McMaster University

Advanced Optimization Laboratory

Title:

Self-Regular Proximities and New Search Directions for Linear and Semidefinite Optimization

Author:

Jiming Peng, Cornelis Roos and Tamás Terlaky

AdvOl-Report No. 2000/1

March 2000, Hamilton, Ontario, Canada

Self-Regular Proximities and New Search Directions for Linear and Semidefinite Optimization

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

March 2000

[†]*Faculty of Information Technology and Systems,*
Delft University of Technology
P.O.Box 5031, 2600 GA Delft, The Netherlands.
Email: C.Roos@its.tudelft.nl.

[‡]*Department of Computing and Software,*
McMaster University, Hamilton, Ontario, Canada, L8S 4L7.
Email: terlaky@cas.mcmaster.ca

Abstract

In this paper, we first introduce the notion of *self-regular* functions. Various appealing properties of self-regular functions are explored and we also discuss the relation between self-regular functions and the well-known *self-concordant* functions. Then we use such functions to define self-regular proximity measure for path-following interior point methods for solving linear optimization (LO) problems. Any self-regular proximity measure naturally defines a primal-dual search direction. In this way a new class of primal-dual search directions for solving LO problems is obtained. Using the appealing properties of self-regular functions, we prove that these new large-update path-following methods for LO enjoy a polynomial, $\mathcal{O}\left(n^{\frac{q+1}{2q}}\log\frac{n}{\varepsilon}\right)$ iteration bound, where $q \geq 1$ is the so-called barrier degree of the self-regular proximity measure underlying the algorithm. When q increases, this bound approaches the best known complexity bound for interior point methods, namely $\mathcal{O}\left(\sqrt{n}\log\frac{n}{\varepsilon}\right)$. Our unified analysis provides also the $\mathcal{O}\left(\sqrt{n}\log\frac{n}{\varepsilon}\right)$ best known iteration bound of small-update IPMs. At each iteration, we need only to solve one linear system. As a byproduct of our results, we remove some limitations of the algorithms presented in [24] and improve their complexity as well. An extension of these results to semidefinite optimization (SDO) is also discussed.

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

AMS Subject Classification: **90C05**

[*]This work was finished when the first author visited the Department of Computing and Software, McMaster University, Canada. Email: pengj@cas.mcmaster.ca, J.Peng@its.tudelft.nl.

1 Introduction

This work deals with interior point methods (IPMs) for linear and semidefinite optimization problems. Since the seminal work of Karmarkar [12], many researchers have proposed and analyzed various IPMs for LO and SDO and a large amount of results have been reported. For an overview of these results we refer to the recent books on the subject ([27, 33, 35]). An important fact is that almost all known polynomial-time variants of IPMs use the so-called *central path* [29] as a guideline to the optimal set, and some variant of Newton's method to follow the central path approximately. These Newton-type methods fall into different groups with respect to the strategies used in the algorithm to follow the central path. In this paper, among others, we consider these so-called path-following methods for LO and SDO. To be more specific we need to go into more detail at this stage. We start with the following linear optimization problem:

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

where $A \in \Re^{m \times n}$ satisfies $\operatorname{rank}(A) = m,\ \ b \in \Re^m,\ \ c \in \Re^n$, 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. Actually, by using the self-dual embedding model, we can further assume that $x^0 = s^0 = e$ and $\mu_0 = {x^0}^T s^0 / n = 1$. For this and some other properties mentioned below, see, e.g., [27].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 by $(x(\mu), y(\mu), s(\mu))$ 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 [29] and Megiddo [16]. If $\mu \to 0$ then the limit of the central path exists and since the limit point satisfies the complementarity condition, the limit yields optimal solutions for both (P) and (D) [27].

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 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 the 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 'closer' to $(x(\mu_+), y(\mu_+), s(\mu_+))$. We redo this procedure until the present iterate is 'close enough' to $(x(\mu_+), y(\mu_+), s(\mu_+))$ and thus we can set $\mu := \mu_+$. Then μ is reduced again by the factor $1-\theta$ and we apply Newton's method again targeting at the new μ-center, and so on. This process is repeated 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 [34] (see also Mehrotra and Ye [18] or [27]).

It may be clear from the above description that in both the analysis and implementation 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 terms of some proximity measures. In fact, the choice of the proximity measure is crucial for both the quality and the elegance of the analysis. Before describing the proximity measure, for simplification of expression, we would like to introduce some notations first. For any strictly feasible primal-dual pair (x, s) and any positive number μ, let us define

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

where $\sqrt{\frac{xs}{\mu}}$ and $\sqrt{\frac{\mu}{xs}}$ denote the vectors whose i^{th} components are $\sqrt{\frac{x_i s_i}{\mu}}$ and $\sqrt{\frac{\mu}{x_i s_i}}$ respectively. The above notations are widely used in the IPM literature to ease the notations and the analysis of IPMs. Two popular proximity measures used in the literature for primal-dual IPMs are defined as follows [27, 33, 35]:

$$\delta(x, s, \mu) := \left\| v - v^{-1} \right\|, \tag{5}$$

$$\Phi(x, s, \mu) := \sum_{i=1}^{n} \phi\left(v_i^2\right), \tag{6}$$

where

$$\phi(t) = t - 1 - \log t, \quad t > 0.$$

It is easy to see that both measures vanish if $v = e$ and go to infinity if v approaches the boundary of the nonnegative orthant. The latter property is known as the *barrier property* of the proximity measure. The second measure Φ is closely related to the *logarithmic barrier function*, with respect to the *barrier parameter* μ; its usefulness, owing to its *barrier property*, has been known already for a long time (cf. Frisch [6], Lootsma [15] and Fiacco and McCormick [5]). The measure δ, up to a factor $\frac{1}{2}$, was introduced by Jansen et al. [9], and thoroughly used in [23, 24, 27, 37]. Its SDO analogue was also used in the analysis of interior point methods for semidefinite optimization [4]. We notice that variants of the proximity $\delta(xs, \mu)$ had been used by Kojima et al. in [13] and Mizuno et al. in [19]. Many other proximities have also been used in the IPM literature [27, 33, 35]. Usually these proximities are closely related to a special class of functions: the so-called *self-concordant* functions introduced by Nesterov and Nemirovskii [21].

The choice of the parameter θ also plays an important role both in the theory and practice of IPMs. Usually, if θ is a constant independent of n the dimension of the problem, for instance

$\theta = \frac{1}{2}$, then we call the algorithm a large-update (or long-step) method. If θ depends on the problem such as $\theta = \frac{1}{\sqrt{n}}$, then the algorithm is named a small-update (or short-step) 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 primal-dual large-update methods, which are the most efficient methods in practice (see, e.g. Andersen et al. [1]). The small-update method has the best known iteration bound as $\mathcal{O}(\sqrt{n}\log\frac{n}{\varepsilon})$, while the large-update method has a worse $\mathcal{O}(n\log\frac{n}{\varepsilon})$ iteration bound [27, 33, 35]. However, large-update IPMs perform much better 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 [8, 10, 20, 27, 36, 37]. Then, at each iteration, one solves some additional equations based on the higher-order approximations to the system (2).

Different from the higher-order approach, in a recent work [24] the authors proposed a new class of search directions for LO and SDO to follow the central path and showed that large-update IPMs based on the new search direction have [1] $\mathcal{O}\left(n^{\frac{4}{3+q}}\log\frac{n}{\varepsilon}\right)$ polynomial iteration bound, where $q \in [1,3]$ is a parameter. As a fist step to the approach in this paper we briefly describe the technique used in [24] . For ease of reference, we introduce the following notations

$$d_x := \frac{v\Delta x}{x}, \quad d_s := \frac{v\Delta s}{s}; \tag{7}$$

$$\bar{d}_x := \frac{\Delta x}{x}, \quad \bar{d}_s := \frac{\Delta s}{s}. \tag{8}$$

Using the above notations and (4), one can state the centrality condition in (2) as $v = v^{-1} = e$ and rewrite the last equation in (3) as

$$d_x + d_s = v^{-1} - v. \tag{9}$$

The new search direction introduced in [24] is a slight modification of the standard Newton direction which is defined by a new system as follows

$$\begin{aligned} \bar{A}d_x &= 0,\\ \bar{A}^T\Delta y + d_s &= 0,\\ d_x + d_s &= v^{-q} - v, \end{aligned} \tag{10}$$

where $\bar{A} = \frac{1}{\mu}AV^{-1}X,\ V = \mathrm{diag}\,(v),\ X = \mathrm{diag}\,(x)$ and $q \in [1,3]$ is a parameter. An interesting observation as pointed out in [24] is that in the special case of $q = 3$, the right side of the third term in the system (10) represents the negative gradient of the proximity measure $\frac{1}{2}\delta^2$ in the v-space. When solving this system, we get the steepest descent direction for the proximity measure δ^2 along which the proximity can be driven to zero. It is also of interest to note that the right hand side of (9) is the negative gradient of the proximity $\frac{1}{2}\Phi$ which means the standard primal-dual Newton method is identical to the steepest descent method for minimizing Φ. In fact, most potential reduction methods for LO utilize the gradient of the objection function to define a search direction. For instance, the primal-dual potential function considered by Ye in [35] (see Chapter 4.3) can equivalently be written as

$$\Phi_1(x,s,\mu) = (n+q_0)\log\|v\|^2 - \sum_{i=1}^{n}\log v_i^2, \quad q_0 \geq \sqrt{n},$$

while the search direction satisfies

$$d_x + d_s = v^{-1} - \frac{n+q_0}{\|v\|^2}v.$$

[1] In [24], the algorithm employed the same search direction as given by (10) while the parameter $q-1$ is denoted by ρ with $\rho \in [0,2]$.

Note that this search direction is proportional to that defined by (9) where v is replaced by $\frac{\sqrt{n+q_0}}{\|v\|}v$, which is similar to an update of μ.

The aim of this paper is to improve the complexity of large-update IPMs. Motivated by the above observations and the results in [24], we reconsider the interrelations between the proximity measure and the search direction in IPMs.For this we first introduce the class of *self-regular* functions. Then the proximity measures are defined based on these *self-regular* functions. We define search directions for LO and SDO based on these new proximity measures. By exploring the properties of the proximities, we will establish the complexity of the corresponding algorithms. As we will see later, this approach provides a unified framework for proving polynomial complexity of a lot of existing IPMs. Moreover, it can also be used to improve the complexity of some IPMs.

The paper is organized as follows. First, in Section 2, we introduce univariate *self-regular* functions and study their properties. The common features and differences between *self-regular* and *self-concordant* functions are addressed as well. In Section 3 we utilize these *self-regular* functions to define the proximity measures and search directions for LO. Based on the properties of the proximity measures, we analyze the complexity of the methods with damped step and show that large-update IPMs using the new search directions have $\mathcal{O}\left(n^{\frac{q+1}{2q}}\log\frac{n}{\varepsilon}\right)$ polynomial iteration bounds, where $q \geq 1$ is the *barrier degree* of the proximity. For some special choices of the proximities, the complexity of the algorithm becomes $\mathcal{O}\left(\sqrt{n}\log n\log\frac{n}{\varepsilon}\right)$. The possibility of relaxing the self-regularity conditions imposed on the proximity is addressed as well. In Section 4 we discuss an extension of the new class of primal-dual algorithms to SDO and study thier complexity. We close this paper by some concluding remarks in Section 5. To make the paper easily readable, most long and cumbersome proofs of the technical results are moved to the appendix.

A few words about the notations. Throughout the paper, $\|\cdot\|$ denotes the Frobenius norm for matrices, and the 2-norm for vectors. 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 $\lambda_{\max}(G)$) to be the minimal (or maximal) eigenvalue of G. For notation simplicity, we denote by $\mathcal{I}$ the index set $\mathcal{I} = \{1, 2, \cdots, n\}$. 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 we have $\lambda_{\min}(G) = \lambda_n(G) \geq 0$, $\lambda_{\max}(G) = \lambda_1(G)$. For any matrix G, we denote by $\varrho_1(G) \geq \varrho_2(G) \geq \cdots \geq \varrho_n(G)$ the singular values of G. Especially if G is symmetric, then one has $\varrho_i(G) = |\lambda_i(G)|$ for all $i \in \mathcal{I}$. For two symmetric matrices G, H, $G \preceq H$ means $H - G$ is positive semidefinite, namely $H - G \succeq 0$.

2 Introduction to Univariate Self-Regular Functions

In this section, we present a comprehensive introduction to univariate *self-regular* functions which will be used later in our analysis for IPMs. As we will see later in Sections 3 and 4, *self-regular* functions in the positive orthant $\Re^n_{++}$ and the cone of positive definite matrices are simply transparent extensions of univariate *self-regular* functions. This section consists of three parts. In the first part we give the basic definition of univariate *self-regular* functions and provide several generating rules. Several concrete examples are presented for demonstration as well. The second part is dedicated to exploiting various appealing properties of *self-regular* functions, including the growth behavior, the barrier behavior of the functions, as well as the relations among these *self-regular* functions and their first and second-order derivatives. In the last part of this section we will discuss the common features and differences between *self-regular* functions and the well-known *self-concordant* functions.

Recall that, as we described earlier in the introduction, by using the so-called v-space, we can rewrite the centrality condition as $v = e$ or, equivalently coordinatewise, $v_i = 1$ for all $i \in \mathcal{I}$. Hence, taking it easy, we can consider consequently a univariate function in $\Re_{++}$ such that it attains its global minimum at 1 and can be used to measure the distance from any point in $\Re_{++}$ to 1. This is the primary goal of the new function. Moreover, it is also desirable for the function to enjoy certain barrier property that prevents the argument from moving to the boundary of $\Re_{++}$. This partly explains the conditions that are introduced below to define the notion of a univariate *self-regular* function.

2.1 Definitions and Examples

Definition 2.1 *A function $\psi(t) \in \mathcal{C}^2 : (0, \infty) \to \Re$ is self-regular if it satisfies the following conditions*

C.1 $\psi(t)$ is strictly convex with respect to $t > 0$ and vanishes at its global minimal point $t = 1$, i.e., $\psi(1) = \psi'(1) = 0$. Further, there exist positive constants $\nu_2 \geq \nu_1 > 0$ and $p \geq 1, q \geq 1$ such that

$$\nu_1(t^{p-1} + t^{-1-q}) \leq \psi''(t) \leq \nu_2(t^{p-1} + t^{-1-q}), \quad \forall t \in (0, \infty); \tag{11}$$

C.2 For any $t_1, t_2 > 0$,

$$\psi(t_1^r t_2^{1-r}) \leq r\psi(t_1) + (1-r)\psi(t_2), \quad \forall r \in [0, 1]. \tag{12}$$

We call parameter q the *barrier degree* and p the *growth degree* of the function $\psi(t)$ if it is *self-regular.*

For notational convenience, we first introduce a specific univariate function as follows.

$$\Upsilon_{p,q}(t) = \frac{1}{p(p+1)}\left(t^{p+1} - 1\right) + \frac{1}{q(q-1)}\left(t^{1-q} - 1\right) + \frac{p-q}{pq}(t-1), \quad p, q \geq 1. \tag{13}$$

Clearly there holds $\Upsilon''(t) = t^{p-1} + t^{-1-q}$. It is also trivial to verify that the function $\Upsilon_{p,q}(t)$ satisfies condition C.1 with $\nu_1 = \nu_2 = 1$. Since any *self-regular* function can be written as $\psi(t) = \int_1^t \int_1^\xi \psi''(\zeta) d\zeta d\xi$, by simple integrations one can easily show that for any function $\psi(t)$ satisfying C.1, there holds

$$\nu_1 \Upsilon_{p,q}(t) \leq \psi(t) \leq \nu_2 \Upsilon_{p,q}(t); \tag{14}$$

and

$$\nu_1 \left|\Upsilon'_{p,q}(t)\right| \leq |\psi'(t)| \leq \nu_2 \left|\Upsilon'_{p,q}(t)\right|. \tag{15}$$

The above two relations demonstrate the crucial role of the function $\Upsilon_{p,q}(t)$ in the class of *self-regular* functions.Further, we mention that (11) can be restated as $\nu_1 \Upsilon''_{p,q}(t) \leq \psi''(t) \leq \nu_2 \Upsilon''_{p,q}(t)$ and one can easily check that this relation, together with either (14) or (15) is equivalent to condition C.1.

On the other hand, it is straightforward to see that

$$\frac{t^{1-q} - 1}{q-1}|_{q=1} = \lim_{q \to 1} \frac{t^{1-q} - 1}{q-1} = -\log t.$$

In light of this relation, from now on the argument $\frac{t^{1-q}-1}{q-1}$ in the paper is endowed as $-\log t$ when $q=1$. Using this notation, one can write $\Upsilon_{1,1}(t) := \frac{1}{2}t^2 - \frac{1}{2} - \log t$ the classical logarithmic barrier function.[2] At the present stage, it is not yet clear that the function $\Upsilon_{p,q}$ is *self-regular*. As asserted earlier in this section, $\Upsilon_{p,q}$ satisfies condition C.1. In Lemma 2.5 we will show that $\Upsilon_{p,q}$ satisfies C.2 as well.

To facilitate our analysis, let us denote by Ω_1, Ω_2 the sets of functions whose elements satisfy conditions C.1 and C.2, respectively[3].

We start with an important lemma that gives a different characterization of condition C.2.

Lemma 2.2 [4]*A twice differentiable function $\psi(t) : (0, +\infty) \to \Re$ belongs to Ω_2 if and only if the function $\psi(\exp(\zeta)) : \Re \to \Re$ is convex in ζ, or equivalently $\psi'(t) + t\psi''(t) \geq 0$ for $t > 0$.*

Proof: The proof is straightforward. First from the definition of convexity, we know that $\psi(\exp(\zeta))$ is convex if and only if for any $\zeta_1, \zeta_2 \in \Re$, the following inequality holds

$$\psi\left(\exp\left(r\zeta_1 + (1-r)\zeta_2\right)\right) \leq r\psi(\exp(\zeta_1)) + (1-r)\psi(\exp(\zeta_2)), \quad r \in [0,1]. \tag{16}$$

Letting $t_1 = \exp(\zeta_1), t_2 = \exp(\zeta_2)$, obviously one has $t_1, t_2 \in (0, +\infty)$. Further, relation (16) can be equivalently rewritten as

$$\psi({t_1}^r {t_2}^{1-r}) \leq r\psi(t_1) + (1-r)\psi(t_2), \quad r \in [0,1].$$

Since $\psi(\exp(\zeta))$ is convex, i.e., $\exp(2\zeta)\,\psi''(\exp(\zeta)) + \exp(\zeta)\,\psi'(\exp(\zeta)) \geq 0$. Substituting by $t = \exp(\zeta)$, one gets $t\psi'(t) + t^2\psi''(t) \geq 0$ which is equivalent to $\psi'(t) + t\psi''(t) \geq 0$ for $t > 0$. This completes the proof of the lemma. □

Note that condition C.2 can be restated equivalently (it is a standard exercise in convexity theory) as

$$\psi(t_1 t_2) \leq \frac{1}{2}\left(\psi(t_1^2) + \psi(t_2^2)\right), \quad \forall t_1, t_2 > 0. \tag{17}$$

From its definition (12) or (17), one gets

Lemma 2.3 *If $\psi(t) \in \Omega_2$, then so is $\psi\left(\frac{1}{t}\right)$.*

Proof: Defining $\phi(t) = \psi(\frac{1}{t})$, from (17) we obtain

$$\phi(t_1 t_2) = \psi\left(\frac{1}{t_1 t_2}\right) \leq \frac{1}{2}\left(\psi\left(\frac{1}{t_1^2}\right) + \psi\left(\frac{1}{t_2^2}\right)\right) = \frac{1}{2}\left(\phi(t_1^2) + \phi(t_2^2)\right).$$

The proof of the lemma is finished. □

It is straightforward to verify that the functions $\log t, -\log t \in \Omega_2$ and for any $\rho \in \Re$, one has $t^\rho \in \Omega_2$ as well. This gives, since Ω_2 is closed under addition,

[2]Observe that (6) and the definition of $\phi(t)$ implies $2\Upsilon_{1,1}(v_i) = \phi(v_i^2)$.

[3]Condition C.1 is a very simple condition which can be verified without much difficulties. Hence in this section we focus only on condition C.2.

[4]Another reason why we give this lemma is that, the second condition in the lemma has been widely used in the analysis of singular value inequalities in matrix theory [3]. However, our original condition C.2 is clearer and more suitable for our specific purpose.

Proposition 2.4 *Let N be any positive integer and*

$$\psi(t) = \beta_0 \log t + \sum_{i=1}^{N} \beta_i (t^{\rho_i} - 1), \quad \beta_0 \in \Re,\ \beta_i \geq 0,\ \rho_i \in \Re,\ i = 1, 2, \cdots, N. \tag{18}$$

Then $\psi(t) \in \Omega_2$.

Hence any function $\psi(t)$ defined by (18) that satisfies condition C.1 is *self-regular*. Now let us return to consider the special case $\Upsilon_{p,q}(t)$. One has

Lemma 2.5 *The function $\Upsilon_{p,q}(t)$ is self-regular when $p, q \geq 1$.*

Proof: We need only to show that $\Upsilon_{p,q}(t) \in \Omega_2$. Through simple calculus, one has

$$\Upsilon'_{p,q}(t) + t\Upsilon''_{p,q}(t) = \frac{p+1}{p} t^p + \frac{q-1}{q} t^{-q} + \frac{1}{q} - \frac{1}{p}.$$

Hence, it follows immediately that for any $p \geq q \geq 1$, there holds

$$\Upsilon'_{p,q}(t) + t\Upsilon''_{p,q}(t) > \frac{1}{q} - \frac{1}{p} \geq 0.$$

Thus, by Lemma 2.2, it remains to prove that

$$\Upsilon'_{p,q}(t) + t\Upsilon''_{p,q}(t) \geq 0, \quad \forall q > p \geq 1, t > 0.$$

Note that when $q > p \geq 1$, one has $1 - \frac{1}{q} \geq \frac{1}{p} - \frac{1}{q}$ which further implies that for any $t > 0$,

$$\begin{aligned} \Upsilon'_{p,q}(t) + t\Upsilon''_{p,q}(t) &= \frac{p+1}{p} t^p + \frac{q-1}{q} t^{-q} + \frac{1}{q} - \frac{1}{p} \geq \frac{p+1}{p} t^p + \frac{q-p}{pq} t^{-q} + \frac{1}{q} - \frac{1}{p} \\ &> \left(\frac{1}{p} - \frac{1}{q}\right)\left(t^p + t^{-q} - 1\right) > 0. \end{aligned}$$

The proof of the lemma is completed. □

To show that $\Upsilon_{p,q}(t)$ is not the only family of *self-regular* functions, we proceed with another example.

Example 2.6 Consider the function

$$\psi(t) = \frac{t^{p+1} - 1}{p+1} + \frac{t^{1-q} - 1}{q-1}, \quad p \geq 1, q > 1.$$

One has

$$\begin{aligned} \psi'(t) &= t^p - t^{-q} \\ \psi''(t) &= p t^{p-1} + q t^{-q-1}. \end{aligned}$$

It is obvious that condition C.1 is satisfied (with $\nu_1 = \min(p, q)$, $\nu_2 = \max(p, q)$). To check C.2, we use Lemma 2.2:

$$\psi'(t) + t\psi''(t) = (p+1)t^p + (q-1)t^{-q} \geq 0.$$

This shows that $\psi(t)$ is self-regular. Note that if $p = q$ then $\psi(t) = p\Upsilon_{p,p}(t)$, but otherwise the functions $\psi(t)$ and $\Upsilon_{p,q}(t)$ are linearly independent.

A very simple but quite useful observation is that any nontrivial nonnegative combination of two *self-regular* functions ψ_1 and ψ_2 is still *self-regular*. This is done in our following proposition whose proof is provided in the appendix of the paper.

Proposition 2.7 *If the functions $\psi_1(t), \psi_2(t)$ are self-regular, then any nonnegative combination $\beta_1\psi_1 + \beta_2\psi_2$ with $\beta_1, \beta_2 \geq 0,\ \beta_1 + \beta_2 > 0$ is self-regular.*

Proof: See Appendix A.1. □

The above proposition means that the set of self-regular functions is a pointed convex cone.

Note that in Lemma 2.3, we have shown that if $\psi(t) \in \Omega_2$, so is the function $\psi\left(\frac{1}{t}\right)$. In what follows we consider a converse question. Suppose that a function satisfies the relation $\psi(t) = \psi\left(\frac{1}{t}\right)$. The issue we want to address is when such a function is *self-regular*. Our next result gives a positive answer to this question and thus provides another way to generate *self-regular* functions. This result will be used in our later discussions about the relations between *self-regular* functions and *self-concordant* functions.

Lemma 2.8 *If $\psi(t) = \psi(t^{-1})$ and $\psi(t) \in \Omega_1$, then $\psi(t)$ is self-regular.*

Proof: It suffices to prove that, for any $t_1, t_2 > 0$,

$$\psi(t_1 t_2) \leq \frac{1}{2}\left(\psi(t_1^2) + \psi(t_2^2)\right).$$

Since $\psi(t) \in \Omega_1$, $\psi(t)$ is increasing for $t \geq 1$. Hence if $t_1 t_2 \geq 1$, then by condition C.1 there holds

$$\psi(t_1 t_2) \leq \psi\left(\frac{t_1^2 + t_2^2}{2}\right) \leq \frac{1}{2}\left(\psi(t_1^2) + \psi(t_2^2)\right).$$

Thus it remains to consider the case $t_1 t_2 < 1$, or in other words $\frac{1}{t_1 t_2} > 1$. Now using the assumption $\psi(t) = \psi(t^{-1})$, we obtain

$$\psi(t_1 t_2) = \psi\left(\frac{1}{t_1 t_2}\right) \leq \frac{1}{2}\left(\psi\left(\frac{1}{t_1^2}\right) + \psi\left(\frac{1}{t_2^2}\right)\right) = \frac{1}{2}\left(\psi(t_1^2) + \psi(t_2^2)\right).$$

The proof of the lemma is completed. □

By way of illustration, consider $\psi(t) = \frac{1}{2}(t - t^{-1})^2$, a function well known in the IPM literature. Clearly, this function satisfies the hypothesis of Lemma 2.8. One may easily check that it satisfies C.1, with $p = \nu_1 = 1$ and $q = \nu_2 = 3$. By Lemma 2.8, the self-regularity follows. Note that this function can also be put in the form given by Example 2.6.

We close this subsection by giving some examples which show the independence of conditions C.1 and C.2. Let $\psi_1(t) = t^2 - t - \log t$, it satisfies condition C.1. However, by simple calculus one gets

$$\psi_1'(t) + t\psi_1''(t) = 4t - 1.$$

¿From Lemma 2.2 we know that $\psi_1(t)$ does not satisfy C.2 since $\psi_1'(t) + t\psi_1''(t) < 0$ whenever $t < \frac{1}{4}$. On the other hand, the simple function t^2 satisfies C.2 but not C.1. Thus the conditions C.1 and C.2 are independent.

2.2 Fundamental Properties of Univariate Self-Regular Functions

This section is committed to investigating various fascinating properties of functions satisfying condition C.1. As we will see later in Section 3.5, in case of LO, by means of using several technical conclusions from this section one can build up the polynomial complexity of IPMs without resorting to condition C.2.

We start by presenting several intrinsic features of any function $\psi(t) \in \Omega_1$ with $q > 1$ which display some intriguing relationships among the function $\psi(t) \in \Omega_1$ and its derivatives. Since condition C.1 is a necessary condition for *self-regularity*, these attractive features are naturally shared by general univariate *self-regular* functions.

Lemma 2.9 *Suppose that $\psi(t) \in \Omega_1$. Then there holds*

$$\left|\frac{1}{t}\psi'(t)\right| \leq \frac{\nu_2}{\nu_1}\psi''(t), \quad t > 0.$$

Proof: See Appendix A.1. □

Recall that by Lemma 2.2, a function $\psi(t) \in \Omega_2$ if and only if $\psi'(t) + t\psi''(t) \geq 0$ for any $t > 0$. It follows from Lemma 2.9

Corollary 2.10 *If a function $\psi(t) \in \Omega_1$ with $\nu_1 = \nu_2$, then it is self-regular.*

The above corollary is a direct generalization of Lemma 2.5, since for any positive $\nu > 0$, the function $\nu\Upsilon_{p,q}(t)$ is also *self-regular.*

Proposition 2.11 *Suppose that the function $\psi(t)$ satisfies the condition C.1 with $q > 1$. Then there is a constant $\mathcal{C}_\nu$ such that*

$$\psi(t)\psi''(t) \leq \mathcal{C}_\nu \psi'(t)^2, \quad \forall t > 0.$$

Proof: See Appendix A.1. □

It is worthwhile to mention that that the above proposition is not true if $q = 1$.

The above relations among the function $\psi(t)$ and its first and second derivatives given by Lemma 2.9 and Proposition 2.11 provide partially an explanation for the name *self-regularity.* As announced before, we will prove that *self-regular* functions can be used to improve the complexity of large-update IPMs. Unfortunately we have not been able to use the above properties for that goal. In that respect the results that follow in this section are more important: they play a crucial role in the analysis of large-update IPMs.

First we give an estimation about a function $\psi(t)$ in Ω_1 in terms of t and $\psi'(t)$, one has

Lemma 2.12 *Suppose that $\psi(t) \in \Omega_1$. Then there holds*

$$\frac{1}{2}(t-1)^2 \leq \frac{\psi(t)}{\nu_1}; \tag{19}$$

$$\frac{t^{p+1}-1}{p(p+1)} - \frac{t-1}{p} \leq \frac{\psi(t)}{\nu_1}; \tag{20}$$

$$\frac{t^{1-q}-1}{q(q-1)} + \frac{t-1}{q} \leq \frac{\psi(t)}{\nu_1}; \tag{21}$$

and

$$\psi(t) \leq \frac{1}{2\nu_1}\psi'(t)^2. \tag{22}$$

Proof: The first three inequalities from (19) to (21) are immediate consequences of the following relations

$$\psi''(t) \geq \nu_1(t^{p-1}+t^{-q-1}) \geq \nu_1 \max\left\{t^{p-1}, t^{-q-1}\right\} = \nu_1 \max\left\{1, t^{p-1}, t^{-q-1}\right\}, \quad \forall t>0, p,q \geq 1.$$

Using condition C.1 again, one can easily see that

$$\frac{\psi''(t)}{\nu_1} > 1.$$

It follows

$$\psi(t) = \int_1^t \int_1^\xi \psi''(\zeta) d\zeta d\xi \leq \frac{1}{\nu_1}\int_1^t \int_1^\xi \psi''(\zeta)\psi''(\xi) d\zeta d\xi = \frac{1}{\nu_1}\int_1^t \psi''(\xi)\left(\psi'(\xi)-\psi'(1)\right) d\xi = \frac{1}{2\nu_1}\psi'(t)^2,$$

which gives (22). The lemma is proved. □

Our next lemma provides some lower bounds for $|\psi'(t)|$ which can be viewed as characterizations of the barrier and growth behaviors of the functions $\psi'(t)$ and $\psi(t)$ as well.

Lemma 2.13 *Suppose that $\psi(t) \in \Omega_1$. Then there holds*

$$|\psi'(t)| \geq \frac{\nu_1}{q}(t^{-q}-1), \quad \forall t<1,$$

and

$$|\psi'(t)| \geq \frac{\nu_1}{p}(t^{p}-1), \quad \forall t>1.$$

Proof: From C.1 we obtain, whenever $t<1$,

$$\psi'(t) = \int_1^t \psi''(\xi) d\xi \leq \nu_1 \int_1^t \left(\xi^{p-1}+\xi^{-1-q}\right) d\xi \leq \nu_1 \int_1^t \xi^{-1-q} d\xi = \frac{\nu_1}{q}(1-t^{-q}).$$

The above relation gives the first inequality of the lemma directly. One can prove the second inequality similarly. □

Finally we close this section by presenting a result about the growth behavior of $\psi(t)$ with respect to t.

Lemma 2.14 *Suppose that $\psi(t) \in \Omega_1$. Then, for any $\vartheta > 1$, we have*

$$\psi(\vartheta t) \leq \frac{\nu_2}{\nu_1}\left(\vartheta^{p+1}\psi(t) + \vartheta \Upsilon'_{p,q}(\vartheta)\sqrt{2\nu_1\psi(t)} + \nu_1 \Upsilon_{p,q}(\vartheta)\right).$$

Proof: See Appendix A.1. □

An immediate consequence of the above lemma is

Corollary 2.15 *Let $\psi(t) \in \Omega_1$ and $\vartheta > 1$. Then there exist two constants $\nu_3, \nu_4 > 0$, depending on p and q, such that*

$$\psi(\vartheta t) \leq \frac{\nu_2 \nu_4}{\nu_1} \left(\psi(t) + (\vartheta - 1)\sqrt{2\nu_1 \psi(t)} + \nu_1 (\vartheta - 1)^2 \right),$$

for all $\vartheta \in [1, 1 + \nu_3]$.

Proof: With l'Hospital's rule we easily obtain the following limits

$$\lim_{\vartheta \to 1} \frac{\Upsilon_{p,q}(\vartheta)}{(\vartheta - 1)^2} = 1, \quad \lim_{\vartheta \to 1} \frac{\vartheta \Upsilon'_{p,q}(\vartheta)}{\vartheta - 1} = 2.$$

The above limits imply the existence of positive constants ν_3 and ν_4, such that for any $\vartheta \in [1, 1 + \nu_3]$

$$\vartheta^{p+1} \leq \nu_4, \quad \vartheta \Upsilon'(\vartheta) \leq \nu_4 (\vartheta - 1), \quad \Upsilon_{p,q}(\vartheta) \leq \nu_4 (\vartheta - 1)^2.$$

Substitution of these inequalities into Lemma 2.14, yields the desired inequality. □

2.3 Relations between Self-Regularity and Self-Concordance

In this section we discuss the relationship between the new class of *self-regular* functions and the class of so-called *self-concordant* functions in $\Re_{++}$. The case $\Re^n_{++}$ (which we will introduce later) follows analogously. Let us first recall the definition of a univariate self-concordant function [21].

Definition 2.16 *A function $f(t) \in \mathcal{C}^3 : (0, \infty) \to (0, \infty)$ is self-concordant if f is convex, three times continuously differentiable, and satisfies the condition*

$$|f'''(t)| \leq \bar{\nu} \left(f''(t)\right)^{\frac{3}{2}}, \quad \forall t \in (0, \infty),$$

for some constant $\bar{\nu} > 0$.

It is worthwhile to point out that in their remarkable book [21], Nesterov and Nemirovskii introduced the notion of so-called *self-concordant* barrier to analyze IPMs for general convex optimization problems.

Definition 2.17 *A univariate function $f(t) \in \mathcal{C}^3$ is said to be a self-concordant barrier for the domain $\Re_+$ if it satisfies the following two conditions:*
(i): The function $f(t)$ is self-concordant.
(ii): There exists a constant $\tilde{\nu}$ such that

$$|f'(t)| \leq \tilde{\nu} \left(f''(t)\right)^{\frac{1}{2}}.$$

An interesting characterization of a univariate *self-concordant* barrier is that, as observed by F. Glineur [7], that Definitions 2.16 and 2.17 imply the existence of a constant $\hat{\nu}$ such that

$$f'(t) f'''(t) \leq \hat{\nu} \left(f''(t)\right)^2.$$

This relation is precisely the conclusion presented in Proposition 2.11 where the function $\psi(t)$ is replaced by $f'(t)$.

Since the set of *self-concordant* barriers is only a subset[5] of the set of *self-concordant* functions, in the sequel we will only focus on the relations between the *self-concordant* functions and *self-regular* functions.

We first observe that the self-regularity condition only requires that $\psi(t)$ is twice differentiable, whereas self-concordancy requires thrice differentiability. This does not mean, however, that the self-regularity condition is weaker than the condition for self-concordancy.

By way of example, consider

$$\psi(t) = t - 1 - \log t,$$

which can be considered as the prototype of a univariate self-concordant function. Since

$$\psi'(t) = 1 - \frac{1}{t}, \quad \psi''(t) = \frac{1}{t^2}, \quad \psi'''(t) = -\frac{2}{t^3},$$

the self-concordancy of $\psi(t)$, with $\bar{\nu} = 2$, readily follows. It is easy to see that $\psi(t)$ does not satisfy condition C.1, since the second derivative $\psi''(t) = t^{-2}$ reduces to zero as t goes to infinity. Hence, $\psi(t)$ is not self-regular.

It is much harder to find a self-regular function that is not self-concordant. In fact, a large class of self-regular functions is self-concordant, as the next result shows. In this lemma we consider the functions defined by

$$\psi(t) = \beta_0 \log t + \sum_{i=1}^{N} \beta_i \left(t^{\rho_i} - 1\right). \tag{23}$$

Proposition 2.18 *If $\psi(t)$, as given by (23), belongs to Ω_1, then $\psi(t)$ is self-concordant.*

Proof: If $\psi(t) \in \Omega_1$, then one has

$$p := \max\{\rho_i - 1 : i = 1, \cdots, N\} \geq 1, \quad q := \max\{1 - \rho_i : i = 1, \cdots, N\} \geq 1.$$

The above definitions imply

$$t^{\rho_i - 2} \leq \max\{t^{p-1}, t^{-q-1}\}, \quad \forall t > 0,$$

which allows us to write, for any $t > 0$,

$$\begin{aligned} t^{\rho_i - 3} &\leq \max\{t^{\frac{3}{2}(\rho_i - 2)}, t^{-3}\} \leq \max\{t^{\frac{3}{2}(p-1)}, t^{-3}, t^{\frac{3}{2}(-q-1)}\} \\ &= \max\{t^{\frac{3}{2}(p-1)}, t^{\frac{3}{2}(-q-1)}\} \leq \left(t^{p-1} + t^{-q-1}\right)^{\frac{3}{2}}, \end{aligned}$$

where the first inequality follows from $t^{\rho_i} \leq \max\{t^{\frac{3}{2}\rho_i}, 1\}$. From the above inequality, one can conclude that there exists a constant $\bar{\nu}$ depending on the constants ν_1, ν_2 from (11), the coefficients β_i and the exponents ρ_i, such that

$$\left|\psi'''(t)\right| \leq \bar{\nu} \left(\psi''(t)\right)^{\frac{3}{2}}.$$

This proves the proposition. □

All examples of self-regular functions given so far, were of the form (23). Thus we conclude that all self-regular functions considered so far, are self-concordant.

We conclude this section by showing that there exist *self-regular* functions that are not *self-concordant*.

[5] As F. Jarre pointed out on page 272 of [11] Definition 2.17 is equivalent to requiring that the function $g(t) = \exp(-f(t)/\hat{\nu})$ is concave. This relation implies that self-concordant barriers are always logarithmic functions, i.e., $f(t) = -\hat{\nu} \log(g(t))$.

Example 2.19 We consider the function

$$\begin{aligned}\psi(t) &= (t-t^{-1})^2 + \frac{1}{3}\left(\int_1^t \sin\zeta^{-3} d\zeta + \int_1^{t^{-1}} \sin\zeta^{-3} d\zeta\right) \\ &= (t-t^{-1})^2 + \frac{1}{3}\int_1^t \left(\sin\zeta^{-3} - \zeta^{-2}\sin\zeta^3\right) d\zeta.\end{aligned}$$

Obviously, $\psi(1) = \psi'(1) = 0$. By direct calculus, we get

$$\begin{aligned}\psi'(t) &= 2t - 2t^{-3} + \tfrac{1}{3}\sin t^{-3} + -\tfrac{1}{3}t^{-2}\sin t^3 \\ \psi''(t) &= 2 + 6t^{-4} - t^{-4}\cos t^{-3} + \frac{2}{3}t^{-3}\sin t^3 - \cos t^3.\end{aligned}$$

Since for any $t > 0$,

$$-1 \le \cos t^{-3}, \sin t^3, \cos t^3 \le 1, \quad t^{-4} \ge \frac{4}{3}t^{-3} - \frac{1}{3},$$

it follows

$$\psi''(t) \ge 1 + 5t^{-4} - \frac{2}{3}t^{-3} \ge \frac{5}{6} + \frac{9}{2}t^{-4} > \frac{5}{6}\left(1 + t^{-4}\right);$$

and

$$\psi''(t) \le 3 + 7t^{-4} + \frac{2}{3}t^{-3} < \frac{15}{2}\left(1 + t^{-4}\right).$$

The above two inequalities imply $\psi(t) \in \Omega_1$. Since $\psi(t) = \psi(t^{-1})$ for any $t > 0$, Lemma 2.8 implies that $\psi(t)$ is in Ω_2. Hence $\psi(t)$ is *self-regular*. It is not hard to verify that $\psi(t)$ is not self-concordant.

From the results of this section it is clear that the set of self-regular functions and the set of self-concordant functions are different, but there is really a large intersection of these two sets. The functions in the intersection share all the good properties of self-concordant functions, but as we will show the additional property of being self-regular enables us to get a better complexity result for large-update IPMs than can be obtained on the basis of self-concordancy alone.

3 New Primal-Dual Methods for LO and Their Complexity

In the present section we propose a new class of primal-dual Newton-type methods for LO and study the complexity of large-update algorithms. The section composes of five parts. In the first subsection we introduce *self-regular* proximity measures for LO and discuss their properties. In the second subsection we describe the new algorithm based on the *self-regular* proximity. In the third subsection we estimate the proximity measure after one step. The complexity of the algorithm is reported in the fourth subsection. In the last subsection we show that the same complexity bound can be obtained if we only assume that condition C.1 is satisfied by the proximity function, and thus deleting the assumption that the proximity function satisfies condition C.2.

3.1 Self-Regular Proximities for LO

We start with the definition of *self-regular* functions on $\Re^n_{++}$.For simplification of exposition, if no confusion occurs, we allow ourselves sometimes abuse some notations such as the function $\psi(\cdot)$.

Definition 3.1 *A function $\Psi(x) : \Re^n_{++} \to (0, \infty)$ is said to be self-regular if*

$$\Psi(x) = \sum_{i=1}^{n} \psi(x_i), \tag{24}$$

where $\psi(t) \in \mathcal{C}^2 : \Re_{++} \to (0, \infty)$ is self-regular.

In view of the above definition, it is easily seen that *self-regular* functions in $\Re^n_{++}$ also enjoy certain attractive features inheriting directly from their ancestors in $\Re_{++}$. Hence, instead of boring repetitions, we will simply omit such discussions and move onto the main thread in this work, the new proximity and search direction based on it.

Keeping in mind the notations (4), we define new proximity measures for LO as follows.

$$\Psi(x, s, \mu) := \Psi(v) = \sum_{i=1}^{n} \psi(v_i). \tag{25}$$

We say the proximity is *self-regular* if the function $\psi(t)$ is *self-regular*. First we give a relation between the proximities in the v-space, the primal and the dual space. One has

Proposition 3.2 *Let the proximity $\Psi(v)$ is defined by (25). If it is self-regular, then*

$$\Psi(v) \le \frac{1}{2}\left(\Psi\left(\frac{x}{\sqrt{\mu}}\right) + \Psi\left(\frac{s}{\sqrt{\mu}}\right)\right).$$

Proof: The conclusion follows readily from the fact that

$$v_i = (\mu^{-\frac{1}{4}} x_i^{\frac{1}{2}})(\mu^{-\frac{1}{4}} s_i^{\frac{1}{2}}), \quad \forall i \in \mathcal{I},$$

and condition C.2. □

The above proposition is very interesting. It is also helpful in our later discussion about the decreasing behavior of the proximity measure. However, for simplicity we prefer to use condition C.2 directly in our later analysis.

Our next result summarizes some appealing properties of the proximity $\Psi(v)$ which are also shared by general *self-regular* functions in $\Re^n_{++}$. For notational convenience, we also define

$$\sigma = \|\nabla \Psi(v)\|. \tag{26}$$

We have

Proposition 3.3 *Let the proximity $\Psi(v)$ be defined by (25). Then there holds*

$$\Psi(v) \le \frac{\sigma^2}{2\nu_1}, \tag{27}$$

$$v_{\min} \ge \left(1 + \frac{q\sigma}{\nu_1}\right)^{-\frac{1}{q}}, \tag{28}$$

and

$$v_{\max} \le \left(1 + \frac{p\sigma}{\nu_1}\right)^{\frac{1}{p}}. \tag{29}$$

If $v_{\max} > 1$ *and* $v_{\min} < 1$, *then*

$$\sigma \geq \nu_1 \left(\frac{(v_{\max}^p - 1)^2}{p^2} + \frac{(v_{\min}^{-q} - 1)^2}{q^2} \right)^{\frac{1}{2}}. \tag{30}$$

For any $\vartheta > 1$,

$$\Psi(\vartheta v) \;\leq \frac{\nu_2}{\nu_1} \left(\vartheta^{p+1} \Psi(v) + \vartheta \Upsilon'_{p,q}(\vartheta) \sqrt{2n\nu_1 \Psi(v)} + n\nu_1 \Upsilon_{p,q}(\vartheta) \right). \tag{31}$$

If $\vartheta \in (1, 1+\nu_3]$, *then*

$$\Psi(\vartheta v) \leq \frac{\nu_2 \nu_4}{\nu_1} \Psi(v) + \frac{\nu_2 \nu_4 \sqrt{2n\nu_1 \Psi(v)}}{\nu_1} (\vartheta - 1) + n\nu_2\nu_4(\vartheta - 1)^2, \tag{32}$$

where ν_3, ν_4 *are the same constants as introduced in Corollary 2.15.*

Proof: The first inequality of the proposition follows from Lemma 2.12. To prove the second inequality, we first note that the inequality holds trivially if $v_{\min} \geq 1$. If $v_{\min} < 1$, from Lemma 2.13 one obtains

$$\sigma \geq |\psi'(v_{\min})| \geq \frac{\nu_1}{q} \left(\frac{1}{v_{\min}^q} - 1 \right),$$

which implies (28). By following an analogous process one can show that both (29) and (30) are true. The inequality (31) follows immediately from Lemma 2.13, by means of applying the Cauchy-Schwartz inequality to the all one vector e and the vector $(\sqrt{\psi(v_1)}, \cdots, \sqrt{\psi(v_n)})^T$. By using (31) and following a similar chain as in the proof of Corollary 2.15, one can readily obtain the last conclusion of the proposition. □

Inequality (32) implies that, if $\Psi(v) \leq \tau$ with $\tau > 0$, and $\vartheta - 1 = \frac{\nu_3}{\sqrt{n}} \leq \nu_3$, then

$$\Psi(\vartheta v) \leq \frac{\nu_2 \nu_4 \tau}{\nu_1} + \nu_2 \nu_3 \nu_4 \sqrt{\frac{2\tau}{\nu_1}} + \nu_2 \nu_4 \nu_3^2. \tag{33}$$

Thus, if $\Psi(v) \leq \tau = \mathcal{O}(1)$, then $\Psi(\vartheta v) \leq \mathcal{O}(1)$.

It is instructive to discuss the relations between the duality gap and the new proximity. Note that by using the first conclusion of Lemma 2.12 we readily obtain

$$\frac{\Psi(v)}{\nu_1} \geq \frac{1}{2} \|v - e\|^2 = \frac{1}{2} \|v\|^2 - \sum_{i=1}^{n} v_i + \frac{n}{2} \geq \frac{1}{2} \|v\|^2 - \sqrt{n} \|v\| + \frac{n}{2},$$

which further implies

$$\|v\| \leq \sqrt{n} + \sqrt{\frac{2\Psi(v)}{\nu_1}}.$$

It follows immediately

$$x^T s = \mu \|v\|^2 \leq n\mu + 2\mu \sqrt{\frac{2n\Psi(v)}{\nu_1}} + \frac{2\Psi(v)}{\nu_1} \mu. \tag{34}$$

Hence if $\Psi(v) = \mathcal{O}(n)$, then $x^T s \leq \mathcal{O}(n\mu)$. This implies that the proximity can be used as a potential function for minimizing the duality gap.

It deserves to consider a little more about which kind of *self-concordant* functions are not *self-regular*. Let us recall a previous example $\psi(t) = t - 1 - \log t$. This function is *self-concordant* but not *self-regular*. To be more specific, let us consider the corresponding proximity $\tilde{\Psi}(v) = \sum_{i=1}^{n}(v_i - 1 - \log v_i)$. Our target is to minimize the duality gap $x^T s$ while staying in a suitable neighborhood of the central path. When the parameter μ is fixed, this is equivalent to minimizing the argument $\|v\|^2$ in the v-space. Naturally we expect the proximity measure can act as a potential function for the scaled duality gap $\|v\|^2$ or the square of the distance to the central path $\|v-e\|^2$, or in other words, $\tilde{\Psi}(v) \geq \hat{c}\max\left(\|v\|^2, \|v-e\|^2\right)$ for some constant $\hat{c} > 0$. However, with a closer look one can easily see that such a constant $\hat{c}$ does not exist for all $v \in \Re^n_{++}$. Hence, taking this point into account, the proximity measure is not 'good'. Thus we consider this function not suitable for our purpose. On the other hand, the function $\psi(t) = t^2 - 2t + 1$ is *self-concordant* but not *self-regular*, since there is no barrier term in it. One can see the corresponding proximity $\Psi(v) = \|v-e\|^2$ does not have a good control on the distance to the boundary of the feasible set, since even when $\|v\|$ goes to zero, the proximity measure $\Psi(v)$ does not increase to infinity. Again we exclude this proximity from our 'regular' group.

3.2 The Algorithm

We progress to outline the procedure of the new algorithm. Let $d_x, \Delta y, d_s$ denote the solution of the following modified Newton equation system for the scaled system (10):

$$\begin{aligned} \bar{A} d_x &= 0, \\ \bar{A}^T \Delta y + d_s &= 0, \\ d_x + d_s &= -\nabla\Psi(v), \end{aligned}$$

or, $(\Delta x, \Delta y, \Delta s)$ the solution of its equivalent system

$$\begin{aligned} A\Delta x &= 0, \\ A^T \Delta y + \Delta s &= 0, \\ s\Delta x + x\Delta s &= -\mu v \nabla\Psi(v). \end{aligned} \tag{35}$$

The last equation in the above system represents the pseudo-Newton system for the equation

$$xs = -\mu v(\nabla\Psi(v) - v).$$

We would like to remark that when we use the new proximity $\Psi(v)$ with a large value q to define the neighborhood and then employ the new system (35), the small entries $x_i s_i$ (or equivalently v_i) are heavily penalized in terms of the neighborhood as well as in the definition of the search direction.

Recall (cf. Chapter 7 in [27]) that if $\operatorname{rank}(A) = m$, then 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

$$x_+ = x + \alpha\Delta x, \quad y_+ = y + \alpha\Delta y, \quad s_+ = s + \alpha\Delta s. \tag{36}$$

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 $\Psi(x^0 s^0, \mu^0) \leq \tau$ for $\mu^0 = 1$. This can be done without loss of generality (cf. [27]).

If, for the current iterates (x, y, s) and barrier parameter value μ the proximity $\Psi(xs, \mu)$ exceeds τ then we use one or more damped Newton steps to recenter; 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 $\tau > \nu_1^{-1}$;
an accuracy parameter $\varepsilon > 0$;
a fixed barrier update parameter θ, $0 < \theta < 1$;
(x^0, s^0) and $\mu^0 = 1$ such that $\Psi(x^0, s^0, \mu^0) \le \tau$.
begin
  $x := x^0$; $s := s^0$; $\mu := \mu^0$;
  while $n\mu \ge \varepsilon$ **do**
  begin
    $\mu := (1-\theta)\mu$;
    while $\Psi(x, s, \mu) \ge \tau$ **do**
    Solve the system (35) and compute $\Delta x, \Delta y, \Delta s$,
    begin
      Compute a step size α;
      $x := x + \alpha\Delta x$;
      $s := s + \alpha\Delta s$;
      $y := y + \alpha\Delta y$
    end
  end
end

Remark 3.4 *The damping parameter α has to be taken such that the proximity measure function Ψ decreases sufficiently. In the next section we determine a default value for α.*

Remark 3.5 *There are various choices for the parameter τ. If τ is a small constant independent of n and $\theta = \mathcal{O}(\frac{1}{\sqrt{n}})$, then the algorithm is called an IPM with small neighborhood which has the best known $\mathcal{O}\left(\sqrt{n}\log\frac{n}{\epsilon}\right)$ iteration bound. If τ is chosen as a number associated with to n, for instance $\tau = \mathcal{O}(n)$, then we call the algorithm an IPM with large neighborhood. For large neighborhood IPMs, the current best known iteration bound, based on a self-concordant function, is $\mathcal{O}\left(n\log\frac{n}{\varepsilon}\right)$. This is the bound we want to improve by using a* self-regular *proximity measure.*

Remark 3.6 *In the algorithm we always assume that $v_{\max} > 1$. This is because when $v_{\max} \le 1$, we can reduce the value of the proximity in the algorithm (or stay in a certain neighborhood of the central path) by appropriately reducing μ. In such case we even do not need to solve the Newton-type system.*

Remark 3.7 *The algorithm terminates with a point satisfying $n\mu \le \varepsilon$ and $\Psi(x, s, \mu) \le \tau$. By using (34), we obtain that*

$$x^T s \le n\mu + 2\mu\sqrt{\frac{2n\tau}{\nu_1}} + \mu\frac{2\tau}{\nu_1}.$$

Hence, if $\tau \le \mathcal{O}(n)$ which means that the algorithm works indeed in a large neighborhood of the central path, then the algorithm finally reports a feasible solution such that $x^T s \le \mathcal{O}(\varepsilon)$. For instance, if we choose the parameter $\tau = n$ and the proximity satisfying condition C.1 with $\nu_1 = 1$, then the algorithm will provide a solution satisfying $x^T s \le 7\varepsilon$.

3.3 Estimate of The Proximity after A Step

In this section, we will estimate the proximity after a strictly feasible step. When the solution of the system (35) or equivalently the search direction is available, usually we want to know how far we can go along the search direction while staying in the feasible region. The maximal feasible step size is helpful not only for the theoretical analysis of IPMs, but also for efficient implementation of IPMs [1]. A useful observation is that, since the displacements Δx and Δs are orthogonal, the scaled displacements d_x and d_s are orthogonal as well, i.e.,

$$d_x^T d_s = 0, \tag{37}$$

which further yields, by (26)

$$\sigma^2 = \|-\nabla\Psi(v)\|^2 = (d_x + d_s)^T (d_x + d_s) = \|d_x\|^2 + \|d_s\|^2. \tag{38}$$

Let us define

$$v_+ = \sqrt{\frac{x_+ s_+}{\mu}} = \sqrt{(v + \alpha d_x)(v + \alpha d_s)},$$

where x_+, s_+ are defined as in (36). It is easy to see that x_+ and s_+ are feasible if and only if $v + \alpha d_x$ and $v + \alpha d_s$ are both nonnegative. Note that from (8) it follows

$$v + \alpha d_x = v(e + \alpha \bar{d_x}), \quad v + \alpha d_s = v(e + \alpha \bar{d_s}).$$

Hence the maximal step size $\alpha_{\max}$ is essentially determined by the condition $e + \alpha_{\max} \bar{d_x} \geq 0$ and $e + \alpha_{\max} \bar{d_s} \geq 0$. Our first aim in this section is to estimate the norm of $(\bar{d_x}, \bar{d_s})$, since this provides a bound for $\alpha_{\max}$ because the maximal feasible step size $\alpha_{\max}$ certainly satisfies $1 \leq \alpha_{\max} \|(\bar{d_x}, \bar{d_s})\|$. We have

Lemma 3.8 *Let $\bar{d_x}, \bar{d_s}$ be defined by (8) and let*

$$\bar{\alpha} := v_{\min} \sigma^{-1}. \tag{39}$$

Then there holds

$$\|(\bar{d_x}, \bar{d_s})\| \leq \bar{\alpha}^{-1},$$

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

$$\alpha_{\max} \geq \bar{\alpha} \geq \sigma^{-1} \left(1 + \frac{q\sigma}{\nu_1}\right)^{-\frac{1}{q}}.$$

Proof: Observe that from (38) it follows

$$\|(d_x, d_s)\|^2 = \sigma^2.$$

By the definition of $(\bar{d_x}, \bar{d_s})$, we have

$$\|(\bar{d_x}, \bar{d_s})\| = \left\|(v^{-1} d_x, v^{-1} d_s)\right\| \leq \frac{1}{v_{\min}} \|(d_x, d_s)\| = \frac{\sigma}{v_{\min}} \leq \sigma \left(1 + \frac{q\sigma}{\nu_1}\right)^{\frac{1}{q}},$$

where the last inequality is derived from (28). This gives the first inequality in the lemma, from which the second inequality in the lemma can be easily derived. □

We proceed to estimate the decrease of the value of the proximity after one step. Naturally the proximity after one step is given by $\Psi(v_+)$. Let us define the gap of the proximity before and after one step as a function of α, i.e.,

$$f(\alpha) = \Psi(v_+) - \Psi(v). \tag{40}$$

From the definition of Ψ we derive, using C.2,

$$f(\alpha) = \sum_{i=1}^{n} (\psi(v_{+i}) - \psi(v_i)) \leq -\Psi(v) + \frac{1}{2}\sum_{i=1}^{n} (\psi(v_i + \alpha[d_x]_i) + \psi(v_i + \alpha[d_s]_i)) =: f_1(\alpha). \tag{41}$$

Obviously both the function $f(\alpha)$ and $f_1(\alpha)$ are twice continuously differentiable with respect to α if the step size α is feasible. To evaluate the decrease of the function $f(\alpha)$ after one feasible step, it suffices to estimate the value of the function $f_1(\alpha)$ for a feasible step size α. An important step in estimating the function $f_1(\alpha)$ is to bound its second derivative which is done in our next lemma.

Lemma 3.9 *Let the function $f_1(\alpha)$ be defined by (41) and let the parameter $\alpha \in [0, \bar{\alpha})$. Then one has*

$$f_1''(\alpha) \leq \frac{\nu_2\sigma^2}{2}\left((v_{\max} + \alpha\sigma)^{p-1} + (v_{\min} - \alpha\sigma)^{-q-1}\right). \tag{42}$$

Proof: From (41) it follows that

$$\begin{aligned}
f_1''(\alpha) &= \frac{1}{2}\sum_{i=1}^{n}\left(\psi''(v_i + \alpha[d_x]_i)[d_x]_i^2 + \psi''(v_i + \alpha[d_s]_i)[d_s]_i^2\right) \\
&\leq \frac{\nu_2}{2}\sum_{i=1}^{n}\left((v_i + \alpha[d_x]_i)^{p-1}[d_x]_i^2 + (v_i + \alpha[d_x]_i)^{-q-1}[d_x]_i^2\right) \\
&\quad + \frac{\nu_2}{2}\sum_{i=1}^{n}\left((v_i + \alpha[d_s]_i)^{p-1}[d_s]_i^2 + (v_i + \alpha[d_s]_i)^{-q-1}[d_s]_i^2\right) \\
&\leq \frac{\nu_2\|d_x\|^2}{2}\left((v_{\max} + \alpha\|d_x\|)^{p-1} + (v_{\min} - \alpha\|d_x\|)^{-1-q}\right) \\
&\quad + \frac{\nu_2\|d_s\|^2}{2}\left((v_{\max} + \alpha\|d_s\|)^{p-1} + (v_{\min} - \alpha\|d_s\|)^{-1-q}\right) \\
&\leq \frac{\nu_2\sigma^2}{2}\left((v_{\max} + \alpha\sigma)^{p-1} + (v_{\min} - \alpha\sigma)^{-1-q}\right),
\end{aligned}$$

where the first inequality is an immediate consequence from condition C.1, the second and third inequalities are given by the choice $p, q \geq 1$ and the following fact that for any $i \in \mathcal{I}$,

$$v_{\min} - \alpha\sigma \leq v_{\min} - \alpha\|d_x\| \leq v_i + \alpha[d_x]_i \leq v_{\max} + \alpha\|d_x\| \leq v_{\max} + \alpha\sigma; \tag{43}$$

$$v_{\min} - \alpha\sigma \leq v_{\min} - \alpha\|d_s\| \leq v_i + \alpha[d_s]_i \leq v_{\max} + \alpha\|d_s\| \leq v_{\max} + \alpha\sigma. \tag{44}$$

The proof of the lemma is done. □

A direct calculation gives

$$f_1(0) = 0; \quad f_1'(0) = f'(0) = -\frac{\sigma^2}{2}.$$

By using $f_1(\alpha) = f_1(0) + f_1'(0)\alpha + \int_0^\alpha \int_0^\xi f_1''(\zeta) d\zeta d\xi$ and (42), we deduce that for any $\alpha \in [0, \bar{\alpha})$, there holds

$$f_1(\alpha) \leq -\frac{\sigma^2}{2}\alpha + \frac{\nu_2 \sigma^2}{2} \int_0^\alpha \int_0^\xi \left((v_{\max} + \zeta\sigma)^{p-1} + (v_{\min} - \zeta\sigma)^{-q-1} \right) d\zeta d\xi =: f_2(\alpha). \tag{45}$$

It is easy to see, via making use of simple calculus, that $f_2(\alpha)$ is convex and twice differentiable for all $\alpha \in [0, \bar{\alpha})$. Let α^* denote the global minimizer of $f_2(\alpha)$ in the interval $[0, \bar{\alpha})$, that is

$$\alpha^* = \text{argmin}_{\alpha \in [0, \bar{\alpha})} f_2(\alpha). \tag{46}$$

There is no difficulty to check that $f_2'(0) < 0$ and $f_2'(\alpha)$ goes to infinity as α approaches $\bar{\alpha}$. From the convexity of $f_2(\alpha)$ in $[0, \bar{\alpha})$ one can easily deduce that α^* is the unique solution of the following equation

$$-\sigma + \frac{\nu_2}{p}\left((v_{\max} + \alpha^* \sigma)^p - v_{\max}^p \right) + \frac{\nu_2}{q}\left((v_{\min} - \alpha^* \sigma)^{-q} - v_{\min}^{-q} \right) = 0. \tag{47}$$

Note that the definition (39) implies that

$$\alpha^* < \bar{\alpha}.$$

In the sequel we will present a lower bound for the value of α^*. Before stating such a result, we insert here a technical lemma which is helpful in estimating α^*.

Lemma 3.10 *Suppose that $\beta \in [0, 1]$. Then there holds*

$$(1+t)^\beta \leq 1 + \beta t, \qquad \forall\, t \geq 0; \tag{48}$$

$$(1-t)^\beta \leq 1 - \beta t, \qquad \forall\, t \in [0, 1]. \tag{49}$$

Proof: The inequality (48) is true since for any fixed $\beta \leq 1$, the function $(1+t)^\beta - 1 - \beta t$ is decreasing for $t \geq 0$ and zero if $t = 0$. Similarly we observe that for any fixed $\beta \leq 1$, the function $(1-t)^\beta - 1 + \beta t$ is decreasing with respect to $t \in [0, 1]$ and zero if $t = 0$. This gives (49). □

Now we are ready to state one of our primary results in this section. This result is essential in establishing the complexity of the algorithm.

Lemma 3.11 *Let the constant α^* be defined by (46). Suppose that $\Psi(v) \geq \nu_1^{-1}$ and $v_{\max} > 1$. Let*

$$\nu_5 := \min\{\frac{\nu_1}{2\nu_1\nu_2 + p(\nu_1 + 2\nu_2)}, \frac{\nu_1^2}{(1+\nu_1)(2\nu_1\nu_2 + q(\nu_1 + 2\nu_2))}\}.$$

Then we have

$$\alpha^* \geq \nu_5 \sigma^{-\frac{q+1}{q}}. \tag{50}$$

In the special case where $\psi(t) = \Upsilon_{p,q}(t)$, the above bound (with $\nu_1 = \nu_2 = 1$) simplifies to

$$\alpha^* \geq \min\left(\frac{1}{3p+2}, \frac{1}{6q+4} \right) \sigma^{-\frac{q+1}{q}}. \tag{51}$$

Proof: See Appendix A.2. □

The following result concerning the minimal value of a specific convex function is useful in our estimation of $f(\alpha)$.

Lemma 3.12 *Suppose that $h(t)$ is a twice differentiable convex function with $h(0) = 0, h'(0) < 0$ and that $h(t)$ attains its global minimum at its stationary point $t^* > 0$. If $h''(t)$ is increasing with respect to t, then for any $t \in [0, t^*]$, there holds*

$$h(t) \leq \frac{h'(0)t}{2}.$$

Proof: Since $h(0) = 0$, we have

$$\begin{aligned} h(t) &= \int_0^t h'(\xi)d\xi = h'(0)t + \int_0^t \int_0^\xi h''(\zeta)d\zeta d\xi \leq h'(0)t + \int_0^t \xi h''(\xi)d\xi \\ &= h'(0)t + (\xi h'(\xi))|_0^t - \int_0^t h'(\xi)d\xi \leq h'(0)t - h(t), \end{aligned}$$

where the inequality is given the assumption that $h''(t)$ is nonnegative and increasing with respect to $t > 0$. The lemma follows directly from the above relation. □

One can easily verify that the function $f_2(\alpha)$ satisfies all the conditions in Lemma 3.12. Hence, at the point α^* defined by (46), since $f_2'(0) = f'(0)$, there holds

$$f(\alpha^*) \leq f_1(\alpha^*) \leq f_2(\alpha^*) \leq \frac{f'(0)}{2}\alpha^*. \tag{52}$$

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

Theorem 3.13 *Let the function $f(\alpha)$ be defined by (40) with $\Psi(v) \geq \nu_1^{-1}$. Then the step-size α given by $\alpha = \alpha^*$ or $\alpha = \nu_5 \sigma^{-\frac{q+1}{q}}$ is strictly feasible. Moreover there holds*

$$f(\alpha) \leq \frac{1}{2} f'(0)\alpha \leq -\frac{\nu_5 \nu_1^{\frac{q-1}{2q}}}{4} \Psi(v)^{\frac{q-1}{2q}}.$$

In the special case where $\psi(t) = \Upsilon_{p,q}(t)$, the above bound (with $\nu_1 = \nu_2 = 1$) simplifies to

$$f(\alpha) \leq -\min\left(\frac{1}{12p+8}, \frac{1}{24q+16}\right) \Psi(v)^{\frac{q-1}{2q}}.$$

Proof: It suffices to prove the first result of the theorem. First observe that for the given step size in the theorem, there holds trivially $\alpha \leq \alpha^*$ and thus it is strictly feasible. Further, making use of Lemma 3.12, together with (52) and (50) we conclude

$$f(\alpha) \leq -\frac{\nu_5}{4}\sigma^{\frac{q-1}{q}} \leq -\frac{\nu_5 \nu_1^{\frac{q-1}{2q}}}{4} \Psi(v)^{\frac{q-1}{2q}},$$

where the second inequality is implied by (27) in Proposition 3.3. This completes the proof of the theorem. □

3.4 Complexity of The Algorithm

In this subsection we derive an upper bound for the number of iterations of the algorithm when at each step the damping factor α is used as in Theorem 3.13. In this situation, each Newton step reduces the proximity by at least $-\frac{\nu_5}{4}\left(\nu_1 \Psi(v)\right)^{\frac{q-1}{2q}}$ which depends on the current proximity.

If the current point v is in the neighborhood defined in the algorithm, i.e., $\Psi(v) \le \tau$, then we need to update the parameter μ by $\mu := (1-\theta)\mu$ and consequently the new point in the v-space becomes $\hat{v} := \frac{v}{\sqrt{1-\theta}}$. Since the proximity after the update is defined by $\Psi(\hat{v})$, it might increase as well. Our next lemma provides an upper bound for $\Psi(\hat{v})$.

Lemma 3.14 *Suppose that $\Psi(v) \le \tau$ and $\hat{v} = \frac{v}{\sqrt{1-\theta}}$. Then one has*

$$\Psi(\hat{v}) \;\le\; \frac{\nu_2 \tau}{\nu_1 (1-\theta)^{\frac{p+1}{2}}} + \nu_2 \Upsilon'_{p,q}\left((1-\theta)^{-\frac{1}{2}}\right)\sqrt{\frac{2n\tau}{\nu_1(1-\theta)}} + n\nu_2 \Upsilon_{p,q}\left((1-\theta)^{-\frac{1}{2}}\right).$$

Proof: Replacing the parameter ϑ by $\frac{1}{\sqrt{1-\theta}}$ in inequality (31) of Proposition 3.3, one obtains the desired inequality. $\square$ For ease of reference, let us denote by $\psi_0(\theta,\tau,n)$ the right-hand side of the inequality in Lemma 3.14, i.e.,

$$\psi_0(\theta,\tau,n) := \frac{\nu_2 \tau}{\nu_1 (1-\theta)^{\frac{p+1}{2}}} + \nu_2 \Upsilon'_{p,q}\left((1-\theta)^{-\frac{1}{2}}\right)\sqrt{\frac{2n\tau}{\nu_1(1-\theta)}} + n\nu_2 \Upsilon_{p,q}\left((1-\theta)^{-\frac{1}{2}}\right). \quad (53)$$

It is trivial to see that if $\tau \le \mathcal{O}(n)$ and $\theta \in (0,1)$ is a constant independent of n, then one has $\psi_0(\theta,\tau,n) \le \mathcal{O}(n)$.

To estimate the total number of inner iterations in the algorithm, we also need the following result from [24]. For self-completeness, we provide a brief proof here as well.

Lemma 3.15 *Suppose that $t_0 > 0$ is a constant. Suppose $\{t_k > 0, k = 0,1,2,\cdots,K\}$ is a sequence satisfying the following inequalities*

$$t_{k+1} \le t_k - \beta t_k^{\gamma}, \quad k = 0,1,\cdots,K \quad (54)$$

with $\gamma \in [0,1)$ and $t_{K+1} < 0$. Then

$$K \le \left\lceil \frac{t_0^{1-\gamma}}{\beta(1-\gamma)} \right\rceil$$

and for any fixed $\rho \ge 0$ one has

$$t_{k+1} \le \rho, \quad \text{for all} \quad k \ge \left\lceil \frac{t_0^{1-\gamma} - \rho^{1-\gamma}}{\beta(1-\gamma)} \right\rceil.$$

Proof: First we note if $\beta \ge t_0^{1-\gamma}$, then 1 step is sufficient. Hence we can assume without loss of generality that $0 < \beta < t_0^{1-\gamma}$. Let us further assume that at the present step (k-th step) there holds $0 < \beta < t_k^{1-\gamma}$. Thus, by means of (54), we have

$$t_{k+1}^{1-\gamma} \le \left(t_k - \beta t_k^{\gamma}\right)^{1-\gamma} = t_k^{1-\gamma}\left(1 - \beta t_k^{\gamma-1}\right)^{1-\gamma} \le t_k^{1-\gamma}\left(1 - \beta(1-\gamma)t_k^{\gamma-1}\right) = t_k^{1-\gamma} - \beta(1-\gamma),$$

where the second inequality follows from (49). The lemma follows immediately from the above relation. $\square$

Combining Theorem 3.13, Lemma 3.14 and Lemma 3.15 together, we obtain the following estimation as an upper bound for the number of inner iterations.

Lemma 3.16 *Let $\Psi(xs,\mu) \le \tau$ and $\tau \ge \nu_1^{-1}$. Then after an update of the barrier parameter no more than*

$$\left\lceil \frac{8q\nu_1^{-\frac{q-1}{2q}}}{\nu_5(q+1)} \left(\psi_0(\theta,\tau,n)\right)^{\frac{q+1}{2q}} \right\rceil$$

iterations are needed to recenter. In the special case where $\psi(t) = \Upsilon_{p,q}(t)$, the above bound (with $\nu_1 = \nu_2 = 1$) simplifies to

$$\left\lceil \frac{8q \max\left(3p+2, 6q+4\right)}{q+1} \left(\psi_0(\theta,\tau,n)\right)^{\frac{q+1}{2q}} \right\rceil$$

inner iterations are needed to recenter.

It follows

Theorem 3.17 *If $\tau \ge \nu_1^{-1}$, the total number of iterations required by the primal-dual Newton algorithm is not more than*

$$\left\lceil \frac{8q\nu_1^{-\frac{q-1}{2q}}}{\nu_5(q+1)} \left(\psi_0(\theta,\tau,n)\right)^{\frac{1+q}{2q}} \right\rceil \left\lceil \frac{1}{\theta} \log \frac{n}{\varepsilon} \right\rceil.$$

In the special case where $\psi(t) = \Upsilon_{p,q}(t)$, the above bound (with $\nu_1 = \nu_2 = 1$) simplifies to

$$\left\lceil \frac{8q \max\left(3p+2, 6q+4\right)}{q+1} \left(\psi_0(\theta,\tau,n)\right)^{\frac{q+1}{2q}} \right\rceil \left\lceil \frac{1}{\theta} \log \frac{n}{\varepsilon} \right\rceil.$$

Proof: The number of barrier parameter updates is given by (cf. Lemma II.17, page 116, in [27])

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

Multiplication of this number by the bound for the number of inner iterations in Lemma 3.16 yields the theorem. □ For large-update IPMs, omitting the round off brackets in

Theorem 3.17 does not change the order of magnitude of the iteration bound. As we asserted earlier in this section, when the parameter τ is chosen such that $\tau \le \mathcal{O}(n)$, then there holds $\psi_0(\theta,\tau,n) \le \mathcal{O}(n)$. Hence we may safely consider the following expression as an upper bound for the number of iterations:

$$\mathcal{O}\left(n^{\frac{q+1}{2q}} \log \frac{n}{\varepsilon}\right).$$

One might notice that our iteration bound depends on some constant which might be very large. But recall that these constants depend on the choice of the function $\psi(t)$ (or p, q, ν_1, ν_2), and not on the data of the underlying problem. It is worthwhile to discuss some specific cases. For instance, if we choose

$$\psi(t) = \frac{1}{2}\left(t^2 - 1\right) + \frac{1}{q-1}\left(t^{1-q} - 1\right),$$

then the constants $\nu_1 = 1, \nu_2 = q$. In such a situation, the analysis can be simplified significantly. We also observe that for the above mentioned choice of $\psi(t)$, the search direction in this paper reduces to the same search direction introduced in [24], while the complexity in this paper is better than the one presented in [24] and the limitation $q \in [1,3]$ is removed as well.

Another interesting class of *self-regular* functions is $\Upsilon_{p,q}(t)$ defined by (13) with $p, q \geq 1$. The function $\Upsilon_{p,q}(t)$ satisfies C.1 and C.2 with $\nu_1 = \nu_2 = 1$. Let us consider a special case $p = 2,\ q = \log n$. In such circumstance, the last statement of Theorem 3.17 provides an upper bound for the number of iterations in the algorithm as

$$\mathcal{O}\left(\sqrt{n}\log n \log\frac{n}{\varepsilon}\right).$$

This gives to date the best complexity bound known for large-update methods. Note that the order of magnitude does not change if we take $\tau = n$, or $\tau = \mathcal{O}(n)$, which is practically more attractive and quite close to what is implemented in most of the existing IPM solvers [1]. We would like to remind the reader that theoretically we can also choose very big p. However, when we take $\tau = \mathcal{O}(n)$ which means the algorithm works indeed in a large neighborhood, then one should use only small parameter p (or as large as $p = \mathcal{O}(1)$) since otherwise after a large update of μ, the proximity $\Psi(v)$ might become much larger than $\mathcal{O}(n)$.

For the small-update method, if we choose θ sufficiently small such that

$$\frac{1}{\sqrt{1-\theta}} \leq 1 + \frac{\nu_3}{\sqrt{n}},$$

then by (33) we know that there exists a constant ν_9 such that

$$\psi_0(\theta, \tau, n) \leq \nu_9.$$

In such a circumstance, Theorem 3.17 shows that the small-update method still has a complexity as $\mathcal{O}\left(\sqrt{n}\log\frac{n}{\varepsilon}\right)$ iteration bound.

3.5 Relaxing The Requirements on The Proximity Function

At the end of this section, we consider the possibility of weakening the requirements on the function $\psi(t)$, or more concretely, the question whether we can build up the complexity of our algorithm based only on condition C.1. By restricting ourselves to the LO case, in the sequel we will present an affirmative answer to this question.

Now let us review how we have obtained the estimation of the proximity after one step. Let $f(\alpha), f_1(\alpha)$ be the functions defined by (40) and (41) respectively. The role of Condition C.2 is fundamental in establishing the relation (41) which enables us to bound the function $f(\alpha)$ from above by $f_1(\alpha)$. A crucial step in evaluating $f_1(\alpha)$ is to prove the inequality (42) for the second derivative $f_1''(\alpha)$ in Lemma 3.9. Since $f'(0) = f_1'(0)$, it is not difficult to see that if we can derive some inequality for the second derivative $f''(\alpha)$ analogous to (42), then in a similar vein we can estimate the proximity after one step (or equivalently $f(\alpha)$) directly without requiring the relation (41). In such a situation condition C.2 becomes superfluous. In what follows we give such an estimation for $f''(\alpha)$.

Lemma 3.18 *Let the function $f(\alpha)$ be defined by (40) and $\psi(t) \in \Omega_1$. Then for any $\alpha \in [0, \frac{1}{2}\bar{\alpha})$, we have*

$$f''(\alpha) \leq \frac{3\nu_2\sigma^2}{2}\left(1 + \frac{\nu_2}{\nu_1}\right)\left((v_{\max} + \alpha\sigma)^{p-1} + (v_{\min} - \alpha\sigma)^{-q-1}\right). \tag{55}$$

Particularly, if $\psi(t) \in \Omega_1$ with parameter $p \geq 3$, then for any $\alpha \in [0, \bar{\alpha})$

$$f''(\alpha) \leq \frac{\nu_2\sigma^2}{2}\left(1 + \frac{\nu_2}{\nu_1}\right)\left((v_{\max} + \alpha\sigma)^{p-1} + (v_{\min} - \alpha\sigma)^{-q-1}\right) \tag{56}$$

holds

Proof: See Appendix A.2. □

The above lemma implies that, at least for linear optimization, condition C.2 is not necessary. We can use any function in Ω_1 to define the proximity and subsequently the search direction. By using the above two technical results, and following analogous arguments as done in Section 3.3-4, for large update IPMs one can get the following expression as an upper bound for the number of iterations

$$\mathcal{O}\left(n^{\frac{q+1}{2q}}\log\frac{n}{\varepsilon}\right) \quad \text{or} \quad \mathcal{O}\left(\sqrt{n}\log n\log\frac{n}{\varepsilon}\right).$$

The details are left to the interested readers.

4 New Primal-Dual Algorithms for Semidefinite Optimization

In this section we undertake the task of generalizing the algorithm posed in the previous section to the case of SDO. The section consists of two parts. In the first part we introduce the new search direction based on the *self-regular* proximity for SDO and describe the algorithm. In the second half we establish the polynomial complexity of the algorithm.

4.1 Self-Regular Proximities and New Search Directions for SDO

We consider the SDO problem in the following standard form:

$$\text{(SDO)} \qquad \begin{aligned} &\min \mathbf{Tr}\,(CX) \\ &\mathbf{Tr}\,(A_jX) = b_j, \quad j = 1,\cdots,m, \quad X \succeq 0, \end{aligned}$$

and its dual problem

$$\text{(SDD)} \qquad \begin{aligned} &\max b^Ty \\ &\sum_{j=1}^{m} y_jA_j + S = C, \quad S \succeq 0. \end{aligned}$$

Here C and A_j $(1 \le j \le m)$ are symmetric $n \times n$ matrices, and $b, y \in \Re^m$. 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 and thus we can further assume that both X and S are diagonal. The concept of the central path can also be extended to SDO. We assume without loss of generality that both SDO and its dual SDD are strictly feasible (this can be done via the so-called self-dual embedding technique [4]). The central path for SDO is defined by the solution sets $\{X(\mu), y(\mu), S(\mu) : \mu > 0\}$ of the following system

$$\begin{cases} \mathbf{Tr}\,(A_jX) &= b_j, \quad j = 1,\ldots,m, \\ \sum_{j=1}^{m} y_jA_j + S &= C, \\ XS &= \mu E, \qquad X, S \succeq 0, \end{cases} \tag{57}$$

where E denotes the $n \times n$ identity matrix and $\mu > 0$. The basic idea of IPMs is to follow this central path and approximate the optimal set of SDO as μ goes to zero. Suppose the point (X, y, S) is strictly feasible. Newton's method amounts to linearizing the system (57), thus yielding the following equation

$$\begin{cases} \mathbf{Tr}\,(A_j \Delta X) & = & 0, & j = 1, \cdots, m, \\ \Delta S + \sum_{j=1}^{m} \Delta y_j A_j & = & 0 & \\ X\Delta S + \Delta X S & = \mu E - XS & & \end{cases} \tag{58}$$

A crucial observation for SDO is that the above Newton system might not have a symmetric solution ΔX. Many researchers have proposed several different ways of symmetrizing the third equation in the Newton system so that the new system have a unique symmetric solution [31, 32]. In this paper we consider the symmetrization scheme from which the NT direction [22, 32] is derived. Let

$$P = X^{\frac{1}{2}}(X^{\frac{1}{2}}SX^{\frac{1}{2}})^{-\frac{1}{2}}X^{\frac{1}{2}} = S^{-\frac{1}{2}}(S^{\frac{1}{2}}XS^{\frac{1}{2}})^{\frac{1}{2}}S^{-\frac{1}{2}}, \tag{59}$$

and $D = P^{\frac{1}{2}}$, where for any symmetric positive definite matrix G, the exponent $G^{\frac{1}{2}}$ denotes its symmetric square root. The matrix D can be used to rescale X and S to the same matrix V defined by [4, 30]

$$V := \frac{1}{\sqrt{\mu}} D^{-1}XD^{-1} = \frac{1}{\sqrt{\mu}} DSD. \tag{60}$$

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

Now let us recall the definition of a matrix function.

Definition 4.1 *[2]* [6] *Suppose the matrix X is diagonalizable with*

$$X = Q_X^{-1} \text{diag}\,(\lambda_1(X), \lambda_2(X), \cdots, \lambda_n(X))\, Q_X,$$

where Q_X is nonsingular and $h(t)$ is a function from $\Re$ into itself. The function $h(X)$ is defined by

$$h(X) = Q_X^{-1} \text{diag}\,(h(\lambda_1(X)), h(\lambda_2(X)), \cdots, h(\lambda_n(X)))\, Q_X. \tag{61}$$

Particularly if X is symmetric, then we can choose Q_X to be orthogonal, i.e., $Q_X^{-1} = Q_X^T$.

Remark 4.2 *In the rest of this work, when we use the function $h(\cdot)$ (or $\psi(\cdot)$) and its derivative function $h'(\cdot)$ (or $\psi'(\cdot)$) without any specification, it always denotes a matrix function if the argument is a matrix and it means a general function from $\Re$ to $\Re$ if the argument is also in $\Re$. It should be noticed that the matrix Q_X might not be unique.* For any matrix G whose eigenvalues are real[7], let us define

$$\Psi(G) = \sum_{i=1}^{n} \psi(\lambda_i(G)) = \mathbf{Tr}\,(\psi(G)), \tag{62}$$

where $\psi(G)$ is defined by (61).

We next extend the notion of self-regularity to the case of matrix functions.

Definition 4.3 *Suppose that X is symmetric positive definite and that the functions $\psi(X)$ and $\Psi(X)$ are defined by (61) and (62), respectively. The function $\psi(X)$ is said to be* self-regular *if the functions $\psi(X)$ and $\Psi(X)$ satisfy the following conditions.*

[6] The matrix Q_X is not unique, but $h(X)$ is well defined whenever the scalar function $h(t)$ is well defined (see [2], page 90).

[7] G might be not symmetric positive definite.

C.3 $\Psi(X)$ *is strictly convex with respect to* $X \succ 0$ *and vanishes at its global minimal point* $X = E$, *i.e.,* $\Psi(E) = 0, \psi'(E) = 0_{n\times n}$. *Further, there exist two positive constants* $\nu_2 > \nu_1 > 0$ *and* $p, q \geq 1$ *such that*

$$\nu_1(X^{p-1} + X^{-1-q}) \preceq \psi''(X) \preceq \nu_2(X^{p-1} + X^{-1-q}), \quad \forall X \succ 0. \tag{63}$$

C.4 For any $X_1, X_2 \succ 0$,

$$\Psi\left([X_1^{\frac{1}{2}} X_2 X_1^{\frac{1}{2}}]^{\frac{1}{2}}\right) \leq \frac{1}{2}\left(\Psi(X_1) + \Psi(X_2)\right). \tag{64}$$

A very important observation is that if the kernel function $\psi(t)$ defining the matrix function $\psi(X)$ is *self-regular*, so is $\psi(X)$. This is proven in our next proposition.

Proposition 4.4 *Let the matrix functions* $\Psi(X), \psi(X)$ *be defined by (65) and (61) respectively. If* $\psi(t)$ *is self-regular, then so is* $\psi(X)$.

Proof: See Appendix A.3. □

The proximity measure we suggest here for SDO is defined by

$$\Psi(X, S, \mu) := \Psi(V) = \mathbf{Tr}\left(\psi(V)\right). \tag{65}$$

Consequently we say that the proximity $\Psi(V)$ is a *self-regular* proximity if the function $\psi(V)$ is *self-regular*. In the sequel we progress to characterize the *self-regular* proximities for SDO. For notational brevity, we also define

$$\sigma^2 = \mathbf{Tr}\left(\psi'(V)^2\right) = \left\|\psi'(V)\right\|^2. \tag{66}$$

Replacing $v_i, v_{\max}$ and $v_{\min}$ used in the case of LO by $\lambda_i(V), \lambda_{\max}(V)$ and $\lambda_{\min}(V)$, respectively if necessary, and by following a similar chain of reasoning as in the proof of Proposition 3.3, one can prove the following result. Those conclusions can also be viewed as common features of general *self-regular* functions in the cone of positive definite matrices.

Proposition 4.5 *Let the proximity* $\Psi(V)$ *be defined by (65) and* σ *by (66). If the function* $\Psi(\cdot)$ *satisfies condition C.3, then there holds*

$$\Psi(V) \leq \frac{\sigma^2}{2\nu_1}, \tag{67}$$

$$\lambda_{\min}(V) \geq \left(1 + \frac{q\sigma}{\nu_1}\right)^{-\frac{1}{q}}, \tag{68}$$

and

$$\lambda_{\max}(V) \leq \left(1 + \frac{p\sigma}{\nu_1}\right)^{\frac{1}{p}}. \tag{69}$$

If $\lambda_{\max}(V) > 1$ *and* $\lambda_{\min}(V) < 1$, *then*

$$\sigma \geq \nu_1 \left(\frac{(\lambda_{\max}(V)^p - 1)^2}{p^2} + \frac{(\lambda_{\min}(V)^{-q} - 1)^2}{q^2}\right)^{\frac{1}{2}}. \tag{70}$$

Bearing in mind that our primary target is to drive the duality gap to zero. With respect to this point, it is desirable for the proximity measure used in the algorithm to provide certain guarantee for the duality gap or the scaled duality gap $\|V\|^2$ in the scaled V-space. In the sequel we consider the relationships between the scaled duality gap and the proximity. In case of SDO, by substituting v_i by $\lambda_i(V)$ if necessary and following a similar vein as the proof of Lemma 2.12, we can easily deduce

$$\frac{\Psi(V)}{\nu_1} \geq \frac{1}{2}\sum_{i=1}^{n}(\lambda_i(V)-1)^2 = \frac{1}{2}\|V\|^2 - \sum_{i=1}^{n}\lambda_i(V) + \frac{n}{2} \geq \frac{1}{2}\|V\|^2 - \sqrt{n}\|V\| + \frac{n}{2}.$$

This reation gives

$$\|V\| \leq \sqrt{n} + \sqrt{\frac{2\Psi(V)}{\nu_1}},$$

which further yields

$$\mathbf{Tr}\,(XS) = \mu\|V\|^2 \leq n\mu + 2\mu\sqrt{\frac{2n\Psi(V)}{\nu_1}} + \frac{2\Psi(V)}{\nu_1}\mu. \tag{71}$$

The above inequality shows it clearly that the proximity acts indeed as a potential function for minimizing the duality gap.

Now let us move onto our discussion about search directions for SDO. Let us further define

$$\begin{aligned} \bar{A}_i &:= DA_iD; \qquad i = 1, \cdots, m; \\ D_X &:= \frac{1}{\sqrt{\mu}}D^{-1}\Delta X D^{-1}, \qquad D_S := \tfrac{1}{\sqrt{\mu}}D\Delta S D. \end{aligned} \tag{72}$$

Then the NT search direction can be written as the solution of the following system [22, 30]

$$\begin{cases} \mathbf{Tr}\,(\bar{A}_j D_X) = 0, \qquad j = 1, \cdots, m, \\ D_S + \sum_{j=1}^{m}\Delta y_i \bar{A}_j = 0 \\ D_X + D_S = V^{-1} - V \end{cases} \tag{73}$$

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

$$\begin{cases} \mathbf{Tr}\,(\bar{A}_j D_X) = 0, \qquad j = 1, \cdots, m, \\ D_S + \sum_{j=1}^{m}\Delta y_j \bar{A}_j = 0 \\ D_X + D_S = -\psi'(V) \end{cases} \tag{74}$$

Having D_X and D_S, ΔX and ΔS can be calculated from (72). Due to the orthogonality of ΔX and ΔS, it is trivial to see that

$$\mathbf{Tr}\,(D_X D_S) = \mathbf{Tr}\,(D_S D_X) = 0. \tag{75}$$

Our new algorithm for SDO can be described as follows.

Large-Update Primal-Dual Algorithm for SDO

Input:
  A proximity parameter $\tau \geq \nu_1^{-1}$;
  an accuracy parameter $\varepsilon > 0$;
  a fixed barrier update parameter $\theta \in (0,1)$;
  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 $\Psi(X, S, \mu) \geq \tau$ **do**
    Solve the system (74),
    begin
      Compute a step size α;
      $X := X + \alpha \Delta X$;
      $S := S + \alpha \Delta S$;
      $y := y + \alpha \Delta y$;
    end
  end
end

Remark 4.6 *The algorithm terminates with a point satisfying $n\mu \leq \varepsilon$ and $\Psi(X, S, \mu) \leq \tau$. By using (71), we obtain that*

$$\mathbf{Tr}\,(XS) \leq n\mu + 2\mu\sqrt{\frac{2n\tau}{\nu_1}} + \mu\frac{2\tau}{\nu_1}.$$

Hence, if $\tau \leq \mathcal{O}(n)$ which means that the algorithm works indeed in a large neighborhood of the central path, then the algorithm finally reports a feasible solution such that $\mathbf{Tr}\,(XS) \leq \mathcal{O}(\varepsilon)$.

We close this subsection by presenting a result about the proximity $\Psi(V)$ under condition C.4. This is the analogue of the result presented in Proposition 3.2.

Proposition 4.7 *Suppose $X, S \succ 0$ and V is the scaled matrix defined by (60). If the function $\Psi(\cdot)$ defined by (62) satisfies condition C.4, then there holds*

$$\Psi(V) \leq \frac{1}{2}\left(\Psi\left(\frac{X}{\sqrt{\mu}}\right) + \Psi\left(\frac{S}{\sqrt{\mu}}\right)\right).$$

Proof: By (60) we have

$$V = \frac{1}{\sqrt{\mu}}\left(D^{-1}XSD\right)^{\frac{1}{2}}.$$

Since the matrix $D^{-1}XSD$ is symmetric positive definite and has the same eigenvalues as the matrix XS, it is unitarily similar to the matrix $X^{\frac{1}{2}}SX^{\frac{1}{2}}$. Hence from (62) and condition C.4

we get

$$\Psi(V) = \Psi\left(\frac{1}{\sqrt{\mu}}\left(X^{\frac{1}{2}}SX^{\frac{1}{2}}\right)^{\frac{1}{2}}\right) \le \frac{1}{2}\left(\Psi\left(\frac{X}{\sqrt{\mu}}\right) + \Psi\left(\frac{S}{\sqrt{\mu}}\right)\right).$$

□

4.2 Complexity of the algorithm

As we have seen in the LO case, the key in evaluating the complexity of the algorithm, is to calculate the decrease of the proximity after one feasible step. To give such an estimation, we need to obtain some bounds for the derivatives of a specific function of the step size. In case of SDO, this requires to use the derivatives of a function involving matrix functions. We next present a result about the derivatives of a function given by the trace of a matrix function. This result will be used repeatedly in our later analysis. To continue our analysis, we need another concept relevant to matrix functions in matrix theory. This is the notion of matrices of functions (or matrix-valued functions) [3].

Definition 4.8 *A matrix $X(t)$ is said to be a matrix of functions (or matrix-valued function) if each entry of $X(t)$ is a function of t, i.e., $X(t) = [X_{ij}(t)]$.*

The usual concepts of continuity, differentiability and integrability have been extended to matrix-valued functions of a scalar by interpreting them entrywisely. Thus we have

$$\frac{d}{dt}X(t) = [\frac{d}{dt}X_{ij}(t)].$$

For simplification of expression, in the rest of this section we denote by $X'(t)$ the gradient of the matrix-valued function $X(t)$. The following relations between matrix-valued functions and their derivatives [3](page 490-491) are immediate consequences of the well known sum and product rule in calculus. These relations are very helpful in the proof of our next lemma. Suppose that the matrix-valued functions $G(t), H(t)$ are differentiable and nonsingular at the point t. Then we have

$$\frac{d}{dt}\mathbf{Tr}\,(G(t)) = \mathbf{Tr}\left(\frac{d}{dt}G(t)\right) = \mathbf{Tr}\,(G'(t))\,, \tag{76}$$

$$\frac{d}{dt}\,(G(t)H(t)) = [\frac{d}{dt}G(t)]H(t) + G(t)[\frac{d}{dt}H(t)] = G'(t)H(t) + G(t)H'(t). \tag{77}$$

For any function $\psi(t)$, let us denote by $\Delta\psi$ the divided difference of $\psi(t)$ such that

$$\Delta\psi(t_1, t_2) = \frac{\psi(t_1) - \psi(t_2)}{t_1 - t_2}, \quad \forall t_1 \ne t_2 \in \Re,$$

and in particular when $t = t_1 = t_2$, there holds $\Delta\psi(t, t) = \psi'(t)$. Now we have

Lemma 4.9 *Suppose that $H(t)$ is a matrix of functions such that the matrix $H(t)$ is positive definite with eigenvalues $\lambda_1(t) \ge \lambda_2(t) \ge \cdots \ge \lambda_n(t) > 0$. If $H(t)$ is twice differentiable with respect to t for all $t \in (l_t, u_t)$ and $\psi(t)$ is also a twice continuously differentiable function in a suitable domain which contains all the eigenvalues of $H(t)$, then*

$$\frac{d}{dt}\mathbf{Tr}\,(\psi(H(t))) = \mathbf{Tr}\,(\psi'(H(t))H'(t))\,, \quad t \in (l_t, u_t),$$

and

$$\frac{d^2}{dt^2}\mathbf{Tr}\left(\psi(H(t))\right) \le \varpi \left\|H'(t)\right\|^2 + \mathbf{Tr}\left(\psi'(H(t))H''(t)\right), \tag{78}$$

where

$$\varpi = \max\{\left|\Delta\psi'(\lambda_j(t), \lambda_k(t))\right|, j, k = 1, 2, \cdots, n\} \tag{79}$$

is a number depending on $H(t)$ and $\psi(t)$ with

$$\Delta\psi'(t_1, t_2) = \frac{\psi'(t_1) - \psi'(t_2)}{t_1 - t_2}, \quad \forall t_1, t_2 \in [l_t, u_t].$$

Proof: See Appendix A.3. □

We progress to estimate the decrement of the proximity after one step. Let us denote by V_+ the scaled matrix defined by (59) and (60) where the matrices X and S are replaced by $X_+ = X + \alpha\Delta X, S_+ = S + \alpha\Delta S$ respectively. It is trivial to verify that V_+^2 is unitarily similar to the matrix $X_+^{\frac{1}{2}} S_+ X_+^{\frac{1}{2}}$ and thus $(V + \alpha D_X)^{\frac{1}{2}} (V + \alpha D_S) (V + \alpha D_X)^{\frac{1}{2}}$. This further implies that the eigenvalues of the matrix V_+ are precisely the same as that of the matrix

$$\tilde{V}_+ := \left((V + \alpha D_X)^{\frac{1}{2}} (V + \alpha D_S) (V + \alpha D_X)^{\frac{1}{2}}\right)^{\frac{1}{2}}. \tag{80}$$

Since the proximity after one step is defined by $\Psi(V_+)$, from (65) it follows immediately

$$\Psi(V_+) = \Psi\left(\tilde{V}_+\right). \tag{81}$$

Again we define the gap between the proximity before and after one step as a function of the step size, that is

$$f(\alpha) := \Psi(V_+) - \Psi(V) = \Psi(\tilde{V}_+) - \Psi(V). \tag{82}$$

Our goal is to estimate the decrement of $f(\alpha)$ for a feasible step size α. The first thing we want to know is how far we can go along the search direction, or equivalently how large is the maximal step size. Let

$$\bar{D}_x = V^{-\frac{1}{2}} D_X V^{-\frac{1}{2}}, \quad \bar{D}_s = V^{-\frac{1}{2}} D_S V^{-\frac{1}{2}}. \tag{83}$$

Since

$$V + \alpha D_X = V^{\frac{1}{2}}(E + \alpha\bar{D}_x)V^{\frac{1}{2}}, \quad V + \alpha D_S = V^{\frac{1}{2}}(E + \alpha\bar{D}_s)V^{\frac{1}{2}},$$

the matrices $V + \alpha D_X$ and $V + \alpha D_S$ are positive definite if and only if the matrices $E + \alpha\bar{D}_x$ and $E + \alpha\bar{D}_s$ are positive definite. Thus the maximal feasible step size is dependent on the eigenvalues of matrices $\bar{D}_x$ and $\bar{D}_s$. Our next result gives some estimations about the norms of the matrices $\bar{D}_x$ and $\bar{D}_s$, and the maximal feasible step size $\alpha_{\max}$ as well.

Lemma 4.10 *Let the matrices $\bar{D}_x$ and $\bar{D}_s$ are defined by (83) and let $\alpha_{\max}$ be the maximal feasible step size. Let*

$$\bar{\alpha} := \lambda_{\min}(V)\sigma^{-1}. \tag{84}$$

Then we have

$$\left\|\bar{D}_x\right\|^2 + \left\|\bar{D}_s\right\|^2 \le \bar{\alpha}^{-2},$$

and

$$\alpha_{\max} \ge \bar{\alpha} \ge \sigma^{-1}\left(1 + \frac{q\sigma}{\nu_1}\right)^{-\frac{1}{q}}.$$

Proof: First we quote an inequality about the singular values and eigenvalues of matrices from [3](3.3.20,page 178). Suppose that the matrices G, H are symmetric. For any $i \in \mathcal{I}$, by using the notations introduced at the end of the introduction in this paper, one gets

$$\varrho_i(GH) \le \min\left(\varrho_i(G)\varrho_1(H), \varrho_1(G)\varrho_i(H)\right) = \min\left(|\lambda_1(G)\lambda_i(H)|, |\lambda_1(H)\lambda_i(G)|\right). \tag{85}$$

It follows immediately

$$|\lambda_i(\bar{D}_x)| \le \frac{1}{\lambda_{\min}(V)^{\frac{1}{2}}}\varrho_i(D_X V^{-\frac{1}{2}}) \le \frac{1}{\lambda_{\min}(V)}|\lambda_i(D_X)|, \quad i \in \mathcal{I};$$

and

$$|\lambda_i(\bar{D}_s)| \le \frac{1}{\lambda_{\min}(V)^{\frac{1}{2}}}\varrho_i(D_S V^{-\frac{1}{2}}) \le \frac{1}{\lambda_{\min}(V)}|\lambda_i(D_S)|, \quad i \in \mathcal{I}.$$

The above two relations mean

$$\begin{aligned}
\|\bar{D}_x\|^2 + \|\bar{D}_s\|^2 &= \sum_{i=1}^{n}\left(\lambda_i^2(\bar{D}_x) + \lambda_i^2(\bar{D}_s)\right) \le \frac{1}{\lambda_{\min}(V)^2}\sum_{i=1}^{n}\left(\lambda_i^2(D_X) + \lambda_i^2(D_X)\right) \\
&= \frac{1}{\lambda_{\min}(V)^2}\left(\|D_X\|^2 + \|D_S\|^2\right) = \frac{1}{\lambda_{\min}(V)^2}\left(\|D_X + D_S\|^2\right) \\
&= \bar{\alpha}^2 \le \sigma^2\left(1 + \frac{q\sigma}{\nu_1}\right)^{\frac{2}{q}},
\end{aligned}$$

where the third equality given by (75), and the last inequality follows from the definition of σ and (68). The second statement of the lemma is an immediate consequence of the first result of the lemma. □

Now we are going to bound the function $f(\alpha)$, given by (82). By using condition C.4 one gets

$$f(\alpha) = \Psi(\tilde{V}_+) - \Psi(V) \le \frac{1}{2}\left(\Psi\left(V + \alpha D_X\right) + \Psi(V + \alpha D_S)\right) - \Psi(V) =: f_1(\alpha).$$

Hence it suffices for us to estimate the decrease of the value of the function $f_1(\alpha)$ after one step. The main difficulty in the estimation of the function $f_1(\alpha)$ is to evaluate its first and second derivatives. From Lemma 4.9 we get

$$f_1'(\alpha) = \frac{1}{2}\mathbf{Tr}\left(\psi'(V + \alpha D_X)D_X + \psi'(V + \alpha D_S)D_S\right), \tag{86}$$

and by definition

$$f_1''(\alpha) = \frac{1}{2}\frac{d^2}{d\alpha^2}\mathbf{Tr}\left(\psi(V + \alpha D_X) + \psi(V + \alpha D_S)\right).$$

Now we are ready to state one of our main results in this section.

Lemma 4.11 *If the step size $\alpha < \bar{\alpha}$, then there holds*

$$f_1''(\alpha) \le \frac{\nu_2\sigma^2}{2}\left((\lambda_{\max}(V) + \alpha\sigma)^{p-1} + (\lambda_{\min}(V) - \alpha\sigma)^{-q-1}\right).$$

Proof: See Appendix A.3. □

The rest of this section follows the procedure in Section 3.4 step by step. Thus in the sequel we only outline the basic schedule and omit most of the detailed proofs. Note that by applying Lemma 4.9 to the function $f(\alpha)$ one easily gets

$$f'(0) = f_1'(0) = -\frac{\sigma^2}{2}.$$

Hence, since $f(0) = f_1(0) = 0$, from Lemma 4.11 we conclude

$$f(\alpha) \leq f_1(\alpha) \leq -\frac{\sigma^2\alpha}{2} + \frac{\nu_2\sigma^2}{2}\int_0^\alpha \int_0^\xi \left((\lambda_{\max}(V)+\zeta\sigma)^{p-1} + (\lambda_{\min}(V)-\zeta\sigma)^{-q-1}\right) d\zeta d\xi,$$

which is essentially the same as its LO analogue (45) where the variables $v_{\max}, v_{\min}$ are replaced by $\lambda_{\max}(V), \lambda_{\min}(V)$, respectively. Let

$$f_2(\alpha) := -\frac{\sigma^2\alpha}{2} + \frac{\nu_2\sigma^2}{2}\int_0^\alpha \int_0^\xi \left((\lambda_{\max}(V)+\zeta\sigma)^{p-1} + (\lambda_{\min}(V)-\zeta\sigma)^{-q-1}\right) d\zeta d\xi.$$

Obviously, $f_2(\alpha)$ is convex and twice differentiable for all $\alpha \in [0, \bar{\alpha})$. It is also easy to see that $f_2(\alpha)$ is decreasing at zero and goes to infinity as $\alpha \to \bar{\alpha}$. Let α^* be the point at which $f_2(\alpha)$ attains its global minimal value, i.e.,

$$\alpha^* = \text{argmin}_{\alpha \in [0,\bar{\alpha})} f_2(\alpha). \tag{87}$$

For this α^*, by means of Lemma 3.12, we obtain

$$f(\alpha^*) \leq f_2(\alpha^*) \leq \frac{1}{2} f_2'(0)\alpha^* = \frac{1}{2} f'(0)\alpha^*. \tag{88}$$

On the other hand, replacing $v_{\max}$ and $v_{\min}$ by $\lambda_{\max}(V), \lambda_{\min}(V)$ respectively if necessary and proceeding similarly as in the proof of Lemma 3.11, one can get the following lower bound for α^*

$$\alpha^* \geq \nu_5 \sigma^{-\frac{q+1}{q}},$$

where ν_5 is a positive constant defined in Lemma 3.11. Thus, if the step-size α^* or $\alpha = \nu_5\sigma^{-\frac{q+1}{q}}$ is employed in the algorithm, then after one Newton step we get

Theorem 4.12 *Let the function $f(\alpha)$ be defined by (82) with $\Psi(V) \geq \nu_1^{-1}$. Then the step-size α given by $\alpha = \alpha^*$ (87) or $\alpha = \nu_5\sigma^{-\frac{q+1}{q}}$ is strictly feasible. Moreover there holds*

$$f(\alpha) \leq \frac{1}{2} f'(0)\alpha \leq -\frac{\nu_5 \nu_1^{\frac{q-1}{2q}}}{4} \Psi(V)^{\frac{q-1}{2q}}.$$

In the special case where $\psi(t) = \Upsilon_{p,q}(t)$, the above bound (with $\nu_1 = \nu_2 = 1$) simplifies to

$$f(\alpha) \leq -\min\left(\frac{1}{12p+8}, \frac{1}{24q+16}\right) \Psi(V)^{\frac{q-1}{2q}}.$$

Proof: The proof is a copy of that for its LO analogue Theorem 3.13, and thus the details are omitted here. □

Since the proximity $\Psi(V)$ is determined by the eigenvalues of the matrix V, the growth behavior of the proximity $\Psi(V)$ is precisely the same as its LO counterpart $\Psi(v)$. If the current point enters the neighborhood again, then we update μ to $(1-\theta)\mu$ for some $\theta \in (0,1)$. Progressing along the way as already done for the LO case, one can show that after the update of μ, the proximity is still bounded above by the number $\psi_0(\theta, \tau, n)$ defined analogously as in (53). It follows immediately

Lemma 4.13 *Let $\Psi(X,S,\mu) \le \tau$ and $\tau \ge \nu_1^{-1}$. Then after an update of the barrier parameter no more than*

$$\left\lceil \frac{8q\nu_1^{-\frac{q-1}{2q}}}{\nu_5(q+1)} \left(\psi_0(\theta,\tau,n)\right)^{\frac{q+1}{2q}} \right\rceil$$

iterations are needed to recenter. In the special case where $\psi(t) = \Upsilon_{p,q}(t)$, the above bound (with $\nu_1 = \nu_2 = 1$) simplifies to

$$\left\lceil \frac{8q \max(3p+2, 6q+4)}{q+1} \left(\psi_0(\theta,\tau,n)\right)^{\frac{q+1}{2q}} \right\rceil.$$

Now we can state the complexity of the algorithm:

Theorem 4.14 *If $\tau \ge \nu_1^{-1}$, the total number of iterations required by the primal-dual Newton algorithm is not more than*

$$\left\lceil \frac{8q\nu_1^{-\frac{q-1}{2q}}}{\nu_5(q+1)} \left(\psi_0(\theta,\tau,n)\right)^{\frac{1+q}{2q}} \right\rceil \left\lceil \frac{1}{\theta} \log \frac{n}{\varepsilon} \right\rceil.$$

In the special case where $\psi(t) = \Upsilon_{p,q}(t)$, the above bound (with $\nu_1 = \nu_2 = 1$) simplifies to

$$\left\lceil \frac{8q \max(3p+2, 6q+4)}{q+1} \left(\psi_0(\theta,\tau,n)\right)^{\frac{q+1}{2q}} \right\rceil \left\lceil \frac{1}{\theta} \log \frac{n}{\varepsilon} \right\rceil.$$

Omitting the round off brackets in Theorem 4.14, one can conclude that by choosing $\theta \in (0,1)$ and suitable $p, q \ge 1$, the complexity of our algorithm with large-update for SDO is in the order of $\mathcal{O}\left(n^{\frac{q+1}{2q}} \log \frac{n}{\varepsilon}\right)$ iterations bound, while the algorithm with small-update ($\theta = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$) still has a complexity $\mathcal{O}(\sqrt{n} \log \frac{n}{\varepsilon})$ iterations bound. Moreover, by means of Theorem 4.14, one can readily verify that if p is a constant and $q = \log n$, then the new large-update algorithm has a complexity $\mathcal{O}\left(\sqrt{n} \log n \log \frac{n}{\varepsilon}\right)$ iterations bound.

Finally we remark that, as we observed in Section 3.5, the condition C.2 can be waived without damaging the complexity of the algorithm for LO. It might also be possible to remove condition C.4 for semidefinite optimization problem[8]. However, in such situation, the analysis will become much more complicated. On the other hand, it is impossible to improve the complexity of the algorithm when allowing such relaxation. Hence we rest on our simple unified framework and leave the exploration of such possibilities to the interested readers.

5 Concluding Remarks

A new class of proximities for path-following methods for solving LO and SDO was proposed. These proximities are derived from so-called *self-regular* functions which satisfy certain conditions. New search directions based on the proximity were suggested for solving the considered LO and SDO problems. By exploring the properties of *self-regular* functions, we were able to prove that, under a unified framework, large-update IPMs based on the new directions have polynomial $\mathcal{O}\left(n^{\frac{q+1}{2q}} \log \frac{n}{\varepsilon}\right)$ iteration bound, where $q \ge 1$ is the so-called *barrier degree* of the

[8]Actually, in [24] the authors established the complexity of the algorithm without referring to condition C.4 albeit the proximity used in [24] is *self-regular*.

self-regular function; while the small-update method still keeps the best known $\mathcal{O}(\sqrt{n}\log\frac{n}{\varepsilon})$ iteration bound. In principle our algorithm can be viewed as a natural extension of primal-dual potential reduction methods. As we will see in the appendix of this work, some well known proximities used in the IPM literature fall into this new family. It is worthwhile to note that a simple idea, using some higher order barrier term in the proximity, and subsequently in the search direction leads to some improvement over the complexity of the IPM algorithm with large-updates. This gives rise to some interesting issues: The first one 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 [17], the target we used is μe (or μE for SDO), what will happen if we replace this by the new target $v\psi'(v)+v^2$? We also mention that, we have implemented a simple version of our algorithm, and run for few problems. Our preliminary numerical results show that the algorithm is promising. Nevertheless, much more work is needed to test the efficiency of the new approach in practice.

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'. An interesting observation mentioned at the end of Section 3, is that even when the same search direction is employed, the proved complexity of the algorithm with large-update in this paper is better than the one presented in [23]. In this work we focus mainly on the new search direction and did not improve to the complexity of the standard large-update Newton method. Due to the above observation, it is reasonable to expect that a new analysis based on some new proximity might improve the complexity of the standard large-update Newton method as well. We note that, albeit the complexity of our algorithm depends only on the choice of the barrier function, or more concretely, relies on the *barrier degree* q. However the constants used in our estimation might become very large whenever the parameters p and q increase. It is of interest to consider how to choose appropriate parameters for a given problem.

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 topic deserves more studies. Lastly we would like to mention that there are also other ways to extend our results. 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 the algorithm will be determined by its local convergence properties.

It is also worthwhile to build similar algorithms for convex quadratic optimization, linear complementarity problems and general convex optimization problems. Some of such extensions will be addressed in our forthcoming papers [25, 26].

Acknowledgement: The authors would like to thank two anonymous referees and an associate editor for their helpful suggestions that lead to several substantial improvements in the paper.

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. Bellman. Introduction to Matrix Analysis. Vol. 12. Classics in Applied Mathematics, SIAM, Philadelphia, 1995.

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

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

[5] A.V. Fiacco and G.P. McCormick. *Nonlinear Programming : Sequential Unconstrained Minimization Techniques.* John Wiley & Sons, New York, 1968. Reprint: Volume 4 of *SIAM Classics in Applied Mathematics*, SIAM Publications, Philadelphia, PA 19104–2688, USA, 1990.

[6] R. Frisch. The logarithmic potential method for solving linear programming problems. Memorandum, University Institute of Economics, Oslo, 1955.

[7] F. Glineur. Private communication.

[8] 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.

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

[10] 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.

[11] F. Jarre. Interior point methods for classes of convex programs. In T. Terlaky, editor, *Interior Point Methods of Mathematical Programming*, pages 189–252. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

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

[13] 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.

[14] 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.

[15] F.A. Lootsma. *Numerical Methods for Nonlinear Optimization.* Academic Press, London, UK, 1972.

[16] 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.

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

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

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

[20] 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.

[21] Y.E. Nesterov and A.S. Nemirovskii. Interior point polynomial algorithms in convex programming. *SIAM Studies in Applied Mathematics*, Vol(13),SIAM, Philadelphia, USA, 1994.

[22] 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.

[23] 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 Informatics, Delft University of Technology, The Netherlands, 1998. To appear in *Annals of Operations Research*

[24] J. Peng, C. Roos and T. Terlaky. A new class of polynomial primal-dual methods for linear and semidefinite optimization. Technical Report, Faculty of Technical Mathematics and Informatics, Delft University of Technology, The Netherlands, 1999.

[25] J. Peng, C. Roos and T. Terlaky. New polynomial primal-dual methods for second-order conic optimization. In Preparation, Faculty of Technical Mathematics and Informatics, Delft University of Technology, The Netherlands, 2000.

[26] J. Peng, C. Roos, T. Terlaky and A. Yoshise. New polynomial methods for nonlinear complementarity problems based on self-regular proximities. In Preparation, Faculty of Technical Mathematics and Informatics, Delft University of Technology, The Netherlands, 2000.

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

[28] W. Rudin, *Principles of Mathematical Analysis.* Mac-Graw Hill Company, New York, 1978.

[29] 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.

[30] J.F. Sturm, S. Zhang, Symmetric primal-dual path following algorithms for semidefinite programming, *Applied Numerical Mathematics*, 29:301-315, 1999.

[31] M.J. Todd, A study of search directions in primal-dual interior-point methods for semidefinite programming, *Optimization Methods and Software*, 11:1-46, 1999.

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

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

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

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

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

[37] 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.

6 Appendix

In this appendix we present the detailed proofs of several important technical conclusions in the paper. The appendix consists of three parts: Appendix A.1, A.2 and A.3. The first part provides the detailed proofs of some results in Section 2. The second part (Appendix A.2) contains the complete proofs of certain statements in Section 3. The last part (Appendix A.3) includes the proofs of several conclusions presented in Section 4 in detail.

6.1 Appendix A.1

Proof of Proposition 2.7:

We need only to consider the case that both β_1 and β_2 are positive. Since the functions $\psi_1(t), \psi_2(t) \in \Omega_1 \cap \Omega_2$, it follows immediately from Lemma 2.2 that the function $\psi(t) := \beta_1\psi_1(t) + \beta_2\psi_2(t) \in \Omega_2$. Obviously $\psi(t)$ is strictly convex and has a global minimizer at $t = 1$. Hence it remains to show that the imposed condition (11) with respect to the second derivative $\psi''(t)$ holds for some constants $\nu_1, \nu_2 > 0$ and $p, q \geq 1$. Since $\psi_1(t)$ and $\psi_2(t)$ are self-regular, there exist constants $\nu_1^1, \nu_2^1, \nu_1^2, \nu_2^2 > 0$ and $p_1, q_1, p_2, q_2 \geq 1$ such that

$$\nu_1^1(t^{p_1-1} + t^{-1-q_1}) \leq \psi_1''(t) \leq \nu_2^1(t^{p_1-1} + t^{-1-q_1}); \tag{89}$$

$$\nu_1^2(t^{p_2-1} + t^{-1-q_2}) \leq \psi_2''(t) \leq \nu_2^2(t^{p_2-1} + t^{-1-q_2}). \tag{90}$$

Let

$$p = \max(p_1, p_2), \quad q = \max(q_1, q_2).$$

Obviously $p, q \geq 1$ holds. From the definition of $\psi(t)$ we obtain

$$\begin{aligned}
\psi''(t) &\geq \beta_1\nu_1^1(t^{p_1-1} + t^{-1-q_1}) + \beta_2\nu_1^2(t^{p_2-1} + t^{-1-q_2}) \\
&= \left(\beta_1\nu_1^1 t^{p_1-1} + \beta_2\nu_1^2 t^{p_2-1}\right) + \left(\beta_1\nu_1^1 t^{q_1-1} + \beta_2\nu_1^2 t^{q_2-1}\right) \\
&> \min(\beta_1, \beta_2) \min\left(\nu_1^1, \nu_1^2\right)\left(t^{p-1} + t^{-1-q}\right).
\end{aligned}$$

Note that by the choice of p and q, we obtain

$$0 < \frac{t^{p_1-1} + t^{-1-q_1}}{t^{p-1} + t^{-1-q}} = \frac{t^{p_1-1}}{t^{p-1} + t^{-1-q}} + \frac{t^{-1-q_1}}{t^{p-1} + t^{-1-q}} < 2;$$

$$0 < \frac{t^{p_2-1} + t^{-1-q_2}}{t^{p-1} + t^{-1-q}} = \frac{t^{p_2-1}}{t^{p-1} + t^{-1-q}} + \frac{t^{-1-q_2}}{t^{p-1} + t^{-1-q}} < 2.$$

This relation, combining with (89) and (90), yields

$$\psi''(t) \leq 2\left(\beta_1\nu_2^1 + \beta_2\nu_2^2\right)\left(t^{p-1} + t^{-1-q}\right).$$

Let us choose

$$\nu_1 = \min(\beta_1, \beta_2) \min\left(\nu_1^1, \nu_1^2\right), \quad \nu_2 = 2\left(\beta_1\nu_2^1 + \beta_2\nu_2^2\right).$$

From our above discussion we can claim that

$$\nu_1\left(t^{p-1} + t^{-1-q}\right) \leq \psi''(t) \leq \nu_2\left(t^{p-1} + t^{-1-q}\right),$$

which completes the proof of Proposition 2.7. □

Proof of Lemma 2.9:
Recall from (11) that $\nu_1 \Upsilon''_{p,q}(t) \leq \psi''(t)$, and from (15) that $|\psi'(t)| \leq \nu_2 \left|\Upsilon'_{p,q}(t)\right|$. Hence it suffices for the proof if we show that

$$\left|\Upsilon'_{p,q}(t)\right| \leq t\Upsilon''_{p,q}(t), \quad t > 0. \tag{91}$$

By Lemma 2.2, $\Upsilon_{p,q}(t) \in \Omega_2$. This means that

$$\Upsilon'_{p,q}(t) + t\Upsilon''_{p,q}(t) = t^p \left(1 + \frac{1}{p}\right) - \frac{1}{p} + t^{-q}\left(1 - \frac{1}{q}\right) + \frac{1}{q} \geq 0. \tag{92}$$

Hence it remains to show that

$$-\Upsilon'_{p,q}(t) + t\Upsilon''_{p,q}(t) \geq 0, \quad t > 0.$$

By simple calculus we can restate this inequality as

$$-\frac{t^p - 1}{p} + \frac{t^{-q} - 1}{q} + t^p + t^{-q} \geq 0,$$

or, equivalently,

$$t^p \left(1 - \frac{1}{p}\right) + \frac{1}{p} + t^{-q}\left(1 + \frac{1}{q}\right) - \frac{1}{q} \geq 0, \quad t > 0,\, p \geq 1,\, q \geq 1. \tag{93}$$

Replacing t by $1/t$, this is equivalent to

$$t^{-p} \left(1 - \frac{1}{p}\right) + \frac{1}{p} + t^{q}\left(1 + \frac{1}{q}\right) - \frac{1}{q} \geq 0, \quad t > 0,\, p \geq 1,\, q \geq 1.$$

When interchanging p and q we observe that this is exactly the same inequality as (92). Hence the proof is complete. □

Proof of Proposition 2.11:
We first prove the proposition for a special case[9] when $\psi(t) = \Upsilon(t) := \Upsilon_{p,q}(t)$. In other words, we want to show that there is a constant $\mathcal{C}_\nu$ such that

$$\Upsilon(t)\Upsilon''(t) \leq \mathcal{C}_\nu \Upsilon'(t)^2, \quad \forall t > 0.$$

The above inequality holds trivially if $t = 1$. Hence it suffices to show that there exists a constant $\mathcal{C}_\nu$ satisfying

$$\Im(t) = \frac{\Upsilon(t)\Upsilon''(t)}{\Upsilon'(t)^2} \leq \mathcal{C}_\nu, \quad \forall 1 \neq t > 0.$$

From the definition of $\Upsilon(t)$, it easily follows

$$\lim_{t\to\infty} \Im(t) = \frac{p}{p+1}, \quad \lim_{t>0, t\to 0} \Im(t) = \frac{q}{q-1}.$$

Moreover, by using l'Hospital's rule, one readily obtains

$$\lim_{t\to 1} \Im(t) = \lim_{t\to 1} \frac{\Upsilon(t)\Upsilon''(t)}{\Upsilon'(t)^2} = \frac{1}{2}.$$

[9]For simplification of expression, we temporarily drop the indices in the function $\Upsilon_{p,q}(t)$ and denote it by $\Upsilon(t)$.

From the above three limits and the continuity of $\Im(t)$ on the intervals $(0,1)$ and $(1,\infty)$, we can conclude that there is a constant $\mathcal{C}_\nu$ satisfying $\Im(t) \le \mathcal{C}_\nu$, which equals to the statement of the proposition when $\psi(t) = \Upsilon(t)$.

Now we consider the case that $\psi(t)$ is a general function satisfying C.1. By (11) one has

$$\nu_1 \Upsilon''(t) \le \psi''(t) \le \nu_2 \Upsilon''(t).$$

Since $\psi'(t) \ge 0$ if $t \ge 1$ and $\psi'(t) < 0$ whenever $t < 1$, from the above relation we obtain

$$\begin{aligned} 0 \le \nu_1 \Upsilon'(t) \le \psi'(t) \le \nu_2 \Upsilon'(t), \qquad & \forall t \ge 1; \\ \nu_2 \Upsilon'(t) \le \psi'(t) \le \nu_1 \Upsilon'(t) < 0, \qquad & \forall t \in (0,1). \end{aligned}$$

It follows

$$\psi'(t)^2 \ge {\nu_1}^2 \Upsilon'(t)^2 \ge \frac{{\nu_1}^2}{\mathcal{C}_\nu} \Upsilon(t) \Upsilon''(t) \ge \frac{{\nu_1}^2}{\mathcal{C}_\nu {\nu_2}^2} \psi(t) \psi''(t),$$

where the second inequality follows from our discussion for the special case $\psi(t) = \Upsilon(t)$, and the last inequality is a consequence of (11) and (14). Replacing $\mathcal{C}_\nu$ by

$$\mathcal{C}_\nu := \max \left(\mathcal{C}_\nu, \frac{{\nu_2}^2}{{\nu_1}^2} \mathcal{C}_\nu \right),$$

one gets the desired inequality in the proposition. □

Proof of Lemma 2.14:

First we observe that if $\vartheta t \le 1$ then $\psi(\vartheta t) < \psi(t)$, and then the inequality in the lemma is trivial, because ψ is strictly convex and minimal at $t = 1$. Thus, we may assume $\vartheta t > 1$. Now, if $t \le 1$, then $1 < \vartheta t \le \vartheta$, whence we have $\psi(\vartheta t) \le \psi(\vartheta) \le \nu_2 \Upsilon_{p,q}(\vartheta)$. Hence, if $t < 1$ the inequality in the lemma certainly holds. Therefore, from now on we assume that $t > 1$.

According to (14), we have

$$\psi(\vartheta t) \le \nu_2 \Upsilon_{p,q}(\vartheta t).$$

The rest of the proof consists of deriving an upper bound for $\Upsilon_{p,q}(\vartheta t)$. To simplify the notation we denote $\Upsilon_{p,q}$ as Υ. From the definition of Υ, we deduce that

$$\Upsilon(\vartheta t) = \frac{\vartheta^{p+1} t^{p+1} - 1}{p(p+1)} + \frac{\vartheta^{1-q} t^{1-q} - 1}{q(q-1)} + \frac{(p-q)(\vartheta t - 1)}{pq}.$$

By straightforward, though cumbersome, calculation, this can be rewritten as

$$\begin{aligned} \Upsilon(\vartheta t) &= \vartheta^{1-q} \left(\frac{t^{p+1} - 1}{p(p+1)} + \frac{t^{1-q} - 1}{q(q-1)} + \frac{p-q}{pq} (t-1) \right) + \frac{(\vartheta^{p+1} - \vartheta^{1-q})}{p(p+1)} (t^{p+1} - 1) \\ &\quad + \frac{(p-q)(\vartheta - \vartheta^{1-q})}{pq} (t-1) + \left(\frac{\vartheta^{p+1} - 1}{p(p+1)} + \frac{\vartheta^{1-q} - 1}{q(q-1)} + \frac{p-q}{pq} (\vartheta - 1) \right) \\ &= \vartheta^{1-q} \Upsilon(t) + \left(\vartheta^{p+1} - \vartheta^{1-q} \right) \frac{t^{p+1} - 1}{p(p+1)} + \frac{(p-q)(\vartheta - \vartheta^{1-q})}{pq} (t-1) + \Upsilon(\vartheta). \end{aligned}$$

By using the second conclusion (20) of Lemma 2.12, and together with the fact that $\vartheta^{p+1} - \vartheta^{1-q} > 0$, we obtain

$$\Upsilon(\vartheta t) \le \vartheta^{1-q} \Upsilon(t) + \left(\vartheta^{p+1} - \vartheta^{1-q} \right) \left(\Upsilon(t) + \frac{t-1}{p} \right) + \frac{(p-q)(\vartheta - \vartheta^{1-q})}{pq} (t-1) + \Upsilon(\vartheta)$$

$$= \vartheta^{p+1}\Upsilon(t) + \frac{\vartheta^{p+1} - \vartheta^{1-q}}{p}(t-1) + \frac{(p-q)(\vartheta - \vartheta^{1-q})}{pq}(t-1) + \Upsilon(\vartheta)$$

$$= \vartheta^{p+1}\Upsilon(t) + \vartheta\left(\frac{\vartheta^{p}-1}{p} + \frac{1-\vartheta^{-q}}{q}\right)(t-1) + \Upsilon(\vartheta)$$

$$= \vartheta^{p+1}\Upsilon(t) + \vartheta\Upsilon'(\vartheta)\,(t-1) + \Upsilon(\vartheta), \tag{94}$$

where the last equality follows from the choice of Υ. Now, using the first inequality of (19) in Lemma 2.12, we can conclude

$$\Upsilon(\vartheta t) \leq \vartheta^{p+1}\Upsilon(t) + \vartheta\Upsilon'(\vartheta)\sqrt{2\Upsilon(t)} + +\Upsilon(\vartheta).$$

Finally, substitution of $\Upsilon(t) \leq \frac{\psi(t)}{\nu_1}$ yields the inequality in the lemma. □

6.2 Appendix A.2

Proof of Lemma 3.11:
Since $\Psi(v) \geq \nu_1^{-1}$, by Proposition 3.3 one has $\sigma \geq 1$. Let us define

$$w_1(\alpha) = -\frac{\sigma}{2} + \frac{\nu_2}{p}\left((v_{\max} + \alpha\sigma)^p - v_{\max}^p\right)$$

and

$$w_2(\alpha) = -\frac{\sigma}{2} + \frac{\nu_2}{q}\left((v_{\min} - \alpha\sigma)^{-q} - v_{\min}^{-q}\right).$$

It is easy to verify that both functions $w_1(\alpha)$ and $w_2(\alpha)$ are increasing for $\alpha \in [0, \bar{\alpha})$. Using these two functions, we can rewrite equation (47) as

$$w_1(\alpha^*) + w_2(\alpha^*) = 0.$$

Through simple calculus, we find that $w_1(\alpha_1^*) = 0$ if

$$\alpha_1^* = \frac{v_{\max}}{\sigma}\left(\left(1 + \frac{p\sigma}{2\nu_2 v_{\max}^p}\right)^{\frac{1}{p}} - 1\right).$$

We now progress to estimate α_1^*. First observe that, since $p \geq 1$, by using (49) in Lemma 3.10, we obtain

$$\left(1 + \frac{p\sigma}{2\nu_2 v_{\max}^p}\right)^{\frac{1}{p}} = \left(1 - \frac{p\sigma}{p\sigma + 2\nu_2 v_{\max}^p}\right)^{-\frac{1}{p}} \geq \left(1 - \frac{\sigma}{p\sigma + 2\nu_2 v_{\max}^p}\right)^{-1}.$$

It follows readily

$$\alpha_1^* \geq \frac{v_{\max}}{\sigma}\left(\left(1 - \frac{\sigma}{p\sigma + 2\nu_2 v_{\max}^p}\right)^{-1} - 1\right) \geq \frac{v_{\max}}{p\sigma + 2\nu_2 v_{\max}^p} \geq \frac{1}{p\sigma + 2\nu_2 v_{\max}^p}, \tag{95}$$

where the last inequality is given by the assumption that $v_{\max} \geq 1$. Because $p \geq 1$, from (30) we can conclude that

$$v_{\max}^p \leq \frac{p\sigma}{\nu_1} + 1 \leq \frac{p\sigma}{\nu_1} + \sigma.$$

This relation, joint with (95), yields

$$\alpha_1^* \geq \frac{\nu_1}{2\nu_1\nu_2 + p(\nu_1 + 2\nu_2)}\sigma^{-1}.$$

We next estimate the root of $w_2(\alpha)$, or in other words the value

$$\alpha_2^* = \frac{v_{\min}}{\sigma}\left(1 - \left(1 + \frac{q\sigma v_{\min}^q}{2\nu_2}\right)^{-\frac{1}{q}}\right).$$

Again, via applying (49), one gets

$$\left(1 + \frac{q\sigma v_{\min}^q}{2\nu_2}\right)^{-\frac{1}{q}} = \left(1 - \frac{q\sigma v_{\min}^q}{2\nu_2 + q\sigma v_{\min}^q}\right)^{\frac{1}{q}} \leq 1 - \frac{\sigma v_{\min}^q}{2\nu_2 + q\sigma v_{\min}^q}.$$

Hence

$$\alpha_2^* \geq \frac{v_{\min}}{\sigma}\frac{\sigma v_{\min}^q}{2\nu_2 + q\sigma v_{\min}^q} = \frac{v_{\min}^{q+1}}{2\nu_2 + q\sigma v_{\min}^q}. \tag{96}$$

Recalling (30) one obtains

$$v_{\min}^q \geq \frac{\nu_1}{\nu_1 + q\sigma}, \tag{97}$$

which, together with (96), further implies

$$\alpha_2^* \ \geq \frac{\nu_1 v_{\min}}{2\nu_1\nu_2 + q(\nu_1 + 2\nu_2)}\sigma^{-1}. \tag{98}$$

Since $\sigma \geq 1$, from (97) it readily follows

$$v_{\min} \geq \left(1 + \frac{q\sigma}{\nu_1}\right)^{-\frac{1}{q}} \geq \left(1 + \frac{q}{\nu_1}\right)^{-\frac{1}{q}}\sigma^{-\frac{1}{q}} \geq \frac{\nu_1}{1+\nu_1}\sigma^{-\frac{1}{q}}. \tag{99}$$

where the last inequality is true because, by using (48) in Lemma 3.10, we have

$$\left(1 + \frac{q}{\nu_1}\right)^{\frac{1}{q}} \leq 1 + \frac{1}{\nu_1}.$$

Combining (98) with (99), one can conclude

$$\alpha_2^* \geq \frac{\nu_1^2}{(1+\nu_1)\left(2\nu_1\nu_2 + q(\nu_1 + 2\nu_2)\right)}\sigma^{-\frac{q+1}{q}}.$$

Now let us recall the fact that both $w_1(\alpha)$ and $w_2(\alpha)$ are increasing functions of α, from the above discussions we easily conclude

$$\alpha^* \geq \min\left(a_1^*, a_2^*\right) \geq \nu_5\sigma^{-\frac{q+1}{q}},$$

where

$$\nu_5 = \min\{\frac{\nu_1}{2\nu_1\nu_2 + p(\nu_1 + 2\nu_2)}, \frac{\nu_1^2}{(1+\nu_1)\left(2\nu_1\nu_2 + q(\nu_1 + 2\nu_2)\right)}\}$$

is a constant depending only on the proximity $\Psi(v)$ and independent of the undertaken problem. This proves the first conclusion of the lemma. The second statement in the lemma is a direct consequence of inequality (50). The proof of the lemma is finished. □

Proof of Lemma 3.18:
First recall that $v_+ = (v+\alpha[d_x])^{\frac{1}{2}}(v+\alpha[d_s])^{\frac{1}{2}}$, from (40) we obtain

$$\begin{aligned} f''(\alpha) = \sum_{i=1}^n \frac{\partial^2 \psi(v_{+i})}{\partial \alpha^2} &= \frac{1}{4}\sum_{i=1}^n \left(\frac{v_i+\alpha[d_s]_i}{v_i+\alpha[d_x]_i}[d_x]_i^2 + 2[d_x]_i[d_s]_i + \frac{v_i+\alpha[d_x]_i}{v_i+\alpha[d_s]_i}[d_s]_i^2\right)\psi''(v_{+i}) \\ &\quad -\frac{1}{4}\sum_{i=1}^n \left(\frac{v_i+\alpha[d_s]_i}{v_i+\alpha[d_x]_i}[d_x]_i^2 - 2[d_x]_i[d_s]_i + \frac{v_i+\alpha[d_x]_i}{v_i+\alpha[d_s]_i}[d_s]_i^2\right)\left(\frac{\psi'(v_{+i})}{v_{+i}}\right). \end{aligned}$$

It is trivial to see that

$$[d_x]_i[d_s]_i \le \frac{1}{2}\left(\frac{v_i+\alpha[d_s]_i}{v_i+\alpha[d_x]_i}[d_x]_i^2 + \frac{v_i+\alpha[d_x]_i}{v_i+\alpha[d_s]_i}[d_s]_i^2\right), \quad \forall i=1,\cdots,n.$$

Hence, by using condition C.1 and Lemma 2.9, one has

$$\begin{aligned} f''(\alpha) &\le \frac{\nu_2}{2}\left(1+\frac{\nu_2}{\nu_1}\right)\sum_{i=1}^n \left(\frac{v_i+\alpha[d_s]_i}{v_i+\alpha[d_x]_i}[d_x]_i^2 + \frac{v_i+\alpha[d_x]_i}{v_i+\alpha[d_s]_i}[d_s]_i^2\right)\left(v_{+i}^{p-1}+v_{+i}^{-q-1}\right) \\ &\le \frac{\nu_2}{2}\left(1+\frac{\nu_2}{\nu_1}\right)\sum_{i=1}^n \left(\frac{1+\alpha\bar{d}_{si}}{1+\alpha\bar{d}_{xi}}[d_x]_i^2 + \frac{1+\alpha\bar{d}_{xi}}{1+\alpha\bar{d}_{si}}[d_s]_i^2\right)\left((v_{\max}+\alpha\sigma)^{p-1}+(v_{\min}-\alpha\sigma)^{-q-1}\right), \end{aligned} \tag{100}$$

where the last inequality is true, because by using the notations (8), from (43) and (44) one can easily derive the following relation

$$v_{\min}-\alpha\sigma \le v_{+i} \le v_{\max}+\alpha\sigma, \quad i \in \mathcal{I}.$$

If $\alpha \in [0, \frac{1}{2}\bar{\alpha}]$, then it is easy to see that for any $i \in \{1,2,\cdots,n\}$, we have

$$0 \le \frac{1+\alpha\bar{d}_{xi}}{1+\alpha\bar{d}_{si}}, \frac{1+\alpha\bar{d}_{si}}{1+\alpha\bar{d}_{xi}} \le 3.$$

This relation, together with (100), (26) and (37), gives the first statement of the lemma.

To prove the second conclusion of the lemma, we observe that, from the first inequality in (100) one can derive

$$\begin{aligned} f''(\alpha) &\le \frac{\nu_2}{2}\sum_{i=1}^n \left((v_i+\alpha[d_s]_i)^{\frac{p+1}{2}}(v_i+\alpha[d_x]_i)^{\frac{p-3}{2}} + (v_i+\alpha[d_s]_i)^{\frac{-q+1}{2}}(v_i+\alpha[d_x]_i)^{\frac{-q-3}{2}}\right)[d_x]_i^2 \\ &\quad + \frac{\nu_2^2}{2\nu_1}\sum_{i=1}^n \left((v_i+\alpha[d_s]_i)^{\frac{p-3}{2}}(v_i+\alpha[d_x]_i)^{\frac{p+1}{2}} + (v_i+\alpha[d_x]_i)^{\frac{-q+1}{2}}(v_i+\alpha[d_s]_i)^{\frac{-q-3}{2}}\right)[d_s]_i^2 \\ &\le \frac{\nu_2}{2}\left(1+\frac{\nu_2}{\nu_1}\right)\left((v_{\max}+\alpha\sigma)^{p-1}+(v_{\min}-\alpha\sigma)^{-q-1}\right)\sum_{i=1}^n\left([d_x]_i^2+[d_s]_i^2\right) \\ &= \frac{\nu_2\sigma^2}{2}\left(1+\frac{\nu_2}{\nu_1}\right)\left((v_{\max}+\alpha\sigma)^{p-1}+(v_{\min}-\alpha\sigma)^{-q-1}\right), \end{aligned}$$

where the second inequality is given by (43) and (44), and the last equality follows by using the definition of σ (26) and (37). The proof of the lemma is completed. □

6.3 Appendix A.3

Proof of Proposition 4.4:
We first show that condition C.3 is true. For this we must prove that $\Psi(X)$ is strictly convex for $X \succ 0$, i.e., for any $X_1 \neq X_2 \succ 0$, the following inequality holds

$$\Psi\left(\frac{X_1 + X_2}{2}\right) < \frac{1}{2}\left(\Psi(X_1) + \Psi(X_2)\right). \tag{101}$$

Since both X_1, X_2 are positive definite, so is the matrix $\frac{1}{2}(X_1 + X_2)$. Let Q be the orthogonal matrix which diagonalizes $\frac{1}{2}(X_1 + X_2)$, i.e.,

$$\Lambda = \text{diag }(\lambda) = \frac{1}{2} Q(X_1 + X_2)Q^T = \Lambda = \frac{1}{2}\left(Q_1\Lambda_1 Q_1^T + Q_2\Lambda_2 Q_2^T\right), \tag{102}$$

where both Q_1, Q_2 are orthogonal, $\Lambda_1 = \text{diag }(\lambda^1), \Lambda_2 = \text{diag }(\lambda^2)$ are positive diagonal matrices, and $\lambda, \lambda^1, \lambda^2$ are vectors whose components are the eigenvalues of $\frac{1}{2}(X_1 + X_2), X_1$ and X_2 respectively. Denote G the matrix whose entries are defined by $G_{ij} = [Q_1]_{ij}^2$ and similarly H with $H_{ij} = [Q_2]_{ij}^2$. From (102) we readily obtain

$$\lambda = \frac{1}{2}(G\lambda^1 + H\lambda^2).$$

Using the orthogonality of the matrices Q_1 and Q_2, one can easily see that G and H are two doubly stochastic matrices whose entries satisfy the following relations

$$\sum_{i=1}^{n} G_{ij} = \sum_{i=1}^{n} H_{ij} = 1, \quad j = 1, 2, \cdots, n;$$
$$\sum_{j=1}^{n} G_{ij} = \sum_{j=1}^{n} H_{ij} = 1, \quad i = 1, 2, \cdots, n.$$

Hence

$$\begin{aligned}\Psi\left(\frac{X_1 + X_2}{2}\right) &= \sum_{i=1}^{n} \psi(\lambda_i) = \sum_{i=1}^{n} \psi\left(\frac{G_i\lambda^1 + H_i\lambda^2}{2}\right) \\ &< \frac{1}{2}\left(\sum_{i=1}^{n} \psi(G_i\lambda^1) + \sum_{i=1}^{n} \psi(H_i\lambda^2)\right),\end{aligned}$$

where the inequality follows from the strict convexity of $\psi(t)$ and the fact that $X_1 \neq X_2$. Using the convexity of $\psi(t)$ again, we have

$$\sum_{i=1}^{n} \psi(G_i\lambda^1) \leq \sum_{i=1}^{n}\sum_{j=1}^{n} G_{ij}\psi(\lambda_j^1) = \sum_{j=1}^{n} \psi(\lambda_j^1) = \Psi(X_1),$$

where the first equality given by the choice of the matrix G. Similarly one has

$$\sum_{i=1}^{n} \psi(H_i\lambda^2) \leq \Psi(X_2).$$

The above two inequalities yield the desired relation (101). From (65) and (61) one can easily verify all other statements in condition C.3.

We next switch to condition C.4. Using the notations introduced earlier at the end of Section 1 in the paper, for any nonsingular matrix $G \in \Re^{n\times n}$ and any $i \in \mathcal{I}$, there holds

$$\varrho_i(G) = \left(\lambda_i(G^T G)\right)^{\frac{1}{2}} = \left(\lambda_i(GG^T)\right)^{\frac{1}{2}}.$$

It follows immediately that for any $X, S \succ 0$,

$$\varrho_i\left(X^{\frac{1}{2}}S^{\frac{1}{2}}\right) = \left(\lambda_i\left(X^{\frac{1}{2}}SX^{\frac{1}{2}}\right)\right)^{\frac{1}{2}} = \lambda_i\left([X^{\frac{1}{2}}SX^{\frac{1}{2}}]^{\frac{1}{2}}\right).$$

Now, invoking the definition of $\Psi(X)$, one gets

$$\Psi\left([X^{\frac{1}{2}}SX^{\frac{1}{2}}]^{\frac{1}{2}}\right) = \sum_{i=1}^{n}\psi\left(\varrho_i(X^{\frac{1}{2}}S^{\frac{1}{2}})\right).$$

We progress by quoting a result in [3] (conclusion (d) of Theorem 3.3.14 on page 176-177) which claims that, for any matrices $G, H \in \Re^{n\times n}$, if the function $f(\exp(t))$ is convex on $(-\infty, \infty)$, and if either G or H is nonsingular, or $f(\cdot)$ is continuous on $[0, \infty)$, then

$$\sum_{i=1}^{n} f(\varrho_i(GH)) \leq \sum_{i=1}^{n} f(\varrho_i(G)\varrho_i(H)). \tag{103}$$

Since the function $\psi(t) \in \Omega_2$, by Lemma 2.2 $\psi(\exp(t))$ is convex on $(-\infty, \infty)$. Replacing the arguments f, G, H in (103) by $\psi, X^{\frac{1}{2}}$ and $S^{\frac{1}{2}}$ respectively, we obtain

$$\begin{aligned}
\Psi\left([X^{\frac{1}{2}}SX^{\frac{1}{2}}]^{\frac{1}{2}}\right) &= \sum_{i=1}^{n}\psi\left(\varrho_i(X^{\frac{1}{2}}S^{\frac{1}{2}})\right) \leq \sum_{i=1}^{n}\psi\left(\varrho_i(X^{\frac{1}{2}})\varrho_i(S^{\frac{1}{2}})\right) \\
&\leq \frac{1}{2}\sum_{i=1}^{n}\left(\psi(\varrho_i^2(X^{\frac{1}{2}})) + \psi(\varrho_i^2(S^{\frac{1}{2}}))\right) \\
&= \frac{1}{2}\sum_{i=1}^{n}\left(\psi(\varrho_i(X)) + \psi(\varrho_i(S))\right) \\
&= \frac{1}{2}\left(\Psi(X) + \Psi(S)\right),
\end{aligned}$$

where the second inequality follows from condition C.2, the second equality is implied by the fact that both X and S are symmetric positive definite. The proof of the proposition is finished.

□

Proof of Lemma 4.9:

The proof is essentially an immediate consequence of (6.6.27) in [3], for self-completeness we write it out here. Since $H(t)$ is positive definite, it is diagonalizable. Hence we can assume without loss of generality that $H(t) = Q_H(t)^T\Lambda_H(t)Q_H(t)$ where $\Lambda_H(t)$ is a diagonal matrix whose elements are the eigenvalues $\{\lambda_1(t), \lambda_2(t), \cdots, \lambda_n(t)\}$ of $H(t)$, and $Q_H(t)$ is orthogonal. Denote $D_j(t)$ the diagonal matrix which has a 1 in its j, j position, and all other entries of $D_j(t)$ are zero. For notation simplicity we will omit the parameter t and use the notations $Q_H, D_j, H, H', \lambda_j$ for $Q_H(t), D_j(t), H(t), H'(t), \lambda_j(t)$ respectively in the proof of this lemma. It follows from (6.6.27) in [3] (page 531) that[10]

$$\frac{d}{dt}\psi(H(t)) = Q_H^T\left(\sum_{j,k=1}^{n}\Delta\psi(\lambda_j, \lambda_k)D_j[Q_H H' Q_H^T]D_k\right)Q_H. \tag{104}$$

[10] We mention that in [3], the authors used the distinct eigenvalues of $H(t)$ to define the block matrix D_j whose i, i-th entry is 1 if $\lambda_i = \lambda_j$ while the constant n in (104) was replaced by the number of distinct eigenvalues of $H(t)$. However, it can be easily verified that our definition (104) is equivalent to that given by (6.6.27) in [3]. We introduce (104) to keep consistence with our definition about a matrix's eigenvalues in the introduction and avoid too many new notations.

Recalling the definition of the trace of a matrix, we know that for any matrices G, H, there holds $\mathbf{Tr}\,(GH) = \mathbf{Tr}\,(HG)$. This simple fact, combining with (104) and (76) gives

$$\mathbf{Tr}\left(\frac{d}{dt}\psi(H(t))\right) = \mathbf{Tr}\left(\sum_{j,k=1}^{n} \Delta\psi(\lambda_j, \lambda_k) D_j [Q_H H' Q_H^T] D_k\right).$$

Now by the choice of D_j, we can claim that for any symmetric matrix G, there holds $\mathbf{Tr}\,(D_j G D_k) = 0$ if $\lambda_j(t) \neq \lambda_k(t)$. Thus it readily follows

$$\begin{aligned}
\mathbf{Tr}\left(\frac{d}{dt}\psi(H(t))\right) &= \mathbf{Tr}\left(\sum_{j,k=1}^{n} \Delta\psi(\lambda_j, \lambda_k) D_j [Q_H H' Q_H^T] D_k\right) \\
&= \mathbf{Tr}\left(\sum_{j=1}^{n} \Delta\psi(\lambda_j, \lambda_j) D_j [Q_H H' Q_H^T] D_j\right) \\
&= \mathbf{Tr}\left(\sum_{j=1}^{n} \psi'(\lambda_j) [Q_H^T D_j^2 Q_H] H'\right),
\end{aligned}$$

where the last equality given by the definition of $\Delta\psi(\cdot,\cdot)$. The above equality further implies

$$\begin{aligned}
\mathbf{Tr}\left(\frac{d}{dt}\psi(H(t))\right) &= \mathbf{Tr}\left(\sum_{j=1}^{n} Q_H^T(t) [D_j \psi'(\lambda_j) D_j] Q_H H'\right) \\
&= \mathbf{Tr}\left(Q_H^T \mathrm{diag}\,\left(\psi'(\lambda_1), \psi'(\lambda_2), \cdots, \psi'(\lambda_n)\right) Q_H H'\right) \\
&= \mathbf{Tr}\,(\psi'(H) H').
\end{aligned}$$

This completes the proof of the first conclusion in the lemma.

We now progress to prove the second statement of the lemma. First note that, by applying the chain rule (77) one easily gets

$$\frac{d^2}{dt^2}\mathbf{Tr}\,(\psi(H(t))) = \frac{d}{dt}\mathbf{Tr}\,(\psi'(H(t))H'(t)) = \mathbf{Tr}\,(\psi'(H(t))H''(t)) + \mathbf{Tr}\left([\frac{d}{dt}\psi'(H(t))]H'(t)\right).$$

Hence it remains to show that

$$\mathbf{Tr}\left([\frac{d}{dt}\psi'(H(t))]H'(t)\right) \leq \varpi \left\|H'(t)\right\|^2.$$

Substituting the function $\psi(t)$ by $\psi'(t)$ in (104) we obtain

$$\begin{aligned}
\mathbf{Tr}\left([\frac{d}{dt}\psi'(H(t))]H'(t)\right) &= \mathbf{Tr}\left(Q_H^T \left(\sum_{j,k=1}^{n} \Delta\psi'(\lambda_j, \lambda_k) D_j [Q_H H' Q_H^T] D_k\right) Q_H H'\right) \\
&= \mathbf{Tr}\left(\left(\sum_{j,k=1}^{n} \Delta\psi'(\lambda_j, \lambda_k) D_j [Q_H H' Q_H^T]\right) D_k [Q_H H' Q_H^T]\right).
\end{aligned}$$

Again, by the choices of D_j, D_k, we can easily verify that, for any symmetric matrix G, $\mathbf{Tr}\,(D_j G D_k G) = G_{j,k}^2$. Hence it immediately follows

$$\begin{aligned}
\mathbf{Tr}\left([\frac{d}{dt}\psi'(H(t))]H'(t)\right) &= \mathbf{Tr}\left(\sum_{j,k=1}^{n} \Delta\psi'(\lambda_j, \lambda_k) [Q_H H' Q_H^T]_{j,k}^2\right) \\
&\leq \varpi \sum_{j,k=1}^{n} [Q_H H'(t) Q_H^T]_{j,k}^2 \\
&= \varpi \left\|H'(t)\right\|^2,
\end{aligned}$$

where the inequality is given by the definition of ϖ. The proof of the lemma is completed. □

Proof of Lemma 4.11:

First notice that by Lemma 4.10 the step size used in the lemma is strictly feasible. From Lemma 4.9 we conclude

$$f_1''(\alpha) \leq \frac{1}{2}\left(\varpi_1 \|D_X\|^2 + \varpi_2 \|D_S\|^2\right),$$

where

$$\varpi_1 = \max\{\left|\Delta\psi'(\lambda_j(V+\alpha D_X), \lambda_k(V+\alpha D_X))\right| : \ j,k \in \mathcal{I}\};$$
$$\varpi_2 = \max\{\left|\Delta\psi'(\lambda_j(V+\alpha D_S), \lambda_k(V+\alpha D_S))\right| : \ j,k \in \mathcal{I}\}.$$

Since $\sigma^2 = \|D_X\|^2 + \|D_S\|^2$, it suffices to prove the following inequality

$$\max(\varpi_1, \varpi_2) \leq \nu_2\left((\lambda_{\max}(V) + \alpha\sigma)^{p-1} + (\lambda_{\min}(V) - \alpha\sigma)^{-q-1}\right). \tag{105}$$

From the choice of ϖ_1 we can safely claim that

$$\varpi_1 = \left|\Delta\psi'(\lambda_{j_*}(V+\alpha D_X), \lambda_{k_*}(V+\alpha D_X))\right|$$

for some index $j_*, k_* \in \mathcal{I}$. By the definition of $\Delta\psi'(\cdot,\cdot)$ and the mean value theorem [28], there exists a constant

$$\zeta_* \in [\min(\lambda_{j_*}(V+\alpha D_X), \lambda_{k_*}(V+\alpha D_X)), \max(\lambda_{j_*}(V+\alpha D_X), \lambda_{k_*}(V+\alpha D_X))]$$

satisfying

$$\psi''(\zeta_*) = \Delta\psi'(\lambda_{j_*}(V+\alpha D_X), \lambda_{k_*}(V+\alpha D_X)).$$

This relation, combining with condition C.1 yields

$$\varpi_1 \leq \nu_2\left(\zeta_*^{p-1} + \zeta_*^{-q-1}\right), \tag{106}$$

for some

$$\zeta_* \in [\min(\lambda_i(V+\alpha D_X) : i \in \mathcal{I}), \max(\lambda_i(V+\alpha D_X) : i \in \mathcal{I})].$$

Now let us recall the assumption in the lemma $\alpha \in [0, \bar{\alpha})$, which further implies that for any $i \in \mathcal{I}$,

$$\lambda_{\min}(V) - \alpha\sigma \leq \lambda_{\min}(V) - \alpha\|D_X\|_2 \leq \lambda_i(V+\alpha D_X) \leq \lambda_{\max}(V) + \alpha\|D_X\|_2 \leq \lambda_{\max}(V) + \alpha\sigma.$$

It follows immediately

$$\lambda_{\min}(V) - \alpha\sigma \leq \zeta_* \leq \lambda_{\max}(V) + \alpha\sigma.$$

As a direct consequence of the above equality and (106), we have

$$\varpi_1 \leq \nu_2\left((\lambda_{\max}(V) + \alpha\sigma)^{p-1} + (\lambda_{\min}(V) - \alpha\sigma)^{-q-1}\right).$$

Similarly one can show

$$\varpi_2 \leq \nu_2\left((\lambda_{\max}(V) + \alpha\sigma)^{p-1} + (\lambda_{\min}(V) - \alpha\sigma)^{-q-1}\right).$$

The above two inequalities give (105), which further concludes the lemma. □