

# Primal-Dual Interior-Point Algorithms for Semidefinite Optimization Based on a Simple Kernel Function <sup>\*</sup>

G. Q. Wang<sup>†</sup>   Y. Q. Bai<sup>‡</sup>   C. Roos<sup>‡</sup>

May 22, 2004

<sup>†</sup> *Department of Mathematics, College Science,  
Shanghai University, Shanghai, 200436.*

<sup>‡</sup> *Faculty of Electrical Engineering, Mathematics and Computer Science,  
Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands  
e-mail: guoq-wang@hotmail.com, [y.bai, c.roos]@ewi.tudelft.nl*

## Abstract

Interior-point methods (IPMs) for semidefinite optimization (SDO) have been studied intensively, due to their polynomial complexity and practical efficiency. Recently, J.Peng et al. [14, 15] introduced so-called self-regular kernel (and barrier) functions and designed primal-dual interior-point algorithms based on self-regular proximity for linear optimization (LO) problems. They have also extended the approach for LO to SDO. In this paper we present a primal-dual interior-point algorithm for SDO problems based on a simple kernel function which was first introduced in [3]. The kernel function in this paper is not self-regular due to its growth term increasing linearly. We derive the complexity analysis for algorithms with large- and small-update methods. The complexity bounds are  $O(qn) \log \frac{n}{\epsilon}$  and  $O(q^2 \sqrt{n}) \log \frac{n}{\epsilon}$ , respectively, which are as good as those in linear case.

**Keywords:** semidefinite optimization, interior-point methods, primal-dual methods, large- and small-update methods, polynomial complexity.

*AMS Subject Classification:*    **90C22, 90C31**

## 1 Introduction

Semidefinite optimization problems (SDO) are convex optimization problems over the intersection of an affine set and the cone of positive semidefinite matrices. In the past decade, SDO has been one of the most active research areas in mathematical programming. There are two major factors that are responsible for this increased interest in SDO. Firstly, SDO has found numbers applications in various fields, such as statistics, structural design, electrical engineering and combinatorial optimization. Secondly, efficient new algorithms, interior-point methods (IPMs) have led to increased interest both in the application and

---

<sup>\*</sup>The second author is on leave from the Department of Mathematics, Shanghai University, Shanghai 200436, China. She kindly acknowledges the support of Dutch Organization for Scientific Researches (NWO grant 613.000.110). Corresponding author.

the research of SDO. In this paper we deal with so-called primal-dual IPMs. It is generally agreed that these IPMs are most efficient from a computational part of view [2].

Many researchers have studied SDO and achieved plentiful and beautiful results. For an overview of these results we refer to [18]. Several interior-point methods designed for LO have been successfully extended to SDO. In particular, primal-dual interior-point algorithms are of highly efficiency both in theory and in practice. An important work in this direction is the paper of Nesterov and Todd [13] who showed that the primal-dual algorithm maintains its theoretical efficiency when the nonnegativity constrains in LO are replaced by a convex cone, as long as the cone is homogeneous and self-dual, or in the terminology of Nesterov and Todd, as long as the cone is self-scaled. Recently, J.Peng et al. [14, 15] introduced so-called self-regular kernel functions. The prototype self-regular function in [14, 15] is given by

$$\Upsilon_{p,q}(t) = \frac{t^{p+1} - 1}{p(p+1)} + \frac{t^{1-q} - 1}{q(q-1)} + \frac{p-q}{pq}(t-1), \quad (1)$$

where  $p \geq 1$  and  $q > 1$ . The parameter  $p$  is called the growth degree and  $q$  the barrier degree of the kernel function. Based on self-regular functions, they designed primal-dual interior-point algorithms for LO problems and also extended the approach to SDO ones. The complexity bounds obtained by them are  $O(\sqrt{n}) \log \frac{n}{\epsilon}$  and  $O(\sqrt{n} \log n) \log \frac{n}{\epsilon}$ , for small-update methods and large-update methods, respectively, which are currently the best known bounds. Motivated by their work, in this paper we present a primal-dual interior-point algorithm for SDO based on the kernel function:

$$\psi(t) = t - 1 + \frac{t^{1-q} - 1}{q - 1}, \quad t > 0, \quad (2)$$

where  $q > 1$  is a parameter. This kernel function determines a matrix-valued function  $\psi(V)$ , which is defined in the usual way (see Section 2.1), and a real-valued function  $\Psi(V)$  defined as follows.

$$\Psi(V) := \text{Tr}(\psi(V)), \quad V \in \mathbf{S}_{++}^n. \quad (3)$$

The kernel function  $\psi(t)$  was introduced in [3] for LO. See also [4]. The kernel function  $\psi(t)$  does not belong to the family of self-regular functions, and as a consequence a large part of the analysis tools in [14, 15] do no longer apply. In this paper we develop some new analysis tools for the analysis of primal-dual interior-point algorithm based on  $\psi(t)$ . Via the special definition of the matrix functions  $\psi(V)$  and  $\Psi(V)$  and nice properties of  $\psi(t)$  the analysis in this paper is much simpler than in [14, 15]. We derive the complexity analysis for algorithms with large- and small-update methods. The complexity bounds are as good as bounds for the linear case.

The outline of the paper is as follows. In Section 2 we first describe special matrix functions used in later sections. Then we briefly recall the basic concepts of interior-point methods for SDO, such as central path, NT-search directions, etc.. In Section 3, we describe a primal-dual interior-point algorithm based on  $\Psi(V)$  for SDO. In Section 4, we present the properties of  $\psi(t)$  and study the matrix functions  $\psi(V)$  and  $\Psi(V)$ . We analyze the algorithm to derive the complexity bound with large- and small-update methods in Section 5. Finally, some concluding remarks follow in Section 6.

Some notations used throughout the paper are as follows.  $\mathbf{R}^n$ ,  $\mathbf{R}_+^n$  and  $\mathbf{R}_{++}^n$  denote the set of vectors with  $n$  components, the set of nonnegative vectors and the set of positive vectors, respectively.  $\|\cdot\|$  denotes the Frobenius norm for matrices, and the 2-norm for vectors.  $\mathbf{R}^{m \times n}$  is the space of all  $m \times n$  matrices.  $\mathbf{S}^n$ ,  $\mathbf{S}_+^n$  and  $\mathbf{S}_{++}^n$  denote the cone of symmetric, symmetric positive semidefinite and symmetric positive definite  $n \times n$  matrices, respectively.  $E$  denotes  $n \times n$  identity matrix. The Löwner partial order " $\succeq$ " on positive semidefinite (or positive definite) matrices means  $A \succeq B$  ( $A \succ B$ ) if  $A - B$  is positive semidefinite (or positive definite). We use the matrix inner product  $A \bullet B = \text{Tr}(A^T B)$ . For any symmetric positive definite matrix  $Q \in \mathbf{S}_{++}^n$ , the expression  $Q^{\frac{1}{2}}$  denotes its symmetric square root. Similarly we can define the power  $Q^\alpha$  for any  $Q \succ 0$  and  $\alpha \in \mathbf{R}$ . When  $\lambda$  is a vector we denote the diagonal matrix  $\Lambda$  with entries  $\lambda_i$  by  $\text{diag}(\lambda)$ . For any  $V \in \mathbf{S}_{++}^n$ , we denote by  $\lambda(V)$  the vector of eigenvalues of  $V$  arranged in non-increasing order, that is,  $\lambda_1(V) \geq \lambda_2(V) \geq \dots \geq \lambda_n(V)$ . For any matrix  $M$ , we denote by  $\sigma_1(M) \geq \sigma_2(M) \geq \dots \geq \sigma_n(M)$  the singular values of  $M$ . Especially if  $M$  is symmetric, then one has  $\sigma_i(M) = |\lambda_i(M)|$ ,  $i = 1, 2, \dots, n$ .

## 2 Preliminaries

### 2.1 Special matrix functions

To introduce special matrix functions which will be useful for designing primal-dual interior-point algorithm, first of all, let us recall some known facts from linear algebra.

**Theorem 2.1** [Spectral theorem for symmetric matrices in [18]] *The real  $n \times n$  matrix  $A$  is symmetric if and only if there exists an orthogonal basis with respect to which  $A$  is real and diagonal, i.e. if and only if there exists a matrix  $U \in \mathbf{R}^{n \times n}$  such that  $U^T U = E$  and  $U^T A U = \Lambda$  where  $\Lambda$  is a diagonal matrix.  $\square$*

The columns  $u_i$  of  $U$  are the eigenvectors of  $A$ , satisfying

$$A u_i = \lambda_i u_i, \quad i = 1, \dots, n,$$

where  $\lambda_i$  is the  $i$ -th diagonal entry of  $\Lambda$ .

Now we are ready to show how a matrix function can be obtained from  $\psi(t)$ .

**Definition 2.2** *Let  $V \in \mathbf{S}_{++}^n$  be any symmetric  $n \times n$  matrix and*

$$V = Q^T \text{diag}(\lambda(V)) Q$$

where  $Q$  is any orthonormal matrix that diagonalizes  $V$ . Let  $\psi(t)$  be defined as in (2). The matrix function  $\psi(V) : \mathbf{S}_{++}^n \rightarrow \mathbf{S}^n$  is defined by

$$\psi(V) = Q^T \text{diag}(\psi(\lambda_1(V)), \psi(\lambda_2(V)), \dots, \psi(\lambda_n(V))) Q. \quad (4)$$

Then we define a matrix function  $\Psi(V) : \mathbf{S}_{++}^n \rightarrow \mathbf{R}_+$  as follows.

**Definition 2.3**

$$\Psi(V) = \text{Tr}(\psi(V)) = \sum_{i=1}^n \psi(\lambda_i(V)) \quad (5)$$

where  $\psi(V)$  is given by (4).

Since  $\psi(t)$  is triple differentiable, the derivatives  $\psi'(t)$ ,  $\psi''(t)$  and  $\psi'''(t)$  are well-defined for  $t > 0$ . Hence, replacing  $\psi(\lambda_i(V))$  in (4) by  $\psi'(\lambda_i(V))$ ,  $\psi''(\lambda_i(V))$  and  $\psi'''(\lambda_i(V))$ , respectively, for each matrix, we obtain that the matrix functions  $\psi'(V)$ ,  $\psi''(V)$  and  $\psi'''(V)$  already are defined.  $\psi(V)$  depends only on the restriction of  $\psi(t)$  on the spectrum (the set of eigenvalues) of  $V$  (See p.278 in [5]).

Just as in the linear case, we call  $\psi(t)$  the kernel function for the matrix functions  $\psi(V)$  and  $\Psi(V)$ . Moreover, we call both  $\psi(V)$  and  $\Psi(V)$  matrix barrier functions.

The following lemma can be found in [15]. For completeness' sake we include the proof below.

**Lemma 2.4**  $\Psi(V)$  is strictly convex with respect to  $V \succ 0$  and vanishes at its global minimal point  $V = E$ , i.e.,  $\psi(E) = \psi'(E) = 0_{n \times n}$ . Moreover,  $\Psi(E) = 0$ .

**Proof:** Firstly, we prove  $\Psi(V)$  is strictly convex for  $V \succ 0$ , i.e., for any  $V_1 \neq V_2 \succ 0$ , the following inequality holds

$$\Psi\left(\frac{V_1 + V_2}{2}\right) < \frac{1}{2}(\Psi(V_1) + \Psi(V_2)). \quad (6)$$

Since both  $V_1, V_2$  are positive definite, so is the matrix  $\frac{1}{2}(V_1 + V_2)$ . Using Theorem 2.1, there exists orthogonal matrices  $Q, \tilde{Q}_1, \tilde{Q}_2$  and diagonal matrices  $\Lambda, \Lambda_1, \Lambda_2$  such that

$$\Lambda = \text{diag}(\lambda) = \frac{1}{2} Q (V_1 + V_2) Q^T = \frac{1}{2} Q V_1 Q^T + \frac{1}{2} Q V_2 Q^T = \frac{1}{2} Q \tilde{Q}_1 \Lambda_1 \tilde{Q}_1^T Q^T + \frac{1}{2} Q \tilde{Q}_2 \Lambda_2 \tilde{Q}_2^T Q^T.$$

Denote that  $Q_1 = Q\tilde{Q}_1$  and  $Q_2 = Q\tilde{Q}_2$ , we have

$$\Lambda = \frac{1}{2}(Q_1\Lambda_1Q_1^T + Q_2\Lambda_1Q_2^T), \quad (7)$$

where  $\Lambda_1 = \text{diag}(\lambda^1)$  and  $\Lambda_2 = \text{diag}(\lambda^2)$  are positive diagonal matrices, and  $\lambda, \lambda^1, \lambda^2$  are vectors whose components are the eigenvalues of  $\frac{1}{2}(V_1 + V_2)$ ,  $V_1$  and  $V_2$ , respectively. We denote

$$M = ([Q_1]_{ij}^2)_{n \times n} \quad \text{and} \quad N = ([Q_2]_{ij}^2)_{n \times n}.$$

From (7) one can easily get

$$\lambda = \frac{1}{2}(M\lambda^1 + N\lambda^2).$$

Using the orthogonality of the matrices  $Q_1$  and  $Q_2$ , one has

$$\begin{aligned} \sum_{i=1}^n M_{ij} &= \sum_{i=1}^n N_{ij} = 1, \quad j = 1, 2, \dots, n, \\ \sum_{j=1}^n M_{ij} &= \sum_{j=1}^n N_{ij} = 1, \quad i = 1, 2, \dots, n. \end{aligned}$$

So  $M$  and  $N$  are double stochastic matrices. From the strict convexity of  $\psi(t)$  and the fact that  $V_1 \neq V_2 \succ 0$ . We readily obtain

$$\Psi\left(\frac{V_1 + V_2}{2}\right) = \sum_{i=1}^n \psi(\lambda_i) = \sum_{i=1}^n \psi\left(\frac{M_i\lambda^1 + N_i\lambda^2}{2}\right) < \frac{1}{2}\left(\sum_{i=1}^n \psi(M_i\lambda^1) + \sum_{i=1}^n \psi(N_i\lambda^2)\right),$$

where  $M_i(N_i)$  denotes the  $i$ -th row of the  $M(N)$ , using the orthogonality of the matrices  $Q_1$  and  $Q_2$  and the convexity of  $\psi(t)$  again. Since  $M$  is double stochastic, we have

$$\sum_{i=1}^n \psi(M_i\lambda^1) \leq \sum_{i=1}^n \sum_{j=1}^n M_{ij} \psi(\lambda_j^1) = \sum_{j=1}^n \psi(\lambda_j^1) = \Psi(V_1),$$

similarly, one has

$$\sum_{i=1}^n \psi(N_i\lambda^2) \leq \Psi(V_2).$$

The above two inequalities yield the desired relation (6). Since  $\psi(1) = \psi'(1) = 0$ , by using (4) one can easily verify other statements.  $\square$

Moreover, we need to deal with another concepts relevant to matrix functions in matrix theory. The related results of matrix of function can be found in [7, 9].

**Definition 2.5** A matrix  $M(t)$  is said to be a matrix of functions if each entry of  $M(t)$  is a function of  $t$ , i.e.,  $M(t) = [M_{ij}(t)]$ .

First, we state two inequalities, which are used in the proof of Lemma 5.2. If  $M, N \in \mathbf{S}^n$ , then

$$|\text{Tr}(MN)| \leq |\lambda_1(M)| \sum_{i=1}^n |\lambda_i(N)|. \quad (8)$$

Furthermore, if  $M_1 \preceq M_2$  and  $N \succeq 0$ , then

$$\text{Tr}(M_1N) \leq \text{Tr}(M_2N). \quad (9)$$

The usual concepts of continuity, differentiability and integrability can be naturally extended to matrices of functions, by interpreting them entry-wise. Let  $M(t)$  and  $N(t)$  be two matrices of functions. Then it can easily be understood that

$$\frac{d}{dt}M(t) = \left[ \frac{d}{dt}M_{ij}(t) \right] = M'(t), \quad (10)$$

$$\frac{d}{dt}\text{Tr}(M(t)) = \text{Tr} \left( \frac{d}{dt}M(t) \right) = \text{Tr}(M'(t)), \quad (11)$$

$$\frac{d}{dt}\text{Tr}(\psi(M(t))) = \text{Tr}(\psi'(M(t))M'(t)), \quad (12)$$

$$\frac{d}{dt}(M(t)N(t)) = \left[ \frac{d}{dt}M(t) \right] N(t) + M(t) \left[ \frac{d}{dt}N(t) \right] = M'(t)N(t) + M(t)N'(t). \quad (13)$$

**Remark 2.6** *In the rest of the section, when we use the function  $\psi(\cdot)$  and its derivatives  $\psi'(\cdot)$  and  $\psi''(\cdot)$ , these denote matrix functions if the argument is a matrix and a univariate function if the argument is in  $\mathbf{R}_+$ .*

## 2.2 The central path for SDO

We consider the standard form for SDO problems:

$$\begin{aligned} (P) \quad & \text{minimize} && C \bullet X \\ & \text{subject to} && A_i \bullet X = b_i, \quad i = 1, 2, \dots, m, \quad X \succeq 0. \end{aligned} \quad (14)$$

and its dual:

$$\begin{aligned} (D) \quad & \text{maximize} && b^T y \\ & \text{subject to} && \sum_{i=1}^m y_i A_i + S = C, \quad S \succeq 0. \end{aligned} \quad (15)$$

where each  $A_i \in \mathbf{S}^n$ ,  $b \in \mathbf{R}^m$ , and  $C \in \mathbf{S}^n$ . The matrices  $A_i$  are linearly independent.

We assume that (P) and (D) satisfy the interior-point condition IPC, i.e., there exists  $X \in \mathcal{P}$ ,  $S \in \mathcal{D}$  with  $X \succ 0$ ,  $S \succ 0$ , where  $\mathcal{P}$  and  $\mathcal{D}$  denote the feasible set of problem (P) and (D), respectively. Under the assumption of IPC, the optimality conditions for (P) and (D) can be written as follows.

$$\begin{aligned} & A_i \bullet X = b_i, \quad i = 1, 2, \dots, m, \quad X \succeq 0, \\ & \sum_{i=1}^m y_i A_i + S = C, \quad S \succeq 0, \\ & XS = 0. \end{aligned} \quad (16)$$

We modify the above system by relaxing the third equation as follows.

$$\begin{aligned} & A_i \bullet X = b_i, \quad i = 1, 2, \dots, m, \quad X \succeq 0, \\ & \sum_{i=1}^m y_i A_i + S = C, \quad S \succeq 0, \\ & XS = \mu E, \end{aligned} \quad (17)$$

with  $\mu > 0$  and  $E$  is the  $n \times n$  unit matrix. Under the assumption that (P) and (D) satisfy the IPC (this can be achieved via the so-called self-dual embedding technique introduced by E. de Klerk et al. in

[18]), the system (17) has a unique solution, denoted by  $(X(\mu), y(\mu), S(\mu))$ . We call  $X(\mu)$  the  $\mu$ -center of  $(P)$  and  $(y(\mu), S(\mu))$  the  $\mu$ -center of  $(D)$ . The set of  $\mu$ -centers (with  $\mu$  running through positive real numbers) gives a homotopy path, which is called *the central path* of  $(P)$  and  $(D)$ . If  $\mu \rightarrow 0$  then the limit of the central path exists and since the limit points satisfy the complementarity condition, the limit yields optimal solutions for  $(P)$  and  $(D)$ . The relevance of the central path for LO and SDO has been discussed in many papers, see, e.g. [18], [10], and [17], etc..

### 2.3 Search directions determined by kernel functions

The core idea of a primal-dual interior-point algorithm is to follow the central path and to approach the optimal set of SDO problems by letting  $\mu$  go to zero. A direct application of Newton's method to (17) produces the following equations for the search direction  $\Delta X, \Delta y$  and  $\Delta S$ :

$$\begin{aligned} A_i \bullet \Delta X &= 0, \quad i = 1, 2, \dots, m, \\ \sum_{i=1}^m \Delta y_i A_i + \Delta S &= 0, \\ X \Delta S + \Delta X S &= \mu E - X S. \end{aligned} \tag{18}$$

It can be showed that this system has a unique solution, see [18]. We can rewrite system (18) as follows.

$$\begin{aligned} A_i \bullet \Delta X &= 0, \quad i = 1, 2, \dots, m, \\ \sum_{i=1}^m \Delta y_i A_i + \Delta S &= 0, \\ \Delta X + X \Delta S S^{-1} &= \mu S^{-1} - X. \end{aligned} \tag{19}$$

It is obvious that  $\Delta S$  is symmetric due to the second equation in (19). However, a crucial observation is that  $\Delta X$  is not necessary symmetric because  $X \Delta S S^{-1}$  may be not symmetric. Many researchers have proposed methods for symmetrizing the third equation in the above Newton system such that the resulting new system has a unique symmetric solution. Among them, the following three directions are the most popular ones, the direction introduced by Alizadeh, Haeberly, Overton (AHO) in [1], by Helmberg, et al., Kojima et al., and Monteiro (HKM) in [6, 8, 11], Nesterov and Todd (NT) in [12, 13], respectively, called AHO, HKM, and NT directions. In this paper we use the symmetrization scheme from which the NT direction [12] is derived. One important reason for this is that the NT scaling technique transfers the primal variable  $X$  and the dual  $S$  into the same space: the so-called  $V$ -space. If we apply the NT-symmetrized scheme, namely, the term  $X \Delta S S^{-1}$  in the third equation is replaced by  $P \Delta S P^T$ , where

$$P := X^{\frac{1}{2}} (X^{\frac{1}{2}} S X^{\frac{1}{2}})^{-\frac{1}{2}} X^{\frac{1}{2}} = S^{-\frac{1}{2}} (S^{\frac{1}{2}} X S^{\frac{1}{2}})^{\frac{1}{2}} S^{-\frac{1}{2}}.$$

Then the above system is replaced by the system

$$\begin{aligned} A_i \bullet \Delta X &= 0, \quad i = 1, 2, \dots, m, \\ \sum_{i=1}^m \Delta y_i A_i + \Delta S &= 0, \\ \Delta X + P \Delta S P^T &= \mu S^{-1} - X. \end{aligned} \tag{20}$$

Obviously, now  $\Delta X$  is a symmetric matrix and system (20) still has a unique solution (See [18]). Let  $D = P^{\frac{1}{2}}$ . Then matrix  $D$  can be used to scale  $X$  and  $S$  to the same matrix  $V$  because

$$V := \frac{1}{\sqrt{\mu}} D^{-1} X D^{-1} = \frac{1}{\sqrt{\mu}} D S D. \tag{21}$$

Therefore we have

$$V^2 := \frac{1}{\mu} D^{-1} X S D. \quad (22)$$

Note that the matrices  $D$  and  $V$  are symmetric and positive definite.

Let us further define

$$\bar{A}_i := D A_i D, \quad i = 1, 2, \dots, m; \quad D_X := \frac{1}{\sqrt{\mu}} D^{-1} \Delta X D^{-1}; \quad D_S := \frac{1}{\sqrt{\mu}} D \Delta S D. \quad (23)$$

Then it follows from (20) that the (scaled) NT search direction  $(D_X, \Delta y, D_S)$  is satisfied the system

$$\begin{aligned} \bar{A}_i \bullet D_X &= 0, \quad i = 1, 2, \dots, m, \\ \sum_{i=1}^m \Delta y_i \bar{A}_i + D_S &= 0, \\ D_X + D_S &= V^{-1} - V. \end{aligned} \quad (24)$$

So far we describe the scheme that defines the classical NT direction. Now, following [15] we change to the new approach of this paper. Given the kernel function  $\psi(t)$  and the associated  $\psi(V)$  and  $\psi'(V)$  as defined in Definition 2.2, instead of the  $V^{-1} - V$ , we replace the right-hand side in the third equation in (24) by  $-\psi'(V)$ . Thus we consider the following system.

$$\begin{aligned} \bar{A}_i \bullet D_X &= 0, \quad i = 1, 2, \dots, m, \\ \sum_{i=1}^m \Delta y_i \bar{A}_i + D_S &= 0, \\ D_X + D_S &= -\psi'(V). \end{aligned} \quad (25)$$

It is straightforward to verify the system (25) has a unique solution. Having  $D_X$  and  $D_S$ , we can calculate  $\Delta X$  and  $\Delta S$  by (23). We will argue below that the matrix function  $\psi(V)$  determines in a natural way an interior-point algorithm. Since  $D_X$  and  $D_S$  are orthogonal, that is,

$$\text{Tr}(D_X D_S) = \text{Tr}(D_S D_X) = 0, \quad (26)$$

we have

$$D_X = D_S = 0_{n \times n} \Leftrightarrow \psi'(V) = 0_{n \times n} \Leftrightarrow V = E \Leftrightarrow \Psi(V) = 0,$$

i.e., if and only if  $X S = \mu E$ , that is, i.e., if and only if  $X = X(\mu)$  and  $S = S(\mu)$ , as it should. Otherwise  $\Psi(V) > 0$ . Hence, if  $(X, y, S) \neq (X(\mu), y(\mu), S(\mu))$  then  $(\Delta X, \Delta y, \Delta S)$  is nonzero. By taking a step along the search direction, with the step size  $\alpha$  defined by some line search rules, one constructs a new triple  $(X, y, S)$  according to

$$X_+ = X + \alpha \Delta X, \quad y_+ = y + \alpha \Delta y, \quad S_+ = S + \alpha \Delta S. \quad (27)$$

### 3 A primal-dual algorithm for SDO

We can now describe the algorithm in a more formal way. The generic form of this algorithm is shown in Figure 1. It is clear from this description that closeness of  $(X, y, S)$  to  $(X(\mu), y(\mu), S(\mu))$  is measured by the value of  $\Psi(V)$ , with  $\tau$  as a threshold value: if  $\Psi(V) \leq \tau$  then we start a new *outer iteration* by performing a  $\mu$ -update, otherwise we enter an *inner iteration* by computing the search directions at the current iterates with respect to the current value of  $\mu$  and apply (27) to get new iterates.

The parameters  $\tau, \theta$  and the step size  $\alpha$  should be chosen in such a way that the algorithm is 'optimized' in the sense that the number of iterations required by the algorithm is as small as possible. The choice

---

## Primal-Dual Algorithm for SDO

---

**Input:**

A threshold parameter  $\tau \geq 1$ ;  
 an accuracy parameter  $\varepsilon > 0$ ;  
 a fixed barrier update parameter  $\theta$ ,  $0 < \theta < 1$ ;  
 a strictly feasible  $(X^0, S^0)$  and  $\mu^0 = 1$  such that  $\Psi(X^0, S^0, \mu^0) \leq \tau$ .

**begin**

$X := X^0$ ;  $S := S^0$ ;  $\mu := \mu^0$ ;

**while**  $n\mu \geq \varepsilon$  **do**

**begin**

$\mu := (1 - \theta)\mu$ ;

**while**  $\Psi(X, S, \mu) > \tau$  **do**

**begin**

Solve system (25) and use (23) for  $\Delta X, \Delta y, \Delta S$ ;

Determine a step size  $\alpha$ ;

$X := X + \alpha\Delta X$ ;

$S := S + \alpha\Delta S$ ;

$y := y + \alpha\Delta y$ ;

$V := \frac{1}{\sqrt{\mu}}(D^{-1}XSD)^{\frac{1}{2}}$ ;

**end**

**end**

**end**

---

Figure 1: Algorithm

of the so-called barrier update parameter  $\theta$  plays an important role both in theory and practice of IPMs. Usually, if  $\theta$  is a constant independent of the dimension  $n$  of the problem, for instance  $\theta = \frac{1}{2}$ , then we call the algorithm a *large-update* (or *long-step*) method. If  $\theta$  depends on the dimension of the problem, such as  $\theta = \frac{1}{\sqrt{n}}$ , then the algorithm is named a *small-update* (or *short-step*) method.

The choice of the step size  $\alpha$  ( $0 \leq \alpha \leq 1$ ) is another crucial issue in the analysis of the algorithm. It has to be taken such that the closeness of the iterates to the current  $\mu$ -center improves by a sufficient amount. In the theoretical analysis the step size  $\alpha$  is usually given a value that depends on the closeness of the current iterates to the  $\mu$ -center.

## 4 Properties of the kernel (barrier) function

In this section, we study properties of the kernel function  $\psi(t)$  and the barrier function  $\Psi(V)$ .

### 4.1 Properties of $\psi(t)$

In this section we recall some properties of  $\psi(t)$  from [3] and [4]. The first three derivatives of  $\psi(t)$  are

$$\psi'(t) = 1 - \frac{1}{t^q}, \quad \psi''(t) = \frac{q}{t^{q+1}} > 0, \quad \psi'''(t) = -\frac{q(q+1)}{t^{q+2}}. \quad (28)$$

It is quite straightforward to verify

$$\psi(1) = \psi'(1) = 0, \quad \psi''(t) > 0, \quad t > 0,$$

$$\lim_{t \rightarrow 0} \psi(t) = \lim_{t \rightarrow \infty} \psi(t) = +\infty.$$

Moreover, from (28),  $\psi(t)$  is strictly convex and  $\psi''(t)$  is monotonically decreasing in  $t \in (0, +\infty)$ .

**Lemma 4.1** [Lemma 2.3 in [3]] *Let  $t_1 > 0$ , and  $t_2 > 0$ . Then*

$$\psi(\sqrt{t_1 t_2}) \leq \frac{1}{2}(\psi(t_1) + \psi(t_2)).$$

□

**Lemma 4.2** *One has*

$$\psi'(t)^2 \geq \psi'\left(\frac{1}{t}\right)^2, \quad 0 < t \leq 1. \quad (29)$$

**Proof:** When  $0 < t \leq 1$ , we have

$$\psi'(t)^2 - \psi'\left(\frac{1}{t}\right)^2 = \left(1 - \frac{1}{t^q}\right)^2 - (1 - t^q)^2 = \left(\frac{1}{t^q} + t^q - 2\right) \left(\frac{1}{t^q} - t^q\right) \geq 0,$$

which implies the lemma. □

**Lemma 4.3** [Lemma 2.6 in [4]] *One has*

$$\psi(t) < \frac{1}{2}\psi''(1)(t-1)^2, \quad \text{if } t \geq 1.$$

□

**Lemma 4.4** [Lemma 3.1 in [4]] *Suppose that  $\psi(t_1) = \psi(t_2)$  with  $t_1 \leq 1 \leq t_2$  and  $\beta \geq 1$ . Then  $\psi(\beta t_1) \leq \psi(\beta t_2)$ . Equality holds if and only if  $\beta = 1$  or  $t_1 = t_2 = 1$ . □*

**Lemma 4.5** *Let  $\varrho: [0, \infty) \rightarrow [1, \infty)$  be the inverse function of  $\psi(t)$  for  $t \geq 1$ . If  $q > 1$ , one has*

$$1 + s \leq \varrho(s) \leq s + \frac{q}{q-1}. \quad (30)$$

*If  $q \geq 2$ , one has*

$$\varrho(s) \leq 1 + \sqrt{s^2 + \frac{q}{q-1}s}. \quad (31)$$

**Proof:** When  $q > 1$ , we have

$$t = \varrho(s) \Leftrightarrow \psi(t) = t - 1 + \frac{t^{1-q} - 1}{q-1} = s, \quad t \geq 1.$$

Using that  $t \geq 1$  one easily sees that

$$1 + s \leq \varrho(s) \leq s + \frac{q}{q-1}.$$

Now we consider the case that  $q \geq 2$ . We first establish that  $t\psi(t) \geq (t-1)^2$  for  $t \geq 1$ . This goes as follows. Defining  $f(t) = t\psi(t) - (t-1)^2$  one has  $f(1) = 0$  and  $f'(t) = \psi(t) + t\psi'(t) - 2(t-1)$ . Hence  $f'(1) = 0$  and  $f''(t) = 2\psi'(t) + t\psi''(t) - 2$ . Since  $f''(t) = (q-2)t^{-q} \geq 0$  the claim follows. Hence we have

$$t \leq 1 + \sqrt{t\psi(t)}.$$

Now substituting  $t \leq \psi(t) + \frac{q}{q-1}$  we obtain

$$\varrho(s) \leq 1 + \sqrt{s^2 + \frac{q}{q-1}s}.$$

This completes the lemma.  $\square$

**Remark 4.6** (31) is tighter than (30) at  $s = 0$ . This will help us in getting a sharp iteration bound for small-update methods.

## 4.2 Properties of $\Psi(V)$

The following lemmas are crucial for the analysis of the SDO algorithm to be stated below.

We need a result from [7] that we state without proof.

**Lemma 4.7** [ Lemma 3.3.14 (c) in [7]] *Let  $M, N \in \mathbf{S}^n$  be two nonsingular matrices and  $f(t)$  be given real-valued function such that  $f(e^t)$  is a convex function. One has*

$$\sum_{i=1}^n f(\sigma_i(MN)) \leq \sum_{i=1}^n f(\sigma_i(M)\sigma_i(N)), \quad (32)$$

where  $\sigma_i(M), i = 1, 2, \dots, n$  denote the singular values of  $M$ .  $\square$

We can apply the above lemma to  $\psi(t)$  because it is easily verify that  $\psi(e^t)$  is convex function.

**Lemma 4.8** [ Proposition 5.2.6 in [15]] *Suppose that matrices  $V_1$  and  $V_2$  are symmetric positive definite, then*

$$\Psi([V_1^{\frac{1}{2}}V_2V_1^{\frac{1}{2}}]^{\frac{1}{2}}) \leq \frac{1}{2}(\Psi(V_1) + \Psi(V_2)). \quad (33)$$

**Proof:** For any nonsingular matrix  $V \in \mathbf{S}^n$ ,

$$\sigma_i(V) = (\lambda_i(V^T V))^{\frac{1}{2}} = (\lambda_i(VV^T))^{\frac{1}{2}}, i = 1, 2, \dots, n.$$

From the above equality, we have

$$\sigma_i\left([V_1^{\frac{1}{2}}V_2V_1^{\frac{1}{2}}]^{\frac{1}{2}}\right) = \left(\lambda_i(V_1^{\frac{1}{2}}V_2V_1^{\frac{1}{2}})\right)^{\frac{1}{2}} = \lambda_i\left([V_1^{\frac{1}{2}}V_2V_1^{\frac{1}{2}}]^{\frac{1}{2}}\right), i = 1, 2, \dots, n.$$

Since  $V_1$  and  $V_2$  are symmetric positive definite, one has

$$\sigma_i(V_1) = \lambda_i(V_1), \quad \sigma_i(V_2) = \lambda_i(V_2), \quad i = 1, 2, \dots, n.$$

Using the definition of  $\Psi(V)$ , Lemma 4.7 and Lemma 4.1, one has

$$\begin{aligned} \Psi\left([V_1^{\frac{1}{2}}V_2V_1^{\frac{1}{2}}]^{\frac{1}{2}}\right) &= \sum_{i=1}^n \psi\left(\sigma_i\left([V_1^{\frac{1}{2}}V_2V_1^{\frac{1}{2}}]^{\frac{1}{2}}\right)\right) \leq \sum_{i=1}^n \psi\left(\sigma_i(V_1^{\frac{1}{2}})\sigma_i(V_2^{\frac{1}{2}})\right) \\ &\leq \frac{1}{2} \sum_{i=1}^n \left(\psi\left(\sigma_i^2(V_1^{\frac{1}{2}})\right) + \psi\left(\sigma_i^2(V_2^{\frac{1}{2}})\right)\right) \\ &= \frac{1}{2} \sum_{i=1}^n (\psi(\sigma_i(V_1)) + \psi(\sigma_i(V_2))) = \frac{1}{2}(\Psi(V_1) + \Psi(V_2)). \end{aligned}$$

This completes the lemma.  $\square$

In the analysis of the algorithm we also use the norm-based proximity measure  $\delta(V)$  defined by

$$\delta(V) := \frac{1}{2} \|\psi'(V)\| = \frac{1}{2} \sqrt{\sum_{i=1}^n \psi'(\lambda_i(V))^2} = \frac{1}{2} \|D_X + D_S\|. \quad (34)$$

Obviously, since  $D_X \perp D_S$ , we have

$$\|D_X + D_S\|^2 = \|D_X\|^2 + \|D_S\|^2. \quad (35)$$

Moreover, recall that  $\Psi(V)$  is strictly convex and minimal at  $V = E$  whereas the minimal value is zero. So we have

$$\Psi(V) = 0 \Leftrightarrow \delta(V) = 0 \Leftrightarrow V = E.$$

The following lemma gives a bound of  $\delta(V)$  in terms of  $\Psi(V)$ .

**Lemma 4.9** *One has*

$$\delta(V) \geq \frac{1}{2} \left( 1 - \frac{1}{(1 + \Psi(V))^q} \right), \quad V \succeq 0. \quad (36)$$

**Proof:** The statement in the lemma is obvious if  $V = E$ , since then  $\delta(V) = \Psi(V) = 0$ , otherwise we have  $\delta(V) > 0$  and  $\Psi(V) > 0$ . To deal with the nontrivial case we consider, for  $K > 0$ , the problem

$$z^K = \min \{ \delta(V)^2 = \frac{1}{4} \sum_{i=1}^n \psi'(\lambda_i(V))^2 : \Psi(V) = K \}.$$

The first order optimality condition becomes

$$\frac{1}{2} \psi'(\lambda_i(V)) \psi''(\lambda_i(V)) = \gamma \psi'(\lambda_i(V)), \quad i = 1, 2, \dots, n.$$

where  $\gamma \in R$  is the Lagrange multiplier. Since  $K > 0$  we have  $\lambda_i(V) \neq 1$  for at least one  $i$ . Therefore,  $\gamma \neq 0$ . It follows that either  $\psi(\lambda_i'(V)) = 0$  or  $\psi''(\lambda_i(V)) = 2\gamma$ , for each  $i$ . Since  $\psi''(t)$  is monotonically decreasing, this implies that all  $\lambda_i(V)$  for which  $\psi''(\lambda_i(V)) = 2\gamma$  have the same value. Denoting this value as  $\eta$ , and observing that all other eigenvalues have value 1, we conclude that, after reordering the eigenvalues,  $V$  has the form

$$V = Q^T \text{diag}(\eta, \dots, \eta, 1, \dots, 1) Q.$$

Now  $\Psi(V) = K$  implies  $k\psi(\eta) = K$ . Given  $k$ , this uniquely determines  $\psi(\eta)$ , where we have

$$4\delta(V)^2 = k\psi'(\eta)^2 = k\left(1 - \frac{1}{\eta^q}\right)^2, \quad \psi(\eta) = \frac{K}{k}.$$

Not that the equation  $\psi(\eta) = \frac{K}{k}$  has two solutions, one smaller than 1 and one larger than 1, that can be written as  $\eta = \lambda^1(V)$  and  $\eta = \lambda^2(V)$  with  $0 < \lambda^1(V) < 1$  and  $\lambda^2(V) > 1$ . By Lemma 4.2 we have  $\psi'(\lambda^1(V))^2 \geq \psi'(\frac{1}{\lambda^1(V)})^2$ . Since we are minimizing  $\delta(V)^2$ , we conclude that  $\eta > 1$ . From the definition of  $\psi(t)$  we deduce that  $\psi(\eta) \leq \eta - 1$ , for  $\eta > 1$ . Hence we obtain that  $\frac{K}{k} \leq \eta - 1$ , whence  $\eta \geq 1 + \frac{K}{k}$ . Since  $\psi'(\eta)$  is monotonically increasing, we conclude that

$$4\delta(V)^2 = k\psi'(\eta)^2 = k\left(1 - \frac{1}{\eta^q}\right)^2 \geq k\left(1 - \frac{1}{(1 + \frac{K}{k})^q}\right)^2 \geq \left(1 - \frac{1}{(1 + K)^q}\right)^2.$$

The last inequality is due to the fact that the middle expression is increasing in  $k$ . Thus we

$$\delta(V) \geq \frac{1}{2} \left( 1 - \frac{1}{(1 + K)^q} \right).$$

Substituting  $\Psi(V)$  for  $K$ , this inequality of the lemma follows.  $\square$

Note that during the course of the algorithm the largest values of  $\Psi(V)$  occur just after the update of  $\mu$ . So next we derive an estimate for the effect of a  $\mu$ -update on the value of  $\Psi(V)$ . We start with an important lemma.

**Lemma 4.10** *Let  $\varrho : [0, \infty) \rightarrow [1, \infty)$  be the inverse function of  $\psi(t)$  for  $t \geq 1$ . Then we have for any positive definite matrix  $V$  and any  $\beta \geq 1$ :*

$$\Psi(\beta V) \leq n\psi\left(\beta\varrho\left(\frac{\Psi(V)}{n}\right)\right). \quad (37)$$

**Proof:** The inequality in the theorem is obvious if  $\beta = 1$ , or when  $V = E$ . So we may assume below that  $\beta > 1$  and  $V \neq E$ . We consider the following maximization problem:

$$\max_V \{\Psi(\beta V) : \Psi(V) = K\},$$

where  $K$  is any nonnegative number. The first order optimality conditions for this problem are

$$\beta\psi'(\beta\lambda_i(V)) = \gamma\psi'(\lambda_i(V)), \quad i = 1, \dots, n,$$

where  $\gamma$  denotes the Lagrange multiplier. Since  $\psi'(1) = 0$  and  $\beta\psi'(\beta) > 0$ , we must have  $\lambda_i(V) \neq 1$  for all  $i$ . We even may assume that  $\lambda_i(V) > 1$  for all  $i$ . To see this, let  $k_i$  be such that  $\psi(V_i) = k_i$ . Given  $k_i$ , this equation has two solutions:  $\lambda_i(V)^{(1)} < 1$  and  $\lambda_i(V)^{(2)} > 1$ . As a consequence of Lemma 4.4 we have  $\psi(\beta\lambda_i(V)^{(1)}) \leq \psi(\beta\lambda_i(V)^{(2)})$ . Since we are maximizing  $\Psi(\beta V)$ , we conclude that  $\lambda_i(V) = \lambda_i(V)^{(2)} > 1$ . Thus we have, for all  $i$ ,

$$\beta\left(1 - \frac{1}{(\beta\lambda_i(V))^q}\right) = \gamma\left(1 - \frac{1}{\lambda_i(V)^q}\right), \quad \lambda_i(V) > 1.$$

Using this and  $\beta > 1$  we easily obtain that  $\gamma > \beta > 1$ , and that all coordinates  $\lambda_i(V)$  are equal and given by

$$\lambda_i(V) = \left(\frac{\gamma\beta^{q-1} - 1}{\beta^{q-1}(\gamma - \beta)}\right)^{\frac{1}{q}}, \quad i = 1, 2, \dots, n.$$

Denoting their common values as  $t$  we deduce from  $\Psi(V) = K$  that  $n\psi(t) = K$ . This implies  $t = \varrho(K/n)$ . Hence the maximal value that  $\Psi(\beta V)$  can attain is given by

$$\Psi(\beta tE) = n\psi(\beta t) = n\psi(\beta\varrho(\frac{K}{n})) = n\psi(\beta\varrho(\frac{\Psi(V)}{n})).$$

This proves the lemma. □

**Corollary 4.11** *Let  $0 \leq \theta \leq 1$  and  $V_+ = \frac{V}{\sqrt{1-\theta}}$ . If  $\Psi(V) \leq \tau$ , then*

$$\Psi(V_+) \leq n\psi\left(\frac{\varrho(\tau/n)}{\sqrt{1-\theta}}\right). \quad (38)$$

**Proof:** With  $\beta \geq 1$  and  $\Psi(V) \leq \tau$  the inequality follows from Lemma 4.10.

We use two upper bounds for  $\varrho(s)$  which are formed by (30) and (31) to make two upper bounds for  $\Psi(V)$ . As we will show in the next section, each subsequent inner iteration will give rise to a decrease of value of  $\Psi(V)$ . Hence, we may already conclude that the numbers

$$L_1 := n\psi\left(\frac{\frac{\tau}{n} + \frac{q}{q-1}}{\sqrt{1-\theta}}\right), \quad q > 1, \quad (39)$$

and

$$L_2 := n\psi\left(\frac{1 + \sqrt{\frac{\tau^2}{n^2} + \frac{q}{q-1}\frac{\tau}{n}}}{\sqrt{1-\theta}}\right), \quad q \geq 2, \quad (40)$$

are two upper bounds for  $\Psi(V)$  during the course of the algorithm.

## 5 Derivation of iteration bound

### 5.1 Decrease of the value of $\Psi(V)$ and selection of step size

In each inner iteration we first compute the search direction  $\Delta X$ ,  $\Delta y$ , and  $\Delta S$  from the system (25), also using (23). After a step with size  $\alpha$  the new iterations are  $X_+ = X + \alpha\Delta X$ ,  $y_+ = y + \alpha\Delta y$  and  $S_+ = S + \alpha\Delta S$ , respectively.

Note that by (23), we have

$$X_+ = X + \alpha\Delta X = X + \alpha\sqrt{\mu}DD_XD = \sqrt{\mu}D(V + \alpha D_X)D$$

and

$$S_+ = S + \alpha\Delta S = S + \alpha\sqrt{\mu}D^{-1}D_S D^{-1} = \sqrt{\mu}D^{-1}(V + \alpha D_S)D^{-1}.$$

Thus we obtain

$$V_+ = \frac{1}{\sqrt{\mu}} (D^{-1}X_+S_+D)^{\frac{1}{2}}.$$

We can verify that  $V_+^2$  is unitarily similar to the matrix  $X_+^{\frac{1}{2}}S_+X_+^{\frac{1}{2}}$  and thus to  $(V + \alpha D_X)^{\frac{1}{2}}(V + \alpha D_S)(V + \alpha D_X)^{\frac{1}{2}}$ . This implies that the eigenvalues of  $V_+$  are precisely the same as those of the matrix

$$\bar{V}_+ := \left( (V + \alpha D_X)^{\frac{1}{2}}(V + \alpha D_S)(V + \alpha D_X)^{\frac{1}{2}} \right)^{\frac{1}{2}}. \quad (41)$$

By the definition of  $\Psi(V)$ , we obtain  $\Psi(V_+) = \Psi(\bar{V}_+)$ .

Hence, by Lemma 4.8, we have

$$\Psi(V_+) = \Psi(\bar{V}_+) \leq \frac{1}{2} (\Psi(V + \alpha D_X) + \Psi(V + \alpha D_S)).$$

Defining,

$$f(\alpha) := \Psi(V_+) - \Psi(V) = \Psi(\bar{V}_+) - \Psi(V),$$

we have  $f(\alpha) \leq f_1(\alpha)$ , where

$$f_1(\alpha) := \frac{1}{2} (\Psi(V + \alpha D_X) + \Psi(V + \alpha D_S)) - \Psi(V).$$

Obviously,

$$f(0) = f_1(0) = 0.$$

By using (11), (12) and (13), we get

$$f_1'(\alpha) = \frac{1}{2} \text{Tr}(\psi'(V + \alpha D_X)D_X + \psi'(V + \alpha D_S)D_S), \quad (42)$$

and

$$f_1''(\alpha) = \frac{1}{2} \text{Tr}(\psi''(V + \alpha D_X)D_X^2 + \psi''(V + \alpha D_S)D_S^2). \quad (43)$$

Using the third inequality of system (25). Obviously,

$$f_1(0) = \frac{1}{2} \text{Tr}(\psi'(V)(D_X + D_S)) = \frac{1}{2} \text{Tr}(-\psi'(V)^2) = -2\delta(V)^2. \quad (44)$$

The following lemma is a slight modification of the Weyl Theorem (see [18]).

**Lemma 5.1** *Let  $A, A + B \in \mathbf{S}_+^n$ , one has*

$$\lambda_i(A + B) \geq \lambda_n(A) - |\lambda_1(B)|, \quad i = 1, 2, \dots, n. \quad (45)$$

**Proof:** By the Rayleigh-Ritz theorem (see [7]), for any  $i = 1, 2, \dots, n$ , there exists  $X_0 \in \mathbf{R}^n$ , such that

$$\begin{aligned} \lambda_i(A+B) &\geq \lambda_n(A+B) = \frac{X_0^T(A+B)X_0}{X_0^T X_0} = \frac{X_0^T A X_0}{X_0^T X_0} + \frac{X_0^T B X_0}{X_0^T X_0} \\ &\geq \frac{X_0^T A X_0}{X_0^T X_0} - \left| \frac{X_0^T B X_0}{X_0^T X_0} \right| \geq \min_{X \neq 0} \frac{X^T A X}{X^T X} - \max_{X \neq 0} \left| \frac{X^T B X}{X^T X} \right| = \lambda_n(A) - |\lambda_1(B)|. \end{aligned}$$

This completes the lemma.  $\square$

Below we use the following notation:

$$\delta := \delta(V).$$

One of the main results in this section is below.

**Lemma 5.2** *One has*

$$f_1''(\alpha) \leq 2\delta^2 \psi''(\lambda_n(V) - 2\alpha\delta). \quad (46)$$

**Proof:** By using (34) and (35), we have  $\|D_X + D_S\|^2 = \|D_X\|^2 + \|D_S\|^2 = 4\delta^2$ . It implies that  $|\lambda_1(D_X)| \leq 2\delta$  and  $|\lambda_1(D_S)| \leq 2\delta$ . Using Lemma 5.1 and  $V + \alpha D_X \succeq 0$ , we have

$$\lambda_i(V + \alpha D_X) \geq \lambda_n(V) - \alpha |\lambda_1(D_X)| \geq \lambda_n(V) - 2\alpha\delta, \quad i = 1, 2, \dots, n.$$

Since  $\psi''(t)$  is monotonically decreasing in  $t \in (0, \infty)$ , we obtain

$$\psi''(\lambda_i(V + \alpha D_X)) \leq \psi''(\lambda_n(V) - 2\alpha\delta),$$

hence, we have

$$\psi''(V + \alpha D_X) \preceq \psi''(\lambda_n(V) - 2\alpha\delta)E.$$

Since  $D_X^2 \in S_+^n$ , by using (9) and (8), we obtain

$$\text{Tr}(\psi''(V + \alpha D_X) D_X^2) \leq \text{Tr}(\psi''((\lambda_n(V) - 2\alpha\delta)E) D_X^2) \leq \psi''(\lambda_n(V) - 2\alpha\delta) \sum_{i=1}^n \lambda_i(D_X^2).$$

Similarly,

$$\text{Tr}(\psi''(V + \alpha D_S) D_S^2) \leq \text{Tr}(\psi''((\lambda_n(V) - 2\alpha\delta)E) D_S^2) \leq \psi''(\lambda_n(V) - 2\alpha\delta) \sum_{i=1}^n \lambda_i(D_S^2).$$

Using (43), from the above two inequalities and  $\|D_X\|^2 + \|D_S\|^2 = 4\delta^2$ , we have

$$f_1''(\alpha) \leq \frac{1}{2} \psi''(\lambda_n(V) - 2\alpha\delta) \sum_{i=1}^n (\lambda_i(D_X^2) + \lambda_i(D_S^2)) = 2\delta^2 \psi''(\lambda_n(V) - 2\alpha\delta).$$

This completes the proof.  $\square$

Next we will choose a suitable step size for the algorithm. This should be chosen such that  $X_+$  and  $S_+$  are feasible and such that  $\Psi(V_+) - \Psi(V)$  decreases sufficiently.

The following strategy for selecting the step size is almost a "word-by-word" extension of the LO case in [3]. Therefore, for the proof of the following lemmas we refer to [3].

**Lemma 5.3** [Lemma 3.2 in [3]] *If the step size  $\alpha$  satisfies*

$$-\psi'(\lambda_n(V) - 2\alpha\delta) + \psi'(\lambda_n(V)) \leq 2\delta,$$

*then*

$$f_1'(\alpha) \leq 0. \quad (47)$$

□

Let  $\rho : [0, \infty) \rightarrow (0, 1]$  denote the inverse function of the restriction of  $-\frac{1}{2}\psi'(t)$  on the interval  $(0, 1]$ , one may easily verify that

$$\rho(y) = \frac{1}{(2y+1)^{\frac{1}{q}}}, \quad y \in [0, \infty). \quad (48)$$

**Lemma 5.4** [Lemma 3.3 in [3]] *The largest possible value of the step size of  $\alpha$  satisfying (48) is given by*

$$\bar{\alpha} := \frac{1}{2\delta}(\rho(\delta) - \rho(2\delta)). \quad (49)$$

□

By using (49), Lemma 5.3 and the well-known Bernoulli inequality  $(1+x)^\alpha \leq 1+\alpha x$ ,  $x \geq -1$  and  $0 \leq \alpha \leq 1$ , we get

$$\bar{\alpha} \geq \tilde{\alpha} = \frac{1}{q(2\delta+1)^{\frac{1}{q}}(4\delta+1)}. \quad (50)$$

We define our default step size as  $\tilde{\alpha}$ .

**Lemma 5.5 (Lemma 3.12 in [14])** *Let  $h(t)$  be a twice differentiable convex function with  $h(0) = 0$ ,  $h'(0) < 0$  and let  $h(t)$  attain its (global) minimum at  $t^* > 0$ . If  $h''(t)$  is increasing for  $t \in [0, t^*]$  then*

$$h(t) \leq \frac{th'(0)}{2}, \quad 0 \leq t \leq t^*.$$

□

Via the above lemma, we have the following lemma.

**Lemma 5.6** [Lemma 3.6 in [3]] *If the step size  $\alpha$  is such that  $\alpha \leq \bar{\alpha}$ , then*

$$f(\alpha) \leq -\alpha\delta^2. \quad (51)$$

□

By Lemma 5.6, the default size  $\tilde{\alpha}$  satisfies  $\tilde{\alpha} \leq \alpha$ , we get the following upper bound for  $f(\tilde{\alpha})$ :

$$f(\tilde{\alpha}) \leq -\frac{\delta^2}{q(2\delta+1)^{\frac{1}{q}}(4\delta+1)}.$$

By Lemma 4.9, assuming  $\Psi(V) \geq \tau \geq 1$ , we obtain

$$\delta = \delta(V) \geq \frac{1}{2} \left( 1 - \frac{1}{(1+\Psi(V))^q} \right) \geq \frac{1}{2} \left( 1 - \frac{1}{(1+\tau)^q} \right) \geq \frac{1}{2} \left( 1 - \frac{1}{1+q\tau} \right) \geq \frac{1}{4}.$$

Since the decrease depends monotonically on  $\delta$ , substitution yields

$$f(\tilde{\alpha}) \leq -\frac{\frac{1}{16}}{2q(1.5)^{\frac{1}{q}}} \leq -\frac{1}{48q}.$$

## 5.2 Iteration bounds

We need to count how many inner iterations are required to return to the situation where  $\Psi(V) \leq \tau$ . We denote the value of  $\Psi(V)$  after the  $\mu$ -update as  $\Psi_0$ , the subsequent values in the same outer iteration are denoted as  $\Psi_k$ ,  $k = 1, 2, \dots, K$ , where  $K$  denotes the total number of inner iterations in the outer iteration. Using (39), we have

$$\Psi_0 \leq n\psi\left(\frac{\frac{\tau}{n} + \frac{q}{q-1}}{\sqrt{1-\theta}}\right).$$

Since  $\psi(t) \leq t - 1$  when  $t \geq 1$ , and  $1 - \sqrt{1-\theta} = \frac{\theta}{1+\sqrt{1-\theta}} \leq \theta$ , we have

$$\Psi_0 \leq \frac{\theta n + \tau + \frac{n}{q-1}}{\sqrt{1-\theta}}. \quad (52)$$

According to decrease of  $f(\tilde{\alpha})$ , we get

$$\Psi_{k+1} \leq \Psi_k - \frac{1}{48q}, \quad k = 0, 1, \dots, K-1. \quad (53)$$

**Lemma 5.7** *One has*

$$K \leq 48q \frac{\theta n + \tau + \frac{n}{q-1}}{\sqrt{1-\theta}}. \quad (54)$$

**Proof:** Using (52) and (53), the proof is trivial.  $\square$

The number of outer iterations is bounded above by  $\frac{1}{\theta} \log \frac{n}{\epsilon}$ , (seeing [16] II.17, page 116). By multiplying the number of outer iterations and the number of inner iterations we get an upper bound for the total number of iterations, namely,

$$48q \frac{\theta n + \tau + \frac{n}{q-1}}{\theta \sqrt{1-\theta}} \log \frac{n}{\epsilon}.$$

In large-update methods one takes for  $\theta$  a constant (independent on  $n$ ), namely  $\theta = \Theta(1)$ , and  $\tau = O(n)$ . The iteration bound then becomes

$$O(qn) \log \frac{n}{\epsilon}. \quad (55)$$

Obviously, the bound suggest to take  $q$  as small as possible, i.e.,  $q = 2$ . Note that the bound is exactly the same as the bound obtained in [3] for LO, with the same kernel function, namely  $O(n \log \frac{n}{\epsilon})$ .

## 5.3 Complexity of small-update methods

It may be noted that in the above case the worse iteration bound is not as good as it can be for small-update methods due to the fact that the used upper bound for  $\varrho(s)$  is not tight at  $s = 0$ : it should be equal to  $\varrho(0) = 1$  when  $s = 0$ . Using Lemma 4.5, we will see below that an upper bound that is tight at  $s = 0$  will lead to the correct iteration bound.

We recall from previous section that the number  $K$  of inner iterations is bound above by

$$K \leq 48q\Psi_0.$$

To estimate  $\Psi_0$ , we use (40) and Lemma 4.3, with  $\psi''(1) = q$ . We then obtain

$$\Psi_0 \leq n\psi\left(\frac{1 + \sqrt{\frac{\tau^2}{n^2} + \frac{q}{q-1} \frac{\tau}{n}}}{\sqrt{1-\theta}}\right) \leq \frac{qn}{2} \left(\frac{1 + \sqrt{\frac{\tau^2}{n^2} + \frac{q}{q-1} \frac{\tau}{n}}}{\sqrt{1-\theta}} - 1\right)^2.$$

Using  $1 - \sqrt{1 - \theta} = \frac{\theta}{1 + \sqrt{1 - \theta}} \leq \theta$  again and  $\frac{q}{q-1} \leq 2$  this can be simplified to

$$\Psi_0 \leq \frac{qn}{2} \left( \frac{\theta + \sqrt{\frac{\tau^2}{n^2} + \frac{2\tau}{n}}}{\sqrt{1 - \theta}} \right)^2 = \frac{q \left( \theta\sqrt{n} + \sqrt{\frac{\tau^2}{n} + 2\tau} \right)^2}{2(1 - \theta)}.$$

We conclude that the total number of iterations is bounded above by

$$\frac{K}{\theta} \log \frac{n}{\epsilon} \leq 24 \frac{q^2 \left( \theta\sqrt{n} + \sqrt{\frac{\tau^2}{n} + 2\tau} \right)^2}{\theta(1 - \theta)} \log \frac{n}{\epsilon}.$$

For small-update methods, namely,  $\theta = \Theta(\frac{1}{\sqrt{n}})$  and  $\tau = O(1)$ , it shows clearly that the iteration bound is  $O(q^2\sqrt{n} \log \frac{n}{\epsilon})$ .

## 6 Conclusions and remarks

We have extended the approach to primal-dual interior-point algorithms for LO, as developed in [3] to SDO. We derive the same complexity bounds for large- and small- update methods for SDO as for the LO case.

We developed some new analysis tools that can deal with non-self-regular kernel functions ( for self-regular function, see [14, 15]). Moreover, our analysis is simple and straightforward.

Some interesting topics remain for further research. Firstly, the search direction used in this paper is based on the NT-symmetrization scheme and it is natural to ask if other symmetrization schemes can be used. Secondly, numerical experiments are necessary to compare the behavior of the algorithm of this paper with existing methods. Thirdly, it is an interesting question whether we can design similar algorithms by using the kernel function  $\Upsilon_{p,q}(t)$  with  $0 \leq p \leq 1$ , and  $q > 1$ . Note that if  $p \geq 1$  then  $\Upsilon_{p,q}(t)$  is self-regular and the case has been considered in [14, 15].

## References

- [1] Alizadeh, F., Haeberly, J.A., and Overton. M.: A new primal-dual interior-point method for semidefinite programming. In Lewis, J.G., Ed., *Proceedings of the fifth SIAM Conference on Applied Linear Algebra*, SIAM, 1994, 113-117.
- [2] Andersen, E.D., Gondzio, J., Mészáros, Cs. and Xu. X.: Implementation of interior point methods for large scale linear programming. In Terlaky, T. Ed., *Interior Point Methods of Mathematical Programming*, Kluwer Academic Publishers, The Netherlands, 1996, 189-252.
- [3] Bai, Y.Q., Roos. C.: A primal-dual interior-point method based on a new kernel function with linear growth rate. *Proceedings of Industrial Optimization Symposium and Optimization Day. Australia, November 2002.*
- [4] Bai, Y. Q., Ghami, M. El, and Roos. C.: A comparative study of kernel functions for primal-dual interior-point algorithms in linear optimization. *Accepted by SIAM Journal on Optimization*, 2004.
- [5] Ben-Tal, A, Nemirovski. A.: *Lectures on Modern Convex Optimization: Analysis, Algorithm, and Engineering Applications*. MPS-SIAM Series on Optimization, Vol. 02, SIAM, Philadelphia, PA, 2001.
- [6] Helmborg, C., Rendl, F., Vanderbei, R. J. and Wolkowicz, H.: An interior-point method for semidefinite programming. *SIAM Journal on Optimization*, 6, 1996, 342-361.
- [7] Horn, R.A. and Charles R. Johnson: *Topics in Matrix Analysis*, Cambridge University Press, 1991.
- [8] Kojima, M. Mizuno, S. and Yoshise, A.: Interior-point methods for the monotone semidefinite linear complementarity problem in symmetric matrices. *SIAM Journal on Optimization*, 7, 1997, 342-361.
- [9] Lütkepohl, H.: *Handbook of matrices*, John Wiley & Sons, 1996.

- [10] Megiddo, N.: Pathways to the optimal set in linear programming. In Megiddo, N. Ed., *Progress in Mathematical Programming: Interior Point and Related Methods*, Springer Verlag, New York, 1989, 131–158. Identical version in : *Proceedings of the 6th Mathematical Programming Symposium of Japan, Nagoya, Japan*, 1986, 1–35.
- [11] Monteiro, R.D.C.: Primal-dual path-following algorithms for semidefinite programming. *SIAM Journal on Optimization*, 7, 1997, 663-678.
- [12] Nesterov, Yu. E. and Todd, M. J.: Self-scaled barriers and interior-point methods for convex programming, *Mathematics of Operations Research*, 22(1), 1997, 1-42.
- [13] Nesterov, Yu. E. and Todd, M. J.: Primal-dual interior-point methods for self-scaled cones. *SIAM Journal on Optimization*, 8(2),(electronic), 1998, pages 324–364.
- [14] Peng, P. Roos, C. and Terlaky, T.: Self-regular functions and new search directions for linear and semidefinite optimization. *Mathematical Programming*, 93, 2002, 129–171.
- [15] Peng, J. Roos, C. and Terlaky. T.: *Self-Regularity: A New Paradigm for Primal-Dual Interior-Point Algorithms*. Princeton University Press, 2002.
- [16] Roos, C. Terlaky, T. and Vial, J.-Ph.: *Theory and Algorithms for Linear Optimization. An Interior-Point Approach*. John Wiley & Sons, Chichester, UK, 1997.
- [17] Sonnevend. G.: An “analytic center” for polyhedrons and new classes of global algorithms for linear (smooth, convex) programming. In Prékopa, A. Szelezsán, J. and Strazicky, B. Eds. *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*, Springer Verlag, Berlin, West–Germany, 1986, 866–876.
- [18] Wolkowicz, H. Saigal, R. and Vandenberghe, L.: *Handbook of Semidefinite Programming, Theory, Algorithms, and Applications*, Kluwer Academic Publishers, 2000.