PREPRINT MCS-P331-1092, MCS DIVISION, ARGONNE NATIONAL LABORATORY

An Infeasible-Interior-Point Algorithm for Linear Complementarity Problems*

Stephen Wright†

October 15, 1992

Abstract

In this paper, we discuss a polynomial and Q-subquadratically convergent algorithm for linear complementarity problems that does not require feasibility of the initial point or the subsequent iterates. The algorithm is a modification of the linearly convergent method of Zhang and requires the solution of at most two linear systems with the same coefficient matrix at each iteration.

1 Introduction

The linear complementarity problem is to find a vector pair $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ such that

$$y = Mx + h, \qquad (x, y) \geq (0, 0), \qquad x^T y = 0, \tag{1}$$

where $h \in \mathbb{R}^n$ and M is an $n \times n$ positive semidefinite matrix. A pair (x, y) is said to be *feasible* if $y = Mx + h$ and $(x, y) \geq (0, 0)$, and *strictly feasible* if the latter inequality is strict.

It is well known that convex quadratic programming problems and linear programming problems can be expressed as linear complementarity problems; the same is true of extended linear-quadratic programming problems (see Rockafellar [9]). Much research has been devoted to interior point methods for (1). Recently, Ji, Potra, and Huang [1] proposed a predictor-corrector algorithm with polynomial complexity and superlinear convergence, while the predictor-corrector algorithm of Ye and Anstreicher [11] is polynomial and Q-quadratic. In the latter paper, it was assumed only that a strictly feasible point and a strictly complementary solution (one for which $\max(x_i^*, y_i^*) > 0$ for $i = 1, \cdots, n$) exist for (1). Both these algorithms generate a sequence of strictly feasible iterates (x^k, y^k). A strictly feasible starting point (x^0, y^0) must therefore be supplied. To find such a point, one often must augment the problem in an artificial way.

*This research was supported by the Applied Mathematical Sciences subprogram of the Office of Energy Research, U. S. Department of Energy, under Contract W-31-109-Eng-38.

†Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439.

More recently, research has focused on algorithms that generate sequences (x^k, y^k) for which $(x^k, y^k) > 0$ but possibly $y^k \neq Mx^k + h$. These *infeasible-interior-point methods* more nearly reflect computational practice (see, for example, Lustig, Marsten, and Shanno [4]). Also, minor modifications to a solution to a "nearby" problem can produce an excellent starting point for the present problem — an advantage when the underlying problem to be solved is nonlinear. Infeasible algorithms for linear programming have been proposed by Kojima, Meggido, and Mizuno [2], Kojima, Mizuno, and Todd [3], and Potra [7, 8]. Of these, only Potra's algorithms have both polynomial complexity and superlinear convergence properties. Potra's methods are of the predictor-corrector type and require three systems of linear equations (two of which have the same coefficient matrix) to be solved at each iteration. The only infeasible-interior-point algorithm for more general problems than linear programs that we are aware of is due to Zhang [12]. He analyzes an algorithm for a class of problems that includes (1) and proves Q-linear convergence of the complementarity gap $\mu_k = (x^k)^T y^k / n$ to zero. Polynomial complexity is obtained for a particular choice of starting point. The algorithm requires the solution of a single system of linear equations at each iteration.

In this paper, we propose modifications of Zhang's algorithm that retain polynomial complexity and have the added feature that the sequence $\{\mu_k\}$ converges superlinearly to zero with Q-order 2. Only the mild assumptions of Ye and Anstreicher [11] are required. Our method requires the solution of at most two linear systems of equations with the same coefficient matrix at each iteration.

When this report was about to be issued, we received a new report by Zhang and Zhang [13] that describes an infeasible-interior-point algorithm that is similar to ours in some respects. They allow relaxed versions of the centering condition and the feasibility dominance condition (cf. below (4e) and (4d), respectively) to be used on some iterations, and they obtain similar convergence properties. However, their algorithm is applicable only to linear programming problems.

Our algorithm is specified in Section 2. Some technical results are proved in Section 3, while in Section 4, we prove Q-linear convergence and polynomial complexity. Results concerning boundedness of the steps are proved in Section 5. Finally, superlinear convergence properties are discussed in Section 6.

Unless otherwise specified, $\|\cdot\|$ denotes the Euclidean norm of a vector. Iteration numbers appear as superscripts on vectors and matrices and as subscripts on scalars.

2 The Algorithm

Given a starting point with $(x^0, y^0) > (0, 0)$, the algorithm generates a sequence of iterates $(x^k, y^k) > (0, 0)$. The desirability of each point is measured by the merit function

$$\phi(x, y) = x^T y + \|y - Mx - h\|$$

whose two terms measure the complementarity gap and infeasibility. Clearly, a vector pair (x^*, y^*) is a solution of (1) if and only if $(x^*, y^*) \geq (0, 0)$ and $\phi(x^*, y^*) = 0$. We use the shorthand notation ϕ_k to denote $\phi(x^k, y^k)$.

In order to describe the step between successive iterates, we define

$$\mu_k = (x^k)^T y^k / n, \qquad e = (1, 1, \cdots, 1)^T,$$

$$X^k = \mathrm{diag}(x_1^k, x_2^k, \cdots, x_n^k), \qquad Y^k = \mathrm{diag}(y_1^k, y_2^k, \cdots, y_n^k).$$

The step is calculated as follows.

Given $\tilde{\gamma} \in (0, 1)$, $\tilde{\beta} \in [0, 1)$, $\tilde{\sigma} \in [0, 1)$, solve

$$\begin{bmatrix} M & -I \\ Y^k & X^k \end{bmatrix} \begin{bmatrix} \Delta x^k \\ \Delta y^k \end{bmatrix} = \begin{bmatrix} -h - Mx^k + y^k \\ -X^k Y^k e + \tilde{\sigma} \mu_k e \end{bmatrix}. \tag{2}$$

Choose

$$\tilde{\alpha} = \arg \min_{\alpha} \phi(x^k + \alpha \Delta x^k, y^k + \alpha \Delta y^k) \tag{3}$$

subject to

$$\begin{aligned}
\alpha &\in [0, 1], && \text{(4a)} \\
x^k + \alpha \Delta x^k &> 0, && \text{(4b)} \\
y^k + \alpha \Delta y^k &> 0, && \text{(4c)} \\
(x^k + \alpha \Delta x^k)^T (y^k + \alpha \Delta y^k) &\geq (1 - \tilde{\beta})(1 - \alpha)(x^k)^T y^k, && \text{(4d)} \\
(x_i^k + \alpha \Delta x_i^k)(y_i^k + \alpha \Delta y_i^k) &\geq (\tilde{\gamma}/n)(x^k + \alpha \Delta x^k)^T (y^k + \alpha \Delta y^k), \qquad i = 1, \cdots, n. && \text{(4e)}
\end{aligned}$$

It has been noted previously that (2) are simply the equations obtained by applying one iteration of Newton's method to the nonlinear equations

$$F(x, y) = \begin{bmatrix} y - Mx - h \\ XYe \end{bmatrix} = \begin{bmatrix} 0 \\ \tilde{\sigma} \mu_k e \end{bmatrix},$$

starting from the point (x^k, y^k). The condition (4e), usually referred to as a *centering condition*, ensures that the iterates do not prematurely approach the edge of the non-negative orthant. The condition (4d) is a relaxation of the condition enforced by Zhang [12, formula (5.7)] to ensure that feasibility is given a higher priority than complementarity (Zhang uses $\tilde{\beta} = 0$). Potra's algorithms replace (4d) with an *equality* condition in which $\tilde{\beta} = 0$. We allow $\tilde{\beta} > 0$ to permit superlinear convergence, as will become clear in the analysis that follows.

We can now state our algorithm.

Given $\bar{\gamma} \in (0, 1/2)$, $\sigma \in (0, 1/2)$, $(x^0, y^0) > (0, 0)$,
$\rho \in (0, \bar{\gamma})$, $\bar{\phi} > 0$, and (x^0, y^0) with $x_i^0 y_i^0 \geq 2\bar{\gamma}\mu_0$;

$t_0 \leftarrow 1$, $\gamma_0 \leftarrow 2\bar{\gamma}$;

for $k = 0, 1, 2, \cdots$

if $\phi_k = \phi(x^k, y^k) \leq \bar{\phi}$
then Solve (2)–(4) with $\tilde{\sigma} = \mu_k$, $\tilde{\beta} = \bar{\gamma}^{t_k}$, $\tilde{\gamma} = \bar{\gamma}(1 + \bar{\gamma}^{t_k})$;
if $\phi(x^k + \tilde{\alpha}\Delta x^k,\, y^k + \tilde{\alpha}\Delta y^k) \leq \rho\phi_k$
then $\alpha_k \leftarrow \tilde{\alpha}$, $\beta_k \leftarrow \tilde{\beta}$, $\sigma_k \leftarrow \tilde{\sigma}$, $\gamma_{k+1} \leftarrow \tilde{\gamma}$;
$t_{k+1} \leftarrow t_k + 1$;
$(x^{k+1}, y^{k+1}) \leftarrow (x^k, y^k) + \alpha_k(\Delta x^k, \Delta y^k)$;
go to next k;
end if
end if

Solve (2)–(4) with $\tilde{\sigma} \in [\sigma, 1/2]$, $\tilde{\beta} = 0$, $\tilde{\gamma} = \gamma_k$;
$\alpha_k \leftarrow \tilde{\alpha}$, $\beta_k \leftarrow 0$, $\sigma_k \leftarrow \tilde{\sigma}$, $\gamma_{k+1} \leftarrow \tilde{\gamma}$;
$t_{k+1} \leftarrow t_k$;
$(x^{k+1}, y^{k+1}) \leftarrow (x^k, y^k) + \alpha_k(\Delta x^k, \Delta y^k)$;
go to next k;

end for.

The idea behind this algorithm is simple. While the merit function exceeds $\bar{\phi}$, it is identical to Zhang's algorithm. One linear system of equations is solved at each iteration to yield what we refer to as a *safe* step. When the merit function falls below the threshold $\bar{\phi}$, the algorithm computes a step that may yield rapid convergence by setting σ_k equal to the current duality gap and allowing $\tilde{\beta} > 0$ in (4d). We refer to such a step as a *fast* step. The coefficient matrices for the fast and safe steps are the same, and the integer variable t_k keeps track of the number of fast steps that have been taken prior to iteration k. If the fast step does not give a significant reduction in ϕ (by at least a factor of ρ), it is discarded, and we take a safe step instead.

This strategy may lead to our rejecting a fast step and taking a safe step that gives a smaller decrease in ϕ at some iterations. Note, however, that we pay a price for taking fast steps in that γ is decreased, and so the Q-linear convergence rate will not be as favorable on subsequent iterations. Therefore, it makes sense to reject fast steps unless they make substantial progress.

In the safe-step calculation, there is considerable scope for user-defined heuristics in the choice of σ_k. In a practical implementation, the value of σ_k may be adjusted according to the merit function decrease $\phi_{k-1} - \phi_k$ on the previous step, and the value of α_{k-1}.

In the remainder of the paper, we analyze the convergence properties of this algorithm under the following assumptions:

Assumption 1 *M is positive semidefinite.*

This assumption implies that if (x^1, y^1) and (x^2, y^2) are any two points satisfying $y = Mx = h$, then

$$(y^2 - y^1)^T(x^2 - x^1) = (x^2 - x^1)^T M (x^2 - x^1) \geq 0. \tag{5}$$

Assumption 2 *The problem (1) has a strictly feasible point $(\bar{x}, \bar{y})$. That is, $(\bar{x}, \bar{y}) > (0, 0)$ and $\bar{y} = M\bar{x} + h$.*

Assumption 3 *The solution set for (1) is nonempty and, moreover, there is a strictly complementary solution (x^*, y^*).*

With respect to (x^*, y^*), we define the partitioning $\{1, 2, \cdots, n\} = N \cup B$, where

$$B = \{i \mid x_i^* > 0\}, \qquad N = \{i \mid y_i^* > 0\}.$$

Assumptions 1, 2, and 3 will be assumed everywhere without being explicitly stated.

3 Technical Results

In this section, we prove a number of results that are needed in the analysis of Sections 4, 5, and 6. In the statement of many of our results, we refer to the pair (x^k, y^k), which is always understood to be an iterate generated by the algorithm of Section 2.

We start by showing that the dominance of feasibility over complementarity is not completely abandoned by our relaxed condition (4d), but still holds to within a certain constant.

Lemma 3.1

$$\hat{\beta} \triangleq \prod_{k=0}^{\infty} (1 - \beta_k) > 0.$$

Proof. Inspection of the algorithm shows that on each iteration, we have either $\beta_k = 0$ (safe steps) or $\beta_k = \bar{\gamma}^t$, $t = 1, 2, \cdots$ (fast steps). Therefore

$$\hat{\beta} \geq \prod_{t=1}^{\infty} (1 - \bar{\gamma}^t),$$

and the right-hand side is bounded away from zero since $\bar{\gamma} \in (0, 1)$. ∎

If we define

$$\nu_0 = 1, \qquad \nu_k = \prod_{i=0}^{k-1} (1 - \alpha_i), \qquad k = 1, 2, \cdots,$$

the following result defines upper and lower bounds on μ_k.

Lemma 3.2 *For all (x^k, y^k) generated by the algorithm and $\mu_k = (x^k)^T y^k/n$,*

$$\begin{aligned} \mu_k &\geq \hat{\beta}\nu_k\mu_0, \\ \mu_k &\leq \mu_0 + \frac{1}{n}\|y^0 - Mx^0 - h\|. \end{aligned}$$

Proof. By (4d),

$$\mu_k \geq (1-\beta_{k-1})(1-\alpha_{k-1})\mu_{k-1} \geq \prod_{i=0}^{k-1}(1-\beta_i)\nu_k\mu_0 \geq \hat{\beta}\nu_k\mu_0,$$

giving the first inequality. The second inequality follows from

$$n\mu_k \leq \phi_k \leq \phi_0 = n\mu_0 + \|y^0 - Mx^0 - h\|.$$

■

Zhang [12] defines an "auxiliary sequence" (u^k, v^k) by defining an initial point (u^0, v^0) such that

$$(u^0, v^0) \leq (x^0, y^0), \qquad v^0 = Mu^0 + h,$$

and subsequent iterates by

$$u^{k+1} = u^k + \alpha_k(\Delta x^k + x^k - u^k), \qquad v^{k+1} = v^k + \alpha_k(\Delta y^k + y^k - v^k).$$

He proves the following result.

Lemma 3.3 *(Zhang [12, Lemma 4.1]) For $k \geq 0$,*

(i) $v^k = Mu^k + h$;

(ii) $x^{k+1} - u^{k+1} = \nu_k(x^0 - u^0) \geq 0$ *and* $y^{k+1} - v^{k+1} = \nu_k(y^0 - v^0) \geq 0$;

(iii) If $\alpha_K = 1$ for some $K \geq 0$, then $(x^k, y^k) = (u^k, v^k)$ and therefore (x^k, y^k) is strictly feasible for all $k > K$.

Using Lemma 3.3, we can bound some components of the iterates (x^k, y^k)

Lemma 3.4 *There is a constant $C_1 > 0$ such that for all iterates (x^k, y^k),*

$$\begin{aligned} i \in N &\Rightarrow x_i^k \leq C_1\mu_k, && (6a) \\ i \in B &\Rightarrow y_i^k \leq C_1\mu_k, && (6b) \end{aligned}$$

and

$$\begin{aligned} i \in B &\Rightarrow x_i^k \geq \bar{\gamma}/C_1, && (7a) \\ i \in N &\Rightarrow y_i^k \geq \bar{\gamma}/C_1. && (7b) \end{aligned}$$

Proof. From the definition of (u^k, v^k), we have

$$\begin{aligned}(x^k - x^*)^T(y^k - y^*) \\ &= (x^k - u^k + u^k - x^*)^T(y^k - v^k + v^k - y^*) \qquad (8) \\ &= (x^k - u^k)^T(y^k - v^k) + (u^k - x^*)^T(y^k - v^k) + (x^k - u^k)^T(v^k - y^*) + (u^k - x^*)^T(v^k - y^*).\end{aligned}$$

Now $v^k = Mu^k + h$ and $y^* = Mx^* + h$, so by (5),

$$(u^k - x^*)^T(v^k - y^*) \geq 0.$$

Therefore, since $(x^*)^T y^* = 0$, we have

$$\begin{aligned}&-(x^k)^T y^* - (x^*)^T y^k + (x^k)^T y^k \\ &\geq (x^k - u^k)^T(y^k - v^k) + (u^k - x^*)^T(y^k - v^k) + (x^k - u^k)^T(v^k - y^*),\end{aligned}$$

and so

$$\begin{aligned}&(x^*)^T y^k + (x^k)^T y^* \\ &\leq (x^k)^T y^k - (x^k - u^k)^T(y^k - v^k) - (u^k - x^*)^T(y^k - v^k) - (x^k - u^k)^T(v^k - y^*) \\ &= (x^k)^T y^k + (x^k - u^k)^T(y^k - v^k) - (x^k - x^*)^T(y^k - v^k) - (x^k - u^k)^T(y^k - y^*). \qquad (9)\end{aligned}$$

Now, by Lemma 3.3,

$$x^k - u^k \geq 0, \qquad y^k - v^k \geq 0, \qquad y^k \geq 0, \qquad x^k \geq 0,$$

and so

$$(y^k)^T(x^k - u^k) \geq 0, \qquad (x^k)^T(y^k - v^k) \geq 0.$$

By substitution in (9), we have

$$\begin{aligned}&(x^*)^T y^k + (x^k)^T y^* \\ &\leq (x^k)^T y^k + (x^k - u^k)^T(y^k - v^k) + (x^*)^T(y^k - v^k) + (y^*)^T(x^k - u^k) \\ &= (x^k)^T y^k \left[1 + \frac{(x^k - u^k)^T(y^k - v^k)}{(x^k)^T y^k} + \frac{(x^*)^T(y^k - v^k)}{(x^k)^T y^k} + \frac{(y^*)^T(x^k - u^k)}{(x^k)^T y^k}\right].\end{aligned}$$

Now, by Lemmas 3.2 and 3.3, we can bound the term in the square brackets by $\bar{C}_1$, where

$$\bar{C}_1 \triangleq \left[1 + \frac{(x^0 - u^0)^T(y^0 - v^0)}{\hat{\beta}(x^0)^T y^0} + \frac{(x^*)^T(y^0 - v^0)}{\hat{\beta}(x^0)^T y^0} + \frac{(y^*)^T(x^0 - u^0)}{\hat{\beta}(x^0)^T y^0}\right],$$

and so

$$(x^*)^T y^k + (x^k)^T y^* \leq n\bar{C}_1\mu_k.$$

Hence, for $i \in N$,

$$x_i^k y_i^* \leq n\bar{C}_1\mu_k \quad \Rightarrow \quad x_i^k \leq \frac{n\bar{C}_1}{y_i^*}\mu_k,$$

while for $i \in B$,

$$y_i^k \le \frac{n\bar{C}_1}{x_i^*}\mu_k.$$

Hence (6) is obtained by taking

$$C_1 = n\bar{C}_1 \max\left(\sup_{i\in B}\frac{1}{x_i^*},\ \sup_{i\in N}\frac{1}{y_i^*}\right).$$

For (7a), we simply note that

$$i \in B,\quad x_i^k y_i^k \ge \gamma_k \mu_k \quad\Rightarrow\quad x_i^k \ge \frac{\gamma_k \mu_k}{y_i^k} \ge \frac{\gamma_k}{C_1} \ge \frac{\bar{\gamma}}{C_1}.$$

The proof of (7b) is similar. ∎

Assumption 2 can be used to show that the iterates remain bounded.

Lemma 3.5 *There is a constant $C_2 > 0$ such that for $k \ge 0$ and $i = 1, 2, \cdots, n$,*

$$0 < x_i^k \le C_2, \qquad 0 < y_i^k \le C_2.$$

Proof. Because of Assumption 2, there is a strictly feasible point $(\bar{x}, \bar{y})$. Now

$$\begin{aligned}
&(x^k - \bar{x})^T(y^k - \bar{y}) \\
&\quad= (x^k - u^k + u^k - \bar{x})^T(y^k - v^k + v^k - \bar{y}) \\
&\quad= (x^k - u^k)^T(y^k - v^k) + (u^k - \bar{x})^T(y^k - v^k) + (x^k - u^k)^T(v^k - \bar{y}) \\
&\qquad +(u^k - \bar{x})^T(v^k - \bar{y}) \\
&\quad= (x^k - u^k)^T(y^k - v^k) + (u^k - x^k)^T(y^k - v^k) + (x^k - \bar{x})^T(y^k - v^k) \\
&\qquad +(x^k - u^k)^T(v^k - y^k) + (x^k - u^k)^T(y^k - \bar{y}) + (u^k - \bar{x})^T(v^k - \bar{y}) \\
&\quad= -(x^k - u^k)^T(y^k - v^k) + (x^k)^T(y^k - v^k) - \bar{x}^T(y^k - v^k) \\
&\qquad +(x^k - u^k)^T y^k - (x^k - u^k)^T\bar{y} + (u^k - \bar{x})^T(v^k - \bar{y}).
\end{aligned} \tag{10}$$

Now, since (u^k, v^k) and $(\bar{x}, \bar{y})$ are both feasible, the last term in (10) is non-negative. Moreover,

$$\begin{aligned}
x^k > 0,\ \ y^k - v^k \ge 0 \ &\Rightarrow\ (x^k)^T(y^k - v^k) \ge 0, \\
y^k > 0,\ \ x^k - u^k \ge 0 \ &\Rightarrow\ (y^k)^T(x^k - u^k) \ge 0, \\
(x^k - u^k) = \nu_k(x^0 - u^0), \qquad &\quad\ (y^k - v^k) = \nu_k(y^0 - v^0).
\end{aligned}$$

Hence, from (10),

$$\begin{aligned}
&(x^k - \bar{x})^T(y^k - \bar{y}) \\
&\quad\ge -(x^k - u^k)^T(y^k - v^k) - \bar{x}^T(y^k - v^k) - \bar{y}^T(x^k - u^k) \\
&\quad= -\nu_k^2(x^0 - u^0)^T(y^0 - v^0) - \nu_k\bar{x}^T(y^0 - v^0) - \nu_k\bar{y}^T(x^0 - u^0),
\end{aligned}$$

and so, using Lemma 3.2 to bound $(x^k)^T y^k$, we obtain

$$
\begin{aligned}
&\bar{x}^T y^k + \bar{y}^T x^k \\
&\quad \le (x^k)^T y^k + \bar{x}^T \bar{y} + (x^0 - u^0)^T (y^0 - v^0) + \bar{x}^T (y^0 - v^0) + \bar{y}^T (x^0 - u^0) \\
&\quad \le \phi_0 + \bar{x}^T \bar{y} + (x^0 - u^0)^T (y^0 - v^0) + \bar{x}^T (y^0 - v^0) + \bar{y}^T (x^0 - u^0) \\
&\quad \triangleq \bar{C}_2.
\end{aligned}
$$

Hence

$$
0 < y_i^k \le \frac{\bar{C}_2}{\bar{x}_i}, \qquad 0 < x_i^k \le \frac{\bar{C}_2}{\bar{y}_i}, \qquad i = 1, 2, \cdots, n.
$$

The result is obtained by setting

$$
C_2 = \bar{C}_2 \max \left(\sup_{i=1,\cdots,n} \frac{1}{\bar{x}_i}, \sup_{i=1,\cdots,n} \frac{1}{\bar{y}_i} \right).
$$

■

We can use Lemma 3.5 to define lower bounds on some other components of (x^k, y^k).

Lemma 3.6 *For all $k \ge 0$,*

$$
\begin{aligned}
i \in B &\Rightarrow y_i^k \ge \frac{1}{C_2} \bar{\gamma} \mu_k, \\
i \in N &\Rightarrow x_i^k \ge \frac{1}{C_2} \bar{\gamma} \mu_k,
\end{aligned}
$$

where C_2 is as defined in Lemma 3.5.

Proof. We have from Lemma 3.5 that

$$
i \in B, \ x_i^k y_i^k \ge \gamma_k \mu_k \Rightarrow y_i^k \ge \frac{\gamma_k \mu_k}{x_i^k} \ge \frac{\gamma_k \mu_k}{C_2} \ge \frac{\bar{\gamma} \mu_k}{C_2}.
$$

■

4 Linear Convergence and Polynomial Complexity

In this section, we modify some results of Zhang [12] to show that the algorithm of Section 2 produces a sequence $\{\phi_k\}$ that converges Q-linearly. When the starting point is chosen appropriately, the method has polynomial complexity.

We start with a result that can be used to derive global bounds on the step $(\Delta x^k, \Delta y^k)$. We define

$$
\begin{aligned}
\xi_k &= \left(\frac{n}{\gamma_k} \right)^{1/2} \frac{(x^k - u^k)^T y^k + (y^k - v^k)^T x^k}{(x^k)^T y^k} \\
\eta_k &= \left(1 - 2\sigma_k + \frac{\sigma_k^2}{\gamma_k} \right) + 2\nu_k \frac{(x^0 - u^0)^T (y^0 - v^0)}{\hat{\beta} (x^0)^T y^0} \\
\omega_k &= \left(\xi_k + \sqrt{\xi_k^2 + \eta_k} \right)^2,
\end{aligned}
$$

and the diagonal matrix

$$D^k = (X^k)^{-1/2}(Y^k)^{1/2}, \tag{11}$$

which also appears in much of the subsequent analysis.

Lemma 4.1 *For all k,*

$$\|D^k \Delta x^k\|^2 + \|(D^k)^{-1}\Delta y^k\|^2 \leq \omega_k (x^k)^T y^k. \tag{12}$$

Moreover, there is a constant $\omega > 0$

$$\omega_k \leq \omega. \tag{13}$$

Proof. Minor modifications of the proofs of Lemma 6.2 and Theorem 7.1 in Zhang [12] yield (12). Since, from Lemma 3.2, we have

$$(x^k)^T y^k \geq \hat{\beta}\nu_k (x^0)^T y^0,$$

we can modify the proof of Zhang [12, Lemma 6.1] to show that

$$\xi_k \leq \left(\frac{n}{\gamma_k}\right)^{1/2}\left[1 + \frac{(x^0-u^0)^T y^* + (y^0-v^0)^T x^* + (x^0-u^0)^T(y^0-v^0)}{\hat{\beta}(x^0)^T y^0}\right].$$

Since $\sigma_k \in [0,1)$ and $\gamma_k \geq \bar{\gamma} > 0$, it is easy to see that $\{\xi_k\}$ and $\{\eta_k\}$ are bounded sequences. Hence $\{\omega_k\}$ is also bounded, and so we have (13). ■

We can now prove linear convergence.

Theorem 4.2 *There is a constant $\delta \in (0,1)$ such that*

$$\phi_{k+1} \leq (1-\delta)\phi_k, \qquad k = 0,1,2,\cdots, \tag{14}$$

that is, the algorithm converges globally and Q-linearly.

Proof. Consider a safe step with $0 < \sigma \leq \sigma_k \leq 1/2$. As in the proof of Theorem 7.1 of Zhang [12], we can show that

$$\phi_{k+1} \leq (1-\delta_k)\phi_k,$$

where

$$\delta_k \geq \left(1 - \frac{2(1-\gamma_k)\sigma}{n}\right)\frac{(1-\gamma_k)\sigma}{n\omega} \geq \left(1 - \frac{2(1-\bar{\gamma})\sigma}{n}\right)\frac{(1-2\bar{\gamma})\sigma}{n\omega} > 0,$$

where the second inequality follows from $\gamma_k \in (\bar{\gamma}, 2\bar{\gamma}]$.

When a successful fast step is taken, we have by definition that

$$\phi_{k+1} \leq \rho\phi_k.$$

The result follows by setting

$$\delta = \min\left(\left(1 - \frac{2(1-\bar{\gamma})\sigma}{n}\right)\frac{(1-2\bar{\gamma})\sigma}{n\omega},\ 1-\rho\right). \tag{15}$$

■

The complexity result depends on a particular choice of starting point. Zhang [12] suggests the following choice. First, define

$$(u^0, v^0) = \arg \min_{(u,v)} \|u\|^2 + \|v\|^2, \qquad \text{subject to} \qquad v = Mu + h, \tag{16}$$

and

$$r^* = \inf \{ \|(x^*, y^*)\| \mid (x^*, y^*) \text{ is a solution of (1) } \}. \tag{17}$$

Now, we choose $r \geq \|(u^0, v^0)\|$ and set

$$(x^0, y^0) = r(e, e). \tag{18}$$

Theorem 4.3 *Suppose that (x^0, y^0) is defined by (16)–(18) and that there is a constant β independent of n such that*

$$r \geq r^*/(\beta\sqrt{n}).$$

Suppose that, for a given $\epsilon > 0$, we have

$$\phi_0 \leq 1/\epsilon^\tau,$$

where τ is a constant independent of n. Then there is an integer K_ϵ such that

$$K_\epsilon = O(n^2 \log(1/\epsilon)),$$

and $\phi_k \leq \epsilon$ for $k \geq K_\epsilon$.

Proof. As in Zhang [12, Lemma 7.1], with minor modifications, it can be shown that if we choose

$$\omega = \limsup_{k \to \infty} \omega_k,$$

then $\omega = O(n)$. Equation (15) then implies that $\delta \geq \bar{\delta}/n^2$, for some $\bar{\delta} > 0$ independent of n. Therefore $\phi_k \leq \epsilon$ when

$$k \geq \frac{\log(\epsilon/\phi_0)}{\log(1-\delta)} = \frac{O(\log(1/\epsilon))}{\delta}.$$

Hence $K_\epsilon = O(n^2 \log(1/\epsilon))$. ■

5 Boundedness of the Steps

In this section, we obtain bounds for components of the steps Δx^k and Δy^k in terms of γ_k, σ_k, and μ_k. These bounds are finer than the global bounds implied by (12) and are important for the analysis of superlinear convergence. We treat components in the basic and nonbasic index sets B and N differently and define

$$\Delta x_B^k = [\Delta x_i^k]_{i \in B}, \qquad \Delta x_N^k = [\Delta x_i^k]_{i \in N},$$

and similarly for Δy_B^k, Δy_N^k, etc. Without loss of generality, we can assume that the components of all vectors (and the rows and columns of M) are rearranged so that the indices in B are listed first. We can then partition M and D^k in an obvious way as

$$M = \begin{bmatrix} M_{BB} & M_{BN} \\ M_{NB} & M_{NN} \end{bmatrix}, \qquad D^k = \begin{bmatrix} D_B^k & \\ & D_N^k \end{bmatrix}.$$

(X^k and Y^k may be partitioned similarly to D^k.) Here, e_B and e_N denote vectors of the appropriate length whose components are all 1.

We start with a simple result that bounds components of Δx_N^k and Δy_B^k.

Lemma 5.1 *There is a constant C_4 such that*

$$i \in N \;\Rightarrow\; |\Delta x_i^k| \le C_4 \mu_k, \tag{19a}$$
$$i \in B \;\Rightarrow\; |\Delta y_i^k| \le C_4 \mu_k. \tag{19b}$$

Proof. From (12) and (13), we have for $i \in N$ that

$$\left| \Delta x_i^k \left(\frac{y_i^k}{x_i^k} \right)^{1/2} \right| \le (n \mu_k \omega)^{1/2},$$

and so

$$|\Delta x_i^k| \le (n\omega)^{1/2} \frac{\mu_k^{1/2}}{(x_i^k y_i^k)^{1/2}} x_i^k \le \frac{(n\omega)^{1/2}}{\gamma_k^{1/2}} x_i^k.$$

If we define $C_4 = C_1 (n\omega/\bar{\gamma})^{1/2}$, the inequality (19a) follows from Lemma 3.4. The proof of (19b) is similar. ■

Bounds on the remaining components Δx_B^k and Δy_N^k are more difficult to obtain. We need the following lemma, which is similar to results obtained by Ye and Anstreicher [11, Lemma 3.5] and Monteiro and Wright [5, Lemma 3.6]. For clarity, we drop the superscript k from vectors and matrices in the following two results.

Lemma 5.2 *The vector pair $(w, z) = (\Delta x_B, \Delta y_N)$ solves the convex quadratic programming problem*

$$\min_{(w,z)} \tfrac{1}{2} \|D_B w\|^2 - \sigma_k \mu_k e_B^T X_B^{-1} w + \tfrac{1}{2} \|D_N^{-1} z\|^2 - \sigma_k \mu_k e_N^T Y_N^{-1} z, \tag{20}$$

subject to

$$M_{BB} w = (-h - Mx + y)_B - M_{BN} \Delta x_N + \Delta y_B, \tag{21a}$$
$$M_{NB} w - z = (-h - Mx + y)_N - M_{NN} \Delta x_N. \tag{21b}$$

Proof. The first-order necessary conditions for (w, z) to solve (20),(21) are also sufficient, since the problem is convex. These conditions are that (w, z) is feasible with respect to (21) and

$$\begin{bmatrix} D_B^2 w - \sigma_k \mu_k X_B^{-1} e_B \\ D_N^{-2} z - \sigma_k \mu_k Y_N^{-1} e_N \end{bmatrix} \in R \left(\begin{bmatrix} M_{BB}^T & M_{NB}^T \\ 0 & -I \end{bmatrix} \right), \tag{22}$$

where $R(.)$ denotes the range space of a matrix. Ye and Anstreicher [11, Lemma 3.4] prove that

$$R\left(\begin{bmatrix} M_{BB}^T & M_{NB}^T \\ 0 & -I \end{bmatrix}\right) = R\left(\begin{bmatrix} M_{BB} & M_{BN} \\ 0 & I \end{bmatrix}\right),$$

and so (22) becomes

$$\begin{bmatrix} D_B^2 w - \sigma_k \mu_k X_B^{-1} e_B \\ D_N^{-2} z - \sigma_k \mu_k Y_N^{-1} e_N \end{bmatrix} \in R\left(\begin{bmatrix} M_{BB} & M_{BN} \\ 0 & I \end{bmatrix}\right). \tag{23}$$

We now show that $(\Delta x_B, \Delta y_N)$ satisfies (23). Equation (2), appropriately partitioned and scaled, yields

$$\begin{aligned} M_{BB}\Delta x_B &= (-h - Mx + y)_B - M_{BN}\Delta x_N + \Delta y_B, & \text{(24a)} \\ M_{NB}\Delta x_B - \Delta y_N &= (-h - Mx + y)_N - M_{NN}\Delta x_N, & \text{(24b)} \\ D_B^2 \Delta x_B + \Delta y_B &= -y_B + \sigma_k \mu_k X_B^{-1} e_B, & \text{(24c)} \\ D_N^{-2} \Delta y_N + \Delta x_N &= -x_N + \sigma_k \mu_k Y_N^{-1} e_N. & \text{(24d)} \end{aligned}$$

It follows from (24a) and (24b) that (w, z) satisfies the constraints (21). Now, since

$$y^* = Mx^* + h \;\Rightarrow 0 = M_{BB} x_B^* + h_B,$$

we have by substitution of (24c) in (24a) that

$$\begin{aligned} & D_B^2 \Delta x_B - \sigma_k \mu_k X_B^{-1} e_B \\ &= -y_B - [M_{BB}\Delta x_B + h_B + M_{BB} x_B + M_{BN} x_N - y_B + M_{BN}\Delta x_N] \\ &= -M_{BB}(x_B + \Delta x_B - x_B^*) - M_{BN}(x_N + \Delta x_N). \end{aligned} \tag{25}$$

From (24d),

$$D_N^{-2} \Delta y_N - \sigma_k \mu_k Y_N^{-1} e_N = -(x_N + \Delta x_N). \tag{26}$$

Together, (25) and (26) imply (23). ∎

Lemma 5.3 *There is a constant $C_5 > 0$ such that*

$$\begin{aligned} \|\Delta x_B\| &\leq \frac{C_5}{2}(\mu_k + \sigma_k). & \text{(27a)} \\ \|\Delta y_N\| &\leq \frac{C_5}{2}(\mu_k + \sigma_k). & \text{(27b)} \end{aligned}$$

Proof. Since the feasible set for (21) is nonempty, there is a feasible vector pair $(\bar{w}, \bar{z})$ such that

$$\left\| \begin{bmatrix} \bar{w} \\ \bar{z} \end{bmatrix} \right\| = O\left(\| -h - Mx + y\|\right) + O\left(\|\Delta x_N\|\right) + O\left(\|\Delta y_B\|\right) = O(\mu_k),$$

where the estimate for the infeasibility term follows from Lemma 3.2, since

$$\| -h - Mx^k + y^k\| = \nu_k \| -h - Mx^0 + y^0\| \le \frac{\mu_k}{\hat{\beta}\mu_0} \| -h - Mx^0 + y^0\|.$$

Note that the objective function (20) may be written as

$$\tfrac{1}{2}\|D_B w - \sigma_k \mu_k D_B^{-1} X_B^{-1} e_B\|^2 + \tfrac{1}{2}\|D_N^{-1} z - \sigma_k \mu_k D_N Y_N^{-1} e_N\|^2 + \text{constant}.$$

Hence

$$\begin{aligned}
&\|\Delta x_B - \sigma_k\mu_k D_B^{-2} X_B^{-1} e_B\|^2 + \|\Delta y_N - \sigma_k \mu_k D_N^2 Y_N^{-1} e_N\|^2 \\
\le\;& \|D_B^{-1}\|^2 \|D_B \Delta x_B - \sigma_k \mu_k D_B^{-1} X_B^{-1} e_B\|^2 + \|D_N\|^2 \|D_N^{-1}\Delta y_N - \sigma_k \mu_k D_N Y_N^{-1} e_N\|^2 \\
\le\;& \bar{C}_5^2 \left\{ \|D_B \Delta x_B - \sigma_k \mu_k D_B^{-1} X_B^{-1} e_B\|^2 + \|D_N^{-1}\Delta y_N - \sigma_k\mu_k D_N Y_N^{-1} e_N\|^2 \right\} \\
\le\;& \bar{C}_5^2 \left\{ \|D_B \bar{w} - \sigma_k \mu_k D_B^{-1} X_B^{-1} e_B\|^2 + \|D_N^{-1}\bar{z} - \sigma_k\mu_k D_N Y_N^{-1} e_N\|^2 \right\},
\end{aligned}$$

where $\bar{C}_5 = \max(\|D_B^{-1}\|, \|D_N\|)$. The last inequality follows from optimality of $(\Delta x_B, \Delta y_N)$ in (20),(21). Hence

$$\left\| \begin{bmatrix} \Delta x_B - \sigma_k \mu_k D_B^{-2} X_B^{-1} e_B \\ \Delta y_N - \sigma_k \mu_k D_N^2 Y_N^{-1} e_N \end{bmatrix} \right\| \le \bar{C}_5 \left\| \begin{bmatrix} D_B \bar{w} - \sigma_k \mu_k D_B^{-1} X_B^{-1} e_B \\ D_N^{-1}\bar{z} - \sigma_k \mu_k D_N Y_N^{-1} e_N \end{bmatrix} \right\|,$$

and so

$$\begin{aligned}
\left\| \begin{bmatrix} \Delta x_B \\ \Delta y_N \end{bmatrix} \right\| \;\le\;& \bar{C}_5 \left\{ \left\| \begin{bmatrix} D_B & \\ & D_N^{-1} \end{bmatrix} \right\| \left\| \begin{bmatrix} \bar{w} \\ \bar{z} \end{bmatrix} \right\| + \sigma_k \mu_k \left\| \begin{bmatrix} D_B^{-1} & \\ & D_N \end{bmatrix} \right\| \left\| \begin{bmatrix} X_B^{-1} e_B \\ Y_N^{-1} e_N \end{bmatrix} \right\| \right\} \\
&+ \sigma_k \mu_k \left\| \begin{bmatrix} D_B^{-2} & \\ & D_N^2 \end{bmatrix} \right\| \left\| \begin{bmatrix} X_B^{-1} e_B \\ Y_N^{-1} e_N \end{bmatrix} \right\|.
\end{aligned}$$

Now

$$\|D_B^{-1}\| = \sup_{i\in B} \left(\frac{x_i^k}{y_i^k}\right)^{1/2} \le C_2 \frac{1}{(\bar{\gamma}\mu_k)^{1/2}}, \qquad \|D_N\| = C_2 \frac{1}{(\bar{\gamma}\mu_k)^{1/2}}, \qquad \text{(Lemmas 3.5, 3.6)},$$

$$\|D_B\| = \sup_{i\in B} \left(\frac{y_i^k}{x_i^k}\right)^{1/2} \le C_1 \left(\frac{\mu_k}{\bar{\gamma}}\right)^{1/2}, \qquad \|D_N^{-1}\| \le C_1 \left(\frac{\mu_k}{\bar{\gamma}}\right)^{1/2}, \qquad \text{(Lemma 3.4)},$$

$$\|X_B^{-1} e_B\| = \sup_{i\in B} \frac{1}{x_i^k} \le \frac{C_1}{\bar{\gamma}}, \qquad \|Y_N^{-1} e_N\| \le \frac{C_1}{\bar{\gamma}}, \qquad \text{(Lemma 3.4)},$$

$$\|\bar{w}\| = O(\mu_k), \qquad \|\bar{z}\| = O(\mu_k), \qquad \bar{C}_5 = \max\left(\|D_B^{-1}\|, \|D_N\|\right) \le \frac{C_2}{(\bar{\gamma}\mu_k)^{1/2}}.$$

Hence

$$\begin{aligned}
\left\| \begin{bmatrix} \Delta x_B \\ \Delta y_N \end{bmatrix} \right\| \;\le\;& O(\mu_k^{-1/2}) \left\{ O(\mu_k^{3/2}) + \sigma_k O(\mu_k^{1/2}) \right\} + \sigma_k \mu_k O(\mu_k^{-1}) \\
\le\;& \frac{C_5}{2}(\mu_k + \sigma_k),
\end{aligned}$$

for an appropriately defined constant C_5. ■

For "fast" steps, the results of Lemmas 5.1 and 5.3 can be combined to give the following result.

Theorem 5.4 *If $\sigma_k = \mu_k$, then*

$$|\Delta x_i^k \Delta y_i^k| \le C_4 C_5 \mu_k^2, \qquad i = 1, \cdots, n.$$

6 Asymptotic Subquadratic Convergence

In this section, we show that all steps after a certain point in the algorithm are fast steps and that the sequences $\{\mu_k\}$ and $\{\phi_k\}$ converge Q-subquadratically to zero.

Throughout the analysis, we will make use of the constant C_6 defined by

$$C_6 \stackrel{\triangle}{=} \max\left(1, 2C_4C_5\right). \tag{28}$$

We start by defining a threshold condition involving μ_k and γ_k, and finding a lower bound on the step length when this condition is satisfied.

Lemma 6.1 *Suppose at iteration k that*

$$\frac{\mu_k}{(\gamma_k - \bar{\gamma})(1-\bar{\gamma})} \le \frac{\rho}{3C_6} \tag{29}$$

and that a fast step is calculated. Then the step length $\tilde{\alpha}$ will satisfy

$$1 \ge \tilde{\alpha} \ge 1 - C_6 \frac{\mu_k}{(\gamma_k - \bar{\gamma})(1-\bar{\gamma})}.$$

Proof. Before proceeding, note that if the fast step is successful, the algorithm sets γ_{k+1} to $\bar{\gamma}(1+\bar{\gamma}^{t_k})$, and so

$$\gamma_k - \gamma_{k+1} = \bar{\gamma}(1+\bar{\gamma}^{t_k-1}) - \bar{\gamma}(1+\bar{\gamma}^{t_k}) = \bar{\gamma}^{t_k}(1-\bar{\gamma}) = (\gamma_k - \bar{\gamma})(1-\bar{\gamma}). \tag{30}$$

Under these circumstances, condition (29) is equivalent to

$$\frac{\mu_k}{\gamma_k - \gamma_{k+1}} \le \frac{\rho}{3C_6}. \tag{31}$$

We use (29) and (31) interchangeably for the remainder of the proof.

The proof is in three stages. First, we show that the relaxed centrality condition (4e) holds for all α satisfying

$$\alpha \in \left[0, 1 - C_6 \frac{\mu_k}{\gamma_k - \gamma_{k+1}}\right]. \tag{32}$$

Second, we show that the other main condition on α, namely, (4d), also holds for α satisfying (32). Together, (4d) and (4e) imply that (4b) and (4c) also hold. Third, we show that

$\phi(x^k + \alpha\Delta x^k, y^k + \alpha\Delta y^k)$ is decreasing for α in the range (32), which, by (3), implies the result.

We start with (4e). Theorem 5.4, (2), (4e), (28), and the fact that $\tilde{\sigma} = \sigma_k = \mu_k$ imply that

$$\begin{aligned} |\Delta x_i^k \Delta y_i^k| &\leq C_4 C_5 \mu_k^2 \leq (C_6/2)\mu_k^2, \\ x_i^k y_i^k &\geq \gamma_k \mu_k, \\ x_i^k \Delta y_i^k + y_i^k \Delta x_i^k &= -x_i^k y_i^k + \sigma_k \mu_k = -x_i^k y_i^k + \mu_k^2. \end{aligned}$$

Hence

$$\begin{aligned} &(x_i^k + \alpha\Delta x_i^k)(y_i^k + \alpha\Delta y_i^k) \\ &\quad = x_i^k y_i^k + \alpha(x_i^k \Delta y_i^k + y_i^k \Delta x_i^k) + \alpha^2 (\Delta x_i^k)^T \Delta y_i^k \\ &\quad \geq \gamma_k \mu_k (1-\alpha) + \alpha\mu_k^2 - \alpha^2 (C_6/2)\mu_k^2. \end{aligned}$$

On the other hand,

$$\begin{aligned} &\frac{1}{n}\gamma_{k+1}(x^k + \alpha\Delta x_k)^T (y^k + \alpha\Delta y_k) \\ &\quad \leq \gamma_{k+1}\left[\mu_k(1-\alpha+\alpha\mu_k) + \alpha^2 (C_6/2)\mu_k^2\right]. \end{aligned}$$

Therefore condition (4e) will be satisfied provided that

$$\begin{aligned} &\gamma_k \mu_k (1-\alpha) + \alpha\mu_k^2 - \alpha^2 (C_6/2)\mu_k^2 \\ &\quad \geq \gamma_{k+1}\left[\mu_k(1-\alpha+\alpha\mu_k) + \alpha^2 (C_6/2)\mu_k^2\right], \end{aligned}$$

or, equivalently,

$$(\gamma_k - \gamma_{k+1})\mu_k(1-\alpha) + \alpha\mu_k^2(1-\gamma_{k+1}) - \alpha^2\mu_k^2(C_6/2)(1+\gamma_{k+1}) \geq 0.$$

Since $\gamma_{k+1} \in (0,1)$ and $\alpha \in (0,1]$, this last inequality will hold if

$$(\gamma_k - \gamma_{k+1})\mu_k(1-\alpha) - \alpha C_6 \mu_k^2 \geq 0 \quad \Rightarrow \quad 0 \leq \alpha \leq \bar{\alpha}_I \stackrel{\triangle}{=} \frac{1}{1 + C_6 \frac{\mu_k}{\gamma_k - \gamma_{k+1}}}.$$

By assumption (31), the second term in the denominator is less than one, so we can write

$$\bar{\alpha}_I \geq 1 - C_6 \frac{\mu_k}{\gamma_k - \gamma_{k+1}}.$$

Therefore, (4e) is satisfied for α in the range (32).

Turning now to condition (4d), we note that

$$\begin{aligned} \frac{1}{n}(x^k + \alpha\Delta x^k)^T (y^k + \alpha\Delta y^k) &= \mu_k(1-\alpha+\alpha\mu_k) + \frac{1}{n}\alpha^2 (\Delta x^k)^T \Delta y^k \\ &\geq \mu_k(1-\alpha+\alpha\mu_k) - \alpha(C_6/2)\mu_k^2, \end{aligned}$$

where we have used Theorem 5.4, (28), and $\alpha \in (0,1]$ to derive the inequality. Therefore, (4d) will hold if

$$\begin{array}{rrcl} & \mu_k(1-\alpha+\alpha\mu_k) - \alpha(C_6/2)\mu_k^2 & \geq & (1-\alpha)(1-\bar{\gamma}^{t_k})\mu_k \\ \Leftrightarrow & \bar{\gamma}^{t_k} - \alpha\left[\bar{\gamma}^{t_k} - \mu_k + (C_6/2)\mu_k\right] & \geq & 0. \end{array} \quad (33)$$

From (28), (29), and (30), we have that

$$\frac{\mu_k}{\bar{\gamma}^{t_k}} \leq \frac{\rho}{3C_6}(1-\bar{\gamma}) < 1.$$

Hence the term in square brackets in (33) is positive, and so (4d) is satisfied if

$$0 \leq \alpha \leq \bar{\alpha}_{II} \triangleq \frac{\bar{\gamma}^{t_k}}{\bar{\gamma}^{t_k} - \mu_k + (C_6/2)\mu_k}.$$

Now

$$\bar{\alpha}_{II} \geq \frac{1}{1+(C_6/2)\mu_k/\bar{\gamma}^{t_k}} \geq 1 - (C_6/2)\frac{\mu_k}{\bar{\gamma}^{t_k}}.$$

Using (30) again, we have that $\bar{\gamma}^{t_k} \geq (\gamma_k - \gamma_{k+1})$, and so

$$\bar{\alpha}_{II} \geq 1 - C_6\frac{\mu_k}{\gamma_k - \gamma_{k+1}}.$$

Therefore (4d) holds for α in the range (32).

Finally, we examine $\phi(x^k + \alpha\Delta x^k, y^k + \alpha\Delta y^k)$ for $\alpha \in [0,1]$. Using the notation $r_k = \|y^k - Mx^k - h\|$, we have that

$$\begin{array}{rcl} \phi(x^k+\alpha\Delta x^k, y^k+\alpha\Delta y^k) & = & (x^k+\alpha\Delta x^k)^T(y^k+\alpha\Delta y^k) + (1-\alpha)r_k \\ & = & n\mu_k + \alpha[(x^k)^T\Delta y^k + (y^k)^T\Delta x^k] + \alpha^2(\Delta x^k)^T\Delta y^k + (1-\alpha)r_k \\ & = & n\mu_k - \alpha n(1-\mu_k)\mu_k + \alpha^2(\Delta x^k)^T\Delta y^k + (1-\alpha)r_k. \end{array}$$

If $(\Delta x^k)^T\Delta y^k \leq 0$, then, since $\mu_k < 1$ by (31), $\phi(x^k+\alpha\Delta x^k, y^k+\alpha\Delta y^k)$ is decreasing on $[0,1]$. Otherwise, the unconstrained minimum occurs at

$$\alpha_{\min} = \frac{n(1-\mu_k)\mu_k + r_k}{2(\Delta x^k)^T\Delta y^k} \geq \frac{(2/3)n\mu_k}{2(\Delta x^k)^T\Delta y^k} \geq \frac{(2/3)n\mu_k}{2nC_4C_5\mu_k^2} \geq \frac{(2/3)}{C_6\mu_k}. \quad (34)$$

But from (31),

$$(3/2)C_6\mu_k \leq \frac{\rho}{2}(\gamma_k - \gamma_{k+1}) \leq 1, \quad (35)$$

and so $\alpha_{\min} \geq 1$. Again, we deduce that $\phi(x^k+\alpha\Delta x^k, y^k+\alpha\Delta y^k)$ is decreasing on $[0,1]$. This implies that the value of $\tilde{\alpha}$ chosen by the algorithm is the largest value that satisfies (4), and so

$$1 \geq \tilde{\alpha} \geq 1 - C_6\frac{\mu_k}{\gamma_k - \gamma_{k+1}},$$

as required. ■

The next two results show that once the threshold condition (29) is satisfied at some iteration, then fast steps may be taken on this and all subsequent iterations, and subquadratic convergence can be attained.

Lemma 6.2 *If (29) is satisfied at iteration k and a fast step is taken, then*

$$\begin{aligned} \mu_{k+1} &\leq \rho\mu_k, &\text{(36a)}\\ \phi_{k+1} &\leq \rho\phi_k, &\text{(36b)} \end{aligned}$$

$$\begin{aligned} \mu_{k+1} &\leq \left(\frac{3C_6}{\gamma_k - \gamma_{k+1}}\right)\mu_k^2, &\text{(37a)}\\ \phi_{k+1} &\leq \left(\frac{3C_6}{n(\gamma_k - \gamma_{k+1})}\right)\phi_k^2, &\text{(37b)} \end{aligned}$$

and

$$\frac{\mu_{k+1}}{(\gamma_{k+1} - \bar{\gamma})(1-\bar{\gamma})} \leq \frac{\mu_k}{(\gamma_k - \bar{\gamma})(1-\bar{\gamma})}. \tag{38}$$

Proof. Consider

$$\frac{1}{n}(x^k + \alpha\Delta x^k)^T(y^k + \alpha\Delta y^k) = \mu_k[1 - \alpha(1-\mu_k)] + \alpha^2\frac{(\Delta x^k)^T\Delta y^k}{n}. \tag{39}$$

The argument in the last part of Lemma 6.1, in particular, formulae (34) and (35), can be applied to show that the quadratic function (39) is decreasing on $[0,1]$. Therefore

$$\alpha_k \geq 1 - C_6\frac{\mu_k}{\gamma_k - \gamma_{k+1}},$$

and so

$$\begin{aligned} \mu_{k+1} &= \frac{1}{n}(x^k + \alpha_k\Delta x^k)^T(y^k + \alpha_k\Delta y^k)\\ &\leq \mu_k\left[1 - \left(1 - C_6\frac{\mu_k}{\gamma_k - \gamma_{k+1}}\right)(1-\mu_k)\right] + \frac{1}{n}\left|(\Delta x^k)^T\Delta y^k\right|\\ &\leq \mu_k\left[C_6\frac{\mu_k}{\gamma_k - \gamma_{k+1}} + \mu_k\right] + (C_6/2)\mu_k^2 &\text{(40)}\\ &\leq \mu_k\left[\frac{\rho}{3} + \frac{\rho}{3}\right] + \frac{\rho}{3}\mu_k\\ &\leq \rho\mu_k, \end{aligned}$$

yielding (36a). For (36b), we again use the notation $r_k = \|y^k - Mx^k - h\|$ and note that

$$r_{k+1} = (1-\alpha_k)r_k \leq C_6\frac{\mu_k}{\gamma_k - \gamma_{k+1}}r_k \leq \rho r_k.$$

Therefore,

$$\phi_{k+1} = n\mu_{k+1} + r_{k+1} \le n\rho\mu_k + \rho r_k = \rho\phi_k.$$

For (37a), using (40) and the fact that

$$1 \le C_6 \le C_6 \frac{1}{\gamma_k - \gamma_{k+1}},$$

we have that

$$\mu_{k+1} \le \mu_k \left[C_6 \frac{\mu_k}{\gamma_k - \gamma_{k+1}} + \mu_k + C_6\mu_k \right] \le \frac{3C_6}{\gamma_k - \gamma_{k+1}} \mu_k^2,$$

as required. Also,

$$\begin{aligned}
\phi_{k+1} &= n\mu_{k+1} + r_{k+1} \\
&\le \frac{3C_6}{\gamma_k - \gamma_{k+1}} n\mu_k^2 + \frac{C_6}{\gamma_k - \gamma_{k+1}} \mu_k r_k \\
&\le \frac{3C_6}{\gamma_k - \gamma_{k+1}} \mu_k [n\mu_k + r_k] \\
&\le \frac{3C_6}{n(\gamma_k - \gamma_{k+1})} [n\mu_k + r_k]^2 \\
&= \frac{3C_6}{n(\gamma_k - \gamma_{k+1})} \phi_k^2,
\end{aligned}$$

giving (37b).

From (37a) and (30), we have

$$\mu_{k+1} \le \frac{3C_6}{\bar{\gamma}^{t_k}(1-\bar{\gamma})} \mu_k^2.$$

Therefore,

$$\frac{\mu_{k+1}}{(\gamma_{k+1} - \bar{\gamma})(1-\bar{\gamma})} = \frac{\mu_{k+1}}{\bar{\gamma}^{t_k+1}(1-\bar{\gamma})} \le \frac{1}{\bar{\gamma}^{t_k+1}(1-\bar{\gamma})} \frac{3C_6}{\bar{\gamma}^{t_k}(1-\bar{\gamma})} \mu_k^2 = \frac{3C_6}{\bar{\gamma}} \left[\frac{\mu_k}{\bar{\gamma}^{t_k}(1-\bar{\gamma})} \right]^2. \quad (41)$$

From (29) and (30), we have

$$\frac{3C_6}{\bar{\gamma}} \frac{\mu_k}{\bar{\gamma}^{t_k}(1-\bar{\gamma})} = \frac{3C_6}{\bar{\gamma}} \frac{\mu_k}{(\gamma_k - \bar{\gamma})(1-\bar{\gamma})} \le \frac{\rho}{\bar{\gamma}} < 1,$$

where the last inequality follows from the definition of ρ and $\bar{\gamma}$ in the algorithm. Substitution of this inequality into (41) gives (38). ■

Theorem 6.3 *Suppose that the condition (29) is satisfied at iteration K and that*

$$\phi_K \le \bar{\phi}.$$

Then

(i) the algorithm will take fast steps at iteration K and at all subsequent iterations, and

(ii) the sequences $\{\mu_k\}$ and $\{\phi_k\}$ converge superlinearly to zero with Q-order 2.

Proof. The first part follows inductively from Lemma 6.2. Since (29) is satisfied at iteration K and since (36b) holds, the fast step will be accepted at this iteration. Formula (38) implies that (29) will again hold at iteration $K+1$. Since $\phi_{K+1} \leq \rho\phi_K < \bar{\phi}$, a fast step will be accepted at iteration $K+1$, and so on.

For (ii), we use (37a). Since

$$\mu_{k+1} \leq \frac{3C_6}{\bar{\gamma}^{t_k}(1-\bar{\gamma})}\mu_k^2,$$

we have

$$\log \mu_{k+1} \leq 2\log\mu_k + \log(3C_6) - t_k \log\bar{\gamma} - \log(1-\bar{\gamma}). \tag{42}$$

We can now use an argument like that of Ye [10]. From (36a) and

$$\log\rho < \log\bar{\gamma} < 0 \qquad \text{and} \qquad 1 \leq t_k \leq k+1,$$

we have for sufficiently large choice of k that

$$\log\mu_k \leq \log\mu_K + (k-K)\log\rho \ll (k+1)\log\bar{\gamma} \leq t_k\log\bar{\gamma},$$

that is, the first term on the right hand side of (42) will eventually dominate the third term and, in fact,

$$\lim_{k\to\infty}\frac{t_k}{\log\mu_k} = 0.$$

Dividing (42) by $\log\mu_k$, we have

$$\liminf_{k\to\infty}\frac{\log\mu_{k+1}}{\log\mu_k} = 2. \tag{43}$$

From Potra [6], (43) implies that the Q-order of convergence for $\{\mu_k\}$ is 2.

The argument for $\{\phi_k\}$ is similar. ∎

Finally, we show that the threshold condition (29) will eventually be met and, hence, that subquadratic convergence will be obtained.

Theorem 6.4 *Define*

$$f \triangleq \frac{\log\bar{\gamma}}{\log\rho} \in (0,1).$$

Define a constant $\hat{\epsilon}$ as follows:

$$\hat{\epsilon} = \begin{cases} \bar{\phi} & \text{if } \frac{\bar{\phi}}{n\bar{\gamma}(1-\bar{\gamma})} \leq \frac{\rho}{3C_6}, \\ \left[\frac{\rho}{3C_6}\frac{n\bar{\gamma}^2(1-\bar{\gamma})}{\bar{\phi}}\right]^{1/(1-f)}\bar{\phi} & \text{otherwise.} \end{cases} \tag{44}$$

Then if K is the smallest positive integer such that

$$\phi_K \le \hat{\epsilon}, \tag{45}$$

we have that

$$\frac{\mu_K}{\bar{\gamma}^{t_K}(1-\bar{\gamma})} \le \frac{\rho}{3C_6} \tag{46}$$

and

$$\phi_K \le \bar{\phi}, \tag{47}$$

and hence the conditions of Theorem 6.3 are satisfied.

Proof. First, consider the case of

$$\frac{\bar{\phi}}{n\bar{\gamma}(1-\bar{\gamma})} \le \frac{\rho}{3C_6}. \tag{48}$$

Then $\hat{\epsilon} = \bar{\phi}$, and so (45) immediately implies (47). Since K is the first iterate below $\bar{\phi}$, the algorithm cannot have taken any fast steps yet, so $t_K = 1$. Hence, by (45), (48), and the fact that $\mu_K \le \phi_K/n$,

$$\frac{\mu_K}{\bar{\gamma}^{t_K}(1-\bar{\gamma})} = \frac{\mu_K}{\bar{\gamma}(1-\bar{\gamma})} \le \frac{\phi_K/n}{\bar{\gamma}(1-\bar{\gamma})} \le \frac{\bar{\phi}/n}{\bar{\gamma}(1-\bar{\gamma})} \le \frac{\rho}{3C_6},$$

which gives (46).

In the remaining case, note that

$$\frac{\rho}{3C_6}\frac{n\bar{\gamma}^2(1-\bar{\gamma})}{\bar{\phi}} \le \frac{\rho}{3C_6}\frac{n\bar{\gamma}(1-\bar{\gamma})}{\bar{\phi}} < 1,$$

and so $\hat{\epsilon} < \bar{\phi}$ in (44), so again (45) implies (47). This inequality also implies that the algorithm may have taken some fast steps prior to iteration K but, since a reduction of at least ρ occurs on each fast step, the number of such steps is bounded by

$$\left\lceil \frac{\log(\hat{\epsilon}/\bar{\phi})}{\log \rho} \right\rceil.$$

Hence

$$t_K \le \left\lceil \frac{\log(\hat{\epsilon}/\bar{\phi})}{\log \rho} \right\rceil + 1 \le \frac{\log \hat{\epsilon} - \log \bar{\phi}}{\log \rho} + 2.$$

Now,

$$\begin{aligned} \bar{\gamma}^{t_K} &\ge \bar{\gamma}^2 \exp\left[\frac{\log \hat{\epsilon} - \log \bar{\phi}}{\log \rho} \log \bar{\gamma}\right] \\ &= \bar{\gamma}^2 \exp\left[f \log \hat{\epsilon} - f \log \bar{\phi}\right] \\ &= \bar{\gamma}^2 \hat{\epsilon}^f \bar{\phi}^{-f}. \end{aligned}$$

Hence

$$\begin{aligned}
\frac{\mu_K}{\bar{\gamma}^{t_K}(1-\bar{\gamma})} &\le \frac{\mu_K}{\bar{\gamma}^2(1-\bar{\gamma})\hat{\epsilon}^f\bar{\phi}^{-f}} \\
&\le \frac{\phi_K/n}{\bar{\gamma}^2(1-\bar{\gamma})\hat{\epsilon}^f\bar{\phi}^{-f}} \le \frac{\hat{\epsilon}}{n\bar{\gamma}^2(1-\bar{\gamma})\hat{\epsilon}^f\bar{\phi}^{-f}} = \hat{\epsilon}^{1-f}\frac{\bar{\phi}^f}{n\bar{\gamma}^2(1-\bar{\gamma})}.
\end{aligned}$$

Substituting in this expression from (44), we have

$$\frac{\mu_K}{\bar{\gamma}^{t_K}(1-\bar{\gamma})} \le \frac{\rho}{3C_6}\frac{n\bar{\gamma}^2(1-\bar{\gamma})}{\bar{\phi}}\bar{\phi}^{1-f}\frac{\bar{\phi}^f}{n\bar{\gamma}^2(1-\bar{\gamma})} = \frac{\rho}{3C_6},$$

and so (46) is satisfied. ∎

The following result is immediate from Theorems 6.3 and 6.4.

Corollary 6.5 *The algorithm converges Q-subquadratically.*

References

[1] J. Ji, F. Potra, and S. Huang, *A predictor-corrector method for linear complementarity problems with polynomial complexity and superlinear convergence*, Tech. Rep. 18, Department of Mathematics, The University of Iowa, Iowa City, Iowa, August 1991.

[2] M. Kojima, N. Meggido, and S. Mizuno, *A primal-dual exterior point algorithm for linear programming*, Tech. Rep. RJ 8500, IBM Alamaden Research Center, San Jose, Calif., 1991.

[3] M. Kojima, S. Mizuno, and M. Todd, *Infeasible interior-point primal-dual potential-reduction algorithms for linear programming*, Tech. Rep. 1023, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, New York, August 1992.

[4] I. J. Lustig, R. E. Marsten, and D. F. Shanno, *Computational experience with a globally convergent primal-dual predictor-corrector algorithm for linear programming*, Tech. Rep. SOR 92–10, Program in Statistics and Operations Research, Department of Civil Engineering and Operations Research, Princeton University, Princeton, New Jersey, September 1992.

[5] R. D. C. Monteiro and S. J. Wright, *A globally and superlinearly convergent potential reduction interior point method for convex programming*, Tech. Rep. MCS–P316–0792, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois, August 1992.

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

[7] F. A. Potra, *An infeasible interior-point predictor-corrector algorithm for linear programming*, Tech. Rep. 26, Department of Mathematics, University of Iowa, Iowa City, Iowa, June 1992.

[8] ——, *A quadratically convergent infeasible interior-point algorithm for linear programming*, Tech. Rep. 28, Department of Mathematics, University of Iowa, Iowa City, Iowa, July 1992.

[9] R. T. Rockafellar, *Computational schemes for large-scale problems in extended linear-quadratic programming*, Mathematical Programming, 48 (1990), pp. 447–474.

[10] Y. Ye, *On the Q-order of convergence of interior-point algorithms for linear programming*, Tech. Rep. 91–17, Department of Management Sciences, University of Iowa, October 1991.

[11] Y. Ye and K. Anstreicher, *On quadratic and* $O(\sqrt{n}L)$ *convergence of a predictor-corrector algorithm for LCP*, Tech. Rep. 91–20, Department of Management Sciences, University of Iowa, November 1991.

[12] Y. Zhang, *On the convergence of an infeasible interior-point algorithm for linear programming and other problems*, Tech. Rep. 92–07, Department of Mathematics and Statistics, University of Maryland, Baltimore County, April 1992.

[13] Y. Zhang and D. Zhang, *Superlinear convergence of infeasible interior-point methods for linear programming*, Tech. Rep. 92–15, Department of Mathematics and Statistics, University of Maryland, Baltimore County, September 1992.