Mathematical Programming 62 (1993) 537–551
North-Holland

On quadratic and $O(\sqrt{n}\,L)$ convergence of a predictor–corrector algorithm for LCP

Yinyu Ye* and Kurt Anstreicher

Department of Management Sciences, University of Iowa, Iowa City, IA, USA

Received 3 December 1991
Revised manuscript received 20 September 1992

Recently several new results have been developed for the asymptotic (local) convergence of polynomial-time interior-point algorithms. It has been shown that the predictor–corrector algorithm for linear programming (LP) exhibits asymptotic quadratic convergence of the primal–dual gap to zero, without any assumptions concerning nondegeneracy, or the convergence of the iteration sequence. In this paper we prove a similar result for the monotone linear complementarity problem (LCP), assuming only that a strictly complementary solution exists. We also show by example that the existence of a strictly complementarity solution appears to be necessary to achieve superlinear convergence for the algorithm.

Key words: Linear complementarity problem, quadratic programming, superlinear convergence, quadratic convergence, polynomial-time algorithm.

1. Introduction

Consider the linear complementarity problem (LCP):

$$\begin{aligned} \min \quad & x^{\mathrm{T}}s \\ \text{s.t.} \quad & s = Mx + q, \quad (x, s) \geqslant 0, \end{aligned}$$

where $M \in \mathbb{R}^{n\times n}$ and $q \in \mathbb{R}^n$. As usual, we assume without losing generality:

(A1) The feasible region of LCP has a nonempty relative interior, i.e., there exists (x^0, s^0) such that $s^0 = Mx^0 + q$ and $x^0 > 0$, $s^0 > 0$.

Note that M may not be symmetric. However, for (x, s) feasible in LCP, the objective may be written as $x^{\mathrm{T}}s = \frac{1}{2}x^{\mathrm{T}}(M+M^{\mathrm{T}})x + q^{\mathrm{T}}x$, and $M+M^{\mathrm{T}}$ is symmetric. The LCP problem is called monotone (convex) if and only if $M+M^{\mathrm{T}}$ is positive semi-definite, which we assume throughout this paper:

Correspondence to: Prof. Yinyu Ye, Department of Management Sciences, University of Iowa, Iowa City, IA 52242, USA.

*Research supported in part by NSF Grants DDM-8922636 and DDM-9207347, and an Interdisciplinary Research Grant of the University of Iowa, Iowa Center for Advanced Studies.

(A2) M is a positive semi-definite matrix, that is $x^{\mathrm{T}}Mx=\frac{1}{2}x^{\mathrm{T}}(M+M^{\mathrm{T}})x\geqslant 0$ for every $x\in\mathbb{R}^n$.

We call a feasible point (x, s) strictly feasible if it is feasible and positive. A feasible point (x^*, s^*) is optimal (complementary) if and only if

$$x_j^* s_j^* = 0 \quad \text{for } j=1, 2, \ldots, n.$$

A strictly complementary solution is an optimal solution satisfying

$$x_j^* + s_j^* > 0 \quad \text{for } j=1, 2, \ldots, n.$$

Consider a sequence of strictly feasible points $\{(x^k, s^k)\}$ such that the (complementary) gap $(x^k)^{\mathrm{T}}s^k \to 0$. Then we say that this gap sequence converges Q-superlinearly to zero if

$$\lim_{k\to\infty} \frac{(x^{k+1})^{\mathrm{T}}s^{k+1}}{(x^k)^{\mathrm{T}}s^k} = 0,$$

and Q-quadratically to zero if

$$\limsup_{k\to\infty} \frac{(x^{k+1})^{\mathrm{T}}s^{k+1}}{((x^k)^{\mathrm{T}}s^k)^2} < +\infty.$$

In the context of the present work it is important to emphasize that the notions of convergence, superlinear convergence, or quadratic convergence of the gap sequence, $\{(x^k)^{\mathrm{T}}s^k\}$, in no way require the convergence of the iteration sequence $\{(x^k, s^k)\}$. Of course, from Hoffman's lemma [7] and Luo and Tseng's theorem [15] it follows that in a particular sense the iteration sequence converges to the optimal solution set with the corresponding R-rate.

Recently, there has been an exciting outbreak of activity in the area of constructing primal–dual interior-point algorithms for either the linear programming problem (LP), or the monotone linear complementarity problem (LCP) possessing a strictly complementary solution, that are demonstrably superlinearly or quadratically convergent under certain assumptions (e.g. Ji et al. [9, 10], Kojima et al. [11, 13], McShane [16, 17], Mehrotra [18], Ye et al. [28–30], and Zhang et al. [31–33]).

The issue of the asymptotic convergence of interior-point algorithms was first raised in Iri and Imai [8]. They showed that their multiplicative barrier function method (also see de Ghellinck and Vial [4]), with an exact line search procedure, possesses quadratic convergence for nondegenerate LP. Then, Yamashita [27] showed that a variant of this method possesses both polynomial $O(nL)$ complexity and quadratic convergence for nondegenerate LP, and Tsuchiya and Tanabe [25] showed that Iri and Imai's method possesses quadratic convergence under a weaker nondegeneracy assumption. Zhang and Tapia [31] showed that a primal–dual algorithm exhibits $O(nL)$ complexity, and superlinear convergence, under the assumption of convergence of the iteration sequence, with quadratic convergence under a nondegeneracy assumption. Kojima et al. [13] also showed quadratic convergence of a path-following algorithm for nonlinear complementarity problems under the nondegeneracy assumption. Other algorithms, interior or exterior, with quadratic convergence for nondegenerate LP include Coleman and Li's [3]. Some negative results on the asymptotic convergence of Karmarkar's original algorithm and a potential reduction method (with separate primal and dual updates) were given by Bayer and Lagarias [2] and Gonzaga and Todd [5], respectively.

Quadratic convergence for general LP was first established by Ye, Güler, Tapia and Zhang [29], Mehrotra [18], and Tsuchiya [24]. The algorithm of Ye et al., and Mehrotra, is based on the predictor–corrector algorithm of Mizuno et al. [19]. Each iteration of the algorithm needs to solve two systems of linear equations or two least squares problems — one in the predictor step and one in the corrector step. Tsuchiya's result is based on Iri and Imai's method, which requires knowledge of the exact optimal objective value in advance. A standard way of dealing with this difficulty is to integrate the primal and dual problems into a single LP problem, whose size is twice that of the original problem. The "currently best" result, to our knowledge, was given in [28] where it is shown that the Q-order of convergence of a variant of the $O(\sqrt{n}\,L)$-iteration predictor–corrector algorithm for general LP, counting each iteration as solving one system of linear equations of the size of the original problem, equals 2.

While superlinear or quadratic convergence results for LP have been established with no assumptions, all superlinear convergence results for LCP to date use some combination of the following assumptions:

(A3) The LCP has a strictly complementary solution.

(A4) The LCP is nondegenerate, meaning it has a unique solution.

(A5) The iteration sequence $\{(x^k, s^k)\}$ generated by the interior-point algorithm converges, and it converges to a strict complementarity solution.

(Note that (A3) automatically holds for LP.) The "currently best" results for LCP are given by Kojima et al. and Ji et al., and they can be cataloged as follows:

- global and quadratic convergence assuming (A3) and (A4) (Kojima et al. [11, 13]).
- $O(nL)$ iteration complexity and superlinear convergence assuming (A3) and (A5) (Ji et al. [10]).
- $O(\sqrt{n}\,L)$ iteration complexity and superlinear convergence assuming (A3) and (A5) (Ji et al. [9]).

In these bounds L represents the data length for a problem with all integer data.

Certainly, the global property of polynomiality and the local property of superlinearity are desirable. However, (A4) is not realistic, and (A5) may not hold in general. Thus, the current asymptotic convergence result for LCP is quite behind that for LP. In what follows we consider the LCP extension of the predictor–corrector algorithm (e.g., Mizuno et al. [19] and Sonnevend et al. [23]) suggested by Ji et al. [9]. We show that this $O(\sqrt{n}\,L)$ iteration algorithm for LCP actually possesses Q-quadratic convergence, without assuming either (A4) or (A5), where one iteration consists of two steps — one predictor and one corrector. Of course, we assume (A1), (A2), and (A3). Among these assumptions, (A1) is the standard assumption for any interior-point algorithm, and (A2) is necessary for the LCP to have a convex objective, but (A3) is restrictive since in general it does not hold for LCP. We will actually show by example, however, that (A3) appears to be necessary in order to achieve superlinear convergence for the algorithm. We also show that a modification of the algorithm achieves Q-order of convergence equals 2 (i.e., Q-subquadratic convergence), counting one iteration as one step. Our results thus completely fill in the asymptotic convergence gap between LP and LCP.

The paper is organized as follows: In Section 2 we review the predictor–corrector algorithm and collect several previously established estimates. Section 3 contains several technical results. Our main convergence results are given in Section 4. A summary and concluding remarks are contained in Section 5.

2. The predictor–corrector algorithm

In this section, we briefly describe the predictor–corrector LCP algorithm (Ji et al. [9]). We employ the notation $X=\mathrm{diag}(x)$, $S=\mathrm{diag}(s)$, etc., and we let Ω denote the collection of all strictly feasible points (x, s). Consider the neighborhood

$$\mathcal{N}(\alpha)=\{(x,s)\in\Omega\colon \|Xs/\mu-e\|\leqslant\alpha\},$$

where $\|\cdot\|$ represents the l_2 norm, $\mu=x^{\mathrm{T}}s/n$, e is the vector of all ones, and α is a constant between 0 and 1.

To begin with choose $0<\beta\leqslant\frac{1}{4}$ (a typical choice would be $\frac{1}{4}$). All search directions d_x and d_s will be defined as the solutions of the following system of linear equations (Kojima et al. [14])

$$\begin{aligned} Xd_s+Sd_x&=\gamma\mu e-Xs,\\ d_s&=Md_x, \end{aligned} \tag{1}$$

where $0\leqslant\gamma\leqslant 1$. A typical iteration of the algorithm proceeds as follows. Given $(x^k, s^k)\in\mathcal{N}(\beta)$, we solve the system (1) with $(x,s)=(x^k,s^k)$ and $\gamma=0$. Denote by d_x^{p} and d_s^{p} the resulting directions. For some step length $\theta\geqslant 0$ let

$$x(\theta)=x^k+\theta d_x^{\mathrm{p}},\qquad s(\theta)=s^k+\theta d_s^{\mathrm{p}},$$

and $\mu(\theta)=x(\theta)^{\mathrm{T}}s(\theta)/n$. This is the predictor step. Our specific choice for θ will be stated after we consider the following lemma [9, 19].

Lemma 2.1. *If for some positive $\theta^k\leqslant 1$ we have*

$$\|X(\theta)s(\theta)/\mu(\theta)-e\|\leqslant\alpha<1\quad \textit{for all } 0\leqslant\theta\leqslant\theta^k, \tag{2}$$

then $(x(\theta^k), s(\theta^k))\in\mathcal{N}(\alpha)\subset\Omega$. □

Lemma 2.1 basically says that the interior feasibility of $(x(\theta^k), s(\theta^k))$ is guaranteed as long as (2) is satisfied. Thus, we can choose the *largest* step length $\theta^k \leqslant 1$ such that (2) is satisfied for $\alpha = \beta + \tau$, for some $0 < \tau \leqslant \beta$, and let

$$\hat{x}^k = x(\theta^k) \quad \text{and} \quad \hat{s}^k = s(\theta^k).$$

We can compute the largest θ^k by finding the smallest positive root, r^k, of a quartic polynomial (if r^k does not exist, then $r^k = \infty$) and assigning $\theta^k = \min(r^k, 1)$, or simply choosing it as the lower bound in Lemma 2.2 below.

Next we solve the system (1) with $(x, s) = (\hat{x}^k, \hat{s}^k) \in \mathcal{N}(\beta + \tau)$, $\mu = (\hat{x}^k)^T \hat{s}^k / n$, and $\gamma = 1$, resulting in d_x^c and d_s^c. Let $x^{k+1} = \hat{x}^k + d_x^c$ and $s^{k+1} = \hat{s}^k + d_s^c$. It has been proved [9, 19] that $(x^{k+1}, s^{k+1}) \in \mathcal{N}(\beta)$ as long as $0 < \beta \leqslant \frac{1}{4}$ and $0 < \tau \leqslant \beta$. This is the corrector (or centering) step.

Observe that the algorithm generates a sequence of feasible points satisfying

$$\|X^k s^k / \mu^k - e\| \leqslant \beta \tag{3}$$

and

$$\begin{aligned} (\hat{x}^k)^T \hat{s}^k &= (1 - \theta^k)(x^k)^T s^k + (\theta^k)^2 (d_x^p)^T d_s^p, \\ (x^{k+1})^T s^{k+1} &= (\hat{x}^k)^T \hat{s}^k + (d_x^c)^T d_s^c. \end{aligned} \tag{4}$$

It has also been shown [9] that

$$\begin{aligned} (d_x^p)^T d_s^p &\leqslant (x^k)^T s^k / 4, \\ (d_x^c)^T d_s^c &\leqslant (\hat{x}^k)^T \hat{s}^k / (8n). \end{aligned} \tag{5}$$

For convenience, let $\delta^k = D_x^p d_s^p / \mu^k$, where $D_x^p = \operatorname{diag}(d_x^p)$, in the predictor step. Then, Mizuno et al. (Lemmas 1, 2 and 4 of [19]) showed that

$$\|\delta^k\| \leqslant \tfrac{1}{4}\sqrt{2}\, n. \tag{6}$$

Ji et al. [9] and Ye et al. [30] essentially developed the following lemma.

Lemma 2.2. *If θ^k is the largest $\theta^k \leqslant 1$ satisfying the conditions of Lemma 2.1 with $\alpha = \beta + \tau$ and $0 < \tau \leqslant \beta$, then $\theta^k \geqslant 2\tau / (\sqrt{\tau^2 + 4\tau \|\delta^k\|} + \tau)$.* □

Clearly, this lemma together with (4), (5) and (6) implies that the iteration complexity of the algorithm, with a suitable initialization, is $O(\sqrt{n}\, L)$ for a constant $0 < \tau \leqslant \beta$. Note that

$$\begin{aligned} 1 - \theta^k &\leqslant 1 - \frac{2\tau}{\sqrt{\tau^2 + 4\tau\|\delta^k\|} + \tau} = \frac{\sqrt{\tau^2 + 4\tau\|\delta^k\|} - \tau}{\sqrt{\tau^2 + 4\tau\|\delta^k\|} + \tau} \\ &= \frac{4\tau\|\delta^k\|}{(\sqrt{\tau^2 + 4\tau\|\delta^k\|} + \tau)^2} \leqslant \frac{\|\delta^k\|}{\tau}. \end{aligned} \tag{7}$$

Relations (4), (5), (6), and (7), and Lemma 2.2, imply

$$\mu^{k+1} \leqslant (1+1/(8n)) \left(\frac{\|\delta^k\|}{\tau} \mu^k + (d_x^{\rm p})^{\rm T} d_s^{\rm p}/n \right)$$
$$\leqslant (1+1/(8n)) \left(\frac{\|D_x^{\rm p} d_s^{\rm p}\|}{\tau} + (d_x^{\rm p})^{\rm T} d_s^{\rm p}/n \right). \tag{8}$$

From (8), we see that if

$$\|d_x^{\rm p}\| = {\rm O}(\mu^k) \quad \text{and} \quad \|d_s^{\rm p}\| = {\rm O}(\mu^k),$$

then the complementarity slackness sequence converges to zero Q-quadratically. (Here and in what follows the "big O" notation represents a positive quantity that may depend on n and/or the LP original data, but which is independent of the iteration k.) This fact was first established for the LP case in Ye et al. [29], and will be established for LCP in the next section.

3. Technical results

For a LCP possessing a strictly complementary solution, a unique partition B and N, where $B \cap N = \{1, 2, \ldots, n\}$ and $B \cup N = \emptyset$, exists such that $x_N^* = 0$ and $s_B^* = 0$ in every complementarity solution and at least one complementarity solution has $x_B^* > 0$ and $s_N^* > 0$. Güler and Ye [6] have shown that for all k, relation (3) implies that

$$\xi \leqslant x_j^k \leqslant 1/\xi \quad \text{for } j \in B, \qquad \xi \leqslant s_j^k \leqslant 1/\xi \quad \text{for } j \in N, \tag{9}$$

where $0 < \xi < 1$ is a fixed positive number that is independent of k.

We now introduce several technical lemmas. For simplicity, we drop the index k and recall the linear system during the predictor step

$$\begin{aligned} Xd_s + Sd_x &= -Xs, \\ d_s &= Md_x. \end{aligned} \tag{10}$$

Let $\mu = x^{\rm T} s/n$, $z = Xs$ and $Z = {\rm diag}(z)$. Note from (3) that we must have

$$(1-\beta)\mu \leqslant z_j \leqslant (1+\beta)\mu \quad \text{for } j = 1, 2, \ldots, n.$$

Define $D = X^{1/2} S^{-1/2}$. We now estimate $\|d_x\|$ and $\|d_s\|$. We start by characterizing the solution to (10).

Lemma 3.1. *If d_x and d_s are obtained from the linear system (10), and $\mu = x^{\rm T} s/n$, then*

$$\|D^{-1} d_x\| \leqslant \|(XS)^{1/2} e\| = \sqrt{n\mu}, \qquad \|Dd_s\| \leqslant \|(XS)^{1/2} e\| = \sqrt{n\mu}.$$

Proof. The proof is straightforward, e.g., see Kojima et al. [14]. □

Lemma 3.2. *If d_x and d_s are obtained from the linear system (10), and $\mu = x^{\rm T} s/n$, then*

$$\|(d_x)_N\| = {\rm O}(\mu) \quad \text{and} \quad \|(d_s)_B\| = {\rm O}(\mu).$$

Proof. From Lemma 3.1 and (9), we obtain

$$\|(d_x)_N\| = \|D_N D_N^{-1}(d_x)_N\| \leqslant \|D_N\| \|D_N^{-1}(d_x)_N\| \leqslant \|D_N\| \mathrm{O}(\sqrt{\mu})$$
$$= \|Z_N^{1/2} S_N^{-1}\| \mathrm{O}(\sqrt{\mu}) \leqslant \|Z_N^{1/2}\| \mathrm{O}(1/\xi)\mathrm{O}(\sqrt{\mu})$$
$$= \mathrm{O}(\sqrt{\mu})\mathrm{O}(\sqrt{\mu}) = \mathrm{O}(\mu).$$

This proves that $\|(d_x)_N\| = \mathrm{O}(\mu)$. The proof that $\|(d_s)_B\| = \mathrm{O}(\mu)$ is similar. □

The proofs of $\|(d_x)_B\| = \mathrm{O}(\mu)$ and $\|(d_s)_N\| = \mathrm{O}(\mu)$ are more involved. Towards this end, we first note

$$S(x+d_x) = -Xd_s, \qquad X(s+d_s) = -Sd_x,$$

and therefore

$$x + d_x = -(XS^{-1})d_s = -D^2 d_s, \qquad s + d_s = -(X^{-1}S)d_x = -D^{-2}d_x. \tag{11}$$

Before proceeding, we need some results regarding (non-symmetric) positive semi-definite (p.s.d.) matrices that may be of independent interest. In what follows, we will consider M to be partitioned (following a re-ordering of rows and columns) as

$$M = \begin{pmatrix} M_{BB} & M_{BN} \\ M_{NB} & M_{NN} \end{pmatrix}. \tag{12}$$

Lemma 3.3. *Let M be a p.s.d. matrix, partitioned as in* (12). *Then $M_{BB}x_B = 0$ if and only if $M_{BB}^{\mathrm{T}} x_B = 0$. Furthermore, $M_{BB}x_B = 0$ implies that $(M_{NB} + M_{BN}^{\mathrm{T}})x_B = 0$.*

Proof. Let $x = (x_B^{\mathrm{T}}, 0^{\mathrm{T}})^{\mathrm{T}}$. If either $M_{BB}x_B = 0$ or $M_{BB}^{\mathrm{T}} x_B = 0$, then $x^{\mathrm{T}}Mx = 0$, so x is a global minimizer of the quadratic form $y^{\mathrm{T}}My$. Consequently $(M + M^{\mathrm{T}})x = 0$, which is exactly

$$(M_{BB} + M_{BB}^{\mathrm{T}})x_B = 0, \qquad (M_{NB} + M_{BN}^{\mathrm{T}})x_B = 0. \qquad \square$$

Lemma 3.4. *Let M be a p.s.d. matrix, partitioned as in* (12). *Then*

$$R\begin{pmatrix} M_{BB} & M_{BN} \\ & I \end{pmatrix} = R\begin{pmatrix} M_{BB}^{\mathrm{T}} & M_{NB}^{\mathrm{T}} \\ & -I \end{pmatrix},$$

where $R(\cdot)$ denotes the range of a matrix.

Proof. From the fundamental theorem of linear algebra, it is equivalent to prove that

$$N\begin{pmatrix} M_{BB}^{\mathrm{T}} & \\ M_{BN}^{\mathrm{T}} & I \end{pmatrix} = N\begin{pmatrix} M_{BB} & \\ M_{NB} & -I \end{pmatrix},$$

where $N(\cdot)$ denotes the nullspace of a matrix. To begin, assume that

$$\begin{pmatrix} M_{BB}^{\mathrm{T}} & \\ M_{BN}^{\mathrm{T}} & I \end{pmatrix} \begin{pmatrix} x_B \\ x_N \end{pmatrix} = 0. \tag{13}$$

From Lemma 3.3, $M_{BB}x_B = 0$. Also $x_N = -M_{BN}^{\mathrm{T}} x_B$, so showing that $M_{NB}x_B - x_N = 0$ is equivalent to showing that $(M_{NB} + M_{BN}^{\mathrm{T}})x_B = 0$, which also holds by Lemma 3.3. Thus

$$\begin{pmatrix} M_{BB} & \\ M_{NB} & -I \end{pmatrix} \begin{pmatrix} x_B \\ x_N \end{pmatrix} = 0. \tag{14}$$

The argument that (14) implies (13) is similar. □

Lemma 3.5. *If d_x and d_s are obtained from the linear system* (10), *and $\mu = x^{\mathrm{T}}s/n$, then $u=(d_x)_B$ and $v=(d_s)_N$ are the solutions to the (weighted) least-squares problem*

$$\begin{aligned} \min_{u,\,v} \quad & \tfrac{1}{2}\|D_B^{-1}u\|^2 + \tfrac{1}{2}\|D_N v\|^2 \\ \text{s.t.} \quad & M_{BB}u = -M_{BN}(d_x)_N + (d_s)_B, \\ & M_{NB}u - v = -M_{NN}(d_x)_N. \end{aligned} \tag{15}$$

Proof. Note that from (10), $u=(d_x)_B$, $v=(d_s)_N$ is certainly feasible in the problem (15). Next, from (10) and (11), we see that

$$x_B + (d_x)_B = -D_B^2 M_{B\cdot} d_x, \qquad s_N + (d_s)_N = -D_N^{-2}(d_x)_N. \tag{16}$$

Since $s_B^* = 0$ for all optimal s^*, with $x_N^* = 0$, we must have $q_B = -M_{BB}x_B^* \in R(M_{BB})$. Therefore,

$$D_B^{-2}x_B = s_B = M_{B\cdot}x + q_B = M_{BB}(x_B - x_B^*) + M_{BN}x_N.$$

Substituting this into the first equation of (16) obtains

$$D_B^{-2}(d_x)_B = -M_{BB}(x_B - x_B^* + (d_x)_B) - M_{BN}(x_N + (d_x)_N). \tag{17}$$

Also $s_N = D_N^{-2}x_N$, which substituted into the second equation of (16) yields

$$D_N^2(d_s)_N = -x_N - (d_x)_N. \tag{18}$$

Then (17) and (18) together imply that

$$\begin{pmatrix} D_B^{-2}(d_x)_B \\ D_N^2(d_s)_N \end{pmatrix} \in R\begin{pmatrix} M_{BB} & M_{BN} \\ & I \end{pmatrix}.$$

Applying Lemma 3.4, we conclude that

$$\begin{pmatrix} D_B^{-2}(d_x)_B \\ D_N^2(d_s)_N \end{pmatrix} \in R\begin{pmatrix} M_{BB}^{\mathrm{T}} & M_{NB}^{\mathrm{T}} \\ & -I \end{pmatrix},$$

which shows exactly that $u=(d_x)_B$, $v=(d_s)_N$ satisfies the Karush–Kuhn–Tucker conditions for optimality in the least squares problem (15). □

The LP version of Lemma 3.5 was first established by Adler and Monteiro [1] and Witzgall et al. [26]. We are now ready to prove the following key result.

Theorem 3.6. *If d_x and d_s are obtained from the linear system* (10), *and $\mu = x^{\mathrm{T}}s/n$, then $\|d_x\| = \mathrm{O}(\mu)$ and $\|d_s\| = \mathrm{O}(\mu)$.*

Proof. Due to Lemma 3.2, we only need to prove

$$\|(d_x)_B\| = O(\mu) \quad \text{and} \quad \|(d_s)_N\| = O(\mu).$$

Since the least-squares problem (15) is always feasible, there must be *feasible* $\bar{u}$ and $\bar{v}$ such that

$$\|\bar{u}\| = O(\|(d_x)_N\| + \|(d_s)_B\|) \quad \text{and} \quad \|\bar{v}\| = O(\|(d_x)_N\| + \|(d_s)_B\|),$$

which together with Lemma 3.2 implies $\|\bar{u}\| = O(\mu)$ and $\|\bar{v}\| = O(\mu)$. Furthermore, from Lemma 3.5 and relations (3) and (9),

$$\begin{aligned}
&\|(d_x)_B\|^2 + \|(d_s)_N\|^2 \\
&\quad = \|D_B D_B^{-1}(d_x)_B\|^2 + \|D_N^{-1} D_N (d_s)_N\|^2 \\
&\quad \leqslant \|D_B^2\| \, \|D_B^{-1}(d_x)_B\|^2 + \|D_N^{-2}\| \, \|D_N(d_s)_N\|^2 \\
&\quad = \|Z_B^{-1} X_B^2\| \, \|D_B^{-1}(d_x)_B\|^2 + + \|Z_N^{-1} S_N^2\| \, \|D_N(d_s)_N\|^2 \\
&\quad \leqslant (\|Z_B^{-1} X_B^2\| + \|Z_N^{-1} S_N^2\|)(\|D_B^{-1}(d_x)_B\|^2 + \|D_N(d_s)_N\|^2) \\
&\quad \leqslant (\|Z_B^{-1} X_B^2\| + \|Z_N^{-1} S_N^2\|)(\|D_B^{-1}\bar{u}\|^2 + \|D_N \bar{v}\|^2) \\
&\quad \leqslant (\|Z_B^{-1} X_B^2\| + \|Z_N^{-1} S_N^2\|)(\|D_B^{-2}\| \, \|\bar{u}\|^2 + \|D_N^2\| \, \|\bar{v}\|^2) \\
&\quad \leqslant O(1/\mu)(\|D_B^{-2}\| \, \|\bar{u}\|^2 + \|D_N^2\| \, \|\bar{v}\|^2) \\
&\quad = O(\mu)(\|D_B^{-2}\| + \|D_N^2\|) \\
&\quad = O(\mu)(\|Z_B X_B^{-2}\| + \|Z_N S_N^{-2}\|) \\
&\quad = O(\mu^2). \qquad \square
\end{aligned}$$

4. Quadratic convergence

Theorem 3.6 indicates that at the kth predictor step, d_x^{p} and d_s^{p} satisfy

$$\|(d_x^{\mathrm{p}})\| = O(\mu^k) \quad \text{and} \quad \|(d_s^{\mathrm{p}})\| = O(\mu^k), \tag{19}$$

where $\mu^k = (x^k)^{\mathrm{T}} s^k / n$. We are now in a position to state our main result.

Theorem 4.1. *Let* $\{(x^k, s^k)\}$ *be the sequence generated by the algorithm. Then, with constants* $0 < \beta \leqslant \frac{1}{4}$ *and* $\alpha = 2\beta$:

(i) *The algorithm has iteration complexity* $O(\sqrt{n}\, L)$.
(ii) $(x^k)^{\mathrm{T}} s^k \to 0$ *Q-quadratically.*

Proof. The proof of (i) is due to Ji et al. [9]. This also establishes

$$\lim_{k \to \infty} \mu^k = 0.$$

The proof of (ii) directly follows from (8) and Theorem 3.6. □

According to the above analysis, we do not need to choose the largest step θ^k in the predictor step, but only the lower bound in Lemma 2.2. Thus, we are not required to find the roots of a quartic polynomial.

As we mentioned before, each iteration of the algorithm needs to solve two systems of the linear equations (1) — one in the predictor step and one in the corrector step. If we count each iteration as solving one system of linear equations, as we usually do in the analysis of interior-point algorithms, then the (average) order of convergence of the algorithm is only $\sqrt{2}$. Similar to the technique used for linear programming [28] we now show how to construct a modification of the algorithm whose Q-order of convergence is at least 2. Recall that the Q-order of convergence of the sequence to zero is defined as

$$\sup\left\{\sigma>1:\ \lim_{k\to\infty}\frac{(x^{k+1})^{\mathrm{T}}s^{k+1}}{((x^k)^{\mathrm{T}}s^k)^{\sigma}}=0\right\}$$

or

$$\inf\left\{\sigma>1:\ \lim_{k\to\infty}\frac{(x^{k+1})^{\mathrm{T}}s^{k+1}}{((x^k)^{\mathrm{T}}s^k)^{\sigma}}=\infty\right\}$$

(see Ortega and Rheinboldt [21] and Potra [22]). Potra further showed that the Q-order of convergence equals

$$\liminf_{k\to\infty}\frac{\log[(x^{k+1})^{\mathrm{T}}s^{k+1}]}{\log[(x^k)^{\mathrm{T}}s^k]}. \tag{20}$$

The convergence with Q-order 2 is also called Q-subquadratic convergence [21].

Variant 1. An iteration of the variant proceeds as follows. Given $(x^k, s^k)\in\mathcal{N}(\beta)$, we perform $T\geqslant 1$ successive predictor steps followed by one corrector step, where in the tth predictor step, $1\leqslant t\leqslant T$, we choose $\tau=\tau_t>0$ such that

$$\sum_{t=1}^{T}\tau_t=\beta. \tag{21}$$

In other words, on the tth predictor step of these T steps, we solve (1) with $(x, s)=(\hat{x}^k, \hat{s}^k)\in\mathcal{N}(\beta+\tau_1+\cdots+\tau_{t-1})$ (the initial $(\hat{x}^k, \hat{s}^k)=(x^k, s^k)\in\mathcal{N}(\beta)$) and $\gamma=0$. Denote by d_x^k and d_s^k the resulting directions. For some $\theta>0$ let

$$x(\theta)=\hat{x}^k+\theta d_x^k,\qquad s(\theta)=\hat{s}^k+\theta d_s^k.$$

Our specific choice for θ is similar as before: The largest θ^k such that (2) is satisfied for

$$\alpha=\beta+\tau_1+\cdots+\tau_{t-1}+\tau_t.$$

From (8) and Theorem 3.6, on the tth predictor step we have

$$x(\theta^k)^{\mathrm{T}}s(\theta^k)\leqslant R[(\hat{x}^k)^{\mathrm{T}}\hat{s}^k]^2/\tau_t \tag{22}$$

where $R \geqslant 1$ is a fixed number that is independent of the iteration count k. Now update $\hat{x}^k := x(\theta^k)$ and $\hat{s}^k := s(\theta^k)$.

After T predictor steps we have $(\hat{x}^k, \hat{s}^k) \in \mathcal{N}(2\beta)$. Now we perform one corrector step as before to generate

$$(x^{k+1}, s^{k+1}) \in \mathcal{N}(\beta).$$

Based on the previous results, each predictor step within an iteration achieves quadratic convergence for any positive constant sequence $\{\tau_t\}$ satisfying (21). For example, one natural choice would be $\tau_t = \beta/T$ for $t = 1, 2, \ldots, T$. Since each iteration solves $T+1$ systems of linear equations, the (average) Q-order of convergence of $(x^k)^{\mathrm{T}} s^k$ to zero in Variant 1 is $2^{T/(T+1)}$ per linear system solver for any constant $T \geqslant 1$.

We now develop a new variant where eventually we let $T = \infty$, that is, no corrector step is used on the remaining iterations of the algorithm. The algorithm becomes the pure Newton method or the primal–dual affine scaling algorithm (e.g., Kojima et al. [11] and Monteiro et al. [20]).

Variant 2. After $(x^K, s^K) \in \mathcal{N}(\beta)$ for some finite K, we perform only the predictor step, where we choose $\tau = \tau_t > 0$ satisfying (21). A particular choice which we will consider is

$$\tau_t = \beta(\tfrac{1}{2})^t \quad \text{for} \quad t = 1, 2, \ldots.$$

For simplicity, let us reset $K := 1$. Then, in the kth iteration we solve (1) with

$$(x, s) = (x^k, s^k) \in \mathcal{N}\left(\beta + \sum_{t=1}^{k-1} \tau_t\right),$$

and $\gamma = 0$. Denote by d_x^k and d_s^k the resulting directions. For some $\theta > 0$ let

$$x(\theta) = x^k + \theta d_x^k, \qquad s(\theta) = s^k + \theta d_s^k.$$

Our specific choice for θ is again the largest θ^k such that (2) is satisfied for

$$\alpha = \beta + \sum_{t=1}^{k} \tau_t.$$

Now directly update

$$x^{k+1} := x(\theta^k) \quad \text{and} \quad s^{k+1} := s(\theta^k).$$

Theorem 4.2. *Let $(x^K)^{\mathrm{T}} s^K > 0$ be sufficiently small such that*

$$0.5 \log_2((x^K)^{\mathrm{T}} s^K) + \log_2(R/\beta) + 1 \leqslant 0 \quad \textit{and} \quad 0.25 \log_2((x^K)^{\mathrm{T}} s^K) + 1 \leqslant 0,$$

where R is as in (22). Then, Variant 2 generates a sequence (x^k, s^k), with $k \geqslant K$ such that

(i) *The Q-order of convergence of $(x^k)^{\mathrm{T}} s^k$ to zero is at least 2;*

(ii) *$\{x^k, s^k\}$ is a Cauchy, and therefore convergent, sequence.*

Before we prove the theorem, we prove a technical lemma.

Lemma 4.3. *Let the sequence* $\{\Gamma^k\}$, $k \geqslant 1$, *be generated as follows*:

$$\Gamma^{k+1} \leqslant 2\Gamma^k + L' + k$$

where $L' > 0$ *is a fixed number. Furthermore, let* Γ^1 *be given such that*

$$0.5\Gamma^1 + L' + 1 \leqslant 0 \quad \text{and} \quad 0.25\Gamma^1 + 1 \leqslant 0.$$

Then, for all $k \geqslant 1$,

$$\Gamma^{k+1} \leqslant 1.5\Gamma^k$$

and

$$0.5\Gamma^{k+1} + L' + (k+1) \leqslant 0.$$

Proof. We use mathematical induction to prove the lemma. For $k = 1$, the result is obviously true due to the choice of Γ^1. Assume the result holds for k. Then we have

$$\Gamma^{k+1} = 2\Gamma^k + L' + k = 1.5\Gamma^k + 0.5\Gamma^k + L' + k \leqslant 1.5\Gamma^k,$$

and

$$\begin{aligned} 0.5\Gamma^{k+1} + L' + (k+1) &\leqslant 0.75\Gamma^k + L' + (k+1) \\ &= (0.5\Gamma^k + L' + k) + 0.25\Gamma^k + 1 \\ &\leqslant 0.25\Gamma^1 + 1 \leqslant 0. \end{aligned}$$

This concludes the induction. □

Proof of Theorem 4.2. Again we reset $K := 1$. At the kth iteration $(k \geqslant 1)$ we have from (22)

$$(x^{k+1})^{\mathrm{T}} s^{k+1} \leqslant \frac{R[(x^k)^{\mathrm{T}} s^k]^2}{\tau_k} = R[(x^k)^{\mathrm{T}} s^k]^2 2^k / \beta,$$

or

$$\log_2[(x^{k+1})^{\mathrm{T}} s^{k+1}] \leqslant 2 \log_2[(x^k)^{\mathrm{T}} s^k] + \log_2(R/\beta) + k. \tag{23}$$

Using Lemma 4.3 with $\Gamma^k = \log_2[(x^k)^{\mathrm{T}} s^k] < 0$ and $L' = \log_2(R/\beta)$, we conclude that $\{\log_2[(x^k)^{\mathrm{T}} s^k]\}$ is a geometric sequence, bounded above by $\{-4(1.5)^{k-1}\}$, tending to $-\infty$. Since L' is fixed and k is just an arithmetic sequence, we must have

$$\lim_{k\to\infty} \frac{L' + k}{|\log_2[(x^k)^{\mathrm{T}} s^k]|} \leqslant \lim_{k\to\infty} \frac{L' + k}{4(1.5)^{k-1}} \to 0. \tag{24}$$

Then (23) and (24) together imply that

$$\liminf_{k\to\infty} \frac{\log[(x^{k+1})^{\mathrm{T}} s^{k+1}]}{\log[(x^k)^{\mathrm{T}} s^k]} \geqslant 2.$$

(Note $\log[(x^k)^{\mathrm{T}} s^k] < 0$ so that the inequality direction is reversed here.) Then, using (20) proves (i) of the theorem. Now from Theorem 3.6,

$$\|x^{k+1}-x^k\| = \theta^k \|d_x^k\| < \|d_x^k\| = O(\mu^k)$$

and

$$\|s^{k+1}-s^k\| = \theta^k \|d_s^k\| < \|d_s^k\| = O(\mu^k).$$

Hence, $\{x^k, s^k\}$ must be a Cauchy sequence, because $\{\mu^k = (x^k)^T s^k / n\}$ converges to zero superlinearly from (i). This proves (ii). □

To actually achieve the Q-order 2 of convergence of the gap $(x^k)^T s^k$, we need to decide when to start the primal–dual affine scaling procedure described in Variant 2. Although R in Theorem 4.2 is unknown, we can start the procedure when $0.25 \log_2[(x^K)^T s^K] + 1 \leqslant 0$ or $\log_2[(x^K)^T s^K] \leqslant -4$. Again for simplicity, let $K := 1$. Then we add a *safety check* to see if for $k = 1, 2, \ldots,$

$$\log[(x^{k+1})^T s^{k+1}] \leqslant 1.5 \log[(x^k)^T s^k]. \tag{25}$$

(We choose 1.5 in (25) because it is used in Lemma 4.3. Actually, 1.5 can be replaced by any positive constant strictly between 1 and 2 to guarantee that $(x^k)^T s^k$ converges to zero Q-subquadratically.) If (25) is satisfied, we *continue* the primal–dual affine scaling procedure. Otherwise we conclude that $(x^K)^T s^K$ was not sufficiently small, do one *corrector step*, and then *restart* the primal–dual affine scaling procedure. This safety check will guarantee that the algorithm maintains $O(\sqrt{n}\, L)$ polynomial complexity, and achieves the Q-order 2 of convergence of the gap to zero, since eventually no corrector (or centering) step will be needed, according to Theorem 4.2 and Lemma 4.3.

Note that we have now shown that after the gap $(x^k)^T s^k$ becomes sufficiently small, the pure primal–dual affine scaling algorithm, or Newton method, with the step size choice in Variant 2 generates an iteration sequence not only polynomially converging to an optimal solution pair, but one whose convergence is at least Q-subquadratic. It is also interesting to see that the step parameter θ^k of the primal–dual affine scaling procedure, or Newton method, in Variant 2 converges to 1 superlinearly, while the solution sequence $\{x^k, s^k\}$ remains "centered" without any explicit centering. For linear programming (M is skew-symmetric), if the step size equals 1, then from (4) the new iterate is a complementary solution and it must hit the boundary of the feasible region Ω. Thus, our step size eventually becomes larger than the step-size choice commonly used in practice: a fixed fraction (say 0.99, or 0.9995) of the way to the boundary.

Note that although the Q-order of convergence of the gap sequence to zero in Variant 2 is at least 2, this sequence fails to meet the standard quadratic convergence criterion, since possibly

$$\limsup_{k\to\infty} \frac{(x^{k+1})^T s^{k+1}}{[(x^k)^T s^k]^2} = \infty.$$

5. Concluding remarks

Recently, several researchers have proved that an interior-point algorithm, while maintaining $O(\sqrt{n}\,L)$ iteration complexity, exhibits quadratic convergence for LP without the assumption of nondegeneracy or the assumption that the iteration sequence converges. In this paper we have demonstrated a similar result for monotone LCP, which includes convex QP. As we see in the above analyses, such an extension is not trivial due to some fundamental differences between LP and LCP. One of these differences is the guaranteed existence of a strictly complementary solution. A related question is whether or not assumption (A3) can be removed in our analysis. In the following we show a negative result:

Proposition 5.1. *There is a monotone LCP problem, where a strictly complementary solution does not exist, for which the predictor–corrector algorithm or affine scaling algorithm possesses no superlinear convergence.*

Proof. Consider the simple monotone LCP with $n=1$, $M=1$ and $q=0$. The unique complementarity solution is $s=x=0$, which is not strictly complementary. Note, the feasible solution $s=x=\varepsilon$ is a perfectly centered pair for any $\varepsilon>0$. The direction in the predictor step (or affine scaling algorithm) is

$$d_x=-\tfrac{1}{2}x \quad \text{and} \quad d_s=-\tfrac{1}{2}s.$$

Thus, even taking the step size $\theta=1$, the new solution will be $s=x=\frac{1}{2}\varepsilon$. Thus, the complementarity slackness sequence is reduced at most linearly, with constant $\frac{1}{4}$, which proves the proposition. □

References

[1] I. Adler and R. Monteiro, "Limiting behavior of the affine scaling continuous trajectories for linear programming problems," *Contemporary Mathematics* 114 (1990) 189–211.

[2] D.A. Bayer and J.C. Lagarias, "Karmarkar's linear programming algorithm and Newton's method," *Mathematical Programming* 50 (1991) 291–332.

[3] T.F. Coleman and Y. Li, "A quadratically-convergent algorithm for the linear programming problem with lower and upper bounds," in: T.F. Coleman and Y. Li, eds., *Large-Scale Numerical Optimization* (SIAM, Philadelphia, PA, 1990) pp. 49–57.

[4] G. de Ghellinck and J.-P. Vial, "As polynomial Newton method for linear programming," *Algorithmica* 1 (1986) 425–454.

[5] C.C. Gonzaga and M.J. Todd, "An $O(\sqrt{n}\,L)$-iteration large-step primal–dual affine algorithm for linear programming," *SIAM Journal on Optimization* 2 (1992) 349–359.

[6] O. Güler and Y. Ye, "Convergence behavior of interior-point algorithms," *Mathematical Programming* 60 (1993) 215–228.

[7] A.J. Hoffman, "On approximate solutions of systems of linear inequalities," *Journal of Research of the National Bureau of Standards* 49 (1952) 263–265.

[8] M. Iri and H. Imai, "A multiplicative barrier function method for linear programming," *Algorithmica* 1 (1986) 455–482.

[9] J. Ji, F. Potra and S. Huang, "A predictor-corrector method for linear complementarity problems with polynomial complexity and superlinear convergence," Department of Mathematics, The University of Iowa (Iowa City, IA, 1991).

[10] J. Ji, F. Potra, R.A. Tapia and Y. Zhang, "An interior-point method with polynomial complexity and superlinear convergence for linear complementary problems," Department of Mathematics, The University of Iowa (Iowa City, IA, 1991); also TR91-23, Department of Mathematical Sciences, Rice University (Houston, TX, 1991).

[11] M. Kojima, Y. Kurita and S. Mizuno, "Large step interior-point algorithms for linear complementarity problems," Research Report B-243, Department of Information Sciences, Tokyo Institute of Technology (Tokyo, Japan, 1991).

[12] M. Kojima, N. Megiddo and S. Mizuno, "Theoretical convergence of large-step primal–dual interior point algorithms for linear programming," *Mathematical Programming* 59 (1993) 1–21.

[13] M. Kojima, N. Megiddo and T. Noma, "Homotopy continuation methods for nonlinear complementarity problems," to appear in: *Mathematics of Operations Research.*

[14] M. Kojima, S. Mizuno and A. Yoshise, "A polynomial-time algorithm for a class of linear complementarity problems," *Mathematical Programming* 44 (1989) 1–26.

[15] Z.Q. Luo and P. Tseng, "Error bounds and convergence analysis of matrix splitting algorithms for the affine variational inequality problem," *SIAM Journal on Optimization* 2 (1992) 43–54.

[16] K. McShane, "A superlinearly convergent $O(\sqrt{n}\, L)$ iteration primal–dual linear programming algorithm," to appear in: *SIAM Journal on Optimization.*

[17] K. McShane, "A superlinearly convergent $O(\sqrt{n}\, L)$ iteration interior point algorithm for LCP," to appear in: *SIAM Journal on Optimization.*

[18] S. Mehrotra, "Quadratic convergence in a primal–dual method," *Mathematics of Operations Research* 18 (1993) 741–751.

[19] S. Mizuno, M.J. Todd and Y. Ye, "On adaptive-step primal–dual interior-point algorithms for linear programming," to appear in: *Mathematics of Operations Research.*

[20] R.C. Monteiro, I. Adler and M.G.C. 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 (1990) 191–214.

[21] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables* (Academic Press, New York, 1970).

[22] F.A. Potra, "On Q-order and R-order of convergence," *Journal of Optimization Theory and Applications* 63 (1989) 415–431.

[23] G. Sonnevend, J. Stoer and G. Zhao, "On the complexity of following the central path by linear extrapolation in linear programs," to appear in: U. Rieder and P. Kleinschmidt, eds., *Proceedings of the 14-th Symposium on Operations Research* (Ulm, 1989).

[24] T. Tsuchiya, "Quadratic convergence of Iri and Imai's algorithm for degenerate linear programming problems," manuscript, The Institute of Statistical Mathematics (Tokyo, Japan, 1991).

[25] T. Tsuchiya and K. Tanabe, "Local convergence properties of new methods in linear programming," *Journal of the Operations Research Society of Japan* 33 (1990) 22–45.

[26] C. Witzgall, P.T. Boggs and P.D. Domich, "On the convergence behavior of trajectories for linear programming," *Contemporary Mathematics* 114 (1990) 161–187.

[27] H. Yamashita, "A polynomially and quadratically convergent method for linear programming," manuscript, Mathematical Systems Institute (Tokyo, Japan, 1986).

[28] Y. Ye, "On the Q-order of convergence of interior-point algorithms for linear programming," in: Wu Fang, ed., *Proceedings of the 1992 Symposium on Applied Mathematics* (Institute of Applied Mathematics, Chinese Academy of Sciences, 1992) pp. 534–543.

[29] 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 (1993) 151–162.

[30] Y. Ye, R.A. Tapia and Y. Zhang, "A superlinearly convergent $O(\sqrt{n}\, L)$-iteration algorithm for linear programming," TR91-22, Department of Mathematical Sciences, Rice University (Houston, TX, 1991).

[31] Y. Zhang, and R.A. Tapia, "A quadratically convergent polynomial primal–dual interior-point algorithm for linear programming," *SIAM Journal on Optimization* 3 (1993) 118–133.

[32] Y. Zhang, R.A. Tapia and J.E. Dennis, "On the superlinear and quadratic convergence of primal–dual interior-point linear programming algorithms," *SIAM Journal on Optimization* 2 (1992) 304–323.

[33] Y. Zhang, R.A. Tapia and F. Potra, "On the superlinear convergence of interior-point algorithms for a general class of problems," *SIAM Journal on Optimization* 3 (1993) 413–422.