A Superlinear Infeasible-Interior-Point Algorithm for Monotone Complementarity Problems

Stephen Wright* and Daniel Ralph†

Abstract

We use the globally convergent framework proposed by Kojima, Noma, and Yoshise to construct an infeasible-interior-point algorithm for monotone nonlinear complementarity problems. Superlinear convergence is attained when the solution is nondegenerate and also when the problem is linear. Numerical experiments confirm the efficacy of the proposed approach.

1   Introduction

We consider the problem of finding a vector pair $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ such that

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

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable in an open set containing the nonnegative orthant of $\mathbb{R}^n$ (denoted by $\mathbb{R}^n_+$) and monotone, that is,

$$(x' - x)^T (f(x') - f(x)) \geq 0 \ \text{ for all } x', x \in \mathbb{R}^n_+.$$

Problem (1) is a *monotone nonlinear complementarity problem*, abbreviated as NCP. We use $\mathcal{S}$ to denote the solution set for (1).

Interior-point algorithms for problems of this type have been considered recently by Kojima, Noma. and Yoshise [4], Güler [3], and Potra and Ye [6]. In [4], the authors consider a broad class of infeasible-interior-point algorithms for (1) and show that, assuming continuous differentiability of f, at least one of three scenarios eventually occurs: The algorithm reaches the vicinity of a solution to (1), it reaches the vicinity of a solution to a nearby problem, or it returns an error condition that indicates that no solution of (1) exists in a certain large nonnegative neighborhood of $(0, 0)$. The algorithm we propose in this paper falls into

---

1

DISCLAIMER

the class considered in [4], and so this global convergence result holds. Under additional assumptions on f and $\mathcal{S}$, our algorithm exhibits superlinear convergence.

Our algorithm is based on the one described in [9] for linear complementarity problems (for which f has the form $f(x) = Ax + q$ for some $n \times n$ positive semidefinite matrix A). It differs in two main respects. First, the step generated by the algorithm of [9] depends on the entire iteration history through its use of a parameter t_k, which counts the number of fast steps taken prior to iteration k. However, in order to fit the framework of [4], the mapping from one iterate to the next must be determined completely by the current iterate (x^k, y^k), so this explicit dependence on t_k must be eliminated. Second, it is no longer practical to choose the step length α to be the largest scalar in $(0, 1]$ that satisfies certain central path/infeasibility conditions, since to do so would require frequent use of expensive root-finding techniques applied to components of $f(x)$. Instead, we use Armijo conditions in conjunction with a backtracking line search technique, in which the initial trial step length is chosen judiciously.

We present some notation and define the algorithm in Section 2. In Section 3, we state the results from Kojima, Noma, and Yoshise [4] and show that the algorithm of Section 2 fits this framework. In Section 4, we prove the rate-of-convergence results under the assumption that the step lengths are of the same order as the current duality gap estimate $x^T y/n$. We show in Section 5 that this assumption holds when the minimizer is unique and nondegenerate. Computational experience is reported in Section 6.

Unless otherwise specified, $\| \cdot \|$ denotes the Euclidean norm of a vector. We frequently use (x, y) as shorthand for the vector $(x^T, y^T)^T \in \mathbb{R}^{2n}$. Iteration indices (usually k) appear as superscripts on vectors and matrices and as subscripts on scalars. Subscripts are used to indicate components of vectors and matrices. The notation $\mathcal{B}((\bar{x}, \bar{y}), \delta)$ is used for the closed ball

$$\mathcal{B}((\bar{x}, \bar{y}), \delta) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mid \|(x, y) - (\bar{x}, \bar{y})\| \le \delta\}.$$

The vector $(1, 1, \cdots, 1)$ is denoted by e, while z_+ is obtained by replacing all negative components in the vector z by zero. If (x^*, y^*) is a solution of (1), we can partition $\{1, 2, \ldots, n\}$ into two index sets B and N, where

$$x_i^* = 0 \ \forall i \in N, \qquad y_i^* = 0 \ \forall i \in B. \tag{2}$$

The solution is *strictly complementary* if $x^* + y^* > 0$.

2 The Algorithm

Given a starting point with $(x^0, y^0) > (0, 0)$, the algorithm generates a sequence of iterates $(x^k, y^k) > (0, 0)$, $k = 1, 2, \cdots$. With each vector pair $(x, y) > 0$ we associate the following quantities:

$$\mu = x^T y/n, \qquad r = y - f(x), \qquad e = (1, 1, \cdots, 1)^T,$$
$$X = \text{diag}(x_1, x_2, \cdots, x_n), \qquad Y = \text{diag}(y_1, y_2, \cdots, y_n).$$

2

When $(x, y) = (x^k, y^k)$, we sometimes attach a subscript or superscript k to the quantities μ, r, X, Y to make the dependence on (x^k, y^k) explicit.

The main computational operation at each iteration is solution of the $2n$-dimensional linear system

$$\begin{bmatrix} Df(x) & -I \\ Y & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} y - f(x) \\ -XYe + \tilde{\sigma}\mu_k e \end{bmatrix}, \tag{3}$$

where $\tilde{\sigma} \in [0, .5]$. The steps generated by the algorithm have the form

$$x(\alpha) = x + \alpha \Delta x, \tag{4a}$$
$$y(\alpha) = y + \alpha \Delta y + g(\alpha), \tag{4b}$$

where

$$g(\alpha) = f(x + \alpha \Delta x) - f(x) - \alpha Df(x)\Delta x. \tag{5}$$

Note that

$$y(\alpha) - f(x(\alpha)) = (1 - \alpha)(y - f(x)) = (1 - \alpha)r. \tag{6}$$

At each iteration, the formulae (3) and (4) are used to calculate a *fast* step and, if it is unsuccessful, a *safe* step. Safe steps ensure that desirable global convergence properties hold, while fast steps ensure rapid local convergence. These two types of steps are distinguished by different choices of the centering parameter $\tilde{\sigma}$ in (3), different choices of the initial trial step size for the Armijo line search, and slightly different acceptance criteria for the step length. While the formal treatment and theoretical utility of fast and safe steps are quite different, the distinction between them need not be so wide in practice. The wide latitude allowed to the user in the choice of $\tilde{\sigma}$ and initial trial step size for safe steps means that safe steps can be made to perform like fast steps during the later stages of the algorithm.

The overall algorithm is parametrized by a variety of positive scalar constants, which we specify now and explain later, as they arise in subsequent discussions:

$$\chi \in (0, 1), \qquad \bar{\sigma} \in (0, \tfrac{1}{2}), \qquad \bar{\alpha} \in (0, 1],$$
$$\kappa \in (0, 1), \qquad \hat{\tau} \in (0, 1), \qquad \beta_0 > 0,$$
$$0 < \gamma_{\min} < \gamma_{\max} \le \tfrac{1}{2}, \qquad \bar{\gamma} \in (0, \tfrac{1}{2}), \qquad \rho \in (0, \min(\bar{\gamma}^{1/\hat{\tau}}, 1 - \kappa)).$$

The starting point (x^0, y^0) is assumed to satisfy

$$\|r^0\| \le \beta_0 \mu_0, \qquad x_i^0 y_i^0 \ge \gamma_{\max}\mu_0. \tag{7}$$

The main algorithm can now be specified.

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

$\quad$ **if** $\quad \mu_k = 0,$
$\quad$ **then** terminate with solution (x^k, y^k);

$$(x^{k+1}, y^{k+1}) \leftarrow \mathbf{fast}(x^k, y^k);$$
$$\mathbf{if} \quad \mu_{k+1} > \rho\mu_k$$
$$\mathbf{then} \ (x^{k+1}, y^{k+1}) \leftarrow \mathbf{safe}(x^k, y^k);$$
$$\mathbf{end \ if}$$
$$\mathbf{end \ for}.$$

Note that the fast step is taken if it produces at least a factor of ρ decrease in the complementarity gap μ. Otherwise, the algorithm reverts to the safe step. The coefficient matrix in (3) is the same for both fast and safe steps, so only one matrix factorization is required per iteration.

The safe step procedure is defined as follows.

$\mathbf{safe}(x, y):$

> choose $\tilde{\sigma} \in [\bar{\sigma}, \frac{1}{2}]$, $\alpha^0 \in [\bar{\alpha}, 1]$;
> solve (3) to find $(\Delta x, \Delta y)$;
> calculate

$$\tilde{\gamma} = \min(\min_{i=1,\cdots,n} x_i y_i / \mu, \ \gamma_{\max}); \tag{8}$$

> choose α to be the first element in the sequence $\alpha^0, \chi\alpha^0, \chi^2\alpha^0, \cdots$,
>   such that the following conditions are satisfied:

$$x_i(\alpha)y_i(\alpha) \geq \tilde{\gamma}\, x(\alpha)^T y(\alpha)/n, \tag{9a}$$
$$\kappa\alpha(1 - \tilde{\sigma})\mu \leq \mu - x(\alpha)^T y(\alpha)/n \leq \alpha\mu; \tag{9b}$$

$\mathbf{return} \ (x(\alpha), y(\alpha)).$

A nonzero centering term is used, allowing us to move a nontrivial distance along the search direction while staying in the set defined by

$$\{(x, y) \mid x_i y_i \geq \gamma\mu\} \tag{10}$$

where $\gamma = \tilde{\gamma}$ (see formula (9a)). In the second acceptance condition (9b), the left inequality ensures a "sufficient decrease" in the objective function. A condition of this kind is present in most optimization algorithms based on line searchs. The purposes of the right inequality in (9b) is to prevent improvement in the complementarity gap μ from outpacing improvement in the infeasibility, measured by r. The relevant result is proved in Lemma (2.2).

Fast-step calculations are a little more complicated. Since it is not permissible in the framework of [4] to maintain a counter t_k of the number of fast steps taken prior to iteration k, we form an estimate $\tilde{t}$ of its value by examining the properties of the current iterate. The integer $\tilde{t}$ is in turn used to form $\hat{\gamma}$ and $\hat{\beta}$, which are used in the acceptance criteria for the step length α. We show in later analysis that $\hat{\gamma}$ and $\hat{\beta}$ have essentially the same properties as the quantities $\tilde{\gamma}$ and $\tilde{\beta}$ in the algorithm of [10].

4

$\mathrm{fast}(x, y)$:

> solve (3) with $\tilde{\sigma} = 0$ to find $(\Delta x, \Delta y)$;
> calculate $\tilde{\gamma}$ as in (8);
> define $\hat{\gamma} = \gamma_{\min} + \bar{\gamma}(\tilde{\gamma} - \gamma_{\min})$;
> calculate

$$\tilde{\beta} = \frac{\beta_0 \mu}{\|r\|}, \tag{11}$$

> and define $\tilde{t}$ to be the smallest positive integer such that

$$\prod_{j=1}^{\tilde{t}} (1 - \bar{\gamma}^j) \leq \tilde{\beta},$$

> (with $\tilde{t} = 0$ if $\tilde{\beta}_k \geq 1$);
> if $\quad \tilde{t} = \infty \quad$ then $\mathrm{return}(x, y)$;
> define $\hat{\beta} = \bar{\gamma}^{\tilde{t}+1}$ (with $\hat{\beta} = 1$ if $r = 0$);
> define

$$\alpha^0 = 1 - \frac{\mu^{\tilde{t}}}{\min(\tilde{\gamma} - \hat{\gamma}, \hat{\beta})}; \tag{12}$$

> if $\quad \alpha^0 \leq 0 \quad$ then $\mathrm{return}(x, y)$;
> choose α to be the first element in the sequence $\alpha^0, \chi\alpha^0, \chi^2\alpha^0, \cdots$,
> such that the following conditions are satisfied:

$$\begin{aligned} x_i(\alpha) y_i(\alpha) &\geq \hat{\gamma}\, x(\alpha)^T y(\alpha)/n, & \text{(13a)} \\ x(\alpha)^T y(\alpha) &\geq (1-\alpha)(1-\hat{\beta})n\mu; & \text{(13b)} \end{aligned}$$

> $\mathrm{return}(x(\alpha), y(\alpha))$.

There is no centering component in this step, since $\tilde{\sigma} = 0$. It is therefore necessary to relax the value of γ in (10) from $\tilde{\gamma}$ to $\hat{\gamma}$ to ensure that we can move a nontrivial distance along this direction while staying in this set. The second acceptance criterion (13b) is again motivated by our wish to not allow improvement in μ to outpace improvement in r. There is no Armijo condition for the fast step. Instead, a "sufficient decrease" condition is enforced in the main program, since the fast step is accepted only if $\mu_{k+1} \leq \rho\mu_k$.

We stress again that we use the subscripted notation $\tilde{\gamma}_k$, $\hat{\gamma}_k$, $\hat{\beta}_k$, $\tilde{\beta}_k$, and $\tilde{t}_k$ to denote the values of $\tilde{\gamma}$, $\hat{\gamma}$, $\hat{\beta}$, $\tilde{\beta}$, and $\tilde{t}$ associated with the k-th iterate (x^k, y^k). We use α_k to denote the

value of α used by the step that is actually accepted, whether it comes from **fast** or **safe**. For the purposes of subsequent analysis, we also define

$$\nu_k = \prod_{j=0}^{k-1}(1 - \alpha_j).$$

From (6), we have that

$$r^k = \nu_k r^0. \tag{14}$$

Suppose that successful fast steps are taken at iterations $k_1, k_2, \cdots$, with

$$0 \le k_1 < k_2 < \cdots,$$

and that safe steps are taken at all other iterations. If we define t_k to be the total number of fast steps accepted prior to iteration k, then clearly

$$t_{k_l} = l - 1, \qquad l = 1, 2, \cdots, \tag{15}$$

and

$$t_k = l, \qquad k = k_l + 1, \cdots, k_{l+1}, \qquad l = 1, 2, \cdots. \tag{16}$$

The following result relates the settings of $\hat{\gamma}_k$ and $\hat{\beta}_k$ on each fast step to t_k. Its proof is rather technical and is relegated to the appendix.

Lemma 2.1 *If $r^0 \ne 0$, then for each iteration k_l, $l = 1, 2, \cdots$, at which a fast step is taken, we have for $l = 1, 2, \cdots$, that the following inequalities hold.*

$$\tilde{t}_{k_l} \le t_{k_l} = l - 1, \tag{17}$$

$$\tilde{\gamma}_{k_l} - \gamma_{\min} \ge \bar{\gamma}^{l-1}(\gamma_{\max} - \gamma_{\min}), \tag{18}$$

$$\tilde{\gamma}_{k_l} - \hat{\gamma}_{k_l} \ge \bar{\gamma}^{l-1}(1 - \bar{\gamma})(\gamma_{\max} - \gamma_{\min}), \tag{19}$$

$$\hat{\beta}_{k_l} \ge \bar{\gamma}^l, \tag{20}$$

$$\min(\tilde{\gamma}_{k_l} - \hat{\gamma}_{k_l}, \hat{\beta}_{k_l}) \ge \bar{\gamma}^l(\gamma_{\max} - \gamma_{\min}). \tag{21}$$

If $r^0 = 0$, the inequalities (18), (19), and (21) are satisfied.

Finally, we show that the improvement in complementarity μ_k/μ_0 cannot exceed the improvement in feasibility $\|r^k\|/\|r^0\|$, modulo a constant factor $\beta_L > 0$.

Lemma 2.2 *If $r^0 \ne 0$, all iterates (x^k, y^k) satisfy*

$$\frac{\mu_k}{\|r^k\|} \ge \frac{\beta_L}{\beta_0}, \tag{22}$$

where

$$1 > \beta_L = \prod_{j=1}^{\infty}(1 - \bar{\gamma}^j) > e^{-3/2}.$$

Proof. The first inequality follows from (9b), (13b), and (6), with similar arguments to those in the proof of Lemma 2.1, which we do not repeat here. The inequality $\beta_L > e^{-3/2}$ is proved in [10, Lemma 3.1]. ∎

6

3 Global Convergence

Kojima, Noma, and Yoshise [4, Section 4.1] analyze an algorithm that would be equivalent to our algorithm if we allowed only safe steps to be taken. They show that it fits into the framework that allows their global convergence result to hold. We show in this section that the use of fast steps does not disqualify our algorithm from the framework of [4], and so the main global convergence result of that paper holds.

Throughout the section we assume only that f is continuously differentiable and monotone.

The model algorithm of Kojima, Noma, and Yoshise consists of three fundamental components:

- An admissible set $\Omega \subset (\mathbb{R}_{++}^n \times \mathbb{R}_{++}^n) \cup \mathcal{S}$ (where $\mathbb{R}_{++}^n$ is the strictly positive orthant in $\mathbb{R}^n$) to which all iterates are confined. In our case,

$$\Omega = \left\{ (x,y) \geq 0 \mid \|y - f(x)\| \leq (\beta_0/\beta_L)(x^T y/n), \ x_i y_i \geq \bar{\gamma}(x^T y/n), \ i = 1, \cdots, n \right\}.$$

- A merit function $\psi(x,y)$, which in our case is simply $\psi(x,y) = x^T y = n\mu$.

- An algorithmic mapping $\mathcal{A}$ that produces a new point $(x^+, y^+) \in \Omega$ from a given $(x,y) \in \Omega$. In our case, $\mathcal{A}$ is the fast step calculation whenever it produces a decrease factor of at least ρ in μ; otherwise, $\mathcal{A}$ is the safe-step calculation.

The main result of [4] is obtained if the following condition is satisfied. (Note that the items in this condition are obtained by combining Conditions 2.1 and 2.3 of [4].)

Condition 1 *The admissible set Ω, merit function ψ, and algorithmic mapping $\mathcal{A}$ satisfy the following conditions:*

(i) $(x^0, y^0) \in \Omega$;

(ii) $\Omega = \Omega_{++} \cup \Omega_{\mathcal{S}}$ *is a subset of* $(\mathbb{R}_{++}^n \times \mathbb{R}_{++}^n) \cup \mathcal{S}$ *and is closed in* $(\mathbb{R}^n \times \mathbb{R}^n)$, *where*

$$\Omega_{++} = \Omega \cap (\mathbb{R}_{++}^n \times \mathbb{R}_{++}^n), \qquad \Omega_{\mathcal{S}} = \Omega \cap \mathcal{S};$$

(iii) ψ *is a real-valued and continuous function on* Ω_{++};

(iv) $\mathcal{A}$ *is a point-to-set mapping from* Ω *into the collection of nonempty subsets of* Ω. *For every* $(\bar{x}, \bar{y}) \in \Omega_{++}$, *there exist positive numbers* δ *and* ϵ *such that if* $(x,y) \in \mathcal{B}((\bar{x}, \bar{y}), \delta) \cap \Omega$ *and* $(x^+, y^+) \in \mathcal{A}(x,y)$, *then either*

$$(x^+, y^+) \in \Omega_{\mathcal{S}}$$

or

$$(x^+, y^+) \in \Omega_{++} \text{ and } \psi(x^+, y^+) \leq \psi(x,y) - \epsilon;$$

7

(v) $\sup\{x^T y \mid (x,y) \in \Omega_{++}, \psi(x,y) \leq T\} < \infty$ *for every sufficiently large* $T \in \mathbb{R}$;

(vi) if $(x^+, y^+) \in \mathcal{A}(x,y)$, *then*

$$y^+ - f(x^+) = (1-\theta)(y - f(x)),$$

for some $\theta \in (0,1]$.

It is immediately clear from our definition of Ω, ψ, and $\mathcal{A}$ that our algorithm satisfies Condition 1(i), (ii), (iii), (v), and (vi). The following two lemmas show that the remaining condition, (iv), is also satisfied. The first of these lemmas essentially shows that the requirements (9) on the step length α_k in a safe-step calculation are satisfied for all α_k sufficiently small. In keeping with the definition of the generic algorithm of [4], in which the mapping $\mathcal{A}$ does not depend explicitly on the iteration history, we state the result without reference to the iteration counter k. The proof of the first result can be found in the appendix.

Lemma 3.1 *Suppose that* $(\bar{x}, \bar{y}) \in \Omega_{++}$. *Then there exist* $\hat{\delta} > 0$ *and* $\hat{\alpha} \in (0,1]$ *such that for any* $(x,y) \in \mathcal{B}((\bar{x}, \bar{y}), \hat{\delta}) \cap \Omega$, *the calculations (3), (4), and (5) applied to the point* (x,y) *with* $\tilde{\sigma} \in [\bar{\sigma}, 1/2]$ *will yield* $(x(\alpha), y(\alpha))$ *satisfying the conditions (9) for all* $\alpha \in [0, \hat{\alpha}]$.

Lemma 3.2 *Given any* $(\bar{x}, \bar{y}) \in \Omega_{++}$, *there are constants* $\delta > 0$ *and* $\epsilon > 0$ *such that if one step of the algorithm is applied to any point* $(x,y) \in \mathcal{B}((\bar{x}, \bar{y}), \delta) \cup \Omega$, *the new point* (x^+, y^+) *generated by this process has* $(x^+)^T y^+ \leq x^T y - \epsilon$. *Hence Condition 1(iv) is satisfied.*

Proof. It is easy to check that the result (x^+, y^+) of any fast or safe step satisfies the conditions for membership in Ω, so we need only find an $\epsilon > 0$ that satisfies the decrease condition.

Let $\hat{\alpha}$ and $\hat{\delta}$ be as defined in Lemma 3.1. We can clearly choose a $\delta \leq \hat{\delta}$ and $C_2 > 0$ such that

$$x^T y \geq C_2 \qquad \forall (x,y) \in \mathcal{B}((\bar{x}, \bar{y}), \hat{\delta}).$$

If a safe step is calculated from the point (x,y), then from Lemma 3.1 we have that (9) are satisfied for all $\alpha \in [0, \hat{\alpha}]$. Because of the backtracking nature of the step length procedure, it is easy to see that the step length parameter α actually generated by the safe-step procedure satisfies

$$\alpha \geq \min(\chi \hat{\alpha}, \bar{\alpha}),$$

where $\bar{\alpha}$ is the lower bound on the initial trial step length. Setting $\epsilon^{(1)} = (\kappa/2) C_2 \min(\chi \hat{\alpha}, \bar{\alpha})$, we have from (9b) that

$$(x^+)^T y^+ = x(\alpha)^T y(\alpha) \leq x^T y - \kappa(1-\tilde{\sigma})\alpha x^T y \leq x^T y - \kappa(1/2) C_2 \min(\chi \hat{\alpha}, \bar{\alpha}) = x^T y - \epsilon^{(1)}.$$

The other algorithmic possibility is that a successful fast step is taken from (x,y). Because of the acceptance criterion associated with such a step, we have that

$$(x^+)^T y^+ \leq \rho x^T y = x^T y - (1-\rho) x^T y \leq x^T y - \epsilon^{(2)},$$

8

where we have defined $\epsilon^{(2)} = (1 - \rho)C_2$.

The result of the lemma follows by taking $\epsilon = \min(\epsilon^{(1)}, \epsilon^{(2)})$. ∎

Having shown that Condition 1 holds in our case, we can state the main results, both of which follow immediately from results in Kojima, Noma, and Yoshise [4]. The first result is essentially [4, Lemma 2.2].

Theorem 3.3 *The sequence (x^k, y^k) behaves in one of the following three ways.*

(A) $(x^k, y^k) \in \mathcal{S}$ for some $k < \infty$.

(B) The sequence $\{(x^k, y^k)\}$ is bounded, and every limiting point of the sequence belongs to $\mathcal{S} \cap \Omega$.

(C) The sequence $\{(x^k, y^k)\}$ is unbounded.

The second result follows immediately from [4, Theorem 2.5].

Theorem 3.4 *Let ϵ be any small positive number and M be any positive number. Then there exists a finite integer p such that at least one of the following three alternatives occurs.*

(A') $(x^p)^T y^p < \epsilon$ and $\nu_p \|r^0\| < \epsilon$.

(B') $r^0 \neq 0$, and $\nu_p \|r^0\| < \epsilon$.

(C') $r^0 \neq 0$, $\nu_p \|r^0\| \geq \epsilon$, and $\nu_p (r^0)^T x^p - (x^p)^T y^p \geq \nu_p M$.

In the case (C'), the set

$$\mathcal{T}(M) = \{(x, y) \in (\mathbb{R}^n_+ \times \mathbb{R}^n_+) \mid (r^0)^T x \leq M\}$$

contains no solutions of (1).

4 Superlinear Local Convergence

We show here that if the algorithm does not terminate finitely or diverge (that is, if alternative (B) of Theorem 3.3 occurs), then the complementarity gap converges superlinearly to zero, under certain assumptions. The Q-order of convergence is at least $1 + \hat{\tau}$, where $\hat{\tau} \in (0, 1)$ is the user-defined parameter in the main algorithm. The analysis in this section is similar to that of Wright [11], which deals with the case of linear f, but there are a number of complications because of the nonlinearity of f and the simpler line search procedure.

We define a near-solution subset of Ω_{++} by

$$\Omega_{++}(\tilde{\delta}) = \{(x, y) \in \Omega_{++} \mid \operatorname{dist}((x, y), \mathcal{S}) \leq \tilde{\delta}\},$$

where

$$\operatorname{dist}((x, y), \mathcal{S}) \triangleq \min\{\|(x^*, y^*) - (x, y)\| \mid (x^*, y^*) \in \mathcal{S}\}.$$

Assumption 1 *The sequence generated by the algorithm is bounded and does not terminate finitely. Moreover, there is $\tilde{\delta} > 0$ such that*

(i) for all $(x, y) \in \Omega_{++}(\tilde{\delta})$ there is a constant $C_3 > 0$ such that a fast step $(\Delta x, \Delta y)$ calculated via (3) with $\tilde{\sigma} = 0$ satisfies $\|(\Delta x, \Delta y)\| \leq C_3 \mu$;

(ii) $Df(\cdot)$ is Lipschitz continuous on the set $\{x \,|\, (x, y) \in \Omega_{++}(\tilde{\delta})$ for some $y\}$.

We show in Wright [11] that this assumption and therefore the conclusions of this section hold when f is linear and S contains a strictly complementary solution (not necessarily unique).

In our first result, we show that the initial step length in the fast steps (12) approaches 1, giving the possibility of substantial progress. The result is a consequence of Assumption 1 and Theorem 3.3.

Lemma 4.1 *If Assumption 1 holds, then*

(i) $\mu_k \downarrow 0$;

(ii) $\mu_k^{\tilde{\tau}}/\bar{\gamma}^{t_k} \downarrow 0$;

(iii) $\mu_k^{\tilde{\tau}}/\min(\tilde{\gamma}_k - \hat{\gamma}_k, \hat{\beta}_k) \to 0$.

Proof. Assumption 1 implies that alternative (B) of Theorem 3.3 holds, and so all limit points of $\{(x^k, y^k)\}$ are in $S \cap \Omega$. Since $\mu = 0$ for all $(x, y) \in S \cap \Omega$ and since $\{\mu_k\}$ is a decreasing sequence, (i) follows.

For (ii), we consider the effect of iteration $k - 1$ on the value of t_k. If a safe step was taken at iteration $k - 1$, then we have $t_k = t_{k-1}$ and $\mu_k < \mu_{k-1}$, so certainly

$$\frac{\mu_k^{\tilde{\tau}}}{\bar{\gamma}^{t_k}} < \frac{\mu_{k-1}^{\tilde{\tau}}}{\bar{\gamma}^{t_{k-1}}}.$$

If a fast step is taken, then $t_k = t_{k-1} + 1$ and $\mu_k \leq \rho \mu_{k-1}$. Hence, by the definition of ρ we have

$$\frac{\mu_k^{\tilde{\tau}}}{\bar{\gamma}^{t_k}} \leq \frac{\rho^{\tilde{\tau}} \mu_{k-1}^{\tilde{\tau}}}{\bar{\gamma} \bar{\gamma}^{t_{k-1}}} < \frac{\bar{\gamma} \mu_{k-1}^{\tilde{\tau}}}{\bar{\gamma} \bar{\gamma}^{t_{k-1}}} = \frac{\mu_{k-1}^{\tilde{\tau}}}{\bar{\gamma}^{t_{k-1}}}. \tag{23}$$

Hence, the sequence $\mu_k^{\tilde{\tau}}/\bar{\gamma}^{t_k}$ is decreasing. To see that it decreases to zero, consider two cases. First, if only a finite number of fast steps are taken, we have that t_k is constant for all k sufficiently large, while $\mu_k \downarrow 0$, so we obtain the result in this case. Second, if there are an infinite number of fast steps, we have from (23) that $\mu_k^{\tilde{\tau}}/\bar{\gamma}^{t_k}$ decreases by a factor of at least $\rho^{\tilde{\tau}}/\bar{\gamma} < 1$ on each such step. Hence, the subsequence corresponding to the fast steps decreases geometrically to zero, so by monotonicity the whole sequence converges to zero, and (ii) holds in this case also.

For (iii), we note from (17) and (21) that

$$\frac{\mu_k^{\tilde{\tau}}}{\min(\tilde{\gamma}_k - \hat{\gamma}_k, \hat{\beta}_k)} \leq \frac{\mu_k^{\tilde{\tau}}}{\bar{\gamma}^{t_k + 2}},$$

10

and hence the sequence in question is majorized by a sequence that is decreasing monotonically to zero, giving the desired result. ∎

If Assumption 1 holds, then we can define positive constants C_4 and C_5 such that for the fast steps $(\Delta x, \Delta y)$ we have

$$|\Delta x_i \Delta y_i| \leq C_4 \mu^2, \qquad i = 1, \cdots, n, \tag{24}$$

and

$$\|\Delta x\| \leq C_5 \mu \tag{25}$$

for all k sufficiently large. Defining $\bar{\xi} = \xi + C_5 \mu_0$, where $\mu_0 = (x^0)^T y^0 / n$, we have under the prevailing assumption that

$$\|x + \alpha \Delta x\| \leq \bar{\xi}. \tag{26}$$

The next result show that the initial step length (12) eventually satisfies the acceptance criteria (13a) and (13b), making reduction of the step length unnecessary.

Lemma 4.2 *Suppose that Assumption 1 holds. Define the constant C_6 by*

$$C_6 = 2(C_4 + \bar{\xi} L C_5^2). \tag{27}$$

Then provided that $(x, y) \in \Omega_{++}(\tilde{\delta})$ and

$$\frac{C_6 \mu}{\min(\hat{\beta}, \tilde{\gamma} - \hat{\gamma})} \leq 1,$$

where $\hat{\beta}$, $\tilde{\gamma}$, and $\hat{\gamma}$ are defined in **fast**, *the conditions (13a) and (13b) are satisfied for all*

$$\alpha \in \left[0, 1 - \frac{C_6 \mu}{\min(\hat{\beta}, \tilde{\gamma} - \hat{\gamma})} \right]. \tag{28}$$

Proof. We start with the condition (13a). Using (3), (4), (5), (8), (24), (25), (26), and Assumption 1, we have for $\alpha \in [0, 1]$ that

$$
\begin{aligned}
x_i(\alpha) & y_i(\alpha) \\
= \ & (x_i + \alpha \Delta x_i)(y_i + \alpha \Delta y_i + f_i(x + \alpha \Delta x) - f_i(x) - \alpha D f_i(x) \Delta x) \\
\geq \ & x_i y_i + \alpha(-x_i y_i) + \alpha^2 \Delta x_i \Delta y_i + \alpha(x_i + \alpha \Delta x_i) \int_0^1 [D f_i(x + \theta \alpha \Delta x) - D f_i(x)] \Delta x \, d\theta \\
\geq \ & \tilde{\gamma} \mu(1 - \alpha) - \alpha^2 C_4 \mu^2 - \alpha \|x + \alpha \Delta x\| \sup_{\theta \in (0,1)} \|D f(x + \theta \alpha \Delta x) - D f(x)\| \|\Delta x\| \\
\geq \ & \tilde{\gamma} \mu(1 - \alpha) - \alpha^2 C_4 \mu^2 - \alpha^2 \bar{\xi} L \|\Delta x\|^2 \\
\geq \ & \tilde{\gamma} \mu(1 - \alpha) - C_4 \mu^2 - \bar{\xi} L C_5^2 \mu^2.
\end{aligned}
\tag{29}
$$

11

Meanwhile, from the right-hand side of (13a), we have

$$
\begin{aligned}
x(\alpha)^T y(\alpha) \\
&= (x + \alpha\Delta x)^T(y + \alpha\Delta y + f(x + \alpha\Delta x) - f(x) - \alpha Df(x)\Delta x) \\
&= x^T y(1 - \alpha) + \alpha^2 \Delta x^T \Delta y + \alpha(x + \alpha\Delta x)^T \int_0^1 [Df(x + \theta\alpha\Delta x) - Df(x)]\,\Delta x\,d\theta \\
&\leq x^T y(1 - \alpha) + \alpha^2 C_4 n\mu^2 + \alpha^2 \bar{\xi} L\|\Delta x\|^2 \\
&\leq x^T y(1 - \alpha) + C_4 n\mu^2 + \bar{\xi} L C_5^2 \mu^2.
\end{aligned}
\tag{30}
$$

Combining (29) and (30), we see that (13a) will hold provided that

$$
\tilde{\gamma}\mu(1 - \alpha) - C_4\mu^2 - \bar{\xi}LC_5^2\mu^2 \geq \hat{\gamma}\mu(1 - \alpha) + \hat{\gamma}C_4\mu^2 + (\hat{\gamma}/n)\bar{\xi}LC_5^2\mu^2.
\tag{31}
$$

By replacing $\hat{\gamma}$ by 1 in the last term and rearranging, we see that (31) holds if

$$
(\tilde{\gamma} - \hat{\gamma})\mu(1 - \alpha) - 2C_4\mu^2 - 2\bar{\xi}LC_5^2\mu^2 \geq 0 \iff 1 - \alpha \geq \frac{2(C_4\mu + \bar{\xi}LC_5^2\mu)}{(\tilde{\gamma} - \hat{\gamma})}.
$$

From (27) we see that this inequality is equivalent to

$$
1 - \alpha \geq \frac{C_6\mu}{(\tilde{\gamma} - \hat{\gamma})},
\tag{32}
$$

and, since (32) holds for all α in the range of (28), the condition (13a) also holds for all α in this range.

Turning now to (13b), we can perform a similar calculation to (30) (but this time seeking a lower bound) to obtain

$$
x(\alpha)^T y(\alpha) \geq x^T y(1 - \alpha) - C_4 n\mu^2 - \bar{\xi}LC_5^2\mu^2.
\tag{33}
$$

From (27) and (33) we deduce that

$$
x(\alpha)^T y(\alpha) \geq x^T y \left(1 - \alpha - C_4\mu - \bar{\xi}LC_5^2\mu/n\right) \geq x^T y(1 - \alpha - C_6\mu).
$$

Therefore a sufficient condition for (13b) is that

$$
x^T y(1 - \alpha - C_6\mu) \geq (1 - \alpha)(1 - \hat{\beta})x^T y \iff 1 - \alpha \geq \frac{C_6\mu}{\hat{\beta}}.
\tag{34}
$$

Condition (34) is certainly satisfied for all α in the range (28), so we obtain the desired result. ∎

We now give some threshold conditions on μ that ensure that the initial trial step length α^0 is accepted not only by the conditions (13a), (13b), but also by the outer loop, that is, it yields a reduction factor of at least ρ in the complementarity μ.

Lemma 4.3 *Suppose that Assumption 1 holds and that $(x, y) \in \Omega_{++}(\tilde{\delta})$ with*

$$\mu \leq \left(\frac{1}{C_6}\right)^{1/(1-\hat{\tau})} \tag{35}$$

and

$$\frac{\mu^{\hat{\tau}}}{\min(\tilde{\gamma} - \hat{\gamma}, \hat{\beta})} \leq \frac{\rho}{2}. \tag{36}$$

Then the initial step length α^0 for the fast step defined by (12) satisfies the acceptance criteria (13), and the fast step is accepted by the main algorithm.

Proof. Note from (35) that

$$\mu^{1-\hat{\tau}} \leq \frac{1}{C_6} \; \Rightarrow \; \mu^{\hat{\tau}} \geq C_6 \mu.$$

Using (21) and (36), we have

$$\frac{C_6 \mu}{\min(\hat{\beta}, \tilde{\gamma} - \hat{\gamma})} \leq \frac{\mu^{\hat{\tau}}}{\min(\hat{\beta}, \tilde{\gamma} - \hat{\gamma})} \leq \frac{\rho}{2} \leq 1. \tag{37}$$

We can use Lemma 4.2 and the first inequality in (37) to deduce that α^0 lies in the range (28) and hence satisfies (13).

To demonstrate acceptance of the fast step by the main algorithm, we need to show that

$$x(\alpha^0)^T y(\alpha^0) \leq \rho x^T y.$$

Using (30) again, we have

$$x(\alpha)^T y(\alpha) \leq x^T y(1 - \alpha) + C_4(x^T y)\mu + \bar{\xi} L C_5^2 (x^T y / n)\mu \leq x^T y(1 - \alpha + C_6 \mu). \tag{38}$$

From this expression, we deduce that the fast step (with $\alpha = \alpha^0$) is accepted by the main algorithm if

$$1 - \alpha^0 + C_6 \mu \leq \rho. \tag{39}$$

Note from (36) that

$$1 - \alpha^0 = \frac{\mu^{\hat{\tau}}}{\min(\tilde{\gamma} - \hat{\gamma}, \hat{\beta})} \leq \frac{\rho}{2}.$$

Using both (35) and (36), we have

$$C_6 \mu \leq C_6 \mu^{1-\hat{\tau}} \mu^{\hat{\tau}} \leq \mu^{\hat{\tau}} \leq \frac{\mu^{\hat{\tau}}}{\min(\tilde{\gamma} - \hat{\gamma}, \hat{\beta})} \leq \frac{\rho}{2}.$$

The bound (39) clearly follows from these expressions, so the fast step is accepted. ∎

We are ready for the main superlinear convergence result.

13

Theorem 4.4 *Suppose that Assumption 1 holds. Then the algorithm eventually always takes fast steps, and*

(i) *the sequence $\{\mu_k\}$ converges superlinearly to zero with Q-order at least $1 + \hat{\tau}$, and*

(ii) *the sequence $\{\nu_k\}$ converges superlinearly to zero with R-order at least $1 + \hat{\tau}$.*

Proof. Because of Lemma 4.1, the threshold conditions (35), (36) will be satisfied for all sufficiently large k, so fast steps will eventually always be taken.

For the rate-of-convergence result, note from (12), (30), and $\hat{\tau} \in (0, 1)$ that there is a constant C_{10} such that

$$\mu_{k+1} \le C_{10} \frac{\mu_k^{1+\hat{\tau}}}{\min(\tilde{\gamma} - \hat{\gamma}, \hat{\beta})} \le C_{10} \frac{\mu_k^{1+\hat{\tau}}}{\bar{\gamma}^{t_k+2}}, \tag{40}$$

for all fast iterates k for k sufficiently large. Superlinear convergence of μ_k with Q-order at least $1 + \hat{\tau}$ follows by standard arguments; see Wright [9, Theorem 6.3] and Wright and Zhang [12, Theorem 5.2]. For (ii), we have from (22) that

$$\nu_k = \frac{\|r^k\|}{\|r^0\|} \le \frac{\beta_0}{\beta_L \|r^0\|} \mu_k,$$

so $\{\nu_k\}$ is majorized by a sequence that converges with Q-order at least $1 + \hat{\tau}$, giving the result. ∎

5 Nondegenerate Problems

We have already noted that Assumption 1 holds if f is linear and $\mathcal{S}$ contains a strictly complementary solution. In this section, we consider nonlinear f. We show that if (1) has a unique, nondegenerate solution, and if there is a vector $\bar{x} > 0$ for which $f(\bar{x}) > 0$, then Assumption 1 is satisfied.

Assumption 2   (i) *The solution set $\mathcal{S}$ contains the single vector pair (x^*, y^*), where $x^* + y^* > 0$;*

(ii) *There is a point $\bar{x} \in \mathbb{R}^n$ such that $(\bar{x}, f(\bar{x})) > 0$;*

(iii) *The submatrix*

$$[Df(x^*)_{ij}]_{i \in B, j \in B}$$

(with B defined in (2)) is nonsingular, and $Df(\cdot)$ is Lipschitz continuous in a neighborhood of x^*.

We start by showing boundedness of the iteration sequence.

Lemma 5.1 *Suppose that Assumption 2 holds. Then the sequence $\{(x^k, y^k)\}$ is bounded.*

Proof. Define $\epsilon^* > 0$ by

$$\epsilon^* = \tfrac{1}{2} \min_{i=1,\cdots,n} f_i(\bar{x}) > 0,$$

and choose $M > 0$ large enough that $(r^0)^T x^* < M$. Then case (C′) of Theorem 3.4 cannot occur, so there is a finite integer K such that we have that $\nu_K \|r^0\| < \epsilon^*$. Since the sequence $\{\nu_k\}$ is decreasing, we have $\nu_k \|r^0\| < \epsilon^*$ for all $k \geq K$.

By monotonicity of f, we have

$$(x^k - \bar{x})^T (f(x^k) - f(\bar{x})) \geq 0. \tag{41}$$

Hence, using (14) and the inequalities $\bar{x} > 0$, $y^k > 0$, and $(x^k)^T y^k < (x^0)^T y^0$, we have

$$
\begin{aligned}
(x^k)^T[f(\bar{x}) - f(x^k)] &\leq \bar{x}^T f(\bar{x}) - \bar{x}^T f(x^k) \\
\Rightarrow \quad (x^k)^T[f(\bar{x}) + r^k] &\leq \bar{x}^T f(\bar{x}) + \bar{x}^T r^k - \bar{x}^T y^k + (x^k)^T y^k \\
\Rightarrow \quad (x^k)^T[f(\bar{x}) + \nu_k r^0] &\leq \bar{x}^T f(\bar{x}) + \nu_k \bar{x}^T r^0 + (x^0)^T y^0.
\end{aligned} \tag{42}
$$

For $k \geq K$ we have from (42) that

$$\epsilon^* e^T x^k \leq (x^k)^T[f(\bar{x}) - \epsilon^* e] \leq (x^k)^T[f(\bar{x}) + \nu_k r^0] \leq \bar{x}^T f(\bar{x}) + \epsilon^* e^T \bar{x} + (x^0)^T y^0.$$

Hence, since $\epsilon^* > 0$ and $x^k > 0$, it follows that $\{x^k\}$ is bounded.

Boundedness of $\{y^k\}$ follows from

$$\|y^k\| = \|f(x^k) + r^k\| \leq \|f(x^k)\| + \|r^0\|.$$

∎

Because of boundedness, case (C) of Theorem 3.3 does not occur, so the sequence either terminates finitely or converges to (x^*, y^*). An estimate of the size of the step from (3) can also be easily obtained.

Lemma 5.2 *Suppose that Assumption 2 holds. Then there is $\tilde{\delta} > 0$ such that for all $(x, y) \in \Omega_{++}(\tilde{\delta})$, the step calculated from (3) satisfies*

$$\|(\Delta x, \Delta y)\| = O(\mu). \tag{43}$$

Proof. Consider the nonlinear system of equations $F(x, y) = (0, 0)$ defined by

$$F(x, y) = \begin{bmatrix} f(x) - y \\ XYe \end{bmatrix},$$

and note that $F(x^*, y^*) = (0, 0)$ with nonsingular Jacobian

$$DF(x^*, y^*) = \begin{bmatrix} Df(x^*) & -I \\ Y^* & X^* \end{bmatrix}.$$

15

We can choose constants $M > 0$ and $\tilde{\delta} > 0$ such that

$$(x, y) \in \mathcal{B}((x^*, y^*), \tilde{\delta}) \implies \left\| (DF(x^*, y^*))^{-1} \right\| \leq M.$$

Also, we know from the definition of Ω that

$$(x, y) \in \Omega_{++} \implies \|(y - f(x), XYe)\| = O(\mu).$$

Combining these estimates with (3), we have for $(x, y) \in \Omega_{++}(\tilde{\delta})$ that

$$\|(\Delta x, \Delta y)\| \leq M \|(y - f(x), XYe - \tilde{\sigma}e)\| = O(\mu),$$

giving the result. ∎

We can now state the main result of this section.

Theorem 5.3 *Suppose that Assumption 2 holds and that finite termination of the algorithm at the solution point (x^*, y^*) does not occur. Then Assumption 1 is satisfied.*

6   Computational Results

The method described here has been implemented and tested on some small NCPs from the collection of Dirkse [2] and some larger extended linear-quadratic programming problems from Zhu and Rockafellar [13]. The following parameter settings are used in our code:

$$\chi = .9, \qquad \bar{\sigma} = .01, \qquad \bar{\alpha} = .95, \qquad \kappa = .1, \qquad \bar{\gamma} = .5,$$

$$\gamma_{\min} = 10^{-4}, \qquad \gamma_{\max} = 10^{-2}, \qquad \hat{\tau} = .9, \qquad \rho = 0.2.$$

We modify the algorithm slightly to use a different value for the reduction factor χ for the fast steps. This value, χ_{fast}, is set to 0.98. The value of sigma is chosen at each safe iteration according to the formula

$$\sigma = \max(\bar{\sigma}, \min(\mu, \sigma_{\max})),$$

where we used $\sigma_{\max} = .25$. We also avoid calculation of fast steps when they are not likely to succeed. In our implementation, the fast step is not calculated as long as $\mu_k > 0.1$; the safe step is always taken when this condition holds.

Successful termination is declared when the criteria

$$\|r^k\| \leq n \max(\text{TOL}, 10^{-9}), \qquad \mu_k \leq \text{TOL}$$

are both satisfied, where $\text{TOL} = 10^{-10}$ in our examples.

We experimented with three nonlinear problems from Dirkse [2]. We omit the details of these problems here, but refer the reader to [2] and the references therein for further details. Briefly, the problems are

nash: Nash equilibrium problem, with $n = 10$;

16

Table 1: Performance of the algorithm on small NCP test problems

Problem	nash		josephy		colvncp	
Starting Point	$x^0 = e$	$x^0 = 10e$	$x^0 = e$	$x^0 = 10e$	$x^0 = e$	$x^0 = 10e$
μ^0	.16(+3)	.21(+5)	.10(+2)	.73(+4)	.59(+2)	.28(+5)
$\|r^0\|_\infty$	.30(+4)	.18(+5)	.12(+2)	.75(+3)	.65(+3)	.31(+5)
Iterations	43	15	9	17	17	24
Solves	47	19	13	22	26	34
Trial steps	702	17	10	17	26	35
Fast steps	2	2	2	3	7	8

josephy: Four-variable problem due to Josephy;

colvncp: A convex programming problem, formulated as an NCP with $n = 15$.

Two starting points, both of which are distant from the solution, are used for each problem. They are

$$x^0 = e, \qquad y^0 = \max(1, \|f(x^0)\|_\infty)e,$$

and

$$x^0 = 10e, \qquad y^0 = \max(1, \|f(x^0)\|_\infty)e.$$

Table 1 summarizes the performance of our code on these problems. The entries for μ^0 and $\|r^0\|_\infty$ are self-explanatory; they indicate that our starting point was distant from the solution for each of the problems. The number of "iterations" equals the number of evaluations of the Jacobian $Df(\cdot)$ and also the number of matrix factorizations. The number of "solves" indicates the number of times the factors were used to compute a (safe or fast) step. The number of solves typically exceeds the number of iterations because both a safe step and a fast step are computed on some iterations. The "trial steps" entry is the total number of candidate step lengths α_k that were tried during the entire algorithm. This number is equal to the total number of evaluations of the function vector f.

The results in the table represent good performance of the method, with the exception of the Nash problem from the first starting point. The large number of unsuccessful trial values of α_k is due to the lack of sophistication of our Armijo line search, which would certainly be replaced by a safeguarded polynomial interpolation scheme in a more practical implementation of the algorithm. When χ is changed from .9 to a more conservative .5, the number of trial steps for this case decreases to 139, without affecting the number of factorizations.

Our second set of test problems is quite different from the first. They are large, mixed linear complementarity problems, in which, given an index set $\mathcal{I} \subset \{1, \cdots, n\}$ and its complement $\bar{\mathcal{I}} = \{1, \cdots, n\} \setminus \mathcal{I}$, we aim to find a vector pair (x, y) such that

$$y = Mx + q, \qquad (x_{\mathcal{I}}, y_{\mathcal{I}}) \geq 0, \qquad x_{\mathcal{I}}^T y_{\mathcal{I}} = 0, \qquad y_{\bar{\mathcal{I}}} = 0, \tag{44}$$

where the coefficient matrix M is positive semidefinite. A few obvious modifications to the algorithm are required to take account of the mixed nature of the problem, and we do not discuss these here. We obtain the problems by reformulating the stagewise extended linear-quadratic programming (ELQP) problems considered by Rockafellar [7] and Zhu and Rockafellar [13]. The motivation for these stagewise problems comes from optimal control, and we refer the reader to the cited references for more details. Stagewise ELQPs are quite complicated to state. They are defined in terms of the quadratic function

$$J(u_0, \cdots, u_N; v_1, \cdots, v_{N+1}) = \sum_{i=1}^{N} \left[p_i^T u_i + q_i^T v_i + \tfrac{1}{2} u_i^T P_i u_i - \tfrac{1}{2} v_i^T Q_i v_i - v_i^T D_i u_i \right]$$

$$+ p_0^T u_0 + q_{N+1}^T v_{N+1} + \tfrac{1}{2} u_0^T P_0 u_0 - \tfrac{1}{2} v_{N+1}^T Q_{N+1} v_{N+1} - \sum_{i=1}^{N+1} [C_i^T v_i + c_i]^T x_{i-1},$$

where the intermediate variables $(x_0, \cdots, x_N)$ satisfy

$$x_0 = B_0 u_0 + b_0, \qquad x_i = A_i x_{i-1} + B_i u_i + b_i, \qquad i = 1, \cdots, N.$$

The matrices P_i and Q_i are all positive semidefinite. The problem is to find a saddle point of $J(u_0, \cdots, u_N; v_1, \cdots, v_{N+1})$ subject to $u_i \in U_i$, $i = 0, \cdots, N$, and $v_i \in V_i$, $i = 1, \cdots, N+1$, where each U_i and V_i is a polyhedral subset of $\mathbb{R}^{n_i}$ and $\mathbb{R}^{m_i}$, respectively. Associated with the ELQP is a primal problem

$$\min_{(u_0, \cdots, u_N) \in U_0 \times \cdots \times U_N} f(u_0, \cdots, u_N),$$

where

$$f(u_0, \cdots, u_N) = \max_{(v_1, \cdots, v_{N+1}) \in V_1 \times \cdots \times V_{N+1}} J(u_0, \cdots, u_N; v_1, \cdots, v_{N+1}),$$

and a dual problem

$$\max_{(v_1, \cdots, v_{N+1}) \in V_1 \times \cdots \times V_{N+1}} g(v_1, \cdots, v_{N+1}),$$

where

$$g(v_1, \cdots, v_{N+1}) = \min_{(u_0, \cdots, u_N) \in U_0 \times \cdots \times U_N} J(u_0, \cdots, u_N; v_1, \cdots, v_{N+1}).$$

By introducing explicit representations of the polyhedral sets U_i and V_i, and introducing intermediate and slack variables, we can formulate the ELQP as a mixed monotone linear complementarity problem of the form (44). Moreover, by a "stagewise" ordering of the variables, we can ensure that the coefficient matrix M is banded, where the bandwidth is independent of N. It is the bandedness that makes the complementarity formulation practical, since the time taken to factor and solve the linear system at each iteration of our algorithm is $O(N)$, rather than the $O(N^3)$ that would be obtained by a more naive formulation. In our code, the LAPACK band solve routines DGBTRF and DGBTRS [1] are used to solve this linear system.

We tested our algorithm on modifications of the random stagewise ELQPs that are described in Section 6 of Zhu and Rockafellar [13]. The problems are obtained by discretizing

Table 2: Performance of the algorithm on smaller ELQP problems

Data Set	$n_i = 10$, $m_i = 10$, $N = 64$, dimension = 2600				
	1	3	5	7	9
μ_0	.19(+4)	.11(+6)	.15(+4)	.43(+4)	.18(+4)
Iterations	15	19	15	23	18
Solves	20	24	18	35	24
Trial steps	16	24	19	35	24
Fast steps	3	4	5	3	5
CPU time (sec)	26.1	33.3	25.4	41.2	31.5

continuous generalized optimal control problems, where the data for the continuous problems is generated randomly. The matrices P_i and Q_i are chosen to be diagonal and positive semidefinite, while the remaining matrices and vectors in the problem are dense. The polyhedra U_i and V_i are rectangles, that is, Cartesian products of intervals on the real line. In [13], the matrices Q_i and P_i are all strictly positive definite, but we modify them here by setting $Q_i \equiv 0$. As discussed in Rockafellar [7], this choice corresponds to "hard" constraints on the primal variables u_i; in fact, the primal problem above reduces to a quadratic program. The algorithms described by Rockafellar [8] and Zhu and Rockafellar [13] tend not to perform well in this important case, however. They lend themselves better to fully quadratic problems in which P_i and Q_i are all positive definite. On the other hand, our interior-point algorithm seems indifferent to this property.

The code for our algorithm does not take advantage of the linear nature of these problems, even though the Armijo line search could be replaced easily by an exact determination of the maximal α_k. Our aim is to demonstrate that the algorithm that we analyze in this paper is quite effective for both linear and nonlinear problems.

We use the code discussed in [13] to generate the data and starting points for our test problems. Besides resetting each Q_i to 0, we modify the starting points slightly to ensure strict interiority. We use the odd-numbered data sets from the problem generator, which accounts for our numbering scheme for the test problems. Tables 2 and 3 contain a summary of our results. The first line of each table contains the dimensions of the problem according to the notation above. The total dimension of the linear system to be solved at each iteration depends not only on the total number of components in $(u_0, \cdots, u_N)$ and $(v_1, \cdots, v_{N+1})$ but also on the number of intermediate and slack variables. Some components of the linear system can be eliminated conveniently, leaving a subproblem whose size is the "dimension" indicated in Tables 2 and 3. The numbers of iterations, solves, and trial step lengths are as in Table 1. The last row contains CPU times on a Sun SPARCstation IPX.

The results indicate good performance, with between 15 and 28 iterations performed for each problem. These results could have been improved by finer tuning of the user-defined parameters, but they substantiate our claim that a single set of parameters can give good performance on very different problems (large and small, linear and nonlinear). Note that the average number of iterations is slightly higher in Table 3, reflecting the oft-made observation

19

Table 3: Performance of the algorithm on larger ELQP problems

Data Set	$n_i = 20$, $m_i = 20$, $N = 128$, dimension $= 10320$				
	1	3	5	7	9
μ_0	.28(+4)	.11(+6)	.14(+5)	.18(+5)	.10(+5)
Iterations	21	26	20	28	22
Solves	29	38	25	45	32
Trial steps	34	48	31	52	36
Fast steps	6	4	6	1	4
CPU time (sec)	435.	548.	406.	592.	452.

that this number grows slowly with problem size in most interior-point methods. Note, too, that a number of fast steps were taken at the tail of the iteration sequence, and rapid local convergence was observed.

7 Conclusions

The local convergence results represent a natural extension of those presented in [11] for the case of linear f. In this case, we require only existence of a strictly complementary solution for $\|(\Delta x, \Delta y)\| = O(\mu)$ to hold. Monteiro and Wright [5] show that the strict complementarity assumption is necessary for superlinear convergence of methods that behave like Newton's method near the solution. There is a considerable gap between the weak assumption of the linear case and the stronger assumptions of Section 5 for nonlinear f, and current research is aimed at bridging this gap as far as possible.

Our results can be extended easily to the case of Hölder continuous $Df(\cdot)$, that is, $\|Df(x^1) - Df(x^2)\| \leq L\|x^1 - x^2\|^\tau$ for some $\tau \in (0,1]$ and x^1, x^2 in a neighborhood of x^*. When $\tau > (\sqrt{5} - 1)/2$, we can show local convergence with Q-order $1 + \hat{\tau}$, provided that $\hat{\tau}$ lies in the range $(0, \tau^2 + \tau - 1)$. The algorithm can be modified to allow reduction of $\hat{\tau}$ after failed fast iterations, so that that the inclusion $\hat{\tau} \in (0, \tau^2 + \tau - 1)$, and therefore superlinear convergence, eventually occurs. Knowledge of τ is not necessary. In the case of $Df(\cdot)$ merely continuous near x^*, we still have global convergence, from the analysis of Section 3.

Acknowledgments

Appendix

Proof of Lemma 2.1. We deal only with the case of $r^0 \neq 0$, since the proofs for the case of $r^0 = 0$ are simple specializations.

We prove (17) and (18) by induction on l. Taking $l = 1$, we know that safe steps are taken for $k = 0, 1, \cdots, k_1 - 1$. Therefore, from (9b), we have

$$\mu_{k_1}/\mu_0 \geq \prod_{j=1}^{k_1-1} (1 - \alpha_j) = \nu_{k_1},$$

while, from (14),

$$r^{k_1} = \nu_{k_1} r^0.$$

Therefore, by (11), we have $\tilde{\beta}_{k_1} \geq 1$ and $\tilde{i}_{k_1} = 0 = t_{k_1}$, yielding (17). For (18), we note that by choice of (x^0, y^0) and (8), we have

$$\tilde{\gamma}_0 = \gamma_{\max}.$$

Since safe steps are taken for $k = 0, 1, \cdots, k_1 - 1$, we have by repeated application of (9a) and (8) that

$$\gamma_{\max} = \tilde{\gamma}_0 = \tilde{\gamma}_1 = \cdots = \tilde{\gamma}_{k_1}.$$

Hence,

$$\tilde{\gamma}_{k_1} - \gamma_{\min} = \gamma_{\max} - \gamma_{\min},$$

so (18) is satisfied for $l = 1$.

We now assume that (17) and (18) are satisfied for $k_1, k_2, \cdots, k_l$ and prove that they continue to be true at k_{l+1}. From (11) and (14), we have that

$$\mu_{k_l}/\mu_0 \geq \nu_{k_l} \tilde{\beta}_{k_l} \geq \nu_{k_l} \prod_{j=1}^{\tilde{i}_{k_l}} (1 - \bar{\gamma}^j).$$

Applying (13b) to the fast step that is taken at iteration k_l, we have that

$$
\begin{aligned}
\mu_{k_l+1}/\mu_0 &\geq (1 - \alpha_{k_l})(1 - \hat{\beta}_{k_l})(\mu_{k_l}/\mu_0) \\
&\geq (1 - \alpha_{k_l})(1 - \hat{\beta}_{k_l})\nu_{k_l}\tilde{\beta}_{k_l} \\
&= \nu_{k_l+1}(1 - \hat{\beta}_{k_l})\tilde{\beta}_{k_l} \\
&= \nu_{k_l+1}(1 - \bar{\gamma}^{\tilde{i}_{k_l}+1})\tilde{\beta}_{k_l} \\
&\geq \nu_{k_l+1} \prod_{j=1}^{\tilde{i}_{k_l}+1} (1 - \bar{\gamma}^j).
\end{aligned}
$$

Since safe steps are taken at iterations $k_l + 1, \cdots, k_{l+1} - 1$, we have from (9b) that the ratio (μ_k/ν_k) is nondecreasing for $k = k_l + 1, \cdots, k_{l+1}$. In particular,

$$\mu_{k_{l+1}}/\mu_0 \geq \nu_{k_{l+1}} \prod_{j=1}^{\tilde{i}_{k_l}+1} (1 - \bar{\gamma}^j),$$

21

and therefore
$$\tilde{\beta}_{k_{l+1}} \geq \prod_{j=1}^{\tilde{t}_{k_l}+1} (1 - \bar{\gamma}^j). \tag{45}$$

From (45), the definition of $\tilde{t}_{k_{l+1}}$, and the inductive hypothesis for (17), we have
$$\tilde{t}_{k_{l+1}} \leq \tilde{t}_{k_l} + 1 \leq t_{k_l} + 1 = l = t_{k_{l+1}},$$

so (17) continues to hold at k_{l+1}.

Combining (13a) at $k = k_l$ with (8) at $k = k_l + 1$, we have that $\tilde{\gamma}_{k_l+1} \geq \hat{\gamma}_{k_l}$. In fact, as observed earlier, $\tilde{\gamma}_k$ is nondecreasing over the safe iterations $k = k_l + 1, \cdots, k_{l+1} - 1$, and so
$$\tilde{\gamma}_{k_{l+1}} \geq \cdots \geq \tilde{\gamma}_{k_l+1} \geq \hat{\gamma}_{k_l}.$$

Hence, by using the definition of $\hat{\gamma}_k$ and the inductive hypothesis for (18), we have that
$$\tilde{\gamma}_{k_{l+1}} - \gamma_{\min} \geq \hat{\gamma}_{k_l} - \gamma_{\min} = \bar{\gamma}(\tilde{\gamma}_{k_l} - \gamma_{\min}) \geq \bar{\gamma}^l(\gamma_{\max} - \gamma_{\min}).$$

Therefore (18) continues to hold at iteration k_{l+1}. We conclude that (17) and (18) hold for all $l = 1, 2, \cdots$.

To prove (19), we note by (18) and the definition of $\hat{\gamma}_{k_l}$ that
$$\tilde{\gamma}_{k_l} - \hat{\gamma}_{k_l} = \tilde{\gamma}_{k_l} - [\gamma_{\min} + \bar{\gamma}(\tilde{\gamma}_{k_l} - \gamma_{\min})] = (1 - \bar{\gamma})(\tilde{\gamma}_{k_l} - \gamma_{\min}) \geq (1 - \bar{\gamma})\bar{\gamma}^{l-1}(\gamma_{\max} - \gamma_{\min}),$$

as required.

For (20), we need only note that
$$\hat{\beta}_{k_l} = \bar{\gamma}^{\tilde{t}_{k_l}+1} \geq \bar{\gamma}^{(l-1)+1} = \bar{\gamma}^l.$$

The final inequality (21) follows immediately from (19) and (20), if we note that $1 - \bar{\gamma} \geq \bar{\gamma}$.
∎

Proof of Lemma 3.1. Define
$$\hat{\delta} = \tfrac{1}{2} \min_{i=1,\cdots,n} (\min(\bar{x}_i, \bar{y}_i)) > 0,$$

and choose positive constants C_3 and C_2 such that
$$x^T y \geq C_2 \text{ and } \|(x, y)\| \leq C_3 \text{ for all } (x, y) \in \mathcal{B}((\bar{x}, \bar{y}), \hat{\delta}).$$

Note that the coefficient matrix
$$\begin{bmatrix} Df(x) & -I \\ Y & X \end{bmatrix}$$

is nonsingular at $(\bar{x}, \bar{y})$ and continuous in an open set containing $\mathcal{B}((\bar{x}, \bar{y}), \hat{\delta}) \subset (\mathbb{R}_+^n \times \mathbb{R}_+^n)$. The right-hand side of (3) is also continuous with respect to both (x, y) and $\tilde{\sigma}$. Therefore we can define a constant $C_1 > 0$ such that
$$\|(\Delta x, \Delta y)\| \leq C_1$$

22

for all $(x, y) \in \mathcal{B}((\bar{x}, \bar{y}), \hat{\delta})$ and $\tilde{\sigma} \in [\bar{\sigma}, 1/2]$.

Taking $\alpha \in [0, 1]$ and $(x, y) \in \mathcal{B}((\bar{x}, \bar{y}), \hat{\delta})$, we have

$$
\begin{aligned}
x(\alpha)^T & y(\alpha) \\
&= (x + \alpha \Delta x)^T (y + \alpha \Delta y + f(x + \alpha \Delta x) - f(x) - \alpha Df(x) \Delta x) \\
&= x^T y - \alpha(1 - \tilde{\sigma}) x^T y + \alpha^2 \Delta x^T \Delta y + \alpha \int_0^1 (x + \alpha \Delta x)^T [Df(x + \theta \alpha \Delta x) - Df(x)] \Delta x \, d\theta \\
&\leq x^T y [1 - \alpha(1 - \tilde{\sigma})] + \alpha^2 C_1^2 + \alpha C_1 (C_3 + C_1) \sup_{\theta \in (0,1)} \|Df(x + \theta \alpha \Delta x) - Df(x)\|. \quad (46)
\end{aligned}
$$

Hence the left inequality in (9b) is certainly satisfied if

$$
\alpha^2 C_1^2 + \alpha C_1(C_3 + C_1) \sup_{\theta \in (0,1)} \|Df(x + \theta \alpha \Delta x) - Df(x)\| \leq (1 - \kappa)\alpha(1 - \tilde{\sigma}) x^T y. \quad (47)
$$

By continuity of Df, we can make the following assertion: for each $(x, y) \in \mathcal{B}((\bar{x}, \bar{y}), \hat{\delta})$, there is $\eta_1(x) > 0$ such that

$$
\|Df(x + d) - Df(x)\| \leq \frac{(1 - \kappa)C_2}{4C_1(C_3 + C_1)}
$$

for all d with $\|d\| \leq \eta_1(x)$. Since $\mathcal{B}((\bar{x}, \bar{y}), \hat{\delta})$ is compact, we have that

$$
\bar{\eta}_1 \triangleq \inf_{(x,y) \in \mathcal{B}((\bar{x}, \bar{y}), \hat{\delta}) \cap \Omega} \eta_1(x) > 0.
$$

Therefore, setting

$$
\hat{\alpha}^{(1)} = \min \left(\frac{(1 - \kappa)C_2}{4C_1^2}, \frac{\bar{\eta}_1}{C_1} \right),
$$

we have for $(x, y) \in \mathcal{B}((\bar{x}, \bar{y}), \hat{\delta})$, $\tilde{\sigma} \in [\bar{\sigma}, 1/2]$, and $\alpha \in [0, \hat{\alpha}^{(1)}]$ that

$$
\begin{aligned}
\alpha^2 C_1^2 &+ \alpha C_1(C_3 + C_1) \sup_{\theta \in (0,1)} \|Df(x + \theta \alpha \Delta x) - Df(x)\| \\
&\leq \alpha \hat{\alpha}^{(1)} C_1^2 + \alpha C_1(C_3 + C_1) \sup_{\theta \in (0,1)} \|Df(x + \theta \alpha \Delta x) - Df(x)\| \\
&\leq \alpha(1 - \kappa)(C_2/4) + \alpha C_1(C_3 + C_1)(1 - \kappa)(C_2/4)/(C_1(C_3 + C_1)) \\
&\leq (1 - \kappa)\alpha(C_2/2) \\
&\leq (1 - \kappa)\alpha(1 - \tilde{\sigma}) x^T y. \quad (48)
\end{aligned}
$$

Hence (47) is satisfied, yielding the left inequality in (9b).

We now show that (9a) holds. For any $i = 1, \cdots, n$, $(x, y) \in \mathcal{B}((\bar{x}, \bar{y}), \hat{\delta})$, and $\alpha \in [0, \hat{\alpha}^{(1)}]$, we have

$$
x_i(\alpha) y_i(\alpha)
$$

23

$$
\begin{aligned}
&= (x_i + \alpha\Delta x_i)(y_i + \alpha\Delta y_i + f_i(x + \alpha\Delta x) - f_i(x) - \alpha Df_i(x)\Delta x) \\
&\geq x_i y_i + \alpha[-x_i y_i + \tilde{\sigma}(x^T y)/n] - \alpha^2 C_1^2 - \alpha C_1(C_3 + C_1) \sup_{\theta\in(0,1)} \|Df_i(x + \theta\alpha\Delta x) - Df_i(x)\| \\
&\geq x_i y_i(1 - \alpha) + \tilde{\sigma}\alpha(x^T y)/n - \alpha^2 C_1^2 - \alpha C_1(C_3 + C_1) \sup_{\theta\in(0,1)} \|Df(x + \theta\alpha\Delta x) - Df(x)\| \\
&\geq \tilde{\gamma}(x^T y/n)(1 - \alpha) + \tilde{\sigma}\alpha(x^T y/n) - \alpha^2 C_1^2 \\
&\qquad -\alpha C_1(C_3 + C_1) \sup_{\theta\in(0,1)} \|Df(x + \theta\alpha\Delta x) - Df(x)\|.
\end{aligned}
\tag{49}
$$

Meanwhile, we have as in (46) that

$$
x(\alpha)^T y(\alpha) \leq x^T y[1 - \alpha(1 - \tilde{\sigma})] + \alpha^2 C_1^2 + \alpha C_1(C_3 + C_1) \sup_{\theta\in(0,1)} \|Df(x + \theta\alpha\Delta x) - Df(x)\|.
$$

By combining the last two expressions, we find that (9a) will be satisfied if

$$
(x^T y/n)\alpha\tilde{\sigma}(1 - \tilde{\gamma}) - 2\alpha^2 C_1^2 - 2\alpha C_1(C_3 + C_1) \sup_{\theta\in(0,1)} \|Df(x + \theta\alpha\Delta x) - Df(x)\| \geq 0,
$$

which in turn is true if

$$
(C_2/n)\bar{\sigma}(1 - \gamma_{\mathbf{max}}) - 2\alpha C_1^2 - 2C_1(C_3 + C_1) \sup_{\theta\in(0,1)} \|Df(x + \theta\alpha\Delta x) - Df(x)\| \geq 0.
\tag{50}
$$

Just as we chose $\bar{\eta}_1$ above, we can find $\bar{\eta}_2$ such that

$$
\|Df(x + d) - Df(x)\| \leq \frac{(C_2/n)\bar{\sigma}(1 - \gamma_{\mathbf{max}})}{4C_1(C_3 + C_1)}
$$

for all $(x, y) \in \mathcal{B}((\bar{x}, \bar{y}), \hat{\delta})$ and $\|d\| \leq \bar{\eta}_2$. Hence, if we choose

$$
\hat{\alpha}^{(2)} = \min\left(\hat{\alpha}^{(1)}, \frac{\bar{\eta}_2}{C_1}, \frac{(C_2/n)\bar{\sigma}(1 - 2\bar{\gamma})}{4C_1^2}\right),
$$

a calculation similar to (48) shows that (50) (and hence (9a)) is satisfied for all $\alpha \in [0, \hat{\alpha}^{(2)}]$ and $(x, y) \in \mathcal{B}((\bar{x}, \bar{y}), \hat{\delta})$.

For the remaining inequality, we have as in (46) that

$$
x(\alpha)^T y(\alpha) \geq x^T y[1 - \alpha(1 - \tilde{\sigma})] - \alpha^2 C_1^2 - \alpha C_1(C_3 + C_1) \sup_{\theta\in(0,1)} \|Df(x + \theta\alpha\Delta x) - Df(x)\|.
$$

Therefore, the right inequality in (9b) holds if

$$
\alpha\tilde{\sigma}x^T y - \alpha^2 C_1^2 - \alpha C_1(C_3 + C_1) \sup_{\theta\in(0,1)} \|Df(x + \theta\alpha\Delta x) - Df(x)\| \geq 0,
$$

which in turn is satisfied if

$$
\bar{\sigma}C_2 - \alpha C_1^2 - C_1(C_3 + C_1) \sup_{\theta\in(0,1)} \|Df(x + \theta\alpha\Delta x) - Df(x)\| \geq 0.
\tag{51}
$$

24

As before, we can choose $\bar{\eta}_3 > 0$ such that

$$\|Df(x + d) - Df(x)\| \leq \frac{\bar{\sigma}C_2}{2C_1(C_3 + C_1)}$$

for all $(x, y) \in \mathcal{B}((\bar{x}, \bar{y}), \hat{\delta})$ and $\|d\| \leq \bar{\eta}_3$. Another calculation like (48) shows that the inequality (51) is satisfied if $\alpha \in [0, \hat{\alpha}^{(3)}]$, where

$$\hat{\alpha}^{(3)} = \min\left(\hat{\alpha}^{(2)}, \frac{\bar{\sigma}C_2}{2C_1^2}, \frac{\bar{\eta}_3}{C_1}\right).$$

The conclusion of the lemma is obtained by setting $\hat{\alpha} = \hat{\alpha}^{(3)}$. ∎

References

[1] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen, *LAPACK User's Guide*, SIAM, Philadelphia, 1992.

[2] S. P. Dirkse, *Solving complementarity problems via GAMS: A model library.* Manuscript, May 1993.

[3] O. Güler, *Existence of interior points and interior paths in nonlinear monotone complementarity problems*, Mathematics of Operations Research, 18 (1993), pp. 128–147.

[4] M. Kojima, T. Noma, and A. Yoshise, *Global convergence and detecting infeasibility in interior-point algorithms*, Research Reports on Information Sciences, Series B: Operations Research B–257, Department of Information Sciences, Tokyo Institute of Technology, Tokyo, Japan, September 1992.

[5] R. D. C. Monteiro and S. J. Wright, *Local convergence of interior-point algorithms for degenerate monotone LCP*, Computational Optimization and Applications, 3 (1994), pp. 131–155.

[6] F. A. Potra and Y. Ye, *Interior point methods for nonlinear complementarity problems*, Technical Report 15, Department of Mathematics, University of Iowa, Iowa City, July 1991.

[7] R. T. Rockafellar, *Linear-quadratic programming and optimal control*, SIAM Journal on Control and Optimization, 25 (1987), pp. 781–814.

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

[9] S. J. WRIGHT, *An infeasible-interior-point algorithm for linear complementarity problems*, Preprint MCS–P331–1092, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., October 1992. To appear in Mathematical Programming, Series A.

[10] ——, *A path-following infeasible-interior-point algorithm for linear complementarity problems*, Optimization Methods and Software, 2 (1993), pp. 79–106.

[11] ——, *A path-following interior-point algorithm for linear and quadratic optimization problems*, Preprint MCS–P401–1293, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., December 1993.

[12] S. J. WRIGHT AND Y. ZHANG, *A superquadratic infeasible-interior-point method for linear complementarity problems*, Preprint MCS–P418–0294, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Ill., February 1994.

[13] C. ZHU AND R. T. ROCKAFELLAR, *Primal-dual projected gradient algorithms for extended linear-quadratic programming*, SIAM Journal on Optimization, 3 (1993), pp. 751–783.