ELSEVIER

European Journal of Operational Research 143 (2002) 257–267

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/dsw

The Mizuno–Todd–Ye algorithm in a larger neighborhood of the central path

Florian A. Potra

Department of Mathematics and Statistics, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA

Received 28 November 2000; accepted 15 September 2001

Abstract

The Mizuno–Todd–Ye predictor–corrector method based on two neighborhoods $\mathscr{D}(\alpha) \subset \mathscr{D}(\bar{\alpha})$ of the central path of a monotone homogeneous linear complementarity problem is analyzed, where $\mathscr{D}(\alpha)$ is composed of all feasible points with δ-proximity to the central path less than or equal to α. The largest allowable value for $\bar{\alpha}$ is ≈ 1.76. For a specific choice of α and $\bar{\alpha}$ a lower bound of $\chi_n/\sqrt{n}$ is obtained for the stepsize along the affine-scaling direction, where χ_n has an asymptotic value greater than 1.08. For $n \geqslant 400$ it is shown that $\chi_n > 1.05$. The algorithm has $O(\sqrt{n}L)$-iteration complexity under general conditions and quadratic convergence under the strict complementarity assumption.

Keywords: Linear programming; Interior-point algorithm; Path-following

1. Introduction

The Mizuno–Todd–Ye predictor–corrector method (MTY) is one of the most remarkable interior-point methods for linear programming (LP), quadratic programming (QP) and linear complementarity problems (LCP). It is based on a very simple and elegant idea that is used in various other fields of computational mathematics such as the numerical solution of differential equations [4] and continuation methods [13]. The general problem is to find a numerical approximation of a curve $\mathscr{C}$ in a finite dimensional space. This curve may represent the central path of an LCP, the solution of a system of differential equations, the set of solutions of a family of static equilibrium problems that depends on a parameter of interest, and so on. A predictor–corrector method produces a sequence of points $z^k \in \mathscr{N}$, where $\mathscr{N}$ is a neighborhood of the curve $\mathscr{C}$. At a typical iteration of a predictor–corrector method one is given a point $z \in \mathscr{N}$ that approximates a point z_τ on the curve $\mathscr{C}$, where the scalar τ parameterizes the curve, and one has to compute a new value curve τ_+ of this parameter and a new point $z^+ \in \mathscr{N}$ that approximates z_{τ_+}. A certain procedure, called predictor, is used to compute a rough approximation $\bar{z} \in \overline{\mathscr{N}}$ of $z_{\tau_+} \in \mathscr{C}$, where $\overline{\mathscr{N}}$ is a larger neighborhood of $\mathscr{C}$. Then another procedure, called corrector, is used to improve this approximation by producing a point $z^+ \in \mathscr{N}$. For example some of the most efficient software packages for solving systems of differential equations use the

E-mail address: potra@math.umbc.edu (F.A. Potra).

0377-2217/02/$ - see front matter
PII: S0377-2217(02)00388-0

Adams–Bashford scheme for predictor and the Adams–Moulton scheme for corrector. The importance of appropriately choosing the new value τ_+ of the parameter τ cannot be overestimated. For example variable stepsize differential equations solvers, where τ_+ is chosen dynamically at each step, are far superior to their fixed stepsize counterparts that had been used in the past. The same observation applies to the MTY. In the predictor step of MTY one takes $\bar{z} = z(\bar{\theta})$ and $\tau_+ = (1 - \bar{\theta})\tau$ with $z(\theta) = z + \theta w$, where w is the affine scaling direction and $\bar{\theta}$ is the largest value of $\theta \in (0, 1)$ such that $z(\theta) \in \overline{\mathcal{N}}$. This adaptive way of choosing τ_+ accounts for most of the success of MTY. While the path following methods considered prior to MTY used a fixed $\bar{\theta}$, and therefore their practical performance was about the same as their worst case computational complexity bounds, the practical performance of MTY is much better than its worst case computational complexity bounds would indicate. In the original paper [9] it is shown that if $\mathcal{N} = \mathcal{N}_{0.25}$ and $\overline{\mathcal{N}} = \mathcal{N}_{0.5}$, where $\mathcal{N}_\alpha$ is an l_2 neighborhood of the central path with radius α (see (2.22) in Section 2), then the parameter $\bar{\theta}$, computed as described above, satisfies the inequality $\bar{\theta} \geqslant \hat{\chi}/\sqrt{n}$, where

$$\hat{\chi} = \frac{1}{\sqrt[4]{8}} \approx = 0.5946.$$

This implies that the sequence $\{\tau_k\}$ convergence to zero with a linear rate at least $1 - \hat{\chi}/\sqrt{n}$, i.e.,

$$\tau_{k+1} \leqslant (1 - \hat{\chi}/\sqrt{n})\tau_k, \quad k = 1, 2, \ldots. \tag{1.1}$$

Using the above inequality it is easily proved that MTY has $O(\sqrt{n}L)$-iteration complexity. The fixed stepsize path following methods considered prior to MTY satisfy a relation similar to (1.1) and have the same iteration complexity. The difference is that the sequence $\{\tau_k\}$ produced by MTY converges much faster than indicated by (1.1). Thus, Ye et al. [16] proved the surprising fact that the sequence $\{\tau_k\}$ converges quadratically to zero in the sense that

$$\tau_{k+1} \leqslant \sigma\tau_k^2, \quad k = 1, 2 \ldots, \quad \text{for some constant } \sigma > 0. \tag{1.2}$$

For large values of k, for example when $\sigma\tau_k \ll 1$, (1.2) indicates a much faster decrease of τ_{k+1} than indicated by (1.1). Paper [16] answered a question that had been open for a while: are there interior point methods for LP that have both polynomial complexity and superlinear convergence? MTY was the first algorithm for which both properties were proved to hold. MTY was generalized to (monotone) LCP in [7] and the resulting algorithm was proved to have $O(\sqrt{n}L)$ iteration complexity under general conditions and superlinear convergence under the assumption that the LCP has a strictly complementary condition and the iteration sequence converges. From [2] it follows that the latter assumption holds under general conditions. Subsequently Ye and Anstreicher [15] proved that MTY converges quadratically assuming only that the LCP has a strictly complementary solution. The strict complementarity assumption is not restrictive, since according to [10] a large class of interior point methods, which contains MTY, can have only linear convergence if this assumption is violated.

A natural question arises: can we take a larger neighborhood of the central path and still have one predictor followed by just one corrector and retain the $O(\sqrt{n}L)$ iteration complexity and quadratic convergence of the original MTY? We will show that this is possible by replacing the l_2-neighborhoods, by δ-neighborhoods, where δ is the proximity measure introduced in [5] (see also [14]). We mention that a predictor–corrector method for LP using δ-neighborhoods is analyzed in [14]. We will improve on the results in [14] by analyzing the algorithm for LCP and by considering slightly larger δ-neighborhoods. Our version of the MTY uses two neighborhoods $\mathcal{D}(\alpha) \subset \mathcal{D}(\bar{\alpha})$ of the central path of a monotone homogeneous LCP, where $\mathcal{D}(\alpha)$ is composed of all feasible points with δ-proximity to the central path less than or equal to α. We will show that $\bar{\alpha}$ can be chosen as close to $(\sqrt{17} - 1)^{1/2} \approx 1.767$ as desired. For $\alpha = 1/4$ and $\bar{\alpha} = 5/6$ we will obtain a lower bound of $\chi_n/\sqrt{n}$ for the stepsize θ along the affine-scaling direction, where $\chi_n \uparrow \sqrt{24}/(33 + \sqrt{65}) \approx 1.081$. We will show that $\chi_n > 1.05$ for all $n \geqslant 400$ and that $\chi_n > 1.075$ for all $n \geqslant 9000$. These are the largest guaranteed lower bounds for the stepsize along the affine-scaling direction that we know of in the interior point literature. Our predictor–corrector algorithm has $O(\sqrt{n}L)$-iteration complexity under general conditions and quadratic convergence under the strict complementarity assumption.

Conventions: We denote by $\mathbb{N}$ the set of all nonnegative integers. $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}_{++}$ denote the set of real, nonnegative real, and positive real numbers respectively. For any real number γ, $\lceil\gamma\rceil$ denotes the smallest integer greater or equal to γ. Given a vector x, the corresponding upper case symbol denotes as usual the diagonal matrix X defined by the vector. The symbol e represents the vector of all ones, with dimension given by the context.

We denote component-wise operations on vectors by the usual notations for real numbers. Thus, given two vectors u, v of the same dimension, uv, u/v, etc. will denote the vectors with components u_iv_i, u_i/v_i, etc. This notation is consistent as long as component-wise operations always have precedence in relation to matrix operations. Note that $uv \equiv Uv$. If A is a matrix then $Auv \equiv AUv$, but in general $Auv \neq (Au)v$. Also if f is a scalar function and v is a vector, then $f(v)$ denotes the vector with components $f(v_i)$. For example if $v \in \mathbb{R}^n_+$, then $\sqrt{v}$ denotes the vector with components $\sqrt{v_i}$, and $1 - v$ denotes the vector with components $1 - v_i$. Traditionally the vector $1 - v$ is written as $e - v$, where e is the vector of all ones. Inequalities are to be understood in a similar fashion. For example if $v \in \mathbb{R}^n$ then $v \geqslant 3$ means that $v_i \geqslant 3$, $i = 1, \ldots, n$. Traditionally this is written as $v \geqslant 3e$. If $\|\cdot\|$ is a vector norm on $\mathbb{R}^n$ and A is a matrix then the operator norm induced by $\|\cdot\|$ is defined by $\|A\| = \max\{\|Ax\|; \|x\| = 1\}$. As a particular case we note that if U is the diagonal matrix defined by the vector u, then $\|U\|_2 = \|u\|_\infty$.

We frequently use the $O(\cdot)$ and $\Omega(\cdot)$ notation to express the relationship between functions. Our most common usage will be associated with a sequence $\{x^k\}$ of vectors and a sequence $\{\tau_k\}$ of positive real numbers. In this case $x^k = O(\tau_k)$ means that there is a constant K (dependent on problem data) such that for every $k \in \mathbb{N}$, $\|x^k\| \leqslant K\tau_k$. Similarly, if $x^k > 0$, $x^k = \Omega(\tau_k)$ means that $(x^k)^{-1} = O(1/\tau_k)$.

Finally introduce a less standard notation. If x, $s \in \mathbb{R}^n$ then the vector $z \in \mathbb{R}^{2n}$ obtained by concatenating x and s will be denoted by $\lceil x, s \rfloor$, i.e.,

$$z = \lceil x, s \rfloor = \begin{bmatrix} x \\ s \end{bmatrix} = \left[x^{\mathrm{T}}, s^{\mathrm{T}}\right]^{\mathrm{T}}. \tag{1.3}$$

Throughout this paper the mean value of xs will be denoted by

$$\mu(z) = \frac{x^{\mathrm{T}}s}{n}. \tag{1.4}$$

2. The Mizuno–Todd–Ye predictor–corrector method

Given two matrices $Q, R \in \mathbb{R}^{n\times n}$ and a vector $b \in \mathbb{R}^n$, the horizontal LCP (HLCP) consists in finding a pair of vectors (x, s) such that

$$\begin{aligned} &xs = 0, \\ &Qx + Rs = b, \\ &x, s \geqslant 0. \end{aligned} \tag{2.1}$$

The standard (monotone) LCP is obtained by taking $R = -I$ and Q positive semidefinite. There are other formulations of the LCPs as well, but as shown in [1] all popular formulations are equivalent and the behaviour of a large class of interior point methods is identical on those formulations, so that it is sufficient to prove results only for one of the formulations. We have chosen HLCP because of its symmetry. The LP problem and the QP can be formulated as an HLCP. Therefore HLCP provides a convenient and general framework for studying interior point methods.

Throughout this paper we assume that the HLCP (2.1) is monotone in the sense that

$$Qu + Rv = 0 \quad \text{implies} \quad u^{\mathrm{T}}v \geqslant 0, \text{ for any } u, v \in \mathbb{R}^n. \tag{2.2}$$

This condition is satisfied if the HLCP is a reformulation of a QP and for many other interesting classes of problems (see the excellent monograph [3]). If the HLCP is a reformulation of an LP then the following stronger condition holds:

$$Qu + Rv = 0 \quad \text{implies} \quad u^{\mathrm{T}}v = 0, \quad \text{for any } u, v \in \mathbb{R}^n. \tag{2.3}$$

In this case we say that the HLCP is skew-symmetric. Since in the skew-symmetric case we can often obtain sharper estimates we are going to consider this particular case as well in this paper. Let us denote the set of all feasible points of HLCP by

$$\mathscr{F} = \{z = \lceil x, s \rfloor \in \mathbb{R}^{2n}_+ : \ Qx + Rs = b\}, \tag{2.4}$$

and the solution set (or the optimal face) of HLCP by

$$\mathscr{F}^* = \{z^* = \lceil x^*, s^* \rfloor \in \mathscr{F} : \ x^*s^* = 0\}. \tag{2.5}$$

The relative interior of $\mathscr{F}$,

$$\mathscr{F}^0 = \mathscr{F} \bigcap \mathbb{R}^{2n}_{++}, \tag{2.6}$$

will be called the set of strictly feasible points, or the set of interior points. It is known (see [8]) that if $\mathscr{F}^0$ is nonempty then for any parameter $\tau > 0$ the nonlinear system

$$\begin{aligned} xs &= \tau e, \\ Qx + Rs &= b \end{aligned} \tag{2.7}$$

has a unique positive solution. The set of all such solutions defines the central path $\mathscr{C}$ of the HLCP. By denoting

$$z = \lceil x, s \rfloor \tag{2.8}$$

and considering the quadratic mapping $F_\tau : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$,

$$F_\tau(z) = \begin{bmatrix} xs - \tau e \\ Qx + Rs - b \end{bmatrix}, \tag{2.9}$$

we can write

$$\mathscr{C} = \{(z, \tau) \in \mathbb{R}^{2n}_{++} \times \mathbb{R}_{++} : F_\tau(z) = 0\}. \tag{2.10}$$

In this paper we will consider neighborhoods of $\mathscr{C}$ of the form

$$\mathscr{D}(\alpha) = \{(z, \tau) \in \mathscr{F}^0 \times \mathbb{R}_{++} : \delta(z, \tau) \leqslant \alpha\}, \tag{2.11}$$

where δ is the proximity measure

$$\delta(z, \tau) := \left\| \sqrt{\frac{\tau}{xs}} - \sqrt{\frac{xs}{\tau}} \right\|_2 \tag{2.12}$$

considered in [5] (see also [14]).

Let α and $\overline{\alpha}$ be two positive numbers such that

$$0 < \alpha < \overline{\alpha} \leqslant \frac{2\sqrt{\alpha}}{\sqrt[4]{\alpha^2 + 2}}. \tag{2.13}$$

At a typical iteration of MTY a point $z = \lceil x, s \rfloor \in \mathscr{F}^0$ has already been computed. Let $z(\theta)$ be defined via a damped Newton step of the form

$$z(\theta) = z + \theta w, \tag{2.14}$$

where

$$w = \lceil u, v \rfloor = -F_0'(z)^{-1} F_0(z) \tag{2.15}$$

is the Newton direction of F_0 at z. This direction is called the affine scaling direction and it can be computed as the solution of the linear system

$$\begin{aligned} su + xv &= -xs, \\ Qu + Rv &= 0. \end{aligned} \tag{2.16}$$

The predictor step of MTY computes an optimal damping parameter

$$\overline{\theta} = \max\{\check{\theta} \in [0, 1] : z(\theta) \in \mathscr{D}(\overline{\alpha}), \forall \theta \in [0, \check{\theta}]\} \tag{2.17}$$

and obtains the predicted point $(\overline{z}, \overline{\tau}) \in \mathscr{D}(\overline{\alpha})$ as

$$\overline{z} = z(\overline{\theta}), \quad \overline{\tau} = (1 - \overline{\theta})\tau. \tag{2.18}$$

The corrector step produces $(z^+, \tau_+) \in \mathscr{D}(\alpha)$ via

$$z^+ = \overline{z} + \overline{w}, \quad \tau_+ = (1 - \overline{\theta})\tau, \tag{2.19}$$

where

$$\overline{w} = \lceil \overline{x}, \overline{s} \rfloor = -F_{\overline{\tau}}'(\overline{z})^{-1} F_{\overline{\tau}}(\overline{z}) \tag{2.20}$$

is the Newton direction of $F_{\overline{\tau}}$ at $\overline{z}$. This direction is called the centering direction and it can be computed as the solution of the following linear system:

$$\begin{aligned} \overline{s}\,\overline{u} + \overline{x}\,\overline{v} &= \overline{\tau} - \overline{x}\,\overline{s}, \\ Q\overline{u} + R\overline{v} &= 0. \end{aligned} \tag{2.21}$$

We can now define the whole algorithm by the following pseudo-code:

Algorithm 1 (*MTY*).
Given α, $\overline{\alpha}$ satisfying (2.13) and $z^0 \in \mathscr{F}^0$ with $(z^0, \mu(z^0)) \in \mathscr{D}(\alpha)$:
Set $k \leftarrow 0$ and $\tau_0 \leftarrow \mu(z^0)$;
repeat

(*predictor step*)
Set $(z, \tau) \leftarrow (z^k, \tau_k)$;
Compute affine scaling direction (2.15) by solving (2.16);
Compute damping parameter (2.17);
Compute $(\bar{z}, \bar{\tau})$ from (2.18);
Set $(\bar{z}^k, \bar{\tau}_k) \leftarrow (\bar{z}, \bar{\tau})$;
(*corrector step*)
Compute centering direction (2.20) by solving (2.21);
Compute (z^+, τ_+) from (2.19);
Set $(z^{k+1}, \tau_{k+1}) \leftarrow (z^+, \tau_+),\ k \leftarrow k+1$;
until some stopping criterion is satisfied.

In Section 4 we will prove that the parameter $\bar{\theta}$ computed in the predictor step is bounded below by a quantity of the form $\chi/\sqrt{n}$, where χ is a positive quantity depending only on α and $\bar{\alpha}$. This implies that Algorithm 1 has $O(\sqrt{n}L)$-iteration complexity.

We note that the original MTY algorithm coincides with Algorithm 1 with the exception of the fact that there the neighborhoods $\mathscr{D}(\alpha)$ and $\mathscr{D}(\bar{\alpha})$, are replaced by $\mathscr{N}(0.25)$ and $\mathscr{N}(0.5)$, where $\mathscr{N}(\alpha)$ is an l_2-neighborhood of $\mathscr{C}$ of the form:

$$\mathscr{N}(\alpha) = \{(z, \tau) \in \mathscr{F}^0 \times \mathbb{R}_{++} :\ \delta_K(z, \tau) \leqslant \alpha\}, \tag{2.22}$$

and δ_K is the well known l_2-proximity measure

$$\delta_K(z, \tau) := \frac{1}{\tau}\|\tau e - xs\|_2. \tag{2.23}$$

We note that

$$\mathscr{N}(\alpha) \subset \mathscr{D}\left(\frac{\alpha}{\sqrt{1-\alpha}}\right)$$

for any $\alpha \in (0, 1)$. For example we have

$$\mathscr{N}(0.5) \subset \mathscr{D}(0.71).$$

The neighborhoods $\mathscr{D}(\alpha)$ and $\mathscr{D}(\bar{\alpha})$ considered in the present paper can be slightly larger. Indeed if we take α and $\bar{\alpha}$ such that

$$0 < \bar{\alpha} < \sqrt{\sqrt{17} - 1}, \quad \frac{\sqrt{2}\bar{\alpha}^2}{\sqrt{16 - \bar{\alpha}^4}} \leqslant \alpha < \bar{\alpha}, \tag{2.24}$$

then (2.13) is satisfied. Therefore we can take $\bar{\alpha}$ as close as we want to $\sqrt{\sqrt{17} - 1} \approx 1.767$.

3. Some technical results

In this section we present some simple inequalities that will be used in the proof of our main result. First we bound the size of the solution of the following linear system:

$$\begin{aligned} su + xv &= a, \\ Qu + Rv &= 0, \end{aligned} \tag{3.1}$$

where $z = \lceil x, s \rfloor \in \mathbb{R}^{2n}_{++}$ and $a \in \mathbb{R}^n$ are given vectors. We associate the following index sets with the solution $w = \lceil u, v \rfloor$ of this system:

$$\begin{aligned} &\mathscr{I}_+(w) = \{i \in \{1, \ldots, n\} : u_i v_i > 0\}, \\ &\mathscr{I}_-(w) = \{i \in \{1, \ldots, n\} : u_i v_i < 0\}. \end{aligned} \tag{3.2}$$

The next result is well known in the interior point literature, but we give a simple proof for the sake of completeness.

Lemma 3.1. *The solution* $w = \lceil u, v \rfloor$ *of* (3.1) *satisfies the following inequalities*:

$$\sum_{i \in \mathscr{I}_+} u_i v_i \leqslant \frac{1}{4}\|(xs)^{-1/2} a\|_2^2, \tag{3.3}$$

$$\|uv\|_\infty \leqslant \frac{1}{4}\|(xs)^{-1/2} a\|_2^2, \tag{3.4}$$

$$\|uv\|_1 \leqslant \frac{1}{2}\|(xs)^{-1/2} a\|_2^2, \tag{3.5}$$

$$\|uv\|_2 \leqslant \frac{1}{\sqrt{8}}\|(xs)^{-1/2} a\|_2^2. \tag{3.6}$$

Proof. Since our HLCP is monotone we have

$$\sum_{i \in \mathscr{I}_+} u_i v_i \geqslant -\sum_{i \in \mathscr{I}_-} u_i v_i,$$

so that

$$\|uv\|_\infty \leqslant \sum_{i \in \mathscr{I}_+} u_i v_i,$$

$$\|uv\|_1 \leqslant 2\sum_{i \in \mathscr{I}_+} u_i v_i.$$

Hence (3.3) implies (3.4) and (3.5). On the other hand (3.4) and (3.5) imply (3.6). Therefore we only have to prove (3.3). By dividing the first equation in (3.1) with $\sqrt{xs}$ we obtain:

$$Du + D^{-1}v = a/\sqrt{xs},$$

where D is the diagonal matrix given by

$$D = X^{-1/2}S^{1/2}.$$

From the above relation we obtain

$$0 < 4u_iv_i \leqslant \frac{a_i^2}{x_is_i} \quad \forall i \in \mathscr{I}_+$$

and summing up over $i \in \mathscr{I}_+$ we get (3.3). The proof is complete. □

The following relation between $\delta(z,\tau)$ and $\delta(z,\mu(z))$ can be easily proved (see [12]).

Lemma 3.2. *For any $z \in \mathbb{R}^{2n}_{++}$ and any $\tau > 0$ we have*

$$\delta^2(z,\tau) = \frac{\tau}{\mu(z)}\delta^2(z,\mu(z)) + \frac{(\tau-\mu(z))^2}{\tau\mu(z)}n.$$

Corollary 3.3. *For any $(z,\tau) \in \mathbb{R}^{2n+1}_{++}$ with $\delta^2(z,\tau) \leqslant 2\eta$ there holds*

$$\frac{1}{1+\eta/n+\sqrt{2\eta/n+\eta^2/n^2}} \leqslant \frac{\mu(z)}{\tau} \leqslant 1+\eta/n+\sqrt{2\eta/n+\eta^2/n^2}. \tag{3.7}$$

Proof. Let us denote $t = \mu/\tau$ where $\mu = \mu(z)$. From Lemma 3.2 we deduce that

$$2\eta/n \geqslant \frac{(\tau-\mu)^2}{\tau\mu} = t^{-1} - 2 + t = \frac{t^2-2t+1}{t}$$

and the corollary follows from the quadratic inequality

$$t^2 - 2(1+\eta/n)t + 1 \leqslant 0. \qquad \square$$

In the following lemma we obtain bounds on the elements of xs in terms of the proximity $\delta(z,\mu(z))$. The result follows from Lemma 3.2 in [6], but we give a simple proof for the sake of completeness.

Lemma 3.4. *If $\delta^2(z,\tau) \leqslant 2\eta$ then*

$$\frac{1}{1+\eta+\sqrt{2\eta+\eta^2}} \leqslant \frac{xs}{\tau} \leqslant 1+\eta+\sqrt{2\eta+\eta^2}.$$

Proof. Let $\mu = \mu(z)$ and let us denote by v an arbitrary entry of $\tau/(xs)$. Without loss of generality we assume $v = \tau(x_1s_1)$. Let us denote by $\check{x}$ and $\check{y}$ the $n-1$ dimensional vectors formed with the last $n-1$ components of the vectors x and s and let $\check{z} = \lceil \check{x}, \check{s} \rfloor$. We have

$$\begin{aligned} 2\eta \geqslant \delta^2(z,\tau) &= v + \sum_{i=2}^{n}\frac{\tau}{x_is_i} - 2(n-1) - 2 + \frac{1}{v} \\ &\quad + \sum_{i=2}^{n}\frac{x_is_i}{\tau} \geqslant v - 2 + \frac{1}{v} + \delta(\check{z},\tau) \\ &\geqslant v - 2 + \frac{1}{v}. \end{aligned}$$

It follows that v satisfies the following quadratic inequality:

$$v^2 - 2(1+\eta)nv + 1 \leqslant 0.$$

Hence v must lie between the roots of the corresponding quadratic equation, i.e.,

$$1+\eta-\sqrt{2\eta+\eta^2} \leqslant v \leqslant 1+\eta+\sqrt{2\eta+\eta^2},$$

and the proof is complete by noticing that the lower bound above is equal to the inverse of the upper bound. □

4. Polynomial complexity

In this section we return to Algorithm 1 and prove that it is globally convergent for any choice of parameters α and $\overline{\alpha}$ satisfying (2.13) (which is equivalent (2.24)). More precisely we will show that at each iteration the steplength $\overline{\theta}$ computed by Algorithm 1 is bounded below by $\chi/\sqrt{n}$ where χ is a constant depending only on α and $\overline{\alpha}$. By a standard argument this implies that Algorithm 1 has $O(\sqrt{n}L)$ iteration complexity.

First we obtain a convenient representation for $\delta^2(z(\theta))$, where $z(\theta)$ is defined by (2.14) and (2.15). By denoting

$$x(\theta) = x + \theta u, \quad s(\theta) = s + \theta v,$$
$$\tau(\theta) = (1 - \theta(\tau)) \tag{4.1}$$

and

$$\varphi = \frac{\theta^2}{1-\theta}, \tag{4.2}$$

we can write

$$\begin{aligned}\delta^2(z(\theta),\tau(\theta)) &= \sum_{i=1}^{n}\frac{\tau(\theta)}{x_i(\theta)s_i(\theta)} - 2n + \sum_{i=1}^{n}\frac{x_i(\theta)s_i(\theta)}{\tau(\theta)} \\ &= \sum_{i=1}^{n}\frac{(1-\theta)\tau}{(1-\theta)x_is_i+\theta^2u_iv_i} - 2n \\ &\quad + \sum_{i=1}^{n}\frac{(1-\theta)x_is_i+\theta^2u_iv_i}{(1-\theta)\tau} \\ &= \sum_{i=1}^{n}\frac{\tau}{x_is_i+\varphi u_iv_i} - 2n \\ &\quad + \sum_{i=1}^{n}\frac{x_is_i+\varphi u_iv_i}{\tau} \\ &= \sum_{i=1}^{n}\frac{\tau}{x_is_i} - 2n + \sum_{i=1}^{n}\frac{x_is_i}{\tau} \\ &\quad - \sum_{i=1}^{n}\frac{\tau\varphi u_iv_i}{x_is_i(x_is_i+\varphi u_iv_i)} + \sum_{i=1}^{n}\frac{\varphi u_iv_i}{\tau}.\end{aligned}$$

It follows that

$$\delta^2(z(\theta),\tau(\theta)) = \delta^2(z,\tau) + f(\varphi), \tag{4.3}$$

where

$$f(\varphi) = \varphi e^{\mathrm{T}}q - \sum_{i=1}^{n}\frac{\varphi q_i}{p_i(p_i+\varphi q_i)}, \tag{4.4}$$

and

$$p = xs/\tau, \quad q = uv/\tau. \tag{4.5}$$

We take as domain of definition of f the interval

$$\Phi = [0,\hat{\varphi}), \quad \hat{\varphi} = \min\left\{\frac{p_i}{|q_i|} : i \in \mathscr{I}_-\right\}. \tag{4.6}$$

It is easily verified that

$$f''(\varphi) = \sum_{i=1}^{n}\frac{2q_i^2}{(p_i+\varphi q_i)^3}.$$

It follows that f is convex on Φ. Since

$$f(0) = 0, \quad \lim_{\varphi\to\hat{\varphi}} f(\varphi) = \infty$$

we deduce that the equation

$$f(\varphi) = \overline{\alpha}^2 - \alpha^2 \tag{4.7}$$

has a unique solution $\overline{\varphi}$ in Φ. Using the relation (4.2) we get the following result.

Lemma 4.1. *The steplength $\overline{\theta}$ defined in* (2.17) *is given by*

$$\overline{\theta} = \frac{2\overline{\varphi}}{\overline{\varphi} + \sqrt{\overline{\varphi}^2 + 4\overline{\varphi}}},$$

where $\overline{\varphi}$ is the unique solution of the Eq. (4.7) *in the interval Φ.*

Proof. From (4.3) it follows that

$$\delta(z(\theta),(1-\theta\tau)) \leqslant \overline{\alpha}, \forall\theta \in (0,\overline{\theta}], \tag{4.8}$$

where $\overline{\theta}$ is defined in our lemma. In order to prove that this is the same as the original definition (2.17) we still have to prove that $z(\theta)$ is strictly feasible for any $\theta \in (0,\overline{\theta}]$. It is easily verified that $Qx(\theta) + Rs(\theta) = b$. We are left to prove that $x(\theta) > 0$, $s(\theta) > 0$, $\forall\theta \in (0,\overline{\theta}]$. Assume by contradiction that there is $\theta_0 \in (0,\overline{\theta}]$ such that $x(\theta_0)s(\theta_0) = 0$. In this case $\lim_{\theta\to\theta_0}\delta(z(\theta),(1-\theta)\tau) = \infty$ which contradicts (4.8). □

We note that $\overline{\varphi}$ can be computed to any given accuracy by applying Newton's method to (4.7). For a hybrid procedure that is guaranteed to produce an ϵ-approximation of $\overline{\varphi}$ with at most $\mathrm{O}(\log\log\epsilon)$ function evaluations see [11]. In what follows we will obtain a lower bound $\hat{\varphi}$ for $\overline{\varphi}$. For our complexity analysis we need only a rough estimate. Therefore we will restrict the domain of f to

$$0 \leqslant \varphi \leqslant \varphi_1 := \frac{3\min p}{8\|q\|_\infty}. \tag{4.9}$$

Let us note that since $(z,\tau) \in \mathscr{D}(\alpha)$ we can apply Lemma 3.4 to obtain the following bounds on the elements of p

$$\frac{1}{m(\alpha)} \leqslant \min p \leqslant \|p\|_\infty \leqslant m(\alpha), \tag{4.10}$$

where

$$m(\alpha) = 1 + \alpha^2/2 + \sqrt{\alpha^2 + \alpha^4/4}. \tag{4.11}$$

We will use this result to find a lower bound of $\overline{\varphi}$ in terms of the following quantity:

$$h(\alpha) = \frac{8(m^2(\alpha)+1)(8m^2(\alpha)-5)}{}0.15 + 40m^2(\alpha). \tag{4.12}$$

Proposition 4.2. *For any φ satisfying* (4.9) *we have*

$$f(\varphi) \leqslant \varphi h(\alpha) \sum_{i\in\mathscr{I}_+} q_i,$$

where $h(\alpha)$ is given by (4.12).

Proof. Using the definition of f we can write

$$\begin{aligned} f(\varphi) &= \sum_{i=1}^{n} \varphi q_i \left(1 - \frac{1}{p_i(p_i + \varphi q_i)}\right) \\ &= \sum_{i\in\mathscr{I}_+} \varphi q_i \left(1 - \frac{1}{p_i(p_i + \varphi q_i)}\right) \\ &\quad \times \sum_{i\in\mathscr{I}_-} \varphi q_i \left(1 - \frac{1}{p_i(p_i + \varphi q_i)}\right) \\ &\leqslant \sum_{i\in\mathscr{I}_+} \varphi q_i \left(1 - \frac{1}{\|p\|_\infty(\|p\|_\infty + \varphi\|q\|_\infty)}\right) \\ &\quad - \sum_{i\in\mathscr{I}_-} \varphi |q_i| \left(1 - \frac{1}{\min p(\min p - \varphi\|q\|_\infty)}\right). \end{aligned}$$

Since our HCLP is monotone we have

$$\sum_{i\in\mathscr{I}_+} q_i \geqslant \sum_{i\in\mathscr{I}_-} |q_i|$$

so that by using (4.9) we obtain

$$\begin{aligned} f(\varphi) &\leqslant \sum_{i\in\mathscr{I}_+} \varphi q_i \left(\frac{1}{\min p(\min p - \varphi\|q\|_\infty)} \right. \\ &\quad \left. - \frac{1}{\|p\|_\infty(\|p\|_\infty + \varphi\|q\|_\infty)}\right) \leqslant \varphi \left[\frac{1}{5/8(\min p)^2} \right. \\ &\quad \left. - \frac{1}{\|p\|_\infty(\|p\|_\infty + 3/8\min p)}\right] \sum_{i\in\mathscr{I}_+} q_i, \end{aligned}$$

and the desired bound follows by applying (4.10). □

Corollary 4.3. *The unique solution $\overline{\varphi}$ of Eq.* (4.7) *in the interval Φ satisfies the inequality*

$$\overline{\varphi} \geqslant \hat{\varphi} := \min\left\{\frac{\overline{\alpha}^2 - \alpha^2}{h(\alpha)\sum_{i\in\mathscr{I}_+} q_i}, \frac{3}{8m(\alpha)\|q\|_\infty}\right\}.$$

In order to obtain a lower bound for $\overline{\theta}$ we need the following result, which follows immediately from (4.2) and the above corollary.

Corollary 4.4. *The steplength $\overline{\theta}$ computed in the predictor step of Algorithm 1 satisfies the inequality*

$$\overline{\theta} \geqslant g(\hat{\varphi})\sqrt{\hat{\varphi}},$$

where

$$g(\varphi) = \frac{\sqrt{\varphi + 4} - \sqrt{\varphi}}{2}$$

is a decreasing function on $[0, \infty)$.

The above results will be used to prove that $\overline{\theta} \geqslant \chi_n/\sqrt{n}$, where

$$\chi_n := 2\sqrt{\frac{\gamma}{c_n}} g\left(\frac{4\gamma}{nc_n}\right), \tag{4.13}$$

$$\gamma := \min\left\{\frac{\overline{\alpha}^2 - \alpha^2}{h(\alpha)}, \frac{3}{8m(\alpha)}\right\}, \tag{4.14}$$

$$c_n := \begin{cases} m(\alpha/n), & \text{if HLCP is monotone,} \\ 1 & \text{if HLCP is skew-symmetric.} \end{cases} \tag{4.15}$$

We note that sequence $\{\chi_n\}$ is increasing and

$$\lim_{n\to\infty} \chi_n = 2\sqrt{\gamma}. \tag{4.16}$$

We are now ready to prove the main result of this section.

Theorem 4.5. *Algorithm* 1 *is well defined for any parameters α and $\overline{\alpha}$ satisfying* (2.13) *or* (2.24) *and the following relations hold for $k = 0, 1, \ldots$:*

(i) $(z^k, \tau_k) \in \mathscr{D}(\alpha)$, $(\overline{z}^k, \tau_{k+1}) \in \mathscr{D}(\overline{\alpha})$;
(ii) $\tau_{k+1} = (1 - \theta_k)\tau_k$;
(iii) $1 \leqslant \frac{\mu(z^k)}{\tau_k} \leqslant c_n$;
(iv) $\theta_k \geqslant \frac{\chi_n}{\sqrt{n}}$, $\chi_1 \leqslant \chi_2 \leqslant \chi_3 \leqslant \cdots \leqslant \chi_n \to 2\sqrt{\gamma}$.

Proof. We drop the superscript k, and use the notation introduced in this section and in Algorithm 1. We assume that $(z, \tau) \in \mathscr{D}(\alpha)$.

(i) From Lemma 4.1 it follows that $(\overline{z}, \overline{\tau}) \in \mathscr{D}(\overline{\alpha})$. Let us denote

$$\overline{q} = \overline{u}\overline{v} = [\overline{u}_1\overline{v}_1, \ldots, \overline{u}_n\overline{v}_n]^{\mathrm{T}}, \quad \overline{\mathscr{I}}_{+-} = \mathscr{I}_{+-}(\overline{w}),$$

where $\mathscr{I}_{+-}$ are given by (3.2) and $\overline{w} = [\overline{u}, \overline{v}]$. We will prove now that the corrector step produces $(z^+, \tau_+) \in \mathscr{D}(\alpha)$. Using (2.13) and (2.21), $\tau_+ = \overline{\tau}$

and the monotony of our HLCP (which implies $\sum_{i\in\overline{\mathscr{I}}_+}\overline{q}_i \geqslant \sum_{i\in\overline{\mathscr{I}}_-}|\overline{q}_i|$) we can write

$$\begin{aligned}
\delta^2(z^+,\tau_+) &= \sum_{i=1}^n \frac{\overline{\tau}}{x_i^+s_i^+} - 2n + \sum_{i=1}^n \frac{x_i^+s_i^+}{\overline{\tau}} \\
&= \sum_{i=1}^n \frac{\overline{\tau}}{\overline{\tau}+\overline{u}_i\overline{v}_i} - 2n + \sum_{i=1}^n \frac{\overline{\tau}+\overline{u}_i\overline{v}_i}{\overline{\tau}} \\
&= \sum_{i=1}^n \left(\frac{\overline{\tau}}{\overline{\tau}+\overline{u}_i\overline{v}_i} - 1\right) + \sum_{i=1}^n \frac{\overline{u}_i\overline{v}_i}{\overline{\tau}} \\
&= \sum_{i=1}^n \frac{-\overline{u}_i\overline{v}_i}{\overline{\tau}+\overline{u}_i\overline{v}_i} + \sum_{i=1}^n \frac{\overline{u}_i\overline{v}_i}{\overline{\tau}} \\
&= \sum_{i=1}^n \frac{(\overline{u}_i\overline{v}_i)^2}{\overline{\tau}(\overline{\tau}+\overline{u}_i\overline{v}_i)} = \sum_{i=1}^n \frac{\overline{q}_i^2}{1+\overline{q}_i} \\
&= \sum_{i\in\overline{\mathscr{I}}_+} \frac{\overline{q}_i^2}{1+\overline{q}_i} + \sum_{i\in\overline{\mathscr{I}}_-} \frac{\overline{q}_i^2}{1-|\overline{q}_i|} \\
&\leqslant \sum_{i\in\overline{\mathscr{I}}_+} \frac{\|\overline{q}\|_\infty\overline{q}_i}{1+\|\overline{q}\|_\infty} + \sum_{i\in\overline{\mathscr{I}}_-} \frac{\|\overline{q}\|_\infty|\overline{q}_i|}{1-\|\overline{q}\|_\infty} \\
&\leqslant \frac{2\|\overline{q}\|_\infty}{1-\|\overline{q}\|_\infty^2}\sum_{i\in\overline{\mathscr{I}}_+}\overline{q}_i.
\end{aligned}$$

Applying Lemma 3.1 with $\overline{w}$ and $\overline{\tau}-\overline{x}\overline{s}$ instead of w and a and using (2.24) we obtain

$$\delta^2(z^+,\tau_+) \leqslant \frac{2\delta^4(\overline{z},\overline{\tau})}{16-\delta^4(\overline{z},\overline{\tau})} \leqslant \frac{2\overline{\alpha}^4}{16-\overline{\alpha}^4} \leqslant \alpha^2.$$

An argument similar to the one used in the proof of Lemma 4.1 shows that z^+ is strictly feasible. Hence $(z^+,\tau_+)\in\mathscr{D}(\alpha)$.

(ii) Is trivially satisfied since Algorithm 1 computes $\tau_+ = \overline{\tau} = (1-\overline{\theta})\tau$.

(iii) In the skew-symmetric case we have $u^{\mathrm{T}}v = \overline{u}^{\mathrm{T}}\overline{v} = 0$, so that by using (2.16) and (2.21) we have

$$\mu(\overline{z}) = (1-\overline{\theta})\mu(z) + \overline{\theta}^2u^{\mathrm{T}}v/n = (1-\overline{\theta})\mu(z),$$

$$\mu(z^+) = \mu(\overline{z}) + \overline{u}^{\mathrm{T}}\overline{v}/n = \mu(\overline{z}).$$

In the monotone case the result follows from Corollary 3.3.

(iv) Lemma 3.1 gives

$$\sum_{i\in\overline{\mathscr{I}}_+} u_iv_i \leqslant \frac{1}{4}\|\sqrt{xs}\|_2^2 = \frac{n\mu(z)}{4},$$

and by using (iii) we deduce that

$$\sum_{i\in\overline{\mathscr{I}}_+} q_i \leqslant \frac{nc_n}{4}.$$

Similarly we can prove that

$$\|q\|_\infty \leqslant \frac{nc_n}{4},$$

and by using Corollary 4.3 we obtain

$$\hat{\varphi} \geqslant \frac{4}{nc_n}\min\left\{\frac{\overline{\alpha}^2-\alpha^2}{h(\alpha)}, \frac{3}{8m(\alpha)}\right\} = \frac{4\gamma}{nc_n}.$$

The desired result follows by applying Corollary 4.4. □

To have an idea about the quality of the bound obtained in point (iv) of the above theorem let us take for example

$$\alpha = \frac{1}{4}, \quad \overline{\alpha} = \frac{5}{6}. \tag{4.17}$$

It is easily seen that conditions (2.13) or (2.24) are satisfied and for this values we have

$$\gamma = \frac{12}{33+\sqrt{65}}. \tag{4.18}$$

This shows that for large values of n we will have

$$\chi_n \approx 2\sqrt{\gamma} > 1.08. \tag{4.19}$$

For $n=2$ we obtain $\chi_2 > 0.74$ in the skew-symmetric case and $\chi_2 > 0.71$ in the monotone case. For larger values of n we obtain larger bounds. Thus it is easily checked that

$$\chi_n > 1.05, \quad \forall n \geqslant 400,$$

and

$$\chi_n > 1.075, \quad \forall n \geqslant 9000,$$

both in the skew-symmetric case and the monotone case. These are the largest guaranteed lower bounds for the stepsize along the affine scaling direction in the interior point literature that we know of.

5. Quadratic convergence

In this section we show that if the HLCP has a strictly complementary solution then for large k

the steplength θ_k is much larger than indicated by the lower bounds obtained in Theorem 4.5. In fact we will prove that $\lim_{k\to\infty}\theta_k = 1$, or more precisely that

$$\theta_k = 1 - \mathrm{O}(\tau_k), \tag{5.1}$$

which implies that the sequences $\{\tau_k\}$ and $\{\mu(z^k)\}$ are quadratically convergent to zero in the sense that

$$\tau_{k+1} = \mathrm{O}(\tau_k^2), \quad \mu(z^{k+1}) = \mathrm{O}(\mu(z^k)^2). \tag{5.2}$$

Let us denote by

$$\mathscr{F}^{\#} := \{z^* = \lceil x^*, s^* \rfloor \in \mathscr{F}^* : x^* + s^* > 0\}, \tag{5.3}$$

the set of all strictly complementary solutions of our HLCP. In what follows we will assume that $\mathscr{F}^{\#}$ is nonempty. As mentioned in the introduction this assumption is not restrictive, since according to [10] strict complementarity is a necessary condition for superlinear convergence for a large class of interior point methods.

The following result was first proved for the standard monotone LCP in [15]. Its extension for the HCLP is immediate, using for example the equivalences proved in [1] (see also [2]).

Lemma 5.1. *If $\mathscr{F}^{\#} \neq \emptyset$ then the solution $w = \lceil u, v \rfloor$ of* (2.16) *satisfies*

$$|u_i v_i| = \mathrm{O}(\mu^2), \quad \forall i \in \{1, 2, \ldots, n\},$$

where $\mu = \mu(z)$ is given by (1.4).

Theorem 5.2. *If the HLCP has a strictly complementary solution, then the sequences $\{z^k\}$ and $\{\bar{z}^k\}$ generated by Algorithm 1 converge to the analytic center of the solution set $\mathscr{F}^*$. Moreover the sequences $\{\tau_k\}$ and $\{\mu(z^k)\}$ converge quadratically to zero in the sense that* (5.2) *is satisfied.*

Proof. We use the same notation as in the proof of Theorem 4.5. In particular, we will drop whenever possible the subscripts and superscripts k. The convergence of the sequences $\{z^k\}$ and $\{\bar{z}^k\}$ follows from the results of [2]. Lemma 5.1 and point (iii) of Theorem 4.5 imply $q_k = \mathrm{O}(\tau)$, so that according to Corollary 4.3 we have $\hat{\varphi} = \Omega(1/\tau)$. Using Corollary 4.4 we deduce that

$$0 \leqslant 1 - \overline{\theta} \leqslant 1 - \frac{2\sqrt{\hat{\varphi}}}{\sqrt{\hat{\varphi}+4} + \sqrt{\hat{\varphi}}} = \frac{4}{(\sqrt{\hat{\varphi}+4} + \sqrt{\hat{\varphi}})^2} = \mathrm{O}(\tau).$$

According to point (iii) of Theorem 4.5 this implies $\tau_{k+1} = \mathrm{O}(\tau_k^2)$, which in turn implies $\mu(z^{k+1}) = \mathrm{O}(\mu(z^k)^2)$, by virtue of point (ii) of Theorem 4.5. The proof is complete. □

Acknowledgements

This work was supported in part by NSF, Grant DMS-9996154.

References

[1] M. Anitescu, G. Lesaja, F.A. Potra, An infeasible-interior-point predictor–corrector algorithm for the P_*-geometric LCP, Applied Mathematics and Optimization 36 (2) (1997) 203–228.
[2] J.F. Bonnans, C.C. Gonzaga, Convergence of interior point algorithms for the monotone linear complementarity problem, Mathematics of Operations Research 21 (1996) 1–25.
[3] R.W. Cottle, J.-S. Pang, R.E. Stone, The Linear Complementarity Problem, Academic Press, Boston, MA, 1992.
[4] E. Hairer, G. Wanner, Solving ordinary differential equations. II, Stiff and differential-algebraic problems, second ed., Springer-Verlag, Berlin, 1996.
[5] B. Jansen, C. Roos, T. Terlaky, J.-P. Vial, Primal–dual algorithms for linear programming based on the logarithmic barrier method, Journal of Optimization Theory and Applications 83 (1) (1994) 1–26.
[6] B. Jansen, C. Roos, T. Terlaky, J.-Ph. Vial, Primal–dual target-following algorithms for linear programming, Annals of Operations Research 62 (1996) 197–231, Interior point methods in mathematical programming.
[7] J. Ji, F.A. Potra, S. Huang, A predictor–corrector method for linear complementarity problems with polynomial complexity and superlinear convergence, Journal of Optimization Theory and Applications 84 (1) (1995) 187–199.
[8] M. Kojima, N. Megiddo, T. Noma, A. Yoshise, A unified approach to interior point algorithms for linear complementarity problems, Lecture Notes in Computer Science 538 (1991).
[9] S. Mizuno, M.J. Todd, Y. Ye, On adaptive-step primal-dual interior-point algorithms for linear programming, Mathematics of Operations Research 18 (4) (1993) 964–981.
[10] R.D.C. Monteiro, S.J. Wright, Local convergence of interior-point algorithms for degenerate monotone LCP,

Computational Optimization and Applications 3 (1994) 131–155.
[11] F.A. Potra, Efficient hybrid algorithms for finding zeros of convex functions, Journal of Complexity 10 (1994) 199–215.
[12] F.A. Potra, A path-following method for linear complementarity problems based on the affine invariant kantorovich theorem. ZIB-Report 00-30, Konrad-Zuse-Zentrum, Berlin, August 2000.
[13] W.C. Rheinboldt, Numerical Analysis of Parametrized Nonlinear Equations, John Wiley and Sons Inc., New York, 1986.
[14] C. Roos, J.-P. Vial, T. Terlaky, Theory and algorithms for linear optimization: An Interior Point Approach, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley and Sons, New York, 1997.
[15] Y. Ye, K. Anstreicher, On quadratic and $O(\sqrt{n}L)$ convergence of predictor–corrector algorithm for LCP, Mathematical Programming 62 (3) (1993) 537–551.
[16] Y. Ye, O. Güler, R.A. Tapia, Y. Zhang, A quadratically convergent $O\sqrt{n}L)$ iteration algorithm for linear programming, Mathematical Programming 59 (2) (1993) 151–162.