

# A superlinearly convergent predictor-corrector method for degenerate LCP in a wide neighborhood of the central path with $O(\sqrt{n}L)$ -iteration complexity \*

Florian A. Potra<sup>†</sup>

February 14, 2006

**Abstract:** An interior point method for monotone linear complementarity problems acting in a wide neighborhood of the central path is presented. The method has  $O(\sqrt{n}L)$ -iteration complexity and is superlinearly convergent even when the problem does not possess a strictly complementary solution.

**Keywords:** linear complementarity problem, interior-point algorithm, large neighbourhood, superlinear convergence

**Mathematics subject classification:** 49M15, 65K05, 90C33

**NOTE:** in this report we correct some errors from the paper with the same title published in *Mathematical Programming, Ser. A*, **100**, 2(2004), 317–337.

## 1 Introduction

A large body of theoretical and experimental work done over the past decade has established primal-dual interior point methods as the most efficient interior point methods for linear programming (LP), quadratic programming (QP), and linear complementarity problems (LCP). An excellent monograph [17] dedicated to this field describes the most relevant theoretical results obtained prior to 1997, and contains insightful comments on software development. Predictor-corrector methods play a special role among primal-dual interior point methods. The classical representative of such methods is the Mizuno-Todd-Ye algorithm (MTY) for LP [10]. This method operates between two  $l_2$  neighborhoods of

---

\*Work supported by the National Science Foundation under Grant No. 0139701 .

<sup>†</sup>University of Maryland Baltimore County, potra@math.umbc.edu

the central path,  $\mathcal{N}_2(.25) \subset \mathcal{N}_2(.5)$ . At a typical iteration of MTY one is given a point  $z \in \mathcal{N}_2(.25)$  and one "predicts" a point  $\bar{z} \in \mathcal{N}_2(.5)$  by carefully choosing a steplength along the affine-scaling direction at  $z$ . It is shown that the predictor reduces the primal dual gap by a factor of at least  $(1 - \hat{\chi}/\sqrt{n})$ , where  $\hat{\chi} = 1/\sqrt[4]{8} \approx .5946$ . The corrector produces a point  $z^+ \in \mathcal{N}_2(.25)$  by taking a unit steplength along the centering direction at  $\bar{z}$ . It is shown that while improving centering, the corrector maintains the same duality gap. Thus the predictor-corrector associates with each point  $z$  from the neighborhood  $\mathcal{N}_2(.25)$  of the central path, a point  $z^+$  in the same neighborhood such that the duality gap at  $z^+$  is smaller by a factor of at least  $(1 - \hat{\chi}/\sqrt{n})$  than at  $z$ . It follows that the corresponding iterative procedure has  $O(\sqrt{n}L)$ -iteration complexity. The same complexity is achieved by many primal-dual algorithms developed prior to MTY. However, it turns out that the duality gap of the sequence generated by MTY converges faster than indicated by the above complexity result, as shown by Ye et al. [20] who proved that the duality gap converges quadratically to zero. This result answered a question that had been open for a while in the interior-point literature : 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 [8], 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 (i.e., when the LCP is nondegenerate) and the iteration sequence converges. From [3] it follows that the latter assumption always holds. Subsequently Ye and Anstreicher [19] proved that MTY converges quadratically assuming only that the LCP is nondegenerate. The nondegeneracy assumption is not restrictive, since according to [11] a large class of interior point methods, which contains MTY, can have only linear convergence if this assumption is violated. However as shown in [15, 16] it is possible to obtain superlinear convergence even in the degenerate case by using higher order predictors.

In a recent paper [13], we have analyzed a predictor-corrector method for LCP where the  $l_2$ -neighborhoods are replaced by  $\delta$ -neighborhoods, where  $\delta$  is the proximity measure introduced in [7] (see also [14]), and proved that in this case the constant  $\hat{\chi}$  determining the guaranteed decrease of the primal dual gap is greater than 1. However the neighborhoods considered in that paper are still "small". Or, as it is well known, the practical performance of primal-dual interior point methods is better in "wide" neighborhoods on the central path. Paradoxically, the iteration complexity of the methods that use wide neighborhoods is worse than the complexity of the corresponding methods for small neighborhoods. The most commonly used wide neighborhoods are based on the  $\delta_\infty$  or the  $\delta_\infty^-$  proximity measures (see the definitions in the next section). The best iteration complexity achieved by interior point methods that act in these neighborhoods and use only first

order derivatives is  $O(nL)$  (see the monographs [14, 17, 18]). By using higher order derivatives the iteration complexity can be reduced arbitrarily closed to  $O(\sqrt{nL})$ , as shown in [6], or even to  $O(\sqrt{nL})$  [21].

Predictor-corrector methods of MTY type are more difficult to develop and analyze in wide neighborhoods of the central path. This is due to the fact that correctors based on the first order centering direction (first order correctors) are rather inefficient in such neighborhoods. It is known for example that one needs  $O(n)$  centering steps in order to reduce the  $\delta_\infty$  proximity measure by a factor of .5 (see [2]).

Using a very elegant analysis, Gonzaga [5] has managed to show that a predictor-corrector method based on a  $\delta_\infty$  neighborhood of the central path has  $O(nL)$ -iteration complexity, the same as the best complexity achieved by any other known interior-point method in this neighborhood that uses only first order derivatives. However his predictor-corrector method does no longer have the simple structure of MTY, where a predictor is followed by just one corrector. In Gonzaga's algorithm a predictor is followed by an a priori unknown number of correctors. In fact the above mentioned complexity result is proved by showing that the total number of correctors is at most  $O(nL)$ . The structure of Gonzaga's algorithm makes it very difficult to analyze the asymptotic convergence properties of the duality gap. No superlinear convergence results have been obtained so far for this method.

In the present paper we propose a predictor-corrector method based on a  $\delta_\infty^-$  neighborhood of the central path that has the same structure as the MTY algorithm: each predictor is followed by exactly one corrector. We prove that this algorithm has  $O(nL)$  iteration complexity under general conditions, and quadratic convergence of the primal-dual gap under the assumption that the LCP is nondegenerate.

By using a higher order predictor we reduce the iteration complexity and prove superlinear convergence even in the degenerate case. More precisely, by using a predictor of order  $m$  we obtain algorithm with  $O(n^{1/2} + (n+1)^{1/(m+1)}L)$  iteration complexity. If  $m > 1$ , then the duality gap is superlinearly convergent under general conditions. The Q-order of convergence of the primal-dual gap is  $m+1$  in the nondegenerate case, and  $(m+1)/2$  in the degenerate case. Here  $m$  can be a constant, or can depend on the problem dimension. Since  $\lim_{n \rightarrow \infty} n^{(1/n^\omega)} = 1$  for any constant  $\omega > 0$  (in fact, as shown in Section 4,  $n^{(1/n^\omega)} \leq e^{1/(\omega e)}$ , for all  $n$ ), it follows that by taking  $m = \lceil (n+1)^\omega - 1 \rceil$ , for some  $\omega \in (0, 1)$ , the iteration complexity of our predictor-corrector algorithm reduces to  $O(\sqrt{nL})$ , the same as the best complexity results for small neighborhoods (actually the best iteration-complexity known so far for any interior point method for LP).

We mention that the numerical implementation of a predictor of order  $m$  involves a matrix factorization, that costs  $O(n^3)$  arithmetic operations and  $m+1$  backsolve, each backsolve requiring  $O(n^2)$  arithmetic operations. Therefore the cost of a predictor of order

$\lceil n^\omega \rceil$  is still  $O(n^3)$  arithmetic operations, and it is dominated by the cost of the matrix factorization. A moderate order for the predictor suffices both for theoretical and practical considerations. For example if we take  $\omega = 0.1$  then the values of  $m = \lceil (n+1)^\omega - 1 \rceil$ , corresponding to  $n = 10^6$ ,  $n = 10^7$ ,  $n = 10^8$ , and  $n = 10^9$  are 3, 5, 6, and 7 respectively.

Higher order interior point methods in the wide neighborhood of the central path have been considered before by Monteiro, Adler and Resende [12], Hung and Ye [6], and Zhao [21]. The algorithms presented in those papers are not predictor-corrector algorithms and no superlinear convergence results have been proved for any of them. In fact it is very unlikely that any of those algorithms is superlinearly convergent.

**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  $\kappa$ ,  $\lceil \kappa \rceil$  denotes the smallest integer greater or equal to  $\kappa$ . 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  $u_i v_i$ ,  $u_i/v_i$ , etc. This notation is consistent as long as component-wise operations always have precedence in relation to matrix operations. Note that  $uv \equiv Uv$  and 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  $v \geq 3$  means that  $v_i \geq 3$ ,  $i = 1, \dots, n$ . Traditionally this is written as  $v \geq 3e$ . 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 order relationships between functions. The 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  $k \in \mathbb{N}$ ,  $\|x^k\| \leq K\tau_k$ . Similarly, if  $x^k > 0$ ,  $x^k = \Omega(\tau_k)$  means that  $(x^k)^{-1} = O(1/\tau_k)$ . If we have both  $x^k = O(\tau_k)$  and  $x^k = \Omega(\tau_k)$ , we write  $x^k = \Theta(\tau_k)$ .

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  $\lceil x, s \rceil$ , i.e.,

$$z = \lceil x, s \rceil = \begin{bmatrix} x \\ s \end{bmatrix} = \begin{bmatrix} x^T, s^T \end{bmatrix}^T. \quad (1.1)$$

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

$$\mu(z) = \frac{x^T s}{n}. \quad (1.2)$$

## 2 The horizontal linear complementarity problem

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

$$\begin{aligned} xs &= 0 \\ Qx + Rs &= b \\ x, s &\geq 0. \end{aligned} \quad (2.1)$$

The standard (monotone) linear complementarity problem (LCP) is obtained by taking  $R = -I$ , and  $Q$  positive semidefinite. There are other formulations of the linear complementarity problems 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 linear programming problem (LP), and the quadratic programming problem (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:

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

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 [4]). If the HLCP is a reformulation of an LP then the following stronger condition holds

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

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. Let us denote the set of all feasible points of HLCP by

$$\mathcal{F} = \{z = \lceil x, s \rceil \in \mathbb{R}_+^{2n} : Qx + Rs = b\},$$

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

$$\mathcal{F}^* = \{z^* = [x^*, s^*] \in \mathcal{F} : x^*s^* = 0\}.$$

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

$$\mathcal{F}^0 = \mathcal{F} \cap \mathbb{R}_{++}^{2n},$$

will be called the set of strictly feasible points, or the set of interior points. It is known (see [9]) that if  $\mathcal{F}^0$  is nonempty, then, for any parameter  $\tau > 0$  the nonlinear system,

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

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

$$z = [x, s]$$

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

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

we can write

$$\mathcal{C} = \{z \in \mathbb{R}_{++}^{2n} : F_\tau(z) = 0, \tau > 0\}.$$

If  $F_\tau(z) = 0$ , then necessarily  $\tau = \mu(z)$ , where  $\mu(z)$  is given by (1.2). The distance of a point  $z \in \mathcal{F}$  to the central path can be quantified by different proximity measures. The following proximity measures have been extensively used in the interior point literature:

$$\delta_2(z) := \left\| \frac{xs}{\mu(z)} - e \right\|_2, \quad \delta_\infty(z) := \left\| \frac{xs}{\mu(z)} - e \right\|_\infty, \quad \delta_\infty^-(z) := \left\| \left[ \frac{xs}{\mu(z)} - e \right]^- \right\|_\infty,$$

where  $[v]^-$  denotes the negative part of the vector  $v$ , i.e.  $[v]^- = \max\{-v, 0\}$ .

By using the above proximity measures we can define the following neighborhoods of central path

$$\begin{aligned} \mathcal{N}_2(\alpha) &= \{z \in \mathcal{F}^0 : \delta_2(z) \leq \alpha\}, \\ \mathcal{N}_\infty(\alpha) &= \{z \in \mathcal{F}^0 : \delta_\infty(z) \leq \alpha\}, \\ \mathcal{N}_\infty^-(\alpha) &= \{z \in \mathcal{F}^0 : \delta_\infty^-(z) \leq \alpha\}, \end{aligned}$$

where  $0 < \alpha < 1$  is a given parameter. We have

$$\mathcal{N}_2(\alpha) \subset \mathcal{N}_\infty(\alpha) \subset \mathcal{N}_\infty^-(\alpha), \quad \text{and} \quad \lim_{\alpha \uparrow 1} \mathcal{N}_\infty^-(\alpha) = \mathcal{F}.$$

Although the last relation above is well known, we give a short proof for the sake of completeness. For any  $z \in \mathcal{F}$  choose some  $z^0 \in \mathcal{F}^0$  and define  $z^0(\tau) = (1 - \tau)z^0 + \tau z$ . Since  $z^0(\tau) \in \mathcal{N}_\infty^-(\delta_\infty^-(z^0(\tau)))$  for any  $\tau \in (0, 1)$ , and  $\lim_{\tau \uparrow 1} z^0(\tau) = z$ , we deduce that  $\mathcal{F} = \bigcup_{\alpha \in (0, 1)} \mathcal{N}_\infty^-(\alpha) =: \lim_{\alpha \uparrow 1} \mathcal{N}_\infty^-(\alpha)$ . In this paper we will work with  $\mathcal{N}_\infty^-(\alpha)$ . We note that this neighborhood can be written under the form:

$$\mathcal{N}_\infty^-(\alpha) = \mathcal{D}(1 - \alpha), \text{ where } \mathcal{D}(\beta) = \{z \in \mathcal{F}^0 : xs \geq \beta\mu(z)\}.$$

### 3 The first order predictor-corrector method

Let  $\beta$  be a given parameter in the interval  $(0, 1)$ . At a typical iteration of our predictor-corrector algorithm a point  $z = [x, s] \in \mathcal{D}(\beta)$  has already been computed. Let  $z(\theta)$  be defined via a damped Newton step of the form

$$z(\theta) = z + \theta w,$$

where

$$w = [u, v] = -F_0'(z)^{-1}F_0(z) \quad (3.1)$$

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

$$\begin{aligned} su + xv &= -xs \\ Qu + Rv &= 0. \end{aligned} \quad (3.2)$$

We want to move along the affine scaling direction  $w$  as far as possible while preserving the condition  $z(\theta) \in \mathcal{D}((1 - \gamma)\beta)$ , where

$$\gamma := \frac{1 - \beta}{n + 1}. \quad (3.3)$$

For any  $\theta \geq 0$  we define

$$x(\theta) = x + \theta u, \quad s(\theta) = s + \theta v, \quad \mu = \mu(z), \quad \mu(\theta) = \mu(z(\theta)) = x(\theta)^T s(\theta) / n. \quad (3.4)$$

Using (3.2) we obtain immediately

$$x(\theta)s(\theta) = (1 - \theta)xs + \theta^2 uv, \quad \mu(\theta) = (1 - \theta)\mu + \theta^2 u^T v / n. \quad (3.5)$$

Let us denote

$$p = \frac{xs}{\mu}, \quad q = \frac{uv}{\mu}.$$

From Lemma 3.1 it follows that  $e^T q \leq .25n$ . It follows that

$$\mu(\theta) > \mu(\bar{\theta}_0) = 0, \quad \text{for all } 0 \leq \theta < \bar{\theta}_0 = \frac{2}{1 + \sqrt{1 - 4e^T q/n}}. \quad (3.6)$$

The relation

$$x(\theta)s(\theta) \geq (1 - \gamma)\beta\mu(\theta) \quad (3.7)$$

is equivalent with the following system of quadratic inequalities

$$(1 - \theta)(p_i - (1 - \gamma)\beta) + \theta^2(q_i - (1 - \gamma)\beta e^T q/n) \geq 0, \quad i = 1, \dots, n, \quad (3.8)$$

Since  $z \in \mathcal{D}(\beta)$  the above inequalities are satisfied for  $\theta = 0$ . Let us denote by  $\Delta_i$  the discriminant of the  $i$ 'th quadratic function in (3.8), i.e.

$$\Delta_i = (p_i - (1 - \gamma)\beta)^2 - 4(p_i - (1 - \gamma)\beta)(q_i - (1 - \gamma)\beta e^T q/n).$$

The  $i$ -th equation in (3.8) is clearly satisfied for all  $\theta \in (0, \bar{\theta}_i]$ , where

$$\bar{\theta}_i = \begin{cases} \infty & \text{if } \Delta_i \leq 0 \\ 1 & \text{if } q_i - (1 - \gamma)\beta e^T q/n = 0 \\ \frac{2(p_i - (1 - \gamma)\beta)}{p_i - (1 - \gamma)\beta + \sqrt{\Delta_i}} & \text{if } \Delta_i > 0 \text{ and } q_i - (1 - \gamma)\beta e^T q/n \neq 0 \end{cases}.$$

Therefore we have

$$x(\theta)s(\theta) \geq (1 - \gamma)\beta\mu(\theta) > (1 - \gamma)\beta\mu(\bar{\theta}) \geq 0, \quad \text{for all } 0 \leq \theta < \bar{\theta}, \quad (3.9)$$

where

$$\bar{\theta} = \min \{ \bar{\theta}_i : i = 0, 1, \dots, n \}. \quad (3.10)$$

From (3.2) and (3.4) it follows that  $Qx(\theta) + Rs(\theta) = b$ , and by using a standard continuity argument we can prove that  $x(\theta) > 0$ ,  $s(\theta) > 0$  for all  $\theta \in (0, \bar{\theta})$ . As a result of the predictor step we obtain the the predicted point

$$\bar{z} = z(\bar{\theta}). \quad (3.11)$$

If  $\bar{\theta} = \bar{\theta}_0$ , then  $\mu(\bar{z}) = 0$ , so that  $\bar{z}$  is an optimal solution of our problem, i.e.  $\bar{z} \in \mathcal{F}^*$ . If  $\mu(\bar{z}) > 0$  then  $\bar{z} \in \mathcal{D}((1 - \gamma)\beta)$  and a corrector step will be performed. Before describing the corrector step, let us give a simple example that shows that it is possible to have  $\bar{z} \in \mathcal{F}^*$ . We remark that if  $e^T q = 0$ , which is always the case if HLCP is skew symmetric, then  $\bar{\theta} \leq 1$ . Indeed in this case there is an  $i$  such that  $q_i \leq 0$  which implies  $\bar{\theta}_i \leq 1$ .



However if  $e^T q > 0$  then we can only guarantee that  $\bar{\theta} \leq 2$ . This upper bound is implied, for example, by the fact that  $\bar{\theta}_0 \leq 2$ . In the following example we have  $\bar{\theta} = 2$  and  $\bar{z} \in \mathcal{F}^*$ .

**Example 1.** Take  $n = 2$  and

$$Q = I, R = -I, b = 0, x = s = e.$$

It is easily seen that  $z = [x, s] \in \mathcal{D}(\beta)$ , for any  $\beta \in (0, 1)$  and that the unique solution of (3.2) is  $u = v = -.5e$ . In this case we have  $\bar{\theta} = 2$  and  $\bar{z} = [0, 0] \in \mathcal{F}^*$ .

Let us assume now that  $\bar{z} \notin \mathcal{F}^*$ . As we have seen above, in this case  $\bar{z} \in \mathcal{D}((1-\gamma)\beta)$ . In the unlikely situation that  $\bar{z} \in \mathcal{D}(\beta)$ , which can never happen if  $u^T v = 0$ , then no corrector is necessary. Otherwise we will use a corrector step based on the Newton direction of  $F_{\mu(\bar{z})}$  at  $\bar{z}$

$$\bar{w} = [\bar{u}, \bar{v}] = -F'_{\mu(\bar{z})}(\bar{z})^{-1} F_{\mu(\bar{z})}(\bar{z}), \quad (3.12)$$

which is also called centering direction at  $\bar{z}$ . It can be computed by solving the following linear system

$$\begin{aligned} \bar{s}\bar{u} + \bar{x}\bar{v} &= \mu(\bar{z}) - \bar{x}\bar{s} \\ Q\bar{u} + R\bar{v} &= 0. \end{aligned} \quad (3.13)$$

Let us denote

$$\bar{x}(\theta) = \bar{x} + \theta\bar{u}, \bar{s}(\theta) = \bar{s} + \theta\bar{v}, \bar{z}(\theta) = [\bar{x}(\theta), \bar{s}(\theta)], \bar{\mu} = \mu(\bar{z}), \bar{\mu}(\theta) = \mu(\bar{z}(\theta)).$$

From (3.13) it follows that

$$\bar{x}(\theta)\bar{s}(\theta) = (1-\theta)\bar{x}\bar{s} + \theta\bar{\mu} + \theta^2\bar{u}\bar{v}, \quad \bar{\mu}(\theta) = \bar{\mu} + \theta^2\bar{u}^T\bar{v}/n. \quad (3.14)$$

The relation

$$\bar{x}(\theta)\bar{s}(\theta) \geq \beta\bar{\mu}(\theta) \quad (3.15)$$

is equivalent to the following system of quadratic inequalities in  $\theta$

$$\bar{f}_i(\theta) := \bar{p}_i - \beta + \theta(1 - \bar{p}_i) + \theta^2(\bar{q}_i - \beta e^T \bar{q}/n) \geq 0, \quad i = 1, \dots, n, \quad (3.16)$$

where

$$\bar{p} = \frac{\bar{x}\bar{s}}{\bar{\mu}}, \quad \bar{q} = \frac{\bar{u}\bar{v}}{\bar{\mu}}.$$

Let

$$\bar{\alpha}_i = \bar{q}_i - \beta e^T \bar{q}/n, \quad \bar{\Delta}_i = (1 - \bar{p}_i)^2 - 4(\bar{q}_i - \beta e^T \bar{q}/n)(\bar{p}_i - \beta)$$

denote the leading coefficient, and the discriminant of  $\bar{f}_i(\theta)$ . If  $\bar{\Delta}_i \geq 0$  and  $\bar{\alpha}_i \neq 0$  we denote by  $\check{\theta}_i$  and  $\hat{\theta}_i$  the smallest and the largest root of  $\bar{f}_i(\theta)$  respectively, i.e.

$$\check{\theta}_i = \frac{\bar{p}_i - 1 - \text{sign}(\bar{\alpha}_i)\sqrt{\bar{\Delta}_i}}{2\bar{\alpha}_i}, \quad \hat{\theta}_i = \frac{\bar{p}_i - 1 + \text{sign}(\bar{\alpha}_i)\sqrt{\bar{\Delta}_i}}{2\bar{\alpha}_i}$$

In the proof of Theorem 3.2 we will show that (3.16) has a solution, so that the following situation cannot occur for any  $i = 1, \dots, n$

$$\bar{\Delta}_i < 0 \text{ and } \bar{\alpha}_i < 0.$$

Therefore the  $i$ 'th inequality in (3.16) will be satisfied for all  $\theta \in \mathcal{T}_i$ , where

$$\mathcal{T}_i = \begin{cases} (-\infty, \infty), & \text{if } \bar{\Delta}_i < 0 \text{ and } \bar{\alpha}_i > 0 \\ [\hat{\theta}_i, \infty), & \text{if } \bar{\Delta}_i \geq 0 \text{ and } \bar{\alpha}_i > 0 \\ [\check{\theta}_i, \hat{\theta}_i], & \text{if } \bar{\Delta}_i \geq 0 \text{ and } \bar{\alpha}_i < 0 \\ (-\infty, (\bar{p}_i - \beta)/(\bar{p}_i - 1)], & \text{if } \bar{\alpha}_i = 0 \text{ and } \bar{p}_i > 1 \\ [(\bar{p}_i - \beta)/(\bar{p}_i - 1), \infty), & \text{if } \bar{\alpha}_i = 0 \text{ and } \bar{p}_i < 1 \\ (-\infty, \infty), & \text{if } \bar{\alpha}_i = 0 \text{ and } \bar{p}_i = 1 \end{cases}.$$

It follows that (3.15) holds for all  $\theta \in \mathcal{T}$  where

$$\mathcal{T} = \bigcap_{i=1}^n \mathcal{T}_i. \quad (3.17)$$

The steplength  $\theta_+$  for the corrector is defined by

$$\theta_+ = \min \mathcal{T}. \quad (3.18)$$

We note that if  $\bar{u}^T \bar{v} > 0$  then  $\theta_+$  minimizes  $\bar{\mu}(\theta)$  on  $\mathcal{T}$ . If  $\bar{u}^T \bar{v} = 0$  then  $\bar{\mu}(\theta)$  is constant on  $\mathcal{T}$ , and  $\theta_+$  could be determined to minimize  $\delta_{\infty}^-(\bar{z}(\theta))$  on  $\mathcal{T}$ . In any case we have

$$\bar{\mu}(\theta_+) \leq \bar{\mu}(\theta), \quad \forall \theta \in \mathcal{T}. \quad (3.19)$$

With  $\theta_+$  determined as above, the corrector step produces a point

$$z^+ = \bar{z} + \theta_+ \bar{w} \in \mathcal{D}(\beta), \quad (3.20)$$

and another predictor-corrector iteration can be performed.

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

**Algorithm 1**

Given  $\beta \in (0, 1)$  and  $z^0 \in \mathcal{D}(\beta)$ :

  Compute  $\gamma$  from (3.3);

  Set  $\mu_0 \leftarrow \mu(z^0)$ ,  $k \leftarrow 0$ ;

**repeat**

*(predictor step)*

    Set  $z \leftarrow z^k$ ;

    Compute affine scaling direction (3.1) by solving (3.2);

    Compute predictor steplength (3.10);

    Compute  $\bar{z}$  from (3.11);

    If  $\mu(\bar{z}) = 0$  then STOP:  $\bar{z}$  is an optimal solution;

    If  $\bar{z} \in \mathcal{D}(\beta)$  then

      set  $z^{k+1} \leftarrow \bar{z}$ ,  $\mu_{k+1} \leftarrow \mu(\bar{z})$ ,  $k \leftarrow k + 1$ , and RETURN;

*(corrector step)*

    Compute centering direction (3.12) by solving (3.13);

    Compute centering steplength (3.18);

    Compute  $z^+$  from (3.20);

    Set  $z^{k+1} \leftarrow z^+$ ,  $\mu_{k+1} \leftarrow \mu(z^+)$ ,  $k \leftarrow k + 1$ , and RETURN;

**until** some stopping criterion is satisfied.

**Polynomial Complexity.** In what follows we will prove that the steplength  $\bar{\theta}$  computed in the predictor step is bounded below by a quantity of the form  $\chi/n$ , where  $\chi$  is a positive constant. This implies that Algorithm 1 has  $O(nL)$ -iteration complexity.

We start by recalling some well known facts from the interior point literature that will be used in the proof of our main result (see, for example, [17]).

**Lemma 3.1** *Assume that HLCP (2.1) is monotone, and let  $w = [u, v]$  be the solution of the following linear system*

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

where  $z = [x, s] \in \mathbb{R}_{++}^{2n}$  and  $a \in \mathbb{R}^n$  are given vectors, and consider the index set:

$$\mathcal{I}_+(w) = \{i \in \{1, \dots, n\} : u_i v_i > 0\}.$$

Then the following inequalities are satisfied:

$$\|uv\|_\infty \leq \sum_{i \in \mathcal{I}_+(w)} u_i v_i \leq \frac{1}{4} \left\| (xs)^{-1/2} a \right\|_2^2.$$

The following result implies that Algorithm 1 has  $O(nL)$  iteration complexity.

**Theorem 3.2** *If HLCP (2.1) is monotone then Algorithm 1 is well defined and*

$$\mu_{k+1} \leq \left(1 - \frac{3\sqrt{(1-\beta)\beta}}{2(n+2)}\right) \mu_k, \quad k = 0, 1, \dots$$

*Proof.* According to Lemma 3.1 we have

$$\|q\|_\infty \leq .25n, \quad e^T q \leq \sum_{i \in \mathcal{I}_+(w)} q_i \leq .25n. \quad (3.21)$$

Since  $z \in \mathcal{D}(\beta)$ , and the function  $t \rightarrow t/(t + \sqrt{t^2 + 4at})$  is increasing on  $(0, \infty)$  for any  $a > 0$ , we deduce that

$$\begin{aligned} \bar{\theta}_i &\geq \frac{2\beta\gamma}{\beta\gamma + \sqrt{\Delta_i}} \geq \frac{2\beta\gamma}{\beta\gamma + \sqrt{\beta^2\gamma^2 + 4\beta\gamma(\|q\|_\infty + e^T q/n)}} \\ &\geq \frac{2}{1 + \sqrt{1 + (\beta\gamma)^{-1}(n+1)}}. \end{aligned}$$

It follows that the quantity defined in (3.10) satisfies

$$\bar{\theta} \geq \frac{2}{1 + \sqrt{1 + (\beta\gamma)^{-1}(n+1)}} = \frac{2\sqrt{\beta(1-\beta)}}{\sqrt{\beta(1-\beta)} + \sqrt{(n+1)^2 + \beta(1-\beta)}}.$$

Since

$$\sqrt{\beta(1-\beta)} + \sqrt{(n+1)^2 + \beta(1-\beta)} \leq .5 + \sqrt{(n+1)^2 + .25} < n+2$$

we deduce that

$$\bar{\theta} > \tilde{\theta} := \frac{2\sqrt{(1-\beta)\beta}}{n+2}. \quad (3.22)$$

Relations (3.9), (3.5) and (3.21) imply

$$\bar{\mu} = \mu(\bar{\theta}) < \mu(\tilde{\theta}) \leq \left((1 - \tilde{\theta}) + .25\tilde{\theta}^2\right) \mu = \left(1 - (1 - .25\tilde{\theta})\tilde{\theta}\right) \mu.$$

Since we assume that  $n \geq 2$ , we have

$$1 - .25\tilde{\theta} = 1 - \frac{\sqrt{(1-\beta)\beta}}{2(n+2)} \geq 1 - \frac{\sqrt{(1-\beta)\beta}}{8} \geq 1 - \frac{1}{16} = \frac{15}{16},$$

and we obtain

$$\bar{\mu} \leq \left(1 - \frac{15\sqrt{(1-\beta)\beta}}{8(n+2)}\right) \mu. \quad (3.23)$$

Let us analyze now the corrector. Since  $\bar{z} \in \mathcal{D}((1-\gamma)\beta)$ , we have

$$\left\| \sqrt{\frac{\bar{x}\bar{s}}{\bar{\mu}}} - \sqrt{\frac{\bar{\mu}}{\bar{x}\bar{s}}} \right\|_2^2 = \sum_{i=1}^n \frac{\bar{\mu}}{\bar{x}_i\bar{s}_i} - 2n + \sum_{i=1}^n \frac{\bar{x}_i\bar{s}_i}{\bar{\mu}} = \sum_{i=1}^n \frac{\bar{\mu}}{\bar{x}_i\bar{s}_i} - n \leq \frac{1-(1-\gamma)\beta}{(1-\gamma)\beta} n,$$

and by applying Lemma 3.1 we deduce that

$$\|\bar{u}\bar{v}\|_\infty \leq \frac{\xi n}{4} \bar{\mu}, \quad \sum_{i \in \mathcal{I}_+(\bar{w})} \bar{u}_i \bar{v}_i \leq \frac{\xi n}{4} \bar{\mu}, \quad \text{where } \xi := \frac{1-(1-\gamma)\beta}{(1-\gamma)\beta}. \quad (3.24)$$

From (3.14) it follows that

$$\frac{\bar{x}(\theta)\bar{s}(\theta)}{\bar{\mu}} \geq (1-\theta)(1-\gamma)\beta + \theta - .25\xi n\theta^2 = (1-\gamma)\beta + \theta(1-(1-\gamma)\beta) - .25\xi n\theta^2$$

and

$$\bar{\mu}(\theta) \leq (1 + .25\xi\theta^2) \bar{\mu}. \quad (3.25)$$

Therefore

$$\frac{\bar{x}(\theta)\bar{s}(\theta) - \beta\bar{\mu}(\theta)}{\bar{\mu}} \geq g(\theta) := -\gamma\beta + \theta(1-(1-\gamma)\beta) - .25\xi(n+\beta)\theta^2.$$

Using the definition of  $\xi$  together with the definition (3.3) of  $\gamma$  we obtain

$$g(\theta) = \frac{1-\beta}{4\beta(n+1)} \left(-4\beta^2 + 4\beta(n+1+\beta)\theta - (n+1)(n+1+\beta)\theta^2\right).$$

Since

$$g\left(\frac{2\beta}{n+1}\right) = \frac{(1-\beta)\beta^2}{(n+1)^2} \geq 0,$$

we deduce that  $2\beta/(n+1) \in \mathcal{T}$ . By using (3.25) and  $n \geq 2$  we have

$$\mu_+ = \bar{\mu}(\theta_+) \leq \bar{\mu}\left(\frac{2\beta}{n+1}\right) \leq \left(1 + \frac{(1-\beta)\beta(n+1+\beta)}{(n+1)^2(n+\beta)}\right) \bar{\mu}.$$

Given that  $\beta(1 - \beta) \leq .25$  and  $n \geq 2$  the above inequality implies that

$$\mu_+ \leq \left(1 + \frac{3(1 - \beta)\beta}{2(n + 1)^2}\right) \bar{\mu}. \quad (3.26)$$

Finally, by using (3.23), we we obtain

$$\begin{aligned} \mu_+ &\leq \left(1 - \frac{15\sqrt{(1 - \beta)\beta}}{8(n + 2)}\right) \left(1 + \frac{3(1 - \beta)\beta}{2(n + 1)^2}\right) \mu \\ &\leq \left(1 - \frac{15\sqrt{(1 - \beta)\beta}}{8(n + 2)}\right) \left(1 + \frac{3\beta(1 - \beta)}{2n(n + 2)}\right) \mu \\ &\leq \left(1 - \frac{15\sqrt{(1 - \beta)\beta}}{8(n + 2)} + \frac{3\beta(1 - \beta)}{2n(n + 2)}\right) \mu \\ &\leq \left(1 - \left(\frac{15}{8} - \frac{3\sqrt{(1 - \beta)\beta}}{2n}\right) \frac{\sqrt{(1 - \beta)\beta}}{(n + 2)}\right) \mu \\ &\leq \left(1 - \frac{3\sqrt{(1 - \beta)\beta}}{2(n + 2)}\right) \mu. \end{aligned}$$

The proof is complete. ■

**Quadratic Convergence** We end this section by showing that if the HLCP has a strictly complementary solution, then for large  $k$  the decrease of  $\mu_k$  is much faster than indicated by the bounds obtained in Theorem 3.2. In fact, we will prove that the sequence  $\{\mu(z^k)\}$  is quadratically convergent in the sense that

$$\mu_{k+1} = O(\mu_k^2). \quad (3.27)$$

Let us denote by

$$\mathcal{F}^\# := \{z^* = [x^*, s^*] \in \mathcal{F}^* : x^* + s^* > 0\}$$

the set of all strictly complementary solutions of our HLCP. In what follows we will assume that  $\mathcal{F}^\#$  is nonempty. In this case we say that our HLCP is nondegenerate. As mentioned in the introduction this assumption is not restrictive, since according to [11] strict complementarity is a necessary condition for superlinear convergence for a large class of interior point methods using only first order derivatives.

The following result was first proved for the standard monotone LCP in [19]. Its extension for the HLCP is immediate, using for example the equivalences proved in [1] (see also [3]).

**Lemma 3.3** *If  $\mathcal{F}^\# \neq \emptyset$  then the solution  $w = [u, v]$  of (3.2) satisfies*

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

where  $\mu = \mu(z)$  is given by (1.2).

**Theorem 3.4** *If the HLCP has a strictly complementary solution, then the sequence  $\{\mu_k\}$  generated by Algorithm 1 converges quadratically to zero in the sense that (3.27) is satisfied.*

*Proof.* We use the same notation as in the proof of Theorem 3.2. In particular, we will drop whenever possible the subscripts and superscripts  $k$ . From Lemma 3.3 it follows that there is a constant  $\varsigma_1$  such that

$$\sum_{i \in \mathcal{I}_+(w)} u_i v_i \leq \varsigma_1 \mu^2, \quad \|uv\|_\infty \leq \varsigma_1 \mu^2.$$

Since  $z \in \mathcal{D}(\beta)$  we deduce from (3.5) that

$$x(\theta)s(\theta) \geq (1 - \theta)\beta\mu - \varsigma_1\theta^2\mu^2, \quad \mu(\theta) \leq (1 - \theta)\mu + \varsigma_1\theta^2\mu^2/n \quad \forall \theta \in (0, 1). \quad (3.28)$$

Therefore (3.7) holds for any  $\theta$  satisfying the inequality

$$(1 - \theta)\beta - \varsigma_1\theta^2\mu \geq (1 - \gamma)\beta(1 - \theta) + (1 - \gamma)\beta\varsigma_1\theta^2\mu/n,$$

which is clearly valid if

$$\gamma\beta(1 - \theta) \geq (1 + (1 - \gamma)\beta/n)\varsigma_1\mu.$$

It follows that the quantity defined in (3.10) satisfies the inequality.

$$\bar{\theta} \geq 1 - \frac{(1 + (1 - \gamma)\beta/n)\varsigma_1}{\gamma\beta}\mu = 1 - \varsigma_2\mu.$$

From (3.9) and (3.28) we deduce that

$$\bar{\mu} = \mu(\bar{\theta}) \leq \mu(1 - \varsigma_2\mu) \leq (\varsigma_2 + \varsigma_1/n)\mu^2,$$

and by using (3.26) we obtain

$$\mu(z^+) \leq (1 + 1.5\beta(1 - \beta)/(n + 1)^2)\bar{\mu} \leq (1 + 1.5\beta(1 - \beta)/(n + 1)^2)(\varsigma_2 + \varsigma_1/n)\mu^2.$$

The proof is complete. ■

## 4 A higher order predictor corrector

In this section we consider a predictor  $z(\theta) = [x(\theta), s(\theta)]$  of order  $m$  of the form

$$z(\theta) = z + \sum_{i=1}^m w^i \theta^i, \quad (4.1)$$

where the vectors  $w^i = [u^i, v^i]$  are obtained as solutions of the following linear systems

$$\begin{cases} su^1 + xv^1 = -(1 + \epsilon)xs \\ Qu^1 + Rv^1 = 0 \end{cases}, \\ \begin{cases} su^2 + xv^2 = \epsilon xs - u^1v^1 \\ Qu^2 + Rv^2 = 0 \end{cases}, \quad (4.2) \\ \begin{cases} su^i + xv^i = -\sum_{j=1}^{i-1} u^jv^{i-j} \\ Qu^i + Rv^i = 0 \end{cases}, \quad i = 3, \dots, m,$$

where

$$\epsilon = \begin{cases} 0, & \text{if HLCP is nondegenerate} \\ 1, & \text{if HLCP is degenerate} \end{cases}. \quad (4.3)$$

The  $m$  linear systems above have the same matrix, so that their numerical solution requires only one matrix factorization and  $m$  backsolves. This involves  $O(n^3) + mO(n^2)$  arithmetic operations.

We note that for  $\epsilon = 0$ ,  $w^1$  is just the affine scaling direction used in the first order predictor. The directions  $w^i$  are related to the higher derivatives of the central path (see, for example, [15]).

Given predictor (4.1) we define

$$\tilde{\theta} = \sup \{ \tilde{\theta} > 0 : z(\theta) \in \mathcal{D}((1 - \gamma)\beta), \forall \theta \in [0, \tilde{\theta}] \}, \quad (4.4)$$

where  $\gamma$  is given by (3.3). From (4.1) - (4.2) we deduce that

$$\begin{aligned} x(\theta)s(\theta) &= (1 - \theta)^{1+\epsilon}xs + \sum_{i=m+1}^{2m} \theta^i h^i, \\ \mu(\theta) &= (1 - \theta)^{1+\epsilon}\mu + \sum_{i=m+1}^{2m} \theta^i (e^T h^i / n), \\ \text{where } h^i &= \sum_{j=i-m}^m u^j v^{i-j}. \end{aligned} \quad (4.5)$$

Therefore the computation of (4.4) involves the solution of a system of polynomial inequalities of order  $2m$  in  $\theta$ . Good lower bounds of the exact solution can be obtained by



a line search procedure. In what follows we will give simple lower bounds in explicit form, that are sufficiently good for proving the theoretical results.

After computing  $\check{\theta}$ , or a suitable convenient lower bound, we define

$$\bar{\theta} = \operatorname{argmin} \{ \mu(\theta) : \theta \in [0, \check{\theta}] \}, \quad \bar{z} = z(\bar{\theta}). \quad (4.6)$$

We have  $\bar{z} \in \mathcal{D}((1 - \gamma)\beta)$  by construction. Using the same corrector as in the previous section we obtain  $z^+ \in \mathcal{D}(\beta)$ . By replacing the predictor in Algorithm 1, with the predictor described above we obtain:

**Algorithm 2**

Given a real number  $\beta \in (0, 1)$ , an integer  $m \geq 2$ , and a vector  $z^0 \in \mathcal{D}(\beta)$ :

  Compute  $\gamma$  from (3.3);

  Set  $k \leftarrow 0$ ;

**repeat**

    (*predictor step*)

    Compute directions  $w^i = \lceil u^i, v^i \rceil$ ,  $i = 1, \dots, m$  by solving (4.2);

    Compute  $\check{\theta}$  from (4.4);

    Compute  $\bar{z}$  from (4.6);

    If  $\mu(\bar{z}) = 0$  then STOP:  $\bar{z}$  is an optimal solution;

    Set  $\bar{z}^k \leftarrow \bar{z}$ ;

    If  $\bar{z} \in \mathcal{D}(\beta)$  then

      set  $z^{k+1} \leftarrow \bar{z}$ ,  $\mu_{k+1} \leftarrow \mu(\bar{z})$ ,  $k \leftarrow k + 1$ , and RETURN;

    (*corrector step*)

    Compute centering direction (3.12) by solving (3.13);

    Compute centering steplength (3.18);

    Compute  $z^+$  from (3.20);

    Set  $z^{k+1} \leftarrow z^+$ ,  $\mu_{k+1} \leftarrow \mu(z^+)$ ,  $k \leftarrow k + 1$ , and RETURN;

**until** some stopping criterion is satisfied.

Let us denote

$$\eta_i = \left\| Du^i + D^{-1}v^i \right\|_2, \quad \text{where } D = X^{-1/2}S^{1/2}.$$

In the following lemma we obtain upper bounds for  $\eta_i$ . Our majorization technique is a slight improvement over the one used in [6, 21].

**Lemma 4.1** *For all  $i = 1, 2, \dots, m$ , we have*

$$\sqrt{\|Du^i\|_2^2 + \|D^{-1}v^i\|_2^2} \leq \eta_i \leq 2\alpha_i \sqrt{\beta\mu} \left( \frac{1 + \epsilon}{2} \sqrt{n/\beta} \right)^i,$$

where the sequence

$$\alpha_i = \frac{1}{i} \binom{2i-2}{i-1} \leq \frac{1}{i} 4^i$$

satisfies the following recurrence scheme

$$\alpha_1 = 1, \quad \alpha_i = \sum_{j=1}^{i-1} \alpha_j \alpha_{i-j}.$$

*Proof.* The first part of the inequality follows immediately, since by using (4.2) and the monotony of the HLCP we deduce that  $u^{iT}v^i \geq 0$ . Hence

$$\|Du^i + D^{-1}v^i\|_2^2 = \|Du^i\|_2^2 + 2u^{iT}v^i + \|D^{-1}v^i\|_2^2 \geq \|Du^i\|_2^2 + \|D^{-1}v^i\|_2^2.$$

By multiplying the first equations of (4.2) with  $(xs)^{-1/2}$  we obtain

$$\begin{aligned} Du^1 + D^{-1}v^1 &= -(1+\epsilon)(xs)^{1/2} \\ Du^2 + D^{-1}v^2 &= \epsilon(xs)^{1/2} - (xs)^{-1/2}u^1v^1 \\ Du^i + D^{-1}v^i &= -(xs)^{-1/2} \sum_{j=1}^{i-1} Du^j D^{-1}v^{i-j}, \quad i = 3, \dots, m. \end{aligned}$$

Because  $z \in \mathcal{D}(\beta)$  we have  $(xs)^{-1/2} \leq 1/\sqrt{\beta\mu}$ , and we deduce that

$$\begin{aligned} \eta_1 &= (1+\epsilon)\sqrt{n\mu}, \\ \eta_2^2 &= \epsilon^2 n\mu - 2\epsilon u^{1T}v^1 + \|(xs)^{-1/2}u^1v^1\|_2^2 \leq \epsilon^2 n\mu + \frac{\|u^1v^1\|_2^2}{\beta\mu} \\ &\leq \epsilon^2 n\mu + \frac{\eta_1^4}{8\beta\mu} \leq \epsilon^2 n\mu + \frac{(1+\epsilon)^4 n^2 \mu}{8\beta} < \frac{(1+\epsilon)^4 n^2 \mu}{4\beta}, \\ \eta_i &\leq \frac{1}{\sqrt{\beta\mu}} \sum_{j=1}^{i-1} \|Du^j\|_2 \|D^{-1}v^{i-j}\|_2, \quad i = 3, \dots, m. \end{aligned}$$

Since

$$\begin{aligned} &\|Du^j\|_2 \|D^{-1}v^{i-j}\|_2 + \|Du^{i-j}\|_2 \|D^{-1}v^j\|_2 \\ &\leq \left( \|Du^j\|_2^2 + \|D^{-1}v^j\|_2^2 \right)^{1/2} \left( \|Du^{i-j}\|_2^2 + \|D^{-1}v^{i-j}\|_2^2 \right)^{1/2} \leq \eta_j \eta_{i-j}, \end{aligned}$$

we obtain

$$\eta_i \leq \frac{1}{2\sqrt{\beta\mu}} \sum_{j=1}^{i-1} \eta_j \eta_{i-j}, \quad i = 3, \dots, m.$$

The required inequalities are then easily proved by mathematical induction. ■

**Theorem 4.2** *If HCLP (2.1) is monotone then Algorithm 2 is well defined and for each  $n \geq 7$  we have*

$$\mu_{k+1} \leq \left(1 - .01 \frac{\sqrt{\beta} \sqrt[3]{1-\beta}}{\sqrt{n}^{m+1} \sqrt{n+1}}\right) \mu_k, \quad k = 0, 1, \dots$$

*Proof.* An upper bound of  $\|h^i\|_2$ ,  $i = m+1, m+2, \dots, 2m$  can be obtained by writing

$$\begin{aligned} \|h^i\|_2 &\leq \sum_{j=i-m}^m \|Du^j\|_2 \|D^{-1}v^{i-j}\|_2 \leq \sum_{j=1}^{i-1} \|Du^j\|_2 \|D^{-1}v^{i-j}\|_2 \\ &= \frac{1}{2} \sum_{j=1}^{i-1} \left( \|Du^j\|_2 \|D^{-1}v^{i-j}\|_2 + \|Du^{i-j}\|_2 \|D^{-1}v^j\|_2 \right) \\ &\leq \frac{1}{2} \sum_{j=1}^{i-1} \sqrt{\|Du^j\|_2^2 + \|D^{-1}v^j\|_2^2} \sqrt{\|Du^{i-j}\|_2^2 + \|D^{-1}v^{i-j}\|_2^2} \\ &\leq \frac{1}{2} \sum_{j=1}^{i-1} \eta_j \eta_{i-j} \leq 2\beta\mu \left( \frac{1+\epsilon}{2} \sqrt{n/\beta} \right)^i \sum_{j=1}^{i-1} \alpha_j \alpha_{i-j} \\ &= 2\beta\mu \left( \frac{1+\epsilon}{2} \sqrt{n/\beta} \right)^i \alpha_i \leq \frac{2\beta\mu}{i} \left( 2(1+\epsilon) \sqrt{n/\beta} \right)^i \leq \frac{2\beta\mu}{i} \left( 4\sqrt{n/\beta} \right)^i. \end{aligned}$$

It follows that

$$\begin{aligned} \left\| \sum_{i=m+1}^{2m} \theta^i h^i \right\|_2 &\leq \frac{2\beta\mu}{m+1} \left( 4\theta \sqrt{n/\beta} \right)^{m+1} \sum_{j=0}^{m-1} \left( 4\theta \sqrt{n/\beta} \right)^j \\ &\leq \frac{2\beta\mu \left( 4\theta \sqrt{n/\beta} \right)^{m+1}}{(m+1) \left( 1 - \left( 4\theta \sqrt{n/\beta} \right) \right)}. \end{aligned}$$

For the remainder of this proof we assume that in the predictor step we have

$$\theta \leq \frac{\sqrt{\beta}}{6\sqrt{n}}. \quad (4.7)$$

In this case, since  $m \geq 2$ , we deduce that

$$\left\| \sum_{i=m+1}^{2m} \theta^i h^i \right\|_2 \leq \frac{6\beta\mu}{m+1} \left( 4\theta \sqrt{n/\beta} \right)^{m+1} \leq 2\beta\mu \left( 4\theta \sqrt{n/\beta} \right)^{m+1}.$$

Using (4.5),  $\|h\|_1 \leq \sqrt{n} \|h\|_2$ , and  $z \in \mathcal{D}(\beta)$  we obtain the following estimates

$$x(\theta)s(\theta) \geq (1-\theta)^{1+\epsilon}xs - \left\| \sum_{i=m+1}^{2m} \theta^i h^i \right\|_2 \geq \left[ (1-\theta)^{1+\epsilon}\beta - 2\beta \left(4\theta\sqrt{n/\beta}\right)^{m+1} \right] \mu,$$

$$\begin{aligned} \mu(\theta) = \frac{x(\theta)^T s(\theta)}{n} &\leq (1-\theta)^{1+\epsilon}\mu + \frac{1}{n} \left\| \sum_{i=m+1}^{2m} \theta^i h^i \right\|_1 \\ &\leq \left[ (1-\theta)^{1+\epsilon} + \frac{2\beta}{\sqrt{n}} \left(4\theta\sqrt{n/\beta}\right)^{m+1} \right] \mu. \end{aligned} \quad (4.8)$$

Therefore the inequality

$$x(\theta)s(\theta) \geq (1-\gamma)\beta\mu(\theta)$$

holds for any  $\theta$  satisfying (4.7) and

$$\gamma(1-\theta)^{1+\epsilon} \geq 2 \left(1 + \frac{1-\gamma}{\sqrt{n}}\right) \left(4\theta\sqrt{n/\beta}\right)^{m+1}.$$

It is easily seen that both the above inequality and (4.7) are satisfied by

$$\tilde{\theta} := \frac{\sqrt{\beta}}{6\sqrt{n}} m^{+1} \sqrt{\frac{\gamma}{2}}. \quad (4.9)$$

From (4.4) it follows that  $\bar{\theta} \leq \tilde{\theta}$ , and according to (4.6),(4.8) we have

$$\begin{aligned} \bar{\mu} &= \mu(\bar{\theta}) \leq \mu(\tilde{\theta}) \leq \left(1 - \frac{\sqrt{\beta}}{6\sqrt{n}} m^{+1} \sqrt{\frac{\gamma}{2}} + \frac{\beta\gamma 2^{m+1}}{\sqrt{n} 3^{m+1}}\right) \mu \\ &= \left(1 - \left(1 - \frac{\sqrt{\beta}\gamma 2^{m+2}}{3^m} m^{+1} \sqrt{\frac{2}{\gamma}}\right) \frac{\sqrt{\beta}}{6\sqrt{n}} m^{+1} \sqrt{\frac{\gamma}{2}}\right) \mu \\ &\leq \left(1 - \left(1 - \frac{2^{4+1/3}\sqrt{\beta}\gamma^{2/3}}{9}\right) \frac{\sqrt{\beta}}{6\sqrt{n}} m^{+1} \sqrt{\frac{\gamma}{2}}\right) \mu. \end{aligned}$$

Since  $\gamma = (1-\beta)/(n+1) \leq (1-\beta)/8$  and  $\sqrt{\beta}(1-\beta)^{2/3}$  is maximized for  $\beta = 3/7$ , we deduce that

$$\bar{\mu} \leq \left(1 - .1 \frac{\sqrt{\beta}}{\sqrt{n}} m^{+1} \sqrt{\frac{\gamma}{2}}\right) \mu = \left(1 - .1 \frac{\sqrt{\beta} m^{+1} \sqrt{(1-\beta)/2}}{\sqrt{n} m^{+1} \sqrt{n+1}}\right) \mu. \quad (4.10)$$

Let us analyze now the corrector. From (3.26) and (4.10) we have

$$\begin{aligned}
\mu_+ &\leq \left(1 + \frac{3\beta(1-\beta)}{2(n+1)^2}\right) \bar{\mu} \leq \left(1 + \frac{3\beta(1-\beta)}{2(n+1)^2}\right) \left(1 - .1 \frac{\sqrt{\beta}^{m+1} \sqrt{(1-\beta)/2}}{\sqrt{n}^{m+1} \sqrt{n+1}}\right) \mu \\
&\leq \left(1 - .1 \frac{\sqrt{\beta}^{m+1} \sqrt{(1-\beta)/2}}{\sqrt{n}^{m+1} \sqrt{n+1}} + \frac{3\beta(1-\beta)}{2(n+1)^2}\right) \mu \\
&\leq \left(1 - \frac{\sqrt{\beta}^3 \sqrt{1-\beta}}{10\sqrt[3]{2} \sqrt{n}^{m+1} \sqrt{n+1}} + \frac{3\beta(1-\beta)}{2n(n+1)}\right) \mu \\
&\leq \left(1 - \left(1 - \frac{30\sqrt[3]{2} \sqrt{\beta}^3 \sqrt{(1-\beta)^2}}{2\sqrt{n}(n+1)^{m/(m+1)}}\right) \frac{\sqrt{\beta}^3 \sqrt{1-\beta}}{10\sqrt[3]{2} \sqrt{n}^{m+1} \sqrt{n+1}}\right) \mu.
\end{aligned}$$

Since  $n \geq 7$  and  $\sqrt{\beta}^3 \sqrt{(1-\beta)^2}$  is maximized for  $\beta = 3/7$ , we deduce that

$$\mu_+ \leq \left(1 - .01 \frac{\sqrt{\beta}^3 \sqrt{1-\beta}}{\sqrt{n}^{m+1} \sqrt{n+1}}\right) \mu,$$

which completes the proof. ■

As an immediate consequence of the above theorem we obtain the following complexity result:

**Corollary 4.3** *Algorithm 2 produces a point  $z^k \in \mathcal{D}(\beta)$  with  $x^k T s^k \leq \varepsilon$  in at most  $O\left(n^{1/2+1/(m+1)} \log\left(x^0 T s^0 / \varepsilon\right)\right)$  iterations.*

Using an unpublished idea of Roos (used also in [21]) we can take

$$m = O(\lceil (n+1)^\omega - 1 \rceil), \quad \text{for some } \omega \in (0, 1),$$

and obtain the following result:

**Corollary 4.4** *Algorithm 2, with  $m = O(\lceil (n+1)^\omega - 1 \rceil)$  produces a point  $z^k \in \mathcal{D}(\beta)$  with  $x^k T s^k \leq \varepsilon$  in a most  $O\left(\sqrt{n} \log\left(x^0 T s^0 / \varepsilon\right)\right)$  iterations.*

*Proof.* From Corollary 5.3 of [21] we have

$$\lim_{n \rightarrow \infty} {}^{(n+1)^\omega} \sqrt[n+1]{} = 1, \quad \text{and} \quad {}^{(n+1)^\omega} \sqrt[n+1]{} \leq e^{1/(\omega e)}, \quad n = 2, \dots,$$

which proves our claim. ■

As we mentioned before, the numerical implementation of a predictor of order  $m$  requires a matrix factorization and  $m$  backsolves. If the matrices  $Q$  and  $R$  are full, the cost of a matrix factorization is  $O(n^3)$  arithmetic operations, while the cost of a backsolve is  $O(n^2)$  arithmetic operations. Therefore, it appears that in order to minimize the cost, we should choose  $\omega$  close to 1. However, this reasoning is based on the "worst case scenario" bounds given in Theorem 4.2. As we will see below, the convergence to zero of the primal-dual gap is asymptotically much faster than indicated by those bounds, so that in practice the optimal order  $m$  of the predictor should be chosen based on numerical experiments. As mentioned in the introduction, a small value of  $\omega$  seems to be reasonable. For example if we take we take  $\omega = 0.1$ , then the values of  $m = \lceil (n+1)^\omega - 1 \rceil$  corresponding to  $n = 10^6$ ,  $n = 10^7$ ,  $n = 10^8$ , and  $n = 10^9$  are 3, 5, 6, and 7 respectively.

We end this paper by showing that the primal-dual gap of the sequence produced by Algorithm 2 is superlinearly convergent. More precisely we will prove that  $\mu_{k+1} = O(\mu^{m+1})$  if the HLCP (2.1) is nondegenerate, and  $\mu_{k+1} = O(\mu^{(m+1)/2})$  otherwise. The main ingredient of our proof is provided by the following lemma, which is an immediate consequence of the results of [15] about the analyticity of the central path:

**Lemma 4.5** *The solution of (4.2) satisfies*

$$u^i = O(\mu^i), \quad v^i = O(\mu^i), \quad i = 1, \dots, m, \quad \text{if HLCP (2.1) is nondegenerate,}$$

and

$$u^i = O(\mu^{i/2}), \quad v^i = O(\mu^{i/2}), \quad i = 1, \dots, m, \quad \text{if HLCP (2.1) is degenerate.}$$

**Theorem 4.6** *The sequence  $\mu_k$  produced by Algorithm 2 satisfies*

$$\mu_{k+1} = O(\mu^{m+1}), \quad \text{if HLCP (2.1) is nondegenerate,}$$

and

$$\mu_{k+1} = O(\mu^{(m+1)/2}), \quad \text{if HLCP (2.1) is degenerate.}$$

*Proof.* Let us denote

$$\nu = \frac{m+1}{1+\epsilon} = \begin{cases} m+1, & \text{if HLCP (2.1) is nondegenerate,} \\ (m+1)/2, & \text{if HLCP (2.1) is degenerate.} \end{cases}$$

From Lemma 4.5 and equation (4.5) it follows that there is a constant  $\varpi$  such that

$$x(\theta)s(\theta) \geq \beta(1-\theta)^{1+\epsilon}\mu - \theta^{m+1}\varpi\mu^\nu, \quad \mu(\theta) \leq (1-\theta)^{1+\epsilon}\mu + \theta^{m+1}\varpi\mu^\nu/\sqrt{n}, \quad (4.11)$$

for all  $\theta \in (0, 1)$ . Hence  $x(\theta)s(\theta) \geq (1 - \gamma)\beta\mu(\theta)$  for any  $\theta \in (0, 1)$  such that

$$\beta\gamma(1 - \theta)^{1+\epsilon} \geq \varpi(1 + (1 - \gamma)\beta/\sqrt{n})\mu^{\nu-1}.$$

Therefore, there is a constant  $\varpi_1$  such that

$$\check{\theta} \geq 1 - \varpi_1\mu^{(\nu-1)/(1+\epsilon)}.$$

According to (4.6), (4.11) we have

$$\bar{\mu} = \mu(\bar{\theta}) \leq \mu(1 - \varpi_1\mu^{\nu-1}) \leq \varpi_1\mu^\nu + \varpi\mu^\nu/n = (\varpi_1 + \varpi/\sqrt{n})\mu^\nu,$$

and by using (3.26) we obtain

$$\mu_+ \leq \left(1 + \frac{\beta(1 - \beta)}{(n + 1)^2}\right) \bar{\mu} \leq \left(1 + \frac{\beta(1 - \beta)}{(n + 1)^2}\right) (\varpi_1 + \varpi/\sqrt{n}) \mu^\nu.$$

The proof is complete. ■

## 5 Conclusions

We have presented a predictor-corrector interior point algorithm acting between two wide neighborhoods of the central path of a monotone HCLP  $\mathcal{D}(\beta) \subset \mathcal{D}((1 - \gamma)\beta)$ , where  $\beta$  can be any number from the interval  $(0, 1)$  and  $\gamma$  is given by (3.3). The predictor starts with a point  $z \in \mathcal{D}(\beta)$  and produces a point  $\bar{z} \in \mathcal{D}((1 - \gamma)\beta)$ . The corrector uses the centering direction at  $\bar{z}$  and produces a point  $z^+ \in \mathcal{D}(\beta)$ . The first order predictor uses the affine scaling direction at  $z$ , and the corresponding algorithm has  $O(nL)$  iteration complexity and is quadratically convergent for nondegenerate problems. If we use a predictor of order  $m \geq 2$ , then the corresponding algorithm is superlinearly convergent even for degenerate problem. If we choose  $m = \lceil (n + 1)^\omega - 1 \rceil$ , for some  $\omega \in (0, 1)$ , then the algorithm has  $O(\sqrt{n}L)$  iteration complexity. To our knowledge this is the first interior point algorithm acting in a wide neighborhood of the central path that has both  $O(\sqrt{n}L)$  iteration complexity and superlinear convergence. An anonymous referee pointed out that our algorithm can be presented as a corrector-predictor algorithm acting in  $\mathcal{D}(\beta)$ . In this variant one starts with a point in  $\mathcal{D}(\beta)$  and performs a corrector step, which produces a point closer to the central path, followed by a predictor step that produces a point on the boundary of  $\mathcal{D}(\beta)$ . In this case the parameter  $\gamma$  does not have to be explicitly given. While this can certainly be done, we have chosen to stay within the "classical" setting of the original MTY predictor-corrector algorithm with two explicitly defined neighborhoods of the central path.

*Acknowledgement.* This paper is dedicated to Lia, an incomparable companion who has been the source of inspiration for my best works .

## References

- [1] M. Anitescu, G. Lesaja, and F. A. Potra. An infeasible–interior–point predictor–corrector algorithm for the  $P_*$ -Geometric LCP. *Applied Mathematics & Optimization*, 36(2):203–228, 1997.
- [2] K.M Anstreicher and R.A. Bosch. A new infinity-norm path following algorithm for linear programming. *SIAM J. Optim.*, 5(2):236–246, 1995.
- [3] J. F. Bonnans and C. C. Gonzaga. Convergence of interior point algorithms for the monotone linear complementarity problem. *Mathematics of Operations Research*, 21:1–25, 1996.
- [4] R. W. Cottle, J.-S. Pang, and R. E. Stone. *The Linear Complementarity Problem*. Academic Press, Boston, MA, 1992.
- [5] C. C. Gonzaga. Complexity of predictor-corrector algorithms for LCP based on a large neighborhood of the central path. *SIAM J. Optim.*, 10(1):183–194 (electronic), 1999.
- [6] P-F. Hung and Y. Ye. An asymptotical ( $O\sqrt{nL}$ )-iteration path-following linear programming algorithm that uses wide neighborhoods. *Siam Journal on Optimization*, 6(3):159–195, August 1996.
- [7] B. Jansen, C. Roos, T. Terlaky, and J.-P. Vial. Primal-dual algorithms for linear programming based on the logarithmic barrier method. *J. Optim. Theory Appl.*, 83(1):1–26, 1994.
- [8] J. Ji, F. A. Potra, and S. Huang. A predictor-corrector method for linear complementarity problems with polynomial complexity and superlinear convergence. *Journal of Optimization Theory and Applications*, 84(1):187–199, 1995.
- [9] M. Kojima, N. Megiddo, T. Noma, and A. Yoshise. A unified approach to interior point algorithms for linear complementarity problems. *Lecture Notes in Comput. Sci.*, 538, 1991.
- [10] S. Mizuno, M. J. Todd, and Y. Ye. On adaptive-step primal-dual interior-point algorithms for linear programming. *Mathematics of Operations Research*, 18(4):964–981, 1993.



- [11] R. D. C. Monteiro and S. J. Wright. Local convergence of interior-point algorithms for degenerate monotone LCP. *Computational Optimization and Applications*, 3:131–155, 1994.
- [12] R.C. Monteiro, I. Adler, and M.G. Resende. A polynomial-time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extension. *Mathematics of Operations Research*, 15:191–214, 1990.
- [13] F. A. Potra. The Mizuno-Todd-Ye algorithm in a larger neighborhood of the central path. *European Journal of Operational Research*, 143:257–267, 2002.
- [14] C. Roos, J.-Ph. Vial, and T. Terlaky. *Theory and Algorithms for Linear Optimization : An Interior Point Approach*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley and Sons, 1997.
- [15] J. Stoer, M. Wechs, and S. Mizuno. High order infeasible-interior-point methods for solving sufficient linear complementarity problems. *Math. Oper. Res.*, 23(4):832–862, 1998.
- [16] J. F. Sturm. Superlinear convergence of an algorithm for monotone linear complementarity problems, when no strictly complementary solution exists. *Mathematics of Operations Research*, 24:72–94, 1999.
- [17] S. J. Wright. *Primal–Dual Interior–Point Methods*. SIAM Publications, Philadelphia, 1997.
- [18] Y. Ye. *Interior Point Algorithms : Theory and Analysis*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley and Sons, 1997.
- [19] Y. Ye and K. Anstreicher. On quadratic and  $O(\sqrt{n}L)$  convergence of predictor-corrector algorithm for LCP. *Mathematical Programming*, 62(3):537–551, 1993.
- [20] Y. Ye, O. Güler, R. A. Tapia, and Y. Zhang. A quadratically convergent  $O(\sqrt{n}L)$ -iteration algorithm for linear programming. *Mathematical Programming*, 59(2):151–162, 1993.
- [21] G. Zhao. Interior point algorithms for linear complementarity problems based on large neighborhoods of the central path. *SIAM J. Optim.*, 8(2):397–413 (electronic), 1998.