

A Full-Newton Step $O(n)$ Infeasible Interior-Point Algorithm for Linear Optimization

C. Roos

February 5, 2005
February 19, 2005

*Faculty of Electrical Engineering, Computer Science and Mathematics
Delft University of Technology
P.O. Box 5031, 2600 GA Delft, The Netherlands
e-mail: C.Roos@ewi.tudelft.nl*

Abstract

We present a full-Newton step infeasible interior-point algorithm. It is shown that at most $O(n)$ (inner) iterations suffice to reduce the duality gap and the residuals by the factor $\frac{1}{e}$. The bound coincides with the best known bound for infeasible interior-point algorithms. It is conjectured that further investigation will improve the above bound to $O(\sqrt{n})$.

Keywords: Linear optimization, infeasible interior-point method, primal-dual method, polynomial complexity.

AMS Subject Classification: 90C05, 90C51

1 Introduction

Interior-Point Methods (IPMs) for solving Linear Optimization (LO) problems were initiated by Karmarkar [6]. They not only have polynomial complexity, but are also highly efficient in practice. One may distinguish between *feasible* IPMs and *infeasible* IPMs (IIPMs). Feasible IPMs start with a strictly feasible interior point and maintain feasibility during the solution process. An elegant and theoretically sound method to find a strictly feasible starting point is to use a self-dual embedding model, by introducing artificial variables. This technique was presented first by Ye et al. [29]. Subsequent references are [1, 16, 27]. Well-known commercial software packages are based on this approach, for example MOSEK [2]¹ and SeDuMi [19]² are based on the use of the self-dual model. Also the leading commercial linear optimization package CPLEX³ includes the self-dual embedding model as a possible option.

Most of the existing software packages use an IIPM. IIPMs start with an arbitrary positive point and feasibility is reached as optimality is approached. The first IIPMs were proposed by Lustig

¹MOSEK: <http://www.mosek.com>

²SeDuMi: <http://sedumi.mcmaster.ca>

³CPLEX: <http://cplex.com>

[9] and Tanabe [20]. Global convergence was shown by Kojima et. al. [7], whereas Zhang [30] proved an $O(n^2L)$ iteration bound for IIPMs under certain conditions. Mizuno [10] introduced a primal-dual IIPM and proved global convergence of the algorithm. Other relevant references are [4, 5, 8, 11, 12, 14, 15, 18, 21, 24, 25] A detailed discussion and analysis of IIPMs can be found in the book by Wright [26] and, with less detail, in the books by Ye [28] and Vanderbei [22]. The performance of existing IIPMs highly depends on the choice of the starting point, which makes these methods less robust than the methods using the self-dual embedding technique.

As usual, we consider the linear optimization (LO) problem in the standard form

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

with its dual problem

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

Here $A \in \mathbf{R}^{m \times n}$, $b, y \in \mathbf{R}^m$ and $c, x, s \in \mathbf{R}^n$. Without loss of generality we assume that $\text{rank}(A) = m$. The vectors x , y and s are the vectors of variables.

The best know iteration bound for IIPMs,

$$O \left(n \log \frac{\max \left\{ (x^0)^T s^0, \|b - Ax^0\|, \|c - A^T y^0 - s^0\| \right\}}{\varepsilon} \right). \quad (1)$$

Here $x^0 > 0$, y^0 and $s^0 > 0$ denote the starting points, and $b - Ax^0$ and $c - A^T y^0 - s^0$ are the initial primal and dual residue vectors, respectively, whereas ε is an upper bound for the duality gap and the norms of residual vectors upon termination of the algorithm. It is assumed in this result that there exists an optimal solution (x^*, y^*, s^*) such that $\|(x^*; s^*)\|_\infty \leq \zeta$, and the initial iterates are $(x^0, y^0, s^0) = \zeta(e, 0, e)$.

Up till 2003, the search directions used in all primal-dual IIPMs were computed from the linear system

$$\begin{aligned} A\Delta x &= b - Ax \\ A^T \Delta y + \Delta s &= c - A^T y - s \\ s\Delta x + x\Delta s &= \mu e - xs, \end{aligned} \quad (2)$$

which yields the so-called primal-dual Newton search directions Δx , Δy and Δs . Recently, Salahi et al. used a so-called 'self-regular' proximity function to define a new search direction for IIPMS [17]. Their modification only involves the third equation in the above system. The iteration bound of their method does not improve the bound in (1).

To introduce the idea underlying the algorithm presented in this paper we make some remarks with a historical flavor. In feasible IPMs, feasibility of the iterates is given, the ultimate goal is to get iterates that are optimal. There is a well known IPM that aims to reach optimality in one step, namely the affine-scaling method. But everybody who is familiar with IPMs knows that this does not yield a polynomial-time method. The last two decades have made it very clear that to get a polynomial-time method one should be less greedy, and work with a search direction that moves the iterates only slowly in the direction of optimality. The reason is that only then one can take full profit of the efficiency of Newton's method, which is the working horse in all IPMs.

In IIPMs, the iterates are not feasible, and apart from reaching optimality one needs to strive for feasibility. This is reflected by the choice of the search direction, as defined by (2). Because, when moving from x to $x^+ := x + \Delta x$ the new iterate x^+ satisfies the primal feasibility constraints, except possibly the nonnegativity constraint. In fact, in general x^+ will have negative components, and, to keep the iterate positive, one is forced to take a damped step of the form $x^+ := x + \alpha \Delta x$, where $\alpha < 1$ denotes the step size. But this should ring a bell. The same phenomenon occurred with the affine-scaling method in feasible IPMs. There it has become clear that the best complexity results hold for methods that are much less greedy and that use full Newton steps (with $\alpha = 1$). Striving to reach feasibility in one step might be too optimistic, and may deteriorate the overall behavior of a method. One may better exercise a little patience and move slower into the direction of feasibility. Therefore, in our approach the search directions are designed in such a way that a full Newton step reduces the sizes of the residual vectors with the same speed as the duality gap. The outcome of the analysis in this paper shows that this is a good strategy.

The paper is organized as follows. As a preparation to the rest of the paper, in Section 2 we first recall some basic tools in the analysis of a feasible IPM. These tools will be used also in the analysis of the IIPM proposed in this paper. Section 3 is used to describe our algorithm in more detail. One characteristic of the algorithm is that it uses intermediate problems. The intermediate problems are suitable perturbations of the given problems (P) and (D) so that at any stage the iterates are feasible for the current perturbed problems; the size of the perturbation decreases at the same speed as the barrier parameter μ . When μ changes to a smaller value, the perturbed problem corresponding to μ changes, and hence also the current central path. The algorithm keeps the iterates close to the μ -center on the central path of the current perturbed problem. To get the iterates feasible for the new perturbed problem and close to its central path, we use a so-called 'feasibility step'. The largest, and hardest, part of the analysis, which is presented in Section 4, concerns this step. It turns out that to keep control over this step, before taking the step the iterates need to be very well centered. Some concluding remarks can be found in Section 5.

Some notations used throughout the paper are as follows. $\|\cdot\|$ denotes the 2-norm of a vector. For any $x = (x_1, x_2, \dots, x_n)^T \in \mathbf{R}^n$, $\min(x)$ denotes the smallest and $\max(x)$ the largest value of the components of x . If $x, s \in \mathbf{R}^n$, then xs denotes the componentwise (or Hadamard) product of the vectors x and s . Furthermore, e denotes the all-one vector of length n . If $z \in \mathbf{R}_+^n$ and $f : \mathbf{R}_+ \rightarrow \mathbf{R}_+$, then $f(z)$ denotes the vector in \mathbf{R}_+^n whose i -th component is $f(z_i)$, with $1 \leq i \leq n$. We write $f(x) = O(g(x))$ if $f(x) \leq \gamma g(x)$ for some positive constant γ .

2 Feasible full-Newton step IPMs

In preparation of dealing with infeasible IPMs, in this section we briefly recall the classical way to obtain a polynomial-time path-following feasible IPM for solving (P) and (D). To solve these problems one needs to find a solution of the following system of equations.

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

In these so-called optimality conditions the first two constraints represent primal and dual feasibility, whereas the last equation is the so-called *complementary condition*. The nonnegativity constraints in the feasibility conditions make the problem already nontrivial: only iterative methods can find solutions of linear systems involving inequality constraints. The complementary condition is nonlinear, which makes it extra hard to solve this system.

2.1 Central path

IPMs replace the complementarity condition by the so-called *centering condition* $xs = \mu e$, where μ may be any positive number. This yields the system

$$\begin{aligned} Ax &= b, & x &\geq 0 \\ A^T y + s &= c, & s &\geq 0 \\ xs &= \mu e. \end{aligned} \tag{3}$$

Surprisingly enough, if this system has a solution, for some $\mu > 0$, then a solution exists for every $\mu > 0$, and this solution is unique. This happens if and only if (P) and (D) satisfy the *interior-point condition* (IPC), i.e., if (P) has a feasible solution $x > 0$ and (D) has a solution (y, s) with $s > 0$ (see, e.g., [16]). If the IPC is satisfied, then the solution of (3) is denoted as $(x(\mu), y(\mu), s(\mu))$, and called the μ -center of (P) and (D) . The set of all μ -centers is called the central path of (P) and (D) . As μ goes to zero, $x(\mu)$, $y(\mu)$ and $s(\mu)$ converge to optimal solution of (P) and (D) . Of course, system (3) is still hard to solve, but by applying Newton's method one can easily find approximate solutions.

2.2 Definition and properties of the Newton step

We proceed by describing Newton's method for solving (3), with μ fixed. Given any primal feasible $x > 0$, and dual feasible y and $s > 0$, we want to find displacements Δx , Δy and Δs such that

$$\begin{aligned} A(x + \Delta x) &= b, \\ A^T(y + \Delta y) + s + \Delta s &= c, \\ (x + \Delta x)(s + \Delta s) &= \mu e. \end{aligned}$$

Neglecting the quadratic term $\Delta x \Delta s$ in the third equation, we obtain the following linear system of equations in the search directions Δx , Δy and Δs .

$$A\Delta x = b - Ax, \tag{4}$$

$$A^T \Delta y + \Delta s = c - A^T y - s, \tag{5}$$

$$s\Delta x + x\Delta s = \mu e - xs. \tag{6}$$

Since A has full rank, and the vectors x and s are positive, one may easily verify that the coefficient matrix in the linear system (4)–(6) is nonsingular. Hence this system uniquely defines the search directions Δx , Δy and Δs . These search directions are used in all existing primal-dual (feasible and infeasible) IPMs and called after Newton.

If x is primal feasible and (y, s) dual feasible, then $b - Ax = 0$ and $c - A^T y - s = 0$, whence the above system reduces to

$$A\Delta x = 0, \tag{7}$$

$$A^T \Delta y + \Delta s = 0, \tag{8}$$

$$s\Delta x + x\Delta s = \mu e - xs, \tag{9}$$

which gives the usual search directions for feasible primal-dual IPMs.

The new iterates are given by

$$x^+ = x + \Delta x,$$

$$y^+ = y + \Delta y,$$

$$s^+ = s + \Delta s.$$

An important observation is that Δx lies in the null space of A , whereas Δs belongs to the row space of A . This implies that Δx and Δs are orthogonal, i.e.,

$$(\Delta x)^T \Delta s = 0.$$

As a consequence we have the important property that after a full Newton step the duality gap assumes the same value as at the μ -centers, namely $n\mu$.

Lemma 2.1 (Lemma II.46 in [16]) *After a primal-dual Newton step one has $(x^+)^T s^+ = n\mu$.*

Assuming that a primal feasible $x^0 > 0$ and a dual feasible pair (y^0, s^0) with $s^0 > 0$ are given that are 'close to' $x(\mu)$ and $(y(\mu), s(\mu))$ for some $\mu = \mu^0$, one can find an ε -solution in $O(\sqrt{n} \log(n/\varepsilon))$ iterations of the algorithm in Figure 1. In this algorithm $\delta(x, s; \mu)$ is a quantity that measures proximity of the feasible triple (x, y, s) to the μ -center $(x(\mu), y(\mu), s(\mu))$. Following [16], this quantity is defined as follows

$$\delta(x, s; \mu) := \delta(v) := \frac{1}{2} \|v - v^{-1}\| \quad \text{where} \quad v := \sqrt{\frac{xs}{\mu}}. \tag{10}$$

The following two lemmas are crucial in the analysis of the algorithm. We recall them without proof. The first lemma describes the effect on $\delta(x, s; \mu)$ of a μ -update, and the second lemma the effect of a Newton step.

Lemma 2.2 (Lemma II.53 in [16]) *Let (x, s) be a positive primal-dual pair and $\mu > 0$ such that $x^T s = n\mu$. Moreover, let $\delta := \delta(x, s; \mu)$ and let $\mu^+ = (1 - \theta)\mu$. Then*

$$\delta(x, s; \mu^+)^2 = (1 - \theta)\delta^2 + \frac{\theta^2 n}{4(1 - \theta)}.$$

Lemma 2.3 (Theorem II.49 in [16]) *If $\delta := \delta(x, s; \mu) \leq 1$, then the primal-dual Newton step is feasible, i.e., x^+ and s^+ are nonnegative. Moreover, if $\delta < 1$, then x^+ and s^+ are positive and*

$$\delta(x^+, s^+; \mu) \leq \frac{\delta^2}{\sqrt{2(1 - \delta^2)}}.$$

Corollary 2.4 *If $\delta := \delta(x, s; \mu) \leq \frac{1}{\sqrt{2}}$, then $\delta(x^+, s^+; \mu) \leq \delta^2$.*

Primal-Dual Feasible IPM

Input:

Accuracy parameter $\varepsilon > 0$;

barrier update parameter θ , $0 < \theta < 1$;

feasible (x^0, y^0, s^0) with $(x^0)^T s^0 = n\mu^0$, $\delta(x^0, s^0, \mu^0) \leq 1/2$.

begin

$x := x^0$; $y := y^0$; $s := s^0$; $\mu := \mu^0$;

while $x^T s \geq \varepsilon$ **do**

begin

μ -update:

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

centering step:

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

end

end

Figure 1: Feasible full-Newton-step algorithm

2.3 Complexity analysis

We have the following theorem, whose simple proof we include, because it slightly improves the complexity result in [16, Theorem II.52].

Theorem 2.5 *If $\theta = 1/\sqrt{2n}$, then the algorithm requires at most*

$$\sqrt{2n} \log \frac{n\mu^0}{\varepsilon}$$

iterations. The output is a primal-dual pair (x, s) such that $x^T s \leq \varepsilon$.

Proof: At the start of the algorithm we have $\delta(x, s; \mu) \leq 1/2$. After the update of the barrier parameter to $\mu^+ = (1 - \theta)\mu$, with $\theta = 1/\sqrt{2n}$, we have, by Lemma 2.2, the following upper bound for $\delta(x, s; \mu^+)$:

$$\delta(x, s; \mu^+)^2 \leq \frac{1 - \theta}{4} + \frac{1}{8(1 - \theta)} \leq \frac{3}{8}.$$

Assuming $n \geq 2$, the last inequality follows since the expression at the left is a convex function of θ , whose value is $3/8$ both in $\theta = 0$ and $\theta = 1/2$. Since $\theta \in [0, 1/2]$, the left hand side does not exceed $3/8$. Since $3/8 < 1/2$, we obtain $\delta(x, s; \mu^+) \leq 1/\sqrt{2}$. After the primal-dual Newton step to the μ^+ -center we have, by Corollary 2.4, $\delta(x^+, s^+; \mu^+) \leq 1/2$. Also, from Lemma 2.1, $(x^+)^T s^+ = n\mu^+$. Thus, after each iteration of the algorithm the properties

$$x^T s = n\mu, \quad \delta(x, s; \mu) \leq \frac{1}{2}$$

are maintained, and hence the algorithm is well defined. The iteration bound in the theorem now easily follows from the fact that in each iteration the value of $x^T s$ is reduced by the factor $1 - \theta$. This proves the theorem. \square

3 Infeasible full-Newton step IPM

We are now ready to present an infeasible start algorithm that generates an ε -solution of (P) and (D) , if it exists, or establishes that no such solution exists.

3.1 Perturbed problems

We start with choosing arbitrarily $x^0 > 0$ and $y^0, s^0 > 0$ such that $x^0 s^0 = \mu^0 e$ for some (positive) number μ^0 . For any ν with $0 < \nu \leq 1$ we consider the perturbed problem (P_ν) , defined by

$$(P_\nu) \quad \min \left\{ (c - \nu(c - A^T y^0 - s^0))^T x : Ax = b - \nu(b - Ax^0), \quad x \geq 0 \right\},$$

and its dual problem (D_ν) , which is given by

$$(D_\nu) \quad \max \left\{ (b - \nu(b - Ax^0))^T y : A^T y + s = c - \nu(c - A^T y^0 - s^0), \quad s \geq 0 \right\}.$$

Note that if $\nu = 1$ then $x = x^0$ yields a strictly feasible solution of (P_ν) , and $(y, s) = (y^0, s^0)$ a strictly feasible solution of (D_ν) . We conclude that if $\nu = 1$ then (P_ν) and (D_ν) satisfy the IPC.

Lemma 3.1 (cf. Theorem 5.13 in [28]) *The original problems, (P) and (D) , are feasible if and only if for each ν satisfying $0 < \nu \leq 1$ the perturbed problems (P_ν) and (D_ν) satisfy the IPC.*

Proof: Suppose that (P) and (D) are feasible. Let \bar{x} be feasible solution of (P) and (\bar{y}, \bar{s}) a feasible solution of (D) . Then $A\bar{x} = b$ and $A^T \bar{y} + \bar{s} = c$, with $\bar{x} \geq 0$ and $\bar{s} \geq 0$. Now let $0 < \nu \leq 1$, and consider

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

One has

$$Ax = A((1 - \nu)\bar{x} + \nu x^0) = (1 - \nu)A\bar{x} + \nu Ax^0 = (1 - \nu)b + \nu Ax^0 = b - \nu(b - Ax^0),$$

showing that x is feasible for (P_ν) . Similarly,

$$\begin{aligned} A^T y + s &= (1 - \nu)(A^T \bar{y} + \bar{s}) + \nu(A^T y^0 + s^0) \\ &= (1 - \nu)c + \nu(A^T y^0 + s^0) = c - \nu(c - A^T y^0 - s^0), \end{aligned}$$

showing that (y, s) is feasible for (D_ν) . Since $\nu > 0$, x and s are positive, thus proving that (P_ν) and (D_ν) satisfy the IPC.

To prove the inverse implication, suppose that (P_ν) and (D_ν) satisfy the IPC for each ν satisfying $0 < \nu \leq 1$. Obviously, then (P_ν) and (D_ν) are feasible for these values of ν . Letting ν go to zero it follows that (P) and (D) are feasible. \square

3.2 Central path of the perturbed problems

We assume that (P) and (D) are feasible. Letting $0 < \nu \leq 1$, Lemma 3.1 implies that the problems (P_ν) and (D_ν) satisfy the IPC, and hence their central paths exist. This means that the system

$$b - Ax = \nu(b - Ax^0), \quad x \geq 0 \quad (11)$$

$$c - A^T y - s = \nu(c - A^T y^0 - s^0), \quad s \geq 0. \quad (12)$$

$$xs = \mu e. \quad (13)$$

has a unique solution, for every $\mu > 0$. In the sequel this unique solution is denoted as $(x(\mu, \nu), y(\mu, \nu), s(\mu, \nu))$. These are the μ -centers of the perturbed problems (P_ν) and (D_ν) .

Note that since $x^0 s^0 = \mu^0 e$, x^0 is the μ^0 -center of the perturbed problem (P_1) and (y^0, s^0) the μ^0 -center of (D_1) . In other words, $(x(\mu^0, 1), y(\mu^0, 1), s(\mu^0, 1)) = (x^0, y^0, s^0)$. In the sequel we will always have $\mu = \nu \mu^0$.

3.3 Idea underlying the algorithm

We just established that if $\nu = 1$ and $\mu = \mu^0$, then $x = x^0$ is the μ -center of the perturbed problem (P_ν) and $(y, s) = (y^0, s^0)$ the μ -center of (D_ν) . These are our initial iterates.

We measure proximity to the μ -center of the perturbed problems by the quantity $\delta(x, s; \mu)$ as defined in (10). Initially we thus have $\delta(x, s; \mu) = 0$. In the sequel we assume that at the start of each iteration, just before the μ -update, $\delta(x, s; \mu)$ is smaller than or equal to a (small) threshold value $\tau > 0$. So this is certainly true at the start of the first iteration.

Now we describe one iteration of our algorithm. Suppose that for some $\mu \in (0, \mu^0]$ we have x, y and s satisfying the feasibility conditions (11) and (12) for $\nu = \mu/\mu^0$, and such that $x^T s = n\mu$ and $\delta(x, s; \mu) \leq \tau$. Roughly spoken, one iteration of the algorithm can be described as follows: reduce μ to $\mu^+ = (1 - \theta)\mu$, with $\theta \in (0, 1)$, and find new iterates x^+, y^+ and s^+ that satisfy (11) and (12), with μ replaced by μ^+ and ν by $\nu^+ = \mu^+/\mu^0$, and such that $x^T s = n\mu^+$ and $\delta(x^+, s^+; \mu^+) \leq \tau$. Note that $\nu^+ = (1 - \theta)\nu$.

To be more precise, this will be achieved as follows. Our first aim is to get iterates (x, y, s) that are feasible for $\nu^+ = \mu^+/\mu^0$, and close to the μ -centers $(x(\mu, \nu^+), y(\mu, \nu^+), s(\mu, \nu^+))$. If we can do this in a such a way that the new iterates satisfy $\delta(x, s; \mu^+) \leq 1/\sqrt{2}$, then we can easily get iterates (x, y, s) that are feasible for (P_{ν^+}) and (D_{ν^+}) and such that $\delta(x, s; \mu^+) \leq \tau$ by performing Newton steps targeting at the μ^+ centers of (P_{ν^+}) and (D_{ν^+}) .

Before proceeding it will be convenient to introduce some new notations. We denote the initial values of the primal and dual residuals as r_b^0 and r_c^0 , respectively, as

$$r_b^0 = b - Ax^0,$$

$$r_c^0 = c - A^T y^0 - s^0.$$

Then the feasibility equations for (P_ν) and (D_ν) are given by

$$Ax = b - \nu r_b^0, \quad x \geq 0 \quad (14)$$

$$A^T y + s = c - \nu r_c^0, \quad s \geq 0. \quad (15)$$

and those of (P_{ν^+}) and (D_{ν^+}) by

$$Ax = b - \nu^+ r_b^0, \quad x \geq 0 \quad (16)$$

$$A^T y + s = c - \nu^+ r_c^0, \quad s \geq 0. \quad (17)$$

To get iterates that are feasible for (P_{ν^+}) and (D_{ν^+}) we need search directions $\Delta^f x$, $\Delta^f y$ and $\Delta^f s$ such that

$$\begin{aligned} A(x + \Delta^f x) &= b - \nu^+ r_b^0 \\ A^T(y + \Delta^f y) + (s + \Delta^f s) &= c - \nu^+ r_c^0. \end{aligned}$$

Since x is feasible for (P_{ν}) and (y, s) is feasible for (D_{ν}) , it follows that $\Delta^f x$, $\Delta^f y$ and $\Delta^f s$ should satisfy

$$\begin{aligned} A\Delta^f x &= (b - Ax) - \nu^+ r_b^0 = \nu r_b^0 - \nu^+ r_b^0 = \theta \nu r_b^0 \\ A^T \Delta^f y + \Delta^f s &= (c - A^T y - s) - \nu^+ r_c^0 = \nu r_c^0 - \nu^+ r_c^0 = \theta \nu r_c^0. \end{aligned}$$

Therefore, the following system is used to define $\Delta^f x$, $\Delta^f y$ and $\Delta^f s$:

$$A\Delta^f x = \theta \nu r_b^0 \quad (18)$$

$$A^T \Delta^f y + \Delta^f s = \theta \nu r_c^0 \quad (19)$$

$$s\Delta^f x + x\Delta^f s = \mu e - xs, \quad (20)$$

and the new iterates are given by

$$x^+ = x + \Delta^f x, \quad (21)$$

$$y^+ = y + \Delta^f y, \quad (22)$$

$$s^+ = s + \Delta^f s. \quad (23)$$

We conclude that after the feasibility step the iterates satisfy (11) and (12), with $\nu = \nu^+$. The hard part in the analysis will be to guarantee that after the feasibility step the iterates satisfy $\delta(x^+, s^+; \mu^+) \leq 1/\sqrt{2}$.

After the feasibility step we perform centering steps in order to get iterates that moreover satisfy $x^T s = n\mu^+$ and $\delta(x, s; \mu^+) \leq \tau$. By using Corollary 2.4, the required number of centering steps can be easily be obtained. Because, assuming $\delta = \delta(x^+, s^+; \mu^+) \leq 1/\sqrt{2}$, after k centering steps we will have iterates (x, y, s) that are still feasible for (P_{ν^+}) and (D_{ν^+}) and such that

$$\delta(x, s; \mu^+) \leq \left(\frac{1}{\sqrt{2}}\right)^{2^k}.$$

So, k should satisfy

$$\left(\frac{1}{\sqrt{2}}\right)^{2^k} \leq \tau,$$

which gives

$$k \geq \log_2 \left(\log_2 \frac{1}{\tau^2} \right). \quad (24)$$

Primal-Dual Infeasible IPM

Input:
 Accuracy parameter $\varepsilon > 0$;
 barrier update parameter θ , $0 < \theta < 1$
 threshold parameter $\tau > 0$.

begin
 $x := x^0 > 0$; $y := y^0$; $s := s^0 > 0$; $x^0 s^0 = \mu^0 e$; $\mu = \mu^0$;
while $\max(x^T s, \|b - Ax\|, \|c - A^T y - s\|) \geq \varepsilon$ **do**
begin
 feasibility step:
 $(x, y, s) := (x, y, s) + (\Delta^f x, \Delta^f y, \Delta^f s)$;
 μ -update:
 $\mu := (1 - \theta)\mu$;
 centering steps:
while $\delta(x, s; \mu) \geq \tau$ **do**
 $(x, y, s) := (x, y, s) + (\Delta x, \Delta y, \Delta s)$;
endwhile
end
end

Figure 2: Algorithm

3.4 Algorithm

A more formal description of the algorithm is given in Figure 2. Note that after each iteration the residuals and the duality gap are reduced by a factor $1 - \theta$. The algorithm stops if the norms of the residuals and the duality gap are less than the accuracy parameter ε .

4 Analysis of the algorithm

Let x , y and s denote the iterates at the start of an iteration, and assume $\delta(x, s; \mu) \leq \tau$. Recall that at the start of the first iteration this is certainly true, because then $\delta(x, s; \mu) = 0$.

4.1 Effect of the feasibility step; choice of θ

As we established in Section 3.3, the feasibility step generates new iterates x^+ , y^+ and s^+ that satisfy the feasibility conditions for (P_{ν^+}) and (D_{ν^+}) . A crucial element in the analysis is to show that after the feasibility step $\delta(x^+, s^+; \mu^+) \leq 1/\sqrt{2}$, i.e., that the new iterates are within the

region where the Newton process targeting at the μ^+ -centers of (P_{ν^+}) and (D_{ν^+}) is quadratically convergent.

Defining

$$v = \sqrt{\frac{xs}{\mu}}, \quad d_x := \frac{v\Delta^f x}{x}, \quad d_s := \frac{v\Delta^f s}{s}, \quad (25)$$

we have

$$x^+ = x + \Delta^f x = x + \frac{xd_x}{v} = \frac{x}{v}(v + d_x) \quad (26)$$

$$s^+ = s + \Delta^f s = s + \frac{sd_s}{v} = \frac{s}{v}(v + d_s). \quad (27)$$

To simplify the presentation we will denote $\delta(x, s; \mu)$ below simply as δ . We assume that $\delta \leq \tau$.

Lemma 4.1 *The new iterates are certainly strictly feasible if*

$$\|d_x\| < \frac{1}{\rho(\delta)} \quad \text{and} \quad \|d_s\| < \frac{1}{\rho(\delta)}, \quad (28)$$

where

$$\rho(\delta) := \delta + \sqrt{1 + \delta^2}. \quad (29)$$

Proof: It is clear from (26) that x^+ is strictly feasible if and only if $v + d_x > 0$. This certainly holds if $\|d_x\| < \min(v)$. Since $2\delta = \|v - v^{-1}\|$, the minimal value t that an entry of v can attain will satisfy $t \leq 1$ and $1/t - t = 2\delta$. The last equation implies $t^2 + 2\delta t - 1 = 0$, which gives $t = -\delta + \sqrt{1 + \delta^2} = 1/\rho(\delta)$. This proves the first inequality in (28). The second inequality is obtained in the same way. \square

In the sequel we denote

$$\omega(v) := \frac{1}{2}\sqrt{\|d_x\|^2 + \|d_s\|^2}. \quad (30)$$

This implies $\|d_x\| \leq 2\omega(v)$ and $\|d_s\| \leq 2\omega(v)$, and moreover,

$$d_x^T d_s \leq \|d_x\| \|d_s\| \leq \frac{1}{2} \left(\|d_x\|^2 + \|d_s\|^2 \right) \leq 2\omega(v)^2 \quad (31)$$

$$|d_{x_i} d_{s_i}| \leq \|d_x\| \|d_s\| \leq 2\omega(v)^2, \quad 1 \leq i \leq n. \quad (32)$$

Lemma 4.2 *One has*

$$4\delta(v^+)^2 \leq \frac{\theta^2 n}{1 - \theta} + \frac{2\omega(v)^2}{1 - \theta} + (1 - \theta) \frac{2\omega(v)^2}{1 - 2\omega(v)^2}.$$

Proof: By definition (10),

$$\delta(x^+, s^+; \mu^+) = \delta(v^+) = \frac{1}{2} \left\| v^+ - \frac{e}{v^+} \right\|, \quad \text{where } v^+ = \sqrt{\frac{x^+ s^+}{\mu^+}}.$$

Using (20), we obtain

$$x^+ s^+ = xs + \left(s\Delta^f x + x\Delta^f s \right) + \Delta^f x \Delta^f s = \mu e + \Delta^f x \Delta^f s. \quad (33)$$

Since $\mu v^2 = xs$, after division of both sides in (33) by μ^+ we get

$$(v^+)^2 = \frac{\mu}{\mu^+} e + \frac{1}{\mu^+} \frac{x d_x}{v} \frac{s d_s}{v} = \frac{e + d_x d_s}{1 - \theta}.$$

Defining for the moment $u = \sqrt{e + d_x d_s}$, we have $v^+ = u/\sqrt{1 - \theta}$. Hence we may write

$$2\delta(v^+) = \left\| \frac{u}{\sqrt{1 - \theta}} - \sqrt{1 - \theta} u^{-1} \right\| = \left\| \frac{\theta u}{\sqrt{1 - \theta}} + \sqrt{1 - \theta} (u - u^{-1}) \right\|.$$

Therefore we have

$$\begin{aligned} 4\delta(v^+)^2 &= \frac{\theta^2}{1 - \theta} \|u\|^2 + (1 - \theta) \|u - u^{-1}\|^2 + 2\theta u^T (u - u^{-1}) \\ &= \left(\frac{\theta^2}{1 - \theta} + 2\theta \right) \|u\|^2 + (1 - \theta) \|u - u^{-1}\|^2 - 2\theta u^T u^{-1} \\ &= \left(\frac{\theta^2}{1 - \theta} + 2\theta \right) e^T (e + d_x d_s) + (1 - \theta) \|u - u^{-1}\|^2 - 2\theta n \\ &= \left(\frac{\theta^2}{1 - \theta} + 2\theta \right) (n + d_x^T d_s) + (1 - \theta) \|u - u^{-1}\|^2 - 2\theta n \\ &= \frac{\theta^2 n}{1 - \theta} + \left(\frac{\theta^2}{1 - \theta} + 2\theta \right) d_x^T d_s + (1 - \theta) \|u^{-1} - u\|^2. \end{aligned}$$

The last term can be reduced as follows.

$$\begin{aligned} \|u^{-1} - u\|^2 &= e^T \left(e + d_x d_s + \frac{e}{e + d_x d_s} - 2e \right) = d_x^T d_s + \sum_{i=1}^n \frac{1}{1 + d_{x_i} d_{s_i}} - n \\ &= d_x^T d_s + \sum_{i=1}^n \left(\frac{1}{1 + d_{x_i} d_{s_i}} - 1 \right) = d_x^T d_s - \sum_{i=1}^n \frac{d_{x_i} d_{s_i}}{1 + d_{x_i} d_{s_i}}. \end{aligned}$$

Substitution gives

$$\begin{aligned} 4\delta(v^+)^2 &= \frac{\theta^2 n}{1 - \theta} + \left(\frac{\theta^2}{1 - \theta} + 2\theta \right) d_x^T d_s + (1 - \theta) \left(d_x^T d_s - \sum_{i=1}^n \frac{d_{x_i} d_{s_i}}{1 + d_{x_i} d_{s_i}} \right) \\ &= \frac{\theta^2 n}{1 - \theta} + \frac{d_x^T d_s}{1 - \theta} - (1 - \theta) \sum_{i=1}^n \frac{d_{x_i} d_{s_i}}{1 + d_{x_i} d_{s_i}}. \end{aligned}$$

Hence, using (31) and (32), we arrive at

$$\begin{aligned} 4\delta(v^+)^2 &\leq \frac{\theta^2 n}{1 - \theta} + \frac{2\omega(v)^2}{1 - \theta} + (1 - \theta) \sum_{i=1}^n \frac{|d_{x_i} d_{s_i}|}{1 - 2\omega(v)^2} \\ &\leq \frac{\theta^2 n}{1 - \theta} + \frac{2\omega(v)^2}{1 - \theta} + \frac{1 - \theta}{2} \sum_{i=1}^n \frac{d_{x_i}^2 + d_{s_i}^2}{1 - 2\omega(v)^2} \\ &= \frac{\theta^2 n}{1 - \theta} + \frac{2\omega(v)^2}{1 - \theta} + (1 - \theta) \frac{2\omega(v)^2}{1 - 2\omega(v)^2}, \end{aligned}$$

which completes the proof. \square

We conclude this section by presenting a value that we not allow $\omega(v)$ to exceed. Because we need to have $\delta(v^+) \leq 1/\sqrt{2}$, it follows from Lemma 4.2 that it suffices if

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

At this stage we decide to choose

$$\theta = \frac{\alpha}{\sqrt{2n}}, \quad \alpha \leq \frac{1}{\sqrt{2}}. \quad (34)$$

Then, assuming $n \geq 2$, one may easily verify that

$$\omega(v) \leq \frac{1}{2} \quad \Rightarrow \quad \delta(v^+) \leq \frac{1}{\sqrt{2}}. \quad (35)$$

We proceed by considering the vectors d_x and d_s more in detail.

4.2 The scaled search directions d_x and d_s

One may easily check that the system (18)–(20), which defines the search directions $\Delta^f x$, $\Delta^f y$ and $\Delta^f s$, can be expressed in terms of the scaled search directions d_x and d_s as follows.

$$\bar{A}d_x = \theta\nu r_b^0, \quad (36)$$

$$\bar{A}^T \frac{\Delta^f y}{\mu} + d_s = \theta\nu v s^{-1} r_c^0, \quad (37)$$

$$d_x + d_s = v^{-1} - v, \quad (38)$$

where

$$\bar{A} = AV^{-1}X. \quad (39)$$

If r_b^0 and r_c^0 are zero, i.e., if the initial iterates are feasible, then d_x and d_s are orthogonal vectors, since then the vector d_x belongs to the null space and d_s to the row space of the matrix \bar{A} . It follows that d_x and d_s form an orthogonal decomposition of the vector $v^{-1} - v$. As a consequence we then have obvious upper bounds for the norms of d_x and d_s , namely $\|d_x\| \leq 2\delta(v)$ and $\|d_s\| \leq 2\delta(v)$, and moreover, $\omega(v) = \delta(v)$, with $\omega(v)$ as defined in (30).

In the infeasible case orthogonality of d_x and d_s may not be assumed, however, and the situation may be quite different. This is illustrated in Figure 3. As a consequence, it becomes much harder to get upper bounds for $\|d_x\|$ and $\|d_s\|$, thus complicating the analysis of the algorithm in comparison with feasible IPMs. To obtain an upper bound for $\omega(v)$ is the subject of several sections to follow.

4.3 Upper bound for $\omega(v)$

Let us denote the null space of the matrix \bar{A} as \mathcal{L} . So,

$$\mathcal{L} := \{\xi \in \mathbf{R}^n : \bar{A}\xi = 0\}.$$

Obviously, the affine space $\{\xi \in \mathbf{R}^n : \bar{A}\xi = \theta\nu r_b^0\}$ equals $d_x + \mathcal{L}$. The row space of \bar{A} equals the orthogonal complement \mathcal{L}^\perp of \mathcal{L} , and $d_s \in \theta\nu v s^{-1} r_c^0 + \mathcal{L}^\perp$.

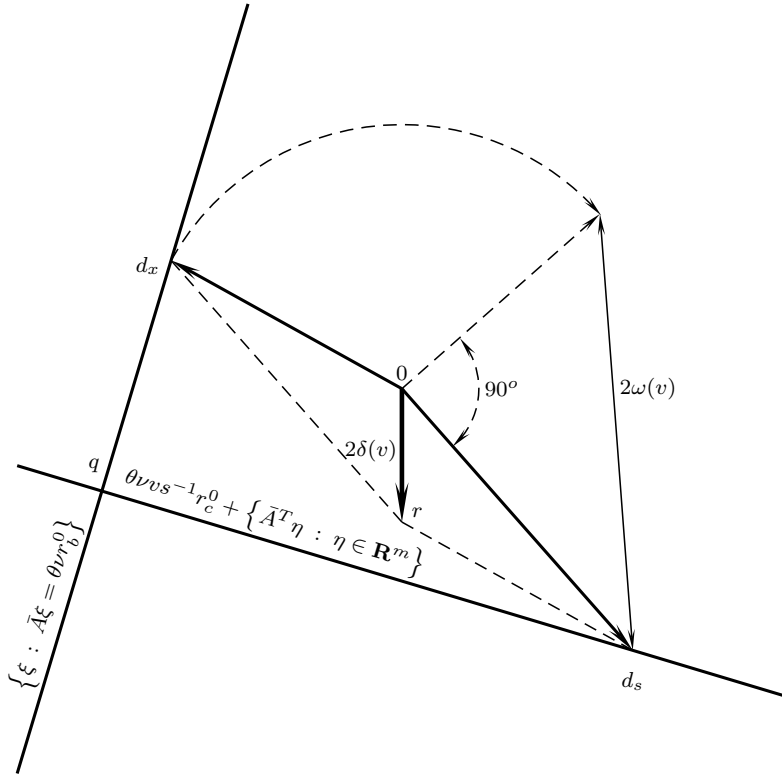


Figure 3: Geometric interpretation of $\omega(v)$.

Lemma 4.3 *Let q be the (unique) point in the intersection of the affine spaces $d_x + \mathcal{L}$ and $d_s + \mathcal{L}^\perp$. Then*

$$2\omega(v) \leq \sqrt{\|q\|^2 + (\|q\| + 2\delta(v))^2}.$$

Proof: To simplify the notation in this proof we denote $r = v^{-1} - v$. Since $\mathcal{L} + \mathcal{L}^\perp = \mathbf{R}^n$, there exist $q_1, r_1 \in \mathcal{L}$ and $q_2, r_2 \in \mathcal{L}^\perp$ such that

$$q = q_1 + q_2, \quad r = r_1 + r_2.$$

On the other hand, since $d_x - q \in \mathcal{L}$ and $d_s - q \in \mathcal{L}^\perp$ there must exist $\ell_1 \in \mathcal{L}$ and $\ell_2 \in \mathcal{L}^\perp$ such that

$$d_x = q + \ell_1, \quad d_s = q + \ell_2.$$

Due to (38) it follows that $r = 2q + \ell_1 + \ell_2$, which implies

$$(2q_1 + \ell_1) + (2q_2 + \ell_2) = r_1 + r_2,$$

from which we conclude that

$$\ell_1 = r_1 - 2q_1, \quad \ell_2 = r_2 - 2q_2.$$

Hence we obtain

$$d_x = q + r_1 - 2q_1 = (r_1 - q_1) + q_2$$

$$d_s = q + r_2 - 2q_2 = q_1 + (r_2 - q_2).$$

Since the spaces \mathcal{L} and \mathcal{L}^\perp are orthogonal we conclude from this that

$$4\omega(v)^2 = \|d_x\|^2 + \|d_s\|^2 = \|r_1 - q_1\|^2 + \|q_2\|^2 + \|q_1\|^2 + \|r_2 - q_2\|^2 = \|q - r\|^2 + \|q\|^2.$$

Assuming $q \neq 0$, since $\|r\| = 2\delta(v)$ the right hand side is maximal if $r = -2\delta(v)q/\|q\|$, and thus we obtain

$$4\omega(v)^2 \leq \left\| \left(1 + \frac{2\delta(v)}{\|q\|} \right) q \right\|^2 + \|q\|^2 = \|q\|^2 + (\|q\| + 2\delta(v))^2,$$

which implies the inequality in the lemma if $q \neq 0$. Since the inequality in the lemma holds with equality if $q = 0$, this completes the proof. \square

In the sequel we denote $\delta(v)$ simply as δ . Recall from (35) that in order to guarantee that $\delta(v^+) \leq \frac{1}{\sqrt{2}}$ we want to have $\omega(v) \leq \frac{1}{2}$. Due to Lemma 4.3 this will certainly hold if $\|q\|$ satisfies

$$\|q\|^2 + (\|q\| + 2\delta)^2 \leq 1. \quad (40)$$

4.4 Upper bound for $\|q\|$

Recall from Lemma 4.3 that q is the (unique) solution of the system

$$\begin{aligned} \bar{A}q &= \theta\nu r_b^0, \\ \bar{A}^T\xi + q &= \theta\nu v s^{-1}r_c^0. \end{aligned}$$

We proceed by deriving an upper bound for $\|q\|$. From the definition (39) of \bar{A} we deduce that $\bar{A} = \sqrt{\mu}AD$, where

$$D = \text{diag} \left(\frac{xv^{-1}}{\sqrt{\mu}} \right) = \text{diag} \left(\sqrt{\frac{x}{s}} \right) = \text{diag} (\sqrt{\mu}vs^{-1}).$$

For the moment, let us write

$$r_b = \theta\nu r_b^0, \quad r_c = \theta\nu r_c^0$$

Then the system defining q is equivalent to

$$\begin{aligned} \sqrt{\mu}ADq &= r_b, \\ \sqrt{\mu}DA^T\xi + q &= \frac{1}{\sqrt{\mu}}Dr_c. \end{aligned}$$

This implies

$$\mu AD^2A^T\xi = AD^2r_c - r_b,$$

whence

$$\xi = \frac{1}{\mu} (AD^2A^T)^{-1} (AD^2r_c - r_b).$$

Substitution gives

$$q = \frac{1}{\sqrt{\mu}} \left(Dr_c - DA^T (AD^2A^T)^{-1} (AD^2r_c - r_b) \right).$$

Observe that

$$q_1 := Dr_c - DA^T (AD^2A^T)^{-1} AD^2r_c = \left(I - DA^T (AD^2A^T)^{-1} AD \right) Dr_c$$

is the orthogonal projection of Dr_c onto the null space of AD . Let (\bar{y}, \bar{s}) be such that $A^T\bar{y} + \bar{s} = c$. Then we may write

$$r_c = \theta\nu r_c^0 = \theta\nu (c - A^T y^0 - s^0) = \theta\nu (A^T(\bar{y} - y^0) + \bar{s} - s^0).$$

Since $DA^T(\bar{y} - y^0)$ belongs to the row space of AD , which is orthogonal to the null space of AD , we obtain

$$\|q_1\| \leq \theta\nu \|D(\bar{s} - s^0)\|.$$

On the other hand, let \bar{x} be such that $A\bar{x} = b$. Then

$$r_b = \theta\nu r_b^0 = \theta\nu (b - Ax^0) = \theta\nu A(\bar{x} - x^0),$$

and the vector

$$q_2 := DA^T (AD^2 A^T)^{-1} r_b = \theta\nu DA^T (AD^2 A^T)^{-1} AD (D^{-1}(\bar{x} - x^0))$$

is the orthogonal projection of $\theta\nu D^{-1}(\bar{x} - x^0)$ onto the row space of AD . Hence it follows that

$$\|q_2\| \leq \theta\nu \|D^{-1}(\bar{x} - x^0)\|.$$

Since $\sqrt{\mu}q = q_1 + q_2$ and q_1 and q_2 are orthogonal, we may conclude that

$$\sqrt{\mu} \|q\| = \sqrt{\|q_1\|^2 + \|q_2\|^2} \leq \theta\nu \sqrt{\|D(\bar{s} - s^0)\|^2 + \|D^{-1}(\bar{x} - x^0)\|^2}, \quad (41)$$

where, as always, $\mu = \mu^0\nu$.

We are still free to choose \bar{x} and \bar{s} , subject to the constraints $A\bar{x} = b$ and $A^T\bar{y} + \bar{s} = c$.

Let \bar{x} be an optimal solution of (P) and (\bar{y}, \bar{s}) of (D) (assuming that these exist). Then \bar{x} and \bar{s} are complementary, i.e., $\bar{x}\bar{s} = 0$. Let ζ be such that $\|\bar{x} + \bar{s}\|_\infty \leq \zeta$. Starting the algorithm with

$$x^0 = s^0 = \zeta e, \quad y^0 = 0, \quad \mu^0 = \zeta^2,$$

the entries of the vectors $x^0 - \bar{x}$ and $s^0 - \bar{s}$ satisfy

$$0 \leq x^0 - \bar{x} \leq \zeta e, \quad 0 \leq s^0 - \bar{s} \leq \zeta e.$$

Thus it follows that

$$\sqrt{\|D(\bar{s} - s^0)\|^2 + \|D^{-1}(\bar{x} - x^0)\|^2} \leq \zeta \sqrt{\|De\|^2 + \|D^{-1}e\|^2} = \zeta \sqrt{e^T \left(\frac{x}{s} + \frac{s}{x} \right)}.$$

Substitution into (41) gives

$$\sqrt{\mu} \|q\| \leq \theta\nu \zeta \sqrt{e^T \left(\frac{x}{s} + \frac{s}{x} \right)}. \quad (42)$$

To proceed we need upper and lower bounds for the elements of the vectors x and s .

4.5 Bounds for x and s ; choice of τ and α

Recall that x is feasible for (P_ν) and (y, s) for (D_ν) and, moreover $\delta(x, s; \mu) \leq \tau$, i.e., these iterates are close to the μ -centers of (P_ν) and (D_ν) . Based on this information we need to estimate the sizes of the entries of the vectors x/s and s/x .

For this we use Theorem A.7, which gives

$$\begin{aligned}\sqrt{\frac{x}{s}} &\leq \frac{1 + \sqrt{2\tau'}}{\chi(\tau')} \frac{x(\mu, \nu)}{\sqrt{\mu}} \\ \sqrt{\frac{s}{x}} &\leq \frac{1 + \sqrt{2\tau'}}{\chi(\tau')} \frac{s(\mu, \nu)}{\sqrt{\mu}},\end{aligned}$$

where

$$\tau' := \psi(\rho(\tau)), \quad \psi(t) = \frac{1}{2}(t^2 - 1) - \log t,$$

and where $\chi : [0, \infty) \rightarrow (0, 1]$ is the inverse function of $\psi(t)$ for $0 < t \leq 1$, as defined in (46).

We choose

$$\tau = \frac{1}{8}. \tag{43}$$

Then $\tau' = 0.016921$, $1 + \sqrt{2\tau'} = 1.18396$ and $\chi(\tau') = 0.872865$, whence

$$\frac{1 + \sqrt{2\tau'}}{\chi(\tau')} = 1.35641 < \sqrt{2}.$$

It follows that

$$\sqrt{\frac{x}{s}} \leq \sqrt{2} \frac{x(\mu, \nu)}{\sqrt{\mu}}, \quad \sqrt{\frac{s}{x}} \leq \sqrt{2} \frac{s(\mu, \nu)}{\sqrt{\mu}}.$$

Substitution into (42) gives

$$\sqrt{\mu} \|q\| \leq \theta \nu \zeta \sqrt{2e^T \left(\frac{x(\mu, \nu)^2}{\mu} + \frac{s(\mu, \nu)^2}{\mu} \right)}.$$

This implies

$$\mu \|q\| \leq \theta \nu \zeta \sqrt{2} \sqrt{\|x(\mu, \nu)\|^2 + \|s(\mu, \nu)\|^2}.$$

Therefore, also using $\mu = \mu^0 \nu = \zeta^2 \nu$ and $\theta = \frac{\alpha}{\sqrt{2n}}$, we obtain the following upper bound for the norm of q :

$$\|q\| \leq \frac{\alpha}{\zeta \sqrt{n}} \sqrt{\|x(\mu, \nu)\|^2 + \|s(\mu, \nu)\|^2}.$$

We define

$$\kappa(\zeta, \nu) = \frac{\sqrt{\|x(\mu, \nu)\|^2 + \|s(\mu, \nu)\|^2}}{\zeta \sqrt{2n}}, \quad 0 < \nu \leq 1, \mu = \mu^0 \nu.$$

Note that since $x(\zeta^2, 1) = s(\zeta^2, 1) = \zeta e$, we have $\kappa(\zeta, 1) = 1$. Now we may write

$$\|q\| \leq \alpha \bar{\kappa}(\zeta) \sqrt{2} \quad \text{where} \quad \bar{\kappa}(\zeta) = \max_{0 < \nu \leq 1} \kappa(\zeta, \nu).$$

We found in (40) that in order to have $\delta(v^+) \leq 1/\sqrt{2}$, we should have $\|q\|^2 + (\|q\| + 2\delta)^2 \leq 1$. Therefore, since $\delta \leq \tau = \frac{1}{8}$, it suffices if q satisfies $\|q\|^2 + (\|q\| + \frac{1}{4})^2 \leq 1$. This holds if and only if $\|q\| \leq 0.570971$. Since $\sqrt{2}/3 = 0.471405$, we conclude that if we take

$$\alpha = \frac{1}{3\bar{\kappa}(\zeta)}, \quad (44)$$

then we will certainly have $\delta(v^+) \leq \frac{1}{\sqrt{2}}$.

There is some computational evidence that if ζ is large enough, then $\bar{\kappa}(\zeta) = 1$. This requires further investigation. We can prove that $\bar{\kappa}(\zeta) \leq \sqrt{2n}$. This is shown in the next section.

4.6 Bound for $\bar{\kappa}(\zeta)$

Due to the choice of the vectors \bar{x} , \bar{y} , \bar{s} and the number ζ (cf. Section 4.4) we have

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

To simplify notation in the rest of this section, we denote $x = x(\mu, \nu)$, $y = y(\mu, \nu)$ and $s = s(\mu, \nu)$. Then x , y and s are uniquely determined by the system

$$\begin{aligned} b - Ax &= \nu(b - A\zeta e), & x \geq 0 \\ c - A^T y - s &= \nu(c - \zeta e), & s \geq 0 \\ xs &= \nu\zeta^2 e. \end{aligned}$$

Hence we have

$$\begin{aligned} A\bar{x} - Ax &= \nu(A\bar{x} - A\zeta e), & x \geq 0 \\ A^T\bar{y} + \bar{s} - A^T y - s &= \nu(A^T\bar{y} + \bar{s} - \zeta e), & s \geq 0 \\ xs &= \nu\zeta^2 e. \end{aligned}$$

We rewrite this system as

$$\begin{aligned} A(\bar{x} - x - \nu\bar{x} + \nu\zeta e) &= 0, & x \geq 0 \\ A^T(\bar{y} - y - \nu\bar{y}) &= s - \bar{s} + \nu\bar{s} - \nu\zeta e, & s \geq 0 \\ xs &= \nu\zeta^2 e. \end{aligned}$$

Using again that the row space of a matrix and its null space are orthogonal, we obtain

$$(\bar{x} - x - \nu\bar{x} + \nu\zeta e)^T (s - \bar{s} + \nu\bar{s} - \nu\zeta e) = 0.$$

Defining

$$a := (1 - \nu)\bar{x} + \nu\zeta e, \quad b := (1 - \nu)\bar{s} + \nu\zeta e,$$

we have $(a - x)^T (b - s) = 0$. This gives

$$a^T b + x^T s = a^T s + b^T x.$$

Since $\bar{x}^T \bar{s} = 0$, $\bar{x} + \bar{s} \leq \zeta e$ and $xs = \nu \zeta^2 e$, we may write

$$\begin{aligned} a^T b + x^T s &= ((1 - \nu)\bar{x} + \nu \zeta e)^T ((1 - \nu)\bar{s} + \nu \zeta e) + \nu \zeta^2 n \\ &= \nu(1 - \nu) (\bar{x} + \bar{s})^T \zeta e + \nu^2 \zeta^2 n + \nu \zeta^2 n \\ &\leq \nu(1 - \nu) (\zeta e)^T \zeta e + \nu^2 \zeta^2 n + \nu \zeta^2 n \\ &= \nu(1 - \nu) \zeta^2 n + \nu^2 \zeta^2 n + \nu \zeta^2 n = 2\nu \zeta^2 n. \end{aligned}$$

Moreover, also using $a \geq \nu \zeta e$, $b \geq \nu \zeta e$, we get

$$\begin{aligned} a^T s + b^T x &= ((1 - \nu)\bar{x} + \nu \zeta e)^T s + ((1 - \nu)\bar{s} + \nu \zeta e)^T x \\ &= (1 - \nu) (s^T \bar{x} + x^T \bar{s}) + \nu \zeta e^T (x + s) \\ &\geq \nu \zeta e^T (x + s) = \nu \zeta (\|s\|_1 + \|x\|_1). \end{aligned}$$

Hence we obtain $\|s\|_1 + \|x\|_1 \leq 2\zeta n$. Since $\|x\|^2 + \|s\|^2 \leq (\|s\|_1 + \|x\|_1)^2$, it follows that

$$\frac{\sqrt{\|x\|^2 + \|s\|^2}}{\zeta \sqrt{2n}} \leq \frac{\|s\|_1 + \|x\|_1}{\zeta \sqrt{2n}} \leq \frac{2\zeta n}{\zeta \sqrt{2n}} = \sqrt{2n},$$

thus proving

$$\bar{\kappa}(\zeta) \leq \sqrt{2n}.$$

4.7 Complexity analysis

In the previous sections we have found that if at the start of an iteration the iterates satisfy $\delta(x, s; \mu) \leq \tau$, with τ as defined in (43), then after the feasibility step, with θ as defined in (34), and α as in (44), the iterates satisfy $\delta(x, s; \mu^+) \leq 1/\sqrt{2}$.

According to (24), at most

$$\log_2 \left(\log_2 \frac{1}{\tau^2} \right) = \log_2 (\log_2 64) = 3$$

centering steps suffice to get iterates that satisfy $\delta(x, s; \mu^+) \leq \tau$. So each iteration consists of at most 4 so-called *inner iterations*, in each of which we need to compute a new search direction. It has become a custom to measure the complexity of an IPM by the required number of inner iterations. In each iteration both the duality gap and the norms of the residual vectors are reduced by the factor $1 - \theta$. Hence, using $x^{0T} s^0 = n\zeta^2$, the total number of iterations is bounded above by

$$\frac{1}{\theta} \log \frac{\max \{n\zeta^2, \|r_b^0\|, \|r_c^0\|\}}{\varepsilon}.$$

Since

$$\theta = \frac{\alpha}{\sqrt{2n}} = \frac{1}{3\bar{\kappa}(\zeta)\sqrt{2n}},$$

the total number of inner iterations is therefore bounded above by

$$12 \bar{\kappa}(\zeta) \sqrt{2n} \log \frac{\max \{n\zeta^2, \|r_b^0\|, \|r_c^0\|\}}{\varepsilon},$$

where $\bar{\kappa}(\zeta) \leq \sqrt{2n}$.

5 Concluding remarks

The current paper shows that the techniques that have been developed in the field of feasible full-Newton step IPMs, and which have now been known for almost twenty years, are sufficient to get a full-Newton step IIPM whose performance is at least as good as the currently best known performance of IIPMs. Following a well-known metaphor of Isaac Newton⁴, it looks like if a “smooth pebble or pretty shell on the sea-shore of IPMs” has been overlooked for a surprisingly long time.

It may be clear that the full-Newton step method presented in this paper is not efficient from a practical point of view, just as the feasible IPMs with the best theoretical performance are far from practical. But just as in the case of feasible IPMs one might expect that computationally efficient large-update methods for IIPMs can be designed whose theoretical complexity is not worse than \sqrt{n} times the iteration bound in this paper. Even better results for large-update methods might be obtained by changing the search direction, by using methods that are based on kernel functions, as presented in [3, 13]. This requires further investigation. Also extensions to second-order cone optimization, semidefinite optimization, linear complementarity problems, etc. seem to be within reach.

Acknowledgements

Hossein Mansouri, PhD student in Delft, forced the author to become more familiar with IIPMs when, in September 2004, he showed him a draft of a paper on large-update IIPMs based on kernel functions. Thank you, Hossein! I am also much indebted to Jean-Philippe Vial. He found a serious flaw in an earlier version of this paper. Thanks are also due Kurt Anstreicher, Florian Potra, Shinji Mizuno and other colleagues for some useful discussions during the Oberwolfach meeting *Optimization and Applications* in January 2005. I also thank Yanqin Bai for pointing out some typos in an earlier version of the paper.

References

- [1] E. D. Andersen and Y. Ye. A computational study of the homogeneous algorithm for large-scale convex optimization. *Computational Applications and Optimization*, 10:243–269, 1998.
- [2] E.D. Andersen and K.D. Andersen. The MOSEK interior-point optimizer for Linear Programming: an implementation of the homogeneous algorithm. In S. Zhang, H. Frenk, C. Roos, and T. Terlaky, editors, *High Performance Optimization Techniques*, pages 197–232. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.
- [3] Y.Q. Bai, M. El Ghami, and C. Roos. A comparative study of kernel functions for primal-dual interior-point algorithms in linear optimization. *SIAM Journal on Optimization*, 15(1):101–128 (electronic), 2004.
- [4] J.F. Bonnans and F.A. Potra. On the convergence of the iteration sequence of infeasible path following algorithms for linear complementarity problems. *Mathematics of Operations Research*, 22(2):378–407, 1997.

⁴“ I do not know what I may appear to the world; but to myself I seem to have been only like a boy playing on the sea-shore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me” [23, p. 863].

- [5] R.M. Freund. A potential-function reduction algorithm for solving a linear program directly from an infeasible 'warm start'. *Mathematical Programming*, 52:441–466, 1991.
- [6] N. Karmarkar. New polynomial-time algorithm for linear programming. *Combinatorica* 4, pages 373–395, 1984.
- [7] M. Kojima, N. Megiddo, and S.Mizuno. A primal-dual infeasible-interior-point algorithm for linear programming. *Mathematical Programming* 61, pages 263–280, 1993.
- [8] M. Kojima, T. Noma, and A. Yoshise. Global convergence in infeasible-interior-point algorithms. *Mathematical Programming, Series A*, 65:43–72, 1994.
- [9] I. J. Lustig. Feasible issues in a primal-dual interior-point method. *Mathematical Programming* 67, pages 145–162, 1990/91.
- [10] S. Mizuno. Polynomiality of infeasible-interior-point algorithms for linear programming. *Mathematical Programming* 67, pages 109–119, 1994.
- [11] S. Mizuno, M.J. Todd, and Y. Ye. A surface of analytic centers and infeasible-interior-point algorithms for linear programming. *Mathematics of Operations Research*, 20:135–162, 1995.
- [12] Yu. Nesterov, M. J. Todd, and Y. Ye. Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. *Mathematical Programming*, 84(2, Ser. A):227–267, 1999.
- [13] J. Peng, C. Roos, and T. Terlaky. *Self-Regularity. A New Paradigm for Primal-Dual Interior-Point Algorithms*. Princeton University Press, 2002.
- [14] F. A. Potra. A quadratically convergent predictor-corrector method for solving linear programs from infeasible starting points. *Mathematical Programming*, 67(3, Ser. A):383–406, 1994.
- [15] F. A. Potra. An infeasible-interior-point predictor-corrector algorithm for linear programming. *SIAM Journal on Optimization*, 6(1):19–32, 1996.
- [16] C. Roos, T. Terlaky, and J.-Ph. Vial. *Theory and Algorithms for Linear Optimization. An Interior-Point Approach*. John Wiley & Sons, Chichester, UK, 1997.
- [17] M. Salahi, T. Terlaky, and G. Zhang. The complexity of self-regular proximity based infeasible IPMs. Technical Report 2003/3, Advanced Optimization Laboratory, Mc Master University, Hamilton, Ontario, Canada, 2003.
- [18] R. Sheng and F. A. Potra. A quadratically convergent infeasible-interior-point algorithm for LCP with polynomial complexity. *SIAM Journal on Optimization*, 7(2):304–317, 1997.
- [19] J.F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods & Software*, 11:625–653, 1999.
- [20] K. Tanabe. Centered Newton method for linear programming: Interior and 'exterior' point method (in Japanese). In: *New Methods for Linear Programming. K. Tone (Ed.)* 3, pages 98–100, 1990.
- [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, Ser. A):1–21, 1998.
- [22] R.J. Vanderbei. *Linear Programming: Foundations and Extensions*. Kluwer Academic Publishers, Boston, USA, 1996.
- [23] R. S. Westfall. *Never at Rest. A biography of Isaac Newton*. Cambridge University Press, Cambridge, UK, 1964.
- [24] S. J. Wright. An infeasible-interior-point algorithm for linear complementarity problems. *Mathematical Programming*, 67(1):29–52, 1994.
- [25] S.J. Wright. A path-following infeasible-interior-point algorithm for linear complementarity problems. *Optimization Methods & Software*, 2:79–106, 1993.

- [26] S.J. Wright. *Primal-Dual Interior-Point Methods*. SIAM, Philadelphia, 1996.
- [27] F. Wu, S. Wu, and Y. Ye. On quadratic convergence of the $O(\sqrt{n}L)$ -iteration homogeneous and self-dual linear programming algorithm. *Annals of Operations Research*, 87:393–406, 1999.
- [28] Y. Ye. *Interior Point Algorithms, Theory and Analysis*. John Wiley & Sons, Chichester, UK, 1997.
- [29] 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.
- [30] Y. Zhang. On the convergence of a class of infeasible-interior-point methods for the horizontal linear complementarity problem. *SIAM Journal on Optimization*, 4:208–227, 1994.

A Some technical lemmas

Given a strictly primal feasible solution x of (P) and a strictly dual feasible solution (y, s) of (D) , and $\mu > 0$, let

$$\Phi(xs; \mu) := \Psi(v) := \sum_{i=1}^n \psi(v_i), \quad v_i := \sqrt{\frac{x_i s_i}{\mu}}, \quad \psi(t) := \frac{1}{2} (t^2 - 1 - \log t^2).$$

It is well known that $\psi(t)$ is the kernel function of the primal-dual logarithmic barrier function, which, up to some constant, is the function $\Phi(xs; \mu)$ (see, e.g., [3]).

Lemma A.1 *One has*

$$\Phi(xs; \mu) = \Phi(xs(\mu); \mu) + \Phi(x(\mu)s; \mu).$$

Proof: The equality in the lemma is equivalent to

$$\sum_{i=1}^n \left(\frac{x_i s_i}{\mu} - 1 - \log \frac{x_i s_i}{\mu} \right) = \sum_{i=1}^n \left(\frac{x_i(\mu) s_i}{\mu} - 1 - \log \frac{x_i(\mu) s_i}{\mu} \right) + \sum_{i=1}^n \left(\frac{x_i s_i(\mu)}{\mu} - 1 - \log \frac{x_i s_i(\mu)}{\mu} \right).$$

Since

$$\log \frac{x_i s_i}{\mu} = \log \frac{x_i s_i}{x_i(\mu) s_i(\mu)} = \log \frac{x_i}{x_i(\mu)} + \log \frac{s_i}{s_i(\mu)} = \log \frac{x_i s_i(\mu)}{\mu} + \log \frac{x_i(\mu) s_i}{\mu},$$

the lemma holds if and only if

$$x^T s - n\mu = (x^T s(\mu) - n\mu) + (s^T x(\mu) - n\mu).$$

Using that $x(\mu)s(\mu)$, whence $x(\mu)^T s(\mu) = n\mu$, this can be written as $(x - x(\mu))^T (s - s(\mu)) = 0$, which holds if the vectors $x - x(\mu)$ and $s - s(\mu)$ are orthogonal. This is indeed the case, because $x - x(\mu)$ belongs to the null space and $s - s(\mu)$ to the row space of A . This proves the lemma. \square

Theorem A.2 *Let $\delta(v)$ be as defined in (10) and $\rho(\delta)$ as in (29). Then $\Psi(v) \leq \psi(\rho(\delta(v)))$.*

Proof: The statement in the lemma is obvious if $v = e$ since then $\delta(v) = \Psi(v) = 0$ and since $\rho(0) = 1$ and $\psi(1) = 0$. Otherwise we have $\delta(v) > 0$ and $\Psi(v) > 0$. To deal with the nontrivial case we consider, for $\tau > 0$, the problem

$$z_\tau = \max_v \left\{ \Psi(v) = \sum_{i=1}^n \psi(v_i) : \delta(v)^2 = \frac{1}{4} \sum_{i=1}^n \psi'(v_i)^2 = \tau^2 \right\}.$$

The first order optimality conditions are

$$\psi'(v_i) = \lambda \psi'(v_i) \psi''(v_i), \quad i = 1, \dots, n,$$

where $\lambda \in \mathbf{R}$. From this we conclude that we have either $\psi'(v_i) = 0$ or $\lambda \psi''(v_i) = 1$, for each i . Since $\psi''(t)$ is monotonically decreasing, this implies that all v_i 's for which $\lambda \psi''(v_i) = 1$ have the same value. Denoting this value as t , and observing that all other coordinates have value 1 (since $\psi'(v_i) = 0$ for these coordinates), we conclude that for some k , and after reordering the coordinates, v has the form

$$v = (\underbrace{t, \dots, t}_k \text{ times}, \underbrace{1, \dots, 1}_{n-k} \text{ times}).$$

Since $\psi'(1) = 0$, $\delta(v) = \tau$ implies $k\psi'(t)^2 = 4\tau^2$. Since $\psi'(t) = t - 1/t$, it follows that

$$t - \frac{1}{t} = \pm \frac{2\tau}{\sqrt{k}},$$

which gives $t = \rho(\tau/\sqrt{k})$ or $t = 1/\rho(\tau/\sqrt{k})$. The first value, which is greater than 1, gives the largest value of $\psi(t)$. This follows because

$$\psi(t) - \psi\left(\frac{1}{t}\right) = \frac{1}{2} \left(t^2 - \frac{1}{t^2} \right) \geq 0, \quad t \geq 1.$$

Since we are maximizing $\Psi(v)$, we conclude that $t = \rho(\tau/\sqrt{k})$, whence we have

$$\Psi(v) = k \psi \left(\rho \left(\frac{\tau}{\sqrt{k}} \right) \right).$$

The question remains which value of k maximizes $\Psi(v)$. To investigate this we take the derivative of $\Psi(v)$ with respect to k . To simplify the notation we write

$$\Psi(v) = k \psi(t), \quad t = \rho(s), \quad s = \frac{\tau}{\sqrt{k}}.$$

The definition of t implies $t = s + \sqrt{1+s^2}$. This gives $(t-s)^2 = 1+s^2$, or $t^2 - 1 = 2st$, whence we have

$$2s = t - \frac{1}{t} = \psi'(t).$$

Some straightforward computations now yield

$$\frac{d\Psi(v)}{dk} = \psi(t) - \frac{s^2 \rho(s)}{\sqrt{1+s^2}} =: f(\tau).$$

We consider this derivative as a function of τ , as indicated. One may easily verify that $f(\tau) = 0$. We proceed by computing the derivative with respect to τ . This gives

$$f'(\tau) = -\frac{1}{\sqrt{k}} \frac{s^2}{(1+s^2)\sqrt{1+s^2}}$$

This proves that $f'(\tau) \leq 0$. Since $f(\tau) = 0$, it follows that $f(\tau) \leq 0$, for each $\tau \geq 0$. Hence we conclude that $\Psi(v)$ is decreasing in k . So $\Psi(v)$ is maximal when $k = 1$, which gives the result in the theorem. \square

Corollary A.3 *Let $\tau \geq 0$ and $\delta(v) \leq \tau$. Then $\Psi(v) \leq \tau'$, where*

$$\tau' := \psi(\rho(\tau)). \tag{45}$$

Proof: Since $\rho(s)$ is monotonically increasing in s , and $\rho(s) \geq 1$ for all $s \geq 0$, and, moreover, $\psi(t)$ is monotonically increasing if $t \geq 1$, the function $\psi(\rho(\delta))$ is increasing in δ , for $\delta \geq 0$. Thus the result is immediate from Theorem A.2. \square

Lemma A.4 *Let $\delta(v) \leq \tau$ and let τ' be as defined in (43). Then*

$$\psi \left(\sqrt{\frac{x_i}{x_i(\mu)}} \right) \leq \tau', \quad \psi \left(\sqrt{\frac{s_i}{s_i(\mu)}} \right) \leq \tau', \quad i = 1, \dots, n.$$

Proof: By Lemma A.1 we have $\Phi(xs; \mu) = \Phi(xs(\mu); \mu) + \Phi(x(\mu)s; \mu)$. Since $\Phi(xs; \mu)$ is always nonnegative, also $\Phi(xs(\mu); \mu)$ and $\Phi(x(\mu)s; \mu)$ are nonnegative. By Corollary A.3 we have $\Phi(xs; \mu) \leq \tau'$. Thus it follows that $\Phi(xs(\mu); \mu) \leq \tau'$ and $\Phi(x(\mu)s; \mu) \leq \tau'$. The first of these two inequalities gives

$$\Phi(xs(\mu); \mu) = \sum_{i=1}^n \psi \left(\sqrt{\frac{x_i s_i(\mu)}{\mu}} \right) \leq \tau'.$$

Since $\psi(t) \geq 0$, for every $t > 0$, it follows that

$$\psi \left(\sqrt{\frac{x_i s_i(\mu)}{\mu}} \right) \leq \tau', \quad i = 1, \dots, n.$$

Due to $x(\mu)s(\mu) = \mu e$, we have

$$\frac{x_i s_i(\mu)}{\mu} = \frac{x_i s_i(\mu)}{x_i(\mu) s_i(\mu)} = \frac{x_i}{x_i(\mu)},$$

whence we obtain the first inequality in the lemma. The second inequality follows in the same way. \square

In the sequel we use the inverse function of $\psi(t)$ for $0 < t \leq 1$, which is denoted as $\chi(s)$. So $\chi : [0, \infty) \rightarrow (0, 1]$, and we have

$$\chi(s) = t \Leftrightarrow s = \psi(t), \quad s \geq 0, 0 < t \leq 1. \quad (46)$$

Lemma A.5 For each $t > 0$ one has $\chi(\psi(t)) \leq t \leq 1 + \sqrt{2\psi(t)}$.

Proof: Suppose $\psi(t) = s \geq 0$. Since $\psi(t)$ is convex, minimal at $t = 1$, with $\psi(1) = 0$, and since $\psi(t)$ goes to infinity both if t goes to zero and if t goes to infinity, there exist precisely two values a and b such that $\psi(a) = \psi(b) = s$, with $0 < a \leq 1 \leq b$.

Since $a \leq 1$ we have $a = \chi(s) = \chi(\psi(t))$, proving the left inequality in the lemma. For b we may write

$$s = \frac{1}{2} (b^2 - 1) - \log b \leq \frac{1}{2} (b^2 - 1),$$

which implies $b \leq 1 + \sqrt{2s}$, proving the right inequality. \square

Corollary A.6 If $\delta(v) \leq \tau$ then

$$\chi(\tau') \leq \sqrt{\frac{x_i}{x_i(\mu)}} \leq 1 + \sqrt{2\tau'}, \quad \chi(\tau') \leq \sqrt{\frac{s_i}{s_i(\mu)}} \leq 1 + \sqrt{2\tau'}$$

Proof: This is immediate from Lemma A.5 and Lemma A.4. \square

Theorem A.7 If $\delta(v) \leq \tau$ then

$$\begin{aligned} \sqrt{\frac{x}{s}} &\leq \frac{1 + \sqrt{2\tau'}}{\chi(\tau')} \frac{x(\mu)}{\sqrt{\mu}} \\ \sqrt{\frac{s}{x}} &\leq \frac{1 + \sqrt{2\tau'}}{\chi(\tau')} \frac{s(\mu)}{\sqrt{\mu}}. \end{aligned}$$

Proof: Using that $x_i(\mu)s_i(\mu) = \mu$, for each i , Corollary A.6 implies

$$\sqrt{\frac{x_i}{s_i}} \leq \frac{(1 + \sqrt{2\tau'}) \sqrt{x_i(\mu)}}{\chi(\tau') \sqrt{s_i(\mu)}} = \frac{1 + \sqrt{2\tau'}}{\chi(\tau')} \sqrt{\frac{x_i(\mu)}{s_i(\mu)}} = \frac{1 + \sqrt{2\tau'}}{\chi(\tau')} \sqrt{\frac{x_i^2(\mu)}{\mu}} = \frac{1 + \sqrt{2\tau'}}{\chi(\tau')} \frac{x_i(\mu)}{\sqrt{\mu}},$$

which implies the first inequality. The second inequality is obtained in the same way. \square