New Complexity Analysis of IIPMs for Linear Optimization Based on a Specific Self-Regular Function

M. Salahi*, M.R. Peyghami†, T. Terlaky‡

June 20, 2005

Abstract

Primal-dual Interior-Point Methods (IPMs) have shown their ability in solving large classes of optimization problems efficiently. Feasible IPMs require a strictly feasible starting point to generate the iterates that converge to an optimal solution. The self-dual embedding model provides an elegant solution to this problem with the cost of slightly increasing the size of the problem. On the other hand, Infeasible Interior Point Methods (IIPMs) can be initiated by any positive vector, and thus are popular in IPM softwares. In this paper we propose an adaptive large-update IIPM based on a specific self-regular proximity function, with barrier degree $1 + \log n$, that operates in the infinity neighborhood of the central path. An $O\left(n^{\frac{3}{2}} \log n \log \frac{n}{\epsilon}\right)$ worst-case iteration bound of our new algorithm is established. This iteration bound improves the so far best $O\left(n^2 \log \frac{n}{\epsilon}\right)$ iterations bound of IIPMs in a large neighborhood of the central path.

Keywords: Linear Optimization, Infeasible Interior Point Methods, Self-Regular Functions, Adaptive Large Update, Polynomial Complexity.

AMS Subject Classification: $90C51, 90C20$

1 Introduction

The landmark paper of Karmarkar [3] revitalized Linear Optimization (LO) as an active area of research. Since then many researchers have suggested and analyzed various Interior

*Corresponding author, Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada, salahim@mcmaster.ca

†Department of Mathematical Sciences, Sharif University of Technology, Tehran, Iran, peyghami@mehr.sharif.edu

‡Department of Computing and Software, McMaster University, Hamilton, Ontario, Canada, terlaky@mcmaster.ca

Point Methods (IPMs). Today, IPMs are not only the most effective methods in practice but also have polynomial time complexity. For surveys of IPMs we refer to [1, 16, 21, 22]. Recently, a new variant of feasible IPMs based on Self-Regular (SR) proximity functions was presented by Peng, Roos and Terlaky [11]. Based on SR-proximities, they provided so far the best worst case theoretical complexity, namely $O\left(\sqrt{n}\log n \log \frac{n}{\epsilon}\right)$, for large neighborhood feasible IPMs, for the case when the barrier degree of the corresponding SR function is $1+\log n$.

We mention that feasible IPMs start with a strictly feasible interior point and keep feasibility during the solution process. One way to find an initial feasible interior point is to use the homogeneous self-dual embedding model by introducing artificial variables as it was first presented by Ye et al. [23]. The embedding approach allows to apply feasible IPMs and automatically preserve the complexity of feasible IPMs [16, 22]. Another answer to the initialization problem is the introduction of Infeasible Interior Point Methods (IIPMs). IIPMs were introduced first by Lustig [5] and later Kojima et al. [4] proved the global convergence of an IIPM. Subsequently, polynomial iteration complexity for variants of this algorithm was established by Zhang [24], Mizuno [8] and Potra [14, 15]. Under certain conditions, an $O(n^2 \log \frac{n}{\epsilon})$-iteration bound for IIPMs was proved by Zhang [24].

In spite of weaker theoretical worst case complexity results, IIPMs are appealing in practice because the computational cost per iteration is higher (one extra backsolve) [16, 23, 25] when the self-dual embedding model is employed. IIPMs start with an arbitrary positive point and feasibility is reached as optimality is approached. Recently, Salahi, Terlaky and Zhang [19] presented an adaptive IIPM based on a specific SR-proximity function with barrier degree 3 that has an $O(n^2 \log \frac{n}{\epsilon})$ iteration complexity. Numerical experiences demonstrate the potential of their algorithm to solve practical problems efficiently.

In this paper we propose an adaptive large-update IIPM based on a specific SR proximity function, with barrier degree $1+\log n$, that uses a different approach to define the search directions compared to simple extension of feasible IPMs based on SR-proximity functions. To this end, we separated the neighborhood of the central path to sub neighborhoods that enables us to have better control on the iterates. An $O(n^{\frac{3}{2}} \log n \log \frac{n}{\epsilon})$ worst case iterations bound of our new algorithm is established. This iteration complexity improves the so far best $O(n^2 \log \frac{n}{\epsilon})$ complexity for large-update IIPMs.

The paper is organized as follows. In Section 1.1, we give a survey of the main ideas behind IPMs. In Section 1.2, by introducing SR-Proximity functions, we present our kernel function that we use in this paper. We explore the role of the target parameter μ w.r.t. the proximity function in Section 2, and then we discuss the change of the complementary gap along the search directions. In Section 3, by introducing an infeasible central path neighborhood, we present an adaptive update algorithm that leads us to our large update IIPM. Some technical results are given, and we investigate a default value for the step size α. The growth behavior of the proximity function and the worst-case complexity estimate of the algorithm are also reported in Section 3. Concluding remarks are given in Section 4. To make the paper easily readable, most detailed technical proofs are moved to the Appendix.

1.1 Review of IPMs

In this paper we deal with the standard LO problem

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

where $A \in \mathbb{R}^{m\times n}$, $c \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$ are given, $\text{rank}(A) = m$ and $x \in \mathbb{R}^n$ is the vector of variables. The dual problem of (P) is given by

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

where $s \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are the vectors of variables. To find an optimal solution of (P) and (D), it is sufficient [16] to solve the system

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

where xs denotes the componentwise (Hadamard) product of the vectors x and s, i.e., $(xs)_i = x_i s_i$ for $i = 1, \ldots, n$. The first and second equations represent primal and dual feasibility and the third equation is called the *complementary condition* for problems (P) and (D).

We use the following notational conventions. Throughout the paper, $\|.\|$ denotes the 2-norm of a vector. For any $x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$, $x_{\min} = \min(x_1, x_2, \ldots, x_n)$ is the smallest component of x. The nonnegative and the positive orthants are denoted by $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$, respectively. We write $f(t) = O(g(t))$ if there exists a positive constant ω such that $f(t) \leq \omega g(t)$, for all $t > 0$.

The fundamental idea of primal-dual IPMs is to replace the third equation in (1), by the parameterized equation[1] $xs = \mu\mathbf{1}$, with $\mu > 0$. Parameter μ is referred to as the *central path parameter*. This replacement leads us to have the following system:

$$\begin{aligned} Ax &= b, \quad x > 0, \\ A^T y + s &= c, \quad s > 0, \\ xs &= \mu\mathbf{1}. \end{aligned} \tag{2}$$

Assuming that both (P) and (D) have an interior feasible solution, for all $\mu > 0$ system (2) has a unique solution [16] and the set of unique solutions $\{(x(\mu), y(\mu), s(\mu)) \mid \mu > 0\}$ of system (2) forms the so-called *central path* of problems (P) and (D) [16, 22]. The central path for LO was recognized first by Sonnevend [20] and Megiddo [6].

We apply Newton's method for getting approximate solutions for system (2). As $\mu \to 0$, the limit of the central path exists and converges to an optimal solution of (P) and (D). Given any $x > 0$, y and $s > 0$, in IIPMs the Newton direction $(\Delta x, \Delta y, \Delta s)$ for system (2) is determined by the following system of linear equations:

$$\begin{aligned} A\Delta x &= -r_b, \\ A^T \Delta y + \Delta s &= -r_c, \\ x\Delta s + s\Delta x &= \mu\mathbf{1} - xs, \end{aligned} \tag{3}$$

[1] $\mathbf{1}$ denotes the all one vector.

where

$$r_b = Ax - b,$$
$$r_c = A^T y + s - c.$$

Since A has full row rank, for any $x > 0$ and $s > 0$, system (3) uniquely defines a Newton search direction $(\Delta x, \Delta y, \Delta s)$. The third equation in (3) is called the *centering equation*. By taking a step along the Newton direction, one constructs a new triple $(x(\alpha), y(\alpha), s(\alpha))$ according to

$$x(\alpha) = x + \alpha\Delta x, \quad y(\alpha) = y + \alpha\Delta y, \quad s(\alpha) = s + \alpha\Delta s, \tag{4}$$

for some $0 < \alpha \leq 1$, satisfying $x(\alpha) > 0$ and $s(\alpha) > 0$. We repeat the aforementioned procedure by decreasing μ to $\mu_+ = (1-\theta)\mu$, for some $\theta \in (0, 1)$. For some sufficiently small $\epsilon > 0$ we find an ϵ-optimal solution of the problems (P) and (D) if $x^T s \leq \epsilon$ and primal/dual infeasibility is small enough.

1.2 SR-IPMs

The choice of θ, the so-called update parameter, plays an important role both in the theory and practice of IPMs. Usually, if θ is a constant independent of the dimension of the problem, e.g., $\theta = \frac{1}{2}$, then we call the algorithm a *large-update* (or *long-step*) method. If θ depends on the dimension of the problem, such as $\theta = \frac{1}{\sqrt{n}}$, then the algorithm is called a *small-update* (or *short-step*) method. Recall that small-update methods have the best iteration bound. They require $O\left(\sqrt{n}\log\frac{n}{\epsilon}\right)$ iterations to produce an ϵ-optimal solution [16]. On the other hand, large-update methods based on the Newton direction are much more efficient in practice than small-update methods, while they have a weaker worst-case iteration bound, namely $O\left(n\log\frac{n}{\epsilon}\right)$ [1, 2, 7, 16, 21, 22]. This phenomenon is called "The gap between theory and practice". To resolve this discrepancy, Peng et al. [10] introduced the class of SR-proximity functions for IPMs and for a special member of this class established an $O\left(\sqrt{n}\log n\log\frac{n}{\epsilon}\right)$ iteration bound for large-update IPMs.

Let us start with the definition of SR kernel functions that induce SR-proximity functions [11].

Definition 1.1 *A two times continuously differentiable function $\psi(t) : \mathbb{R}_{++} \to \mathbb{R}_+$ is self-regular (SR) 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*

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

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

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

We refer to parameter q as the *barrier degree* and p as the *growth degree* of the function $\psi(t)$. Since $\psi(1) = \psi'(1) = 0$, we can determine $\psi(t)$ by its second derivative as follows:

$$\psi(t) = \int_1^t \int_1^\xi \psi''(\zeta) d\zeta d\xi.$$

A popular family of SR functions that is given by

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

where $\nu_1 = \min(p, q)$ and $\nu_2 = \max(p, q)$, with $p \geq 1$ and $q > 1$.

In this paper, we modify the feasible IPMs approach suggested in [12, 17] to IIPMs. We relate to any triple (x, y, s) the vectors

$$v := \sqrt{\frac{xs}{\mu}}, \quad v^{-1} := \sqrt{\frac{\mu \mathbf{1}}{xs}}, \tag{5}$$

where the i^{th} components of v and v^{-1} are given by $\sqrt{\frac{x_i s_i}{\mu}}$ and $\sqrt{\frac{\mu}{x_i s_i}}$, respectively. In infeasible SR-proximity based IPMs, system (3) is modified as

$$\begin{aligned} A\Delta x &= -r_b, \\ A^T \Delta y + \Delta s &= -r_c, \\ x\Delta s + s\Delta x &= -\mu v \nabla \Psi(v), \end{aligned} \tag{6}$$

where the SR-proximity $\Psi(v)$ is defined by $\Psi(v) = \sum_{i=1}^n \psi(v_i)$, where $\psi(t)$ is a univariate SR kernel function.

As it was proven in [10], among various large-update IPMs based on SR-proximity functions, the SR-IPM with $p = 1$ and barrier degree $q = O(\log n)$ has the best worst-case complexity result. Therefore, in this paper we focus on the kernel function:

$$\psi_\ell(t) = \Gamma_{1,1+\log n}(t) = \frac{t^2 - 1}{2} + \frac{t^{-\log n} - 1}{\log n} = \frac{t^2 - 1}{2} + \frac{n^{-\log t} - 1}{\log n}.$$

Thus, the proximity function induced by this kernel function is given by

$$\Phi_\ell(x, s, \mu) = \Psi_\ell(v) = \frac{\|v\|^2 - n}{2} + \frac{\left\| v^{\frac{-\log n}{2}} \right\|^2 - n}{\log n} = \frac{\|v\|^2 - n}{2} + \sum_{i=1}^n \frac{n^{-\log v_i} - 1}{\log n}. \tag{7}$$

2 Properties of the SR proximity function $\Psi_\ell(v)$

In this section we recall from [17] some properties of the SR-proximity function (7) as a function of μ, when x and s are fixed positive vectors. For notational convenience let $\mu_g = \frac{x^T s}{n}$, the so called duality gap parameter, and let

$$\mu_h^\ell = \left(\frac{n}{\left(x^{\frac{-\log n}{2}}\right)^T s^{\frac{-\log n}{2}}} \right)^{\frac{2}{\log n}}. \tag{8}$$

It is easy to verify that the proximity function $\Psi_\ell(v)$ is convex w.r.t. parameter μ. Using the optimality conditions of convex optimization problems, one can easily prove the following theorem.

Theorem 2.1 (**Proposition 2.1 in [17]**) *For any fixed $(x,s) > 0$, the SR-proximity function $\Psi_\ell(v)$ is convex w.r.t. μ and it has a unique global minimizer at*

$$\mu_*^\ell = \left(\mu_g \left(\mu_h^\ell\right)^{\frac{\log n}{2}}\right)^{\frac{2}{2+\log n}}. \tag{9}$$

Theorem 2.1 implies that $\Psi_\ell(v)$ is a decreasing function w.r.t. μ when $\mu \le \mu_*^\ell$, and it is an increasing function w.r.t. μ if $\mu > \mu_*^\ell$. By considering the relation between the arithmetic and generalized harmonic means, one can easily prove the following relation.

$$\mu_h^\ell \le \mu_*^\ell \le \mu_g.$$

The following lemma plays an important role in the definition of our SR neighborhood.

Lemma 2.2 (**Lemma 3.3 in [17]**) *Let $\tau \ge 2$ be a constant. Then, for the proximity function $\Psi_\ell(v)$ induced by the SR-kernel function $\Gamma_{1,\ell}(t)$, the following statements are equivalent:*
i) $\frac{\mu_g}{\mu_h^\ell} \le \tau$,
ii) $\Phi_\ell\left(x, s, \frac{\mu_g}{\tau}\right) \le \frac{(\tau-1)n}{2}$,
iii) $\Phi_\ell\left(x, s, \mu_g\right) \le \frac{\tau^{\frac{\log n}{2}}-1\ \ n}{\log n}$.

IIPMs originally use an infinity neighborhood, that is defined as

$$\mathcal{N}_\infty^-(\rho) := \left\{(x,s) > 0 \middle|\ \|(r_b, r_c)\| \le \frac{\beta\|(r_b^0, r_c^0)\|\mu_g}{\mu_g^0},\ \left\|(v^2-e)^-\right\|_\infty \le \rho\right\}, \tag{10}$$

where $a^- = \min\{a, 0\}$, $\beta \ge 1$ is a constant and $\rho \in (0,1)$ is a constant independent of n. We define the SR neighborhood in a way that it contains the infinity neighborhood $\mathcal{N}_\infty^-(\rho)$ and these two neighborhoods almost match each other [13]. An infeasible SR neighborhood can be defined as

$$\mathcal{N}(\tau, \beta) := \left\{(x,s) > 0 \middle|\ \|(r_b, r_c)\| \le \frac{\beta\|(r_b^0, r_c^0)\|\mu_g}{\mu_g^0},\ \Phi_\ell(x, s, \mu_g) \le \eta(n,\tau)\right\}, \tag{11}$$

where $\eta(n,\tau) = \frac{\tau^{\frac{\log n}{2}}-1\ \ n}{\log n}$ is a positive function that depends on a constant τ and the dimension of the underlying problem. Now we describe how $\eta(n,\tau)$ has been derived. Assume that $(x,s) \in \mathcal{N}_\infty^-(\rho)$, then for this (x,s) pair, with $\mu = \mu_g$, we have

$$\Phi_\ell(x,s,\mu_g) = \frac{e^T v^{-\log n} - n}{\log n} \le \frac{n(1-\rho)^{\frac{-\log n}{2}} - n}{\log n} = \frac{(\tau^{\frac{\log n}{2}} - 1)n}{\log n} = \eta(n,\tau),$$

where $\tau = \frac{1}{1-\rho}$. With this choice of τ the neighborhood $\mathcal{N}(\tau,\beta)$ contains the neighborhood $\mathcal{N}_\infty^-(\rho)$, and the equivalence relations outlined in Lemma 2.2 can be utilized in the sequel.

Now we are interested to determine what is the smallest μ value for which the equality $\Phi_\ell(x,s,\mu) = \frac{(\tau-1)n}{2}$ holds. We denote this value by $\mu = \mu_t^\ell$. It is easy to verify (see [17]) that μ_t^ℓ exists whenever $\Phi_\ell(x,s,\mu_g) \le \eta(n,\tau)$ and it is the smaller positive root of the equation

$$f(\mu) = 2\left(x^{\frac{-\log n}{2}}\right)^T s^{\frac{-\log n}{2}} \mu^{\frac{2+\log n}{2}} - (2+\tau\log n)\, n\mu + x^T s \log n = 0. \tag{12}$$

Since $n \ge 1$ and $(x,s) > 0$, the function $f(\mu)$ is a convex function of μ. Thus, whenever $\Phi_\ell(x,s,\mu_g) \le \eta(n,\tau)$, equation (12) has two positive real roots, one is less than or equal to μ_*^ℓ and the other is greater than or equal to μ_*^ℓ. We choose the smaller root of this equation as a default value for μ_t^ℓ. Assuming $\frac{\mu_g}{\mu_h^\ell} \le \tau$, Lemma 2.2 and the decreasing property of $\Psi_\ell(v)$ w.r.t. $\mu \le \mu_*^\ell$ imply that

$$\Phi_\ell(x,s,\mu_h^\ell) \le \Phi_\ell\left(x,s,\frac{\mu_g}{\tau}\right) \le \frac{(\tau-1)\, n}{2} = \Phi_\ell(x,s,\mu_t^\ell).$$

Hence, when $\mu_g \le \tau\mu_h^\ell$, we have $\mu_t^\ell \le \mu_h^\ell \le \mu_*^\ell$, and equality holds if $\mu_g = \tau\mu_h^\ell$. The relation between μ_t^ℓ and μ_g is established in the following lemma.

Lemma 2.3 *Let $\tau \ge 2$ and $\Phi_\ell(x,s,\mu_g) \le \eta(n,\tau)$. Then, we have $\tau \le \frac{\mu_g}{\mu_t^\ell} \le \tau + \frac{2}{\log n}$.*

Proof: The function $f(\mu)$ given by (12) is a convex function of μ, thus, by substituting $\mu = \frac{\mu_g \log n}{2+\tau\log n}$, one can show that

$$f\left(\frac{\mu_g \log n}{2+\tau\log n}\right) = \frac{2n\left(\mu_g \log n\right)^{\frac{2+\log n}{2}}}{\left(\mu_h^\ell\right)^{\frac{\log n}{2}}\left(2+\tau\log n\right)^{\frac{2+\log n}{2}}} \ge 0.$$

So, we have $\mu_g \le \left(\tau + \frac{2}{\log n}\right)\mu_t^\ell$. Analogously, one can prove that $f\left(\frac{\mu_g}{\tau}\right) \le 0$ and therefore we have $\mu_g \ge \tau\mu_t^\ell$. □

The following remark is crucial in our algorithm design.

Remark 2.4 *In IIPMs it is crucial to keep balance between infeasibility and the complementarity gap. To enhance our ability to influence this balance we introduce a modified Newton system. When the iterates are not in a certain neighborhood of the central path, then instead of (6), our algorithm uses the following system of equations*

$$\begin{aligned} A\Delta x &= -r_b, \\ A^T\Delta y + \Delta s &= -r_c, \\ x\Delta s + s\Delta x &= -\frac{\mu}{2} v\nabla\Psi(v). \end{aligned} \tag{13}$$

The reduction of the right hand side of the third equation in (6) is crucial in order to get a reasonable step size which enables us to prove better iteration complexity. Then system (6) and (13) can be written in a modified form

$$\begin{aligned} A\Delta x &= -r_b, \\ A^T\Delta y + \Delta s &= -r_c, \\ x\Delta s + s\Delta x &= -\frac{\mu}{\chi} v\nabla\Psi(v), \end{aligned} \tag{14}$$

where $\chi = 1$ or 2.

Now we proceed to discuss the change of the complementary gap along the SR search direction for different updates of μ. For this, we rewrite (14) in the original space. Due to our specific choice of the SR proximity, system (14) has the form

$$\begin{aligned} A\Delta x &= -r_b, \\ A^T\Delta y + \Delta s &= -r_c, \\ s\Delta x + x\Delta s &= \frac{1}{\chi}(\mu^{\frac{2+\log n}{2}} x^{\frac{-\log n}{2}} s^{\frac{-\log n}{2}} - xs). \end{aligned} \tag{15}$$

Let us denote by $(\Delta x(\mu), \Delta y(\mu), \Delta s(\mu))$ the solution of system (15). The following results show how the complementary gap changes with the choice of the targeted parameter μ.

Lemma 2.5 (Lemmas 3.4, 3.6 and 3.9 in [18]) *Let μ_h^ℓ, μ_*^ℓ, and μ_t^ℓ be defined by (8), (9) and (12), respectively. Then*
1) $x^T\Delta s(\mu_*^\ell) + s^T\Delta x(\mu_*^\ell) = 0$,
2) $s^T\Delta x(\mu_h^\ell) + x^T\Delta s(\mu_h^\ell) = \frac{n}{\chi}\left(\mu_h^\ell - \mu_g\right)$,
3) $s^T\Delta x(\mu_t^\ell) + x^T\Delta s(\mu_t^\ell) = \frac{1}{\chi}\left(\left(\mu_t^\ell\right)^{\frac{2+\log n}{2}}\left(x^{\frac{-\log n}{2}}\right)^T s^{\frac{-\log n}{2}} - x^T s\right)$.

By using the previous lemma one can easily derive the following theorem as it is derived in [17].

Theorem 2.6 *Let μ_h^ℓ, μ_*^ℓ and μ_t^ℓ be defined by (8), (9) and (12). The following statements hold.*
1) If the target parameter μ is μ_*^ℓ, then for any feasible step size α, we have

$$\left(x + \alpha\Delta x(\mu_*^\ell)\right)^T\left(s + \alpha\Delta s(\mu_*^\ell)\right) = x^T s\left(1 + \alpha^2\frac{\Delta x(\mu_*^\ell)^T\Delta s(\mu_*^\ell)}{x^T s}\right).$$

2) If the target parameter μ is μ_h^ℓ, then for any feasible step size α, we have

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

3) If the target parameter μ is μ_t^ℓ, then for any feasible step size α, we have

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

3 An Adaptive Large Update SR-IIPM

In this section a new variant of SR-IIPMs is considered. For simplicity, we use the notations

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

Correspondingly, we define

$$\mu_g(\alpha) = \frac{x(\alpha)^T s(\alpha)}{n}, \qquad \mu_h^\ell(\alpha) = \left(\frac{n}{\left((x(\alpha))^{\frac{-\log n}{2}} \right)^T (s(\alpha))^{\frac{-\log n}{2}}} \right)^{\frac{2}{\log n}}$$

and

$$\mu_*^\ell(\alpha) = \left(\mu_g(\alpha) (\mu_h^\ell(\alpha))^{\frac{\log n}{2}} \right)^{\frac{2}{2+\log n}}.$$

For convenience, we use the notation v as defined by (5) and the *scaled search directions*

$$d_x := \frac{v\Delta x}{x} \text{ and } d_s := \frac{v\Delta s}{s}.$$

Thus, the third equation of system (14) can be rewritten as

$$d_x + d_s = -\frac{1}{\chi}\nabla\Psi_\ell(v).$$

We also define the norm-based proximity measure σ_ℓ by

$$\sigma_\ell = \|\nabla\Psi_\ell(v)\| = \|v - v^{-1-\log n}\|. \tag{16}$$

In our algorithm, regardless of the iterate is close to, or far away from the central path, we always make an adaptive large update $\mu_+ = (1-\theta)\mu$ of the central path parameter μ and we use the neighborhood given by (11). We utilize a constant $\tau \geq e^2$ to keep control on the distance of the iterate to the central path and to force the value of the proximity function to satisfy the following relation

$$\Phi_\ell(x, s, \mu_g) \leq \eta(n, \tau). \tag{17}$$

At each step we stipulate that the step size should be chosen so that the proximity function $\Phi_\ell(x(\alpha),\ s(\alpha),\ \mu_+)$ has a sufficient decrease, while the new iterate is still in $\mathcal{N}(\tau, \beta)$.

Thus, the new algorithm can be outlined as follows.

Algorithm SR-IIPM

Input:
Proximity parameters $\tau \geq e^2$;
$\eta_1(n, \tau, \sigma_\ell^k)$, a positive function, is given by (39);
parameter $\beta \geq 1$;
neighborhood $\mathcal{N}(\tau, \beta)$;
an accuracy parameter $\varepsilon > 0$;
$(x^0, s^0) \in \mathcal{N}(\tau, \beta)$; $k = 0$.
begin
  while $\max\left\{(x^k)^T s^k, \left\|r_b^k\right\|, \left\|r_c^k\right\|\right\} \geq \epsilon$ **do**
  begin
    $\mu := \left(\mu_t^\ell\right)^k$ defined by (12);
    Let $\chi_k = 1$ when $\frac{\mu_g^k}{\left(\mu_t^\ell\right)^k} \leq \tau + \eta_1(n, \tau_1, \sigma_\ell)$
    and $\chi_k = 2$ when $\frac{\mu_g^k}{\left(\mu_t^\ell\right)^k} > \tau + \eta_1(n, \tau_1, \sigma_\ell)$.
    Solve system (15) for $(\Delta x^k, \Delta y^k, \Delta s^k)$.
    begin
      Determine the step size α_k such that
      $\alpha_k = \max\left\{\alpha \middle| \Phi_\ell^+ \leq \Phi_\ell - \frac{(\sigma_\ell^k)^2}{4\chi_k}\alpha; (x(\alpha), s(\alpha)) \in \mathcal{N}(\tau, \beta)\right\}$[a];
      $x^{k+1} := x(\alpha_k)$; $y^{k+1} := y(\alpha_k)$; $s^{k+1} := s(\alpha_k)$;
      $k = k + 1$.
    end
  end
end

---

[a] $\Phi_\ell := \Phi_\ell(x^k, s^k, (\mu_t^\ell)^k)$, $\Phi_\ell^+ := \Phi_\ell(x(\alpha), s(\alpha), (\mu_t^\ell)^k)$.

We proceed to analyze the complexity of the algorithm. The key element of the analysis is to estimate the value of the step size α in Algorithm SR-IIPM. To do so, we need to estimate the second order term $\frac{\left(\Delta x^k\right)^T \Delta s^k}{\mu^k}$ at the k^{th} iterate. Let

$$\varphi_k = \prod_{i=1}^{k-1}(1 - \alpha_i), \ k = 1, 2, \cdots, \tag{18}$$

where α_i is the step size in the i^{th} iterate. Because the first two components of system (15) are linear, we have

$$(r_b^k, r_c^k) = \varphi_k (r_b^0, r_c^0). \tag{19}$$

Since $(x^k, s^k) \in \mathcal{N}(\tau, \beta)$, from (18) we derive

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

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

$$\varphi_k \leq \frac{\beta \mu_g^k}{\mu_g^0}. \tag{20}$$

To prove polynomial complexity of Algorithm SR-IIPM one need to prove a lower bound on the step size which is an inverse polynomial function of the dimension n. As it is customary [21] in IIPMs, we choose the starting point to satisfy

$$(x^0, y^0, s^0) = (\zeta e, 0, \zeta e), \tag{21}$$

where ζ is a scalar for which

$$\|(x^*, s^*)\|_\infty \leq \zeta, \tag{22}$$

for some primal-dual optimal solution (x^*, y^*, s^*). Usually we do not know the value $\|(x^*, s^*)\|_\infty$, because we do not know any optimal solution 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 reasons: a well-centered starting point for which the ratio

$$\frac{\|(r_b^0, r_c^0)\|}{\mu^0} \tag{23}$$

is small, leads to faster convergence than poorly centered points that are much closer to the solution set. The point (21) satisfies these criteria. It is perfectly centered, and the ratio (23) is bounded from above.

3.1 Some Technical Results

We mention that the orthogonality of the vectors Δx and Δs does not hold in IIPMs. In this section, we estimate the second order term $\frac{(\Delta x^k)^T \Delta s^k}{\mu^k}$. First we provide a lower bound for $v_{\min}^k = \min\{v_i^k |\ 1 \leq i \leq n\}$, the smallest coordinate of the vector v^k. This result will be used frequently in the rest of the paper (e.g., in the proof of Lemma 3.2).

Lemma 3.1 *Suppose that the present iterate (x^k, s^k) is in the neighborhood $\mathcal{N}(\tau, \beta)$. Let $\mu = (\mu_t^\ell)^k$, $n \geq 5$ and $\tau \geq e^2$, then*

$$(\sigma_\ell^k)^2 \geq (\tau - 1)n \geq 31, \tag{24}$$

$$v_{\min}^k \geq (e\tau)^{-1}, \tag{25}$$

$$(v_{\min}^k)^{1+\log n} \sigma_\ell^k \geq \frac{8}{10}. \tag{26}$$

Proof: By Proposition 3.1.5 of [11] we have

$$(\sigma_\ell^k)^2 \geq 2\Phi_\ell\left(x^k, s^k, (\mu_t^\ell)^k\right) = (\tau - 1)n,$$

that proves the first statement of the lemma. Using Lemma 2.3 and the fact that $(\mu_t^\ell)^k \leq (\mu_*^\ell)^k$, we have

$$(v_{\min}^k)^{-\log n} \leq \left\| (v^k)^{-\frac{\log n}{2}} \right\|^2 \leq \|v^k\|^2 = \frac{(x^k)^T s^k}{(\mu_t^\ell)^k} \leq n\left(\tau + \frac{2}{\log n}\right),$$

that for $\tau \geq e^2,\ n \geq 5$ implies

$$v_{\min}^k \geq e^{-1}\left(\tau + \frac{2}{\log n}\right)^{\frac{-1}{\log n}} \geq (\tau e)^{-1}.$$

This way (25) is proved. To prove (26) we recall again Proposition 3.1.5 of [11] and inequality (24), that imply

$$v_{\min}^k \geq (1+\sigma_\ell^k)^{\frac{-1}{1+\log n}} \quad \text{and} \quad \sigma_\ell^k \geq \sqrt{31}.$$

Therefore we have

$$(v_{\min}^k)^{1+\log n}\sigma_\ell^k \geq \frac{\sigma_\ell^k}{1+\sigma_\ell^k} \geq \frac{8}{10},$$

that completes the proof of the Lemma. □

The following lemma gives an upper bound for $\varphi_k\|(x^k, s^k)\|_1$, where φ_k is defined by (18) and this result will be used in the proof of Lemma 3.2. The proof is analogous to the proof of Lemma 5.1 in [19].

Lemma 3.2 *Suppose that $n \geq 5, \tau \geq e^2$ and the initial point (x^0, y^0, s^0) is chosen so that it satisfies (21) and (22), and for the current iterate $(x^k, s^k) \in \mathcal{N}(\tau, \beta)$ holds. Then*

$$\zeta\varphi_k\|(x^k, s^k)\|_1 \leq C_1 n\mu_g^k,$$

where $C_1 = 3\beta + 1$.

Let us define $D^k = \operatorname{diag}\left((x^k)^{\frac{1}{2}}(s^k)^{-\frac{1}{2}}\right)$. We also use the matrix norm defined for a matrix $M \in \mathbb{R}^{p\times q}$ as

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

It is known that $\|Mu\| \leq \|M\|\|u\|$, for all $u \in \mathbb{R}^q$. By the following lemma we give upper bounds for the scaled vectors $(D^k)^{-1}\Delta x^k$ and $D^k\Delta s^k$ that enable us to bound $(\Delta x^k)^T\Delta s^k$.

Lemma 3.3 *Suppose that $n \geq 5,\ \tau \geq e^2$, the initial point is chosen so that it satisfies (21) and (22), and for the current iterate $(x^k, s^k) \in \mathcal{N}(\tau, \beta)$ holds. Then,*

$$\|(D^k)^{-1}\Delta x^k\| \leq C_2\sqrt{n\mu_g^k}\sigma_\ell^k, \quad \|D^k\Delta s^k\| \leq C_2\sqrt{n\mu_g^k}\sigma_\ell^k, \tag{27}$$

where $C_2 = \sqrt{6}C_1 e\tau + 1$.

Proof: See the Appendix. □

Corollary 3.4 *Suppose that the assumptions of Lemma 3.3 are satisfied. Then*

$$\frac{\left|\left(\Delta x^k\right)^T \Delta s^k\right|}{\mu_g^k} \leq C_2^2 n(\sigma_\ell^k)^2. \tag{28}$$

Proof: By Lemma 3.3 we have

$$\left|\left(\Delta x^k\right)^T \Delta s^k\right| = \left|\left((D^k)^{-1}\Delta x^k\right)^T D^k \Delta s^k\right| \leq C_2^2 \mu_g^k n(\sigma_\ell^k)^2$$

that completes the proof. □

The following lemma will be used to derive a lower bound for the step size in Algorithm SR-IIPM.

Lemma 3.5 *Suppose that $n \geq 5$, $\tau \geq e^2$, the initial point is chosen so that it satisfies (21) and (22), and for the current iterate $(x^k, s^k) \in \mathcal{N}(\tau, \beta)$ holds. Then,*

$$(\sigma_1^k)^2 \leq C_3 n(\sigma_\ell^k)^2, \tag{29}$$

where $(\sigma_1^k)^2 := \|d_{x^k}\|^2 + \|d_{s^k}\|^2$ and $C_3 = 2(\tau+1)C_2^2 + 1$.

Proof: Using (16) we have

$$\begin{aligned}(\sigma_\ell^k)^2 = \chi_k^2\|d_{x^k} + d_{s^k}\|^2 &= \chi_k^2(\|d_{x^k}\|^2 + \|d_{s^k}\|^2) + 2\chi_k^2 \left(d_{x^k}\right)^T d_{s^k} \\ &= \chi_k^2(\sigma_1^k)^2 + 2\chi_k^2 \frac{\left(\Delta x^k\right)^T \Delta s^k}{(\mu_t^\ell)^k}.\end{aligned} \tag{30}$$

Thus, by (28) we have

$$\begin{aligned}(\sigma_1^k)^2 &= \frac{(\sigma_\ell^k)^2}{\chi_k^2} - 2\frac{\left(\Delta x^k\right)^T \Delta s^k}{(\mu_t^\ell)^k} \leq \frac{(\sigma_\ell^k)^2}{\chi_k^2} + 2\frac{\left|\left(\Delta x^k\right)^T \Delta s^k\right|}{(\mu_t^\ell)^k} \\ &\leq \frac{(\sigma_\ell^k)^2}{\chi_k^2} + 2(\tau+1)\frac{\left|\left(\Delta x^k\right)^T \Delta s^k\right|}{\mu_g^k} \\ &\leq \frac{(\sigma_\ell^k)^2}{\chi_k^2} + 2(\tau+1)C_2^2 n(\sigma_\ell^k)^2 \leq C_3 n(\sigma_\ell^k)^2,\end{aligned}$$

where the second inequality follows from Lemma 2.3 and the third inequality follows from Corollary 3.4. The proof of the lemma is completed. □

3.2 Estimating the step size

In this section we discuss in details how to derive a lower bound for the step size in the specified neighborhood that leads us to polynomiality of the algorithm. The following theorem provides a lower bound for the maximal feasible step size and gives a certain reduction for the proximity function.

Theorem 3.6 *Let $n \geq 5,\ \tau \geq e^2,\ (x^k, s^k) \in \mathcal{N}(\tau, \beta)$, and $(\Delta x^k, \Delta y^k, \Delta s^k)$ be the solution of system (15) with $\mu^k = \left(\mu_t^\ell\right)^k$ and χ_k is chosen as in Algorithm SR-IIPM. Then, the maximal feasible step size $\alpha_{\max}$ satisfies*

$$\alpha_{\max} \geq \bar{\alpha}_k = \frac{v_{\min}^k}{\sigma_1^k} \geq \frac{1}{e\tau\sigma_1^k}.$$

Moreover, for any step size

$$\alpha \leq \bar{\alpha}_k^* = \frac{4\sigma_\ell^k}{e\tau\sigma_1^k \log n \left(10\chi_k\sigma_1^k + 8\sigma_\ell^k\right)} \tag{31}$$

the following relation holds:

$$\Phi_\ell\left(x(\alpha), s(\alpha), \left(\mu_t^\ell\right)^k\right) \leq \Phi_\ell\left(x^k, s^k, \left(\mu_t^\ell\right)^k\right) - \frac{(\sigma_\ell^k)^2}{4\chi_k}\alpha. \tag{32}$$

Proof: See the Appendix. □

Corollary 3.7 *Suppose that all the assumptions of Theorem 3.6 are satisfied. Then*

$$\hat{\alpha}_k = \frac{4}{11\chi_k e\tau C_3 n\sigma_\ell^k \log n} \tag{33}$$

is strictly feasible and (32) also holds for $\alpha = \hat{\alpha}_k$.

Proof: By (29) one has

$$\bar{\alpha}_k^* \geq \hat{\alpha}_k = \frac{4}{11\chi_k e\tau C_3 n\sigma_\ell^k \log n},$$

that completes the proof. □

In what follows we further analyze whether $\hat{\alpha}_k$ can be used as the default step size, or it has to be somehow reduced. To do so, we need to derive an upper bound for σ_ℓ^k. This bound also allows us to find lower and upper bounds for the complementarity gap, which is important to control the iterates and to derive the iteration bound. For simplicity we omit the super and subscripts k in this lemma.

Lemma 3.8 *Let $n \geq 5,\ \tau \geq e^2$. If $\Psi_\ell(v) \leq \frac{(\tau-1)n}{2}$, then*

$$\sigma_\ell \leq \frac{3}{2}e\tau^{1+\frac{1}{\log n}} n \log n. \tag{34}$$

Proof: To derive an upper bound for σ_ℓ we need to find the optimal value of the problem:

$$\begin{aligned} &\max\ \sigma_\ell^2 \\ &\text{s.t.}\quad \Psi_\ell(v) \leq \frac{(\tau-1)\,n}{2}. \end{aligned}$$

This can be done by solving the following system of optimality conditions:

$$
\begin{aligned}
\psi_\ell'(v_i)\left(\psi_\ell''(v_i)-\lambda\right) &= 0, \quad i=1,\ldots,n,\\
\Psi_\ell(v) &\le \frac{(\tau-1)\,n}{2},\\
\lambda\left(\Psi_\ell(v)-\frac{(\tau-1)\,n}{2}\right) &= 0.
\end{aligned}
\tag{35}
$$

It is obvious that $\lambda=0$ if and only if $v=e$, which implies $\sigma_\ell=0$. Now, if $v\neq e$, and therefore $\lambda\neq 0$, then the following two cases may happen in the solution of system (35).
(i) $v_1=\ldots=v_{n-j}=1$, $v_{n-j+1}=\ldots=v_n>1$,
(ii) $v_1=\ldots=v_{n-j}=1$, $v_{n-j+1}=\ldots=v_n<1$.
For case (i) we have

$$
\frac{jv_n^2-j}{2}-\frac{n}{\log n}\le\Psi_\ell(v)=\frac{(\tau-1)\,n}{2},
$$

that implies $v_n^2\le\frac{n\log n(\tau-1)+2n}{j\log n}+1$. Thus,

$$
\begin{aligned}
\sigma_\ell^2=\sum_{i=1}^{n}\left(v_i-v_i^{-1-\log n}\right)^2 &= j\left(v_n-v_n^{-1-\log n}\right)^2\le jv_n^2\\
&\le\frac{n\tau\log n+2n}{\log n}\le\frac{3}{2}e\tau\tau^{\frac{1}{\log n}}n\log n.
\end{aligned}
$$

For case (ii) we have

$$
\frac{jv_n^{-\log n}-j}{\log n}-\frac{n}{2}\le\Psi_\ell(v)=\frac{(\tau-1)\,n}{2},
$$

that implies $v_n^{-\log n}\le\frac{n\tau\log n+2j}{2j}$. Thus,

$$
\begin{aligned}
\sigma_\ell^2 &= j\left(v_n-v_n^{-1-\log n}\right)^2\le jv_n^{-2-2\log n}=j\left(v_n^{-\log n}\right)^{\frac{2+2\log n}{\log n}}\\
&\le j\left(\frac{n\tau\log n+2j}{2j}\right)^{\frac{2+2\log n}{\log n}}\le\left(\frac{n\tau\log n+2n}{2}\right)\left(\frac{n\tau\log n+2}{2}\right)^{\frac{2+\log n}{\log n}}.
\end{aligned}
$$

This upper bound is larger than the bound derived in (i), therefore, we have the following upper bound for σ_ℓ,

$$
\sigma_\ell\le\frac{3}{2}e\tau^{1+\frac{1}{\log n}}n\log n.
$$

This completes the proof. □

In the sequel, we will see that in the worst case one might not be able to determine a unique lower bound for the step size in the entire neighborhood i.e., in some part of the neighborhood one can have much better lower bound of the step size than in the other part. It will be also motivated why for an iterate in $\mathcal{N}(\tau,\beta)$ we use system (15) with $\chi_k=1$, if $\frac{\mu_g^k}{(\mu_t^\ell)^k}\le\tau+\eta_1(n,\tau_1,\sigma_\ell)$, and with $\chi_k=2$ if $\frac{\mu_g^k}{(\mu_t^\ell)^k}<\tau+\eta_1(n,\tau_1,\sigma_\ell)$.

Lemma 3.9 *Let $n \geq 5,\ \tau \geq e^2,\ (x^k, s^k) \in \mathcal{N}(\tau, \beta),\ \frac{\mu_g^k}{(\mu_h^\ell)^k} \leq 0.99\tau$ and $(\Delta x^k, \Delta y^k, \Delta s^k)$ be the solution of system (15) with $\mu = \left(\mu_t^\ell\right)^k$ and χ_k is chosen as it Algorithm SR-IIPM. Then $\alpha_k = \dfrac{4}{11\chi_k e\tau C_3 n\sigma_\ell^k \log n}$ is strictly feasible and*

$$\Phi_\ell(x(\alpha_k), s(\alpha_k), \mu_g(\alpha_k)) \leq \eta(n, \tau).$$

Proof: See the Appendix. □

The following lemma deals with the case when $\mu_g^k = \tau(\mu_h^\ell)^k$.

Lemma 3.10 *Let $n \geq 5,\ \tau \geq e^2,\ (x^k, s^k) \in \mathcal{N}(\tau, \beta),\ \frac{\mu_g^k}{(\mu_h^\ell)^k} = \tau$ and $(\Delta x^k, \Delta y^k, \Delta s^k)$ be the solution of system (15) with $\mu = \left(\mu_t^\ell\right)^k$ and χ_k is chosen as it Algorithm SR-IIPM. Then $\alpha_k = \dfrac{4}{11\chi_k e\tau^2 C_3 n\sigma_\ell^k \log n}$ is strictly feasible and*

$$\Phi_\ell(x(\alpha_k), s(\alpha_k), \mu_g(\alpha_k)) \leq \eta(n, \tau).$$

Proof: See the Appendix. □

In the following Lemma we show that for a certain upper bound of the step size the inequality corresponding to the residuals in the definition of $\mathcal{N}(\tau, \beta)$ holds. The proof of the lemma gives the main reason why the Newton system is changed in a certain part of the neighborhood.

Lemma 3.11 *Let $n \geq 5,\ \tau \geq e^2,\ (x^k, s^k) \in \mathcal{N}(\tau, \beta)$ and $(\Delta x^k, \Delta y^k, \Delta s^k)$ be the solution of system (15) with $\mu = \left(\mu_t^\ell\right)^k$ and χ_k is chosen as specified in Algorithm SR-IIPM. Then for $\alpha_k \leq \alpha_k^* := \dfrac{4}{11e\tau^2 C_3 n\sigma_\ell^k \log n}$ as the step size one has*

$$\frac{\left\|(r_b^{k+1}, r_c^{k+1})\right\|}{\mu_g^{k+1}} \leq \beta \frac{\|(r_b^0, r_c^0)\|}{\mu_g^0}. \tag{36}$$

Proof: Since after each step with α_k as the step size one has

$$r_b^{k+1} = (1 - \alpha_k) r_b^k \quad \text{and} \quad r_c^{k+1} = (1 - \alpha_k) r_c^k,$$

it suffices to show that

$$\mu_g(\alpha_k) \geq (1 - \alpha_k)\mu_g^k.$$

Let us consider the case when $\alpha_k = \alpha_k^*$. Then from the third part of Theorem 2.6, we know that

$$\mu_g(\alpha_k^*) = \mu_g^k \left(1 - \frac{\alpha_k^*}{\chi_k} + \frac{\alpha_k^* \left(\left(\mu_t^\ell\right)^k\right)^{\frac{2+\log n}{2}}}{\chi_k \mu_g^k \left(\left(\mu_h^\ell\right)^k\right)^{\frac{\log n}{2}}} + (\alpha_k^*)^2 \frac{(\Delta x^k)^T \Delta s^k}{n\mu_g^k} \right). \tag{37}$$

Let us define

$$Z(\alpha) := \frac{\left(\left(\mu_t^\ell\right)^k\right)^{\frac{2+\log n}{2}}}{\chi_k \mu_g^k \left(\left(\mu_h^\ell\right)^k\right)^{\frac{\log n}{2}}} + \alpha \frac{(\Delta x^k)^T \Delta s^k}{n\mu_g^k}. \tag{38}$$

Using (12) we have

$$Z(\alpha_k^*) \geq \frac{(2+\tau \log n)\left(\mu_t^\ell\right)^k - \mu_g^k \log n}{2\chi_k \mu_g^k} - \alpha_k^* \frac{|(\Delta x^k)^T \Delta s^k|}{n\mu_g^k}.$$

Applying the assumptions of the theorem and relations (28) and (34) to this inequality by $\chi_k = 1$ one has

$$Z(\alpha_k^*) \geq 0,$$

if $\frac{\mu_g^k}{\left(\mu_t^\ell\right)^k} \leq \tau + \eta_1(n, \tau, \sigma_\ell^k)$, where

$$\eta_1(n, \tau, \sigma_\ell^k) = \frac{22C_3 e\tau^2 n \log n - 8\tau^2 C_2^2 \sigma_\ell^k}{11C_3 e\tau^2 n(\log n)^2 + 8C_2^2 \sigma_\ell^k}. \tag{39}$$

This implies the statement of the theorem when $\frac{\mu_g^k}{\left(\mu_t^\ell\right)^k} \leq \tau + \eta_1(n, \tau, \sigma_\ell^k)$. Now let us consider the case when $\tau + \eta_1(n, \tau, \sigma_\ell^k) < \frac{\mu_g^k}{\left(\mu_t^\ell\right)^k}$. By letting $\chi_k = 2$ in (37) and using (38) one has

$$\mu_g(\alpha_k^*) = \mu_g^k \left(1 - \frac{\alpha_k^*}{2} + \alpha_k^* Z(\alpha_k^*)\right).$$

Therefore, to show that $\mu_g(\alpha_k^*) \geq (1 - \alpha_k^*)\mu_g^k$, it suffices to have $Z(\alpha_k^*) \geq -\frac{1}{2}$. Since $\frac{\mu_g^k}{\left(\mu_t^\ell\right)^k} \leq \tau + \frac{2}{\log n}$, the inequality $Z(\alpha_k^*) \geq -\frac{1}{2}$ follows easily from (28), (34) and from the definition of α_k^*. Finally, for any $\alpha_k \leq \alpha_k^*$ one has

$$\begin{aligned} Z(\alpha_k) &\geq \frac{(2+\tau \log n)\left(\mu_t^\ell\right)^k - \mu_g^k \log n}{2\chi_k \mu_g^k} - \alpha_k \frac{|(\Delta x^k)^T \Delta s^k|}{n\mu_g^k} \\ &\geq \frac{(2+\tau \log n)\left(\mu_t^\ell\right)^k - \mu_g^k \log n}{2\chi_k \mu_g^k} - \alpha_k^* \frac{|(\Delta x^k)^T \Delta s^k|}{n\mu_g^k}. \end{aligned}$$

Using this inequality analogous to the case when $\alpha_k = \alpha_k^*$ one can show that $\mu_g(\alpha_k) \geq (1 - \alpha_k)\mu_g^k$ that completes the proof of the lemma. □

Remark 3.12 *From now on, we call a step good if the maximum step size α_k in $\mathcal{N}(\tau, \beta)$ for which inequality (32) holds is greater than or equal to α_k^*.*

Remark 3.13 *If after a step the new iterate is in* $\mathcal{N}(\tau, \beta)$ *with* $\frac{\mu_g^k}{(\mu_h^\ell)^k} \leq 0.99\tau$ *or* $\frac{\mu_g^k}{(\mu_h^\ell)^k} = \tau$*, then the results of Lemmas 3.9, 3.10 and 3.11 give an estimate of the step size that leads to the polynomiality of the algorithm. However the algorithm might get to a point for which* $0.99\tau < \frac{\mu_g^k}{(\mu_h^\ell)^k} < \tau$*, because if we would make a large step toward the boundary of the neighborhood, then inequality (32) might get violated before we hit the boundary. In this case the algorithm might do a slightly shorter step that will be analyzed in the sequel.*

In what follows we discuss the cases that can happen when for the current iterate $0.99\tau < \frac{\mu_g^k}{(\mu_h^\ell)^k} < \tau$.

Case 1: The maximum step size of the next iterate is greater than or equal to α_k^*. Then the algorithm makes a good step.

Case 2: The algorithm might make a step with $\alpha_k < \alpha_k^*$ for which $\frac{\mu_g^k(\alpha_k)}{\mu_h^\ell(\alpha_k)} \leq 0.99\tau$. Then by Lemma 3.8 we know that the next step will be again a good step.

Case 3: The algorithm might make a step with $\alpha_k < \alpha_k^*$ for which $\frac{\mu_g^k(\alpha_k)}{\mu_h^\ell(\alpha_k)} = \tau$. Then by Lemma 3.9 we know that the next step will be again a good step.

Case 4: The algorithm might make a step with $\alpha_k < \alpha_k^*$ for which $0.99\tau < \frac{\mu_g^k(\alpha_k)}{\mu_h^\ell(\alpha_k)} < \tau$.

Case 4 is the very crucial one which is discussed in the following lemma.

Lemma 3.14 *Suppose that for the current iterate in* $\mathcal{N}(\tau, \beta)$*,* $0.99\tau < \frac{\mu_g^k}{(\mu_h^\ell)^k} < \tau$*. If the algorithm does a step with* $\alpha_k < \alpha_k^*$*, then Case 4 can not happen.*

Proof: Suppose the algorithm makes a step with $\alpha_k < \alpha_k^*$. Then by inequality (32) and Lemma 3.11 the only thing that can restricts the step size to be less than α_k^* is inequality (17). Then in this situation one can conclude that equality holds in (17) that implies $\frac{\mu_g^k(\alpha_k)}{\mu_h^\ell(\alpha_k)} = \tau$. Therefor, Case 4 can not happen that completes the proof of the lemma. □

The following Lemma is used to estimate the reduction of μ_t^ℓ after a good step. The proof is a direct consequence of Theorem 3.6 and Lemma 3.1.

Lemma 3.15 *Suppose that all the assumptions of Theorem 3.6 are satisfied. Then*

$$\Phi_\ell\left(x(\alpha_k^*), s(\alpha_k^*), \left(\mu_t^\ell\right)^k\right) \leq \Phi_\ell\left(x^k, s^k, \left(\mu_t^\ell\right)^k\right) - \frac{\sqrt{\tau - 1}}{11e\tau^2 C_3 n^{\frac{1}{2}} \log n}. \tag{40}$$

By summarizing the previous results, we have:

Theorem 3.16 *Let* $n \geq 5$*,* $\tau \geq e^2$*,* $(x^k, s^k) \in \mathcal{N}(\tau, \beta)$*, and* $(\Delta x^k, \Delta y^k, \Delta s^k)$ *be the solution of system (15) with* $\mu = \left(\mu_t^\ell\right)^k$ *and* χ_k *is chosen as specified in Algorithm SR-IIPM. Then, for the good step, the default step size* α_k^* *is strictly feasible and* $(x(\alpha_k^*), s(\alpha_k^*)) \in \mathcal{N}(\tau, \beta)$*. In the worst case each good step might be followed by at most one short step for which there exist a positive step size* $\alpha_k \leq \alpha_k^*$ *such that* $(x(\alpha_k), s(\alpha_k)) \in \mathcal{N}(\tau, \beta)$*.*

To obtain an upper bound for the total number of iterations of Algorithm SR-IIPM, we need to investigate the growth behavior of $\Phi_\ell(x^k, s^k, (\mu_t^\ell)^k)$ w.r.t. $(\mu_t^\ell)^k$. Since the step size in the short step in the worst case might be much smaller than the step size in the good step, therefore, we do not take into account how much μ_t^ℓ reduces at a short step, we just use the fact that it reduces at each short step.

By the choice of $(\mu_t^\ell)^k$ we know that the proximity function value $\Phi_\ell(x^k, s^k, (\mu_t^\ell)^k)$ keeps invariant for all iterates. Let us denote by $\left(\mu_t^\ell\right)^+$ the value of $(\mu_t^\ell)^k$ after one step, thus

$$\Phi_\ell(x^k, s^k, (\mu_t^\ell)^k) = \Phi_\ell\left(x(\alpha), s(\alpha), \left(\mu_t^\ell\right)^+\right) = \frac{(\tau-1)n}{2}. \tag{41}$$

The following theorem ensures the sufficient reduction of μ_t^ℓ at a good step.

Theorem 3.17 *Let $n \geq 5$, $\tau \geq e^2$, $(x^k, s^k) \in \mathcal{N}(\tau, \beta)$, and $(\Delta x^k, \Delta y^k, \Delta s^k)$ be the solution of system (15) with $\mu = \left(\mu_t^\ell\right)^k$ and χ_k chosen as it is specified in Algorithm SR-IIPM. Then after each good step one has*

$$\Phi_\ell\left(x(\alpha_k^*), s(\alpha_k^*), (1-\theta)\left(\mu_t^\ell\right)^k\right) \leq \Phi_\ell\left(x^k, s^k, \left(\mu_t^\ell\right)^k\right),$$

where

$$\theta = \frac{2\sqrt{\tau-1}}{11e(\tau+1)\tau^2 C_3 n^{\frac{3}{2}} \log n}. \tag{42}$$

Proof: See the Appendix. □

Corollary 3.18 *Under the assumptions of Theorem 3.17, we have*

$$\left(\mu_t^\ell\right)^+ \leq (1-\theta)\left(\mu_t^\ell\right)^k,$$

where θ is given by (42).

Proof: Using (41) and Theorem 3.17, one can easily conclude the statement. □

Corollary 3.19 *By Theorem 3.6 one has a certain reduction of the proximity measure independent of where the iterates are in $\mathcal{N}(\tau, \beta)$. Therefore, analogous result to Theorem 3.17 also holds after each short step.*

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

Theorem 3.20 *Let $n \geq 5$, $\tau \geq e^2$, $(x^0, s^0) \in \mathcal{N}(\tau, \beta)$, and $t_0 = \max\left\{1, \frac{\|(r_b^0, r_c^0)\|}{\mu_g^0}\right\}$. Then after at most*

$$O\left(n^{\frac{3}{2}} \log n \log \frac{\beta t_0(\tau+1)n}{\epsilon}\right)$$

iterations SR-IIPM terminates with a solution satisfying $(x^k)^T s^k \leq \epsilon$ and $\|(r_b^k, r_c^k)\| \leq \epsilon$.

Proof: Using Lemma II.17 [16], Corollary 3.18 and Corollary 3.19 we need at most

$$\left\lceil \frac{2}{\theta} \log \frac{\beta t_0(\tau+1)n}{\epsilon} \right\rceil$$

iterations to have $\left(\mu_t^\ell\right)^k \leq \frac{\epsilon}{\beta t_0 n(\tau+\frac{2}{\log n})}$, where θ is given by (42). Now, Lemma 2.3 implies that $\mu_g^k \leq (\tau + \frac{2}{\log n}) \left(\mu_t^\ell\right)^k \leq \frac{\epsilon}{n\beta t_0}$. Thus, we have $(x^k)^T s^k \leq \frac{\epsilon}{\beta t_0} \leq \epsilon$. Using Lemma 3.11, we also have $\|(r_b^k, r_c^k)\| \leq \epsilon$. □

4 Conclusion

In this paper based on a specific SR proximity function a new large-update IIPM is proposed. Our algorithm utilizes several properties of the proximity function w.r.t. an adaptively chosen target value μ_t^ℓ. An $O\left(n^{\frac{3}{2}} \log n \log \frac{n}{\epsilon}\right)$ worst-case iterations bound of our new algorithm is established. Our result improves the so far best $O\left(n^2 \log \frac{n}{\epsilon}\right)$ iterations bound of IIPMs in a large neighborhood of the central path. It is worth mentioning that for relatively large value of n the constant in the iteration complexity can be reduced. Extension of this approach to Second Order Conic, Semidefinite Optimization, and possibly to Nonlinear Complementarity problems, is left for the interested reader.

Acknowledgments: The research is supported by the NSERC Discovery Grant DG:5-48923, the Canada Research Chair program and by MITACS. A part of the research was done when the second author was visiting the Advanced Optimization Lab at McMaster University.

References

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

[2] J. Gondzio (1996). Multiple centerality corrections in a primal-dual method for linear programming, *Computational Optimization and Applications*, 6, pp. 137-156.

[3] N.K. Karmarkar (1984). A new polynomial-time algorithm for linear programming, *Combinatorica*, 4, pp. 373-395.

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

[5] I.J. Lustig (1990/91). Feasiblity issues in a primal-dual interior-point methods for linear programming, *Mathematical Programming*, 49, pp. 145-162.

[6] N. Megiddo (1989). Pathways to the optimal set in linear programming, In: N. Megiddo, editor, *Progress in Mathematical Programming: Interior Point and Related Methods*, pp. 131-158. Springer Verlag, New York.

[7] S. Mehrota (1992). On the implementation of a (primal-dual) interior point method, *SIAM Journal on Optimization*, 2, pp. 575-601.

[8] S. Mizuno (1994). Polynomiality of infeasible-interior-point algorithms for linear programming, *Mathematical Programming*, 67, pp. 109-119.

[9] J. Peng, C. Roos, and T. Terlaky (2001). A new and efficient large-update interior point method for linear optimzation, *Journal of Computational Technology*, 6, pp. 61-80.

[10] J. Peng, C. Roos, and T. Terlaky (2002). Self-regular functions and new search directions for linear and semidefinite optimization, *Mathematical Programming*, 93, pp. 129-171.

[11] J. Peng, C. Roos, and T. Terlaky (2002). *Self-Regularity: A New Paradigm for Primal-Dual Interior-Point Algorithms,* Princeton University Press. Princeton, NJ.

[12] J. Peng and T. Terlaky (2002). A dynamic large-update primal-dual interior-point method for linear optimization. *Optimization Methods and Software*, 17, pp. 1077-1104.

[13] J. Peng, T. Terlaky, and Y.B. Zhao (2003). A predictor-corrector algorithm for linear optimization based on a specific self-regular proximity function. To appear in *SIAM Journal on Optimization.*

[14] F.A. Potra (1994). A quadratically convergent predictor-corrector method for linear programs from infeasible starting points, *Mathematical Programming*, 67(3), pp. 383-406.

[15] F.A. Potra (1996). An infeasible interior-point predictor-corrector algorithm for linear programming, *SIAM Journal on Optimization*, 6(1), pp. 19-32.

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

[17] M. Salahi and T. Terlaky (2005). An adaptive self-regular proximity based large-update IPM for LO, *Optimization Methods and Software,* 20(1), pp. 168-195.

[18] M. Salahi and T. Terlaky (2004). Adaptive large neighboIeq19od self-regular predictor-corrector IPMs for LO, *Technical Report* 2004/7*, Advanced Optimization Lab. Department of Computing and Software, McMaster University. http://www.cas.mcmaster.ca/∼oplab/publication.* To appear in *Journal of Optimization Theorey and Applications.*

[19] M. Salahi, T. Terlaky, and G. Zhang (2003). The complexity of Self-Regular proximity based infeasible IPMs, *Technical Report* 2003/3, *Advanced Optimization Lab. Department of Computing and Software, McMaster University. http://www.cas.mcmaster.ca/∼oplab/publication.* To appear in *Computational Optimization and Applications,* 2005.

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

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

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

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

[24] Y. Zhang (1994). On the convergence of a class of infeasible-interior-point methods for the horizontal linear complementary problem, *SIAM Journal on Optimization*, 4, pp. 208-227.

[25] X. Zhu, J. Peng, T. Terlaky, and G. Zhang (2003). On implementing self-regular proximity based feasible IPMs, *Technical Report* 2003/2, *Advanced Optimization Lab. Department of Computing and Software, McMaster University. http://www.cas.mcmaster.ca/∼oplab/publication.*

5 Appendix

In this section we provide the proofs of some technical results.

Proof of Lemma 3.3:

Let us define $(\bar{x}, \bar{y}, \bar{s})$ as follows

$$(\bar{x}, \bar{y}, \bar{s}) = \varphi_k[(x^0, y^0, s^0) - (x^*, y^*, s^*)] + (\Delta x^k, \Delta y^k, \Delta s^k),$$

It is not hard to check that $(\bar{x}, \bar{y}, \bar{s})$ satisfies

$$\begin{aligned} A\bar{x} &= 0, \\ A^T\bar{y} + \bar{s} &= 0. \end{aligned}$$

Thus, we have $\bar{x}^T\bar{s} = 0$, i.e.,

$$\left(\Delta x^k + \varphi_k(x^0 - x^*)\right)^T \left(\Delta s^k + \varphi_k(s^0 - s^*)\right) = 0. \tag{43}$$

From the third equation system of (14), we have

$$x^k\left(\Delta s^k+\varphi_k(s^0-s^*)\right)\ +s^k\left(\Delta x^k+\varphi_k(x^0-x^*)\right)$$
$$=-\frac{\mu^k}{\chi_k}v^k\nabla\Psi_\ell(v^k)+\varphi_k s^k(x^0-x^*)+\varphi_k x^k(s^0-s^*).$$

By multiplying this equation by $\left(X^kS^k\right)^{-\frac{1}{2}}$ and considering $(D^k)^{-1}=(X^k)^{-\frac{1}{2}}(S^k)^{\frac{1}{2}}$ and $D^k=(X^k)^{\frac{1}{2}}(S^k)^{-\frac{1}{2}}$, one has

$$(D^k)^{-1}\left(\Delta x^k+\varphi_k(x^0-x^*)\right)\ +D^k\left(\Delta s^k+\varphi_k(s^0-s^*)\right) \tag{44}$$
$$=\frac{\sqrt{\mu^k}}{\chi_k}\nabla\Psi_\ell(v^k)+\varphi_k(D^k)^{-1}(x^0-x^*)+\varphi_k D^k(s^0-s^*).$$

Then, from (43), we have

$$\|(D^k)^{-1}\ \left(\Delta x^k+\varphi_k(x^0-x^*)\right)+D^k\left(\Delta s^k+\varphi_k(s^0-s^*)\right)\|^2$$
$$=\|(D^k)^{-1}\left(\Delta x^k+\varphi_k(x^0-x^*)\right)\|^2+\|D^k\left(\Delta s^k+\varphi_k(s^0-s^*)\right)\|^2.$$

By taking squared norm in both sides of (44), one has

$$\|(D^k)^{-1}\ \left(\Delta x^k+\varphi_k(x^0-x^*)\right)\|^2+\|(D^k)^{-1}\left(\Delta s^k+\varphi_k(s^0-s^*)\right)\|^2$$
$$\le\left(\frac{1}{\chi_k}\|\sqrt{\mu^k}\nabla\Psi_\ell(v^k)\|+\varphi_k\|(D^k)^{-1}(x^0-x^*)\|+\varphi_k\|D^k(s^0-s^*)\|\right)^2.$$

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

$$\left\|(D^k)^{-1}\left(\Delta x^k+\varphi_k(x^0-x^*)\right)\right\|\ \le\frac{1}{\chi_k}\left\|\sqrt{\mu^k}\nabla\Psi_\ell(v^k)\right\|$$
$$+\varphi_k\left(\left\|(D^k)^{-1}(x^0-x^*)\right\|+\left\|D^k(s^0-s^*)\right\|\right). \tag{45}$$

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

$$\left\|(D^k)^{-1}\Delta x^k\right\|\le\frac{1}{\chi_k}\left\|\sqrt{\mu^k}\nabla\Psi_\ell(v^k)\right\|+2\varphi_k\left(\left\|(D^k)^{-1}(x^0-x^*)\right\|+\left\|D^k(s^0-s^*)\right\|\right). \tag{46}$$

Next we show that the magnitude of each term on the right-hand-side of (46) is $O\left(\sqrt{\mu_g^k}\right)$. From Lemmas 2.3 and 3.1, we have

$$\|(X^kS^k)^{-\frac{1}{2}}\|=\max_{i=1,\dots,n}\frac{1}{(x_i^ks_i^k)^{\frac{1}{2}}}\le\frac{e\tau}{((\mu_t^\ell)^k)^{\frac{1}{2}}}\le e\tau\sqrt{\frac{2+\tau\log n}{\mu_g^k\log n}}. \tag{47}$$

On the other hand, from the definition of the matrix norm we have

$$\|(D^k)^{-1}\|=\|(X^kS^k)^{-\frac{1}{2}}S^ke\|_\infty\le\|(X^kS^k)^{-\frac{1}{2}}\|\|s^k\|_1,$$

and similarly $\|D^k\|\le\|(X^kS^k)^{-\frac{1}{2}}\|\|x^k\|_1$. So, we have

$$\varphi_k\left(\|(D^k)^{-1}(x^0-x^*)\|+\|D^k(s^0-s^*)\|\right)\ \le\varphi_k\zeta\left(\|(D^k)^{-1}e\|+\|D^ke\|\right)$$
$$=\varphi_k\zeta\left(\|(X^kS^k)^{-\frac{1}{2}}s^k\|+\|(X^kS^k)^{-\frac{1}{2}}x^k\|\right) \tag{48}$$
$$\le\varphi_k\zeta\|(X^kS^k)^{-\frac{1}{2}}\|\|(x^k,s^k)\|_1.$$

From (47), (48) and Lemma 3.2 we have

$$\varphi_k \left(\|(D^k)^{-1}(x^0 - x^*)\| + \|D^k(s^0 - s^*)\|\right) \leq C_1 ne\tau\sqrt{\frac{(2+\tau\log n)\mu_g^k}{\log n}}. \tag{49}$$

Finally, using inequality (49) and relation (24), from (46) we have

$$\|(D^k)^{-1}\Delta x^k\| \leq \frac{\sqrt{\mu_g^k}}{\chi_k}\sigma_\ell^k + 2C_1 ne\tau\sqrt{\frac{(2+\tau\log n)\mu_g^k}{\log n}} \leq C_2\sqrt{n\mu_g^k}\sigma_\ell^k,$$

where $C_2 = \sqrt{6}C_1 e\tau + 1$. Analogously, one can derive

$$\|(D^k)^{-1}\Delta s^k\| \leq C_2\sqrt{n\mu_g^k}\sigma_\ell^k,$$

that completes the proof of the lemma. □

Proof of Theorem 3.6:

Let $\bar{\alpha} = v_{\min}^k(\sigma_1^k)^{-1}$. Then using Lemma 3.1 and Lemma 3.3.1 of [11], one has that the maximal feasible step size $\alpha_{\max}$ satisfies

$$\alpha_{\max} \geq \bar{\alpha} \geq \frac{1}{e\tau\sigma_1^k}.$$

This proves the first statement. We proceed with the proof of the second statement that is analogous to the proof of Lemma 3.3.3 [11]. For simplicity we drop the subscript and the superscript k. Now, let us define

$$f(\alpha) = \Psi_\ell(v(\alpha)) - \Psi_\ell(v),$$

where $v(\alpha) = \sqrt{\frac{x(\alpha)s(\alpha)}{\mu_t^\ell}} = \sqrt{(v+\alpha d_x)(v+\alpha d_s)}$. Then by the definition (7) one has

$$f(\alpha) = \frac{1}{2}v^T(d_x + d_s)\alpha + \frac{\left\|v^{-\frac{\log n}{2}}(\alpha)\right\|^2}{\log n} - \frac{\left\|v^{-\frac{\log n}{2}}\right\|^2}{\log n}.$$

Since $\left\|v^{-\frac{\log n}{2}}(\alpha)\right\|^2$ satisfies the SR.2 condition of Definition 1.1, then

$$\begin{aligned} f(\alpha) \leq\ & \frac{1}{2}v^T(d_x + d_s)\alpha + \frac{1}{2\log n}\sum_{i=1}^n [(v_i + \alpha(d_x)_i)^{-\log n} + (v_i + \alpha(d_s)_i)^{-\log n}] \\ & - \frac{\left\|v^{-\log n/2}\right\|^2}{\log n} := f_1(\alpha). \end{aligned}$$

The first and second derivatives of $f_1(\alpha)$ are

$$f_1'(\alpha) = \frac{1}{2}v^T(d_x + d_s) - \frac{1}{2}\sum_{i=1}^n [(v_i + \alpha(d_x)_i)^{-\log n - 1}(d_x)_i + (v_i + \alpha(d_s)_i)^{-\log n - 1}(d_s)_i]$$

and

$$\begin{aligned} f_1''(\alpha) &= \frac{1+\log n}{2}\sum_{i=1}^{n}[(v_i+\alpha(d_x)_i)^{-2-\log n}(d_x)_i^2+(v_i+\alpha(d_s)_i^2)^{-2-\log n}(d_s)_i] \\ &\le \frac{\sigma_1^2(1+\log n)}{2}(v_{\min}-\alpha\sigma_1)^{-2-\log n}. \end{aligned} \tag{50}$$

By the definition of f and f_1 also one has

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

Then we may write

$$f_1(\alpha)\le -\frac{\sigma_\ell^2}{2\chi}\alpha+\frac{\sigma_1^2(1+\log n)}{2}\int_0^\alpha\int_0^\xi(v_{\min}-\eta\sigma_1)^{-2-\log n}d\eta d\xi := f_2(\alpha).$$

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 $[0,\bar{\alpha})$. Then it is the unique solution of the equation

$$-\frac{\sigma_\ell^2}{2\chi}+\frac{\sigma_1}{2}(v_{\min}-\alpha\sigma_1)^{-1-\log n}-\frac{\sigma_1}{2}v_{\min}^{-1-\log n}=0, \tag{51}$$

which is equal to

$$\alpha^*=\frac{v_{\min}}{\sigma_1}\left(1-\left(\frac{\chi\sigma_1}{\chi\sigma_1+\sigma_\ell^2 v_{\min}^{1+\log n}}\right)^{\frac{1}{1+\log n}}\right). \tag{52}$$

Now, by using Lemma 1.3.1 of [11] we have

$$\begin{aligned} \left(\frac{\chi\sigma_1}{\chi\sigma_1+\sigma_\ell^2 v_{\min}^{1+\log n}}\right)^{\frac{1}{1+\log n}} &= \left(1-\frac{\sigma_\ell^2 v_{\min}^{1+\log n}}{\chi\sigma_1+\sigma_\ell^2 v_{\min}^{1+\log n}}\right)^{\frac{1}{1+\log n}} \\ &\le 1-\frac{\sigma_\ell^2 v_{\min}^{1+\log n}}{(1+\log n)\left(\chi\sigma_1+\sigma_\ell^2 v_{\min}^{1+\log n}\right)}. \end{aligned}$$

By substituting this in (52) we conclude that

$$\alpha^*\ge\frac{\sigma_\ell^2 v_{\min}v_{\min}^{1+\log n}}{(1+\log n)\sigma_1\left(\chi\sigma_1+v_{\min}^{1+\log n}\sigma_\ell^2\right)}.$$

Using Lemma 3.1, one has

$$\alpha^*\ge\frac{4\sigma_\ell}{e\tau\sigma_1\log n\,(10\chi\sigma_1+8\sigma_\ell)}.$$

Thus, for $\alpha\le\alpha^*$, we have

$$f(\alpha)\le f_1(\alpha)\le f_2(\alpha)\le\frac{f_2'(0)}{2}\alpha=-\frac{\sigma_\ell^2}{4\chi}\alpha,$$

where the last inequality follows from Lemma 1.3.3 of [11]. □

Proof of Theorem ??:

Let

$$g_1(\alpha_k) = \Phi_\ell(x(\alpha_k), s(\alpha_k), \mu_*^\ell) = \frac{1}{2}\frac{\mu_t^\ell}{\mu_*^\ell}\|v(\alpha_k)\|^2 - \frac{n}{2} + \frac{1}{\log n}\left(\frac{\mu_*^\ell}{\mu_t^\ell}\right)^{\frac{\log n}{2}}\left\|v(\alpha_k)^{\frac{-\log n}{2}}\right\|^2 - \frac{n}{\log n}.$$

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

$$\frac{\mu_t^\ell}{\mu_*^\ell}\|v(0)\|^2 = \left(\frac{\mu_*^\ell}{\mu_t^\ell}\right)^{\frac{\log n}{2}}\left\|v(0)^{\frac{-\log n}{2}}\right\|^2 = n(\tau_0)^{\frac{\log n}{2+\log n}},$$

where $\tau_0 := \frac{\mu_g}{\mu_h}$. On the other hand, we have

$$\|v(\alpha_k)\|^2 = \|v\|^2 + \alpha_k v^T(d_x + d_s) + \alpha_k^2 d_x^T d_s \le \|v\|^2 + \alpha_k^2 d_x^T d_s,$$

because $v^T(d_x + d_s) = \left\|v^{\frac{-\log n}{2}}\right\|^2 - \|v\|^2 \le 0$. It is easy to verify that

$$\left\|v(\alpha_k)^{\frac{-\log n}{2}}\right\|^2 \le (1 - \alpha_k v_{\min}^{-1}\sigma_1)^{-\log n}\left\|v^{\frac{-\log n}{2}}\right\|^2.$$

We also have $(1 - \alpha_k v_{\min}^{-1}\sigma_1)^{-\log n} \le 1.0005$ and $\frac{\alpha_k^2\left|d_x^T d_s\right|}{2} \le 10^{-4}$. Then we have

$$g_1(\alpha) \le \frac{n(\tau_0)^{\frac{\log n}{2+\log n}}}{2} - \frac{n}{2} + 10^{-4} + \frac{1.0005(\tau_0)^{\frac{\log n}{2+\log n}}n}{\log n} - \frac{n}{\log n}.$$

Since $\Phi_\ell(x(\alpha_k), s(\alpha_k), \mu_*^\ell(\alpha_k)) \le \Phi_\ell(x(\alpha_k), s(\alpha_k), \mu_*^\ell)$, then the next iterate satisfies the inequality corresponding to the proximity function if

$$\frac{(\tau_0)^{\frac{\log n}{2+\log n}}}{2} + 10^{-4} + \frac{1.0005(\tau_0)^{\frac{\log n}{2+\log n}}}{\log n} \le \frac{\tau^{\frac{\log n}{2+\log n}}}{2} + \frac{\tau^{\frac{\log n}{2+\log n}}}{\log n}.$$

This implies that $\tau_0 = \tau_2\tau$, where $\tau_2 \le 0.99$ that completes the proof. □

Proof of Lemma 3.10: From the assumption that $\mu_g = \tau\mu_\ell^h = \tau\mu_\ell^t$ one has $\mu_\ell^* = \tau^{\frac{2}{2+\log n}}\mu_\ell^h$. Now, let us define

$$\begin{aligned}
g(\alpha) :&= \Phi_\ell(x(\alpha), s(\alpha), \mu_\ell^*) - \Phi_\ell(x, s, \mu_\ell^*) \\
&= \frac{1}{2}\frac{\mu_\ell^h}{\mu_\ell^*}v^T(d_x + d_s)\alpha + \frac{1}{\log n}\left(\frac{\mu_\ell^*}{\mu_\ell^h}\right)^{\frac{\log n}{2}}\left\|v(\alpha)^{\frac{-\log n}{2}}\right\|^2 - \frac{1}{\log n}\left\|v^{\frac{-\log n}{2}}\right\|^2 \\
&\le \frac{1}{2}\frac{\mu_\ell^h}{\mu_\ell^*}v^T(d_x + d_s)\alpha + \frac{1}{2\log n}\left(\frac{\mu_\ell^*}{\mu_\ell^h}\right)^{\frac{\log n}{2}}\left\|(v + \alpha d_x)^{\frac{-\log n}{2}}\right\|^2 \\
&\quad + \frac{1}{2\log n}\left(\frac{\mu_\ell^*}{\mu_\ell^h}\right)^{\frac{\log n}{2}}\left\|(v + \alpha d_s)^{\frac{-\log n}{2}}\right\|^2 - \frac{1}{\log n}\left\|v^{\frac{-\log n}{2}}\right\|^2 := g_1(\alpha),
\end{aligned}$$

where the last inequality follows from the fact that $\left\|v(\alpha)^{\frac{-\log n}{2}}\right\|^2$ satisfies the SR.2 condition of Definition 1.1.

From the definition of v, we have $\left\|v^{\frac{-\log n}{2}}\right\|^2 = n$. Using the inequality that $\sigma_\ell^2 \geq (\tau-1)n$ further implies

$$\left\|v^{-1-\log n}\right\|^2 - \left\|v^{\frac{-\log n}{2}}\right\|^2 \geq 0.$$

This inequality, together with the fact that

$$\|v\|^2 = \tau \left\|v^{\frac{-\log n}{2}}\right\|^2,$$

gives

$$g_1^{'}(0) = \frac{1}{2}\frac{\mu_\ell^h}{\mu_\ell^*}v^T(d_x+d_s) - \frac{1}{2}\left(\frac{\mu_\ell^*}{\mu_\ell^h}\right)^{\frac{\log n}{2}}(v^{-1-\log n})^T(d_x+d_s) \leq -\frac{\sigma_\ell^2}{2\tau^{\frac{2}{2+\log n}}}.$$

Analogous to Theorem 3.6 one has the following upper bound for the second derivative of $g(\alpha)$

$$g^{''}(\alpha) \leq \frac{\tau^{\frac{\log n}{2+\log n}}(1+\log n)\sigma_1^2}{2}(v_{\min} - \alpha\sigma_1)^{-2-\log n}.$$

Using the facts that

$$g(\alpha) = g(0) + g^{'}(0)\alpha + \int_0^\alpha \int_0^\xi g^{''}(\xi)d\eta d\xi$$

and $g(0) = 0$ one has

$$g(\alpha) \leq -\frac{\sigma_\ell^2}{2\tau^{\frac{2}{2+\log n}}}\alpha + \frac{\tau^{\frac{\log n}{2+\log n}}(1+\log n)\sigma_1^2}{2}\int_0^\alpha \int_0^\xi (v_{\min} - \eta\sigma_1)^{-2-\log n} d\eta d\xi := g_2(\alpha).$$

The function $g_2(\alpha)$ is a continuously differentiable function of α. Let us denote its global minimum by α^*, then analogous to the proof of Theorem 3.6 one can derive the following lower bound for the maximum step size

$$\alpha^* \geq \frac{4}{11\chi_k e\tau^2 C_3 n\sigma_\ell^k \log n}.$$

This completes the proof of the lemma. □

Proof of Theorem 3.17:

The proof of the theorem is based on the following technical lemma that can be found in [18].

Lemma 5.1 *Let* $0 < \theta_k < 1$ *and* $v^+ = \frac{v^k}{\sqrt{1-\theta_k}}$. *Then*

$$\Psi_\ell(v^+) \leq \frac{1}{1-\theta_k}\Psi_\ell(v^k) + \frac{n\theta_k}{(1-\theta_k)}. \tag{53}$$

Using (53), it suffices to choose θ_k satisfying the following inequality

$$\Phi_\ell\left(x(\alpha_k^*), s(\alpha_k^*), \left(\mu_t^\ell\right)^k\right) + n\theta_k \leq (1-\theta_k)\Phi_\ell\left(x^k, s^k, \left(\mu_t^\ell\right)^k\right).$$

Considering (40), this inequality is certainly satisfied if

$$\theta_k\Phi_\ell\left(x^k, s^k, \left(\mu_t^\ell\right)^k\right) + n\theta_k \leq \frac{\sqrt{\tau-1}}{11e\tau^2 C_3 n^{\frac{1}{2}}\log n}.$$

By using the fact that $\Phi_\ell\left(x^k, s^k, \left(\mu_t^\ell\right)^k\right) = \dfrac{(\tau-1)n}{2}$, we can rewrite this inequality as:

$$\theta_k\left(\frac{(\tau-1)n}{2} + n\right) \leq \frac{\sqrt{\tau-1}}{11e\tau^2 C_3 n^{\frac{1}{2}}\log n},$$

or

$$\theta_k \leq \theta = \frac{2\sqrt{\tau-1}}{11e(\tau+1)\tau^2 C_3 n^{\frac{3}{2}}\log n}.$$

By choosing $\theta_k = \theta$ one has

$$\Phi_\ell\left(x(\alpha_k^*), s(\alpha_k^*), (1-\theta_k)\left(\mu_t^\ell\right)^k\right) \leq \Phi_\ell\left(x^k, s^k, \left(\mu_t^\ell\right)^k\right),$$

that completes the proof of the theorem. □