PREPRINT MCS-P401-1293, MCS DIVISION, ARGONNE NATIONAL LABORATORY

A Path-Following Interior-Point Algorithm for Linear and Quadratic Problems*

Stephen J. Wright†

December 1993 (revised January, 1995)

Abstract

We describe an algorithm for the monotone linear complementarity problem (LCP) that converges from any positive, not necessarily feasible, starting point and exhibits polynomial complexity if some additional assumptions are made on the starting point. If the problem has a strictly complementary solution, the method converges subquadratically. We show that the algorithm and its convergence properties extend readily to the mixed monotone linear complementarity problem and, hence, to all the usual formulations of the linear programming and convex quadratic programming problems.

1 Introduction

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

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

where $q \in \mathbb{R}^n$ and M is an $n \times n$ positive semidefinite (p.s.d.) matrix. The *mixed* monotone linear complementarity problem (MLCP) is to find a vector triple $(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$ such that

$$\begin{bmatrix} y \\ 0 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} + \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} \tag{2a}$$

$$(x, y) \geq 0, \qquad x^T y = 0, \tag{2b}$$

where

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

*This research was supported by the Office of Scientific Computing, U.S. Department of Energy, under Contract W-31-109-Eng-38.

†Mathematics and Computer Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439.

is p.s.d. All conventional formulations of the linear programming (LP) and convex quadratic programming (QP) problems can be posed in the form (2) by writing out their conditions for optimality. For instance, consider the QP problem given by

$$\min_{w \in \mathrm{I\!R}^p} \tfrac{1}{2} w^T Q w + c^T w \tag{3}$$

subject to

$$\begin{aligned} w_i \geq l_i \;\; (i \in \mathcal{L} \subset \{1, \cdots, p\}), \qquad & w_i \leq u_i \;\; (i \in \mathcal{U} \subset \{1, \cdots, p\}), \\ Cw \geq d, \qquad & Aw = b, \end{aligned} \tag{4}$$

where Q is symmetric p.s.d., $C \in \mathrm{I\!R}^{m_I \times p}$, $A \in \mathrm{I\!R}^{m_E \times p}$, and so on. If we define

$$E_{\mathcal{L}} = [e_i^T]_{i \in \mathcal{L}}, \qquad E_{\mathcal{U}} = [e_i^T]_{i \in \mathcal{U}},$$

where e_i is the i-th unit vector from the standard basis, and

$$l = [l_i]_{i \in \mathcal{L}}, \qquad u = [u_i]_{i \in \mathcal{U}},$$

then we can state the optimality conditions for (3),(4) in the form (2) by defining

$$n = |\mathcal{L}| + |\mathcal{U}| + m_I, \qquad m = p + m_E,$$

and

$$M_{22} = \begin{bmatrix} Q & -A^T \\ A & 0 \end{bmatrix}, \qquad M_{12} = \begin{bmatrix} E_{\mathcal{L}} & 0 \\ -E_{\mathcal{U}} & 0 \\ C & 0 \end{bmatrix}, \qquad M_{21} = \begin{bmatrix} -E_{\mathcal{L}}^T & E_{\mathcal{U}}^T & -C^T \\ 0 & 0 & 0 \end{bmatrix},$$

$$M_{11} = 0, \qquad q_1 = \begin{bmatrix} -l \\ u \\ -d \end{bmatrix}, \qquad q_2 = \begin{bmatrix} c \\ -b \end{bmatrix}.$$

In this paper, we focus on an algorithm for (1) and its convergence properties. We then show, using recent work involving the relationship between problems (1) and (2), that this algorithm can be extended painlessly to (2) and, hence, to all the usual LP and QP formulations. Little loss of efficiency is involved in solving LPs and QPs by embedding them in algorithms for (2), provided the linear algebra takes account of the particular structure of each problem. Hence we feel that the linear complementarity formulation is the best one to consider because of its generality, simplicity of notation, and practical efficiency of the algorithms on all its special cases.

In two recent papers [10, 9], we described algorithms for (1) that are globally convergent, have polynomial complexity when the starting point (x^0, y^0) satisfies certain assumptions, and exhibit subquadratic convergence of the complementarity gap $\mu_k = (x^k)^T y^k / n$ to zero. Neither the starting point nor the iterates are feasible in general. In both algorithms, most of the work at each iteration consists of a single matrix factorization and between one and three triangular solves with the computed factors. These methods were the first interior-point methods with this desirable combination of properties.

The local analysis in both [10] and [9] requires existence of a strictly complementary solution, that is, (x^*, y^*) solving (1) such that $x^* + y^* > 0$. This assumption is always satisfied if the LCP is a reformulated linear programming problem and, as shown in Monteiro and Wright [7], it is *necessary* for superlinear convergence of Newton-based primal-dual algorithms. The earlier paper [10] makes an additional assumption: existence of a strictly feasible point, that is, $(\bar{x}, \bar{y})$ such that $\bar{y} = M\bar{x} + q$, $(\bar{x}, \bar{y}) > 0$. This assumption is undesirable because it is usually not satisfied by large practical problems.

In this paper, we describe an algorithm that is quite similar to the one in [10], except that it does not use the clumsy merit function $\phi(x, y) = x^T y + \|y - Mx - q\|$ and allows more flexibility in the choice of parameters and starting point. As we mention at the end of Section 3, we can also allow more flexibility in the choice of steplength, bringing the algorithm close to current computational practice. The analysis here is considerably stronger than in [10]. The strict feasibility assumption is no longer required, technical arguments are streamlined, and R-subquadratic convergence of the iterates (x^k, y^k) to a strictly complementary solution is proved.

The new algorithm is quite different from the one in [9], where the ratio of complementarity gap μ to infeasibility norm $\|r\|$ is kept constant until the "fast" phase, when small variations are allowed. To achieve this effect, a correction to the basic path-following step is computed (at a cost of one additional back-substitution) and a messy planar search procedure is performed. The complications in [9] made it possible to prove boundedness of the iteration sequence $\{(x^k, y^k)\}$, which was a vital element in the asymptotic convergence analysis. Our use of a stronger technical result (Lemma 4.3) removes the dependence on boundedness. Instead, as mentioned above, we prove convergence of the iteration sequence as a *consequence* of the main results.

Our algorithm is specified in Section 2. In Section 3 we prove global linear convergence and polynomial complexity. Some technical results are proved in Section 4; these are used to prove superlinear convergence in Section 5. Section 6 shows that the algorithm and its convergence properties can be extended to the mixed problem (2) because (2) can be reformulated as (1). We stress at the outset that this reformulation need not be performed explicitly; it suffices to observe that the (x, y) iterates generated by our extended algorithm are the same as those that would be obtained by reformulating the problem as (1) and applying the algorithm of Section 2 directly, except possibly for some swapping of components to be discussed later.

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. To avoid notational clutter in Sections 3, 4, and 5, we drop the iteration index k from vector and matrix quantities in the proofs. It is retained explicitly in the statement of each result.

We denote the solution set and strictly complementary solution set by

$$\mathcal{S} = \{(x^*, y^*) \,|\, (x^*, y^*) \ \text{ solves } \ (1)\}, \qquad \mathcal{S}^c = \{(x^*, y^*) \in \mathcal{S} \,|\, x^* + y^* > 0\},$$

respectively. The range space of a matrix is denoted by $R(\cdot)$.

2 The Algorithm

The algorithm generates a sequence of strictly positive iterates (x^k, y^k). To describe the step between successive iterates, we define

$$\mu_k = (x^k)^T y^k/n, \qquad r^k = y^k - Mx^k - q, \qquad e = (1, 1, \cdots, 1)^T,$$
$$X^k = \mathrm{diag}(x_1^k, x_2^k, \cdots, x_n^k), \qquad Y^k = \mathrm{diag}(y_1^k, y_2^k, \cdots, y_n^k).$$

We refer to μ_k as the *complementarity gap* and to r^k as the *residual.* Each step is calculated as follows.

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

$$\begin{bmatrix} M & -I \\ Y^k & X^k \end{bmatrix} \begin{bmatrix} u^k \\ v^k \end{bmatrix} = \begin{bmatrix} r^k \\ -X^k Y^k e + \tilde{\sigma}\mu_k e \end{bmatrix}. \tag{5}$$

Choose

$$\tilde{\alpha} = \arg \min_{\alpha \in [0,\hat{\alpha}]} \mu_k(\alpha) \triangleq (x^k + \alpha u^k)^T (y^k + \alpha v^k)/n, \tag{6}$$

where $\hat{\alpha}$ is the largest number in $[0,1]$ such that the following inequalities are satisfied for all $\alpha \in [0, \hat{\alpha})$:

$$(x^k + \alpha u^k)^T (y^k + \alpha v^k) \;\geq\; (1-\tilde{\beta})(1-\alpha)(x^k)^T y^k, \qquad \text{if } r^k \neq 0, \tag{7a}$$
$$(x_i^k + \alpha u_i^k)(y_i^k + \alpha v_i^k) \;\geq\; (\tilde{\gamma}/n)(x^k + \alpha u^k)^T (y^k + \alpha v^k), \qquad i = 1, \cdots, n. \tag{7b}$$

The search direction obtained from (5) is simply the Newton step for the system of nonlinear equations

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

from the point (x^k, y^k).

The inequality (7b), usually referred to as a *centering condition*, ensures that the iterates do not approach the boundary of the nonnegative orthant too closely. Because we restrict $\tilde{\gamma}$ to the range $[\gamma_{\min}, \gamma_{\max}]$ for $0 < \gamma_{\min} < \gamma_{\max} \leq 1/2$, (7b) implies that

$$x_i y_i \geq \gamma_k \mu_k \geq \gamma_{\min} \mu_k, \qquad \forall k. \tag{8}$$

The condition (7a) is used to ensure that improvement in the complementarity gap μ_k does not outstrip improvement in the infeasibility $\|r^k\|$ by too much; a vector pair (x, y) that is complementary but not feasible is of no interest. Note that we need not enforce condition (7a) if the current point is already feasible.

We can now state our algorithm.

Given $\bar{\gamma} \in (0, 1/2)$, $\gamma_{\min}$ and $\gamma_{\max}$ with $0 < \gamma_{\min} < \gamma_{\max} \le 1/2$, $\bar{\sigma} \in (0, 1/2)$, $\rho \in (0, \bar{\gamma})$, $\bar{\mu} \in (0, 1]$, and (x^0, y^0) with $x_i^0 y_i^0 \ge \gamma_{\max}\mu_0 > 0$;

$t_0 \leftarrow 1$, $\gamma_0 \leftarrow \gamma_{\max}$;

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

if $\mu_k = 0$ **then** stop;

(* attempt a fast step *)
Solve (5)–(7) with $\tilde{\sigma} = 0$, $\tilde{\beta} = \bar{\gamma}^{t_k}$, $\tilde{\gamma} = \gamma_{\min} + \bar{\gamma}^{t_k}(\gamma_{\max} - \gamma_{\min})$;
if $(x^k + \tilde{\alpha} u^k)^T (y^k + \tilde{\alpha} v^k)/n \le \rho\mu_k$
then $\alpha_k \leftarrow \tilde{\alpha}$, $\beta_k \leftarrow \tilde{\beta}$, $\sigma_k \leftarrow \tilde{\sigma}$, $\gamma_{k+1} \leftarrow \tilde{\gamma}$;
$t_{k+1} \leftarrow t_k + 1$;
$(x^{k+1}, y^{k+1}) \leftarrow (x^k, y^k) + \alpha_k(u^k, v^k)$;
go to next k;
end if

(* revert to a safe step *)
Solve (5)–(7) with $\tilde{\sigma} \in [\bar{\sigma}, 1/2]$, $\tilde{\beta} = 0$, $\tilde{\gamma} = \gamma_k$;
$\alpha_k \leftarrow \tilde{\alpha}$, $\beta_k \leftarrow 0$, $\sigma_k \leftarrow \tilde{\sigma}$, $\gamma_{k+1} \leftarrow \tilde{\gamma}$;
$t_{k+1} \leftarrow t_k$;
$(x^{k+1}, y^{k+1}) \leftarrow (x^k, y^k) + \alpha_k(u^k, v^k)$;
go to next k;

end for.

The algorithm can be motivated in a few sentences. We begin each major iteration by trying to take a *fast* step, which uses an affine scaling search direction. To encourage longer steps to be taken we use a strictly positive value of $\tilde{\beta}$ and a value $\tilde{\gamma}$ smaller than the current γ_k. The fast steps are accepted only if they produce a reduction in μ_k of at least a factor of ρ. Otherwise, the algorithm reverts to taking a *safe* step, whose major distinguishing feature is its use of a strictly positive value $\tilde{\sigma} \ge \bar{\sigma}$ of the centering parameter.

Safe steps tend to be taken on early iterations, while fast steps are taken toward the tail of the sequence. There may be a gray area in which both safe and fast steps are taken.

The algorithm can be modified to try fast steps only when there is some reasonable hope that they will be accepted. (Earlier versions of the algorithm used a threshold criterion $\mu_k \leq \bar{\mu}$, with $\bar{\mu}$ a user-defined parameter, to decide whether to calculate the safe step.) For the sake of simplicity, we do not consider such modifications here, but note simply that the superlinear convergence properties of the algorithm will hold provided that the fast step is *eventually* tried on every iteration. Besides omitting the threshold $\bar{\mu}$, the algorithm above differs from the one described in [10] in that the duality gap μ is used directly in place of the merit function ϕ, and the particular choices $\gamma_{\min} = \bar{\gamma}$ and $\gamma_{\max} = 2\bar{\gamma}$ are relaxed.

3 Global Convergence and Polynomial Complexity

In this section, we show that the algorithm converges globally to the solution set of (1) from any starting point $(x^0, y^0) > 0$. When the algorithm is initialized in a certain way, the number of iterations is quadratic in the problem dimension n. Throughout the section, we make the following assumption.

Assumption 1 *The LCP (1) is feasible; that is, there is a pair (x, y) such that $y = Mx + q$ and $(x, y) \geq 0$.*

Assumption 1 implies that $\mathcal{S} \neq \emptyset$ (see, for example, [1, Theorem 3.1.2]).

If we define the monotonically decreasing sequence $\{\nu_k\}$ by

$$\nu_0 = 1, \qquad \nu_{k+1} = (1 - \alpha_k)\nu_k,$$

it is easy to see that $r^k = \nu_k r^0$. We have the following simple result, whose proof follows that of [9, Lemma 3.1].

Lemma 3.1 *The constant $\hat{\beta}$ defined by*

$$\hat{\beta} \triangleq \prod_{k=0}^{\infty}(1 - \beta_k) \geq \prod_{k=1}^{\infty}(1 - \bar{\gamma}^k)$$

is strictly positive, and we have for all $k \geq 0$ that

$$\mu_k \geq \hat{\beta}\nu_k\mu_0. \tag{9}$$

We can also show that all iterates remain strictly positive, except when finite termination occurs.

Lemma 3.2 *For all iterates generated by the algorithm, we have either $(x^k, y^k) > 0$ or $\mu_k = 0$.*

Proof. We prove the result by induction. Note first that the assertion is trivially satisfied by the initial iterate $(x^0, y^0) > 0$. If $\mu_k = 0$, the algorithm terminates at the k-th iterate. For the remainder of the proof, we assume that $(x^k, y^k) > 0$ and prove that either $(x^{k+1}, y^{k+1}) > 0$ or $\mu_{k+1} = 0$. We consider the cases $r^k \neq 0$ and $r^k = 0$ separately.

If $r^k \neq 0$, the constraint (7a) is applied to the choice of α_k. Hence, combining (7a) and (7b), we have that

$$(x_i^k + \alpha u_i^k)(y_i^k + \alpha v_i^k) \geq (\tilde{\gamma}/n)(1-\tilde{\beta})(1-\alpha)({x^k}^T y^k), \qquad \forall \alpha \in [0, \alpha_k).$$

Since $\alpha_k \leq 1$, $\tilde{\gamma} > 0$, $\tilde{\beta} \in [0,1)$, and ${x^k}^T y^k > 0$, the right-hand side of this expression is strictly positive for all $\alpha \in [0, \alpha_k)$. Since $(x_i^k, y_i^k) > 0$, it follows that $(x_i^k + \alpha u_i^k, y_i^k + \alpha v_i^k) > 0$ for all $\alpha \in [0, \alpha_k)$. If $x_i^k + \alpha_k u_i^k = 0$ or $y_i^k + \alpha_k v_i^k = 0$ for some index i, we have from (7b) that

$$\mu_{k+1} = (x^k + \alpha_k u^k)^T (y^k + \alpha_k v^k)/n = 0.$$

The alternative case is

$$(x^{k+1}, y^{k+1}) = (x^k + \alpha_k u^k, y^k + \alpha_k v^k) > 0,$$

so our claim holds.

Consider now the remaining case $r^k = 0$, for which (7a) is not enforced. Suppose first that there exist α values in the range $[0, \alpha_k]$ such that

$$x_i^k + \alpha u_i^k = 0 \qquad \text{or} \qquad y_i^k + \alpha v_i^k = 0, \tag{10}$$

for some index $i = 1, \cdots, n$, and let $\bar{\alpha}_k$ be the smallest of these values. It follows from (7b) that

$$(x^k + \bar{\alpha}_k u^k)^T (y^k + \bar{\alpha}_k v^k) = 0. \tag{11}$$

Since $r^k = 0$, equation (5) implies that $v^k = Mu^k$. Hence, by positive semidefiniteness of M, we have ${u^k}^T v^k \geq 0$. Using the second part of (5), we find that

$$(x^k + \alpha u^k)^T (y^k + \alpha v^k) = ({x^k}^T y^k)(1 - \alpha(1-\tilde{\sigma})) + \alpha^2 {u^k}^T v^k \geq ({x^k}^T y^k)(1 - \alpha(1-\tilde{\sigma})). \tag{12}$$

Now since $\bar{\alpha}_k \in [0, \alpha_k] \subset [0,1]$, the relations (11) and (12) can be satisfied simultaneously for $\alpha = \bar{\alpha}_k$ only if $\tilde{\sigma} = 0$ and $\bar{\alpha}_k = \alpha_k = 1$. Hence we are left with two possibilities. Either $\bar{\alpha}_k = \alpha_k = 1$ and $\mu_{k+1} = 0$, or there are no $\alpha \in [0, \alpha_k]$ with the property (10), so $(x^{k+1}, y^{k+1}) > 0$. Therefore our claim holds again for the case of $r^k = 0$, and we are done. ∎

Finite termination of the algorithm with $\mu_k = 0$ and $r^k = 0$ is, of course, the simple case. Throughout the remainder of the paper, we make the implicit assumption that finite termination does not occur and that the algorithm generates an infinite sequence of iterates $\{(x^k, y^k)\}$, $k = 0, 1, \cdots$.

The next lemma contains an inequality that is used in a number of places in the analysis. Similar results appear in Potra [8, Lemma 4.1], Mizuno [5, Lemma 3.3], and Wright [9, Lemma 3.2].

Lemma 3.3 *Let $(x^*, y^*) \in \mathcal{S}$. Then for all $k \geq 0$ there is a constant C_1 such that*

$$\|y^k\|_1 + \|x^k\|_1 \leq C_1 \left[\mu_0 + \mu_k/\nu_k + \|y^*\|_1 + \|x^*\|_1\right]. \tag{13}$$

When the initial point satisfies

$$x^0 = \xi_x e, \qquad y^0 = \xi_y e, \tag{14}$$

for some positive ξ_x and ξ_y, we have

$$\xi_x\|y^k\|_1 + \xi_y\|x^k\|_1 \leq n\mu_0 + n\mu_k/\nu_k + \xi_x\|y^*\|_1 + \xi_y\|x^*\|_1. \tag{15}$$

Proof. As in the proof of [9, Lemma 3.2] we obtain the inequalities

$$\begin{aligned} 0 \;\leq\; & \nu_k^2(x^0)^T y^0 + (1-\nu_k)^2(x^*)^T y^* + x^T y + \nu_k(1-\nu_k)((x^0)^T y^* + (x^*)^T y^0) \\ & -\nu_k((x^0)^T y + (y^0)^T x) - (1-\nu_k)((x^*)^T y + x^T y^*) \end{aligned} \tag{16}$$

and

$$\nu_k((x^0)^T y + (y^0)^T x) \leq \nu_k^2(x^0)^T y^0 + x^T y + \nu_k(1-\nu_k)((x^0)^T y^* + (y^0)^T x^*). \tag{17}$$

By defining

$$\bar{C}_1 = \min_{i=1,\cdots,n} \min(x_i^0, y_i^0),$$

and noting that $1 - \nu_k \leq 1$, we have

$$\|y\|_1 + \|x\|_1 \leq \bar{C}_1^{-1}\left[\nu_k n\mu_0 + n\mu_k/\nu_k + \|x^0\|_\infty\|y^*\|_1 + \|y^0\|_\infty\|x^*\|_1\right],$$

which implies (13) for appropriately defined C_1. The other inequality (15) follows trivially from (14) and (17). ∎

For purposes of polynomial complexity, we assume that the initial point is defined by (14), where ξ_x and ξ_y satisfy the following assumptions:

$$\|x^*\|_\infty \leq \xi_x, \qquad \|y^*\|_\infty \leq \xi_y, \qquad \xi_y \geq \|q\|_\infty, \qquad \xi_y \geq \|Me\|_\infty \xi_x = \|Mx^0\|_\infty. \tag{18}$$

We also find the diagonal matrix D^k defined by

$$D^k = (X^k)^{-1/2}(Y^k)^{1/2}$$

useful in the subsequent analysis. The next lemma allows us to bound quantities involving the steps u^k and v^k.

Lemma 3.4 *For all $k \geq 0$, there is a constant ω such that*

$$\|D^k u^k\|^2 + \|(D^k)^{-1} v^k\|^2 \leq \omega\mu_k. \tag{19}$$

If the initial step is chosen according to (14) and (18), then

$$\omega = \frac{2n^2}{\gamma_{\min}}\left(5 + \frac{9}{\hat{\beta}}\right)^2, \tag{20}$$

and so $\omega = O(n^2)$.

Proof. It is easy to see that the step (u, v) can be partitioned as

$$(u, v) = (\bar{u}, \bar{v}) + (\hat{u}, \hat{v}),$$

where

$$\begin{bmatrix} M & -I \\ Y & X \end{bmatrix} \begin{bmatrix} \bar{u} \\ \bar{v} \end{bmatrix} = \begin{bmatrix} 0 \\ -XYe + \sigma_k \mu_k e \end{bmatrix} \tag{21}$$

and

$$\begin{bmatrix} M & -I \\ Y & X \end{bmatrix} \begin{bmatrix} \hat{u} \\ \hat{v} \end{bmatrix} = \begin{bmatrix} r \\ 0 \end{bmatrix}. \tag{22}$$

(Because of nonsingularity of the coefficient matrix in (21) and (22), both $(\bar{u}, \bar{v})$ and $(\hat{u}, \hat{v})$ are well defined.) As in [9, Lemma 3.3], we can show that

$$\|D\bar{u}\| \leq \frac{2n\mu_k^{1/2}}{\gamma_{\min}^{1/2}}, \qquad \|D^{-1}\bar{v}\| \leq \frac{2n\mu_k^{1/2}}{\gamma_{\min}^{1/2}}. \tag{23}$$

For the other component $(\hat{u}, \hat{v})$, we have from the second part of (22) that

$$D\hat{u} + D^{-1}\hat{v} = 0 \quad \Rightarrow \quad \hat{v} = -D^2\hat{u}.$$

Hence from the first part of (22) and positive semidefiniteness of M, we have

$$M\hat{u} - \hat{v} = r \quad \Rightarrow \quad (M + D^2)\hat{u} = r \quad \Rightarrow \quad \hat{u}^T D^2 \hat{u} \leq \hat{u}^T r.$$

Therefore

$$\|D\hat{u}\|^2 \leq \|D\hat{u}\| \|D^{-1}r\|.$$

Using (8) together with Lemmas 3.1 and 3.3, we have that

$$\begin{aligned}
\|D\hat{u}\| \leq \|D^{-1}r\| &\leq \sum_i \left(\frac{x_i}{y_i}\right)^{1/2} \|r\|_\infty \\
&= \sum_i (x_i y_i)^{-1/2} x_i \nu_k \|r^0\|_\infty \\
&\leq \frac{\|r^0\|_\infty}{\gamma_{\min}^{1/2}\mu_k^{1/2}} \nu_k \|x\|_1 \\
&\leq \frac{\|r^0\|_\infty}{\gamma_{\min}^{1/2}\mu_k^{1/2}} C_1 \left[\nu_k \mu_0 + \mu_k + \nu_k \|y^*\|_1 + \nu_k \|x^*\|_1\right] \\
&\leq \frac{\|r^0\|_\infty}{\gamma_{\min}^{1/2}\mu_k^{1/2}} C_1 \mu_k \left[1 + (\mu_0 + \|y^*\|_1 + \|x^*\|_1)/(\hat{\beta}\mu_0)\right] \\
&= C_2 \mu_k^{1/2},
\end{aligned} \tag{24}$$

for some appropriately defined constant C_2. Since $\|D\hat{u}\| = \|D^{-1}\hat{v}\|$, we therefore have

$$\|D^{-1}\hat{v}\| = \|D\hat{u}\| \leq C_2 \mu_k^{1/2}. \tag{25}$$

By combining (23) and (25), we obtain

$$
\begin{aligned}
&\|Du\|^2 + \|D^{-1}v\|^2 \\
&\quad \le \ (\|D\bar{u}\| + \|D\hat{u}\|)^2 + (\|D^{-1}\bar{v}\| + \|D^{-1}\hat{v}\|)^2 \\
&\quad \le \ 2\left(\frac{2n\mu_k^{1/2}}{\gamma_{\min}^{1/2}} + C_2\mu_k^{1/2}\right)^2 \\
&\quad \le \ \omega\mu_k,
\end{aligned}
$$

where ω is defined in an obvious way.

The special result (20) is proved by analysis similar to that of Lemma 3.4 in Wright [9]. ∎

The key theorem of this section shows that there is a uniform lower bound on the step length α_k on each safe step.

Theorem 3.5 *If a safe step is taken at iteration k, then*

$$\alpha_k \ge \frac{\bar{\sigma}(1-\gamma_{\max})}{2\omega},$$

where ω is as defined in Lemma 3.4.

Proof. We prove the result by showing that the conditions (7a) and (7b) hold for all α in the range

$$\left[0, \frac{\bar{\sigma}(1-\gamma_{\max})}{2\omega}\right]. \tag{26}$$

We further show that the complementarity gap $\mu_k(\alpha)$ defined in (6) is decreasing on the interval (26). These observations are sufficient to prove the result.

Because of (5), we have that (7a) is satisfied if

$$
\begin{aligned}
(x+\alpha u)^T(y+\alpha v) \ &\ge\ (1-\alpha)x^Ty \\
\Leftrightarrow\ x^Ty(1-\alpha+\alpha\sigma_k) + \alpha^2u^Tv \ &\ge\ (1-\alpha)x^Ty \\
\Leftrightarrow\ \sigma_k x^Ty + \alpha u^Tv \ &\ge\ 0.
\end{aligned} \tag{27}
$$

From Lemma 3.4 we have

$$|u^Tv| \le \|Du\|\|D^{-1}v\| \le \omega\mu_k. \tag{28}$$

Since $\sigma_k \ge \bar{\sigma}$ and α is in the range (26), we have

$$
\begin{aligned}
\sigma_k x^Ty + \alpha u^Tv \ &\ge\ \bar{\sigma}x^Ty - \alpha|u^Tv| \\
&\ge\ \bar{\sigma}x^Ty - \frac{\bar{\sigma}(1-\gamma_{\max})}{2\omega}\omega\mu_k \\
&=\ \bar{\sigma}x^Ty\left[1 - \frac{1-\gamma_{\max}}{2n}\right] \\
&>\ 0.
\end{aligned}
$$

Therefore (27) and hence (7a) are satisfied for α in the range (26).

By using (5), we note that (7b) is satisfied if

$$\begin{aligned} (x_i+\alpha u_i)(y_i+\alpha v_i) &\geq \gamma_k(x+\alpha u)^T(y+\alpha v)/n \\ \Leftrightarrow \quad x_i y_i(1-\alpha)+\alpha\sigma_k\mu_k+\alpha^2 u_i v_i &\geq (\gamma_k/n)\left[x^T y(1-\alpha+\sigma_k\alpha)+\alpha^2 u^T v\right]. \end{aligned} \tag{29}$$

Because of (8), (29) is true provided that

$$\begin{aligned} \gamma_k(1-\alpha)\mu_k+\alpha\sigma_k\mu_k+\alpha^2 u_i v_i &\geq \gamma_k\left[(1-\alpha+\sigma_k\alpha)\mu_k+\alpha^2 u^T v/n\right] \\ \Leftrightarrow \quad \sigma_k(1-\gamma_k)\mu_k+\alpha(u_i v_i-\gamma_k u^T v/n) &\geq 0. \end{aligned}$$

Now from (28) and

$$|u_i v_i| = |D_{ii}u_i||D_{ii}^{-1}v_i| \leq \|Du\|\|D^{-1}v\| \leq \omega\mu_k,$$

we have

$$\sigma_k(1-\gamma_k)\mu_k+\alpha(u_i v_i-\gamma_k u^T v/n) \geq \sigma_k(1-\gamma_k)\mu_k-\alpha(\omega\mu_k+\omega\mu_k/n) \geq \bar{\sigma}(1-\gamma_k)\mu_k-2\alpha\omega\mu_k.$$

Since α lies in the range (26) and $\gamma_k \in [\gamma_{\min}, \gamma_{\max}]$, we have

$$\bar{\sigma}(1-\gamma_k)\mu_k-2\alpha\omega\mu_k \geq \bar{\sigma}(1-\gamma_k)\mu_k-2\omega\frac{\bar{\sigma}(1-\gamma_{\max})}{2\omega}\mu_k \geq 0.$$

Therefore (29) and hence (7b) hold for the interval in question.

Finally, from (6), (28), and (26), we have

$$\begin{aligned} n\mu_k'(\alpha) &= (x^T v+y^T u)+2\alpha u^T v \\ &\leq (\sigma_k-1)x^T y+2\alpha\omega\mu_k \\ &\leq (\sigma_k-1)n\mu_k+\bar{\sigma}(1-\gamma_{\max})\mu_k \\ &\leq (\sigma_k+\bar{\sigma}/n-1)n\mu_k \\ &< 0, \end{aligned}$$

since $\sigma_k \in [\bar{\sigma}, 1/2]$. Hence the minimizer of $\mu_k(\alpha)$ subject to the conditions (7a) and (7b) lies beyond the interval (26), and we have the result. ■

We can now show that μ decreases by a factor strictly less than one on each safe iteration.

Theorem 3.6 *If a safe step is taken at iteration k, then*

$$\mu_{k+1} \leq \mu_k\left[1-\frac{\bar{\sigma}(1-\gamma_{\max})}{8\omega}\right], \tag{30}$$

where ω is as defined in Lemma 3.4.

Proof. By Theorem 3.5, $\mu_k(\alpha)$ is decreasing for

$$\alpha \in \left[0, \frac{\bar{\sigma}(1-\gamma_{\max})}{2\omega}\right],$$

and α_k lies beyond this interval. Because of (28), we have

$$\begin{aligned}\mu_k(\alpha) &= \mu_k[1-\alpha(1-\sigma_k)] + \alpha^2 u^T v/n \\ &\leq \mu_k\left[1 - \frac{\bar{\sigma}(1-\gamma_{\max})}{2\omega}(1-\sigma_k)\right] + \frac{\bar{\sigma}^2(1-\gamma_{\max})^2}{4n\omega^2}\omega\mu_k.\end{aligned}$$

Since $1-\sigma_k \geq 1/2$, $\bar{\sigma} \in (0,1/2)$, $1-\gamma_{\max} < 1$, and $n \geq 1$, we have

$$\mu_k(\alpha) \leq \mu_k\left[1 - \frac{\bar{\sigma}(1-\gamma_{\max})}{4\omega}\right] + \frac{\bar{\sigma}(1-\gamma_{\max})}{8\omega}\mu_k = \mu_k\left[1 - \frac{\bar{\sigma}(1-\gamma_{\max})}{8\omega}\right],$$

giving the desired result. ■

If we take fast iterations into account, we find that

$$\mu_{k+1} \leq \mu_k \max\left(1 - \frac{\bar{\sigma}(1-\gamma_{\max})}{8\omega}, \rho\right), \tag{31}$$

so we have geometric convergence of $\{\mu_k\}$ to zero from any starting point $(x^0, y^0) > 0$. For the special choice of starting point (14), (18), we have the following polynomial complexity result.

Corollary 3.7 *Let $\epsilon > 0$ be given. Suppose that the starting point is defined by (14), (18) where $\mu_0 = \xi_x\xi_y \leq 1/\epsilon^\tau$ for some constant $\tau \geq 0$ independent of n. Then there is an integer K_ϵ with*

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

such that $\mu_k \leq \epsilon$ for all $k \geq K_\epsilon$.

Proof. From (30) and the fact that $\omega = O(n^2)$ (Lemma 3.4), we find that there is a constant δ independent of n such that

$$\mu_{k+1} \leq (1-\delta/n^2)\mu_k \tag{32}$$

when a safe step is taken on iteration k. By adjusting δ if necessary, the inequality (32) also holds for fast steps. The result follows from this inequality by a standard argument (see, for example, Zhang [14, Theorem 7.2]).

We note in passing that the same global convergence and complexity results can be obtained even if the exact minimizing $\tilde{\alpha}$ is not found in (6). Instead, we can merely require $\tilde{\alpha}$ to satisfy conditions like those often seen in line search methods for unconstrained minimization. One such pair of conditions is

$$\begin{aligned}\mu_k(\tilde{\alpha}) &\leq \mu_k(0) + \epsilon\tilde{\alpha}\mu_k'(0), &\text{(33a)}\\ \tilde{\alpha} &\geq \tau\tilde{\alpha}_{\max}, &\text{(33b)}\end{aligned}$$

where $\epsilon \in (0, \frac{1}{2})$, $\tau \in (0,1)$, and $\tilde{\alpha}_{\max}$ is the largest value in $[0, \hat{\alpha}]$ for which (33a) holds (where $\hat{\alpha}$ is defined in (7)). It is easy to show, using the techniques in this section, that the exact minimum from (6) satisfies (33); that Theorem 3.5 is still true if we scale the lower bound by τ; and that Corollary 3.7 continues to hold.

If the relaxed conditions (33) are adopted, we can use techniques more like those in practical codes to choose $\tilde{\alpha}$. In particular, we can use an Armijo-like procedure in which we calculate the step length α_{top} along (u^k, v^k) to the boundary of the positive orthant, then back off from this length until (33) and (7) are all satisfied. For appropriate choices of the parameters, the popular choice $\tilde{\alpha} = .9995 \min(1, \alpha_{\text{top}})$ is usually accepted.

∎

4 Bounds for the Fast-Step Components

In this section we show that when $\sigma_k = 0$, the step norms $\|u^k\|$ and $\|v^k\|$ are both $O(\mu_k)$ for all sufficiently large k. This result is essential to the local convergence analysis of the next section. For notational convenience, we use u_B^k to denote the vector whose components are u_i^k for $i \in B$, u_N^k as the subvector made up of u_i^k, $i \in N$, and so on.

We make the following assumption throughout the remainder of this section.

Assumption 2 *$\mathcal{S}^c \neq \emptyset$.*

Given any strictly complementary solution (x^*, y^*), we can define the partition

$$\{1, 2, \cdots, n\} = N \cup B,$$

where

$$B = \{i \,|\, x_i^* > 0\}, \qquad N = \{i \,|\, y_i^* > 0\}. \tag{34}$$

(It is well known that B and N are independent of the particular choice of (x^*, y^*).)

We start by showing that x_N^k and y_B^k can be bounded in terms of μ_k.

Lemma 4.1 *Let $K_{1/2}$ be the smallest integer such that $\nu_k \leq 1/2$ for all $k \geq K_{1/2}$. Then there is a constant $C_4 > 0$ such that*

$$0 < x_i^k \leq C_4 \mu_k, \quad \forall i \in N; \qquad 0 < y_i^k \leq C_4 \mu_k, \quad \forall i \in B. \tag{35}$$

Proof. Note that $K_{1/2}$ is well defined, by the results of Section 3 and the fact that $\{\nu_k\}$ is a decreasing sequence. Let (x^*, y^*) be a strictly complementary solution. By rearranging (16) and noting that

$$(x^0)^T y + x^T y^0 > 0, \qquad (x^*)^T y^* = 0,$$

we have

$$\begin{aligned} x^T y^* + y^T x^* &\leq \frac{\nu_k^2}{1 - \nu_k} n \mu_0 + \frac{n \mu_k}{1 - \nu_k} + \nu_k \left((x^0)^T y^* + (x^*)^T y^0 \right) \\ &\leq 2\nu_k^2 n \mu_0 + 2n\mu_k + \nu_k \left((x^0)^T y^* + (x^*)^T y^0 \right). \end{aligned}$$

By (9), we can therefore define a constant $\bar{C}_4 > 0$ such that

$$x^T y^* + (x^*)^T y \leq \bar{C}_4 \mu_k.$$

Since $(x, y) > 0$ and $(x^*, y^*) \geq 0$ we have

$$\begin{aligned} i \in N \quad &\Rightarrow \quad x_i y_i^* \leq \bar{C}_4 \mu_k \quad \Rightarrow \quad x_i \leq \frac{\bar{C}_4 \mu_k}{y_i^*} \\ i \in B \quad &\Rightarrow \quad x_i^* y_i \leq \bar{C}_4 \mu_k \quad \Rightarrow \quad y_i \leq \frac{\bar{C}_4 \mu_k}{x_i^*}. \end{aligned}$$

The result (35) follows when we define

$$C_4 = \bar{C}_4 \max \left(\max_{i \in B} \frac{1}{x_i^*}, \max_{i \in N} \frac{1}{y_i^*} \right).$$

■

The next result gives the required bounds on half the components of the vector pair (u^k, v^k).

Lemma 4.2 *Let $K_{1/2}$ be as defined in Lemma 4.1. Then there is a constant $C_5 > 0$ such that for all $k \geq K_{1/2}$ we have*

$$\|u_N^k\| \leq C_5 \mu_k, \qquad \|v_B^k\| \leq C_5 \mu_k. \tag{36}$$

Proof. For $i \in N$ we have from (19) that

$$\frac{y_i}{x_i} (u_i)^2 = |D_{ii} u_i|^2 \leq \|Du\|^2 \leq \omega \mu_k.$$

Hence from (35) and (8), we have

$$(u_i)^2 \leq \frac{\omega \mu_k x_i}{y_i} = \frac{\omega \mu_k x_i^2}{x_i y_i} \leq \frac{\omega \mu_k C_4^2 \mu_k^2}{\gamma_{\min} \mu_k} = \frac{\omega C_4^2}{\gamma_{\min}} \mu_k^2.$$

Therefore

$$\|u_N\| \leq \left(\frac{n \omega}{\gamma_{\min}} \right)^{1/2} C_4 \mu_k,$$

giving the first inequality in (36). The proof of the second inequality is similar. ■

To obtain bounds on the remaining components of (u^k, v^k), we need the following two technical lemmas.

Lemma 4.3 (Monteiro and Wright [6, Lemma 2.2].) *Let $f \in \mathbb{R}^q$ and $H \in \mathbb{R}^{p \times q}$ be given. Then there exists a nonnegative constant $L = L(f, H)$ with the property that for any diagonal matrix $S > 0$ and any vector $h \in Range(H)$, the (unique) optimal solution $\bar{w} = \bar{w}(S, h)$ of*

$$\min_w \; f^T w + \frac{1}{2} \|Sw\|^2, \qquad \text{subject to } Hw = h, \tag{37}$$

satisfies

$$\|\bar{w}\|_\infty \leq L \left\{ |f^T \bar{w}| + \|h\|_\infty \right\}.$$

The second technical lemma identifies the components u_B^k and v_N^k as the solution of a quadratic program. It is an extension of a result of Ye and Anstreicher [12, Lemma 3.5] and is proved in Wright [10]. We use D_B^k to denote the diagonal submatrix composed of the elements D_{ii}^k for $i \in B$, and so on.

Lemma 4.4 (Wright [10, Lemma 5.2]) *The vector pair* (u_B^k, v_N^k) *solves the convex quadratic program*

$$\min_{(w,z)} \tfrac{1}{2}\|D_B^k w\|^2 - \sigma_k \mu_k e_B^T (X_B^k)^{-1} w + \tfrac{1}{2}\|(D_N^k)^{-1} z\|^2 - \sigma_k \mu_k e_N^T (Y_N^k)^{-1} z, \tag{38}$$

subject to

$$\begin{aligned} M_{BB} w &= r_B^k - M_{BN} u_N^k + v_B^k, && (39\text{a}) \\ M_{NB} w - z &= r_N^k - M_{NN} u_N^k. && (39\text{b}) \end{aligned}$$

The main result of this section is as follows.

Theorem 4.5 *Let* $K_{1/2}$ *be as defined in Lemma 4.1. Then if* $\sigma_k = 0$*, there is a constant* $C_6 > 0$ *such that*

$$\|u^k\| \le C_6 \mu_k, \qquad \|v^k\| \le C_6 \mu_k.$$

Proof. It follows from Lemmas 4.3 and 4.4 and the inequality (9) that there is a constant $L > 0$ such that

$$\|(u_B^k, v_N^k)\| \le L(\|r^k\| + \|u_N^k\| + \|v_B^k\|) \le L(\nu_k \|r^0\| + 2C_5 \mu_k) \le L\left(\|r^0\|/(\hat{\beta}\mu_0) + 2C_5\right)\mu_k.$$

By combining this inequality with (36), we obtain the result. ∎

Note that the bound on the solution of (37) is independent of the positive diagonal matrix S. This feature is not present in the analysis of [10, Section 5] and [9, Section 5], where estimates of the elements of D_B^k and D_N^k are used in the bound for (u_B^k, v_N^k). In [10], these estimates follow from boundedness of the iterates (x^k, y^k) which, in turn, follow from existence of a strictly feasible point. Since the estimates are no longer required, the strict feasibility assumption is not required either. In [9], the estimates are obtained by synchronizing reductions in μ_k and $\|r_k\|$, which complicates the algorithm considerably. Awkward technical devices such as auxiliary sequences are used in both paper, but are not needed here.

5 Local Superlinear Convergence

In this section, we show that for all k sufficiently large, step k is a fast step and that consequently the sequence $\{\mu_k\}$ converges subquadratically to zero. The treatment in this section follows that of Wright [10, Section 6] and [9, Section 6].

Throughout the analysis, we will make use of the constant C_7 defined by

$$C_7 \triangleq \max\left(1, 2C_6^2\right). \tag{40}$$

We start with a simple technical result.

Lemma 5.1 *For all $k \geq 0$, we have*

$$\frac{\mu_{k+1}}{\bar{\gamma}^{t_{k+1}}} \leq \eta \frac{\mu_k}{\bar{\gamma}^{t_k}},$$

where

$$\eta = \max\left(1 - \frac{\bar{\sigma}(1-\gamma_{\max})}{8\omega}, \frac{\rho}{\bar{\gamma}}\right) < 1,$$

and so the sequence

$$\left\{\frac{\mu_k}{\bar{\gamma}^{t_k}}\right\}$$

converges monotonically and geometrically to zero.

Proof. When a safe step is taken, we have from (30) and $t_k = t_{k+1}$ that

$$\frac{\mu_{k+1}}{\bar{\gamma}^{t_{k+1}}} \leq \left[1 - \frac{\bar{\sigma}(1-\gamma_{\max})}{8\omega}\right] \frac{\mu_k}{\bar{\gamma}^{t_k}}.$$

For a fast step, we have $\mu_{k+1} \leq \rho\mu_k$ and $t_{k+1} = t_k + 1$ and so

$$\frac{\mu_{k+1}}{\bar{\gamma}^{t_{k+1}}} \leq \left(\frac{\rho}{\bar{\gamma}}\right) \frac{\mu_k}{\bar{\gamma}^{t_k}}.$$

The result follows immediately from these two bounds. ∎

We now show that fast steps are taken for all k sufficiently large.

Theorem 5.2 *Define K to be the smallest index such that*

$$\frac{\mu_k}{\bar{\gamma}^{t_k}(\gamma_{\max} - \gamma_{\min})} \leq \frac{\rho}{2C_7} \tag{41}$$

for all $k \geq K$. Then a fast step is taken on iteration k, with step length α_k satisfying

$$1 \geq \alpha_k \geq 1 - C_7 \frac{\mu_k}{\bar{\gamma}^{t_k}(\gamma_{\max} - \gamma_{\min})}. \tag{42}$$

Moreover, we have

$$\mu_{k+1} \leq \frac{2C_7}{\bar{\gamma}^{t_k}(\gamma_{\max} - \gamma_{\min})} \mu_k^2 \leq \rho\mu_k. \tag{43}$$

Proof. The proof is structured like that of Theorem 3.5, in that we show that the conditions (7a) and (7b) hold for all α satisfying

$$\alpha \in \left[0, 1 - C_7 \frac{\mu_k}{\bar{\gamma}^{t_k}(\gamma_{\max} - \gamma_{\min})}\right] \tag{44}$$

and then show that $\mu_k(\alpha)$ is decreasing on this interval.

We start with condition (7a). Note from Theorem 4.5, (40), and $\alpha \in (0,1]$, we have

$$(x+\alpha u)^T(y+\alpha v)/n = (1-\alpha)\mu_k + \alpha^2 u^T v/n \geq (1-\alpha)\mu_k - \frac{C_7}{2}\mu_k^2. \tag{45}$$

Now for α in the interval (44), we have

$$\alpha \leq 1 - C_7 \frac{\mu_k}{\bar{\gamma}^{t_k}(\gamma_{\max}-\gamma_{\min})} \quad \Rightarrow \quad \frac{C_7\mu_k}{2} \leq (1-\alpha)\bar{\gamma}^{t_k}\frac{\gamma_{\max}-\gamma_{\min}}{2} \leq (1-\alpha)\bar{\gamma}^{t_k}.$$

Therefore

$$(1-\alpha)\mu_k - \frac{C_7\mu_k}{2}\mu_k \geq (1-\alpha)\mu_k - \bar{\gamma}^{t_k}(1-\alpha)\mu_k = (1-\bar{\gamma}^{t_k})(1-\alpha)\mu_k = (1-\beta_k)(1-\alpha)\mu_k, \tag{46}$$

since $\beta_k = \bar{\gamma}^{t_k}$ for a fast step. Relation (7a) follows from (45) and (46).

For (7b), we again use Theorem 4.5, (5), (40), and (8) to derive

$$(x_i+\alpha u_i)(y_i+\alpha v_i) = x_i y_i(1-\alpha) + \alpha^2 u_i v_i \geq \gamma_k(1-\alpha)\mu_k - \|u\|\|v\| \geq \gamma_k(1-\alpha)\mu_k - \frac{C_7}{2}\mu_k^2. \tag{47}$$

Meanwhile, assuming that a fast step is computed, we have

$$\gamma_{k+1}(x+\alpha u)^T(y+\alpha v)/n = \gamma_{k+1}(1-\alpha)\mu_k + \gamma_{k+1}\alpha^2 u^T v/n \leq \gamma_{k+1}(1-\alpha)\mu_k + \frac{C_7}{2}\mu_k^2. \tag{48}$$

Combining (47) and (48), we find that (7b) is satisfied if

$$(\gamma_k - \gamma_{k+1})(1-\alpha) \geq C_7\mu_k. \tag{49}$$

Now, since $\bar{\gamma} \in (0, 1/2]$, we have

$$\gamma_k - \gamma_{k+1} = \bar{\gamma}^{t_k-1}(\gamma_{\max}-\gamma_{\min}) - \bar{\gamma}^{t_k}(\gamma_{\max}-\gamma_{\min}) = \bar{\gamma}^{t_k-1}(1-\bar{\gamma})(\gamma_{\max}-\gamma_{\min}) \geq \bar{\gamma}^{t_k}(\gamma_{\max}-\gamma_{\min}).$$

Therefore (49) is satisfied if

$$\bar{\gamma}^{t_k}(\gamma_{\max}-\gamma_{\min})(1-\alpha) \geq C_7\mu_k. \tag{50}$$

But (50) is clearly satisfied for all α in the interval (44), so we deduce that (7b) holds.

Finally, we examine $\mu_k(\alpha)$ from (6). For $\alpha \in [0,1]$ we have

$$\mu_k'(\alpha) = -\mu_k + 2\alpha u^T v/n \leq -\mu_k + C_7\mu_k^2 = -(1-C_7\mu_k)\mu_k.$$

Now from (41), we clearly have $C_7\mu_k < 1$, and therefore $\mu_k'(\alpha) < 0$. Hence the complementarity gap is certainly decreasing on the interval (44). We deduce that the step length α_k lies above the upper bound of the interval (44), so the proof of (42) is complete.

For (43), we use (42) to obtain

$$\begin{aligned}
\mu_{k+1} &= (x+\alpha_k u)^T(y+\alpha_k v)/n \\
&= (1-\alpha_k)\mu_k + \alpha_k^2 u^T v/n \\
&\le \frac{C_7}{\bar{\gamma}^{t_k}(\gamma_{\max}-\gamma_{\min})}\mu_k^2 + \frac{C_7}{2}\mu_k^2 \\
&\le \frac{2C_7}{\bar{\gamma}^{t_k}(\gamma_{\max}-\gamma_{\min})}\mu_k^2, \qquad (51)
\end{aligned}$$

giving the first inequality in (43). The second inequality is an immediate consequence of (41). ■

We can now state our asymptotic rate-of-convergence result.

Theorem 5.3 *The sequence $\{\mu_k\}$ converges superlinearly to zero with Q-order 2.*

Proof. See [10, Theorem 6.3 (ii)]. ■

So far, we have used the term "convergence" to denote convergence of μ_k and $\|r^k\|$ to zero. Fast convergence of the actual iterates (x^k, y^k) to a solution was not proved in Wright [10], [9], but it is easy to show. A key result is a bound on the distance of (x^k, y^k) to the solution set $\mathcal{S}$.

Lemma 5.4 *Suppose that Assumption 2 holds. Then there is a constant C_3 such that*

$$\operatorname{dist}\left((x^k, y^k), \mathcal{S}\right) \triangleq \min_{(x,y)\in\mathcal{S}} \|(x^k, y^k) - (x,y)\|_\infty \le C_3\mu_k.$$

Proof. Note first that for any $(x^*, y^*) \in \mathcal{S}$, we have

$$y - y^* = (Mx + q + r) - (Mx^* + q) = M(x - x^*) + r,$$

and so from Lemma 3.1

$$\begin{aligned}
\|(x,y) - (x^*,y^*)\|_\infty &\le (1+\|M\|_\infty)\|x-x^*\|_\infty + \nu_k\|r^0\|_\infty \\
&\le (1+\|M\|_\infty)\|x-x^*\|_\infty + \mu_k\|r^0\|_\infty/(\hat{\beta}\mu_0). \qquad (52)
\end{aligned}$$

Now from Mangasarian [3, Theorem 2.6], we have that there is a constant $\bar{C}_3$ such that

$$\begin{aligned}
\min_{(x^*, Mx^*+q)\in\mathcal{S}} \|x - x^*\|_\infty &\le \bar{C}_3 \left\|(-Mx-q, -x, x^T(Mx+q))_+\right\|_2 \\
&\le \bar{C}_3 \left\|(r-y, x^T(y-r))_+\right\|_2 \\
&\le \bar{C}_3(\|r\|_2 + x^T y + \|x\|_1\|r\|_\infty) \\
&\le \bar{C}_3(\nu_k\|r^0\|_2 + n\mu_k + \nu_k\|x\|_1\|r^0\|_\infty).
\end{aligned}$$

If we substitute from (13) and (9), we obtain from the last inequality that there is a constant $\hat{C}_3$ such that

$$\min_{(x^*, Mx^*+q)\in\mathcal{S}} \|x - x^*\|_\infty \le \hat{C}_3\mu_k. \qquad (53)$$

The result follows by combining (52) with (53). ■

Theorem 5.5 *If Assumption 2 holds, the iterates (x^k, y^k) converge R-subquadratically to a point in $\mathcal{S}$.*

Proof. We show first that the sequence $\{(x^k, y^k)\}$ is Cauchy. Since fast steps are taken for all $k \geq K$ (Theorem 5.2), then for any k_1, k_2 with $K \leq k_1 < k_2$, we have from Theorem 4.5 that

$$\|(x^{k_2}, y^{k_2}) - (x^{k_1}, y^{k_1})\| \leq \sum_{k=k_1}^{k_2-1} \|(u^k, v^k)\| \leq \sqrt{2} C_6 \sum_{k=k_1}^{k_2-1} \mu_k.$$

From (43), we find that $\{\mu_k\}$ is bounded by a geometric sequence and therefore

$$\sum_{k=k_1}^{k_2-1} \mu_k \leq \sum_{k=k_1}^{k_2-1} \rho^{k-k_1} \mu_{k_1} \leq \frac{\mu_{k_1}}{1-\rho}.$$

Hence,

$$\|(x^{k_2}, y^{k_2}) - (x^{k_1}, y^{k_1})\| \leq \frac{\sqrt{2} C_6}{1-\rho} \mu_{k_1}, \tag{54}$$

which approaches zero as $k_1, k_2 \to \infty$. Hence, the sequence is Cauchy and therefore convergent to a limit point (x^*, y^*) (say). Because of Lemma 5.4 and the fact that $\mathcal{S}$ is closed, we have $(x^*, y^*) \in \mathcal{S}$.

By using the same arguments that led to (54), we have for $k \geq K$ that

$$\|(x^k, y^k) - (x^*, y^*)\| \leq \frac{\sqrt{2} C_6}{1-\rho} \mu_k.$$

The R-subquadratic convergence of (x^k, y^k) to (x^*, y^*) follows from this bound and Theorem 5.3. ∎

6 Extension to MLCP, LP, and QP

We conclude by showing that the algorithm of this paper can be extended to the mixed LCP (2) and, hence, to all the usual formulations of linear and convex quadratic programming problems. The crucial result here is due to Güler [2], who showed that any generalized linear complementarity problem involving a maximal monotone operator can be reformulated as a standard LCP (1). The analysis in this section shows that the operator represented by (2a) is in fact maximal monotone and, hence, satisfies the assumptions of Theorem 3.2 in [2]. The following extension of Assumption 1 is essential to our analysis.

Assumption 3 *The MLCP (2) is feasible; that is, there is a vector triple (x, y, z) with $(x, y) \geq 0$ satisfying (2a).*

We define T to be a *multivalued mapping* from $\mathbb{R}^n$ to subsets of $\mathbb{R}^n$ such that $y \in T(x)$ whenever there exists a $z \in \mathbb{R}^m$ such that (x, y, z) satisfies (2a). The graph of T is given by

$$G(T) = \{(x, y) \mid y = M_{11}x + M_{12}z + q_1, \;\; 0 = M_{21}x + M_{22}z + q_2, \;\; \text{some } z \in \mathbb{R}^m\}. \tag{55}$$

Because of Assumption 3, we have $G(T) \neq \emptyset$. It is easy to check that T is *monotone*, that is, for all $(x, y) \in G(T)$ and $(\bar{x}, \bar{y}) \in G(T)$, we have $(x - \bar{x})^T (y - \bar{y}) \geq 0$. In Theorem 6.3 below, we show that T is in fact *maximal monotone*, that is, there is no other monotone operator $\bar{T} : \mathbb{R}^n \to \mathbb{R}^n$ such that $G(T)$ is strictly contained in $G(\bar{T})$. The following two technical lemmas lay the foundation for this theorem.

Lemma 6.1 *Suppose M is positive semidefinite. Then $Mu = 0 \Leftrightarrow M^T u = 0$.*

Proof.

$$Mu = 0 \quad \Rightarrow \quad u^T M u = 0 \quad \Rightarrow \quad u^T (M + M^T) u = 0,$$

and so u is a minimizer of the convex quadratic function $f(u) = u^T (M + M^T) u$. Therefore

$$\nabla f(u) = 0 \quad \Rightarrow \quad (M + M^T) u = 0 \quad \Rightarrow \quad M^T u = 0,$$

and the forward implication is proved. The converse is similar. ■

Lemma 6.2 *Let the p.s.d. matrix M be partitioned as*

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

where M_{11} and M_{22} are square and

$$M_{\cdot 2} \triangleq \begin{bmatrix} M_{12} \\ M_{22} \end{bmatrix}$$

has full column rank. Then if D is a diagonal matrix with strictly positive diagonal entries, then

$$\begin{bmatrix} M_{11} + D & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$$

is nonsingular.

Proof. Let (x, z) be a vector pair such that

$$\begin{bmatrix} M_{11} + D & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{56}$$

Then, using the p.s.d. property of M, we have

$$\begin{aligned} & \begin{bmatrix} x^T & z^T \end{bmatrix} \begin{bmatrix} M_{11} + D & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = 0 \\ \Rightarrow \quad & \begin{bmatrix} x^T & z^T \end{bmatrix} M \begin{bmatrix} x \\ z \end{bmatrix} + x^T D x = 0 \quad \Rightarrow \quad x^T D x \leq 0 \quad \Rightarrow \quad x = 0. \end{aligned}$$

From (56) we have $M_{\cdot 2} z = 0$, and thus $z = 0$ by the full rank assumption. Hence (56) is satisfied only by $(x, z) = (0, 0)$, and the result is proved. ■

Our main result is the following.

Theorem 6.3 *Suppose that Assumption 3 holds. Then the mapping T whose graph is given by (55) is maximal monotone.*

Proof. We prove the result by appealing to a theorem of Minty [4], which states that T is maximal monotone if and only if $R(I+T)=\mathbb{R}^n$. In the remainder of the proof, we show that for any $w \in \mathbb{R}^n$, there is an $(x,y)\in G(T)$ such that $x+y=w$, and hence $w\in R(I+T)$. In other words, for any w the following linear system must have a solution triple (x,y,z):

$$\begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} - \begin{bmatrix} y \\ 0 \end{bmatrix} = \begin{bmatrix} -q_1 \\ -q_2 \end{bmatrix}, \qquad x+y=w. \tag{57}$$

We show first that solutions to (57) can be obtained from solutions to the following system:

$$\begin{bmatrix} M_{11} & \bar{M}_{12} & -I \\ \bar{M}_{21} & \bar{M}_{22} & 0 \\ I & 0 & I \end{bmatrix} \begin{bmatrix} x \\ \bar{z} \\ y \end{bmatrix} = \begin{bmatrix} -q_1 \\ -\bar{q}_2 \\ w \end{bmatrix}, \tag{58}$$

where $\bar{z}\in\mathbb{R}^{\bar{m}}$, and $\bar{q}_2\in\mathbb{R}^{\bar{m}}$ and the matrices $\bar{M}_{12}$, $\bar{M}_{21}$, $\bar{M}_{22}$ are quantities to be defined, with the crucial property that the submatrix

$$\bar{M}_{\cdot 2} = \begin{bmatrix} \bar{M}_{12} \\ \bar{M}_{22} \end{bmatrix} \in \mathbb{R}^{(n+\bar{m})\times\bar{m}} \tag{59}$$

has full column rank.

Let $v\in\mathbb{R}^m$ be any vector such that $M_{\cdot 2}v=0$, where $M_{\cdot 2}$ is defined as in Lemma 6.2. Then

$$M \begin{bmatrix} 0 \\ v \end{bmatrix} = 0$$

and hence, by Lemma 6.1,

$$M^T \begin{bmatrix} 0 \\ v \end{bmatrix} = 0 \;\Rightarrow\; \begin{bmatrix} M_{21}^T \\ M_{22}^T \end{bmatrix} v = 0. \tag{60}$$

Let $V_2\in\mathbb{R}^{m\times d}$ $(0\le d\le m)$ be the matrix whose columns form a basis for the subspace with the property $M_{\cdot 2}v=0$, and define $V_1\in\mathbb{R}^{m\times(m-d)}$ so that the columns of V_1 are a basis for $R(V_2)^\perp$. Then, defining the nonsingular matrix $V=[V_1\,|\,V_2]$, we have

$$\begin{aligned}
& \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} y-q_1 \\ -q_2 \end{bmatrix} \\
\Leftrightarrow\; & \begin{bmatrix} I & 0 \\ 0 & V^T \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & V \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & V^{-1} \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} = \begin{bmatrix} y-q_1 \\ -V^T q_2 \end{bmatrix} \\
\Leftrightarrow\; & \begin{bmatrix} M_{11} & \bar{M}_{12} & 0 \\ \bar{M}_{21} & \bar{M}_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ (V^{-1}z)_1 \\ (V^{-1}z)_2 \end{bmatrix} = \begin{bmatrix} y-q_1 \\ -V_1^T q_2 \\ -V_2^T q_2 \end{bmatrix},
\end{aligned} \tag{61}$$

where

$$\bar{M}_{12} = M_{12}V_1, \qquad \bar{M}_{22} = V_1^T M_{22} V_1, \qquad \bar{M}_{21} = V_1^T M_{21}.$$

To verify that $\bar{M}_{\cdot 2}$ has full column rank, let w be a vector for which $\bar{M}_{\cdot 2} w = 0$. Then

$$\begin{bmatrix} M_{12} \\ V_1^T M_{22} \end{bmatrix} V_1 w = 0 \;\Rightarrow\; \begin{bmatrix} M_{12} \\ V^T M_{22} \end{bmatrix} V_1 w = 0 \;\Rightarrow\; M_{\cdot 2} V_1 w = 0 \;\Rightarrow\; V_1 w \in R(V_2),$$

and hence $w = 0$ by definition of V_1 and V_2, proving the assertion.

For (61) to be consistent, it is necessary that $V_2^T q_2 = 0$. This follows, however, from Assumption 3 and (60), since

$$q_2 \in R([M_{21} \,|\, M_{22}]) \quad\Rightarrow\quad V_2^T q_2 = 0.$$

Defining $\bar{m} = m - d$ and

$$\begin{aligned} \bar{z} &= (V^{-1}z)_1 && \text{(first } \bar{m} \text{ components of } V^{-1}z\text{)}, \\ \hat{z} &= (V^{-1}z)_2 && \text{(last } m - \bar{m} \text{ components of } V^{-1}z\text{)}, \\ \bar{q}_2 &= V_1^T q_2, \end{aligned}$$

we find that any solution $(x, y, \bar{z})$ of (58) can be transformed into a solution (x, y, z) of (57) by setting

$$z = V \begin{bmatrix} \bar{z} \\ \hat{z} \end{bmatrix},$$

where $\hat{z}$ is chosen arbitrarily.

It remains to show that (58) does in fact have a solution. By adding the third block row in (58) to the first block row, we obtain the equivalent system

$$\begin{bmatrix} M_{11} + I & \bar{M}_{12} & 0 \\ \bar{M}_{21} & \bar{M}_{22} & 0 \\ I & 0 & I \end{bmatrix} \begin{bmatrix} x \\ \bar{z} \\ y \end{bmatrix} = \begin{bmatrix} -q_1 + w \\ -\bar{q}_2 \\ w \end{bmatrix}. \tag{62}$$

Since, by our choice of transformations V_1 and V_2, we clearly have that

$$\bar{M} = \begin{bmatrix} M_{11} & \bar{M}_{12} \\ \bar{M}_{21} & \bar{M}_{22} \end{bmatrix}$$

is p.s.d., while $\bar{M}_{\cdot 2}$ has full rank, it follows from Lemma 6.2 that the upper left 2×2 block in the coefficient matrix of (62) is nonsingular and, hence, that (62) has a unique solution triple $(x, y, \bar{z})$. Hence we have identified $(x, y) \in G(T)$ with $x + y = w$, and our proof is complete. ∎

Because of (61), the graph of T can be restated as

$$G(T) = \{(x, y) \,|\, y = M_{11}x + \bar{M}_{12}\bar{z} + q_1, \;\; 0 = \bar{M}_{21}x + \bar{M}_{22}\bar{z} + \bar{q}_2, \;\; \text{some } \bar{z} \in \mathbb{R}^{\bar{m}}\},$$

with $\bar{M}_{\cdot 2}$ full rank. If we define a matrix $W \in \mathbb{R}^{(n+\bar{m})\times n}$ such that the columns of W form a basis for $R(\bar{M}_{\cdot 2})^{\perp}$, we can eliminate $\bar{z}$ from the definition of $G(T)$ altogether and write

$$G(T) = \{(x, y) \mid Fx - Gy = a\}, \tag{63}$$

where

$$F = W^T \begin{bmatrix} M_{11} \\ \bar{M}_{21} \end{bmatrix}, \qquad G = W^T \begin{bmatrix} I \\ 0 \end{bmatrix}, \qquad a = W^T \begin{bmatrix} -q_1 \\ -\bar{q}_2 \end{bmatrix}.$$

The form (63) is the canonical form used by Güler. It is not difficult to verify that a condition used by Güler to prove maximal monotonicity of T—namely, nonsingularity of $F + G$—is satisfied by (63). As shown in [2, Theorem 3.2] conversion of (63) to the form (1) can now be achieved by premultiplying $Fx - Gy = a$ by a nonsingular operator and possibly swapping some components of x and y. As Güler notes, this reformulation process need not actually be carried out to apply the algorithm of Section 2 to the problem (2). Instead, we can extend our algorithm to (2) and note that the sequence of (x^k, y^k) iterates generated by the extended algorithm is the same as the sequence that would be generated by the basic algorithm applied to the LCP reformulation, subject possibly to the swapping of components between x^k and y^k just mentioned.

The extension of our algorithm to (2) is fairly obvious, so we omit the details, noting simply that the linear system to be solved at each iteration is

$$\begin{bmatrix} M_{11} & M_{12} & -I \\ M_{21} & M_{22} & 0 \\ Y^k & 0 & X^k \end{bmatrix} \begin{bmatrix} u^k \\ w^k \\ v^k \end{bmatrix} = \begin{bmatrix} y^k - M_{11}x^k - M_{12}z^k - q_1 \\ -M_{21}x^k - M_{22}z^k - q_2 \\ -X^kY^ke + \tilde{\sigma}\mu_k e \end{bmatrix}, \tag{64}$$

which is a generalization of (5). Existence of a solution to (64) follows from the same reasoning as in the proof of Theorem 6.3. We need to note that when (2) is feasible, we have

$$-M_{21}x^k - M_{22}z^k - q_2 \in R([M_{21} \mid M_{22}])$$

and, hence,

$$V_2^T[-M_{21}x^k - M_{22}z^k - q_2] = 0.$$

Also, we need to apply Lemma (6.2) with $D = (X^k)^{-1}Y^k$. Although the step components u^k and v^k are uniquely determined by (64), the w^k components are not, unless the submatrix $M_{\cdot 2}$ has full rank.

Finally, we note that our analysis depends crucially on the existence of feasible points for (1) and (2). When (2) arises from an LP or QP, a feasible primal-dual point must exist. This requirement is a little troubling, since in practice many LPs are either primal or dual infeasible. Ye, Todd, and Mizuno [13] and Xu, Hung, and Ye [11] have alleviated this difficulty in the case of linear programming by describing augmentation/reformulation of an LP in standard form, which has the property that the resulting mixed LCP possesses both a feasible point and a strictly complementary solution. Since all the assumptions of this paper are satisfied by these reformulations, our algorithm can be applied with confidence.

References

[1] R. W. Cottle, J.-S. Pang, and R. E. Stone, *The Linear Complementarity Problem* (Academic Press, 1992).

[2] O. Güler, Generalized linear complementarity problems, Manuscript (1993).

[3] O. L. Mangasarian, Error bounds for nondegenerate monotone linear complementarity problems, Math. Progr. 48(1990) 437–445.

[4] G. J. Minty, Monotone (nonlinear) operators in Hilbert space, Duke J. Math. 29(1992) 341–346.

[5] S. Mizuno, Polynomiality of Kojima-Meggido-Mizuno infeasible-interior-point algorithm for linear programming, Tech. Report 1006, School of Operations Research and Industrial Engineering, Cornell University, Ithaca, N.Y. (1993).

[6] R. D. C. Monteiro and S. J. Wright, Superlinear primal-dual affine scaling algorithms for LCP, Preprint MCS–P360–0693, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill. (1993). To appear in Math. Progr., Series A.

[7] R. D. C. Monteiro and S. J. Wright, Local convergence of interior-point algorithms for degenerate monotone LCP, Comput. Optim. Appls 3(1994) 131–155.

[8] F. A. Potra, An infeasible interior-point predictor-corrector algorithm for linear programming, Tech. Report 26, Department of Mathematics, University of Iowa, Iowa City, Iowa (1992). To appear in SIAM J. Optim.

[9] S. J. Wright, A path-following infeasible-interior-point algorithm for linear complementarity problems, Optim. Methods Softw. 2(1993) 79–106.

[10] S. J. Wright, An infeasible-interior-point algorithm for linear complementarity problems, Math. Progr. 67(1994) 29–52.

[11] X. Xu, P. Hung, and Y. Ye, A simplified homogeneous and self-dual linear programming algorithm and its implementation, Manuscript (1993).

[12] Y. Ye and K. Anstreicher, On quadratic and $O(\sqrt{n}L)$ convergence of a predictor-corrector algorithm for LCP, Math. Progr. Series A 62(1993) 537–551.

[13] Y. Ye, M. J. Todd, and S. Mizuno, An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm, Working Paper 92-11, College of Business Administration, The University of Iowa (1992). To appear in Math. Oper. Res.

[14] Y. Zhang, On the convergence of a class of infeasible-interior-point methods for the horizontal linear complementarity problem, SIAM J. Optim. 4(1994) 208–227.