McMaster University

Advanced Optimization Laboratory

Title:

A New Class of Polynomial Primal-Dual Methods for Linear and Semidefinite Optimization

Authors:

Jiming Peng, Cornelis Roos, Tamás Terlaky

AdvOl-Report No. 1999/1

December 1999, Hamilton, Ontario, Canada

A New Class of Polynomial Primal-Dual Methods for Linear and Semidefinite Optimization

Jiming Peng[†] Cornelis Roos[†] Tamás[‡] Terlaky

December 29, 1999

[†]*Faculty of Information Technology and Systems,*
Delft University of Technology
P.O.Box 5031, 2600 GA Delft, The Netherlands

[‡]*Department of Computing and Software,*
McMaster University, Hamilton, Ontario, Canada, L8S 4L7

Abstract

We propose a new class of primal-dual methods for linear optimization (LO). By using some new analysis tools, we prove that the large update method for LO based on the new search direction has a polynomial complexity $O\left(n^{\frac{4}{4+\rho}} \log \frac{n}{\varepsilon}\right)$ iterations where $\rho \in [0, 2]$ is a parameter used in the system defining the search direction. If $\rho = 0$, our results reproduce the well known complexity of the standard primal dual Newton method for LO. At each iteration, our algorithm needs only to solve a linear equation system. An extension of the algorithms to semidefinite optimization is also presented.

Keywords: Linear Optimization, Semidefinite Optimization, Interior Point Method, Primal-Dual Newton Method, Polynomial Complexity.

AMS Subject Classification: **90C05**

1 Introduction

Interior point methods (IPMs) are among the most effective methods for solving wide classes of optimization problems. Since the seminal work of Karmarkar [7], many researchers have proposed and analyzed various IPMs for Linear and Semidefinite Optimization (LO and SDO) and a large amount of results have been reported. For a survey we refer to recent books on the subject ([17], [22], [24]). An interesting fact is that almost all known polynomial-time variants of IPMs use the so-called *central path* [18] as a guideline to the optimal set, and some variant of Newton's method to follow the central path approximately. Therefore, the theoretical analysis of IPMs consists for a great deal of analyzing Newton's method. At present there is still a gap between the practical behavior of the algorithms and the theoretical performance results, in favor of the practical behavior. This is especially true for so-called primal-dual large-update methods, which are the most efficient methods in practice (see, e.g. Andersen et al. [1]).

The aim of this paper is to present a new class of primal dual Newton-type algorithms for LO and SDO. To be more concrete we need to go into more detail at this stage. We consider first the following linear optimization problem:

$$(P) \qquad \min\{c^T x : Ax = b, x \geq 0\},$$

where $A \in \Re^{m\times n}$ satisfies $\text{rank}(A) = m,\ \ b \in \Re^m,\ \ c \in \Re^n$, and its dual problem

$$(D) \qquad \max\{b^T y : A^T y + s = c, s \geq 0\}.$$

We assume that both (P) and (D) satisfy the interior point condition (IPC), i.e., there exists (x^0, s^0, y^0) such that

$$Ax^0 = b,\ x^0 > 0, \quad A^T y^0 + s^0 = c, s^0 > 0.$$

It is well known that the IPC can be assumed without loss of generality. For this and some other properties mentioned below, see, e.g., [17]). Finding an optimal solution of (P) and (D) is equivalent to solving the following system.

$$\begin{aligned} Ax &= b, \quad x \geq 0, \\ A^T y + s &= c, \quad s \geq 0, \\ xs &= 0. \end{aligned} \tag{1}$$

Here xs denotes the coordinatewise product of the vectors x and s. The basic idea of primal-dual IPMs is to replace the third equation in (1), the so-called *complementarity condition* for (P) and (D) by the parameterized equation $xs = \mu e$, where e denotes the all-one vector and $\mu > 0$. Thus we consider the system

$$\begin{aligned} Ax &= b, \quad x \geq 0, \\ A^T y + s &= c, \quad s \geq 0, \\ xs &= \mu e. \end{aligned} \tag{2}$$

If the IPC holds, then for each $\mu > 0$, the parameterized system (2) has a unique solution. This solution is denoted as $(x(\mu), y(\mu), s(\mu))$ and we call $x(\mu)$ the μ-center of (P) and $(y(\mu), s(\mu))$ the μ-center of (D). The set of μ-centers (with μ running through all positive real numbers) gives a homotopy path, which is called *the central path* of (P) and (D). The relevance of the central path for LO was recognized first by Sonnevend [18] and Megiddo [10]. If $\mu \to 0$ then the limit of the central path exists and since the limit points satisfy the complementarity condition $xs = 0$, the limit yields optimal solutions for (P) and (D).

IPMs follow the central path approximately. Let us briefly indicate how this goes. Without loss of generality we assume that $(x(\mu), y(\mu), s(\mu))$ is known for some positive μ. We first update μ to $\mu := (1-\theta)\mu$, for some $\theta \in (0, 1)$. Then we solve the following well-defined Newton system

$$\begin{aligned} A\Delta x &= 0, \\ A^T \Delta y + \Delta s &= 0, \\ s\Delta x + x\Delta s &= \mu e - xs, \end{aligned} \tag{3}$$

and get a unique search direction $(\Delta x, \Delta s, \Delta y)$. By taking a step along the search direction where the step size is defined by some line search rules, one constructs a new triple (x, y, s) that is 'close' to $(x(\mu), y(\mu), s(\mu))$. This process is repeated until the point (x, y, s) is in a certian neighborhood of the central path. Then μ is again reduced by the factor $1 - \theta$ and we apply Newton's method targeting at the new μ-centers, and so on. We continue this process until μ is small enough. Most practical algorithms then construct a basic solution and produce an optimal basic solution by *crossing-over* to the Simplex method. An alternative way is to apply a rounding procedure as described by Ye [23] (see also Mehrotra and Ye [12] [17]).

The choice of the parameter θ plays an important role both in the theory and practice of IPMs. Usually, if θ is a constant independent of n, for instance $\theta = \frac{1}{2}$, then we call the algorithm a large update (or long-step) method. If θ depends on n such as $\theta = \frac{1}{\sqrt{n}}$, then the algorithm is named a small update (or short-step) method. It is now known that small update methods have an $O(\sqrt{n} \log \frac{n}{\varepsilon})$ iteration bound and the large update ones have a worse case iteration bound as $O(n \log \frac{n}{\varepsilon})$ [17, 22, 24]. The reason for the worse bound for large update methods is that, in both the analysis and implementation of large update IPMs, we usually use some proximities (or potential functions) to control the iteration, and up to now we can only prove that the proximity (or the potential function) has at least a constant decrease after one step. For instance, considering the primal dual Newton method, after one step the proximity δ used in this paper satisfies [1] $\delta_+^2 - \delta^2 \leq -\beta$ for some constant β [16]. On the other hand, contrary to the theoretical results, large update IPMs work much more efficient in practice than small update methods [1]. Several authors have suggested to use so called higher-order methods to improve the complexity of large update IPMs [4, 6, 14, 25, 26]. Then, at each iteration, one solves some additional equations based on the higher-order approximations to the system (2).

The motivation of this work is to improve the complexity of large update IPMs. Different from the higher-order approach, we reconsider the Newton system for (2), keeping in mind that our target is to get closer to the μ-center. Now let us focus on the Newton step. For notational convenience, we introduce the following notations:

$$v := \sqrt{\frac{xs}{\mu}}, \quad v^{-1} := \sqrt{\frac{\mu}{xs}}; \tag{4}$$

$$d_x := \frac{v\Delta x}{x}, \quad d_s := \frac{v\Delta s}{s}; \tag{5}$$

$$\bar{d}_x := \frac{\Delta x}{x}, \quad \bar{d}_s := \frac{\Delta s}{s}. \tag{6}$$

Using the above notations, one can state the central condition in (2) as $v = v^{-1} = e$. Denote $d_v = d_x + d_s$, the last equation in (3) is equivalent to

$$d_v = v^{-1} - v.$$

Observe that we can also decompose the above system into two systems as the predictor direction which is obtained by solving

$$(d_v)_{Pred} = -v,$$

and the corrector direction by

$$(d_v)_{Corr} = v^{-1}.$$

[1] Here δ and δ_+ denote the proximity before and after one step,respectively.

The corrector direction serves the purpose of centering, it points towards the "analytic center" of the feasible set, while the predictor direction aims to decrease the duality gap. It is straightforward to verify that $(d_v)_i \leq 0$ for all the components $v_i \geq 1$ and $(d_v)_i > 0$ for the components $v_i < 1$. This means that if $v_i < 1$ then the Newton step increases v_i and v_i decreases whenever $v_i > 1$ to get more close to the μ-center. It is reasonable to expect that if we can increase the small components and decrease the large components of v more, we might arrive our target the μ-center faster. Motivated by this observation, we reconsider the right hand side of the equation defining the corrector direction according to the current point v. The new corrector direction is defined by

$$(d_v)_{Corr} = v^{-1-\rho}, \quad \rho \geq 0,$$

thus yielding a new system as follows

$$\begin{aligned} \bar{A}d_x &= 0, \\ \bar{A}^T \Delta y + d_s &= 0, \\ d_x + d_s &= v^{-1-\rho} - v, \end{aligned} \tag{7}$$

where $\bar{A} = AV^{-1}X, V = \operatorname{diag}(v), X = \operatorname{diag}(x)$ and $\rho \geq 0$ is a parameter. In this work we consider only the case that the parameter ρ is restricted to the interval $[0, 2]$. Note that if $\rho = 0$, the new system is identical to the standard Newton system.

It may be clear from the above description that in the analysis of IPMs we need to keep control on the 'distance' from the current iterates to the current μ-centers. In other words, we need to quantify the 'distance' from the vector xs to the vector μe in some *proximity measure*. In fact, the choice of the proximity measure is crucial for both the quality and the elegance of the analysis. The proximity measure we use here is defined as follows.

$$\delta(xs, \mu) := \left\| v - v^{-1} \right\|. \tag{8}$$

Note that the measure vanishes if $xs = \mu e$ and is positive otherwise. An interesting observation is that in the special case of $\rho = 2$, the right hand side of the third term in the system (7) represents the negative gradient of the proximity measure $\frac{1}{2}\delta^2$ in the v-space. When solving this system with $\rho = 2$, we get the steepest descent direction for the proximity measure along which the proximity can be driven to zero. As we will see later, after one step using the new search direction, the proximity will decrease at least as large as $\beta\delta^{\frac{2}{3}}$ where β is a constant. Consequently we get an improvement over the complexity of the algorithm. We also mention that in [16], the authors have shown that $\delta_+^2 - \delta^2 \leq -\beta\delta$ after one feasible standard Newton step if $v_{\min} \geq 1$, which is exactly the same as we will state in Lemma 3.8 of this work. However, we failed to prove a similar inequality for the case $v_{\min} < 1$ and hence could not improve the complexity of the large update IPM in [16]. The measure δ, up to a factor $\frac{1}{2}$, was introduced by Jansen et al. [5], and thoroughly used in [17], Zhao [26], more recently [16]. Its SDO analogue was also used in the analysis of interior point methods for semidefinite optimization [3]. We notice that variants of the proximity $\delta(xs, \mu)$ had been used by Kojima et al. in [9] and Mizuno et al. in [13].

The paper is organized as follows. First, in Section 2, we present some technical results which will be used in our analysis later. In Section 3 we analyze the method with damped step and show that it has an $O\left(n^{\frac{4}{4+\rho}} \log \frac{n}{\varepsilon}\right)$ polynomial iteration bounds. In Section 4 we discuss an extension of the new primal dual algorithm for SDO and study its complexity. Finally we close this paper by some concluding remarks in Section 5.

A few words about the notations. Throughout the paper, $\|\cdot\|$ denotes the 2-norm for vectors and the Frobenius norm for matrices while $\|\cdot\|_\infty$ denotes the infinity norm. For any $x = (x_1, x_2, \cdots, x_n)^T \in \Re^n$, $x_{\min} = \min(x_1, x_2, \cdots, x_n)$ (or $x_{\max}$) is the component of x which takes the minimal (or maximal) value. For any symmetric matrix G, we also define $\lambda_{\min}(G)$ (or $\lambda_{\max}(G)$) the minimal (or maximal) eigenvalue of G. Furthermore we also assume that the eigenvalues of G are listed according to the order of their absolute values such that $|\lambda_1(G)| \geq |\lambda_2(G)| \geq \cdots \geq |\lambda_n(G)|$. If G is positive semidefinite, then it holds $0 \leq \lambda_{\min}(G) = \lambda_n(G)$, $\lambda_{\max}(G) = \lambda_1(G)$. For any symmetric matrix G, we also denote $|G| = (GG)^{\frac{1}{2}}$. For two symmetric matrices G, H, the relation $G \preceq H$ means $H - G$ is positive semidefinite, or equivalently $H - G \succeq 0$.

2 Technical Results

As we stated in the introduction, a key issue in the analysis of an interior point method, particularly for a large update IPM, is the decreasing property of a positive sequence of the proximity measure values. This is crucial for the complexity of the algorithm. In this section we consider a general positive decreasing sequence. First we give a technical result which will be used in our later analysis.

Lemma 2.1 *Suppose that $\alpha \geq 1$. Then*

$$1 - \alpha t \leq (1-t)^\alpha, \quad t \in [0,1]; \tag{9}$$

$$\alpha(t-1) \leq t^\alpha - 1, \quad t \geq 0. \tag{10}$$

If $\alpha_1 \geq \alpha_2 > 0$, then

$$\left|t - t^{-\alpha_1}\right| \geq \left|t - t^{-\alpha_2}\right|, \quad t > 0. \tag{11}$$

Proof: The inequality (9) follows since for any fixed $\alpha \geq 1$, the function $(1-t)^\alpha - 1 + \alpha t$ is increasing for $t \in [0,1]$ and zero if $t = 0$. When $\alpha \geq 1$, the function $t^\alpha - 1 - \alpha(1-t)$ is convex with respect to $t \geq 0$ and has a global minimum zero at $t = 1$, which gives (10). One can easily verify (11). □

An equivalent statement of (9) is

$$(t-1)^\alpha \geq t^\alpha - \alpha t^{\alpha-1}, \quad t \geq 1, \ \alpha \geq 1. \tag{12}$$

Now we can state our main result in this section.

Proposition 2.2 *Suppose that $t_0 > 1$ is a constant. Suppose $\{t_k, k = 0, 1, 2 \cdots\}$ is a sequence satisfying the following inequalities*

$$t_{k+1} \leq t_k - \beta t_k^\gamma, \quad k = 0, 1, \cdots, \tag{13}$$

with $\gamma \in [0,1)$ and $\beta > 0$. Then one has

$$t_{k+1} \leq 0,$$

for all $k \geq \left\lceil \frac{t_0^{1-\gamma}}{\beta(1-\gamma)} \right\rceil$.

Proof: First we note if $\beta \geq t_0^{1-\gamma}$, then 1 step is sufficient. Hence we can assume without loss of generality that $0 < \beta < t_0^{1-\gamma}$. We begin the proof by considering the simple case that $\beta = 1$ and $\tau = 0$. It follows from (13) that

$$t_{k+1} \leq t_k - t_k^{\gamma}. \tag{14}$$

Assume the current point $t_k \geq 1$, we get

$$\begin{aligned}
(t_k^{1-\gamma} - (1-\gamma))^{\frac{1}{1-\gamma}} &= (1-\gamma)^{\frac{1}{1-\gamma}} \left(\frac{t_k^{1-\gamma}}{1-\gamma} - 1 \right)^{\frac{1}{1-\gamma}} \\
&\geq (1-\gamma)^{\frac{1}{1-\gamma}} \left(\left(\frac{t_k^{1-\gamma}}{1-\gamma} \right)^{\frac{1}{1-\gamma}} - \frac{1}{1-\gamma} \left(\frac{t_k^{1-\gamma}}{1-\gamma} \right)^{\frac{1}{1-\gamma}-1} \right) \\
&= t_k - t_k^{\gamma} \geq t_{k+1},
\end{aligned}$$

where the first inequality follows from (12). The above inequality further implies

$$t_{k+1}^{1-\gamma} \leq t_k^{1-\gamma} - (1-\gamma).$$

Hence, since $\beta = 1$, after at most $\left\lceil \frac{t_0^{1-\gamma}}{1-\gamma} \right\rceil$ steps we have that $t_{k+1} \leq 0$. Now we turn to the general case that $\beta \neq 1$. Let us define a new sequence by

$$\bar{t}_k = t_k \beta^{-\frac{1}{1-\gamma}}, \quad k = 0, 1, \cdots$$

Then the inequality (13) gives

$$\bar{t}_{k+1} \leq \bar{t}_k - \bar{t}_k^{\gamma}.$$

By our discussion about the special case that $\beta = 1$, we know that after at most $\left\lceil \frac{\bar{t}_0^{1-\gamma}}{1-\gamma} \right\rceil$ steps it holds $\bar{t}_{k+1} \leq 0$, which is equivalent to say that, after at most $\left\lceil \frac{t_0^{1-\gamma}}{\beta(1-\gamma)} \right\rceil$ steps we get $t_{k+1} \leq 0$. This completes the proof of the proposition. □

3 New Primal Dual Methods for LO and Their Complexity

In the present section we discuss the new primal dual Newton methods for LO and study the complexity of the large update algorithm. The section consists of four parts. In the first subsection we describe the new algorithm. In the second subsection we estimate the magnitude of the search direction and the maximum value of a feasible step size. The third subsection is devoted to estimate the proximity measure after one step. The complexity of the algorithm is given in the last subsection.

3.1 The algorithm

Let $\Delta x, \Delta y, \Delta s$ denote the solution of the following modified Newton equation system for the parameterized system (2):

$$\begin{aligned}
A\Delta x &= 0, \\
A^T \Delta y + \Delta s &= 0, \\
s\Delta x + x\Delta s &= \mu^{1+\frac{\rho}{2}} \frac{e}{(xs)^{\frac{\rho}{2}}} - xs.
\end{aligned} \tag{15}$$

This is the Newton system for the equation $xs = \mu \left(\frac{\mu e}{xs}\right)^{\frac{\rho}{2}}$. Note that on the central path it holds $\frac{xs}{\mu} = e = \left(\frac{\mu}{xs}\right)^{\frac{\rho}{2}}$. From the definitions of v, d_x, d_s one can easily check that the above system is equivalent to (7). Recall (cf. Chapter 7 in [17]) that since $\text{rank}(A) = m$, for any $\mu > 0$, the above equation system has a unique solution $\Delta x, \Delta y, \Delta s$. The result of a damped Newton step with damping factor α is denoted as

$$x_+ = x + \alpha\Delta x, \quad y_+ = y + \alpha\Delta y, \quad s_+ = s + \alpha\Delta s. \tag{16}$$

In the algorithm we use a threshold value τ for the proximity and we assume that we are given a triple (x^0, y^0, s^0) such that $\delta(x^0 s^0, \mu^0) \leq \tau$ for $\mu^0 = 1$. This can be done without loss of generality (cf. [17]).

If, for the current iterates (x, y, s) and barrier parameter value μ the proximity $\delta(xs, \mu)$ exceeds τ then we use one or more damped Newton steps to recenter while keeping μ temporarily fixed; otherwise μ is reduced by the factor $1 - \theta$. This is repeated until $n\mu < \varepsilon$. Thus the algorithm can be stated as follows.

Large Update Primal-Dual Algorithm for LO

Input:
  A proximity parameter τ;
  an accuracy parameter $\varepsilon > 0$;
  a variable damping factor α;
  a fixed barrier update parameter θ, $0 < \theta < 1$;
  (x^0, s^0) and $\mu^0 = 1$ such that $\delta(x^0 s^0; \mu^0) \leq \tau$.
begin
  $x := x^0$; $s := s^0$; $\mu := \mu^0$;
  while $n\mu \geq \varepsilon$ **do**
  begin
    $\mu := (1 - \theta)\mu$;
    while $\delta(xs; \mu) \geq \tau$ **do**
    Solve the system (15),
    begin
      $x := x + \alpha\Delta x$;
      $s := s + \alpha\Delta s$;
      $y := y + \alpha\Delta y$
    end
  end
end

Remark 3.1 *The damping parameter α has to be taken such that the proximity measure function δ decreases sufficiently. In the next section we determine a default value for α.*

3.2 Magnitude of the search direction and feasible step size

First recall that the proximity $\delta(xs,\mu)$ before the step satisfies

$$\delta(xs,\mu)^2 = \left\|v - v^{-1}\right\|^2 = e^T\left(v - v^{-1}\right)^2 = e^T\left(v^2 + v^{-2} - 2e\right). \tag{17}$$

We have, using (15) and (16),

$$\begin{aligned} x_+s_+ &= (x+\alpha\Delta x)(s+\alpha\Delta s) \\ &= xs + \alpha(x\Delta s + s\Delta x) + \alpha^2\Delta x\Delta s \\ &= xs + \alpha\left(\mu^{1+\frac{\rho}{2}}\frac{e}{(xs)^{\frac{\rho}{2}}} - xs\right) + \alpha^2\Delta x\Delta s. \end{aligned}$$

Defining

$$v_+ = \sqrt{\frac{x_+s_+}{\mu}},$$

and using that $xs = \mu v^2$ and $\Delta x\Delta s = \mu d_x d_s$ we obtain

$$v_+^2 = \frac{x_+s_+}{\mu} = v^2 + \alpha\left(v^{-\rho} - v^2\right) + \alpha^2 d_x d_s. \tag{18}$$

Since the displacements Δx and Δs are orthogonal, the scaled displacements d_x and d_s are orthogonal as well. Hence we have

$$d_x^T d_s = 0. \tag{19}$$

Thus we get the following expression for the proximity after the step.

$$\begin{aligned} \delta(x_+s_+,\mu)^2 &= e^T\left(v_+^2 + \frac{e}{v_+^2} - 2e\right) \\ &= e^T\left(v^2 + \alpha\left(v^{-\rho} - v^2\right) + \alpha^2 d_x d_s + \frac{e}{v^2 + \alpha\left(v^{-\rho} - v^2\right) + \alpha^2 d_x d_s} - 2e\right) \\ &= e^T\left(v^2 + \alpha\left(v^{-\rho} - v^2\right) + \frac{e}{v^2 + \alpha\left(v^{-\rho} - v^2\right) + \alpha^2 d_x d_s} - 2e\right). \end{aligned} \tag{20}$$

In the sequel we will denote $\delta(xs,\mu)$ simply as δ and $\delta(x_+s_+,\mu)$ simply as δ_+.

Recall that the term $d_x d_s$ represents the second order effect in the Newton step. It may be worthwhile to consider the case where this term is zero, i.e., when the Newton process is exact. In that case after a step of size α, x_+s_+/μ is given by

$$w(\alpha) := v^2 + \alpha\left(v^{-\rho} - v^2\right) = \alpha v^{-\rho} + (1-\alpha)v^2.$$

We have the following result.

Lemma 3.2 *If* $d_x d_s = 0$ *then*

$$\delta_+^2 \le (1-\alpha)\delta^2 + \alpha\left\|v^{\frac{\rho}{2}} - v^{-\frac{\rho}{2}}\right\|^2, \quad \forall\alpha\in[0,1].$$

Particularly, if $\alpha = \frac{2}{\rho+2}$, *then*

$$\delta_+^2 \le \frac{\rho\delta^2}{\rho+2}.$$

Proof: If $d_x d_s = 0$ then

$$\delta_+^2 = e^T \left(v^2 + \alpha \left(v^{-\rho} - v^2 \right) + \frac{e}{v^2 + \alpha \left(v^{-\rho} - v^2 \right)} - 2e \right). \tag{21}$$

Defining

$$\chi(t) := t + \frac{1}{t} - 2, \quad t > 0, \tag{22}$$

one may easily verify that $\chi(t)$ is strictly convex on its domain and minimal at $t = 1$ where $\chi(1) = 0$. Moreover it holds $\chi(t) = \chi(\frac{1}{t})$. It follows

$$\delta^2 = \sum_{i=1}^{n} \chi \left(v_i^2 \right) = \sum_{i=1}^{n} \chi \left(v_i^{-2} \right), \quad \delta_+^2 = \sum_{i=1}^{n} \chi \left(w_i(\alpha) \right). \tag{23}$$

Therefore, since $w_i(\alpha) = v_i^2 + \alpha(v_i^{-\rho} - v_i^2) = \alpha v_i^{-\rho} + (1 - \alpha) v_i^2$, we may write

$$\delta_+^2 = \sum_{i=1}^{n} \chi \left(w_i(\alpha) \right) \le \sum_{i=1}^{n} (1 - \alpha) \chi(v_i^2) + \alpha \chi \left(v_i^{-\rho} \right).$$

This proves the first statement of the lemma. To prove the second conclusion of the lemma, we observe first that, when the Newton process is exact, the step size $\alpha = \frac{2}{\rho+2}$ is feasible. We need only to consider the case $\rho > 0$. It follows from (21)

$$\begin{aligned}
\delta_+^2 &= \sum_{i=1}^{n} \frac{\rho}{\rho+2} \left(v_i^2 + v_i^{-2} - 2 \right) + \left(\frac{2}{\rho+2} v_i^{-\rho} - \frac{\rho}{\rho+2} v_i^{-2} - \frac{4}{\rho+2} + \frac{\rho+2}{\rho v_i^2 + 2 v_i^{-\rho}} \right) \\
&= \frac{\rho}{\rho+2} \delta^2 + \sum_{i=1}^{n} \left(\frac{2}{\rho+2} v_i^{-\rho} - \frac{\rho}{\rho+2} v_i^{-2} - \frac{4}{\rho+2} + \frac{\rho+2}{\rho v_i^2 + 2 v_i^{-\rho}} \right) \\
&\le \frac{\rho}{\rho+2} \delta^2 + \sum_{i=1}^{n} \left(\frac{2}{\rho+2} v_i^{-\rho} - \frac{\rho}{\rho+2} \left(\frac{2}{\rho} (v_i^{-\rho} - 1) + 1 \right) - \frac{4}{\rho+2} + \frac{\rho+2}{\rho v_i^2 + 2 v_i^{-\rho}} \right) \\
&= \frac{\rho}{\rho+2} \delta^2 + \sum_{i=1}^{n} \left(\frac{\rho+2}{\rho v_i^2 + 2 v_i^{-\rho}} - 1 \right) \le \frac{\rho}{\rho+2} \delta^2 + \sum_{i=1}^{n} \left(\frac{\rho+2}{\rho - 2 + 2 v_i^{\rho} + 2 v_i^{-\rho}} - 1 \right) \\
&\le \frac{\rho}{\rho+2} \delta^2,
\end{aligned}$$

where the first and second inequalities are given by (10) in Lemma 2.1 where "that" $\alpha = \frac{\rho}{2}$, and the last one implied by the fact that $v_i^{\rho} + v_i^{-\rho} \ge 2$ for all $i = 1, 2, \cdots, n$. The proof of the lemma is finished. □

Now we consider the practical case where the Newton step is not exact. This is the case where the vector $d_x d_s$ is nonzero. According to (18), after a step of size α we then have (see (4)-(6))

$$v_+^2 = \frac{x_+ s_+}{\mu} = v^2 (e + \alpha v^{-1} d_x)(e + \alpha v^{-1} d_s) = v^2 (e + \alpha \bar{d}_x)(e + \alpha \bar{d}_s). \tag{24}$$

Hence the maximal feasible step size is determined by the vector $(\bar{d}_x, \bar{d}_s)$. For notation convenience, we also define

$$\sigma := \left(\sum_{i=1}^{n} (v_i - v_i^{-3})(v_i - v_i^{-1-\rho}) \right)^{\frac{1}{2}}. \tag{25}$$

Lemma 3.3 *Let δ and σ be defined by (8) and (25) respectively. It holds that $\sigma \geq \delta$ for all $\rho \geq 0$, and if $\rho \in [0,2]$ then*

$$\sigma \geq \|(d_x, d_s)\| \geq \delta.$$

Proof: Combining (7) with (19) we obtain

$$\|(d_x, d_s)\|^2 = d_x^T d_x + d_s^T d_s = (d_x + d_s)^T (d_x + d_s) = \left\| v - v^{-1-\rho} \right\|^2.$$

Since $\rho \in [0,2]$, it follows from (11) in Lemma 2.1 that

$$\begin{aligned} \sigma^2 &= \sum_{i=1}^{n} \left(v_i - v_i^{-3} \right) \left(v_i - v_i^{-1-\rho} \right) \\ &\geq \sum_{i=1}^{n} \left(v_i - v_i^{-1-\rho} \right)^2 = \|(d_x, d_s)\|^2 \\ &\geq \sum_{i=1}^{n} \left(v_i - v_i^{-1} \right)^2 = \delta^2, \end{aligned}$$

which yields the second statement of the lemma. Skipping the first inequality and the second equality in the above proof gives the first conclusion of the lemma. □

Our next lemma estimates the norm $\|v^{-1}\|_\infty$ in terms of σ. We have

Lemma 3.4 *Let σ be defined by (25). Then*

$$\left\| v^{-1} \right\|_\infty \leq (1+\sigma)^{\frac{2}{4+\rho}}.$$

Proof: First note that $\|v^{-1}\|_\infty = \frac{1}{v_{\min}}$. Hence it suffices to show that

$$v_{\min} \geq (1+\sigma)^{-\frac{2}{4+\rho}}.$$

This is trivial if $v_{\min} \geq 1$. Now we consider the case that $v_{\min} < 1$. From (25) we derive

$$\sigma^2 \geq \left| \left(v_i^{-3} - v_i \right) \left(v_i^{-1-\rho} - v_i \right) \right|, \quad \forall i = 1, \cdots, n,$$

which further implies, since $\rho \in [0,2]$,

$$\begin{aligned} \sigma^2 &\geq \left(v_{\min}^{-3} - v_{\min} \right) \left(v_{\min}^{-1-\rho} - v_{\min} \right) = \left(v_{\min}^{-\frac{4+\rho}{2}} - v_{\min}^{\frac{4-\rho}{2}} \right) \left(v_{\min}^{-\frac{4+\rho}{2}} - v_{\min}^{\frac{\rho}{2}} \right) \\ &\geq \left(v_{\min}^{-\frac{4+\rho}{2}} - 1 \right)^2. \end{aligned}$$

The statement of the lemma follows directly from the above inequality. □

Now we are ready to state our main result in this section.

Lemma 3.5 *Let $\bar{d}_x, \bar{d}_s$ be defined by (6). Then it holds*

$$\|(\bar{d}_x, \bar{d}_s)\| \leq \sigma (1+\sigma)^{\frac{2}{4+\rho}}.$$

Furthermore, the maximal feasible step size $\alpha_{\max}$ satisfies

$$\alpha_{\max} \geq \sigma^{-1} (1+\sigma)^{-\frac{2}{4+\rho}}.$$

Proof: Since the current point (x, y, s) is strictly feasible, from (6) one can easily see that the step size α is feasible if and only if both the vectors $e + \alpha\bar{d}_x$ and $e + \alpha\bar{d}_s$ are strictly feasible. It follows

$$\|(\bar{d}_x, \bar{d}_s)\| = \left\|\left(\frac{d_x}{v}, \frac{d_s}{v}\right)\right\| \leq \frac{\|(d_x, d_s)\|}{v_{\min}} \leq \frac{\sigma}{v_{\min}} \leq \sigma(1+\sigma)^{\frac{2}{4+\rho}},$$

where the last two inequalities follow from Lemma 3.3 and Lemma 3.4, which concludes the statements of the lemma. □

3.3 Estimate of the proximity after a step

We estimate the decreasing value of the proximity after one step in this section. Let

$$d_x = (d_x^1, d_x^2, \ldots, d_x^n)^T$$

and similarly for d_s. We define the difference between the proximity before and after one step as a function of α, i.e., $f(\alpha) = \delta_+^2 - \delta^2$. From (19) and the definitions of δ (8) and δ_+ (20), we derive

$$\begin{aligned} f(\alpha) &= \sum_{i=1}^{n}\left(\alpha(v_i^{-\rho} - v_i^2) + \frac{1}{v_i^2 + \alpha(v_i^{-\rho} - v_i^2) + \alpha^2 d_x^i d_s^i} - \frac{1}{v_i^2}\right) \\ &= \sum_{i=1}^{n}\alpha(v_i^{-\rho} - v_i^2) + \frac{1}{v_i^2}\left(\frac{1}{(1+\alpha\bar{d}_{xi})(1+\alpha\bar{d}_{si})} - 1\right). \end{aligned} \tag{26}$$

Obviously $f(\alpha)$ is a twice continuously differentiable function of α if the step size α is feasible. Our next result says that in the interval $[0, \alpha_{\max})$ the function $f(\alpha)$ is a convex function of α, where $\alpha_{\max}$ is the maximal feasible step size.

Lemma 3.6 *Let the function $f(\alpha)$ be defined by (26) and let the parameter $\alpha \in [0, \alpha_{\max})$. Then $f(\alpha)$ is convex. Furthermore, it holds*

$$\sum_{i=1}^{n}\frac{1}{v_i^2}\left(\frac{\bar{d}_{xi}^{\,2}}{(1+\alpha\bar{d}_{xi})^3(1+\alpha\bar{d}_{si})} + \frac{\bar{d}_{si}^{\,2}}{(1+\alpha\bar{d}_{si})^3(1+\alpha\bar{d}_{xi})}\right) \leq f''(\alpha); \tag{27}$$

and that

$$f''(\alpha) \leq 3\sum_{i=1}^{n}\frac{1}{v_i^2}\left(\frac{\bar{d}_{xi}^{\,2}}{(1+\alpha\bar{d}_{xi})^3(1+\alpha\bar{d}_{si})} + \frac{\bar{d}_{si}^{\,2}}{(1+\alpha\bar{d}_{si})^3(1+\alpha\bar{d}_{xi})}\right). \tag{28}$$

Proof: By direct algebraic calculus, we have

$$f''(\alpha) = \sum_{i=1}^{n}\frac{1}{v_i^2}\left(\frac{2\bar{d}_{xi}^{\,2}}{(1+\alpha\bar{d}_{xi})^3(1+\alpha\bar{d}_{si})} + \frac{2\bar{d}_{xi}\bar{d}_{si}}{(1+\alpha\bar{d}_{xi})^2(1+\alpha\bar{d}_{si})^2} + \frac{2\bar{d}_{si}^{\,2}}{(1+\alpha\bar{d}_{si})^3(1+\alpha\bar{d}_{xi})}\right).$$

By using the well-known inequality $|2t_1t_2| \leq t_1^2 + t_2^2$, we get

$$\left|\frac{2\bar{d}_{xi}\bar{d}_{si}}{(1+\alpha\bar{d}_{xi})^2(1+\alpha\bar{d}_{si})^2}\right| \leq \frac{\bar{d}_{xi}^{\,2}}{(1+\alpha\bar{d}_{xi})^3(1+\alpha\bar{d}_{si})} + \frac{\bar{d}_{si}^{\,2}}{(1+\alpha\bar{d}_{si})^3(1+\alpha\bar{d}_{xi})},$$

which implies the statements of the lemma. □

Denote $\omega_i = \sqrt{\bar{d_x}_i^{\,2} + \bar{d_s}_i^{\,2}}$ and $\omega = \|(\omega_1, \ldots, \omega_n)\|$. Obviously it holds $\omega = \|(\bar{d_x}, \bar{d_s})\|$. From Lemma 3.5,

$$\omega \leq \frac{\sigma}{v_{\min}} \leq \sigma(1+\sigma)^{\frac{2}{4+\rho}}. \tag{29}$$

Now recalling (27) and (28) we can conclude that for any $\alpha \in [0, \alpha_{\max})$,

$$\sum_{i=1}^{n} \frac{1}{v_i^2} \frac{\omega_i^2}{(1+\alpha\omega)^4} \leq f''(\alpha) \leq 3\sum_{i=1}^{n} \frac{1}{v_i^2} \frac{\omega_i^2}{(1-\alpha\omega)^4} \leq \frac{3\omega^2}{v_{\min}^2(1-\alpha\omega)^4}. \tag{30}$$

A direct calculation gives

$$f'(0) = -\sigma^2.$$

It follows from (30) and the convexity of $f(\alpha)$ that

$$\begin{aligned} f(\alpha) &\leq f'(0)\alpha + \frac{3\omega^2}{v_{\min}^2} \int_0^\alpha \int_0^\xi \frac{1}{(1-\zeta\omega)^4} d\zeta d\xi \\ &= f'(0)\alpha + \frac{\omega}{v_{\min}^2} \int_0^\alpha \left(\frac{1}{(1-\xi\omega)^3} - 1 \right) d\xi \\ &= -\frac{1}{2}\sigma^2\alpha + f_1(\alpha), \end{aligned} \tag{31}$$

where

$$f_1(\alpha) = -\frac{1}{2}\sigma^2\alpha + \frac{\omega}{v_{\min}^2} \int_0^\alpha \left(\frac{1}{(1-\xi\omega)^3} - 1 \right) d\xi.$$

It is easy to see that $f_1(\alpha)$ is also convex and twice differentiable in the interval $[0, \alpha_{\max})$. We are interested in the point $\alpha > 0$ at which the function $f_1(\alpha)$ has the value zero. By direct calculus, one has

$$\begin{aligned} f_1(\alpha) &= -\frac{1}{2}\sigma^2\alpha - \frac{\omega\alpha}{v_{\min}^2} + \frac{1}{2v_{\min}^2} \left(\frac{1}{(1-\alpha\omega)^2} - 1 \right) \\ &= \frac{\omega\alpha}{2v_{\min}^2} \left(\frac{2-\alpha\omega}{(1-\alpha\omega)^2} - 2 - \eta \right), \end{aligned} \tag{32}$$

where

$$\eta = \frac{\sigma^2 v_{\min}^2}{\omega}. \tag{33}$$

This means the function $f_1(\alpha) = 0$ if

$$\alpha = \frac{1}{\omega} \left(1 - \frac{1 + (4\eta+9)^{\frac{1}{2}}}{2(\eta+2)} \right) = \frac{3 + 2\eta - (4\eta+9)^{\frac{1}{2}}}{2\omega(\eta+2)}. \tag{34}$$

For this α we have

$$f(\alpha) \leq -\frac{3 + 2\eta - (4\eta+9)^{\frac{1}{2}}}{4\omega(\eta+2)} \sigma^2.$$

Now we are in a position to state our main result in this section.

Theorem 3.7 *Let the function $f(\alpha)$ be defined by (26) with $\delta \geq 1$. Then the step size $\alpha = \frac{3+2\eta-(4\eta+9)^{\frac{1}{2}}}{2\omega(\eta+2)}$ defined by (33) is feasible. Moreover it holds*

$$f(\alpha) \leq \frac{1}{2} f'(0)\alpha \leq -\frac{1}{30}\delta^{\frac{2\rho}{4+\rho}}.$$

Proof: The first part of the theorem follows directly from inequality (31) and the choice of α. The second conclusion of the theorem depends on several technical results which will be derived below. We first discuss the case that $v_{\min} \geq 1$.

Lemma 3.8 *Let the function $f(\alpha)$ be defined by (26) with $\delta \geq 1$ and let the step size defined by (34). If $v_{\min} \geq 1$, then*

$$f(\alpha) < -\frac{5-\sqrt{13}}{12}\delta < -\frac{1}{30}\delta^{\frac{2\rho}{4+\rho}}.$$

Proof: Since $v_{\min} \geq 1$, it follows from (29) that $\omega \leq \sigma$. By the choice of η we have that

$$\eta \;=\; \frac{\sigma^2 v_{\min}^2}{\omega} \geq \sigma. \tag{35}$$

Now recalling the fact that $\delta \geq 1$, it follows from Lemma 3.3 that

$$\eta \geq \sigma \geq \delta \geq 1,$$

which implies

$$\alpha \geq \frac{5-\sqrt{13}}{6\omega}.$$

The above inequality gives

$$\begin{aligned} f(\alpha) &\leq \frac{1}{2} f'(0)\alpha \leq -\frac{5-\sqrt{13}}{12}\frac{\sigma^2}{\sigma} \\ &\leq -\frac{5-\sqrt{13}}{12}\sigma \leq -\frac{5-\sqrt{13}}{12}\delta. \end{aligned} \tag{36}$$

It is easy to verify the second inequality in the lemma. This completes the proof of the lemma. □

In what follows we consider the case that $v_{\min} < 1$. First we want to estimate the constant η and the step size α in Theorem 3.7.

Lemma 3.9 *Let the constants η be defined by (33) and α by (34). If $\delta \geq 1$ then it holds*

$$\eta^{-1} \leq 2\,(1+\sigma)^{\frac{2-\rho}{4+\rho}}, \tag{37}$$

and

$$\alpha \geq \frac{2}{15}\sigma^{-1}(1+\sigma)^{\frac{\rho-4}{\rho+4}}. \tag{38}$$

Proof: From Lemma 3.4, (29) and (33) we obtain

$$\begin{aligned}\eta &= \frac{\sigma^2 v_{\min}^2}{\omega} \geq \sigma v_{\min}^3 \geq \sigma(1+\sigma)^{-\frac{6}{4+\rho}} \\ &\geq \frac{\sigma}{1+\sigma}\,(1+\sigma)^{-\frac{2-\rho}{4+\rho}} \geq \frac{1}{2}\,(1+\sigma)^{-\frac{2-\rho}{4+\rho}}\,,\end{aligned}$$

where the last two inequalities are implied by the fact $\sigma \geq \delta \geq 1$ and because the function $g(t) = \frac{t}{1+t}$ is increasing with respect to t for $t \geq 0$. This proves the first inequality in the lemma. Now we consider the second inequality. From (34) we derive that

$$\begin{aligned}\alpha &= \frac{3+2\eta-(4\eta+9)^{\frac{1}{2}}}{2\omega(\eta+2)} = \frac{2\eta}{\omega\left(3+2\eta+(4\eta+9)^{\frac{1}{2}}\right)} \\ &= \frac{2}{\omega\left(3\eta^{-1}+2+(4\eta^{-1}+9\eta^{-2})^{\frac{1}{2}}\right)} \geq \frac{2}{\omega\,(6\eta^{-1}+3)} \\ &\geq \frac{2}{\omega}\frac{1}{12\,(1+\sigma)^{\frac{2-\rho}{4+\rho}}+3} \geq \frac{2}{15\omega\,(1+\sigma)^{\frac{2-\rho}{4+\rho}}} \\ &\geq \frac{2}{15}\sigma^{-1}\,(1+\sigma)^{\frac{\rho-4}{\rho+4}}\,,\end{aligned}$$

where the first inequality follows by direct calculus and the second one from (37), the third is true since $(1+\sigma)^{\frac{2-\rho}{4+\rho}} \geq 1$ when $\delta \geq 1$, and the last given by (29). The proof of the lemma is finished. □

Now we can prove the following lemma.

Lemma 3.10 *Let the function $f(\alpha)$ be defined by (26) with $\delta \geq 1$ and let the step size defined by (34) and $v_{\min} < 1$. Then it holds*

$$f(\alpha) \leq \frac{1}{2}f'(0)\alpha \leq -\frac{1}{30}\delta^{\frac{2\rho}{4+\rho}}.$$

Proof: From (38) we obtain

$$\begin{aligned}f(\alpha) &\leq \frac{1}{2}f'(0)\alpha = -\frac{\alpha\sigma^2}{2} \leq -\frac{1}{15}\sigma\,(1+\sigma)^{\frac{\rho-4}{\rho+4}} \\ &= -\frac{1}{15}\sigma^{\frac{2\rho}{4+\rho}}\left(1+\sigma^{-1}\right)^{\frac{\rho-4}{\rho+4}} \leq -\frac{1}{30}\sigma^{\frac{2\rho}{4+\rho}} \\ &\leq -\frac{1}{30}\delta^{\frac{2\rho}{4+\rho}},\end{aligned}$$

where the third inequality is true since $\left(1+\sigma^{-1}\right)^{\frac{\rho-4}{\rho+4}} \geq \frac{1}{2}$ when $\delta \geq 1$, and the last by Lemma 3.3. This completes the proof for the lemma. □

Theorem 3.7 follows from Lemma 3.8 and Lemma 3.10. □

3.4 Complexity of the algorithm

In this section we derive an upper bound for the number of iterations of the algorithm if in each step the damping factor α is as in Theorem 3.7, namely $\alpha = \frac{3+2\eta-(4\eta+9)^{\frac{1}{2}}}{2\omega(\eta+2)}$. Then each damped Newton step reduces the squared proximity by at least $-\frac{1}{30}\delta^{\frac{2\rho}{4+\rho}}$ which depends on the current proximity. We first recall a lemma that estimates the proximity after a barrier parameter update.

Lemma 3.11 *Let (x, y, s) be strictly feasible and $\mu > 0$. If $\mu_+ = (1-\theta)\mu$ then*

$$\delta(xs, \mu_+) \leq \frac{\delta(xs, \mu) + \theta\sqrt{n}}{\sqrt{1-\theta}}.$$

Proof: The lemma is a slight modification of Lemma IV.36 (page 359) in [17]. The only difference exists in the definition of δ where $\delta = \|v - v^{-1}\|$ in this work while $\delta = \frac{1}{2}\|v - v^{-1}\|$ in [17], hence the details are omitted here. □

Lemma 3.12 *Let $\delta(xs, \mu) \leq \tau$ and $\tau \geq 1$. Then after an update of the barrier parameter no more than*

$$\left\lceil \frac{15(4+\rho)}{2} \left(\frac{\tau + \theta\sqrt{n}}{1-\theta} \right)^{\frac{8}{4+\rho}} \right\rceil$$

iterations are needed to recenter, namely to reach $\delta(xs, \mu_+) \leq \tau$ again.

Proof: By Lemma 3.11, after the update,

$$\delta(xs, \mu_+)^2 \leq \frac{(\tau + \theta\sqrt{n})^2}{1-\theta}.$$

Each damped Newton step decreases δ^2 by at least $\frac{1}{30}\delta^{\frac{2\rho}{4+\rho}}$. It follows from Proposition 2.2 that after at most

$$\left\lceil \frac{30}{1-\frac{\rho}{4+\rho}} \left(\frac{\tau + \theta\sqrt{n}}{\sqrt{1-\theta}} \right)^{2-\frac{2\rho}{4+\rho}} \right\rceil = \left\lceil \frac{15(4+\rho)}{2} \left(\frac{\tau + \theta\sqrt{n}}{\sqrt{1-\theta}} \right)^{\frac{8}{4+\rho}} \right\rceil$$

inner iterations, the proximity will have passed the threshold value τ. This implies the lemma. □

Theorem 3.13 *If $\tau \geq 1$, the total number of iterations required by the primal-dual Newton algorithm is no more than*

$$\left\lceil \frac{15(4+\rho)}{2} \left(\frac{\tau + \theta\sqrt{n}}{\sqrt{1-\theta}} \right)^{\frac{8}{4+\rho}} \right\rceil \left\lceil \frac{1}{\theta} \log \frac{n}{\varepsilon} \right\rceil.$$

Proof: It can easily be shown that the number of barrier parameter updates is given by (cf. Lemma II.17, page 116, in [17])

$$\left\lceil \frac{1}{\theta} \log \frac{n}{\varepsilon} \right\rceil .$$

Multiplication of this number by the bound in Lemma 3.12 yields the theorem. □

For large update IPMs, omitting the round off brackets in Theorem 3.13 does not change the order of magnitude of the iteration bound. Hence we may safely consider the following expression as an upper bound for the number of iterations in the case $\tau = O(\sqrt{n})$, $\theta \in (0,1)$ independent of n, and $\rho = 2$:

$$O\left(n^{\frac{2}{3}} \log \frac{n}{\varepsilon} \right).$$

This gives the best bound known for large-update methods with large neighborhoods. Note that if $\rho = 0$, we obtain the to date best known complexity bounds for both small and large update methods. Moreover, our analysis allows extremely agressive updates of μ while preserving $O(n \log \frac{n}{\epsilon})$ complexity. For instance, we may take $\theta = 1 - \frac{1}{\sqrt{n}}$, $\tau = O\left(\sqrt{n}\right)$ and $\rho = 2$, resulting in $O(n \log \frac{n}{\epsilon})$ complexity bound.

4 New Primal Dual Algorithms for Semidefinite Optimization

In this section we consider an extension of the algorithms posed in the previous section to the case of SDO. We consider the SDO given in the following standard form:

$$\begin{array}{ll} \text{(SDO)} & \min \mathbf{Tr}\,(CX) \\ & \mathbf{Tr}\,(A_i X) = b_i \,(1 \le i \le m), \quad X \succeq 0, \end{array}$$

and its dual problem

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

Here C and A_i $(1 \le i \le m)$ are symmetric $n \times n$ matrices, and $b, y \in \mathbb{R}^m$. Furthermore, '$X \succeq 0$' means that X is symmetric positive semidefinite. The matrices A_i are assumed to be linearly independent. SDO is a generalization of LO where all the matrices A_i and C are diagonal which implies S is automatically diagonal and so X might also be assumed to be diagonal. The concept of the central path can also be extended to SDO. We assume both the SDO and its dual SDD are strictly feasible. The central path for SDO is defined by the solution sets $\{X(\mu), y(\mu), S(\mu), \mu > 0\}$ of the following system

$$\begin{aligned} \mathbf{Tr}\,(A_i X) &= b_i, \quad i = 1, \ldots, m, \\ \sum_{i=1}^{m} y_i A_i + S &= C, \\ XS &= \mu E, \quad X, S \succeq 0, \end{aligned} \tag{39}$$

where E denotes the $n \times n$ identity matrix and $\mu > 0$. Suppose the point (X, y, S) is strictly feasible, so $X \succ 0$ and $S \succ 0$. Newton's method amounts to linearizing the system (39), thus yielding the following equation

$$\begin{aligned} \mathbf{Tr}\,(A_i \Delta X) &= 0, \quad i = 1, \cdots, m, \\ -\sum_{i=1}^{m} \Delta y_i A_i &= \Delta S \\ X\Delta S + \Delta X S &= \mu E - XS. \end{aligned} \tag{40}$$

A crucial observation for SDO is that the above Newton system might have no symmetric solution ΔX. Many researchers have proposed different ways of symmetrizing the third equation in the Newton system so that the new system have a unique symmetric solution [20, 21]. In this paper we consider the symmetrization scheme that yields the NT direction [15, 21]. Let us define the matrix

$$P = X^{\frac{1}{2}}(X^{\frac{1}{2}} S X^{\frac{1}{2}})^{-\frac{1}{2}} X^{\frac{1}{2}} = S^{-\frac{1}{2}}(S^{\frac{1}{2}} X S^{\frac{1}{2}})^{\frac{1}{2}} S^{-\frac{1}{2}}, \tag{41}$$

and $D = P^{\frac{1}{2}}$. The matrix D can be used to rescale X and S to the same matrix V defined by ([3, 15, 20, 19])

$$V := \frac{1}{\sqrt{\mu}} D^{-1} X D^{-1} = \frac{1}{\sqrt{\mu}} DSD. \tag{42}$$

Obviously the matrices D and V are symmetric, and positive definite. Also defining

$$\begin{aligned} \bar{A}_i &:= DA_iD; \qquad i = 1, \cdots, m; \\ D_X &:= \frac{1}{\sqrt{\mu}} D^{-1} \Delta X D^{-1}, \quad D_S := \tfrac{1}{\sqrt{\mu}} D \Delta S D. \end{aligned} \tag{43}$$

Then the NT search direction can be written as the solution of the following system

$$\begin{aligned} \mathbf{Tr}\,(\bar{A}_i D_X) &= 0, \ i = 1, \cdots, m, \\ -\sum_{i=1}^{m} \Delta y_i \bar{A}_i &= D_S \\ D_X + D_S &= V^{-1} - V. \end{aligned} \tag{44}$$

Similarly to the case of LO, the new search direction we suggest for SDO is a slight modification of the NT direction, which is defined by the solution of the following system

$$\begin{aligned} \mathbf{Tr}\,(\bar{A}_i D_X) &= 0, \ i = 1, \cdots, m, \\ -\sum_{i=1}^{m} \Delta y_i \bar{A}_i &= D_S \\ D_X + D_S &= V^{-1-\rho} - V, \end{aligned} \tag{45}$$

where we choose $\rho \in [0, 2]$. Then ΔX and ΔS can be calculated from (43). Due to the orthogonality of ΔX and ΔS, it is trivial to see that

$$\mathbf{Tr}\,(D_X D_S) = \mathbf{Tr}\,(D_S D_X) = 0. \tag{46}$$

The proximity measure we use here is

$$\delta\,(XS, \mu) := \left\| V - V^{-1} \right\|, \tag{47}$$

where $\|\cdot\|$ is the Frobenius norm.

The algorithm can be stated as follows.

Large Update Primal-Dual Algorithm for SDO

Input:
A proximity parameter τ;
an accuracy parameter $\varepsilon > 0$;
a variable damping factor α;
a fixed barrier update parameter $\theta \in (0, 1)$;
a strictly feasible (X^0, S^0) and $\mu^0 > 0$ such that $\delta(X^0 S^0; \mu^0) \leq \tau$.

begin
$X := X^0$; $S := S^0$; $\mu := \mu^0$;
while $n\mu \geq \varepsilon$ **do**
begin
$\mu := (1 - \theta)\mu$;
while $\delta(X, S; \mu) \geq \tau$ **do**
Solve the system (45),
begin
$X := X + \alpha \Delta X$;
$S := S + \alpha \Delta S$;
$y := y + \alpha \Delta y$;
end
end
end

Now we begin to estimate the decrease of the proximity after one step. Let us define

$$\sigma^2 = \mathbf{Tr} \left(\left(V - V^{-3} \right) \left(V - V^{-1-\rho} \right) \right). \tag{48}$$

Note that the matrices $V - V^{-3}$ and $V - V^{-1-\rho}$ commute. Hence they admit a similarity transformation that simultaneously diagonalizes both matrices. Then by using similar arguments as in the LO case, one can easily derive the following result

Lemma 4.1 *Let δ and σ be defined by (47) and (48) respectively. It holds*

$$\sigma \geq \|(D_X, D_S)\| \geq \delta,$$

and

$$\left\| V^{-1} \right\| = \lambda_{\max}(V^{-1}) = \frac{1}{\lambda_{\min}(V)} \leq (1 + \sigma)^{\frac{2}{4+\rho}}. \tag{49}$$

Defining δ_+ as the proximity measure after a feasible step, we have

$$\begin{aligned} {\delta_+}^2 &= \mathbf{Tr}\left((V + \alpha D_X)(V + \alpha D_S)\right) - 2n + \mathbf{Tr}\left((V + \alpha D_X)^{-1}(V + \alpha D_S)^{-1}\right) \\ &= \mathbf{Tr}\left(V^2\right) + \alpha \mathbf{Tr}\left(V(D_X + D_S)\right) - 2n + \mathbf{Tr}\left((V + \alpha D_X)^{-1}(V + \alpha D_S)^{-1}\right), \end{aligned} \tag{50}$$

where the equality follows from (46). Our goal is to estimate the decreasing value of

$$f(\alpha) = \delta_+^2 - \delta^2. \tag{51}$$

for a feasible step size α. The main difficulty in the estimation of the function $f(\alpha)$ is to evaluate its first and second derivatives. For this we need some knowledge about matrix functions [2](page 490-491). Suppose the matrix functions $G(t), H(t)$ are differentiable and nonsingular at the point t. Then we have

$$\begin{aligned} \frac{d}{dt}G^{-1}(t) &= -G^{-1}(t)[\frac{d}{dt}G(t)]G^{-1}(t); && (52)\\ \frac{d}{dt}\mathbf{Tr}\,(G(t)) &= \mathbf{Tr}\left(\frac{d}{dt}G(t)\right), && (53)\\ \frac{d}{dt}G(t)H(t) &= [\frac{d}{dt}G(t)]H(t) + G(t)[\frac{d}{dt}H(t)]. && (54)\end{aligned}$$

The following inequalities about matrix eigenvalues are also necessary for our analysis [2](3.3.25, page 183, 3.3.46a,page 192). Suppose that the matrices G, H are symmetric, we have

$$\begin{aligned} |\mathbf{Tr}\,(GH)| &\leq \sum_{i=1}^{n} |\lambda_i(G)\lambda_i(H)|\,, && (55)\\ |\lambda_i(GH)| &\leq \min\left(|\lambda_1(G)\lambda_i(H)|\,, |\lambda_1(H)\lambda_i(G)|\right). && (56)\end{aligned}$$

The inequality (56) also implies that

$$|\mathbf{Tr}\,(GH)| \leq \min\left(|\lambda_1(G)|\sum_{i=1}^{n}|\lambda_i(H)|\,, \quad |\lambda_1(H)|\sum_{i=1}^{n}|\lambda_i(G)|\right). \tag{57}$$

Particularly, if $G_1 \preceq G_2$ and $H \succeq 0$, then it holds

$$\mathbf{Tr}\,(G_1 H) \leq \mathbf{Tr}\,(G_2 H)\,. \tag{58}$$

In fact, for symmetric matrices G, H, it is not difficult to prove the following result which is a refinement of (55).

Lemma 4.2 *Suppose both G and H are symmetric. Then it holds*

$$|\mathbf{Tr}\,(GH)| \leq \mathbf{Tr}\,(|G|\,|H|)\,. \tag{59}$$

Proof: The proof is inductive. First we observe that for any orthogonal matrix Q, it holds $\mathbf{Tr}\left(QGHQ^T\right) = \mathbf{Tr}\,(GH)$. Premultipling by an orthogonal matrix Q and postmultipling by its transpose Q^T if necessary, we can assume without loss of generality that G is diagonal. Now recalling the definition of the operator $|\cdot|$ for matrices, we can claim that $H \preceq |H|$ which implies $|H_{i,i}| \leq |H|_{i,i}$ for all $i = 1, 2, \cdots, n$. It follows

$$|\mathbf{Tr}\,(GH)| = \left|\sum_{i=1}^{n} G_{i,i}H_{i,i}\right| \leq \sum_{i=1}^{n}|G_{i,i}H_{i,i}| \leq \sum_{i=1}^{n}|G_{i,i}|\,|H|_{i,i} = \mathbf{Tr}\,(|G|\,|H|)\,.$$

This completes the proof of the lemma. □

Using (52), (53) and (54) we obtain that

$$f'(0) = -\sigma^2.$$

Now we are going to estimate the second derivative $f''(\alpha)$ of $f(\alpha)$. For notation convenience we also define

$$\bar{D}_x = V^{-\frac{1}{2}} D_X V^{-\frac{1}{2}}, \quad \bar{D}_s = V^{-\frac{1}{2}} D_S V^{-\frac{1}{2}}. \tag{60}$$

We insert here a technical result about the norms of the matrices $\bar{D}_x$ and $\bar{D}_s$.

Lemma 4.3 *Let the matrices $\bar{D}_x$ and $\bar{D}_s$ are defined by (60). Then it holds*

$$\|\bar{D}_x\|^2 + \|\bar{D}_s\|^2 \le \sigma^2 (1+\sigma)^{\frac{4}{4+\rho}}.$$

Proof: From (56) we conclude that

$$|\lambda_i(\bar{D}_x)| \le \frac{1}{\lambda_{\min}(V)} |\lambda_i(D_X)|,$$

and

$$|\lambda_i(\bar{D}_s)| \le \frac{1}{\lambda_{\min}(V)} |\lambda_i(D_S)|.$$

It follows

$$\begin{aligned}
\|\bar{D}_x\|^2 + \|\bar{D}_s\|^2 &= \sum_{i=1}^{n} \left(\lambda_i^2(\bar{D}_x) + \lambda_i^2(\bar{D}_s)\right) \le \frac{1}{\lambda_{\min}(V)^2} \sum_{i=1}^{n} \left(\lambda_i^2(D_X) + \lambda_i^2(D_X)\right) \\
&= \frac{1}{\lambda_{\min}(V)^2} \left(\|D_X\|^2 + \|D_S\|^2\right) = \frac{1}{\lambda_{\min}(V)^2} \left(\|D_X + D_S\|^2\right) \\
&\le \sigma^2 (1+\sigma)^{\frac{4}{4+\rho}},
\end{aligned}$$

where the third equality given by (46), and the last inequality follows from the definition of σ and (49). □

Note that the last term in (50) can be written as

$$\mathbf{Tr}\left(V^{-\frac{1}{2}}(E+\alpha\bar{D}_x)^{-1} V^{-1} (E+\alpha\bar{D}_s)^{-1} V^{-\frac{1}{2}}\right).$$

A useful observation is that the matrices $\bar{D}_x$ and $(E+\alpha\bar{D}_x)^{-1}$ commute, and similarly $\bar{D}_s$ and $(E+\alpha\bar{D}_s)^{-1}$. Now we are ready to state one of our main results in this section.

Lemma 4.4 *Suppose the step size α is strictly feasible. Then it holds*

$$f''(\alpha) \le \frac{3}{\lambda_{\min}(V)^2} \sum_{i=1}^{n} \left(\frac{\lambda_i^2(\bar{D}_x)}{(1-\alpha\lambda_i(|\bar{D}_x|))^3 (1-\alpha\lambda_i(|\bar{D}_s|))} + \frac{\lambda_i^2(\bar{D}_s)}{(1-\alpha\lambda_i(|\bar{D}_s|))^3 (1-\alpha\lambda_i(|\bar{D}_x|))} \right).$$

Proof: By applying (52), (53) and (54) to the function $f(\alpha)$, we obtain

$$\begin{aligned} f''(\alpha) &= 2\mathbf{Tr}\left(V^{-\frac{1}{2}}(E+\alpha\bar{D}_x)^{-3}\bar{D}_x^2V^{-1}(E+\alpha\bar{D}_s)^{-1}V^{-\frac{1}{2}}\right) \\ &+2\mathbf{Tr}\left(V^{-\frac{1}{2}}(E+\alpha\bar{D}_x)^{-1}V^{-1}(E+\alpha\bar{D}_s)^{-3}\bar{D}_s^2V^{-\frac{1}{2}}\right) \\ &+2\mathbf{Tr}\left(V^{-\frac{1}{2}}(E+\alpha\bar{D}_x)^{-2}\bar{D}_xV^{-1}(E+\alpha\bar{D}_s)^{-2}\bar{D}_sV^{-\frac{1}{2}}\right). \end{aligned} \tag{61}$$

We proceed by considering the first term in the above formulae. By the definition of the $|\cdot|$ operator for a matrix, we conclude that

$$E+\alpha\bar{D}_x \succeq E-\alpha\left|\bar{D}_x\right|, \quad E+\alpha\bar{D}_s \succeq E-\alpha\left|\bar{D}_s\right|,$$

which equivalent to

$$(E+\alpha\bar{D}_x)^{-1} \preceq (E-\alpha\left|\bar{D}_x\right|)^{-1}, \quad (E+\alpha\bar{D}_s)^{-1} \preceq (E-\alpha\left|\bar{D}_s\right|)^{-1}.$$

Hence

$$\begin{aligned} (E+\alpha\bar{D}_x)^{-3}\bar{D}_x^2 &= \bar{D}_x(E+\alpha\bar{D}_x)^{-3}\bar{D}_x \preceq \left|\bar{D}_x(E+\alpha\bar{D}_x)^{-3}\bar{D}_x\right| \\ &= \left|\bar{D}_x\right|^2\left|(E+\alpha\bar{D}_x)^{-3}\right| \preceq \left|\bar{D}_x\right|^2(E-\alpha\left|\bar{D}_x\right|)^{-3}. \end{aligned}$$

Since $\left|\bar{D}_x\right|^2=\bar{D}_x^2$, one has

$$V^{-\frac{1}{2}}(E+\alpha\bar{D}_x)^{-3}\bar{D}_x^2V^{-\frac{1}{2}} \preceq V^{-\frac{1}{2}}(E-\alpha\left|\bar{D}_x\right|)^{-3}\bar{D}_x^2V^{-\frac{1}{2}}.$$

It follows

$$\begin{aligned} &\mathbf{Tr}\left(V^{-\frac{1}{2}}(E+\alpha\bar{D}_x)^{-3}\bar{D}_x^2V^{-1}(E+\alpha\bar{D}_s)^{-1}V^{-\frac{1}{2}}\right) \\ &\le \mathbf{Tr}\left(V^{-\frac{1}{2}}(E-\alpha\left|\bar{D}_x\right|)^{-3}\bar{D}_x^2V^{-1}(E+\alpha\bar{D}_s)^{-1}V^{-\frac{1}{2}}\right) \\ &\le \mathbf{Tr}\left(V^{-\frac{1}{2}}(E-\alpha\left|\bar{D}_x\right|)^{-3}\bar{D}_x^2V^{-1}(E-\alpha\left|\bar{D}_s\right|)^{-1}V^{-\frac{1}{2}}\right), \end{aligned}$$

where the inequalities given by (58). Now recalling (56) we can claim that

$$\lambda_i[V^{-\frac{1}{2}}(E-\alpha\left|\bar{D}_x\right|)^{-3}\bar{D}_x^2V^{-\frac{1}{2}}] \le \frac{1}{\lambda_{\min}(V)}\frac{\lambda_i^2(\bar{D}_x)}{[1-\alpha\lambda_i(\left|\bar{D}_x\right|)]^3}, \quad i=1,\cdots,n;$$

and

$$\lambda_i[V^{-\frac{1}{2}}(E-\alpha\left|\bar{D}_s\right|)^{-1}V^{-\frac{1}{2}}] \le \frac{1}{\lambda_{\min}(V)(1-\alpha\lambda_i(\left|\bar{D}_x\right|)}, \quad i=1,\cdots,n.$$

The above two inequalities, combining with (55) yield

$$\begin{aligned} &\mathbf{Tr}\left(V^{-\frac{1}{2}}(E+\alpha\bar{D}_x)^{-3}\bar{D}_x^2V^{-1}(E+\alpha\bar{D}_s)^{-1}V^{-\frac{1}{2}}\right) \\ &\le \mathbf{Tr}\left(V^{-\frac{1}{2}}(E-\alpha\left|\bar{D}_x\right|)^{-3}\bar{D}_x^2V^{-1}(E-\alpha\left|\bar{D}_s\right|)^{-1}V^{-\frac{1}{2}}\right) \\ &\le \frac{1}{\lambda_{\min}(V)^2}\sum_{i=1}^n\frac{\lambda_i^2(\bar{D}_x)}{(1-\alpha\lambda_i(\left|\bar{D}_x\right|))^3(1-\alpha\lambda_i(\left|\bar{D}_s\right|))}. \end{aligned}$$

Similarly we have

$$\begin{aligned}
&\mathbf{Tr}\left(V^{-\frac{1}{2}}(E+\alpha\bar{D}_s)^{-3}\bar{D}_s^2V^{-1}(E+\alpha\bar{D}_x)^{-1}V^{-\frac{1}{2}}\right)\\
&\le \mathbf{Tr}\left(V^{-\frac{1}{2}}(E-\alpha\left|\bar{D}_s\right|)^{-3}\bar{D}_s^2V^{-1}(E-\alpha\left|\bar{D}_x\right|)^{-1}V^{-\frac{1}{2}}\right)\\
&\le \frac{1}{\lambda_{\min}(V)^2}\sum_{i=1}^{n}\frac{\lambda_i^2(\bar{D}_s)}{(1-\alpha\lambda_i(|\bar{D}_s|))^3(1-\alpha\lambda_i(|\bar{D}_x|))}.
\end{aligned}$$

The proof of the Lemma will be finished if we can show that the last term in (61) satisfies the following inequality

$$\begin{aligned}
&\mathbf{Tr}\left(V^{-\frac{1}{2}}(E+\alpha\bar{D}_x)^{-2}\bar{D}_xV^{-1}(E+\alpha\bar{D}_s)^{-2}\bar{D}_sV^{-\frac{1}{2}}\right)\\
&\le \frac{1}{2\lambda_{\min}(V)^2}\sum_{i=1}^{n}\left(\frac{\lambda_i^2(\bar{D}_x)}{(1-\alpha\lambda_i(|\bar{D}_x|))^3(1-\alpha\lambda_i(|\bar{D}_s|))}+\frac{\lambda_i^2(\bar{D}_s)}{(1-\alpha\lambda_i(|\bar{D}_s|))^3(1-\alpha\lambda_i(|\bar{D}_x|))}\right).
\end{aligned} \tag{62}$$

Using the definition of the operator $|\cdot|$ again, one can easily see that

$$\begin{aligned}
V^{-\frac{1}{2}}(E+\alpha\bar{D}_x)^{-2}\bar{D}_xV^{-\frac{1}{2}} &\preceq \left|V^{-\frac{1}{2}}(E+\alpha\bar{D}_x)^{-2}\bar{D}_xV^{-\frac{1}{2}}\right| \preceq V^{-\frac{1}{2}}(E-\alpha\left|\bar{D}_x\right|)^{-2}\left|\bar{D}_x\right|V^{-\frac{1}{2}},\\
V^{-\frac{1}{2}}(E+\alpha\bar{D}_s)^{-2}\bar{D}_sV^{-\frac{1}{2}} &\preceq \left|V^{-\frac{1}{2}}(E+\alpha\bar{D}_s)^{-2}\bar{D}_sV^{-\frac{1}{2}}\right| \preceq V^{-\frac{1}{2}}(E-\alpha\left|\bar{D}_s\right|)^{-2}\left|\bar{D}_s\right|V^{-\frac{1}{2}}.
\end{aligned}$$

These two inequalities, together with (55) and Lemma 4.2 give

$$\begin{aligned}
&\left|\mathbf{Tr}\left(V^{-\frac{1}{2}}(E+\alpha\bar{D}_x)^{-2}\bar{D}_xV^{-1}(E+\alpha\bar{D}_s)^{-2}\bar{D}_sV^{-\frac{1}{2}}\right)\right|\\
&\le \mathbf{Tr}\left(\left|V^{-\frac{1}{2}}(E+\alpha\bar{D}_x)^{-2}\bar{D}_xV^{-\frac{1}{2}}\right|\left|V^{-\frac{1}{2}}(E+\alpha\bar{D}_s)^{-2}\bar{D}_sV^{-\frac{1}{2}}\right|\right)\\
&\le \mathbf{Tr}\left(V^{-\frac{1}{2}}(E-\alpha\left|\bar{D}_x\right|)^{-2}\left|\bar{D}_x\right|V^{-1}(E-\alpha\left|\bar{D}_s\right|)^{-2}\left|\bar{D}_s\right|V^{-\frac{1}{2}}\right)\\
&\le \sum_{i=1}^{n}\frac{1}{\lambda_{\min}(V)^2}\frac{\left|\lambda_i(\bar{D}_x)\lambda_i(\bar{D}_s)\right|}{(1-\alpha\lambda_i(|\bar{D}_x|))^2(1-\alpha\lambda_i(|\bar{D}_s|))^2}\\
&\le \frac{1}{2\lambda_{\min}(V)^2}\sum_{i=1}^{n}\left(\frac{\lambda_i^2(\bar{D}_x)}{(1-\alpha\lambda_i(|\bar{D}_x|))^3(1-\alpha\lambda_i(|\bar{D}_s|))}+\frac{\lambda_i^2(\bar{D}_s)}{(1-\alpha\lambda_i(|\bar{D}_s|))^3(1-\alpha\lambda_i(|\bar{D}_x|))}\right),
\end{aligned}$$

where the last inequality follows from Cauchy inequality. This completes the proof of Lemma 4.4. □

Similar to the LO case, we also define $\omega_i = \sqrt{\lambda_i^2(\bar{D}_x)+\lambda_i^2(\bar{D}_x)}$ and $\omega = \|(\omega_1,\omega_2,\cdots,\omega_n)\|$. A direct consequence of Lemma 4.4 is

$$f''(\alpha) \le \frac{3\omega^2}{\lambda_{\min}(V)^2(1-\alpha\omega)^4},$$

which is the same as in the case of LO, except that $v_{\min}$ is replaced by $\lambda_{\min}(V)$. Because $f'(0) = -\sigma^2$, we can use the same arguments as in the LO case to get

$$f(\alpha) \ \le -\frac{1}{2}\sigma^2\alpha + f_1(\alpha), \tag{63}$$

where

$$f_1(\alpha) = -\frac{1}{2}\sigma^2\alpha + \frac{\omega}{\lambda_{\min}(V)^2}\int_0^{\alpha}\left(\frac{1}{(1-\xi\omega)^3} - 1\right)d\xi,$$

and $f_1(\alpha)$ has the value zero at the point

$$\alpha = \frac{1}{\omega}\left(1 - \frac{1 + (4\eta+9)^{\frac{1}{2}}}{2(\eta+2)}\right) = \frac{3 + 2\eta - (4\eta+9)^{\frac{1}{2}}}{2\omega(\eta+2)}, \tag{64}$$

where

$$\eta = \frac{\sigma^2\lambda_{\min}(V)^2}{\omega}. \tag{65}$$

For this α we have

$$f(\alpha) \leq -\frac{3 + 2\eta - (4\eta+9)^{\frac{1}{2}}}{4\omega(\eta+2)}\sigma^2.$$

Our next result estimates the above-defined constant η and the step size α in terms of σ.

Lemma 4.5 *Let the constants η be defined by (65) and α by (64). If $\delta \geq 1$ then it holds*

$$\eta^{-1} \leq 2\,(1+\sigma)^{\frac{2-\rho}{4+\rho}}, \tag{66}$$

and

$$\alpha \geq \frac{2}{15}\sigma^{-1}(1+\sigma)^{\frac{\rho-4}{\rho+4}}. \tag{67}$$

Proof: First note that from Lemma 4.3 and its proof we obtain

$$\omega \leq \frac{\sigma}{\lambda_{\min}(V)} \leq \sigma(1+\sigma)^{\frac{2}{4+\rho}}. \tag{68}$$

The above relation means

$$\eta = \frac{\sigma^2\lambda_{\min}(V)^2}{\omega} \geq \frac{\sigma^2}{\omega} \geq \sigma$$

if $\lambda_{\min}(V) \geq 1$. Hence (66) is true whenever $\lambda_{\min}(V) \geq 1$. Now we consider the case $\lambda_{\min}(V) < 1$. Using (68), together with Lemma 4.1, (64) and (65), and by following an analogous process as in the proof of (37) in Lemma 3.9, one gets the desirable inequality (66).

By using (66) and (68), and following similar arguments as in the proof of Lemma 3.9, one can easily prove (67). This completes the proof of the lemma. □

Now we can state our another main result in this section, which is a direct consequence of (63) and Lemma 4.5.

Theorem 4.6 *Let the function $f(\alpha)$ be defined by (51) with $\delta \geq 1$. Then the step size $\alpha = \frac{3+2\eta-(4\eta+9)^{\frac{1}{2}}}{2\omega(\eta+2)}$ defined by (65) is feasible. Moreover it holds*

$$f(\alpha) \leq \frac{1}{2}f'(0)\alpha \leq -\frac{1}{30}\delta^{\frac{2\rho}{4+\rho}}.$$

We proceed to estimate the complexity of the algorithm. If the damping factor α is defined as in Theorem 4.6, namely $\alpha = \frac{3+2\eta-(4\eta+9)^{\frac{1}{2}}}{2\omega(\eta+2)}$. Then each damped Newton step reduces the squared proximity by at least $-\frac{1}{30}\delta^{\frac{2\rho}{4+\rho}}$. From Lemma 3.11 we know that the proximity $\delta(xs,\mu_+) \leq \frac{\delta(xs,\mu)+\theta\sqrt{n}}{\sqrt{1-\theta}}$ after the update of μ. It follows from Lemma 3.12 that after an update of the barrier parameter no more than

$$\left\lceil \frac{15(4+\rho)}{2} \left(\frac{\tau+\theta\sqrt{n}}{1-\theta} \right)^{\frac{8}{4+\rho}} \right\rceil$$

iterations are needed to reach $\delta(xs,\mu_+) \leq \tau$ again. This means the algorithm has an

$$\left\lceil \frac{15(4+\rho)}{2} \left(\frac{\tau+\theta\sqrt{n}}{\sqrt{1-\theta}} \right)^{\frac{8}{4+\rho}} \right\rceil \left\lceil \frac{1}{\theta} \log \frac{n}{\varepsilon} \right\rceil$$

polynomial complexity bound.

For large update IPMs, omitting the round off brackets in the above estimation, one can see that the complexity of our algorithm for SDO is in the order of $O\left(n^{\frac{4}{4+\rho}} \log \frac{n}{\varepsilon}\right)$.

5 Concluding Remarks

A new class of search directions was proposed for solving LO and SDO problems. The new directions are a slight modification of the classical Newton direction. By using some new analysis tools, we proved that the large update method based on the new direction has a complexity of order $O\left(n^{\frac{4}{4+\rho}} \log \frac{n}{\varepsilon}\right)$. It is worthwhile to note that a simple idea, to change slightly the corrector direction improves the complexity of the algorithm. This gives rise to some interesting issues. The first issue is whether the new algorithm works well in practice and how to incorporate the idea in the paper into the implementation of IPMs? For instance, in the Mehrotra-type predictor-corrector algorithm [11], the target we used is μe (or μE for SDO), what will happen if we replace this by the new target $v^{-\rho}$? We also mention that we have implemented a simple version of our algorithm, and tested on few problems. Our preliminary numerical results show that the algorithm is promising. Nevertheless, much more work is needed to test the new approach.

The second question is related to the proximity and the search direction. As we claimed in the introduction: 'The proximity is crucial for both the quality and elegance of the analysis'. In the preparation of this work, we tried to give a proof of the complexity of the algorithm based on the logarithmic barrier approach. However, the complexity we obtained is not as good as we presented here. As we observed in the introduction, when $\rho = 2$, the right hand side of the equation system defining the new search direction is the negative gradient of the proximity we used. It is also of interests to note that the right hand side defining the classical Newton direction is the negative gradient of the proximity based on logarithmic barrier approach. This indicates some interrelations between the search direction and the proximity used in the analysis. Will a new analysis based on a new proximity give a better complexity for the standard Newton method? Are there new IPMs with large updates whose complexity is equal to or less than the best known complexity for IPMs? The complexity of our algorithm depends on the parameter

ρ. In this work we were forced to restrict ourselves to the case $\rho \in [0, 2]$. If we could remove this restriction, we might be able to approach the best known complexity bound as close as we wish by letting ρ goes to infinity. This will be a topic for further research.

Our third question is about the algorithm for SDO. The new search direction in this paper is based on the NT symmetrizing scheme: is it possible to design similar algorithms using other schemes? What is the complexity of these new algorithms? This is a topic deserving more research. Lastly we would like to mention that there are also other ways to extend our results here, for instance to study the local convergence properties of these algorithms. This is particularly important since usually, when the point is close to the solution set, then the performance of a algorithm will be determined by its local convergence properties. It may be also worthwhile to build similar algorithms for classes of linear complementarity problems and convex programming.

References

[1] E.D. Andersen, J. Gondzio, Cs. Mészáros, and X. Xu. Implementation of interior point methods for large scale linear programming. In T. Terlaky, editor, *Interior Point Methods of Mathematical Programming*, pages 189–252. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

[2] R.A. Horn, C.R. Johnson. Topics in Matrix Analysis. Cambridge University Press, 1991.

[3] E. de Klerk. *Interior Point Methods for Semidefinite Programming.* Ph.D. Thesis, Faculty of ITS/TWI, Delft University of Technology, The Netherlands, 1997.

[4] P. Hung and Y. Ye. An asymptotically $O(\sqrt{n}L$-iteration path-following linear programming algorithm that uses long steps. *Siam J. on Optimization*, 6:570-586,1996.

[5] B. Jansen, C. Roos, T. Terlaky, and J.-Ph. Vial. Primal-dual algorithms for linear programming based on the logarithmic barrier method. *Journal of Optimization Theory and Applications*, 83:1–26, 1994.

[6] B. Jansen, C. Roos, T. Terlaky, and Y. Ye. Improved complexity using higher order correctors for primal-dual Dikin affine scaling, *Mathematical Programming, Series B*, 76:117–130, 1997.

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

[8] M. Kojima, S. Mizuno, and A. Yoshise. A primal-dual interior point algorithm for linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming: Interior Point and Related Methods*, pages 29–47. Springer Verlag, New York, 1989.

[9] M. Kojima, N. Megiddo, T. Noma and A. Yoshise. *A unified approach to interior point algorithms for linear complementarity problems*, volume 538 of *Lecture Notes in Computer Science.* Springer Verlag, Berlin, Germany, 1991.

[10] N. Megiddo. Pathways to the optimal set in linear programming. In N. Megiddo, editor, *Progress in Mathematical Programming: Interior Point and Related Methods*, pages 131–158. Springer Verlag, New York, 1989. Identical version in : *Proceedings of the 6th Mathematical Programming Symposium of Japan, Nagoya, Japan*, pages 1–35, 1986.

[11] S. Mehrotra. On the implementation of a (primal-dual) interior point method. *SIAM Journal on Optimization*, 2(4):575–601, 1992.

[12] S. Mehrotra and Y. Ye. On finding the optimal facet of linear programs. *Mathematical Programming*, 62:497–515, 1993.

[13] S. Mizuno and A. Nagasawa. A primal-dual affine scaling potential reduction algorithm for linear programming. *Mathematical Programming*, 62:119-131, 1993.

[14] R.D.C. Monteiro, I. Adler and M.G.C. Resende. A polynomial-time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series. *Mathematics of Operations Research*, 15:191–214, 1990.

[15] Y.E. Nesterov and M.J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Mathematics of Operations Research*, 22(1):1–42, 1997.

[16] J. Peng, C. Roos and T. Terlaky. New complexity analysis of the primal-dual Newton method for linear optimization. Technical Report No. 98-05, Faculty of Technical Mathematics and Information, Delft University of Technology, The Netherlands, 1998. To appear in *Annals of Operations Research*

[17] C. Roos, T. Terlaky, and J.-Ph.Vial. *Theory and Algorithms for Linear Optimization. An Interior Approach.* John Wiley & Sons, Chichester, UK, 1997.

[18] G. Sonnevend. An "analytic center" for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. In A. Prekopa, J. Szelezsan, and B. Strazicky, editors, *System Modelling and Optimization : Proceedings of the 12th IFIP-Conference held in Budapest, Hungary, September 1985*, volume 84 of *Lecture Notes in Control and Information Sciences*, pages 866–876. Springer Verlag, Berlin, West–Germany, 1986.

[19] J.F. Sturm, S. Zhang, Symmetric primal-dual path following algorithms for semidefinite programming, Technical Report 9554/A, Tinbergen Institute, Erasmus University, Rotterdam, The Netherlands, 1995.

[20] M.J. Todd, A study of search directions in primal-dual interior-point methods for semidefinite programming, Technical Report 1205, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853, October 1997.

[21] M.J. Todd, K.C. Toh and R.H. Tütüncü, On the Nesterov-Todd direction in semidefinite programming, *SIAM J. Optimization*, 8(1998), pp. 769-796.

[22] S. J. Wright. *Primal-Dual Interior-Point Methods.* SIAM, Philadelphia, USA, 1997.

[23] Y. Ye. On the finite convergence of interior-point algorithms for linear programming. *Mathematical Programming*, 57:325–335, 1992.

[24] Y. Ye. *Interior Point Algorithms, Theory and Analysis.* John Wiley & Sons, Chichester, UK, 1997.

[25] L.L. Zhang and Y. Zhang. On polynomiality of the Mehrotra-type predictor-corrector interior-point algorithms. *Mathematical Programming*, 68:303-318, 1995.

[26] G.Y. Zhao. Interior point algorithms for linear complementarity problems based on large neighborhoods of the central path. *SIAM J. on Optimization*, 8:397-413, 1998.