDO. On the other hand, Nesterov and Nemirovskii [17] presented a deep and unified theory of interior-point methods for solving the more general conic optimization problems using the notation of self-concordant barriers; see their book [17] for a comprehensive treatment of this

*Department of Mathematics, Ferdowsi University of Mashhad, P.O. Box 91775-1159, Mashhad, Iran. Research was done when the author was visiting the Advanced Optimization Lab., McMaster University, Hamilton, Canada, email: koulaei@optlab.mcmaster.ca.

†School of Computational Engineering and Science, McMaster University, Hamilton, Canada, email: terlaky@mcmaster.ca. The work of this author was partially supported by the NSERC Grant DG:5-48923, the Canada Research Chair Program and MITACS.

subject. Other IPMs for solving SDO can be found e.g., in [3, 7, 9, 10, 11, 12, 15, 19, 18, 20, 25, 26, 28]. Most of these works are concentrated on primal-dual methods.

Due to their successful implementations predictor-corrector methods are the most popular primal-dual methods. The theoretically most appealing variant is due to Mizuno, Todd and Ye [14] for LO which was extended to SDO by Nesterov and Todd [19]. Mehrotra-type [13] predictor corrector algorithms for SDO have been implemented in the softwares SeDuMi by Sturm [24] and in SDPT3 by Todd et al. [27]. In spite of extensive use of this variant in IPM based optimization packages, not much about its complexity was known before the recent paper by Salahi, Peng and Terlaky [22], which presents a new variant of Mehrotra predictor-corrector algorithm for LO. This variant incorporates a safeguard in the algorithm that keeps the iterates in a prescribed neighborhood and allows to get a reasonably large step size. This safeguard strategy is also used when the affine scaling step performs poorly that effectively forces the algorithm to take pure centering steps.

This paper studies the extension of this Mehrotra-type algorithm for SDO. The analysis for SDO has appeared to be more involved than for LO. A large part of the theoretical difficulty is due to the issue of maintaining symmetry in the linearized complementarity condition. This issue will be discussed in the next section. Zhang [30] has introduced a general symmetrization scheme using an arbitrary nonsingular scaling matrix P , which leads to a class of search directions corresponding to the different P matrices. Among these search directions we would like to highlight the one originally proposed by Nesterov and Todd (NT) [19] which is a member of this class.

The goal of this paper is to establish iteration-complexity bounds for a generalization of the Mehrotra-type algorithm of [22] based on the NT direction. Our derived iteration-complexity bound is $\mathcal{O}(n^2 \log \frac{X^0 \bullet S^0}{\epsilon})$ for the algorithm and $\mathcal{O}(n \log \frac{X^0 \bullet S^0}{\epsilon})$ for its modified version, that are analogous to the linear case.

This paper is organized as follows. In the Section 2, we introduce the SDO problem. We discuss the issue of linearizing the complementarity condition, the NT symmetrization scheme and state our Mehrotra-type algorithm. In Section 3, we state and prove some technical results and based on these results the iteration-complexity bound of the algorithm is established. The modified version of the algorithm and its complexity analysis is presented in Section 4. Finally, conclusion and final remarks are given in Section 5.

The following notation and terminology are used throughout the paper:

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

$\mathbb{R}^{n \times n}$: the set of all $n \times n$ matrices;

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

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

\mathcal{S}_{++}^n : the set of all $n \times n$ symmetric positive definite matrices;

$\text{Tr}(M)$: the trace of matrix $M \in \mathbb{R}^{n \times n}$, $\text{Tr}(M) = \sum_{i=1}^n M_{ii}$;

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

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

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

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

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

$\rho(M) = \max\{|\lambda_i(M)|, i = 1, \dots, n\}$;

$\text{cond}(M) \equiv \lambda_{\max}/\lambda_{\min}$;

$\|\cdot\|$: Euclidean norm of a vector and the corresponding norm of a matrix,

i.e., $\|y\| \equiv \sqrt{\sum_{i=1}^n y_i^2}$, $\|M\| \equiv \max\{\|My\| : \|y\| = 1\}$;

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

$\text{vec}(A)$: the mn -vector obtained by stacking the columns of $A \in \mathbb{R}^{m \times n}$, one by one from the first to the last.

$X \otimes S$: the Kroneker product of $X, S \in \mathbb{R}^{n \times n}$, $X \otimes S \equiv \begin{bmatrix} x_{11}S & \cdots & x_{1n}S \\ \vdots & \ddots & \vdots \\ x_{n1}S & \cdots & x_{nn}S \end{bmatrix}$.

2 The SDO problem and preliminary discussions

In this section we introduce the SDO problem that will be used in this paper and state the symmetrization scheme which is used to derive the Newton direction. We also give some existence results and describe our variant of the Mehrotra-type predictor-corrector algorithm.

2.1 The SDO problem

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

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

where $C \in \mathcal{S}^n$, $A_i \in \mathcal{S}^n$, $i = 1, \dots, m$ and $b = (b_1, \dots, b_m)^T \in \mathbb{R}^m$ are the data, and $X \in \mathcal{S}^n$ represents the primal variables. The corresponding dual SDO problem is

$$\begin{aligned} \max \quad & b^T y \\ \text{s.t.} \quad & \sum_{i=1}^m A_i y_i + S = C, \\ & S \succeq 0, \end{aligned} \tag{2}$$

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

The set of primal dual feasible solutions is denoted by

$$\mathcal{F} = \left\{ (X, y, S) \in \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n : A_i \bullet X = b_i, \quad i = 1, \dots, m, \quad \sum_{i=1}^m A_i y_i + S = C, \quad X \succeq 0, \quad S \succeq 0 \right\},$$

and the sets of interior feasible solutions of (1) and (2) are,

$$\mathcal{F}_{\text{pri}}^0 = \{X \in \mathcal{S}^n : A_i \bullet X = b_i, \quad i = 1, \dots, m, \quad X \succ 0\},$$

$$\mathcal{F}_{\text{dua}}^0 = \{(S, y) \in \mathcal{S}^n \times \mathbb{R}^m : \sum_{i=1}^m A_i y_i + S = C, S \succ 0\},$$

respectively. We make the following assumptions throughout the paper:

Assumption 1. $\mathcal{F}_{\text{pri}}^0 \times \mathcal{F}_{\text{dua}}^0 \neq \emptyset$.

Assumption 2. The matrices A_i , $i = 1, \dots, m$, are linearly independent; i.e., they span an m -dimensional linear space in \mathcal{S}^n .

It is known that neither Assumption 1, nor Assumption 2 are restrictive. See e.g. de Klerk et al. [6] that Assumption 1 can be made without loss of generality. The self-dual embedding model provides an elegant model that always has a strictly feasible primal-dual pair. Further, Assumption 2 can be made true by simple Gaussian elimination.

Under these assumptions, X^* and (y^*, S^*) are solutions of (1) and (2) if and only if they are solutions of the following nonlinear system:

$$\begin{aligned} A_i \bullet X &= b_i, \quad i = 1, \dots, m, \quad X \succeq 0, \\ \sum_{i=1}^m A_i y_i + S &= C, \quad S \succeq 0, \\ XS &= 0, \end{aligned} \tag{3}$$

where the last equality is called the complementarity equation. The *central path* consists of points (X^μ, y^μ, S^μ) satisfying the perturbed system

$$\begin{aligned} A_i \bullet X &= b_i, \quad (i = 1, \dots, m), \quad X \succeq 0, \\ \sum_{i=1}^m A_i y_i + S &= C, \quad S \succeq 0, \\ XS &= \mu I, \end{aligned} \tag{4}$$

where $\mu \in \mathbb{R}$, $\mu > 0$. It is well known [17, 11] that under Assumptions 1 and 2 the solution (X^μ, y^μ, S^μ) of system (4) exists and it is unique for all $\mu > 0$, and that the limit

$$(X^*, y^*, S^*) = \lim_{\mu \rightarrow 0} (X^\mu, y^\mu, S^\mu),$$

exists and it is a primal and dual optimal solution, i.e., (X^*, y^*, S^*) is a solution of (3).

2.2 The Newton direction and the Mehrotra-type algorithm

In this subsection we derive the Newton direction for the system (4) and based on it state our Mehrotra-type algorithm.

Since for $X, S \in \mathcal{S}^n$, the product XS is generally not in \mathcal{S}^n , so, the left-hand side of (4) is a map from $\mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$ to $\mathbb{R}^{n \times n} \times \mathbb{R}^m \times \mathcal{S}^n$. Thus, the system (4) is not a square system when X and S are restricted to \mathcal{S}^n , which it is needed for applying Newton-like methods. A remedy for this is to make the perturbed optimality system (4) square by modifying the left-hand side to a map from $\mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n$ to itself. To achieve this Zhang [30] introduced a general symmetrization scheme based on the so-called *similar symmetrization* operator $H_P : \mathbb{R}^{n \times n} \rightarrow \mathcal{S}^n$ defined as

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

where $P \in \mathbb{R}^{n \times n}$ is some nonsingular matrix. Zhang [30] also observes that

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

for any nonsingular matrix P and any matrix M with real spectrum, (e.g., $M = XS$ with $X, S \in \mathcal{S}_+^n$) and any $\tau \in \mathbb{R}$. It follows that for any given nonsingular matrix P , (4) is equivalent to

$$\begin{aligned} A_i \bullet X &= b_i, \quad i = 1, \dots, m, \quad X \succeq 0, \\ \sum_{i=1}^m A_i y_i + S &= C, \quad S \succeq 0, \\ H_P(XS) &= \mu I. \end{aligned} \tag{5}$$

A Newton-like method applied to system (5) leads to the following linear system:

$$\begin{aligned} A_i \bullet \Delta X &= 0, \quad i = 1, \dots, m, \\ \sum_{i=1}^m A_i \Delta y_i + \Delta S &= 0, \\ H_P(X\Delta S + \Delta XS) &= \sigma \mu_g I - H_P(XS), \end{aligned} \tag{6}$$

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

Search directions obtained by solving (6) are called unified-MZ-search directions [30]. The matrix P used in (6) is called the scaling matrix for the search direction. It is well known that taking $P = I$ results in the AHO search direction [3], $P = S^{1/2}$ corresponds to the HRVW/KSH/M search direction [7, 11], usually referred to as the H..K..M direction, and the case of $P^T P = X^{-\frac{1}{2}}(X^{\frac{1}{2}} S X^{\frac{1}{2}})^{\frac{1}{2}} X^{-\frac{1}{2}}$, in which P is considered to be

$$P = (X^{-\frac{1}{2}}(X^{\frac{1}{2}} S X^{\frac{1}{2}})^{\frac{1}{2}} X^{-\frac{1}{2}})^{\frac{1}{2}}, \tag{7}$$

gives the NT search direction [18]. Monteiro and Zhang [16] established the polynomiality of a long-step path following method based on search directions defined by scaling matrices belonging to the class

$$\mathcal{P}(X, S) \equiv \{P \in \mathcal{S}_{++}^n : P^2 X S = S X P^2\}. \tag{8}$$

Following [16], Sheng, Potra, and Ji [23] proved the polynomiality of a Mizuno-Todd-Ye type predictor-corrector algorithm for SDO by imposing the scaling matrices to be chosen from the class

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

In what follows we describe the classic variant of Mehrotra's predictor-corrector algorithm. Let define

$$(X(\alpha), y(\alpha), S(\alpha)) = (X, y, S) + \alpha(\Delta X, \Delta y, \Delta S), \tag{9}$$

$$\mu_g(\alpha) = \frac{X(\alpha) \bullet S(\alpha)}{n}. \tag{10}$$

This algorithm uses the so-called negative infinity neighborhood:

$$\mathcal{N}_\infty^-(\gamma) = \{(X, y, S) \in \mathcal{F}^0(P) \times \mathcal{F}^0(D) : \lambda_{\min}[XS] \geq \gamma\mu_g\},$$

where $\gamma \in (0, 1)$ is a given constant. The algorithm, in the predictor step, computes the affine scaling search direction, i.e.,

$$A_i \bullet \Delta X^a = 0, \quad i = 1, \dots, m, \quad (11a)$$

$$\sum_{i=1}^m A_i \Delta y_i^a + \Delta S^a = 0, \quad (11b)$$

$$H_P(X \Delta S^a + \Delta X^a S) = -H_P(XS), \quad (11c)$$

then it computes the maximum feasible step size α_a that ensures $X(\alpha_a) = X + \alpha_a \Delta X^a$, $S(\alpha_a) = S + \alpha_a \Delta S^a \succeq 0$. However the algorithm does not take such a step. Based on this step size it chooses $\sigma = (1 - \alpha_a)^3$ to compute the centering direction that is defined as the solution of the system

$$A_i \bullet \Delta X = 0, \quad i = 1, \dots, m, \quad (12a)$$

$$\sum_{i=1}^m A_i \Delta y_i + \Delta S = 0, \quad (12b)$$

$$H_P(X \Delta S + \Delta X S) = \sigma \mu_g I - H_P(\Delta X^a \Delta S^a) - H_P(XS). \quad (12c)$$

Then the algorithm computes the maximum feasible step size α_c that keeps the next iteration in $\mathcal{N}_\infty^-(\gamma)$.

In case of LO, it has been shown in [22] by an example that Mehrotra's heuristic may force the algorithm to make very small steps to keep the iterates in a certain neighborhood of the central path which may cause too many iterations to converge. This may also happen when the affine scaling step size is very small in which case the algorithm might make pure centering steps that only marginally reduce the duality gap. Therefore a safeguard strategy has been used to have control on the minimal warranted step size. This variant of the algorithm for SDO can be stated as follows.

Algorithm 1: Mehrotra-type algorithm

input:

A proximity parameter $\gamma \in (0, \frac{1}{2})$;

an accuracy parameter $\epsilon > 0$;

$(X^0, y^0, S^0) \in \mathcal{N}_\infty^-(\gamma)$;

while $X \bullet S \geq \epsilon$ **do**

 Compute P corresponding to the NT direction from (7);

Predictor step

begin

 Solve (11) and compute the maximum step size α_a such that $(X(\alpha_a), y(\alpha_a), S(\alpha_a)) \in \mathcal{F}$;

end

Corrector step

begin

if $\alpha_a \geq 0.1$ **then**

 Solve (12) with $\sigma = (1 - \alpha_a)^3$ and

 compute the maximum step size α_c such that $(X(\alpha_c), y(\alpha_c), S(\alpha_c)) \in \mathcal{N}_\infty^-(\gamma)$;

end

if $\alpha_a < 0.1$ **or** $\alpha_c < \frac{2\gamma^2}{3n^2}$ **then**

 Solve (12) with $\sigma = \frac{\gamma}{1-\gamma}$ and

 compute the maximum step size α_c such that $(X(\alpha_c), y(\alpha_c), S(\alpha_c)) \in \mathcal{N}_\infty^-(\gamma)$;

end

end

end

3 Complexity analysis of the algorithm

In this section we are going to establish the iteration-complexity of Algorithm 1. We may write the complementarity equation (12c) in the form

$$H(\hat{X}\Delta\hat{S} + \Delta\hat{X}\hat{S}) = \sigma\mu_g I - H(\Delta\hat{X}^a\Delta\hat{S}^a) - H(\hat{X}\hat{S}) \quad (13)$$

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

$$\hat{X} \equiv PXP, \quad \Delta\hat{X} \equiv P\Delta XP, \quad \hat{S} \equiv P^{-1}SP^{-1}, \quad \Delta\hat{S} \equiv P^{-1}\Delta SP^{-1}. \quad (14)$$

Moreover, in terms of Kronecker product (see page 3.), equation (13) becomes

$$\hat{E} \operatorname{vec} \Delta\hat{X} + \hat{F} \operatorname{vec} \Delta\hat{S} = \operatorname{vec}(\sigma\mu_g I - H(\Delta\hat{X}^a\Delta\hat{S}^a) - H(\hat{X}\hat{S})), \quad (15)$$

where

$$\hat{E} \equiv \frac{1}{2}(\hat{S} \otimes I + I \otimes \hat{S}), \quad \hat{F} \equiv \frac{1}{2}(\hat{X} \otimes I + I \otimes \hat{X}). \quad (16)$$

Using (8) and (14), the set $\mathcal{P}(X, S)$ can be written as follows:

$$\mathcal{P}(X, S) = \{P \in \mathcal{S}_{++}^n : \hat{X}\hat{S} = \hat{S}\hat{X}\},$$

in which the commutativity of \hat{X} and \hat{S} implies that \hat{E} and \hat{S} commute too.

We need to find a lower bound for the maximum step size α_c in the corrector step in order to establish the iteration complexity of Algorithm 1. The following lemmas are needed to derive a lower bound on the size of the centering step.

Lemma 3.1. *Let $(X, y, S) \in \mathcal{S}_{++}^n \times \mathbb{R}^m \times \mathcal{S}_{++}^n$, and let $(\Delta X, \Delta y, \Delta S)$ be the solution of (12). Then*

$$H_P(X(\alpha)S(\alpha)) = (1 - \alpha)H_P(XS) + \alpha\sigma\mu_g I - \alpha H_P(\Delta X^a \Delta S^a) + \alpha^2 H_P(\Delta X \Delta S), \quad (17)$$

$$\mu_g(\alpha) = (1 - \alpha + \alpha\sigma)\mu_g. \quad (18)$$

Proof. By equation (9) we have

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

This expression, together with the linearity of $H_P(\cdot)$ and the complementarity equation (12c) implies that

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

and hence (17) holds. Using (10) and the identity $\text{Tr}(H_P(M)) = \text{Tr}(M)$, we obtain

$$\begin{aligned} X(\alpha) \bullet S(\alpha) &= \text{Tr}[(1 - \alpha)H_P(XS) + \alpha\sigma\mu_g I - \alpha H_P(\Delta X^a \Delta S^a) + \alpha^2 H_P(\Delta X \Delta S)] \\ &= (1 - \alpha)\text{Tr}(H_P(XS)) + \alpha\sigma\mu_g n - \alpha\text{Tr}(H_P(\Delta X^a \Delta S^a)) + \alpha^2 \text{Tr}(H_P(\Delta X \Delta S)) \\ &= (1 - \alpha)X \bullet S + \alpha\sigma\mu_g n - \alpha\Delta X^a \bullet \Delta S^a + \alpha^2 \Delta X \bullet \Delta S. \end{aligned} \quad (19)$$

Using the relations (12a) and (12b) and the fact that (X, y, S) is a primal-dual feasible solution, we can conclude that $\Delta X \bullet \Delta S = 0$. Thus dividing (19) by n gives (18). \square

Lemma 3.2. *Let $P \in \mathcal{P}(X, S)$ and $G \equiv \hat{E}^{-1}\hat{F}$. Then*

$$\|G^{-\frac{1}{2}} \text{vec } \Delta \hat{X}\|^2 + \|G^{\frac{1}{2}} \text{vec } \Delta \hat{S}\|^2 + 2\Delta \hat{X} \bullet \Delta \hat{S} \leq \text{cond}(G) \left\{ \left[\left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right) \gamma \right]^{\frac{1}{2}} + \frac{1}{4} \right\}^2 \frac{n^2 \mu_g}{\gamma}. \quad (20)$$

Proof. By applying Lemma A.3 to relation (15) we obtain

$$\begin{aligned} \|(\hat{F}\hat{E})^{-\frac{1}{2}}\hat{E} \text{vec } \Delta \hat{X}\|^2 + \|(\hat{F}\hat{E})^{-\frac{1}{2}}\hat{F} \text{vec } \Delta \hat{S}\|^2 + 2\Delta \hat{X} \bullet \Delta \hat{S} \\ = \|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(\sigma\mu_g I - H(\Delta \hat{X}^a \Delta \hat{S}^a) - H(\hat{X}\hat{S}))\|^2. \end{aligned}$$

Since $P \in \mathcal{P}(X, S)$, so \hat{E} and \hat{F} commute, which implies that

$$(\hat{F}\hat{E})^{-\frac{1}{2}}\hat{E} = \hat{E}^{-1}\hat{F}^{-\frac{1}{2}} = G^{-\frac{1}{2}}, \quad (\hat{F}\hat{E})^{-\frac{1}{2}}\hat{F} = \hat{E}^{-1}\hat{F}^{-\frac{1}{2}} = G^{\frac{1}{2}}. \quad (21)$$

Hence for the proof of (20) we have

$$\begin{aligned} & \|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(\sigma\mu_g I - H(\Delta\hat{X}^a \Delta\hat{S}^a) - H(\hat{X}\hat{S}))\|^2 \\ & \leq \|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(\sigma\mu_g I - H(\hat{X}\hat{S}))\|^2 + \|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(H(\Delta\hat{X}^a \Delta\hat{S}^a))\|^2 \\ & \quad + 2\|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(\sigma\mu_g I - H(\hat{X}\hat{S}))\| \|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(H(\Delta\hat{X}^a \Delta\hat{S}^a))\|. \end{aligned}$$

The upper bound for the first expression of the right hand side is concluded by (39) in Lemma A.6. A bound for the second expression follows from (41) of Corollary A.8, where $\|A\| = (\rho(A^T A))^{1/2}$, and (35) of Lemma A.2.

$$\begin{aligned} \|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(H(\Delta\hat{X}^a \Delta\hat{S}^a))\|^2 & \leq \|(\hat{F}\hat{E})^{-\frac{1}{2}}\|^2 \| \text{vec}(H(\Delta\hat{X}^a \Delta\hat{S}^a)) \|^2 \\ & = \rho((\hat{F}\hat{E})^{-1}) \| \text{vec}(H(\Delta\hat{X}^a \Delta\hat{S}^a)) \|^2 = \rho((\hat{F}\hat{E})^{-1}) \|H(\Delta\hat{X}^a \Delta\hat{S}^a)\|_F^2 \\ & \leq \frac{1}{4\lambda_1} \|H(\Delta\hat{X}^a \Delta\hat{S}^a)\|_F^2 \leq \frac{1}{4\lambda_1} \left[\frac{\sqrt{\text{cond}(G)}}{2} n\mu_g \right]^2 \\ & \leq \frac{\text{cond}(G)}{4} \frac{n^2\mu_g^2}{4\gamma\mu_g} \leq \frac{\text{cond}(G)}{16} \frac{n^2\mu_g}{\gamma}. \end{aligned} \quad (22)$$

For the third expression, (39) and (22) implies

$$\begin{aligned} \|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(\sigma\mu_g I - H(\hat{X}\hat{S}))\| \|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(H(\Delta\hat{X}^a \Delta\hat{S}^a))\| & \leq \left[\left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right) n\mu_g \frac{\text{cond}(G)}{16} \frac{n^2\mu_g}{\gamma} \right]^{\frac{1}{2}} \\ & \leq \left[\left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right) \frac{\text{cond}(G)}{16} \frac{1}{\gamma} \right]^{\frac{1}{2}} n^2\mu_g \\ & \leq \frac{\sqrt{\text{cond}(G)}}{4\sqrt{\gamma}} \left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right)^{\frac{1}{2}} n^2\mu_g. \end{aligned} \quad (23)$$

From (22), (23) and (39) we obtain

$$\begin{aligned} & \|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(\sigma\mu_g I - H(\Delta\hat{X}^a \Delta\hat{S}^a) - H(\hat{X}\hat{S}))\|^2 \\ & \leq \left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right) n\mu_g + \frac{\text{cond}(G)n^2\mu_g}{16\gamma} + 2\frac{\sqrt{\text{cond}(G)}}{4\sqrt{\gamma}} \left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right)^{\frac{1}{2}} n^2\mu_g \\ & \leq \text{cond}(G) \left\{ \left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right) + \frac{1}{16\gamma} + \frac{1}{2\sqrt{\gamma}} \left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right)^{\frac{1}{2}} \right\} n^2\mu_g \\ & \leq \text{cond}(G) \left\{ \left[\left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right) \gamma \right]^{\frac{1}{2}} + \frac{1}{4} \right\}^2 \frac{n^2\mu_g}{\gamma}, \end{aligned}$$

that completes the proof. \square

The following lemma gives an upper bound for the second order term.

Lemma 3.3. *Let a point $(X, y, S) \in \mathcal{N}_\infty^-(\gamma)$ and $P \in \mathcal{P}(X, Y)$ be given, and define $G = \hat{E}^{-1}\hat{F}$. Then the Newton step corresponding system (12) satisfies*

$$\|H_P(\Delta X \Delta S)\|_F \leq \frac{(\text{cond}(G))^{\frac{3}{2}}}{2} \left\{ \left[\left(1 - 2\sigma + \frac{\sigma^2}{\gamma} \right) \gamma \right]^{\frac{1}{2}} + \frac{1}{4} \right\}^2 \frac{n^2 \mu_g}{\gamma}.$$

Proof. We have

$$\begin{aligned} \|H_P(\Delta X \Delta S)\|_F &= \|H_I(\Delta \hat{X} \Delta \hat{S})\|_F \leq \|\Delta \hat{X} \Delta \hat{S}\|_F \leq \|\Delta \hat{X}\|_F \|\Delta \hat{S}\|_F \\ &= \|\text{vec}(\Delta \hat{X})\| \|\text{vec}(\Delta \hat{S})\| \end{aligned} \quad (24)$$

Using Lemmas A.4, (20) and $\Delta \hat{X} \bullet \Delta \hat{S} = 0$, we obtain

$$\begin{aligned} \|\text{vec}(\Delta \hat{X})\| \|\text{vec}(\Delta \hat{S})\| &\leq \frac{\text{cond}(G)}{2} (\|G^{-\frac{1}{2}} \text{vec} \Delta \hat{X}\|^2 + \|G^{\frac{1}{2}} \text{vec} \Delta \hat{S}\|^2) \\ &\leq \frac{(\text{cond}(G))^{\frac{3}{2}}}{2} \left\{ \left[\left(1 - 2\sigma + \frac{\sigma^2}{\gamma} \right) \gamma \right]^{\frac{1}{2}} + \frac{1}{4} \right\}^2 \frac{n^2 \mu_g}{\gamma}. \end{aligned} \quad (25)$$

The result follows by combining (24) and (25). \square

As a consequence, we obtain the following result for a proper choice of constant σ as a function of parameter γ .

Corollary 3.4. *If $\sigma = \frac{\gamma}{1-\gamma}$ where $0 \leq \gamma \leq \frac{1}{2}$ then*

$$\|H_P(\Delta X \Delta S)\|_F \leq \frac{3(\text{cond}(G))^{\frac{3}{2}} n^2 \mu_g}{2 \gamma}.$$

For NT scaling, we have $\hat{X} = \hat{S}$ and consequently $\hat{E} = \hat{F}$, which implies $\text{cond}(G) = 1$. This specific choice gives the following corollary.

Corollary 3.5. *Let P be the NT scaling. Then*

$$\|H_P(\Delta X \Delta S)\|_F \leq \frac{3 n^2 \mu_g}{2 \gamma}. \quad (26)$$

The difficulty of analyzing Mehrortra-type algorithms arises from the second order term $\Delta X^a \Delta S^a$ in the corrector step of the algorithm, which is one of the main differences of this algorithm and other IPM algorithms. To overcome this difficulty the following lemma is an important technical result that gives the relationship between this term and the matrix XS . The result is used in the proof of the next theorem.

Lemma 3.6. *Let P be the NT scaling matrix and t be defined as follows:*

$$t = \max_{\|u\|=1} \left\{ \frac{u^T H_P(\Delta X^a \Delta S^a) u}{u^T H_P(XS) u} \right\}. \quad (27)$$

Then t satisfies $t \leq \frac{1}{4}$.

Proof. For the NT direction the scaling matrix is $P = [X^{-\frac{1}{2}}(X^{\frac{1}{2}}SX^{\frac{1}{2}})^{\frac{1}{2}}X^{-\frac{1}{2}}]^{\frac{1}{2}} = [S^{\frac{1}{2}}(S^{\frac{1}{2}}XS^{\frac{1}{2}})^{-\frac{1}{2}}S^{\frac{1}{2}}]^{\frac{1}{2}}$ and then from [5] we have

$$D_X + D_S = -V,$$

where $D_X = P\Delta XP$ and $D_S = P^{-1}\Delta SP^{-1}$ and $PXP = P^{-1}SP^{-1} := V$. So, $V^2 = PXSP^{-1}$, i.e., they are similar, and

$$\frac{1}{2}(D_X D_S + D_S D_X) = \frac{1}{4}[(D_X + D_S)^2 - (D_X - D_S)^2],$$

which implies

$$\frac{1}{2}(D_X D_S + D_S D_X) \preceq \frac{1}{4}V^2.$$

It follows that

$$(P\Delta X^a \Delta S^a P^{-1} + P^{-1}\Delta S^a \Delta X^a P) \preceq \frac{1}{4}(PXSP^{-1} + P^{-1}SXP),$$

i.e.,

$$H_P(\Delta X^a \Delta S^a) \preceq \frac{1}{4}H_P(XS).$$

Thus for the NT direction for $\forall u, \|u\| = 1$, we have $4u^T H_P(\Delta X^a \Delta S^a)u \leq u^T H_P(XS)u$, implying

$$t = \max_{\|u\|=1} \left\{ \frac{u^T H_P(\Delta X^a \Delta S^a)u}{u^T H_P(XS)u} \right\} \leq \frac{1}{4}.$$

□

The following theorem gives a lower bound for the maximum feasible step size α_c in the corrector step.

Theorem 3.7. *Suppose that the current iterate $(X, y, S) \in \mathcal{N}_{\infty}^-(\gamma)$, where $\gamma \in (0, \frac{1}{2})$ and $(\Delta X, \Delta y, \Delta S)$ is the solution of (12) with $\sigma = \frac{\gamma}{1-\gamma}$. Then the maximum step size α_c , that keeps $(X(\alpha_c), y(\alpha_c), S(\alpha_c))$ in $\mathcal{N}_{\infty}^-(\gamma)$ satisfies*

$$\alpha_c \geq \frac{2\gamma^2}{3n^2}. \quad (28)$$

Proof. If inequality (28) holds for an $\alpha > \frac{1}{1+t}$, where t is defined by (27), we are done. Otherwise, it follows from (17) and from the fact that $\lambda_{\min}(\cdot)$ is a homogeneous concave function on the space of symmetric matrices [8], that

$$\begin{aligned} \lambda_{\min}(H_P(X(\alpha)S(\alpha))) &= \lambda_{\min}(\alpha\sigma\mu_g I + (1-\alpha)H_P(XS) - \alpha H_P(\Delta X^a \Delta S^a) + \alpha^2 H_P(\Delta X \Delta S)) \\ &\geq \alpha\sigma\mu_g + \lambda_{\min}((1-\alpha)H_P(XS) - \alpha H_P(\Delta X^a \Delta S^a)) + \alpha^2 \lambda_{\min}(H_P(\Delta X \Delta S)). \end{aligned}$$

Let

$$Q(\alpha) = (1-\alpha)H_P(XS) - \alpha H_P(\Delta X^a \Delta S^a).$$

Since $Q(\alpha)$ is symmetric so we have $\lambda_{\min}(Q(\alpha)) = \min_{\|u\|=1} u^T Q(\alpha)u$, therefore there is a vector \bar{u} with $\|\bar{u}\| = 1$, such that, $\lambda_{\min}(Q(\alpha)) = \bar{u}^T Q(\alpha)\bar{u}$, which implies

$$\begin{aligned} \lambda_{\min}(H_P(X(\alpha)S(\alpha))) &\geq \alpha\sigma\mu_g + \bar{u}^T [(1-\alpha)H_P(XS) - \alpha H_P(\Delta X^a \Delta S^a)]\bar{u} + \alpha^2 \lambda_{\min}(H_P(\Delta X \Delta S)) \\ &\geq \alpha\sigma\mu_g + (1-\alpha)\bar{u}^T H_P(XS)\bar{u} - \alpha\bar{u}^T H_P(\Delta X^a \Delta S^a)\bar{u} + \alpha^2 \lambda_{\min}(H_P(\Delta X \Delta S)). \end{aligned}$$

The fact that $H_P(XS)$ is positive definite and $\text{Tr}(H_P(\Delta X^a \Delta S^a)) = 0$ imply $t \geq 0$ in (27) and thus it follows that

$$u^T H_P(\Delta X^a \Delta S^a) u \leq t u^T H_P(XS) u, \quad \forall u, \quad \|u\| = 1, \quad (29)$$

which enables us to derive

$$\begin{aligned} \lambda_{\min}(H_P(X(\alpha)S(\alpha))) &\geq \alpha \sigma \mu_g + (1 - \alpha) \bar{u}^T H_P(XS) \bar{u} - \alpha t \bar{u}^T H_P(XS) \bar{u} + \alpha^2 \lambda_{\min}(H_P(\Delta X \Delta S)) \\ &\geq \alpha \sigma \mu_g + [1 - \alpha(1 + t)] \bar{u}^T H_P(XS) \bar{u} + \alpha^2 \lambda_{\min}(H_P(\Delta X \Delta S)) \\ &\geq \alpha \sigma \mu_g + [1 - \alpha(1 + t)] \lambda_{\min}(H_P(XS)) + \alpha^2 \lambda_{\min}(H_P(\Delta X \Delta S)), \end{aligned}$$

where the last inequality follows from $[1 - \alpha(1 + t)] \geq 0$. Since $\|M\| \leq \|M\|_F$, for all $M \in \mathbb{R}^{n \times n}$ [8] and the previous iteration is in $\mathcal{N}_{\infty}^-(\gamma)$, by using the relation (26) one has

$$\begin{aligned} \lambda_{\min}(H_P(X(\alpha)S(\alpha))) &\geq \alpha \sigma \mu_g + [1 - \alpha(1 + t)] \gamma \mu_g - \alpha^2 \|(H_P(\Delta X \Delta S))\| \\ &\geq \alpha \sigma \mu_g + [1 - \alpha(1 + t)] \gamma \mu_g - \alpha^2 \frac{3n^2 \mu_g}{2\gamma}. \end{aligned}$$

In order the iterations to stay in $\mathcal{N}_{\infty}^-(\gamma)$, by Lemma A.5 it is sufficient to have

$$\alpha \sigma \mu_g + [1 - \alpha(1 + t)] \gamma \mu_g - \alpha^2 \frac{3n^2 \mu_g}{2\gamma} \geq \gamma(1 - \alpha + \alpha \sigma) \mu_g,$$

that certainly holds whenever

$$\alpha \leq \frac{2\gamma[(1 - \gamma)\sigma - \gamma t]}{3n^2} = \frac{2\gamma^2[1 - t]}{3n^2} \leq \frac{2\gamma^2}{3n^2},$$

and this completes the proof. \square

Now we are ready to give the iteration-complexity of Algorithm 1.

Theorem 3.8. *Algorithm 1 stops after at most*

$$\mathcal{O}\left(n^2 \log \frac{X^0 \bullet S^0}{\epsilon}\right)$$

iterations with a solution for which $X \bullet S \leq \epsilon$.

Proof. If $\alpha_a < 0.1$ or $\alpha_c < \frac{\gamma^2}{2n^2}$, then the algorithm uses the safeguard strategy, and by Theorem 3.7 and relation (18) we have

$$\mu_g(\alpha) \leq \left(1 - \frac{2\gamma^2(1 - 2\gamma)}{3(1 - \gamma)n^2}\right) \mu_g.$$

If $\alpha_a \geq 0.1$ and $\alpha_c \geq \frac{\gamma^2}{2n^2}$, then the algorithm uses Mehrotra's updating strategy, and thus we have

$$\mu_g(\alpha) \leq \left(1 - \frac{\gamma^2}{6n^2}\right) \mu_g.$$

This completes the proof by Lemma I.36 of [21]. \square

4 Modified version of Algorithm 1

In this section we state a modified version of Algorithm 1 and analyze its worst case complexity. It will be proved that this modified version has better iteration complexity.

The following system is solved in the corrector step of the modified version of Algorithm 1

$$\begin{aligned} A_i \bullet \Delta X &= 0, \quad i = 1, \dots, m, \\ \sum_{i=1}^m A_i \Delta y_i + \Delta S &= 0, \\ H_P(X \Delta S + \Delta X S) &= \sigma \mu_g I - \alpha_a H_P(\Delta X^a \Delta S^a) - H_P(XS), \end{aligned} \quad (30)$$

where $\mu_g = X \bullet S/n$, $(\Delta X^a, \Delta y^a, \Delta S^a)$ is the search direction from the predictor step, and α_a is the maximum feasible step size in the predictor step. Using (14) and (16), in terms of Kronecker product, the complementarity equation in (30) may be written in the following way:

$$\hat{E} \operatorname{vec} \Delta \hat{X} + \hat{F} \operatorname{vec} \Delta \hat{S} = \operatorname{vec}(\sigma \mu_g I - \alpha_a H(\Delta \hat{X}^a \Delta \hat{S}^a) - H(\hat{X} \hat{S})). \quad (31)$$

The next theorem gives a lower bound for the maximum feasible step size in the worst case for the predictor step.

Theorem 4.1. *Suppose the current iterate $(X, y, S) \in \mathcal{N}_\infty^-$ and let $(\Delta X^a, \Delta y^a, \Delta S^a)$ be the solution of (11). Then the maximum feasible step size, α_a , so that $X(\alpha_a), S(\alpha_a) \in \mathcal{S}_+^n$, satisfies*

$$\alpha_a \geq \frac{\sqrt{\gamma^2 + 2\gamma n \sqrt{\operatorname{cond}(G)} - \gamma}}{n \sqrt{\operatorname{cond}(G)}}. \quad (32)$$

Proof. It follows from (41) given in Corollary A.8 that

$$\lambda_{\min}(\Delta X^a \Delta S^a) \geq -\|\Delta X^a \Delta S^a\|_F \geq -\frac{\sqrt{\operatorname{cond}(G)}}{2} n \mu_g.$$

These inequalities, together with the fact that $(X, y, S) \in \mathcal{N}_\infty^-(\gamma)$ give

$$\begin{aligned} \lambda_{\min}(X(\alpha_a)S(\alpha_a)) &\geq (1 - \alpha_a) \lambda_{\min}(XS) + \alpha_a^2 \lambda_{\min}(\Delta X^a \Delta S^a) \\ &\geq \left[1 - \alpha_a - \alpha_a^2 \frac{\sqrt{\operatorname{cond}(G)}}{2\gamma} n \right] \gamma \mu_g. \end{aligned}$$

To make sure that $\lambda_{\min}(X(\alpha_a)S(\alpha_a)) \geq 0$, it suffices to require that

$$\left[1 - \alpha_a - \alpha_a^2 \frac{\sqrt{\operatorname{cond}(G)}}{2\gamma} n \right] \gamma \mu_g \geq 0,$$

which is equivalent to

$$n \sqrt{\operatorname{cond}(G)} \alpha_a^2 + 2\gamma \alpha_a - 2\gamma \leq 0.$$

This inequality holds when

$$\alpha_a \in \left[\frac{-\sqrt{\gamma^2 + 2\gamma n \sqrt{\operatorname{cond}(G)} - \gamma}}{n \sqrt{\operatorname{cond}(G)}}, \frac{\sqrt{\gamma^2 + 2\gamma n \sqrt{\operatorname{cond}(G)} - \gamma}}{n \sqrt{\operatorname{cond}(G)}} \right],$$

which implies

$$\det(X(\alpha)S(\alpha)) > 0, \forall 0 \leq \alpha \leq \bar{\alpha},$$

where

$$\bar{\alpha} = \frac{\sqrt{\gamma^2 + 2\gamma n\sqrt{\text{cond}(G)} - \gamma}}{n\sqrt{\text{cond}(G)}},$$

and this completes the proof according to Lemma A.9. \square

The following lemma is useful for finding the maximum feasible step size in the corrector step.

Lemma 4.2. *Let a point $(X, y, S) \in \mathcal{N}_\infty^-(\gamma)$ and $P \in \mathcal{P}(X, Y)$ be given, and define $G = \hat{E}^{-1}\hat{F}$. Then the Newton step corresponding to system (30) satisfies*

$$\|H_P(\Delta X \Delta S)\|_F \leq \frac{\text{cond}(G)}{2} \left[\left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right)^{\frac{1}{2}} + \frac{1}{2} \right]^2 n\mu_g.$$

Proof. For the proof we need to show that

$$\|G^{-\frac{1}{2}} \text{vec } \Delta \hat{X}\|^2 + \|G^{\frac{1}{2}} \text{vec } \Delta \hat{S}\|^2 + 2\Delta \hat{X} \bullet \Delta \hat{S} \leq \sqrt{\text{cond}(G)} \left[\left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right)^{\frac{1}{2}} + \frac{1}{2} \right]^2 n\mu_g, \quad (33)$$

which together with (24), Lemma A.4, and $\Delta \hat{X} \bullet \Delta \hat{S} = 0$ gives the result. The proof for (33) is analogous to the proof of Lemma 3.2. Applying Lemma A.3 to relation (31), and using (21) we have

$$\begin{aligned} \|G^{-\frac{1}{2}} \text{vec } \Delta \hat{X}\|^2 + \|G^{\frac{1}{2}} \text{vec } \Delta \hat{S}\|^2 + 2\Delta \hat{X} \bullet \Delta \hat{S} = \\ \|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(\sigma\mu_g I - \alpha_a H(\Delta \hat{X}^a \Delta \hat{S}^a) - H(\hat{X}\hat{S}))\|^2. \end{aligned}$$

So it remains to show that

$$\|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(\sigma\mu_g I - \alpha_a H(\Delta \hat{X}^a \Delta \hat{S}^a) - H(\hat{X}\hat{S}))\|^2 \leq \sqrt{\text{cond}(G)} \left[\left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right)^{\frac{1}{2}} + \frac{1}{2} \right]^2 n\mu_g.$$

We have

$$\begin{aligned} & \|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(\sigma\mu_g I - \alpha_a H(\Delta \hat{X}^a \Delta \hat{S}^a) - H(\hat{X}\hat{S}))\|^2 \\ & \leq \|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(\sigma\mu_g I - H(\hat{X}\hat{S}))\|^2 + \|\alpha_a (\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(H(\Delta \hat{X}^a \Delta \hat{S}^a))\|^2 \\ & \quad + 2\|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(\sigma\mu_g I - H(\hat{X}\hat{S}))\| \|\alpha_a (\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(H(\Delta \hat{X}^a \Delta \hat{S}^a))\|. \end{aligned}$$

By using (32), the upper bound for each part can be obtained in the same way as in the proof of Lemma

3.2. So we may derive

$$\begin{aligned}
& \|(\hat{F}\hat{E})^{-\frac{1}{2}} \text{vec}(\sigma\mu_g I - \alpha_a H(\Delta\hat{X}^a \Delta\hat{S}^a) - H(\hat{X}\hat{S}))\|^2 \\
& \leq \left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right) n\mu_g + \frac{\sqrt{\text{cond}(G)}}{4} n\mu_g + \sqrt[4]{\text{cond}(G)} \left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right)^{\frac{1}{2}} n\mu_g \\
& \leq \sqrt{\text{cond}(G)} \left[\left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right) + \frac{1}{4} + \left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right)^{\frac{1}{2}} \right] n\mu_g \\
& \leq \sqrt{\text{cond}(G)} \left[\left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right)^{\frac{1}{2}} + \frac{1}{2} \right]^2 n\mu_g.
\end{aligned}$$

□

The following corollary follows from the previous lemma by a specific choice of σ and γ .

Corollary 4.3. *If $\sigma = \frac{\gamma}{1-\gamma}$ with $0 \leq \gamma \leq \frac{1}{2}$ then*

$$\|H_P(\Delta X \Delta S)\|_F \leq 2 \text{cond}(G) n\mu_g.$$

The use of the NT scaling matrix gives $\text{cond}(G) = 1$, and consequently the following corollary holds.

Corollary 4.4. *Let P be the NT scaling matrix, then*

$$\|H_P(\Delta X \Delta S)\|_F \leq 2n\mu_g. \quad (34)$$

Theorem 4.5. *Suppose that the current iterate $(X, y, S) \in \mathcal{N}_\infty^-(\gamma)$, where $\gamma \in (0, \frac{1}{2})$, and let $(\Delta X, \Delta y, \Delta S)$ be the solution of (30) with*

$$\sigma = \frac{\gamma}{1-\gamma}.$$

Then the maximum step size α_c , that keeps $(X(\alpha_c), y(\alpha_c), S(\alpha_c))$ in $\mathcal{N}_\infty^-(\gamma)$ satisfies

$$\alpha_c \geq \frac{3\gamma}{8n}.$$

Proof. If the previous inequality is true for $\alpha > \frac{1}{1+\alpha_a t}$, in which t is defined by (27), we are done. Otherwise, analogous to the proof of Theorem 3.16, we have,

$$\begin{aligned}
\lambda_{\min}(H_P(X(\alpha)S(\alpha))) &= \lambda_{\min}(\alpha\sigma\mu_g I + (1-\alpha)H_P(XS) - \alpha H_P(\alpha_a \Delta X^a \Delta S^a) + \alpha^2 H_P(\Delta X \Delta S)) \\
&\geq \alpha\sigma\mu_g + \bar{u}^T [(1-\alpha)H_P(XS) - \alpha H_P(\alpha_a \Delta X^a \Delta S^a)] \bar{u} + \alpha^2 \lambda_{\min}(H_P(\Delta X \Delta S))
\end{aligned}$$

where \bar{u} is the eigenvector corresponding to the smallest eigenvalue of $(1-\alpha)H_P(XS) - \alpha H_P(\alpha_a \Delta X^a \Delta S^a)$.

It follows from (29) that

$$\begin{aligned}
\lambda_{\min}(H_P(X(\alpha)S(\alpha))) &\geq \alpha\sigma\mu_g + (1-\alpha)\bar{u}^T H_P(XS)\bar{u} - \alpha\alpha_a t \bar{u}^T H_P(XS)\bar{u} + \alpha^2 \lambda_{\min}(H_P(\Delta X \Delta S)) \\
&\geq \alpha\sigma\mu_g + [1 - \alpha(1 + \alpha_a t)] \bar{u}^T H_P(XS)\bar{u} + \alpha^2 \lambda_{\min}(H_P(\Delta X \Delta S)) \\
&\geq \alpha\sigma\mu_g + [1 - \alpha(1 + \alpha_a t)] \lambda_{\min}(H_P(XS)) + \alpha^2 \lambda_{\min}(H_P(\Delta X \Delta S)),
\end{aligned}$$

where the last inequality follows from $[1 - \alpha(1 + \alpha_a t)] \geq 0$. The facts that $\|M\| \leq \|M\|_F$ for all $M \in \mathbb{R}^{n \times n}$ [8], the previous iteration is in $\mathcal{N}_\infty^-(\gamma)$, and using inequality (34) imply

$$\begin{aligned} \lambda_{\min}(H_P(X(\alpha)S(\alpha))) &\geq \alpha\sigma\mu_g + [1 - \alpha(1 + \alpha_a t)]\gamma\mu_g - \alpha^2\|(H_P(\Delta X \Delta S))\| \\ &\geq \alpha\sigma\mu_g + [1 - \alpha(1 + \alpha_a t)]\gamma\mu_g - \alpha^2 2n\mu_g. \end{aligned}$$

In order to stay in $\mathcal{N}_\infty^-(\gamma)$, from (18) and Lemma A.5 we have to have

$$\alpha\sigma\mu_g + [1 - \alpha(1 + \alpha_a t)]\gamma\mu_g - \alpha^2 2n\mu_g \geq \gamma(1 - \alpha + \alpha\sigma)\mu_g,$$

that certainly holds whenever

$$\alpha \leq \frac{(1 - \gamma)\sigma - \gamma\alpha_a t}{2n} = \frac{(1 - \alpha_a t)\gamma}{2n} \leq \frac{3\gamma}{8n},$$

and this completes the proof of the theorem. \square

Theorem 4.6. *The modified version of Algorithm 1 stops after at most*

$$\mathcal{O}\left(n \log \frac{X^0 \bullet S^0}{\epsilon}\right)$$

iterations with a solution for which $X \bullet S \leq \epsilon$.

Proof. If $\alpha_a < 0.1$ or $\alpha_c < \frac{3\gamma}{8n}$, then the algorithm uses the safeguard strategy, and by the definition (18) and Theorem 4.6 we have

$$\mu_g(\alpha) \leq \left(1 - \frac{3\gamma(1 - 2\gamma)}{8(1 - \gamma)n}\right) \mu_g.$$

If $\alpha_a \geq 0.1$ and $\alpha_c \geq \frac{3\gamma}{8n}$, then the algorithm uses the Mehrotra's updating strategy, and thus we have

$$\mu_g(\alpha) \leq \left(1 - \frac{\gamma}{10n}\right) \mu_g.$$

This completes the proof by Lemma I.36. of [21]. \square

5 Conclusion

In this paper, we have extended the recently proposed Mehrotra-type predictor corrector algorithm of Salahi, Peng and Terlaky [22] to SDO. We used the symmetrization operator, introduced by Zhang [30], and the NT scaling [18] matrix. We observe that the iteration-complexity of the resulting algorithm for the SDO case is $\mathcal{O}\left(n^2 \log \frac{X^0 \bullet S^0}{\epsilon}\right)$, that is analogous to the LO case. By slightly modifying the algorithm the iteration-complexity of the algorithm is improved to $\mathcal{O}\left(n \log \frac{X^0 \bullet S^0}{\epsilon}\right)$.

References

- [1] F. Alizadeh, *Combinatorial Optimization with Interior-Point Methods and Semi-Definite Matrices*, Ph.D. Thesis, Computer Science Department, University of Minnesota, Minneapolis, MN, 1991.
- [2] ———, *Interior point methods in semidefinite programming with applications to combinatorial optimization*, *SIAM Journal on Optimization* **5** (1995), 13–51.
- [3] F. Alizadeh, J.A Haeberly, and M. Overton, *Primal-dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results*, *SIAM Journal on Optimization* **8** (1998), 746–768.
- [4] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM, Philadelphia, PA, 1994.
- [5] E. de Klerk, *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002.
- [6] E. de Klerk, C. Roos, and T. Terlaky, *Initialization in semidefinite programming via a self-dual, skew-symmetric embedding*, *OR Letters* **20** (1997), 213–221.
- [7] C. Helmberg, F. Rendl, R.J. Vanderbei, and H. Wolkowicz, *An interior-point method for semidefinite programming*, *SIAM Journal on Optimization* **6** (1996), 342–361.
- [8] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, New York, 1990.
- [9] F. Jarre, *An interior-point method for minimizing the maximum eigenvalue of a linear combination of matrices*, *SIAM Journal on Control and Optimization* **31** (1993), 1360–1377.
- [10] M. Kojima, M. Shida, and S. Shindoh, *Local convergence of predictor-corrector infeasible-interior-point algorithms for SDPs and SDLCPs*, *Mathematical Programming* **80** (1998), 129–160.
- [11] M. Kojima, S. Shindoh, and S. Hara, *Interior point methods for the monotone semidefinite linear complementarity problem in symmetric matrices*, *SIAM Journal on Optimization* **7** (1997), 86–125.
- [12] Z.-Q. Luo, J.F. Sturm, and S. Zhang, *Superlinear convergence of a symmetric primal-dual path-following algorithm for semidefinite programming*, *SIAM Journal on Optimization* **8** (1998), 59–81.
- [13] S. Merhrotra, *On the implementation of a (primal-dual) interior point method*, *SIAM Journal on Optimization* **2** (1992), 575–601.
- [14] S. Mizuno, M.J. Todd, and Y. Ye, *On adaptive step primal-dual interior-point algorithms for linear programming*, *Mathematics of Operations Research* **18** (1993), 945–981.
- [15] R.D.C. Monteiro, *Primal-dual path-following algorithms for semidefinite programming*, *SIAM journal on Optimization* **7** (1997), 663–678.

- [16] R.D.C. Monteiro and Y. Zhang, *A unified analysis for a class of long-step primal-dual path-following interior-point algorithms for semidefinite programming*, *Mathematical Programming* **81** (1998), 281–299.
- [17] Y.E. Nesterov and A.S. Nemirovsky, *Interior Point Methods in Convex Programming: Theory and Applications*, SIAM, Philadelphia, PA, 1994.
- [18] Y.E. Nesterov and M.J. Todd, *Self-scaled barriers and interior-point methods for convex programming*, *Mathematics of Operations Research* **22** (1997), 1–42.
- [19] _____, *Primal-dual interior-point methods for self-scaled cones*, *SIAM Journal on Optimization* **8** (1998), 324–364.
- [20] F.A. Potra and R. Sheng, *A superlinearly convergent primal-dual infeasible-interior-point algorithm for semidefinite programming*, *SIAM Journal on Optimization* **8** (1998), 1007–1028.
- [21] C. Roos, T. Terlaky, and J.-Ph. Vial, *Interior Point Methods for Linear Optimization*, second ed., Springer Science, Boston, USA, 2006.
- [22] M. Salahi, J. Peng, and T. Terlaky, *On Mehrotra-type predictor-corrector algorithms*, Technical report 2005/4, Advanced Optimization Lab., Department of Computing and Software, McMaster University, Hamilton, ON, Canada.
- [23] 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*, Report. comput. math. 89, Department of Mathematics, The University of Iowa, Iowa City, IA, 1996.
- [24] J.F. Sturm, *Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones*, *Optimization Methods and Software* **11-12** (1999), 625–653.
- [25] J.F. Sturm and S. Zhang, *On the long-step path-following method for semidefinite programming*, *Operations Research Letters* **22** (1998), 145–150.
- [26] _____, *Symmetric primal-dual path-following algorithms for semidefinite programming*, *Applied Numerical Mathematics* (1999), 301–315.
- [27] M.J. Todd, K.C. Toh, and R.H. Tütüncü, *On the Nesterov-Todd direction in semidefinite programming*, *SIAM Journal on Optimization* **8** (1998), 769–796.
- [28] L. Vandenberghe and S. Boyd, *A primal-dual potential reduction method for problems involving matrix inequalities*, *Mathematical Programming* **69** (1995), 205–236.
- [29] Y. Ye, *A class of projective transformations for linear programming*, *SIAM Journal on Computing* **19** (1990), 457–466.
- [30] Y. Zhang, *On extending some primal-dual interior-point algorithms from linear programming to semidefinite programming*, *SIAM Journal on Optimization* **8** (1998), 365–386.

A Appendix

The following results, introduced in [16], are used during the analysis.

Proposition A.1. *For any $P \in \mathcal{P}(X, S)$ there exists an orthogonal matrix Q_P and diagonal matrices $\Lambda(\hat{X})$ and $\Lambda(\hat{S})$ such that:*

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

The following lemma is proved based on the decomposition of the matrix $(\hat{F}\hat{E})^{-1}$, as it is stated in [16].

Lemma A.2. *Let λ_1 be the smallest eigenvalue of the matrix $\hat{X}\hat{S}$. Then for any $P \in \mathcal{P}(X, S)$ one has*

$$\rho((\hat{F}\hat{E})^{-1}) = \frac{1}{4\lambda_1}. \quad (35)$$

Proof. Using (17) and result *ii.* of Proposition A.1, we find the spectral decomposition of \hat{E} to be

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

In the same way, using result *i.* of Proposition A.1, the spectral decomposition of \hat{F} is

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

Using this equality and result *iii.* of Proposition A.1 we obtain

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

where the matrix in the middle is diagonal with the property that its first diagonal element is $1/(4\lambda_1)$. This matrix is also the inverse of the summation of the four diagonal matrix which have λ_1 as a smallest eigenvalue. So $1/(4\lambda_1)$ is the largest element of the matrix in the middle and consequently the largest eigenvalue of $(\hat{F}\hat{E})^{-1}$ what completes the proof. \square

Lemma A.3. *Let $u, v, r, \in \mathbb{R}^n$ and $E, F \in \mathbb{R}^{n \times n}$ satisfy $Eu + Fv = r$. If $FE^T \in \mathcal{S}_{++}$, then*

$$\|(FE^T)^{-1/2}Eu\|^2 + \|(FE^T)^{-1/2}Fv\|^2 + 2u^T v = \|(FE^T)^{-1/2}r\|^2. \quad (36)$$

Lemma A.4. *For any $u, v \in \mathbb{R}^n$ and $G \in \mathcal{S}_{++}^n$ we have*

$$\|u\|\|v\| \leq \frac{\sqrt{\text{cond}(G)}}{2}(\|G^{-1/2}u\|^2 + \|G^{1/2}v\|^2). \quad (37)$$

Let the spectrum of XS be $\{\lambda_i : i = 1, \dots, n\}$. Then following lemma holds.

Lemma A.5. Suppose that $(X, y, S) \in \mathcal{S}_{++}^n \times \mathcal{S}_{++}^n \times \mathbb{R}^m$, $P \in \mathcal{S}_{++}^n$, and $Q \in \mathcal{P}(X, S)$. Then

$$\lambda_{\min}[H_P(XS)] \leq \lambda_{\min}[XS] = \lambda_{\min}[H_Q(XS)]. \quad (38)$$

Lemma A.6. Let $P \in \mathcal{P}(X, S)$ be given. Then

$$\|(\hat{F}\hat{E})^{-1/2} \text{vec}(\sigma\mu I - H(\hat{X}\hat{S}))\|^2 \leq \left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right) n\mu. \quad (39)$$

Lemma A.7. Let $(X, y, S) \in \mathcal{N}_{\infty}^{-}(\gamma)$ and $P \in \mathcal{P}(X, Y)$ be given, and define $G = \hat{E}^{-1}\hat{F}$. Then the Newton step corresponding to system (6) satisfies

$$\|H_P(\Delta X \Delta S)\|_F \leq \frac{\sqrt{\text{cond}(G)}}{2} \left(1 - 2\sigma + \frac{\sigma^2}{\gamma}\right) n\mu. \quad (40)$$

Corollary A.8. If we set $\sigma = 0$ in (23), then the search direction in the predictor step satisfies

$$\|H_P(\Delta X^a \Delta S^a)\|_F \leq \frac{\sqrt{\text{cond}(G)}}{2} n\mu. \quad (41)$$

Finally, the lemma is proved in [5].

Lemma A.9. Let $X \succ 0$ and $S \succ 0$. If ones has

$$\det(X(\alpha)S(\alpha)) > 0, \quad \forall 0 \leq \alpha \leq \bar{\alpha},$$

then $X(\bar{\alpha}) \succ 0$ and $S(\bar{\alpha}) \succ 0$.