McMaster University

Advanced Optimization Laboratory

Title:

The Complexity of Self-Regular Proximity Based Infeasible IPMs

Authors:

Maziar Salahi, Tamás Terlaky, Guoqing Zhang

AdvOl-Report No. 2003/3

May 2003, Hamilton, Ontario, Canada

The Complexity of Self-Regular Proximity Based Infeasible IPMs

Maziar Salahi*† , Tamás Terlaky†, Guoqing Zhang‡

May 5, 2003

Abstract

Primal-Dual Interior-Point Methods (IPMs) have shown their power in solving large classes of optimization problems. In this paper a self-regular proximity based Infeasible Interior Point Method (IIPM) is proposed for linear optimization problems. First we mention some interesting properties of a specific self-regular proximity function, studied recently by Peng and Terlaky, and use it to define infeasible neighborhoods. These simple but interesting properties of the proximity function indicate that, when the current iterate is in a large neighborhood of the central path, large-update IIPMs emerge as the only natural choice. Then, we apply these results to design a specific self-regularity based dynamic large-update IIPM in large neighborhood. The new dynamic IIPM always takes large-updates and does not utilize any inner iteration to get centered. An $O(n^2 \log \frac{n}{\epsilon})$ worst-case iteration bound of the algorithm is established. Finally, we report the main results of our computational experiments.

Keywords: Linear Optimization, Infeasible Interior Point Method, Self-Regular Proximity Function, Polynomial Complexity.

1 Introduction

Interior Point Methods (IPMs) initiated by Karmarkar [3] not only have polynomial complexity, but are also highly efficient in practice for solving Linear Optimization (LO) problems. A new paradigim of IPMs based on Self-Regular (SR) proximity functions was presented by Peng, Roos and Terlaky [14]. They proved that SR-proximity based feasible IPMs for LO enjoy the best worst case theoretical complexity of large-update IPMs. In this paper, we aim to develop an SR-proximity based Infeasible IPM (IIPM) for LO and establish its polynomial complexity.

Feasible IPMs start with a strictly feasible interior point and keep feasibility during the solution process. It is not trivial how to find an initial feasible interior point. One method to overcome this problem is to use the homogeneous embedding model by introducing artificial variables. Such a homogenous self-dual formulation was presented first by Ye et al.[28]. The other option is

*Department of Mathematical sciences, Sharif University of Technology, P.O. Box 11365-9415, Tehran, Iran, email: salahi@mehr.sharif.edu,

†Advanced Optimization Lab, Department of Computing and Software, McMaster University, Hamilton, Ontario, Canada, L8S 4L7. email: terlaky@mcmaster.ca, msalahi@oplab.cas.mcmaster.ca

‡Department of Industrial and Manufacturing Systems Engineering, University of Windsor, Windsor, Ontario, Canada N9B 3P4. email: gzhang@uwindsor.ca

to develop IIPMs. Infeasible IPMs are widely adopted in many efficient software packages. IIPMs start with an arbitrary positive point and feasibility is reached as optimality is approached. The choice of the starting point in IIPMs is crucial for better performance. Lustig [7] and Tanabe [19] were the first to present IIPMs. Kojima at. al [5] prove the global convergence of an IIPM. Zhang [29] proved an $O(n^2L)$-iteration bound for IIPMs under certain conditions. Mizuno [9] introduced a primal-dual IIPM and proved global convergence of the algorithm.

In this paper we consider primal-dual LO problems in the following the standard form:

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

and the dual problem is given by

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

where $A \in R^{m\times n},\ b \in R^m,\ y \in R^m,\ c \in R^n,\ x, s, \in R^n$ and without loss of generality [16] we may assume that rank $(A) = m$. The vectors $y,\ x$ and s are the vectors of variables.

In this paper we adopt a new SR proximity based search direction to develop new SR proximity based IIPMs.

Our paper is organized as follows. In Section 2 we introduce an infeasible primal-dual self-regular proximity based IPM. Then, in Section 3 we will discus some properties of our SR proximity measure. In Section 4 we specify our new algorithm. In Section 5 we establish some technical results for further complexity analysis. Then, in Section 6 the polynomial complexity of our new algorithm is established. Some of our computational results are presented in Section 7. Conclusions are given in Section 8. To make the paper easily readable, we move most detailed proofs of the technical results to the Appendix.

2 Self-Regular Infeasible IPMs

2.1 Self-Regular Functions

In this section we recall the definition of self-regular functions [12, 14] and some of their properties. The family of univariate self-regular functions are defined as follows.

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

SR.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

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

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

$$\psi(t_1^r t_2^{1-r}) \leq r\psi(t_1) + (1-r)\psi(t_2), \quad \forall r \in [0, 1]. \tag{2}$$

If $\psi(t)$ is self-regular (SR), the parameter q is called the *barrier degree* and p the *growth degree* of the function $\psi(t)$.

There are two popular families of SR functions. The first family is given by

$$\Upsilon_{p,q}(t) = \frac{t^{p+1}-1}{p(p+1)} + \frac{t^{1-q}-1}{q(q-1)} + \frac{p-q}{pq}(t-1), \quad p,q \geq 1, \tag{3}$$

with $\nu_1 = \nu_2 = 1$. The second family is defined as

$$\Gamma_{p,q}(t) = \frac{t^{p+1}-1}{p+1} + \frac{t^{1-q}-1}{q-1}, \;\; p \geq 1, \; q > 1, \tag{4}$$

with $\nu_1 = 1$ and $\nu_2 = q$.

Let $v \in R^n_{++}$. Then an SR-proximity measure $\Psi : R^n_{++} \to R_+$ measures the discrepancy between v and $e = (1, 1, \cdots, 1)^T$, and is defined as $\Psi(v) = \sum_{i=1}^{n} \psi(v_i)$, where $\psi(t)$ is a univariate SR function, called the kernel function of the SR-proximity.

A new paradigim of IPMs is introduced by Peng et.al., in [12, 14], where SR-proximity measures are used to define search directions and to control the iterative process. In the rest of this section an general scheme of SR proximity based IIPMs is developed.

To solve the primal-dual problems (P) and (D), we need to solve the following optimality conditions

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

where xs denotes the coordinatewise (Hadamard) product of the two vectors. In the optimality conditions the first two constraints represent primal and dual feasibility, while the last one is the so-called complementary condition. Instead of solving the problem directly, IPMs relax the complementary condition by using a centrality parameter μ. The central path is defined as the set of unique solutions $\{(x(\mu)|\ \mu > 0\}$ and $\{(y(\mu), s(\mu))|\ \mu > 0\}$ of the system:

$$\begin{aligned} Ax \quad &= b \quad\; x > 0, \\ A^T y + s &= c \quad\; s > 0, \\ xs \quad &= \mu e, \end{aligned} \tag{5}$$

where $\mu > 0$ is a centrality parameter. System (5) is solved approximately by applying Newton's method for getting approximate solutions. As μ goes to zero, $x(\mu), y(\mu), s(\mu)$ converge to an optimal solution. Given any $x > 0$, y and $s > 0$, the Newton direction for (5) in IIPM is determined by the following linear system of equations:

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} -r_b \\ -r_c \\ \mu e - xs \end{bmatrix}, \tag{6}$$

where X and S denote the diagonal matrices diag(x) and diag(s), and r_b and r_c are residuals defined by

$$\begin{aligned} r_b &= Ax - b, \\ r_c &= A^T y + s - c. \end{aligned}$$

In SR-IIPMs the Newton system (6) is modified. Let

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

whose i^{th} components are $\sqrt{\frac{x_i s_i}{\mu}}$ and $\sqrt{\frac{\mu}{x_i s_i}}$, respectively. Then, the Newton system for SR-IIPM is given as:

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} -r_b \\ -r_c \\ -\mu v \nabla \Psi(v), \end{bmatrix}. \tag{7}$$

where $v\nabla\Psi(v) = (v_1 \nabla\psi(v_1), \cdots, v_n \nabla\psi(v_n))^T$. Observe that the right-hand-side of the last set of equations is exactly the same as in feasible SR-IPMs [12, 14]. We solve the Newton system as follows. ¿From the last equation of system (7), we derive

$$\Delta s = x^{-1}(-\mu v \nabla\Psi(v) - s\Delta x).$$

Then, we have the so-called augmented system:

$$\begin{bmatrix} 0 & A \\ A^T & -D^{-2} \end{bmatrix} \begin{bmatrix} \Delta y \\ \Delta x \end{bmatrix} = \begin{bmatrix} -r_b \\ r_h \end{bmatrix}, \tag{8}$$

where

$$D^2 = (X^{-1}S)^{-1},$$
$$r_h = -r_c + x^{-1}\mu v \nabla\Psi(v)).$$

The normal equation is derived by making a block-pivot on $-D^{-2}$:

$$AD^2A^T\Delta y = AD^2 r_h - r_b.$$

We obtain Δy by solving the normal equation and we get Δx by the formula

$$\Delta x = D^2(A^T \Delta y - r_h), \tag{9}$$

finally Δs is computed by using (8). Having the Newton direction, we make a damped step, with a certain step length α to get the new iterate.

$$x(\alpha) := x + \alpha\Delta x,$$
$$y(\alpha) := y + \alpha\Delta s,$$
$$s(\alpha) := s + \alpha\Delta y.$$

3 Properties of the Proximity Function

In this section we investigate the properties of a specific SR-proximity function

$$\Phi(x, s, \mu) = \Psi(v) = \frac{1}{2}\left\|v - v^{-1}\right\|^2,$$

with respect to the argument μ. Observe that $\Psi(v)$ is an SR-proximity function with the kernel function $\psi(t) = \frac{t^2}{2} - 1 + \frac{t^{-2}}{2}$. We are particularly interested in the case when the present iterate (x, s) is far away from the central path. For notational convenience, $\mu_g := \frac{x^T s}{n}$ denotes the parameter value associated with the current duality gap and $\mu_h := \frac{n}{x^{-T}s^{-1}}$ denotes the harmonic mean of $(x_1 s_1, \cdots, x_n s_n)$. Note that when the point (x, s) is fixed, then we can cast the proximity function $\Phi(x, s, \mu)$ as a function of μ, i.e.,

$$\Phi(x, s, \mu) := \frac{x^T s}{2\mu} - n + \frac{\mu}{2} x^{-T} s^{-1}.$$

If $\mu = \mu_g$, from the choice of the kernel function we know that the value of the proximity function is determined by the product

$$x^T s x^{-T} s^{-1} = \frac{n^2 \mu_g}{\mu_h}.$$

By using the Cauchy-Schwartz inequality and (7), one has

$$n^2 = (v^T v)^2 \leq \|v\|^2 \left\|v^{-1}\right\|^2 = x^T s x^{-T} s^{-1},$$

with equality if and only if (x, s) is on the central path. This result implies that $\mu_g \geq \mu_h$ and the inequality holds strictly when (x, s) is not on the central path. Next we consider the behavior of the function $\Phi(x, s, \mu)$ w.r.t. μ. Because both x and s are positive vectors, the function $\Phi(x, s, \mu)$ is convex with respect to μ. Using the optimality conditions for convex optimization problems, we can easily prove the following results.

Proposition 3.1 *For any fixed iterate $(x, s) > 0$, the proximity function $\Phi(x, s, \mu)$ as a function of μ has a global minimizer at the geometric mean of μ_g and μ_h*

$$\mu^* = \sqrt{\frac{x^T s}{x^{-T} s^{-1}}} = \sqrt{\mu_g \mu_h}.$$

It is easy to verify the following interesting relation that plays a crucial role later in the design of our algorithmic scheme.

Proposition 3.2 *Suppose that the iterate $(x, s) > 0$ is fixed. Then we have*

$$\Phi(x, s, \mu_g) = \Phi(x, s, \mu_h).$$

In fact, with simple calculus, we can obtain the following relation between the values of the proximity function w.r.t. μ^* and μ_g

$$\Phi(x, s, \mu_g) = \Phi(x, s, \mu^*) + \frac{\Phi(x, s, \mu^*)^2}{2n}. \tag{10}$$

Now, let us recall that when the primal-dual pair (x, s) is in a large neighborhood of the central path, then $\mu_g \gg \mu_h$ holds. Moreover, we can write

$$\mu_h := \left(1 - \frac{\mu_g - \mu_h}{\mu_g}\right) \mu_g = (1 - \theta)\mu_g,$$

where $\theta > \frac{1}{2}$ when $\mu_g > 2\mu_h$. This, considering $\Phi(x, s, \mu_g) = \Phi(x, s, \mu_g)$, leads naturally to a large-update method. In practical implementations of IPMs, a large-update is used whenever the iterate is in a certain neighborhood of the central path or, equivalently, when the proximity function $\Phi(x, s, \mu_g)$, or the ratio $\frac{\mu_g}{\mu_h}$ is bounded above by a certain number $\tau \gg 1$. Thus, after one update of $\mu := \frac{\mu}{\tau}$, we need also to investigate the growth behavior of the proximity function, which is demonstrated by the following lemma [13]. For ease of understanding the proof is included.

Lemma 3.3 *Let $\tau > 1$ be a constant. If*

$$\frac{\mu_g}{\mu_h} = \frac{x^T s x^{-T} s^{-1}}{n^2} \leq \tau,$$

then

$$\Phi\left(x, s, \frac{\mu_g}{\tau}\right) \leq \frac{(\tau - 1)n}{2}.$$

Proof: By the assumption of the lemma we can write $\mu_g = \bar{\tau}\mu_h$ for some $\bar{\tau} \leq \tau$. It follows that

$$\begin{aligned} \Phi\left(x, s, \frac{\mu_g}{\tau}\right) &= \frac{\tau n}{2} - n + \frac{n\mu_g}{2\tau\mu_h} \\ &= \frac{(\tau - 1)n}{2} - \frac{n}{2} + \frac{n\bar{\tau}}{2\tau} \leq \frac{(\tau - 1)n}{2}, \end{aligned}$$

which further concludes the lemma. □

In fact, for any positive $\tau > 1$, one can see that $\Phi(x, s, \mu) = \frac{(\tau-1)n}{2}$ if and only if the target $\mu = \mu_t$ satisfies

$$\mu_t := \frac{2x^T s}{(\tau + 1)n + \sqrt{(\tau + 1)^2 n^2 - 4x^T s \sum_{i \in \mathcal{I}} x_i^{-1} s_i^{-1}}} = \frac{2\mu_g}{\tau + 1 + \sqrt{(\tau + 1)^2 - 4\mu_g/\mu_h}}. \quad (11)$$

It is also easy to verify that μ_t can be cast as a decreasing function of τ. In particular, $\mu_t = \mu_h$ if and only if $\mu_g = \tau\mu_h$ and $\mu_h > \mu_t$ whenever $\mu_g < \tau\mu_h$. Now we proceed to discuss the properties of the search direction based on our specific self-regular proximity function for different updates of μ. Note that, due to the specific choice of the kernel function $\psi(t)$, we can rewrite the Newton system (7) in the original space as

$$\begin{aligned} A\Delta x &= -r_b, \\ A^T \Delta y + \Delta s &= -r_c, \\ s\Delta x + x\Delta s &= \mu^2 x^{-1} s^{-1} - xs. \end{aligned} \quad (12)$$

Let us denote by $(\Delta x(\mu), \Delta y(\mu), \Delta s(\mu))$ the solution of system (12). The following lemma discusses the change of the duality gap along the search direction $(\Delta x(\mu), \Delta s(\mu))$ for $\mu = \mu_h$.

Lemma 3.4 *Let $(\Delta x(\mu_h), \Delta s(\mu_h))$ be the solution of system (12) with $\mu = \mu_h$. Then we have*

$$x^T \Delta s(\mu_h) + s^T \Delta x(\mu_h) = \frac{n^2}{x^{-T} s^{-1}} - x^T s = n(\mu_h - \mu_g).$$

Recall that in traditional IPMs based on the standard Newton direction, we need to solve equation system (6) at each iteration. In this case, if we set the target to $\mu_+ = \mu_h$, then the solution of system (6) will satisfy

$$x^T \Delta s + s^T \Delta x = n\mu_h - x^T s = n(\mu_h - \mu_g) < 0.$$

This implies that if the targeted parameter is μ_h, then the search direction based on our specific self-regular proximity function and the standard Newton direction will predict the change of the duality gap in the same way, i.e.,

$$(x + \alpha\Delta x(\mu_h))^T(s + \alpha\Delta s(\mu_h)) = (x + \alpha\Delta x)^T(s + \alpha\Delta s) = x^T s\left(1 - \alpha + \frac{\mu_h\alpha}{\mu_g} + \alpha^2\frac{\Delta x^T\Delta s}{x^T s}\right).$$

4 A Dynamic Large-Update Infeasible IPM

In this section we introduce our new algorithm. Our algorithm works with an infeasible central path neighborhood $\mathcal{N}(\tau, \beta)$. Although the neighborhood can be defined for any SR-proximity function, in this paper we choose the particular case $\Psi(v) = \frac{1}{2}\left\|v - v^{-1}\right\|^2$. This neighborhood with an SR-proximity $\Phi(x, s, \mu) = \Psi(v)$ is defined by

$$\mathcal{N}(\tau, \beta) = \left\{(x, y, s) \mid \Psi(v) \le \frac{(\tau - 1)n}{2},\ \|r_b\| \le \left\|r_b^0\right\|\frac{\mu_g}{\mu^0}\beta,\ \|r_c\| \le \left\|r_c^0\right\|\frac{\mu_g}{\mu^0}\beta\right\}, \tag{13}$$

where, $\mu_g = \frac{x^T s}{n}$, $\mu^0 = \mu_g^0$, (x^0, y^0, s^0) is an arbitrary triple with x^0, s^0 strictly positive and $\beta \ge 1$ so that the initial point (x^0, y^0, s^0) belongs to the neighborhood $\mathcal{N}(\tau, \beta)$. If $(x, y, s) \in \mathcal{N}(\tau, \beta)$, then infeasibility is bounded by a multiple of μ and the initial infeasibility.

We utilize a parameter $\tau \ge 10$ to keep control on the update of the duality gap parameter μ and to force the value of the proximity function to satisfy the relation

$$\Phi(x, s, \mu_g) = \Phi(x, s, \mu_h) \le \frac{(\tau - 1)n}{2}. \tag{14}$$

We also stipulate that when the proximity function $\Phi(x, s, \mu_g)$ has a relatively small value, for instance, if $\Phi(x, s, \mu_g) = \Phi(x, s, \mu_h) \le \frac{(\tau-2)n}{4}$, then we choose μ_t defined by (11) as our targeted centering parameter in system (12) for the search direction. Otherwise, we choose μ_h as the targeted parameter. Moreover, the step size is carefully chosen so that all the iterates remain in the neighborhood $\mathcal{N}(\tau, \beta)$. Note that in both cases, we have $\mu_g \ge \frac{\tau}{2}\mu_+$, where μ_+ is the targeted parameter. Hence, our algorithm is indeed a large-update one if $\tau \ge 10$. For simplicity we use the notation $x(\alpha) := x + \alpha\Delta x$, $y(\alpha) := y + \alpha\Delta y$ and $s(\alpha) := s + \alpha\Delta s$. Correspondingly we also define

$$\mu_g(\alpha) = \frac{x(\alpha)^T s(\alpha)}{n}, \quad \mu_h(\alpha) = \frac{n}{x(\alpha)^{-T} s(\alpha)^{-1}}, \quad \mu^*(\alpha) = \sqrt{\mu_g(\alpha)\mu_h(\alpha)}.$$

At each step, we stipulate that the step size should be chosen such that the proximity function $\Phi(x(\alpha), s(\alpha), \mu_t^+)$ has a sufficient decrease so that after the step the proximity function still satisfies (14) at the new iterate. The algorithm can be outlined as follows.

Algorithm SR-IIPM

Input:
Proximity parameters $\tau \geq 10$ and $\beta \geq 1$;
neighborhood $\mathcal{N}(\tau, \beta)$;
an accuracy parameter $\varepsilon > 0$;
$(x^0, y^0, s^0) \in \mathcal{N}(\tau, \beta)$; $k = 0$.
begin
while $\max\left\{(x^k)^T s^k, \left\|r_b^k\right\|, \left\|r_c^k\right\|\right\} \geq \epsilon$ **do**
begin
If $\frac{\mu_g^k}{\mu_h^k} \geq \frac{\tau}{2}$ then $\mu := \mu_h^k$;
otherwise $\mu := \mu_t^k$ given by (11).
Solve system (12) for $\Delta x^k, \Delta y^k, \Delta s^k$.
begin
Determine a step size α_k such that
$\Phi(x(\alpha_k), s(\alpha_k), \mu_t^k) \leq \Phi(x^k, s^k, \mu_t^k) - \frac{\alpha^*}{2}\Phi(x, s, \mu_t)$
and $(x(\alpha_k), y(\alpha_k), s(\alpha_k)) \in \mathcal{N}(\tau, \beta)$;
$x^{k+1} := x(\alpha_k)$; $y^{k+1} := y(\alpha_k)$; $s^{k+1} := s(\alpha_k)$;
$k = k + 1$.
end
end
end

For future use we introduce the scalar quantity ρ_k defined by

$$\rho_k = \prod_{i=0}^{k-1}(1 - \alpha_i).$$

Because the first two components of system (12) are linear, we have

$$\begin{aligned}(r_b^k, r_c^k) &= (1 - \alpha_{k-1})(r_b^{k-1}, r_c^{k-1}) \\ &= (1 - \alpha_{k-1}) \cdots (1 - \alpha_0)(r_b^0, r_c^0) \\ &= \rho_k(r_b^0, r_c^0).\end{aligned} \tag{15}$$

Since $(x^k, y^k, s^k) \in \mathcal{N}(\tau, \beta)$, from (15) we derive

$$\rho_k \frac{\|(r_b^0, r_c^0)\|}{\mu_g^k} = \frac{\left\|(r_b^k, r_c^k)\right\|}{\mu_g^k} \leq \beta \frac{\|(r_b^0, r_c^0)\|}{\mu^0}.$$

Provided that $(r_b^0, r_c^0) \neq 0$, it follows from this inequality that

$$\rho_k \leq \frac{\beta \mu_g^k}{\mu^0}. \tag{16}$$

Polynomial complexity of Algorithm SR-IIPM follows from the result that the lower bound on step size is an inverse polynomial function of n if we choose the starting point to be

$$(x^0, y^0, s^0) = (\zeta e, 0, \zeta e), \tag{17}$$

where ζ is scalar for which

$$\|(x^*, s^*)\|_\infty \leq \zeta \tag{18}$$

for some primal-dual optimal solution (x^*, y^*, s^*). Usually we don't know $\|(x^*, s^*)\|_\infty$, because we do not know any optimal solutions a priori. However, these conditions are still relevant. Theoretically we can choose $\zeta = O(2^L)$ where L is the input length of the LO problem [16]. Such an initial point with sufficiently large ζ is helpful for computational practice for the following reason: a well-centered starting point for which the ratio

$$\frac{\|(r_b^0, r_c^0)\|}{\mu^0} \tag{19}$$

is small leads to faster convergence than do poorly centered points that are much closer to the solution set. The point (17) satisfies these criteria. It is perfectly centered, and the ratio (19) is bounded above.

5 Preliminary Technical Results

In contrast to feasible SR-proximity based IPMs given in [12], the orthogonality property of Δx, and Δs does not hold in IIPMs, i.e., $\Delta x^T \Delta s \neq 0$. In this section we estimate the second order term $\frac{\Delta x^T \Delta s}{\mu}$.

5.1 Technical Results I: Bounding $\nu_k \|(x, s)\|$

The following orthogonality property is useful in several places in the analysis. If $(\bar{x}, \bar{y}, \bar{s})$ is any vector that satisfies the conditions

$$A\bar{x} = 0, \quad A^T\bar{y} + \bar{s} = 0, \tag{20}$$

then

$$\bar{x}^T\bar{s} = -\bar{x}^T A^T \bar{y} = 0. \tag{21}$$

The first result establishes a bound on $\rho_k \left\|(x^k, s^k)\right\|$.

Lemma 5.1 *Suppose that the initial point is chosen to satisfy (17) and (18). Then, there is a positive constant C_1 such that for all $k \geq 0$*

$$\zeta\rho_k \left\|(x^k, s^k)\right\|_1 \leq C_1 n \mu_g^k.$$

Proof: Let (x^*, y^*, s^*) be a primal-dual optimal solution and $(\bar{x}, \bar{y}, \bar{s})$ be the vector defined by

$$(\bar{x}, \bar{y}, \bar{s}) = \rho_k(x^0, y^0, s^0) + (1 - \rho_k)(x^*, y^*, s^*) - (x, y, s).$$

It is not hard to check that $(\bar{x}, \bar{y}, \bar{s})$ satisfies (20). Hence, from (21), we have

$$\begin{aligned}
0 = \bar{x}^T\bar{s} &= (\rho_k x^0 + (1-\rho_k)x^* - x^k)^T(\rho_k s^0 + (1-\rho_k)s^* - s^k) \\
&= \rho_k^2 (x^0)^T s^0 + (1-\rho_k)^2 (x^*)^T s^* + \rho_k(1-\rho_k)((x^0)^T s^* + (s^0)^T x^*) \\
&\quad + (x^k)^T s^k - \nu_k((s^k)^T x^0 + (x^k)^T s^0) - (1-\rho_k)((s^k)^T x^* + (x^k)^T s^*).
\end{aligned}$$

Since $(s^k)^T x^* + (x^k)^T s^* \geq 0$, and $(x^*)^T s^* \ = \ 0$, we have

$$\rho_k((s^k)^T x^0 + (x^k)^T s^0) \leq \rho_k^2 (x^0)^T s^0 + (x^k)^T s^k + \rho_k(1-\rho_k)((x^0)^T s^* + (s^0)^T x^*). \tag{22}$$

Let us now define the constant ξ by

$$\xi = \min(x^0_{\min}, s^0_{\min}) > 0,$$

where $x^0_{min} = \min_{i=1,\cdots,n} x_i^0,\ s^0_{min} = \min_{i=1,\cdots,n} s_i^0$. From (17) and (18) we have $\zeta = \xi$. Since for any iteration we have $(x^k, s^k) > 0$, one can verify that

$$\zeta \left\|(x^k, s^k)\right\|_1 \leq (s^0_{\min}) \left\|x^k\right\|_1 + (x^0_{\min}) \left\|s^k\right\|_1 \leq (x^k)^T s^0 + (s^k)^T x^0.$$

Now from (22), and using $\rho_k \in (0,1)$ we have

$$\begin{aligned} \rho_k \zeta \left\|(x^k, s^k)\right\|_1 &\leq \rho_k^2 n\mu^0 + (x^k)^T s^k + \rho_k(1-\rho_k)\left(\left\|x^0\right\|_\infty \|s^*\|_1 + \left\|s^0\right\|_\infty \|x^*\|_1\right) \\ &\leq \rho_k n\mu^0 + (x^k)^T s^k + \rho_k \left\|(x^0, s^0)\right\|_\infty \|(x^*, s^*)\|_1 . \end{aligned} \tag{23}$$

Form (16), by our choice of ζ and (x^0, y^0, s^0), we also have

$$\begin{aligned} &\left\|(x^0, s^0)\right\|_\infty = \zeta, \\ &\|(x^*, s^*)\| \leq \sqrt{2n}\, \|(x^*, s^*)\|_\infty \leq \sqrt{2n}\zeta, \\ &\mu^0 = \frac{(x^0)^T s^0}{n} = \zeta^2. \end{aligned} \tag{24}$$

Substituting these values into (23) and using $\beta \geq 1$ and (16), we obtain

$$\zeta\rho_k \left\|(x^k, s^k)\right\|_1 \leq \beta\mu_g^k n + (x^k)^T s^k + \sqrt{2n}\beta\mu_g^k.$$

¿From (3.17) of [12] with $p = 1$, $q = 3$ and $\nu_1 = 1$ we have

$$\begin{aligned} (x^k)^T s^k &= \mu^k \left\|v^k\right\|^2 \\ &\leq n\mu_k + 2\mu^k \sqrt{\frac{2n\Psi(v^k)}{\nu_1}} + 2\frac{\Psi(v^k)\mu^k}{\nu_1} \\ &\leq n\mu^k(\tau + 2\sqrt{\tau - 1}), \end{aligned} \tag{25}$$

where μ_k is the target value at this iteration. Regardless if $\mu^k = \mu_t^k$ or μ_h^k, in both cases $\mu^k \leq \mu_g^k$.

We can further estimate (23) by using (25) and (24). We obtain

$$\zeta\rho_k \left\|(x^k, s^k)\right\|_1 \ \leq \beta\mu_g^k n + n\mu_g^k(\tau + 2\sqrt{\tau-1}) + 2n\beta\mu_g^k \leq C_1 n\mu_g^k,$$

where $C_1 = 3\beta + \tau + 2\sqrt{\tau - 1}$. □

5.2 Technical Results II: Bounding $\|\nabla\Psi(v)\|$

We introduce the following notations $d_x := \dfrac{v\Delta x}{x}, \quad d_s := \dfrac{v\Delta s}{s}$ and define

$$\sigma := \|-\nabla\Psi(v)\| = \|d_x + d_s\|$$

and

$$\sigma_1^2 := \|d_x\|^2 + \|d_s\|^2.$$

Then, we have

$$\sigma^2 = \|-\nabla\Psi(v)\|^2 = \|d_x\|^2 + \|d_s\|^2 + 2d_x^T d_s = \sigma_1^2 + 2d_x^T d_s = \sigma_1^2 + 2\frac{\Delta x^T \Delta s}{\mu}. \tag{26}$$

The following lemma gives an upper bound for σ.

Lemma 5.2 *If* $\Psi(v) \leq \frac{(\tau-1)n}{2}$ *then* $\sigma^2 \leq (\tau n + 1)^3$.

Proof: An upper bound for the optimal value of the problem

$$\begin{aligned} \max \;\; & \sigma^2 \\ \text{s.t.} \;\; & \Psi(v) \leq \frac{(\tau-1)n}{2} \end{aligned}$$

is $(\tau n + 1)^3$.

This can be proved by examining the KKT conditions of the problem

$$\begin{aligned} \psi'(v_i)[\psi''(v_i) - \lambda] &= 0, \quad i = 1, ..., n, \\ \Psi(v) &\leq \frac{(\tau-1)n}{2}, \\ \lambda\left(\Psi(v) - \frac{(\tau-1)n}{2}\right) &= 0. \end{aligned} \tag{27}$$

From the third condition we see that $\lambda = 0$ if and only if $v = e$, $\sigma = 0$ and $\psi'(v_i) = \psi''(v_i) = 0$. In this case all the conditions of (27) are satisfied. On the other hand, due to the choice of $\psi(v)$ we have that

$$\psi''(v_i) = 1 + 3v_i^{-4}$$

is a strictly decreasing convex function. If $v \neq e$ and $\lambda \neq 0$ then the following two cases may happen:
(i) $v_1 = v_2 = ... = v_{n-k} = 1, v_{n-k+1} = v_{n-k+2} = ... = v_n > 1$,
(ii) $v_1 = v_2 = ... = v_{n-k} = 1, v_{n-k+1} = v_{n-k+2} = ... = v_n < 1$.
For case (i) we have

$$\frac{kv_n^2 - k}{2} - \frac{n}{2} \leq \frac{kv_n^2 - k}{2} + \frac{kv_n^{-2} - k}{2} = \Psi(v) = \frac{(\tau-1)n}{2},$$

that imply $v_n \leq \sqrt{\frac{\tau n + k}{k}} \leq \sqrt{\tau n + 1}$. Now

$$\sigma^2 = \sum_{i=1}^{n} (v_i - v_i^{-3})^2 = k(v_n - v_n^{-3})^2 \leq kv_n^2 \leq k\frac{\tau n + k}{k} \leq (\tau + 1)n.$$

For case (ii) we have

$$\frac{kv_n^{-2} - k}{2} - \frac{n}{2} \leq \frac{kv_n^2 - k}{2} + \frac{kv_n^{-2} - k}{2} = \Psi(v) = \frac{(\tau-1)n}{2},$$

that imply $v_n^{-2} \leq \frac{\tau n + k}{k}$. Now

$$\sigma^2 = \sum_{i=1}^{n} (v_i - v_i^{-3})^2 = k(v_n - v_n^{-3})^2 \leq kv_n^{-6} \leq k(\frac{\tau n + k}{k})^3 \leq (\tau n + 1)^3.$$

The proof is completed. □

5.3 Technical Results III: Bounding $(D^k)^{-1}\Delta x^k$ and $D^k\Delta s^k$

Consistent with the notation introduced at the augmented system (8), we define D^k as:

$$D^k = (X^k)^{\frac{1}{2}}(S^k)^{-\frac{1}{2}} = \operatorname{diag}((x^k)^{\frac{1}{2}}(s^k)^{-\frac{1}{2}}).$$

We also make repeated use of matrix norms defined for a matrix as $M \in R^{p\times q}$

$$\|M\| = \max_{u\in R^q:\|u\|=1} \|Mu\|.$$

This definition holds for all the three norms $\|.\|_1$, $\|.\|_2$ and $\|.\|_\infty$. A simple consequence of this definition is that

$$\|Mu\| \le \|M\|\,\|u\|, \text{ for all } \ u \in R^q.$$

The next lemma gives bounds on the scaled vectors $(D^k)^{-1}\Delta x^k$ and $D^k\Delta s^k$.

Lemma 5.3 *Suppose that the starting point is chosen as in (17) and (18). Then there is a constant C_2 independent of n, such that for all $k \ge 0$*

$$\left\|(D^k)^{-1}\Delta x^k\right\| \le C_2\sqrt{\mu_g^k}\; n\sigma^{\frac{1}{3}} \;\; and \;\; \left\|D^k\Delta s^k\right\| \le C_2\sqrt{\mu_g^k}\; n\sigma^{\frac{1}{3}}. \tag{28}$$

Proof: Let us define

$$(\bar x, \bar y, \bar s) = (\Delta x^k, \Delta y^k, \Delta s^k) + \rho_k(x^0, y^0, s^0) - \rho_k(x^*, y^*, s^*).$$

It is easy to verify that $(\bar x, \bar y, \bar s)$ satisfies (20), thus from (21) we have

$$[\Delta x^k + \rho_k(x^0 - x^*)]^T[\Delta s^k + \rho_k(s^0 - s^*)] = 0. \tag{29}$$

From $s^k\Delta x^k + x^k\Delta s^k = -\mu^k v^k\nabla\Psi(v^k)$, we have

$$s^k(\Delta x^k + \rho_k(x^0 - x^*) + x^k(\Delta s^k + \rho_k(s^0 - s^*)) \; = -\mu^k v^k\nabla\Psi(v^k) + \rho_k s^k(x^0 - x^*) + \rho_k x^k(s^0 - s^*).$$

If we multiply this system by $(X^kS^k)^{-\frac{1}{2}}$ and note that $(X^k)^{-\frac{1}{2}}(S^k)^{\frac{1}{2}} = (D^k)^{-1}$ and $(X^k)^{\frac{1}{2}}(S^k)^{-\frac{1}{2}} = D^k$, then we have

$$\begin{aligned}(D^k)^{-1}((D^kx^k + \rho_k(x^0 - x^*)) \;\; &+D^k(\Delta s^k + \rho_k(s^0 - s^*)) = \\ &\sqrt{\mu^k}\nabla\Psi(v^k) + \rho_k(D^k)^{-1}(x^0 - x^*) + \rho_k D^k(s^0 - s^*).\end{aligned} \tag{30}$$

Now, because of (29), we have

$$\begin{aligned}&\left\|(D^k)^{-1}(\Delta x^k + \rho_k(x^0 - x^*)) + D^k(\Delta s^k + \rho_k(s^0 - s^*))\right\|^2 \\ &\quad = \left\|(D^k)^{-1}(\Delta x^k + \rho_k(x^0 - x^*))\right\|^2 + \left\|D^k(\Delta s^k + \rho_k(s^0 - s^*))\right\|^2.\end{aligned} \tag{31}$$

Taking squared norms of both sides in (30), and using (28) and (31) we have

$$\begin{aligned}&\left\|(D^k)^{-1}(\Delta x^k + \rho_k(x^0 - x^*))\right\|^2 + \left\|D^k(\Delta s^k + \rho_k(s^0 - s^*))\right\|^2 \\ &\quad \le \left(\left\|\sqrt{\mu^k}\nabla\Psi(v^k)\right\| + \rho_k\left\|(D^k)^{-1}(x^0 - x^*)\right\| + \rho_k\left\|D^k(s^0 - s^*)\right\|\right)^2.\end{aligned} \tag{32}$$

Let us now isolate the first term on the left-hand-side and write

$$\left\|(D^k)^{-1}(\Delta x^k + \rho_k(x^0 - x^*))\right\| \leq \ \|\sqrt{\mu_k}\nabla\Psi(v)\| + \rho_k \left\|(D^k)^{-1}(x^0 - x^*)\right\| + \rho_k \left\|D^k(s^0 - s^*)\right\|. \quad (33)$$

A simple application of the triangle inequality and addition of an extra term $\rho_k \left\|D^k(s^0 - s^*)\right\|$ to the right-hand-side give

$$\left\|(D^k)^{-1}\Delta x^k\right\| \leq \left\|\sqrt{\mu^k}\nabla\Psi(v^k)\right\| + 2\rho_k \left\|(D^k)^{-1}(x^0 - x^*)\right\| + 2\rho_k \left\|D^k(s^0 - s^*)\right\|. \quad (34)$$

Next, we show that each term on the right-hand-side of (33) is $O(\sqrt{\mu^k})$ in magnitude. We have

$$\left\|(X^kS^k)^{\frac{1}{2}}\right\| = \max_{i=1,\cdots,n} \frac{1}{(x_i^k s_i^k)^{-\frac{1}{2}}} = \frac{1}{\min\limits_{i=1,\cdots,n} (x_i^k s_i^k)^{\frac{1}{2}}}.$$

Since $v_{\min} \geq (1+3\sigma)^{-\frac{1}{3}}$, i.e., $\min\limits_{i=1,\cdots,n} \sqrt{\frac{x_i^k s_i^k}{\mu^k}} \geq (1+3\sigma)^{-\frac{1}{3}}$ then by using (3.11) of [12], we have

$$\left\|(X^kS^k)^{-\frac{1}{2}}\right\| \leq (\mu^k)^{-\frac{1}{2}}(1+3\sigma)^{\frac{1}{3}} \leq \sqrt{\frac{\tau+1}{\mu_g^k}}(1+3\sigma)^{\frac{1}{3}}. \quad (35)$$

From the matrix norm $\|(D)^{-1}\|$, we have

$$\begin{aligned}\left\|(D^k)^{-1}\right\| &= \max_{i=1,\cdots,n} \left\|(D_{ii}^k)^{-1}\right\| \\ &= \left\|(D^k)^{-1}e\right\|_\infty = \left\|(X^kS^k)^{-\frac{1}{2}}S^ke\right\|_\infty \\ &\leq \left\|(X^kS^k)^{-\frac{1}{2}}\right\| \left\|s^k\right\|_1, \end{aligned} \quad (36)$$

and similarly:

$$\left\|D^k\right\| \leq \left\|(X^kS^k)^{-\frac{1}{2}}\right\| \left\|x^k\right\|_1.$$

For the last two terms in (34), we have

$$\begin{aligned}\rho_k \left\|(D^k)^{-1}(x^0 - x^*)\right\| + \rho_k \left\|D^k(s^0 - s^*)\right\| &\leq \rho_k\zeta\left(\left\|(D^k)^{-1}e\right\| + \left\|D^ke\right\|\right) \\ &= \rho_k\zeta\left(\left\|(X^kS^k)^{-\frac{1}{2}}s^k\right\| + \left\|(X^kS^k)^{-\frac{1}{2}}x^k\right\|\right) \\ &\leq \rho_k\zeta\left\|(X^kS^k)^{-\frac{1}{2}}\right\| \left\|(x^k, s^k)\right\|_1.\end{aligned}$$

From (35), (36), (37) and Lemma 5.1 we have

$$\rho_k \left\|(D^k)^{-1}(x^0 - x^*)\right\| + \ \rho_k \left\|D^k(s^0 - s^*)\right\| \leq C_1 n \sqrt{\mu_g^k}\sqrt{\tau+1}(1+3\sigma)^{\frac{1}{3}}. \quad (37)$$

Finally, by combining (33) and (37)we obtain the bound on $\left\|(D^k)^{-1}\Delta x^k\right\|$ in (34)

$$\begin{aligned}\left\|(D^k)^{-1}\Delta x^k\right\| &\leq \sqrt{\mu_g^k}\sigma + 2C_1\sqrt{\mu_g^k}\ n(1+3\sigma)^{\frac{1}{3}}\sqrt{\tau+1} \\ &\leq C_2\sqrt{\mu_g^k}\ n\sigma^{\frac{1}{3}},\end{aligned} \quad (38)$$

where $C_2 = 5C_1\sqrt{\tau+1}$. One can analogously derive

$$\left\|(D^k)^{-1}\Delta s^k\right\| \le C_2\sqrt{\mu_g^k}\; n\sigma^{\frac{1}{3}},$$

that completes the proof of the lemma. □ Now we are ready to derive a bound for σ_1. From (28) we have

$$(\Delta x^k)^T\Delta s^k = ((D^k)^{-1}\Delta x^k)^T(D^k\Delta s^k) \le \left\|(D^k)^{-1}\Delta x^k\right\|\left\|D^k\Delta s^k\right\| \le C_2^2\mu_g^k n^2\sigma^{\frac{2}{3}},$$

that gives

$$\frac{(\Delta x^k)^T\Delta s^k}{\mu_g^k} \le C_2^2 n^2\sigma^{\frac{2}{3}}. \tag{39}$$

Since

$$\sigma_1^2 = \sigma^2 - 2\frac{(\Delta x^k)^T\Delta s^k}{\mu^k} \le \sigma^2 + 2(\tau+1)C_2^2n^2\sigma^{\frac{2}{3}},$$

where μ^k is the target value, we have

$$\sigma^2 - 2(\tau+1)C_2^2n^2\sigma^{\frac{2}{3}} \le \sigma_1^2 \le \sigma^2 + 2(\tau+1)C_2^2n^2\sigma^{\frac{2}{3}} \le C_3n^2\sigma^{\frac{2}{3}}, \tag{40}$$

where $C_3 = 3(\tau+1)C_2^2 = 75(\tau+1)^2(3\beta+\tau+2\sqrt{\tau-1})^2$.

6 Complexity analysis

In this section we derive a polynomial upper bound for the number of iterations of Algorithm SR-IIPM with a strictly positive step size α.

Theorem 6.1 *Let $\tau \ge 10$, $\frac{\mu_g^k}{\mu_h^k} < \frac{\tau}{2}$ and $(\Delta x, \Delta y, \Delta s)$ be the solution of system (12) with $\mu = \mu_t^k$. Then the maximal feasible step size $\alpha_{\max}$ satisfies*

$$\alpha_{\max} \ge \frac{9}{10}\sigma^{\frac{-1}{3}}\sigma_1^{-1}.$$

Moreover, for any step size $\alpha \le \alpha_1^* = \frac{3}{10}\frac{\sigma^{\frac{2}{3}}\sigma_1^{-1}}{\sigma+4\sigma_1}$, the following relation holds:

$$\Phi(x(\alpha), s(\alpha), \mu_t^k) \le \Phi(x, s, \mu_t^k) - \frac{\alpha}{4}\sigma^2.$$

Proof: See Appendix. □

We proceed to estimate the proximity function $\Phi(x(\alpha), s(\alpha), \mu_g(\alpha))$ or, equivalently, the function $\Phi(x(\alpha), s(\alpha), \mu^*(\alpha))$ for a feasible step size when μ_t is used in the algorithm as the targeted parameter. Due to the inequality $\Phi(x(\alpha), s(\alpha), \mu^*(\alpha)) \le \Phi(x(\alpha), s(\alpha), \mu^*)$, it suffices to consider the function $\Phi(x(\alpha), s(\alpha), \mu^*)$.

Theorem 6.2 *Let $\tau \ge 10$, $\frac{\mu_g^k}{\mu_h^k} < \frac{\tau}{2}$ and $(\Delta x, \Delta y, \Delta s)$ be the solution of system (12) with $\mu = \mu_t^k$. Then the step size $\alpha_t^* = \min\left(\frac{1}{C_2\sqrt{\tau+1}n\sigma^{\frac{1}{3}}}, \frac{1}{9}\sigma^{-\frac{1}{3}}\sigma_1^{-1}\right)$ is strictly feasible. Moreover, for any step size $\alpha \le \alpha_t^*$ the following relation hold:*

$$\Phi(x(\alpha), s(\alpha), \mu_g(\alpha)) \le \Phi(x, s, \mu_t^k). \tag{41}$$

Proof: See Appendix. □

It remains to consider the behaviors of $\Phi(x(\alpha), s(\alpha), \mu_t^k)$ and $\Phi(x(\alpha), s(\alpha), \mu_g(\alpha))$ when μ_h is used as the targeted duality gap parameter in Algorithm SR-IIPM.

Theorem 6.3 *Let $\tau \geq 10$, $\frac{\tau}{2} \leq \frac{\mu_g^k}{\mu_h^k} \leq \tau$ and $(\Delta x, \Delta y, \Delta s)$ be the solution of system (12) with $\mu = \mu_h^k$. Then the step size $\alpha_h^* = \min\left(\frac{\sigma^{-\frac{2}{3}}}{4C_2^2 n}, \frac{3}{10}\frac{\sigma^{\frac{2}{3}}\sigma_1^{-1}}{\sigma + 4\tau\sigma_1}\right)$ is strictly feasible. Moreover, for any step size $\alpha \leq \alpha_h^*$, we have*

$$\Phi(x(\alpha), s(\alpha), \mu_g(\alpha)) \leq \Phi(x, s, \mu_t^k), \tag{42}$$

$$\Phi(x(\alpha), s(\alpha), \mu_t) \leq \Phi(x, s, \mu_t^k) - \frac{\alpha}{2}\Phi(x, s, \mu_t^k). \tag{43}$$

Proof: See Appendix. □

For the complexity analysis of Algorithm SR-IIPM one need to have an estimate of σ_1. When $d_x^T d_s \geq 0$, then from (26) we have $\sigma_1 \leq \sigma$, but when $d_x^T ds < 0$ then $\sigma_1 > \sigma$ and form (40) we have a bigger upper bound. Combining the results of Theorems 6.1, 6.2 and 6.3 we have the following conclusion.

Corollary 6.4 *Let $\tau \geq 10$, and $(\Delta x, \Delta y, \Delta s)$ be the solution of system (12) with $\mu = \mu_t^k$ if $\frac{\mu_g^k}{\mu_h^k} < \frac{\tau}{2}$, or $\mu = \mu_h^k$ if $\frac{\tau}{2} \leq \frac{\mu_g^k}{\mu_h^k} \leq \tau$. Then the step size $\alpha^* = \frac{3}{10}\frac{1}{C_3 n^2(1+4\tau)}$, is strictly feasible and*

$$\Phi(x(\alpha^*), s(\alpha^*), \mu_t^k) \leq \Phi(x, s, \mu_t^k) - \frac{\alpha^*}{2}\Phi(x, s, \mu_t^k). \tag{44}$$

$$\mu_g^{k+1} \geq (1 - \alpha^*)\mu_g^k. \tag{45}$$

Proof: If $d_x^T d_s \geq 0$, then $\sigma_1 \leq \sigma$ and from Theorem 6.1 we have $\alpha_1^* \geq \frac{3}{50(\tau+1)^2 n^2}$, form Theorem 6.2 we have $\alpha_t^* \geq \frac{1}{9(\tau+1)^3 n^2}$, and from Theorem 6.3 we have $\alpha_h^* \geq \frac{1}{8C_2^2(\tau+1)n^2}$. Now if $\alpha^* = \min(\alpha_1^*, \alpha_t^*, \alpha_h^*)$ then $\alpha^* \geq \frac{1}{8C_2^2(\tau+1)n^2}$. If in at least one iteration $d_x^T d_s \leq 0$, then $\sigma_1^2 \geq \sigma^2$. From Theorem 6.1 we have $\alpha_1^* \geq \frac{3}{50}\sigma^{\frac{2}{3}}\sigma_1^{-2} \geq \frac{3}{50C_3 n^2}$, from Theorem 6.2 and (40) we have $\alpha_t^* \geq \frac{1}{2(\tau+1)^{\frac{3}{2}} C_2 n^2}$ and from Theorem 6.3 and (40) we have $\alpha_h^* \geq \frac{3}{10}\frac{1}{C_3 n^2(1+4\tau)}$, then

$$\alpha^* = \min(\alpha_1^*, \alpha_t^*, \alpha_h^*) \geq \frac{3}{10}\frac{1}{C_3 n^2(1+4\tau)}.$$

For (45), when target value is μ_t^k we have

$$\mu_g^{k+1} = \mu_g^k\left(1 - \alpha + \alpha\frac{(\mu_t^k)^2}{\mu_g^k\mu_h^k} + \alpha^2\frac{\Delta x^T\Delta s}{n\mu_g^k}\right)$$

and thus it suffices to show that

$$\frac{(\mu_t^k)^2}{\mu_g^k\mu_h^k} + \alpha\frac{\Delta x^T\Delta s}{n\mu_g^k} \geq 0. \tag{46}$$

Using the definition μ_t^k and (39) one can easily derive that (46) holds for any $\alpha \leq \alpha^*$. When target value is μ_h^k one can analogously prove (45). The proof is completed. $\square$

Finally, summarizing the results we have:

Theorem 6.5 *Let $\tau \geq 10$ and $(\Delta x, \Delta y, \Delta s)$ be the solution of system (12) and μ is as it is defined in Algorithm SR-IIPM. Then the step size $\alpha^* = \dfrac{3}{10} \dfrac{1}{C_3 n^2(1+4\tau)}$ is strictly feasible. Moreover, we have*

$$\Phi(x(\alpha^*), s(\alpha^*), \mu_g(\alpha^*)) \leq \Phi(x, s, \mu_t), \tag{47}$$

$$\Phi(x(\alpha^*), s(\alpha^*), \mu_t^k) \leq \Phi(x, s, \mu_t^k) - \frac{\alpha^*}{2}\Phi(x, s, \mu_t^k), \tag{48}$$

$$\mu_g^{k+1} \geq (1-\alpha^*)\mu_g^k. \tag{49}$$

To obtain an upper bound for the total number of iterations of Algorithm SR-IIPM we need to estimate the value of the step size α^* or, the change of the parameter μ_t before and after one iteration. The following technical lemma will be used in our estimation about μ_t. The lemma is a direct consequence of Lemma IV.36 in [16].

Lemma 6.6 *Let $v_+ = \frac{v}{\sqrt{1-\theta}}$ for some $\theta \in (0,1)$. Then we have*

$$\Psi(v_+) \leq \frac{1}{1-\theta}\Psi(v) + \frac{\theta\sqrt{2n\Psi(v)}}{1-\theta} + \frac{n\theta^2}{1-\theta}.$$

By applying Lemma 6.6 to Theorem 6.5, we can prove the following theorem.

Theorem 6.7 *Let $\tau \geq 10$ and $(\Delta x, \Delta y, \Delta s)$ be the solution of system (12), where μ is as it is defined in Algorithm SR-IIPM, and α^* is the default step size defined in Theorem 6.5. Let $\theta \leq \frac{\alpha^*}{4}$, Then*

$$\Phi(x(\alpha^*), s(\alpha^*), \mu_t(1-\theta)) \leq \Phi(x, s, \mu_t^k).$$

Proof: Form Lemma 6.6 we can see that to prove the theorem, it suffices to choose θ so that it satisfies the following inequality

$$\Phi(x(\alpha^*), s(\alpha^*), \mu_t^k) + \theta\sqrt{2n\Phi(x(\alpha^*), s(\alpha^*), \mu_t^k)} + n\theta^2 \leq (1-\theta)\Phi(x, s, \mu_t^k).$$

Using Theorem 6.5, we can conclude that the above inequality will be satisfied if

$$\theta\sqrt{2n\Phi(x, s, \mu_t^k)} + \theta\Phi(x, s, \mu_t^k) + n\theta^2 \leq \frac{\alpha^*}{2}\Phi(x, s, \mu_t^k). \tag{50}$$

Recall the fact that $\Phi(x, s, \mu_t^k) = \frac{(\tau-1)n}{2}$, thus we can rewrite (50) as

$$\theta n\sqrt{\tau-1} + \frac{(\tau-1)n}{2} + n\theta^2 \leq \frac{\alpha^*}{4}(\tau-1)n = \frac{n(\tau-1)\alpha^*}{4}.$$

Note that for $\theta < 1$ and $\tau \geq 10$, we have

$$\theta\sqrt{\tau-1} + \frac{(\tau-1)\theta}{2} + \theta^2 \leq (\tau-1)\theta.$$

This relation implies that if we choose $\theta = \frac{\alpha^*}{4}$, then we have

$$\Phi(x(\alpha^*), s(\alpha^*), \mu_t^k(1-\theta)) \le \Phi(x, s, \mu_t^k).$$

The proof is completed. □

Now we can proceed to discuss the complexity of Algorithm SR-IIPM. By the choice of μ_t^k we know that the proximity function $\Phi(x, s, \mu_t)$ keeps invariable for all the iterates. Let us denote by μ_t^+ the target parameter value after one step. Then we have

$$\Phi(x(0), s(0), \mu_t^k) = \Phi(x(\alpha^*), s(\alpha^*), \mu_t^+).$$

On the other hand, we know that

$$\Phi(x(\alpha^*), s(a^*), \mu_g(\alpha^*)) \le \Phi(x(a^*), s(\alpha^*), \mu_t^+),$$
$$\Phi(x(\alpha^*), s(\alpha^*), \mu_t^k) \le \Phi(x(\alpha^*), s(\alpha^*), \mu_t^k) \le \Phi(x(\alpha^*), s(\alpha^*), \mu_t^+).$$

Because $\mu_t^+ \le \mu_g(\alpha^*)$, from this two inequalities we get $\mu_t^k \ge \mu_t^+$. Therefore, by using Theorem 6.7 we can claim that

$$\mu_t^+ \le \left(1 - \frac{\alpha^*}{4}\right) \mu_t^k. \tag{51}$$

Now we are ready to prove the complexity of Algorithm SR-IIPM.

Theorem 6.8 *Let* $\tau \ge 10$ *and* $t_0 = \max\left(1, \frac{\|(r_b^0, r_c^0)\|}{\mu_0}\right)$. *Then Algorithm SR-IIPM will terminate after at most*

$$\left\lceil \frac{40}{3}(1+4\tau)C_3 n^2 \log \frac{n(\tau+1)t_0}{\epsilon} \right\rceil$$

iterations with a solution satisfying $x^T s \le \epsilon$ *and* $\|(r_b, r_c)\| \le \epsilon$.

Proof: In light of relation (51) we know that after at most

$$\left\lceil \frac{4}{\alpha^*} \log \frac{n(\tau+1)t_0}{\epsilon} \right\rceil$$

iterations we have $\mu_t^k \le \frac{\epsilon}{n(\tau+1)t_0}$ that implies $x^T s \le \frac{\epsilon}{t_0}$. Using Theorems 6.2 and 6.5 we have $\mu_g^{k+1} \ge (1-\alpha)\mu_g^k$, therefore we may write

$$\frac{\left\|(r_b^{k+1}, r_c^{k+1})\right\|}{\mu_g^{k+1}} \le \frac{(1-\alpha^*)\left\|(r_b^k, r_c^k)\right\|}{\mu_g^{k+1}} \le \frac{\left\|(r_b^k, r_c^k)\right\|}{\mu_g^k} \le \beta \frac{\|(r_b^0, r_c^0)\|}{\mu_0}.$$

Thus when the specified number of iterations is reached, we have $\left\|(r_b^{k+1}, r_c^{k+1})\right\| \le \beta \mu_g^{k+1} t_0 \le \epsilon$. □

7 Implementation and Numerical Results

We implemented the algorithms in C by using the WSMP [2] sparse matrix package and some elements of OSL [11]. First we call some OSL subroutines for data input and preprocessing, and then we transfer the data of LO problems to our solver; then we solve the problem with our SR-Proximity based IIPM solver and the results are feed back to OSL for postprocessing that gives the final solution. WSMP is utilized to solve the normal equation system at each iteration and some ESSL subroutines are used for matrix and vector operations. The system environment is an IBM RS/6000 44P Model 270 workstation with operating system AIX 4.3. We tested all the benchmark problems in NETLIB. The results are highly encouraging. The average iteration number is less than that of OSL and LIPSOL [30] while the solutions have a little higher precision.

As an illustration, Table 1 compares the iteration numbers and precision with OSL and LIPSOL for the problems of the Kinnington set from the NETLIB library. In the table, the column "Digits" shows how many digits of the objective value are the same as the standard reference solutions in NETLIB.

For a total of 95 problems of the standard set in NETLIB, our algorithm gets an average of 9.93 digits precision by an average of 21.33 iterations, while in average OSL gives 9.01 precise digits by 21.68 iterations and LIPSOL gives 9.53 precise digits by 21.34 iterations. Our results for 34 problems are better than LIPSOL according to less iterations (with the same precision) or higher precision (with the same number of iterations) or both. For 19 problems our algorithm is worse than LIPSOL, while for the other 42 problems, the results are equally good.

8 Conclusions

In this paper, a self-regular proximity based infeasible IPM is presented and polynomial complexity of the algorithm is established. The number of iterations is bounded by $\mathcal{O}\left(n^2 \log \frac{n}{\varepsilon}\right)$. Limited computational results are reported as well. Numerical experiences demonstrate the potential of the algorithm to solve practical problems efficiently.

Acknowledgement: The first author would like to thank the Iranian Ministry of Science, Research and Technology for supporting his research. This research also supported by an NSERC Discovery grant and by MITACS and IBM.

References

[1] E.D. Anderson, C. Roos, T. Terlaky, T. Trafalis and J.P. Warners. The use of low-rank updates in interior-point methods. Technical Report, Delft University of Technology, The Netherlands, 1996.

[2] A. Gupta. *WSMP: Watson Sparse Matrix Package* (Part I: direct solution of symmetric sparce systems). Technical Report RC 21866(98462), IBM T.J. Watson Research Center, Yorktown Heights, NY, 2000. http://www.cs.umn.edu/~agupta/wsmp.html.

[3] N. Karmarkar. New polynomial-time algorithm for linear programming. *Combinatorica*, **4**, 373–395, 1984.

[4] M. Kojima. Basic lemmas in polynomial-time infeasible-interior-point methods for linear programs. *Annals of Operations Research*, **62**, 1–28, 1996.

[5] M. Kojima, N. Megiddo, and S. Mizuno. A primal-dual infeasible-interior-point algorithm for linear programming. *Mathematical Programming*, **61**, 263–280, 1993.

[6] I.J. Lustig. Interior point methods: computational state of the art. *Linear Algebra and Its Applications*, **152**, 191–222, 1993.

Table 1: Comparison of Iteration No and Precision

SR-IIPM			OSL		LIPSOL	
Problem	Iter	Digits	Iter	Digits	Iter	Digits
cre-a	29	12	35	10	30	11
cre-b	42	12	48	10	42	11
cre-c	28	12	34	10	30	11
cre-d	41	11	51	9	38	11
ken-07	14	11	16	11	16	11
ken-11	18	10	21	11	22	11
ken-13	23	11	25	11	27	11
ken-18	32	11	31	11	?	?
osa-07	29	12	24	10	27	11
osa-14	41	11	25	11	37	11
osa-30	31	12	36	11	36	10
osa-60	34	10	32	10	?	?
pds-02	21	12	22	11	29	11
pds-06	33	11	34	11	43	11
pds-10	45	12	46	11	53	11
pds-20	49	12	58	11	69	11

[7] I.J. Lustig. Feasible issues in a primal-dual interior-point methods for linear programming. *Mathematical Programming*, **49**, 145–162, 1990/91.

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

[9] S. Mizuno. Polynomiality of infeasible-interior-point algorithms for linear programming. *Mathematical Programming*, **67**, 109–119, 1994.

[10] R.D.C. Monteiro and I. Adler. Interior path following primal-dual algorithms. Part I: linear programming. *Mathematical Programming*, **44**, 27–41, 1989.

[11] OSL: The Optimization Solution Library.
Home page: `http://www-3.ibm.com/software/data/bi/index.html`

[12] J. Peng *New Design and Analysis of Interior-Point Methods.* Ph. D. Thesis, Faculty of Information Technology and Systems, Delft University of Technology, The Netherlands, 2001.

[13] J. Peng and T. Terlaky. A dynamic large-update primal-dual interior-point methods for linear optimizaion. To appear in *Optimization Methods and Software.*

[14] J. Peng, C. Roos, and T. Terlaky. Self-Regular proximities and new search directions for linear and semidefinite optimization. *Mathematical Programming*, **93**, 129–171, 2002.

[15] F.A. Potra. A quadratically convergence predictor-corrector method for solving linear programs from infeasible starting points. *Mathematical Programming*, **67**, 383–406, 1994.

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

[17] D. Shanno and E. Simantiraki. An infeasible interior-point method for linear complementarity problems. *SIAM Journal on Optimization*, **7**, no. 3, 620–640, 1997.

[18] J. Stoer, M. Wechs, and S. Mizuno. High order infeasible-interior-point methods for solving sufficient linear complementarity problems. *In system modeling and optimization* (Detroit, MI, Chapman Hall/CRC Res. Notes Math., 396, 245–252, 1997.

[19] K. Tanabe. Centered Newton method for linear programming: Interior and 'exterior' point method (in Janpanese). In: *New Methods for Linear Programming.* K. Tone, (Ed.) **3**, 98–100, 1990.

[20] T. Terlaky (Ed.). *Interior Point Methods of Mathematical Programming.* Kluwer Academic Publishers, 1996.

[21] M.J. Todd, and Y. Ye. Approximate Farkas lemmas and stopping rules for iterative infeasible-point algorithms for linear programming. *Mathematical Programming*, **81**, 1–21, 1998.

[22] P. Tseng. Analysis of an infeasible interior path -following method for complementarity problems. *SIAM Journal on Optimization*, **7**, no. 2, 386–402, 1997.

[23] R.H. Tütüncü. An infeasible-interior-point potential-reduction algorithm for linear programming. *Mathematical Programming*, **86**, no. 2, 313–334, 1999.

[24] S.J. Wright. *Primal-Dual Interior-Point Methods.* SIAM, Philadelphia, 1996.

[25] S.J. Wright. An infeasible-interior-point algorithm for linear complementarity problems. *Mathematical Programming,* **67**, 29–52, 1994.

[26] S.J. Wright and Y. Zhang. A superquadratic infeasible-interior-point method for linear complementarity problems. *Mathematical Programming,* **73**, no. 3, 269-289, 1996.

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

[28] Y. Ye, M.J. Todd, and S. Mizuno. An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Mathematics of Operations Research*, **19**, 53–67, 1994.

[29] Y. Zhang. On the convergence of a class of infeasible-interior-point methods for the horizantal linear complementarity problem. *SIAM Journal on Optimization*, **4**, 208–227, 1994.

[30] Y. Zhang. Solving large-scale linear programs by interior-point methods under the MATLAB environment. *Optimization Methods and software*, **10**, 1–31, 1999.

[31] Y. Zhang User's Guide to LIPSOL Linear-programming Interior Point solvers V0.4. *Optimization Methods and Software*, **11/12**, 385–396, 1999.

9 Appendix

In this section we present the detailed proofs of several technical results.

Proof of Theorem 6.1.

Before we start the proof, we quote a technical result from [12, 14].

Lemma 9.1 *Suppose that* $\gamma \in [0,1]$. *Then*

$$\begin{aligned} (1+t)^\gamma &\le 1+\gamma t, \qquad \forall\ t \in [-1,\infty) \\ (1+t)^\gamma &\le 1-\gamma t, \qquad \forall\ t \in [0,1]. \end{aligned} \tag{52}$$

Now we can outline the key steps in the proof. The reader is referred to [12] or [14] for the detailed proof. In order to estimate the decrease of the proximity function after a step, we first give some bounds for the smallest component $v_{\min}$ of v and the maximal feasible step size $\alpha_{\max}$. Recall that by the definition of σ we have

$$\sigma = \left\| v - v^{-3} \right\| \ge \left| v_{\min} - v_{\min}^{-3} \right|.$$

Then we have

$$v_{\min} \ge (1+\sigma)^{\frac{-1}{3}} \ge \frac{3\sigma}{1+3\sigma}\sigma^{-\frac{1}{3}} \ge \frac{9}{10}\sigma^{-\frac{1}{3}}, \tag{53}$$

where the second inequality follows from Lemma 9.1, and the last inequality is given by the fact that when $\tau \ge 10$ is used in the algorithm, then the relation

$$\sigma^2 \ge \left\| v - v^{-1} \right\|^2 \ge (\tau - 1)n \ge 9 \tag{54}$$

holds. Let $\bar{\alpha} = v_{\min}\sigma_1^{-1}$. It follows immediately that

$$\alpha_{\max} \ge v_{\min}\sigma_1^{-1} \ge \frac{9}{10}\sigma^{-\frac{1}{3}}\sigma_1^{-1}. \tag{55}$$

This proves the first statement of the theorem. Now let us define

$$f(\alpha) = \Psi(v_+(\alpha)) - \Psi(v).$$

Using Condition SR.2. of Definition 2.1., we have

$$\begin{aligned} f(\alpha) &\leq \frac{1}{2}\Psi(v+\alpha d_x) + \frac{1}{2}\Psi(v+\alpha d_s) - \Psi(v) := f_1(\alpha) \\ &= -\Psi(v) + \frac{1}{2}\sum_{i\in I}(\psi(v_i+\alpha[d_x]_i) + \psi(v_i+\alpha[d_s]_i)). \end{aligned} \tag{56}$$

Obviously, both functions $f(\alpha)$ and $f_1(\alpha)$ are twice continuously differentiable with respect to α if $\alpha \leq \alpha_{\max}$. Analogously to Lemma 3.3.2 and Lemma 3.3.3 of [12, 14] we have

$$f_1^{''}(\alpha) \leq \frac{\sigma_1^2}{2}(1 + 3(v_{\min} - \alpha\sigma_1)^{-4}),$$

and

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

Then we may write

$$f_1(\alpha) \leq -\frac{\sigma^2}{2} + \frac{\sigma_1^2}{2}\int_0^{\alpha}\int_0^{\xi}[1+3(v_{\min} - \eta\sigma_1)^{-4}]d\eta d\xi := f_2(\alpha).$$

The function $f_2(\alpha)$ is convex and twice continuously differentiable in the interval $[0, \bar{\alpha})$. Let us denote by α^* the global minimizer of $f_2(\alpha)$ in the interval $[0, \bar{\alpha})$, that is

$$\alpha^* = \arg\min\{f_2(\alpha) : \alpha \in [0, \bar{\alpha}]\}.$$

Analogously to the proof of Lemma 3.3.3 of [12] we can see that α^* is the solution of the equation,

$$-\sigma^2 + \sigma_1^2\alpha + \sigma_1((v_{\min} - \alpha\sigma_1)^{-3} - v_{\min}^{-3}) = 0. \tag{57}$$

Let us define

$$\begin{aligned} \omega_1(\alpha) &= -\frac{\sigma^2}{2} + \sigma_1^2\alpha, \\ \omega_2(\alpha) &= -\frac{\sigma^2}{2} + \sigma_1((v_{\min} - \alpha\sigma_1)^{-3} - v_{\min}^{-3}). \end{aligned}$$

It is easy to verify that both functions $\omega_1(\alpha)$ and $\omega_2(\alpha)$ are increasing for $\alpha \in [0, \bar{\alpha})$. Using this two functions we can write equation (57) as

$$\omega_1(\alpha^*) + \omega_2(\alpha^*) = 0.$$

The root α_1^* of $\omega_1(\alpha) = 0$ is $\alpha_1^* = \frac{\sigma^2}{2\sigma_1}$. Through simple calculus, we find that the root α_2^* of $\omega_2(\alpha) = 0$ is

$$\alpha_2^* = \frac{v_{\min}}{\sigma_1}\left(1 - \left(1 + \frac{v_{\min}^3\sigma^2}{2\sigma_1}\right)^{-\frac{1}{3}}\right). \tag{58}$$

Now by using (52) we can write

$$\left(1 + \frac{v_{\min}^3\sigma^2}{2\sigma_1}\right)^{-\frac{1}{3}} = \left(1 - \frac{v_{\min}^3\sigma^2}{2\sigma_1 + v_{\min}^3\sigma^2}\right)^{\frac{1}{3}} \leq 1 - \frac{v_{\min}^3\sigma^2}{3(2\sigma_1 + v_{\min}^3\sigma^2)}.$$

By using a similar proof as given for Lemma 3.3.3 in [12] we can conclude that

$$\alpha_2^* \geq \frac{3}{10} \frac{\sigma^{\frac{2}{3}} \sigma_1^{-1}}{\sigma + 4\sigma_1}.$$

Now we have $\alpha^* \geq \min(\alpha_1^*, \alpha_2^*) = \alpha_2^*$. □

Proof of Theorem 6.2 It suffices to estimate the interval in which the proximity function satisfies

$$\Phi(x(\alpha), s(\alpha), \mu_g(\alpha)) \leq \frac{(\tau - 1)n}{2}.$$

We start by considering the function

$$g_1(\alpha) = \Phi(x(\alpha), s(\alpha), \mu^*) = \frac{1}{2}\left(\frac{\mu_t^k}{\mu^*}\|v_+(\alpha)\|^2 - 2n + \frac{\mu^*}{\mu_t^k}\left\|v_+(\alpha)^{-1}\right\|^2\right).$$

Note that, from the definition of μ^* and μ_t, we know that

$$\frac{\mu_t^k}{\mu^*}\|v_+(0)\|^2 = \frac{\mu^*}{\mu_t^k}\left\|v_+(0)^{-1}\right\|^2 = n\sqrt{\tau}.$$

On the other hand, we have

$$\|v_+(\alpha)\|^2 = \|v\|^2 + \alpha v^T(d_x + d_s) + \alpha^2 d_x^T d_s \leq \|v\|^2 + \alpha^2 d_x^T d_s.$$

Because $\alpha v^T(d_x + d_s) = \alpha(\left\|v^{-1}\right\|^2 - \|v\|^2) \leq 0$ for any strictly positive step size α. It is easy to verify that

$$\left\|v_+(\alpha)^{-1}\right\|^2 \leq (1 - \alpha v_{\min}^{-1}\sigma_1)^{-2}\left\|v^{-1}\right\|^2.$$

If we choose α such that $(1 - \alpha v_{\min}^{-1}\sigma_1)^{-2} \leq \sqrt{2} \leq \sqrt{\frac{\tau}{\tau_0}}$. Then because $1 - 2^{-\frac{1}{4}} \geq \frac{1}{8}$ we have $v_{\min} \geq \frac{9}{10}\sigma^{-\frac{1}{3}}$ and $\alpha \leq \frac{1}{9}\sigma^{-\frac{1}{3}}\sigma_1^{-1}$. Then we have

$$g_1(\alpha) \leq \frac{1}{2}\left(n\sqrt{\tau}_0 - 2n + \alpha^2 d_x^T d_s + n\sqrt{\tau}_0\sqrt{\frac{\tau}{\tau_0}}\right) \leq \frac{1}{2}\left(2n\sqrt{\tau} - 2n + \alpha^2(\tau+1)C_2^2 n^2 \sigma^{\frac{2}{3}}\right),$$

so if we take $\hat{\alpha}_t^* = \min\left(\frac{1}{C_2\sqrt{\tau+1}n\sigma^{\frac{1}{3}}}, \frac{1}{9}\sigma^{-\frac{1}{3}}\sigma_1^{-1}\right)$, we have $g_1(\alpha) \leq \frac{1}{2}(2n\sqrt{\tau} - 2n + n) \leq \frac{(\tau-1)n}{2}$. The proof is completed. □

Proof of Theorem 6.3.
First we observe from the assumption that $\mu_g^k = \tau_1\mu_h^k$ where $\frac{\tau}{2} < \tau_1 \leq \tau$, we obtain $\mu^* = \sqrt{\tau}_1\mu_h^k$. Let us define

$$g(\alpha) = \Phi(x(\alpha), s(\alpha), \mu^*) - \Phi(x, s, \mu^*).$$

Using condition SR.2 we have

$$g(\alpha) \leq \frac{1}{2}\Phi\left(\frac{v + \alpha d_x}{\tau_1^{\frac{1}{4}}}\right) + \frac{1}{2}\Phi\left(\frac{v + \alpha d_s}{\tau_1^{\frac{1}{4}}}\right) - \Phi(v) := g_1(\alpha).$$

Moreover, from the definition of v we have $\left\|v^{-1}\right\|^2 = n$, which further implies

$$\left\|v^{-3}\right\|^2 - \left\|v^{-1}\right\|^2 = \left\|v^{-3} - v^{-1}\right\|^2 + 2\left\|v^{-2} - e\right\|^2 \geq 0.$$

This inequality, together with the fact that $\|v\|^2 = \tau_1 \|v^{-1}\|^2$, gives

$$g_1^{'}(0) \leq -\frac{\sigma^2}{2\sqrt{\tau_1}}.$$

The same way as Lemma 3.3.3 in [12] is proved we have:

$$g_1(\alpha) \leq -\frac{\sigma^2}{2\sqrt{\tau_1}} + \frac{\sigma_1^2}{2\sqrt{\tau}_1}\int_0^\alpha \int_0^\xi (1 + 3\tau_1(v_{min} - \eta\sigma_1)^{-4})d\eta d\xi := g_2(\alpha).$$

It is easy to see, via making use of simple calculus, that $g_2(\alpha)$ is convex and twice differentiable for all α. Let $\hat{\alpha}_2^*$ denote the global minimizer of $g_2(\alpha)$ in the interval. Now like in Lemma 3.3.3 [12] we have $\hat{\alpha}_2^*$ to be the solution of the following equation

$$-\frac{\sigma^2}{\sqrt{\tau_1}} + \frac{\sigma_1^2\alpha}{\sqrt{\tau}_1} + \sigma_1\sqrt{\tau}_1((v_{\min} - \alpha\sigma_1)^{-3} - v_{\min}^{-3}) = 0.$$

Now, Theorem 6.1 gives $\alpha_2^* = \frac{\sigma^2}{2\sigma_1^2}$ and $\alpha_2^* \geq \frac{3}{10}\frac{\sigma^{\frac{2}{3}}\sigma_1^{-1}}{\sigma+4\tau_1\sigma_1}$. It is obvious that $\hat{\alpha}_2^* \geq \min(\alpha_1^*, a_2^*) = \alpha_2^*$. Hence, whenever $\alpha \leq \hat{\alpha}_2^*$, the relation

$$\Phi(x(\alpha), s(\alpha), \mu^*(\alpha)) \leq \Phi(x(\alpha), s(\alpha), \mu^*(0)) \leq \Phi(x, s, \mu^*)$$

holds. Now, by using equation (10) we can claim

$$\Phi(x(\alpha), s(\alpha), \mu_g(\alpha)) \leq \Phi(x, s, \mu_g) \leq \Phi(x, s, \mu_t^k),$$

where the last inequality follows from and the fact that $\mu_g^k \geq \mu_h^k \geq \mu_t^k$. Now we focus on inequality (43). In order to investigate the behavior of the proximity function $\Phi(x(\alpha), s(\alpha), \mu_t)$ for a feasible step size α, we define

$$\begin{aligned} h(\alpha) &:= \Phi(x(\alpha), s(\alpha), \mu_t^k) - \Phi(x(0), s(0), \mu_t^k) \\ &= \frac{1}{2}\left(\frac{\mu_h^k}{\mu_t^k} - \frac{\mu_t^k}{\mu_h^k}\right)\left[\alpha v^T(d_x + d_s) + \alpha^2 d_x^T d_s\right] + \frac{\mu_t}{2\mu_h}[\alpha v^T(d_x + d_s) - \left\|v^{-1}\right\|^2 \\ &\quad + \sum_{i\in\mathcal{I}} \frac{1}{[v + \alpha d_x]_i[v + \alpha d_s]_i} + \alpha^2 d_x^T d_s]. \end{aligned}$$

Now, by applying a procedure similar to the proof of Theorem 6.1 to the second term in the above formula, we can prove that for $\alpha \leq \frac{3}{10}\frac{\sigma^{\frac{2}{3}}\sigma_1^{-1}}{\sigma+4\sigma_1}$ the following relation holds

$$h(\alpha) \leq \frac{1}{2}\left(\frac{\mu_h^k}{\mu_t^k} - \frac{\mu_t^k}{\mu_h^k}\right)\left[\alpha v^T(d_x + d_s) + \alpha^2 d_x^T d_s\right] - \frac{\alpha\mu_t^k}{4\mu_h^k}\sigma^2.$$

Moreover, for any $\alpha \leq \frac{\sigma^{-\frac{2}{3}}}{2(\tau+1)C_2^2}$ we have

$$v^T(d_x + d_s) + \alpha d_x^T d_s \leq -\left(\frac{\mu_g^k}{\mu_h^k} - 1\right)\left\|v^{-1}\right\|^2 + \alpha(\tau+1)C_2^2 n^2\sigma^{\frac{2}{3}} \leq -\left\|v^{-1}\right\|^2.$$

Finally, if we choose $\alpha \leq \min\left(\frac{3}{10}\frac{\sigma^{\frac{2}{3}}\sigma_1^{-1}}{\sigma+4\sigma_1}, \frac{\sigma^{-\frac{2}{3}}}{4C_2^2 n}\right)$, then

$$h(\alpha) \leq -\frac{\alpha}{2}\Phi(x, s, \mu_t^k).$$

The proof is completed. □