SELF-CONCORDANT EXPONENTIAL KERNEL FUNCTION BASED INTERIOR-POINT ALGORITHM FOR SEMIDEFINITE OPTIMIZATION*

JING ZHANG, YANQIN BAI[†] AND PENGFEI MA

Abstract: Semidefinite optimization (SDO) problem is an extension of linear optimization (LO) problem by replacing the vector of variables with a symmetric matrix and the nonnegativity constraints with a positive semidefinite constraint. Most of interior-point methods (IPMs) with polynomial-time complexity can be generalized to the case of semidefinite optimization. In this paper, we present a primal-dual interior-point algorithm for semidefinite optimization problems based on a new self-concordant (SC) exponential kernel function. Combining both properties of self-concordance and kernel function properties for this function, we design and analyze the algorithm and derive the complexity bound for large-update methods. The obtained complexity bounds are analogous to the result in [6] for LO. Finally, we implement the method with different kernel functions and use it to solve the numerical examples. We also provide a comparison of the performances of these different versions of the algorithm when applied to the numerical examples.

Key words: *semidefinite optimization, interior-point methods, self-concordant function*

Mathematics Subject Classification: *90C05 90C51*

---

1 Introduction

Semidefinite optimization (SDO) problem is a convex optimization problem with the feasible region being the intersection of an affine set and the cone of positive semidefinite matrices. SDO arises from many scientific and engineering fields. The handbook [19] listed SDO applications in combinatorial optimization, systems and control theory, structural design, matrix completion problems, and problems in statistics. SDO has efficient algorithms with polynomial complexity bounds. Interior-point methods (IPMs) are powerful and efficient algorithms to solve SDO. Most IPMs for SDO can be viewed as natural extensions of IPMs for linear optimization (LO) problem and have the similar polynomial complexity bounds. The landmark work of generalizations of IPMs from LO to SDO was due to Nesterov and Nemirovskii [13]. They proposed IPMs for solving convex programs in polynomial time based on self-concordant (SC) barrier functions. As special cases of convex programs, LO and SDO have explicit and easily computable SC barrier functions and hence can be solved in polynomial time. Independently, Alizadeh [1] extended in a direct way the IPMs for LO

to SDO. Vandenberghe and Boyd [18] proved that IPMs seem to be the best algorithms for solving SDO from both theoretical and practical viewpoints.

Peng et al. [16] proposed a class of primal-dual IPMs based on self-regular functions for solving LO. The authors extended it to the case of SDO and obtained the best known complexity bounds. Subsequently, Bai et al. [4] introduced a family of kernel functions called eligible kernel functions and presented primal-dual IPMs based on the proposed kernel functions. Up to now, there are plentiful results in this field because IPMs provide a powerful approach for solving SDO. A comprehensive list of publications on SDO can be found on the SDO homepage maintained by Alizadeh [2]. For an overview of these and related results, we refer to [8], [10], [11], [19], [20].

The concept of SC exponential kernel function was first introduced by Bai et al. in [6]. They proved it is SC function and proposed IPMs for LO based on it. They derived the complexity bounds for large-update methods, which coincide with that of logarithmic function. Motivated by the fact that SC exponential kernel function performs well in the framework of IPMs for LO, we extended the kernel function-based IPM from LO to SDO. Specifically, Boyd and Vandenberghe [7] applied Newton's methods to minimize a self-concordant function. They obtained complexity bounds without the assumption of strong convexity and the Lipschitz condition on the Hessian of the objective function. In this paper, we use Newton's method to minimize a self-concordant function in estimating the decrease of barrier function in the process of inner iteration.

The paper is organized as follows. In section 2, we briefly recall the preliminaries and investigate properties of SC exponential kernel function and some other existing kernel functions. In section 3, we present a primal-dual IPM based on SC exponential kernel function. We analyze the algorithm and obtain the complexity bounds for large-update methods. In Section 4, we show the results of numerical examples. Finally, Section 5 contains some concluding remarks.

We use the following notational conventions throughout the paper. The superscript T denotes transpose. $\mathbf{R^n}$, $\mathbf{R^n_+}$ and $\mathbf{R^n_{++}}$ denote the set of vectors with n components, the set of nonnegative vector and the set of positive vectors, respectively. $\mathbf{R^{m \times n}}$ 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. The symbol E denotes $n \times n$ identity matrix. Use the notation int$\mathcal{C}$ and ri($\mathcal{C}$) denote the interior and the relative interior of a convex set $\mathcal{C}$. We use the classical Löwner partial order $\succeq$ for symmetric matrices. So $A \succeq B$ ($A \succ B$) means that $A - B$ is positive semidefinite (positive definite). The sign $\sim$ denotes similarity of two matrices. The matrix inner product is defined by $A \bullet B = \mathbf{Tr}(A^T B)$. For any $Q \in \mathbf{S^n_{++}}$, the expression $Q^{\frac{1}{2}}$ denotes the symmetric square root of Q. For any symmetric matrix G, $\lambda_{\min}(G)(\lambda_{\max}(G))$ denotes the minimal (maximal) eigenvalue of G. When λ is vector we denote the diagonal matrix diag(λ) with entries λ_i by Λ. For any $V \in \mathbf{S^n_{++}}$, we denote by $\lambda(V)$ the vector of eigenvalues of V arranged in non-increasing order, i.e.,

$$\lambda_{\max}(V) = \lambda_1(V) \geq \lambda_2(V) \geq \cdots \geq \lambda_n(V) = \lambda_{\min}(V).$$

For any V, we denote by

$$\eta_{\max}(V) = \eta_1(V) \geq \eta_2(V) \geq \cdots \geq \eta_n(V) = \eta_{\min}(V)$$

the singular values of V; If V is symmetric, one has $\eta_i(V) = |\lambda_i(V)|$, $i = 1, 2, ..., n$. The Frobenius matrix norm is given by

$$\|U\|^2 := \sum_{i=1}^{m} \sum_{j=1}^{n} U_{ij} = \mathbf{Tr}(U^T U).$$

We use $\mathcal{P}$ and $\mathcal{D}$ to denote the feasible sets of primal and dual problems, respectively. The notation $\mathcal{F}^*$ denotes the set of optimal solutions with zero duality gap, i. e.,

$$\mathcal{F}^* := \{(X, y, S) \in \mathcal{P} \times \mathcal{D} : \mathbf{Tr}(XS) = 0\}.$$

Furthermore, if $v \in \mathbf{R}_+^{\mathbf{n}}$ and $f : \mathbf{R} \to \mathbf{R}$, then $f(v)$ denotes the vector in $\mathbf{R}_+^{\mathbf{n}}$ whose i-th component is $f(v_i)$, with $1 \le i \le n$. We write $f(x) = O(g(x))$ if $f(x) \le cg(x)$ for some positive constant c and $f(x) = \Theta(g(x))$ if $c_1 g(x) \le f(x) \le c_2 g(x)$ for positive constants c_1 and c_2.

2 Preliminaries

In this section, we briefly recall some preliminaries used in the subsequent analysis of the method.

2.1 The central path for SDO

Consider the following standard SDO problem:

$$(SDO) \quad \begin{array}{ll} \min & C \bullet X, \\ \text{s.t.} & A_i \bullet X = b_i, \quad i = 1, 2, ..., m, \\ & X \succeq 0, \end{array} \tag{2.1}$$

and its dual problem

$$(SDD) \quad \begin{array}{ll} \max & b^T y, \\ \text{s.t.} & \sum_{i=1}^{m} y_i A_i + S = C, \\ & S \succeq 0, \end{array} \tag{2.2}$$

where C and A_i are symmetric $n \times n$ matrices, $b, y \in \mathbf{R^m}$, and $X \succeq 0$ means that X is symmetric positive semidefinite. Without loss of generality, the matrices A_i are further assumed to be linearly independent.

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

$$\begin{array}{c} A_i \bullet X = b_i, \quad i = 1, 2, ..., m, \quad X \succeq 0, \\ \sum_{i=1}^{m} y_i A_i + S = C, \quad S \succeq 0, \\ XS = 0. \end{array} \tag{2.3}$$

The basic idea of primal-dual IPMs is to replace the above complementarity condition $XS = 0$ by the parameterized equation $XS = \mu E$, $\mu > 0$. So the above system is transformed to

$$\begin{array}{c} A_i \bullet X = b_i, \quad i = 1, 2, ..., m, \quad X \succeq 0, \\ \sum_{i=1}^{m} y_i A_i + S = C, \quad S \succeq 0, \\ XS = \mu E. \end{array} \tag{2.4}$$

This system has a unique solution for each $\mu > 0$ under the assumption of IPC, see [10]. Let $(x(\mu), y(\mu), s(\mu))$ denote the solution of (2.4) for each $\mu > 0$, where $x(\mu)$ is called the μ-center of (SDO) and $(y(\mu), s(\mu))$ the μ-center of (SDD). The set of μ-centers (with μ running through all positive real numbers) is called the central path of (SDO) and (SDD). If $\mu \to 0$ then the limit of the central path exists and since the limit points satisfy the complementarity condition, it yields optimal solutions for (SDO) and (SDD). Traditionally Newton's method is used to find the search direction ΔX, Δy and ΔS by solving a Newton's system below.

$$A_i \bullet \Delta X = 0, \quad i = 1, 2, ..., m,$$
$$\sum_{i=1}^{m} \Delta y_i A_i + \Delta S = 0, \tag{2.5}$$
$$X\Delta S + \Delta X S = \mu E - XS.$$

The above system has a unique solution $(\Delta X, \Delta y, \Delta S)$ (which can be found in [19]). Note that ΔS is symmetric, due to the second equation in (2.5). However, a crucial point is that ΔX may be not symmetric. Many researchers have proposed various ways of 'symmetrizing' the third equation in the Newton system so that the new system has a unique symmetric solution. All these proposals can be described by using a symmetric nonsingular scaling matrix P and by replacing (2.5) by the following system

$$A_i \bullet \Delta X = 0, \quad i = 1, 2, ..., m,$$
$$\sum_{i=1}^{m} \Delta y_i A_i + \Delta S = 0, \tag{2.6}$$
$$\Delta X + P\Delta S P^T = \mu S^{-1} - X.$$

Now ΔX is automatically a symmetric matrix. However, obtaining valid search directions is much more difficult in the SDO than in the LO case. Some popular choices for the matrix P are listed in Table 1. In the sequel, we describe how the usual search directions are obtained for primal-dual methods for solving SDO problems.

Table 1: Choices for the scaling matrix P.

P	References
E	Alizadeh, Haeberley and Overton [2];
X^{-1}	Monteiro [12], Kojima et al. [11];
S	Monteiro [12], Kojima et al. [11];
$X^{\frac{1}{2}}(X^{\frac{1}{2}}SX^{\frac{1}{2}})^{-\frac{1}{2}}X^{\frac{1}{2}}$	Nesterov and Todd [14]

In the Nesterov-Todd (NT)-scheme, we use

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

Let $D = P^{\frac{1}{2}}$. Then matrix D can be used to scale X and S to the same matrix V, defined by

$$V := \frac{1}{\sqrt{\mu}}D^{-1}XD^{-1} = \frac{1}{\sqrt{\mu}}DSD. \tag{2.7}$$

Therefore we have

$$V^2 := \frac{1}{\mu} D^{-1} X S D. \tag{2.8}$$

Obviously the matrices D and V are symmetric and positive definite. Let

$$\bar{A}_i := \frac{1}{\sqrt{\mu}} D A_i D, \quad i = 1, 2, ..., m,$$

$$D_X := \frac{1}{\sqrt{\mu}} D^{-1} \Delta X D^{-1}, \tag{2.9}$$

$$D_S := \frac{1}{\sqrt{\mu}} D \Delta S D.$$

Then it follows from (2.6)

$$\bar{A}_i \bullet D_X = 0, \quad i = 1, 2, ..., m,$$

$$\sum_{i=1}^m \Delta y_i \bar{A}_i + D_S = 0, \tag{2.10}$$

$$D_X + D_S = V^{-1} - V.$$

Note that the right-hand side of the third equation in (2.10) is the negative gradient of the logarithmic barrier function defined by kernel function

$$\psi_{\log}(t) := \frac{t^2 - 1}{2} - \log t, \quad t > 0.$$

The introduction of kernel functions and associated barrier functions different than logarithmic kernel and barrier function give rise to the flexibility in calculating different search directions which may lead to the improved complexity of IPMs. The definition and the properties of the general kernel function and the associated barrier function will be discussed in the next subsection. Using the concept of general kernel function and the barrier function, Bai [5] obtained the following system

$$\bar{A}_i \bullet D_X = 0, \quad i = 1, 2, ..., m,$$

$$\sum_{i=1}^m \Delta y_i \bar{A}_i + D_S = 0,$$

$$D_X + D_S = -\psi'(V).$$

Since D_X and D_S are orthogonal, i.e.,

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

So 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 $XS = \mu E$, which is equivalent to $X = X(\mu)$ and $S = S(\mu)$. So if the iterate is on the central path, then the search directions at that point are zero. Otherwise we have $\Psi(V) > 0$, which means $(\Delta X, \Delta y, \Delta S)$ is nonzero. By taking a step along the search

direction $(\Delta X, \Delta y, \Delta S)$, with the step size α which will be calculated in later analysis, the new iteration is obtained by

$$X_+ = X + \alpha \Delta X, \quad y_+ = y + \alpha \Delta y, \quad S_+ = S + \alpha \Delta S. \tag{2.11}$$

Analogous to the case of LO, the kernel-function-based approach to SDO is obtained by modifying Nesterov-Todd direction. Thus we use the following system to define the (scaled) search directions.

$$\bar{A}_i \bullet D_X = 0, \quad i = 1, 2, ..., m,$$

$$\sum_{i=1}^{m} \Delta y_i \bar{A}_i + D_S = 0, \tag{2.12}$$

$$D_X + D_S = D_V,$$

where $D_V = -V^{-1} \psi''(V)^{-\frac{1}{2}} \psi'(V) V$. We obtain D_X and D_S from (2.12) and then ΔX and ΔS can be calculated from (2.9). Due to the orthogonality of ΔX and ΔS, it is trivial to see that $D_X \perp D_S$, and so

$$\mathbf{Tr}(D_X D_S) = \mathbf{Tr}(D_S D_X) = 0, \tag{2.13}$$

which is an important fact in the design of the algorithm.

2.2 Special matrix functions

Definition 2.1 (Kernel function defined in [4]). A univariate function $\psi(t) : \mathbf{R}_{++} \to \mathbf{R}_+$ is called a kernel function if $\psi(t)$ is twice differentiable and is satisfied the following conditions.

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

2. $\psi(1) = \psi'(1) = 0.$

3. $\psi''(t) > 0.$

Having a kernel function $\psi(t)$, we can define an n-dimensional barrier function $\Psi(v)$ as follows.

Definition 2.2. Define $\Psi(v) : \mathbf{R^n} \to \mathbf{R}$ by

$$\Psi(v) = \sum_{i=1}^{n} \psi(v_i), \quad v_i > 0. \tag{2.14}$$

Let us focus on one-dimension case since kernel function is defined on $\mathbf{R}$. As a special case, the definition of SC function (distinguished to SC barrier function) for a function on $\mathbf{R}$ is given in [7].

Definition 2.3 (SC function on $\mathbf{R}$ defined in [7]). A univariate convex function $\psi(t) : \mathbf{R} \to \mathbf{R}$ is called SC if $\psi(t)$ is three times differentiable and there exists a positive k, the following condition is satisfied

$$|\psi'''(t)| \le 2k\psi''(t)^{\frac{3}{2}}. \tag{2.15}$$

Definition 2.4 (SC barrier function on **R**). A univariate convex function $\psi(t) : \mathbf{R} \to \mathbf{R}$ is called SC barrier if it is SC and and there exists a positive ρ, the following condition is satisfied

$$\psi'(t)^2 \le \rho\psi''(t). \tag{2.16}$$

Note that the definition of SC function differs from the definition of SC barrier function.

Based on the above definitions, we investigate whether the kernel functions proposed in [3], [5], [6] are SC functions. It suffices to verify that $\frac{\psi'''(t)^2}{\psi''(t)^3}$ has an upper bound, which is equivalent to (2.15). However, these kernel functions are not SC barrier functions because their first and second derivatives do not satisfy condition (2.16). These five kernel functions and their first to three derivatives are listed in Table 2. We also show the upper bound of $\frac{\psi'''(t)^2}{\psi''(t)^3}$ of these functions and the limits of $\frac{\psi'(t)^2}{\psi''(t)}$ when t approaches to the boundary. We illustrate the SC property of these functions in Figure 1.

Table 2: Kernel functions and derivatives and SC properties

i	1	2	3
$\psi_i(t),\ t>0$	$\frac{t^2-1}{2} - \log t$	$e^t + e^{\frac{1}{t}} - 2e$	$\frac{t^2-1}{2} + \frac{e^{\frac{1}{t}}-e}{e}$
$\psi_i'(t),\ t>0$	$t - \frac{1}{t}$	$e^t - \frac{1}{t^2}e^{\frac{1}{t}}$	$t - \frac{e^{\frac{1}{t}-1}}{t^2}$
$\psi_i''(t),\ t>0$	$1 + \frac{1}{t^2}$	$e^t + \frac{(2t+1)}{t^4}e^{\frac{1}{t}}$	$1 + \frac{1+2t}{t^4}e^{\frac{1}{t}-1}$
$\psi_i'''(t),\ t>0$	$-\frac{2}{t^3}$	$e^t - \frac{6t^2+6t+1}{t^6}e^{\frac{1}{t}}$	$-\frac{1+6t+6t^2}{t^6}e^{\frac{1}{t}-1}$
$\frac{\psi_i'''(t)^2}{\psi_i''(t)^3},\ t>0$	≤ 3.9988	≤ 1.1014	≤ 2.9551
$\frac{\psi_i'(t)^2}{\psi_i''(t)},\quad t\to 0$	1	$\to +\infty$	$\to +\infty$
$\frac{\psi_i'(t)^2}{\psi_i''(t)},\quad t\to +\infty$	$\to +\infty$	$\to +\infty$	$\to +\infty$
i	4	5	
$\psi_i(t),\ t>0$	$\frac{t^2-1}{2} + \frac{(e-1)^2}{e}\frac{1}{e^t-1} - \frac{e-1}{e}$	$\frac{t^2-1}{2} + (\frac{1}{t}-1)e^{\frac{1}{t}-1}$	
$\psi_i'(t),\ t>0$	$t - \frac{(e-1)^2}{e}\frac{e^t}{(e^t-1)^2}$	$t - \frac{1}{t^3}e^{\frac{1}{t}-1}$	
$\psi_i''(t),\ t>0$	$1 + \frac{(e-1)^2}{e}\frac{e^t(1+e^t)}{(e^t-1)^3}$	$1 + (\frac{1}{t^5}+\frac{3}{t^4})e^{\frac{1}{t}-1}$	
$\psi_i'''(t),\ t>0$	$-\frac{(e-1)^2}{e}\frac{e^t(1+4e^t+e^{2t})}{(e^t-1)^4}$	$-(\frac{1}{t^7}+\frac{8}{t^6}+\frac{12}{t^5})e^{\frac{1}{t}-1}$	
$\frac{\psi_i'''(t)^2}{\psi_i''(t)^3},\ t>0$	≤ 1.8887	≤ 3.5445	
$\frac{\psi_i'(t)^2}{\psi_i''(t)},\quad t\to 0$	$\to +\infty$	$\to +\infty$	
$\frac{\psi_i'(t)^2}{\psi_i''(t)},\quad t\to +\infty$	$\to +\infty$	$\to +\infty$	

Table 3: Verification of the SC barrier property of $\psi_b(t)$

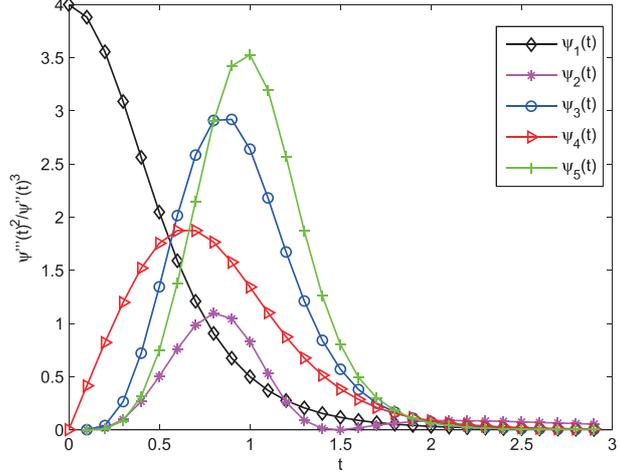| i	$\psi_{bi}(t), t>0$	$\psi_{bi}'(t), t>0$	$\psi_{bi}''(t), t>0$	$\frac{\psi_{bi}'(t)^2}{\psi_{bi}''(t)}, t\to 0$	$\frac{\psi_{bi}'(t)^2}{\psi_{bi}''(t)}, t\to +\infty$
1	$-\log t$	$-\frac{1}{t}$	$\frac{1}{t^2}$	1	1
2	$e^{\frac{1}{t}} - 2e$	$-\frac{1}{t^2}e^{\frac{1}{t}}$	$\frac{(2t+1)}{t^4}e^{\frac{1}{t}}$	$\to +\infty$	$\to 0$
3	$\frac{e^{\frac{1}{t}}-e}{e}$	$-\frac{e^{\frac{1}{t}-1}}{t^2}$	$\frac{1+2t}{t^4}e^{\frac{1}{t}-1}$	$\to +\infty$	$\to 0$
4	$\frac{(e-1)^2}{e}\frac{1}{e^t-1} - \frac{e-1}{e}$	$-\frac{(e-1)^2}{e}\frac{e^t}{(e^t-1)^2}$	$\frac{(e-1)^2}{e}\frac{e^t(1+e^t)}{(e^t-1)^3}$	$\to +\infty$	$\to 0$
5	$(\frac{1}{t}-1)e^{\frac{1}{t}-1}$	$-\frac{1}{t^3}e^{\frac{1}{t}-1}$	$(\frac{1}{t^5}+\frac{3}{t^4})e^{\frac{1}{t}-1}$	$\to +\infty$	$\to 0$

Figure 1: The SC property of kernel functions

Note that kernel functions above except $\psi_2(t)$ consist of two terms:

$$\psi(t) = \frac{t^2 - 1}{2} + \psi_b(t), \quad \forall\, t > 0, \tag{2.17}$$

where $\frac{t^2-1}{2}$ is called growth term and $\psi_b(t)$ called the barrier term. The growth term dominates the behavior of $\psi(t)$ when t approaches infinity, whereas the barrier term is monotonically decreasing in t and dominates its behavior when t approaches zero. We further investigate the SC barrier property of $\psi_b(t)$ and provide the results in Table 3. We observe that only $\psi_{b1}(t)$ is SC barrier function, while the others are not. Thus we can not use these functions to construct IPMs introduced by Nesterov in [15]. Although the IPM based on SC barriers cannot be constructed, a different type of IPM similar to the ones designed in [4] can be constructed. For the SC function $\psi_2(t)$, the IPM for LO was designed and analyzed in [6] and complexity bounds for large-update methods was obtained. In this paper the results of [6] are extended to SDO.

In this paper, we consider the following SC exponential kernel function defined by

$$\psi(t) = e^t + e^{\frac{1}{t}} - 2e, \quad t > 0. \tag{2.18}$$

Based on it, we define an n-dimensional separable barrier function by

$$\Psi(v) = \sum_{i=1}^{n} (e^{v_i} + e^{\frac{1}{v_i}} - 2e), \quad v_i > 0. \tag{2.19}$$

The Newton decrement is defined by

$$\nu(v) = (\nabla\Psi(v)^T \nabla^2\Psi(v)^{-1} \nabla\Psi(v))^{\frac{1}{2}}. \tag{2.20}$$

By spectral theorem for symmetric matrices in [19], we show how a matrix function can be obtained from a kernel function $\psi(t)$.

Definition 2.5 (Definition 5.2 in [5])**.** Let $V \in \mathbf{S}^{\mathbf{n}}_{++}$ and

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

where Q is any orthogonal matrix that diagonalize V and $\lambda(V)$ denote the eigenvalues of V. Let $\psi(t)$ be a kernel function. The (matrix valued) matrix function $\psi(V) : \mathbf{S}^{\mathbf{n}}_{++} \to \mathbf{S}^{\mathbf{n}}$ is defined by

$$\psi(V) = Q^T \text{diag}(\psi(\lambda_1(V)), \psi(\lambda_2(V)), \cdots, \psi(\lambda_n(V)))Q. \tag{2.21}$$

The real valued matrix function $\Psi(V)$ is defined as follows.

Definition 2.6 (Definition 5.3 in [5])**.** Define $\Psi(V) : \mathbf{S}^{\mathbf{n}}_{++} \to \mathbf{R}$ by

$$\Psi(V) = \mathbf{Tr}(\psi(V)) = \sum_{i=1}^{n} \psi(\lambda_i(V)), \tag{2.22}$$

where $\psi(V)$ is given by (2.21).

Note that $\psi(V)$ depends only on the restriction of $\psi(t)$ to the set of eigenvalues of V. 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 (2.21) by $\psi'(\lambda_i(V))$, $\psi''(\lambda_i(V))$ and $\psi'''(\lambda_i(V))$, respectively, we obtain that the matrix functions $\psi'(V)$, $\psi''(V)$ and $\psi'''(V)$ are defined as well.

Remark 2.7. The notation $\psi(\cdot)$ and its derivatives $\psi'(\cdot)$ and $\psi''(\cdot)$ denote matrix functions if the argument is a matrix, vector functions if the argument is a vector, and a univariate function if the argument is in $\mathbf{R}$.

3 Algorithm and Analysis

In this section, we first describe the primal-dual IPM based on $\psi(t)$. Then the computation of the step size and the decrease of $\Psi(V)$ (in one inner iteration) is analyzed. We also analyze the growth behavior of $\Psi(V)$ and obtain the complexity bounds for large-update methods. The generic form of the algorithm is shown in Figure 2.

3.1 Growth behavior of SC function

Recall that $\psi(t)$ is e-convex (The proof can be found in [6]). We also need some results from [9], [16] that we state here without proof.

Lemma 3.1 (Lemma 3.3.14(c) in [9])**.** *Let M, $N \in \mathbf{S}^n$ be two nonsingular matrices and $\psi(t)$ a real-valued function such that $\psi(e^t)$ is a convex function. Then,*

$$\sum_{i=1}^{n} \psi(\eta_i(MN)) \leq \sum_{i=1}^{n} \psi(\eta_i(M)\eta_i(N)), \tag{3.1}$$

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

Lemma 3.2 (Proposition 5.2.6 in [16])**.** *Suppose V_1, $V_2 \in \mathbf{S}^n$ are symmetric positive definite matrices, then,*

$$\Psi([V_1^{\frac{1}{2}} V_2 V_1^{\frac{1}{2}}]^{\frac{1}{2}}) \leq \frac{1}{2}(\Psi(V_1) + \Psi(V_2)). \tag{3.2}$$

Generic Primal-Dual Algorithm for SDO.

Input:
  A threshold parameter $\tau > 0$;
  an accuracy parameter $\varepsilon > 0$;
  a fixed barrier update parameter θ, $0 < \theta < 1$;
  a strictly feasible pair (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 $n\mu \geq \epsilon$ **do**
  begin
    $\mu := (1 - \theta)\mu;$
    while $\Psi(V) > \tau$ **do**
    begin
      Solve system (2.12) and (2.9) for $\Delta X, \Delta y, \Delta S$;
      Determine a step size α;
      $X := X + \alpha\Delta X;$
      $S := S + \alpha\Delta S$;
      $y := y + \alpha\Delta y$;
      $V := \dfrac{1}{\sqrt{\mu}}(D^{-1}XSD)^{\frac{1}{2}};$
    end
  end
end

Figure 2: Generic Primal-Dual Algorithm for SDO.

$\Psi(V)$ is used to measure the closeness of (X, y, S) to $(X(\mu), y(\mu), S(\mu))$, with τ as a threshold value: if $\Psi(V) \leq \tau$, then we start a new outer iteration by preforming a μ-update, otherwise we enter an inner iteration by computing the search directions at the current iterates with respect to the current value of μ and apply (2.11) to get new iterates. Hence, we need to derive an upper bound for the increase of $\Psi(V)$ after the μ-update. We have the following result.

Theorem 3.3. *Let* $\varrho : [0, \infty) \to (0, 1]$ *be the inverse function of* $\psi(t)$ *for* $0 < t \leq 1$*. Then for any positive definite matrix* V *and any* $\gamma \geq 1$ *the following inequality holds*

$$\Psi(\gamma V) \leq n\psi(\frac{1}{\gamma}\varrho(\frac{\Psi(V)}{n})). \tag{3.3}$$

Proof. When $\gamma = 1$ or when $V = E$, we observe that the inequality is obvious. Thus we just need to consider the case where $\gamma > 1$ and $V \neq E$. Let $v_i := \lambda_i(V)$, $1 \leq i \leq n$. Then $v > 0$ and

$$\Psi(\gamma V) = \sum_{i=1}^{n} \psi(\gamma\lambda_i(v)) = \sum_{i=1}^{n} \psi(\gamma v_i) = \Psi(\gamma v).$$

In order to derive the upper bound of $\Psi(\gamma V)$, which equals to $\Psi(\gamma v)$, we consider solving the following maximization problem

$$\max_{v}\{\Psi(\gamma v) : \Psi(v) = z\},$$

where z is any nonnegative number. We can use the similar method as in [6] to get the solution of this maximization problem and finally obtain the upper bound for $\Psi(\gamma v)$. By Theorem 4.2 in [6], we have

$$
\begin{aligned}
\Psi(\gamma V) = \Psi(\gamma v) &\leq n\psi(\frac{1}{\gamma}\varrho(\frac{\Psi(v)}{n})) \\
&= n\psi(\frac{1}{\gamma}\varrho(\frac{\sum_{i=1}^{n}\psi(v_i)}{n})) \\
&= n\psi(\frac{1}{\gamma}\varrho(\frac{\sum_{i=1}^{n}\psi(\lambda_i(v))}{n})) \\
&= n\psi(\frac{1}{\gamma}\varrho(\frac{\Psi(v)}{n})) \\
&= n\psi(\frac{1}{\gamma}\varrho(\frac{\Psi(V)}{n})).
\end{aligned}
$$

The proof is completed.   □

As a result of (3.3) we have that if $\Psi(V) \leq \tau$ and $\gamma = \dfrac{1}{\sqrt{1-\theta}}$, then

$$
\Psi_0 := n\psi(\sqrt{1-\theta}\varrho(\frac{\tau}{n}))
$$

is an upper bound for $\Psi(\gamma V)$, which is the value of $\Psi(V)$ after the $\mu-$update. Moreover, we get an upper bound for $\psi(\sqrt{1-\theta}\varrho(\frac{\tau}{n}))$ as follow.

$$
\begin{aligned}
\psi(\sqrt{1-\theta}\varrho(\frac{\tau}{n})) &= e^{\sqrt{1-\theta}\varrho(\frac{\tau}{n})} + e^{\frac{1}{\sqrt{1-\theta}\varrho(\frac{\tau}{n})}} - 2e \\
&= (e^{\varrho(\frac{\tau}{n})})^{\sqrt{1-\theta}} + (e^{\frac{1}{\varrho(\frac{\tau}{n})}})^{\frac{1}{\sqrt{1-\theta}}} - 2e \\
&\leq (e^{\log(\frac{\tau}{n}+e)})^{\sqrt{1-\theta}} + (e^{\log(\frac{\tau}{n}+2e-1)})^{\frac{1}{\sqrt{1-\theta}}} - 2e \\
&= (\frac{\tau}{n} + e)^{\sqrt{1-\theta}} + (\frac{\tau}{n} + 2e - 1)^{\frac{1}{\sqrt{1-\theta}}} - 2e.
\end{aligned}
$$

Hence we obtain

$$
\Psi_0 \leq n((\frac{\tau}{n} + e)^{\sqrt{1-\theta}} + (\frac{\tau}{n} + 2e - 1)^{\frac{1}{\sqrt{1-\theta}}} - 2e), \tag{3.4}
$$

which is an important inequality that will be used to derive the iteration bound in the sequel.

3.2 Computation of the step size

After a damped step we have

$$
X_+ = X + \alpha\Delta X, \quad y_+ = y + \alpha\Delta y, \quad S_+ = S + \alpha\Delta S,
$$

where the search directions $(\Delta X, \Delta y, \Delta S)$ are computed by (2.9) and (2.12). Furthermore, due to (2.9), we may write

$$
\begin{aligned}
X_+ &= X + \alpha\Delta X = X + \alpha\sqrt{\mu}DD_X D = \sqrt{\mu}D(V + \alpha D_X)D, \\
S_+ &= S + \alpha\Delta S = S + \alpha\sqrt{\mu}D^{-1}D_S D^{-1} = \sqrt{\mu}D^{-1}(V + \alpha D_S)D^{-1}.
\end{aligned}
$$

According to (2.7) we obtain

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

Note 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

$$\bar{V}_+ := ((V + \alpha D_X)^{\frac{1}{2}}(V + \alpha D_S)(V + \alpha D_X)^{\frac{1}{2}})^{\frac{1}{2}}. \tag{3.5}$$

By the definition of $\Psi(V)$, we have $\Psi(V_+) = \Psi(\bar{V}_+)$. Hence, by Lemma 3.2, we obtain that

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

In order to estimate the upper bound of $\Psi(V_+) - \Psi(V)$, we introduce the convex function $f_1(\alpha)$ which is easier to deal with than $\Psi(V_+) - \Psi(V)$. Let

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

We thus have

$$\Psi(V_+) - \Psi(V) = \Psi(\bar{V}_+) - \Psi(V) \le f_1(\alpha).$$

Let

$$v_i^{(X)} = \lambda_i(V + \alpha D_X), \quad v_i^{(S)} = \lambda_i(V + \alpha D_S), \quad v_i = \lambda_i(V),$$
$$d_{Xi} = \lambda_i(D_X), \qquad d_{Si} = \lambda_i(D_S), \quad i = 1, 2, ..., n.$$

From the system (2.9), we have that matrices V, D_X and D_S are symmetric positive definite matrices. By using the eigenvalue theorem 4.1.6 in [9], we have

$$\begin{cases} \lambda_i(V + \alpha D_X) = \lambda_i(V) + \alpha\lambda_i(D_X) = v_i + \alpha d_{Xi}, \\ \lambda_i(V + \alpha D_S) = \lambda_i(V) + \alpha\lambda_i(D_S) = v_i + \alpha d_{Si}, \end{cases} \tag{3.6}$$

That is

$$v^{(X)} = v + \alpha d_X, \quad v^{(S)} = v + \alpha d_S. \tag{3.7}$$

Using Lemma 4.1 in [6], let $\alpha \le \frac{1}{4 + \sqrt{2}\nu(v)}$. Then we have

$$\begin{aligned} 2f_1(\alpha) &= \Psi(V + \alpha D_X) - \Psi(V) + \Psi(V + \alpha D_S) - \Psi(V) \\ &= \sum_{i=1}^n \psi(\lambda_i(V + \alpha D_X)) - \sum_{i=1}^n \psi(\lambda_i(V)) + \sum_{i=1}^n \psi(\lambda_i(V + \alpha D_S)) - \sum_{i=1}^n \psi(\lambda_i(V)) \\ &= \sum_{i=1}^n \psi(v_i^{(X)}) - \sum_{i=1}^n \psi(v_i) + \sum_{i=1}^n \psi(v_i^{(S)}) - \sum_{i=1}^n \psi(v_i) \\ &= \Psi(v^{(X)}) - \Psi(v) + \Psi(v^{(S)}) - \Psi(v) \\ &\le -\frac{1}{2}\alpha\nu^2(v). \end{aligned} \tag{3.8}$$

The last inequality of (3.8) is obtained by the SC property of kernel function. Next to prove the feasibility of the new iterates, we need to analyze the requirement for α. From (2.12), we have $\|(d_X, d_S)\| = \nu(v)$. Hence, $\|d_X\| \leq \nu(v)$ and $\|d_S\| \leq \nu(v)$. Let $v_{\min}$ denote the smallest eigenvalue of the matrix V. From (3.6) and (3.7), we have

$$\begin{cases} v_i^{(X)} = v_i + \alpha d_{Xi} \geq v_{\min} - \alpha\nu(v), \\ v_i^{(S)} = v_i + \alpha d_{Si} \geq v_{\min} - \alpha\nu(v), \end{cases} \quad 1 \leq i \leq n.$$

To ensure the feasibility of the new iterates, it is required that $\alpha < \dfrac{v_{\min}}{\nu(v)}$.

Let $\alpha_0 = \min\{\dfrac{1}{4 + \sqrt{2}\nu(v)}, \dfrac{v_{\min}}{\nu(v)}\}$ and $\alpha = \beta\alpha_0$ for $\beta \in (0, 1)$. For a constant $0 < \eta \leq \frac{1}{4}$, if $\|d_X + d_S\| \leq \eta$, then $\|d_X\|$ and $\|d_S\|$ are small enough to ensure that the iterates remain in the τ-neighborhood of central path. If $\|d_X + d_S\| > \eta$, then we obtain the decrease of $\Psi(V)$:

$$\Psi(V_+) - \Psi(V) \leq -\frac{1}{4}\beta\alpha_0\nu^2(v) \leq -\frac{1}{4}\beta\alpha_0\eta^2. \tag{3.9}$$

3.3 Complexity bounds

In the inner iterate, we decrease the value of $\Psi(V)$ until it satisfies $\Psi(V) \leq \tau$. This means the iterate goes back to the τ-neighborhood of the central path. Once the inner iteration stops, we actually get a relatively 'accurate' value of $\Psi(V)$, i.e. whose value is small and relatively close to 0 (This ensures $\Psi(V) \leq \tau$ holds for a given threshold τ). In addition, the decrease of $\Psi(V)$ is derived from (3.9), the upper bound of $\Psi(V)$ after μ-update is derived from (3.4). Let the symbol $Iter_{inner}$ denotes the number of inner iterations. Hence, we get an upper bound of the number of inner iterations

$$Iter_{inner} \leq \frac{\Psi_0}{\frac{1}{4}\beta\alpha_0\eta^2} = \frac{n((\frac{\tau}{n} + e)^{\sqrt{1-\theta}} + (\frac{\tau}{n} + 2e - 1)^{\frac{1}{\sqrt{1-\theta}}} - 2e)}{\frac{1}{4}\beta\alpha_0\eta^2}. \tag{3.10}$$

Multiplying the upper bound number of (3.10) by $\dfrac{1}{\theta}\log\dfrac{n}{\epsilon}$, which is the upper bound number of barrier parameter updates, (refer to [17]. Lemma II.17, page 116), we get an upper bound for the total number of iterations

$$\frac{n((\frac{\tau}{n} + e)^{\sqrt{1-\theta}} + (\frac{\tau}{n} + 2e - 1)^{\frac{1}{\sqrt{1-\theta}}} - 2e))}{\frac{1}{4}\theta\beta\alpha_0\eta^2}\log\frac{n}{\epsilon}.$$

Set $\tau = O(n)$ and $\theta = \Theta(1)$, we finally obtain the complexity bounds $O(n\log\dfrac{n}{\epsilon})$ for large-update methods.

4 Numerical Examples

In this section we describe the numerical results of our algorithm for SDO problems. All our numerical examples are carried out on a workstation with Intel Core 2 Duo CPU at 3GHz and 4GB of physical memory. The workstation runs MATLAB version 7.11.0 (R2010a) on Windows 7 Professional operating system.

The theoretical complexity bounds for large-update methods for SDO is given in Table 4. For our experiment, the SDO problems are randomly generated as follows. First, we generate

a series of matrices $A_i(i = 1, 2, ..., m)$ which are linearly independent. Since our algorithm needs a starting point that satisfies IPC, we generate a starting point $(X^0, y^0, S^0) \in \mathbf{S}_+^{\mathbf{n}} \times \mathbf{R}^{\mathbf{m}} \times \mathbf{S}_+^{\mathbf{n}}$ randomly. Then, we set $b_i = A_i \bullet X^0$ for $i = 1, 2, ..., m$ to obtain the vector b, and let $C = \sum_{i=1}^{m} y_i A_i + S^0$. Clearly, the resulting SDO problems have primal-dual optimal solutions. We show the numerical results in Table 5 and 6. Let n denotes the dimension of SDO problems and we use the boldtype to represent the number of iterations of IPM based on SC exponential kernel function.

Table 4: Complexity bounds.

i	$\psi_i(t)$	Large-update methods	References
1	$\frac{t^2-1}{2} - \log t$	$O(n \log \frac{n}{\epsilon})$	[4]
2	$e^t + e^{\frac{1}{t}} - 2e$	$O(n \log \frac{n}{\epsilon})$	new
3	$\frac{t^2-1}{2} + \frac{e^{\frac{1}{t}}-e}{e}$	$O(\sqrt{n} \log^2 n \log \frac{n}{\epsilon})$	[5]
4	$\frac{t^2-1}{2} + \frac{(e-1)^2}{e} \frac{1}{e^t-1} - \frac{e-1}{e}$	$O(n^{\frac{3}{4}} \log \frac{n}{\epsilon})$	[5]
5	$\frac{t^2-1}{2} + (\frac{1}{t} - 1)e^{\frac{1}{t}-1}$	$O(\sqrt{n} \log^2 n \log \frac{n}{\epsilon})$	[3]

Table 5: Number of iterations based on different functions with $\theta = 0.5$

Problems	$\psi_1(t)$	$\psi_2(t)$	$\psi_3(t)$	$\psi_4(t)$	$\psi_5(t)$
n=10	28	**27**	27	29	28
n=20	45	**46**	46	49	47
n=30	56	**56**	57	58	53
n=40	79	**81**	78	85	79
n=50	91	**91**	93	95	90

Table 6: Number of iterations based on different functions with $\theta = 0.9$

Problems	$\psi_1(t)$	$\psi_2(t)$	$\psi_3(t)$	$\psi_4(t)$	$\psi_5(t)$
n=10	30	**32**	30	33	31
n=20	49	**51**	53	55	50
n=30	62	**62**	64	66	62
n=40	83	**87**	84	87	83
n=50	96	**98**	95	98	97

From Table 5 and Table 6, we notice that the performance of our algorithm based on $\psi_2(t)$ is comparable with IPMs based on different kernel functions, most notable with IPM based on logarithmic kernel function. Moreover, the total number of iterations are related with θ for large-update methods. When $\theta = 0.5$, the number of iterations are less than that of $\theta = 0.9$.

5 Concluding Remarks

In this paper, we have extended IPM based on an SC exponential kernel function for LO to SDO. Under this framework, we analyzed the algorithm and obtained the complexity bounds for large-update methods, which were analogous to the results in [6]. Finally, the numerical examples showed the algorithm was comparable with IPMs based on different kernel functions, most notable with IPM based on logarithmic kernel function.

References

[1] F. Alizadeh, Combinatorial Optimization with Interior Point Methods and Semidefinite Matrices, PhD thesis, University of Minnesota, Minneapolis, USA, 1991.

[2] F. Alizadeh, J.-P. Haeberley and M. Overton, Primal-dual interior point methods for semidefinite programming: convergence rates, stability and numirical results, *SIAM J. Optim.* 8(1998) 746–768.

[3] Y.Q. Bai, W. Xie and J. Zhang, New parameterized kernel functions for linear optimization, *J. Global Optim.* 54 (2012) 353–366.

[4] Y.Q. Bai, M.E. Ghami and C. Roos, A comparative study of kernel functions for primal-dual interior-point algorithms in linear optimization, *SIAM J. Optim.* 15 (2004) 101–128.

[5] Y.Q. Bai, *Kernel Function-based Interior-point Algorithms for Conic Optimization,* Science Press, Beijing, China, 2010.

[6] Y.Q. Bai, J. Zhang and P.F. Ma, A self-concordant exponential kernel function for primal-dual interior-point algorithm, *Optimization* 63 (2014) 931–953.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization,* Cambridge University Press, New York, 2004.

[8] M.E. Ghami, New Primal-dual Interior-point Methods based on Kernel Functions, PhD thesis, Technische Universiteit Delft, 2005.

[9] R.A. Horn and C.R. Johnson, *Matrix Analysis,* Cambridge University Press, Cambridge, UK, 1985.

[10] E.D. Klerk, *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications,* Kluwer Academic Publishers, Dordrecht, 2002.

[11] M. Kojima, S. Shindoh and S. Hara, Interior point methods for the monotone linear complementarity problem in symmetric matrices, *SIAM J. Optim.* 7 (1997) 86–125.

[12] R. Monteiro, Primal-dual path-following algorithms for semidefinite programming, *SIAM J. Optim.* 7 (1997) 663–678.

[13] Y.E. Neterrov and A.S. Nemirovskii, Interior Point a Ploynomial Methods in Convex Programming: Theory and Algorithms, *Number 13 in Studies in Applied Mathematics,* SIAM, Philadephia, USA, 1994.

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

[15] Y.E. Neterrov, Constructing self-concordant barriers for convex cones, *Core Discussion Paper* 30 (2006).

[16] J.M. Peng, C. Roos and T. Terlaky, Self-regular functions and new search directions for linear and semidefinite optimization, *Math. Program.* 93 (2002) 129–171.

[17] C. Roos, T. Terlaky and J.-Ph. Vial, *Theory and Algorithm for Linear Optimization: an* Iterior-*point Approach,* John Wiley & Sons, Chichester, UK, 1997.

[18] L. Vandenberghe and S. Boyd, Semidefinite programming, *SIAM Review.* 38 (1996) 49–95.

[19] H. Wolkowicz, R. Saigal and L. Vandenberghe, *Handbook of Semidefinite Programming, Theory, Algorithms, and Applications,* Kluwer Academic Publishers, Dordrecht, 2000.

[20] Y. Zhang, On extending some primal-dual interior-point algorithms from linear programming to semidefintie programming, *SIAM J. Optim.* 8 (1998) 365–386.

JING ZHANG
Department of Mathematics, Shanghai University, Shanghai, 200444, China
Department of Mathematics, Zhejiang A & F University, Zhejiang, 311300, China
E-mail address: zhangjingsecond@163.com

YANQIN BAI
Department of Mathematics, Shanghai University, Shanghai, 200444, China
E-mail address: yqbai@shu.edu.cn

PENGFEI MA
Department of Mathematics, Shanghai University, Shanghai, 200444, China
E-mail address: mathpengfeima@126.com