ANALYSIS OF AN INFEASIBLE INTERIOR PATH-FOLLOWING METHOD FOR COMPLEMENTARITY PROBLEMS [1]

August 12, 1997 (revised September 16, 1997)

PAUL TSENG
Department of Mathematics, University of Washington
Seattle, Washington 98195, U.S.A. (tseng@math.washington.edu)

ABSTRACT

We study an infeasible interior path-following method for complementarity problems. The method uses a wide neighborhood and takes one or two Newton steps per iteration. We show that the method attains global convergence, assuming the iterates are defined and bounded (which occurs when the function is a P_0-R_0-function or when the function is monotone and a sufficiently positive solution exists). Moreover, the convergence rate is globally (respectively, locally) linear if the function is a uniformly P-function (respectively, at each cluster point, a certain submatrix of the Jacobian of the function is a P-matrix). By introducing an inexpensive active-set strategy in computing the second Newton direction, we show that the method attains local superlinear convergence without requiring the existence of a strictly complementary solution. The proof of this uses a local error bound on the distance from an iterate to a solution in terms of the square root of the centering parameter.

1 Introduction

We consider the well-known complementarity problem (CP for short) of finding an $(x, y) \in \Re^n \times \Re^n$ satisfying

$$x \geq 0, \quad y \geq 0, \quad x^T y = 0, \quad F(x) - y = 0, \tag{1}$$

where $F = (F_1, ..., F_n)^T$ is a given continuous function from $\Re^n_+$ to $\Re^n$ that is continuously differentiable on $\Re^n_{++}$ [4, 6, 23]. In the case where F is affine on $\Re^n$, CP is called the linear complementarity problem or LCP [5, 13]. We denote by S the set of of solutions of CP, i.e., $S := \{(x, y) \in \Re^n \times \Re^n : (x, y) \text{ satisfy } (1)\}$, which we assume is nonempty.

There has been proposed various approaches for solving CP, of which one based on interior continuation/path-following has received much attention. In this approach, we approximate (1) by a set of smooth equations of the form $H_\mu(x, y) = 0$ with

[1] This research is supported by National Science Foundation Grant CCR-9311621.

$(x, y) \in \Re^{2n}_{++}$, where $\mu \in \Re_{++}$ is the "centering" parameter and

$$H_\mu(x, y) := \begin{bmatrix} Xy - \mu e \\ F(x) - y \end{bmatrix}, \tag{2}$$

with X denoting the diagonal matrix with diagonal entries $x_1, ..., x_n$ and e denoting the vector of 1s. [Thus, H_μ converges to H_0 pointwise as $\mu \to 0$. We note that e can be replaced by any vector with positive components, though for simplicity we will not do so here.] Then, starting with any $\mu \in \Re_{++}$ and $(x, y) \in \Re^{2n}_{++}$, we fix μ and apply one or two Newton-type steps for $H_\mu(x, y) = 0$ to update (x, y), and then we decrease μ and re-iterate. This approach, beginning with the work of Kojima et al. [10], has been widely used [9, 11, 12, 14, 22, 24, 26, 27, 29, 30, 32] and is the focus of our study. Methods based on this interior path-following approach are further distinguished by whether they require the constraints $F(x) - y = 0$ to be satisfied at all iterations ("feasible" methods) or not ("infeasible" methods). In the case where F is monotone and satisfies a scaled Lipschitz condition or a relative Lipschitz condition, a number of feasible path-following methods (using narrow or wide neighborhood) have been proposed for which global linear convergence, with convergence ratio depending on n and the Lipschitz constant only, can be shown [9, 22, 24, 27, 28, 29]. If F is monotone or, more generally, a P_0-function and no other assumption is made on F, infeasible path-following methods have been proposed for which convergence can be shown assuming only boundedness of the iterates [12, 26, 30, 32] (also see [11] for boundedness of homotopy path associated with H_μ and see [14] for a general framework for infeasible interior methods and infeasibility detection in the case F is monotone). Moreover, the methods in [12, 24, 27, 30, 32] (also see [25, 28] for extensions to monotone variational inequalities) are superlinearly convergent near a strictly complementary nondegenerate solution, and the method in [30] is globally linearly convergent when F is a uniformly nondegenerate function on $\Re^n_{++}$. A related approach, based on using a trust-region method to minimize $\|H_0(x, y)\|^2$ subject to $(x, y) \in \Re^{2n}_{+}$, is considered in [20] (also see [18] for a different interior approach based on NE/SQP iterations with the mapping $\min\{x, F(x)\}$). There recently is also much interest in non-interior continuation approach for solving CP. In this approach, the expression $Xy - \mu e$ is replaced by a different expression of x, y, μ that is well behaved even when x or y have nonpositive components. [However, this approach cannot be applied when F is defined on $\Re^n_+$ only, as in the case of some economic equilibrium problems.] While this approach has some fundamental differences with the interior approach, there is also enough similarities in the two approaches to allow the convegence analysis of one to be adapted to the other, on which we will elaborate below. Lastly, we note that there is an extensive literature on interior path-following methods for the special cases of LCP and linear programs, which we will not survey here (see, e.g, [13]).

Recently, there has been advances in the global convergence and local superlinear convergence analysis of non-interior methods for which there was no counterpart in interior methods. In particular, there has been proposed non-interior continuation methods that are globally convergent when F is a P_0-function, is linearly convergent when F is a uniformly P-function, and is locally superlinearly convergent without requiring the existence of a strictly complementary solution (see [1, 2, 3, 31] and references therein). In this paper, we show that an interior path-following method can be developed with global and local convergence properties as strong as those obtained for non-interior continuation methods. In particular, we consider an infeasible interior path-following method that maintains (x, y, μ) to lie in a neighborhood of the central path, defined by

$$\mathcal{N}_\beta := \{ \ (x, y, \mu) \in \Re^{2n}_{++} \times \Re_{++} : \beta_1 \mu e \leq Xy \leq \beta_2 \mu e, \|F(x) - y\| \leq \beta_3 \sqrt{\mu} \ \}, \tag{3}$$

where $\beta = (\beta_1, \beta_2, \beta_3) \in (0, 1) \times (1, \infty) \times \Re_{++}$. [If $\beta_1 = \beta_2 = 1$ and $\beta_3 = 0$, then $\mathcal{N}_\beta$ coincides with the central path for CP [11, 12, 16].] This method iteratively moves $(z, \mu) = (x, y, \mu)$ along the Newton direction $-(H'_\mu(z)^{-1} H_\mu(z), \sigma\mu)$ ($\sigma \in (0, 1)$) while maintaining it to remain in $\mathcal{N}_\beta$ and, to accelerate convergence, a further move is made towards an "improving guess" $(\hat{z}, \hat{\mu})$ (see Algorithm 1). We show that the method converges globally and linearly under conditions similar to those required by non-interior continuation methods [1, 2, 3, 31] (see Proposition 1 and Corollaries 1 and 2 and Lemma 3). Lastly, we propose a choice of $\hat{z}$ that extends $z - H'_\mu(z)^{-1} H_0(z)$ used in [12, 24, 27, 30, 32] via an active-set strategy, and we show that the associated method attains local superlinear convergence without requiring the existence of a strictly complementary solution, as is needed for other interior path-following methods (see Proposition 2). Moreover, this choice of $\hat{z}$ can be computed relatively inexpensively. Our analysis uses in part a new local error bound on the distance from z to a solution z^* in terms of $\sqrt{\mu}$ (see Lemma 4). Preliminary computational experience with the method is reported in Section 5.

Much of our results are motivated by a recent study by the author of a non-interior continuation method based on the Chen-Mangasarian smoothing function [31]. However, while some of our proofs are patterned after those in [31] (notably Corollary 1 and Lemmas 4, 6), a number of new techniques are needed to deal with a different smoothing function $Xy - \mu e$, the unboundedness of $\|H'_\mu(x, y)^{-1}\|$ (see (7) and Corollary 2) and the need to maintain the iterates approximately centered (see (16) and Lemma 5). Also, our choice of neighborhood (3) differs from existing choices for infeasible path-following methods in that $\|F(x) - y\|$ is required to be in the order of $\sqrt{\mu}$, rather than μ [30, 32], which appears to be needed for the superlinear convergence analysis near a non-strictly complementary solution. [Intuitively, if (x, y) converges to an $(x^*, y^*) \in S$ with $x^*_i = y^*_i = 0$ for some i, then $x_i y_i \geq \beta_1 \mu$ would imply $\|x - x^*\| + \|y - y^*\|$ tends to zero no faster than $\sqrt{\mu}$, so the same holds for $\|F(x) - y\| = \|F(x) - F(x^*) - (y - y^*)\|$.] It further differs from the wide neighborhood used in [32] in that Xy is bounded above by a constant multiple of μ (which appears to be needed for the global linear convergence analysis) and from the narrow neighborhood used in [30] in that $\|Xy - \mu e\|/\mu$ is allowed to exceed 1.

In our notation, all vectors are column vectors, $\Re^n$ denotes the space of n-dimensional real column vectors, $\Re^{n \times n}$ denotes the space of $n \times n$ matrices with real entries, ^{T} denotes transpose, $\Re_+$ and $\Re_{++}$ denote the nonnegative and the positive reals, respectively, and $N := \{1, ..., n\}$. For convenience, we sometimes write $\begin{bmatrix} x \\ y \end{bmatrix}$

as "(x, y)" for any vectors x and y. For any vector $x \in \Re^n$, we denote by x_i the ith component of x and, for any $I \subset N$, by x_I the vector obtained after removing from x those x_i with $i \notin I$. We also denote by $\|x\|$ and $\|x\|_\infty$ the 2- and ∞-norm of x, so $\|x\| = \sqrt{x^T x}$ and $\|x\|_\infty = \max_{i \in N} |x_i|$. For any $M \in \Re^{n \times n}$ and any $I \subset N$ and $J \subset N$, we denote by M_{IJ} the submatrix of M obtained by removing all rows of M with indices outside of I and removing all columns of M with indices outside of J. Also, we denote by $|I|$ the cardinality of I and denote $I^c := N \backslash I$ and $\|M\| := \max_{x \in \Re^n : \|x\|=1} \|Mx\|$. For any continuously differentiable function $H = (H_1, ..., H_m)^T : \Re^m \mapsto \Re^m$, we denote its Jacobian by $H' = (H_1', ..., H_m')^T$, where H_i' denotes the gradient of H_i for $i = 1, ..., m$.

2 Algorithm Description

In this section we formally describe our method, parameterized by $\beta = (\beta_1, \beta_2, \beta_3) \in (0,1) \times (1, \infty) \times \Re_{++}$. This method is motivated by a non-interior continuation method in [31] and an interior path-following method in [30], differing from the latter mainly in the choice of the neighborhood $\mathcal{N}_\beta$ and a simpler stepsize rule for θ^t and α^t (see [26, 32] for related infeasible methods).

Algorithm 1 *Choose any* $(z^0, \mu^0) \in \mathcal{N}_\beta$, *any* ψ, σ *in* $(0,1)$ *satisfying* $(1-\sigma)\beta_2 > 1$. *For* $t = 0, 1, ...,$ *we generate* (z^{t+1}, μ^{t+1}) *from* (z^t, μ^t) *as follows:*
Iteration t. *Let* $w^t \in \Re^{2n}$ *satisfy*

$$H'_{\mu^t}(z^t) w^t = -H_{\mu^t}(z^t). \tag{4}$$

Choose θ^t *to be the largest* $\theta \in \{1, \psi, \psi^2, ...\}$ *such that* $(z^t + \theta w^t, (1-\sigma\theta)\mu^t) \in \mathcal{N}_\beta$. *Choose any* $\hat{z}^t \in \Re^{2n}_{++}$, *any* $\hat{\mu}^t \in (0, (1-\sigma\theta^t)\mu^t]$ *and any* $\nu^t \in \{0, 1, ...\}$. *Choose* α^t *to be the largest* $\alpha \in \{1, \psi, \psi^2, ..., \psi^{\nu^t}, 0\}$ *such that*

$$\Big((z^t + \theta^t w^t)(1-\alpha) + \hat{z}^t \alpha, (1-\sigma\theta^t)\mu^t(1-\alpha) + \hat{\mu}^t \alpha\Big) \in \mathcal{N}_\beta, \tag{5}$$

and let

$$z^{t+1} := (z^t + \theta^t w^t)(1-\alpha^t) + \hat{z}^t \alpha^t, \quad \mu^{t+1} := (1-\sigma\theta^t)\mu^t(1-\alpha^t) + \hat{\mu}^t \alpha^t. \tag{6}$$

Roughly speaking, at iteration t of Algorithm 1, we first compute a Newton direction w^t by solving the linear equations (4), and next we move (z^t, μ^t) in the direction $(w^t, -\sigma\mu^t)$ by as "large" a stepsize θ^t as possible while remaining in the neighborhood $\mathcal{N}_\beta$ (this is done using an Armijo-Goldstein-type line search rule), and lastly we move the resulting pair as near to $(\hat{z}^t, \hat{\mu}^t)$ as possible while remaining in $\mathcal{N}_\beta$. [If $\alpha^t = 1$, then $(z^{t+1}, \mu^{t+1}) = (\hat{z}^t, \hat{\mu}^t)$. The integer ν^t controls the accuracy and the work in computing α^t.] This last move is designed to accelerate the convergence of the method and, while $\hat{z}^t$ can be chosen arbitrarily without affecting the global convergence properties of the method, we would like $\hat{z}^t$ to be near the solution set S for faster local convergence. In Section 4 we will consider a choice of $(\hat{z}^t, \hat{\mu}^t)$ that is relatively cheap to compute and yields local superlinear convergence under fairly mild assumptions. Also, in Proposition 1 of the next section, we will show that if $H'_{\mu^t}(z^t)$ is nonsingular (so w^t is uniquely defined), then θ^t is defined and positive (due to our choice of σ) and, since α^t is well defined ($\alpha = 0$ always satisfies (5)), hence (z^{t+1}, μ^{t+1}) is well defined.

3 Global (Linear) Convergence Analysis

We begin with the following proposition showing that Algorithm 1 attains global convergence provided that $H'_\mu(z)$ is nonsingular and $\|z\|$ is bounded for all $(z, \mu) \in \mathcal{N}_\beta$ with μ bounded, and if in addition $H'_\mu(z)^{-1}$ is in the order of $1/\sqrt{\mu}$ in a weak sense (see (7)) and F is sufficiently smooth, then linear convergence is also achieved. The global convergence result is analogous to one obtained by Kojima et al. [12] and by Simantiraki and Shanno [26] for their methods. The linear convergence result and its assumption (7) appear to be new for interior methods. Conditions on F, under which the assumptions of the proposition are satisfied, will be discussed in subsequent lemmas and corollaries.

Proposition 1 *Fix any $\beta = (\beta_1, \beta_2, \beta_3) \in (0,1) \times (1,\infty) \times \Re_{++}$. Assume $H'_\mu(z)$ is nonsingular for all $(z, \mu) \in \mathcal{N}_\beta$ with $\mu \le \mu^0 \in \Re_{++}$. Then $\{(z^t, \mu^t, \theta^t)\}_{t=0,1\ldots}$ generated by Algorithm 1 is well defined and satisfies $(z^t, \mu^t) \in \mathcal{N}_\beta$ and $\mu^{t+1} \le (1 - \sigma\theta^t)\mu^t$ for all t, where σ and ψ are chosen in the method. Moreover, the following hold:*
(a) If $\{z^t\}$ has a convergent subsequence, then $\{\mu^t\} \to 0$ and the limit is in S.
(b) If there exists $\kappa \in \Re_{++}$ such that

$$H'_\mu(x,y)\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} r \\ s \end{bmatrix} \implies \max\left\{\|u\|, \|X^{-1}u\|\sqrt{\mu}\right\} \le \kappa\left(\frac{\|r\|}{\sqrt{\mu}} + \|s\|\right) \tag{7}$$

for all $(x, y, \mu) \in \mathcal{N}_\beta$ and all $(r, s) \in \Re^{2n}$ and there exists $\lambda \in \Re_{++}$ such that $\|F(x + u) - F(x) - F'(x)u\| \le \lambda\|u\|^2$ for all $x \in \Re^n_{++}, u \in \Re^n$, then $\mu^{t+1} \le c\mu^t$ for all t, where $c \in (0,1)$ depends on $\beta_1, \beta_2, \psi, \sigma, \hat{\kappa}, \sqrt{\mu^0}\lambda/\beta_3$ only, with $\hat{\kappa} := \kappa(\max\{1-\beta_1, \beta_2 - 1\}\sqrt{n} + \beta_3)$.

Proof. At the start of each iteration $t = 0, 1, ...,$ we have $0 < \mu^t \le \mu^0$ (since μ^t is monotonically decreasing with t) and $(z^t, \mu^t) \in \mathcal{N}_\beta$, so $H'_{\mu^t}(z^t)$ is nonsingular by assumption, implying w^t is well defined. We show below that θ^t is well defined and positive. Then, since α^t is well defined, so is (z^{t+1}, μ^{t+1}) given by (6). Moreover, our choice (5) of α^t ensures that $(z^{t+1}, \mu^{t+1}) \in \mathcal{N}_\beta$ and the fact $\hat{\mu}^t \le (1 - \sigma\theta^t)\mu^t$ ensures that $\mu^{t+1} \le (1 - \sigma\theta^t)\mu^t$.

For any $(z, \mu) = (x, y, \mu) \in \mathcal{N}_\beta$ with $\mu \le \mu^0$ and any $\theta \in [0,1]$, let $w = (u, v) \in \Re^{2n}$ satisfy $H'_\mu(z)w = -H_\mu(z)$ or, by (2), $Yu + Xv = \mu e - Xy$, $F'(x)u - v = y - F(x)$.

Then, $(x^+, y^+) := z + \theta w = (x + \theta u, y + \theta v)$ satisfies

$$\begin{aligned} X^+y^+ &= (I + \theta D)X(y + \theta v) \\ &= \theta\mu e + (1 - \theta)Xy + \theta^2 DXv \\ &= \theta\mu e + (1 - \theta)Xy + \theta^2(\mu d - XYd - XYDd), \end{aligned}$$

where $d := X^{-1}u$. Since $\beta_1\mu e \le Xy \le \beta_2\mu e$ so that $\|XY\| \le \beta_2\mu$, this implies

$$\begin{aligned} & \theta\mu e + (1 - \theta)\beta_1\mu e - \theta^2(\|d\| + \beta_2\|d\| + \beta_2\|d\|^2)\mu e \\ \le\ X^+y^+ \le\ & \theta\mu e + (1 - \theta)\beta_2\mu e + \theta^2(\|d\| + \beta_2\|d\| + \beta_2\|d\|^2)\mu e, \end{aligned}$$

so, for $\mu^+ := (1 - \sigma\theta)\mu$, we have $\beta_1\mu^+e \le X^+y^+ \le \beta_2\mu^+e$ whenever θ is sufficiently small and, in particular,

$$\theta(\|d\| + \beta_2\|d\| + \beta_2\|d\|^2) \le \min\{1 - \beta_1(1 - \sigma), \beta_2(1 - \sigma) - 1\}. \tag{8}$$

Also, $F'(x)u - v = y - F(x)$ implies

$$\begin{aligned} \|F(x^+) - y^+\| &= \|F(x + \theta u) - F(x) - \theta F'(x)u + (1 - \theta)(F(x) - y)\| \\ &\le \|F(x + \theta u) - F(x) - \theta F'(x)u\| + (1 - \theta)\|F(x) - y\| \\ &\le r(x, \theta u) + (1 - \theta)\beta_3\sqrt{\mu}, \end{aligned}$$

where we denote the 1st-order remainder $r(x, a) := \|F(x + a) - F(x) - F'(x)a\|$ for all $a \in \Re^n$. Since the right-hand side is below $\beta_3\sqrt{\mu^+}$ whenever

$$r(x, \theta u) \le \beta_3\sqrt{\mu}(\sqrt{1 - \sigma\theta} - (1 - \theta)) = \beta_3\sqrt{\mu}\theta\frac{2 - \sigma - \theta}{\sqrt{1 - \sigma\theta} + 1 - \theta} \le \beta_3\sqrt{\mu}\theta\sqrt{1 - \sigma} \tag{9}$$

which, by our choice of σ and the fact $r(x, \theta u)/\theta \to 0$ as $\theta \to 0$, occurs whenever θ is sufficiently small, it follows from our choice of θ^t that θ^t is well defined and positive for all t. Moreover, either $\theta^t = 1$ or else (8) or (9) (with $x = x^t, u = u^t, \mu = \mu^t$, where $(u^t, v^t) = w^t$) fail for $\theta = \theta^t/\psi$, implying

$$\theta^t/\psi > \frac{\min\{1 - \beta_1(1 - \sigma), \beta_2(1 - \sigma) - 1\}}{\|d^t\| + \beta_2\|d^t\| + \beta_2\|d^t\|^2} \quad \text{or} \quad \frac{r(x^t, (\theta^t/\psi)u^t)}{\theta^t/\psi} > \beta_3\sqrt{\mu^t}\sqrt{1 - \sigma}, \tag{10}$$

where $d^t := (X^t)^{-1}u^t$.

(a) By assumption, there exist some subsequence T of $\{0, 1, ...\}$ and some $z^\infty = (x^\infty, y^\infty) \in \Re^{2n}$ such that $\{z^t\}_{t\in T} \to z^\infty$. Since μ^t is monotonically decreasing, $\{\mu^t\} \to$ some $\mu^\infty \ge 0$. Since $(z^t, \mu^t) \in \mathcal{N}_\beta$ for all $t \in T$, then if $\mu^\infty = 0$, we would have in the limit $z^\infty \ge 0$, $X^\infty y^\infty = 0, F(x^\infty) - y^\infty = 0$, implying $z^\infty \in S$ (see (1) and (3)). Thus, it remains to consider the case $\mu^\infty > 0$. Since $\hat{\mu}^t \le (1 - \sigma\theta^t)\mu^t$ so that, by (6), $\mu^{t+1} \le (1 - \sigma\theta^t)\mu^t$ for all t, we see from $\{\mu^t\} \to \mu^\infty > 0$ that $\{\theta^t\} \to 0$. Moreover, a continuity argument shows that $(z^\infty, \mu^\infty) \in \mathcal{N}_\beta$ with $\mu^\infty \le \mu^0$, so $H'_{\mu^\infty}(z^\infty)$ is nonsingular. Since $H'_\mu(z)$ is continuous entrywise in z and independent of

μ, this and (4) imply $\{w^t\}_{t\in T} \to -[H'_{\mu^\infty}(z^\infty)]^{-1}H_{\mu^\infty}(z^\infty)$ so $\{u^t\}_{t\in T}$ converges. Since $X^\infty y^\infty \geq \beta_1\mu^\infty e$ so that $x^\infty > 0$, this and $\{x^t\}_{t\in T} \to x^\infty$ imply $\{d^t\}_{t\in T}$ converges. Thus, the left-hand sides of (10) tend to zero as $t \to \infty$, $t \in T$, while, by our choice of σ and $\mu^t \geq \mu^\infty > 0$, the right-hand sides of (10) are bounded away from zero for all $t \in T$, a contradiction.

(b) By (4) and (7), we have $\|d^t\| \leq \kappa(\|X^ty^t - \mu^te\|/\mu^t + \|F(x^t) - y^t\|/\sqrt{\mu^t}) \leq \kappa(\max\{1-\beta_1, \beta_2 - 1\}\sqrt{n}+\beta_3) = \hat{\kappa}$, as well as $\|u^t\| \leq \kappa(\|X^ty^t - \mu^te\|/\sqrt{\mu^t} + \|F(x^t) - y^t\|) \leq \hat{\kappa}\sqrt{\mu^t}$ for all t, so our assumption on F yields

$$r(x^t, (\theta^t/\psi)u^t) \leq \lambda(\theta^t/\psi)^2\|u^t\|^2 \leq \lambda(\theta^t/\psi)^2\hat{\kappa}^2\mu^t.$$

This together with (10) yields that θ^t is bounded below by a positive constant depending on $\beta_1, \beta_2, \psi, \sigma, \hat{\kappa}, \sqrt{\mu^0}\lambda/\beta_3$ only. Since $\mu^{t+1} \leq (1 - \sigma\theta^t)\mu^t$ for all t, the global linear convergence of $\{\mu^t\}$ follows. ∎

If in addition $\|\hat{z}^t - z^t\|$ is in the order of $\sqrt{\mu^t}$ (such as when $\hat{z}^t$ is given by (15)–(18) and the assumptions of Proposition 2 hold), then the global linear convergence of μ^t in Proposition 1(b) yields, as a byproduct, that $\{z^t\}$ converges linearly in the root sense. Also, the convergence ratio c depends on λ, μ^0 through the product $\sqrt{\mu^0}\lambda/\beta_3$ only and, in the case where F is affine (so $\lambda = 0$), c does not depend on μ^0.

We next have a lemma giving a characterization of P_0-matrices [5] and an accompanying corollary showing that $F'(x)$ being a P_0-matrix is sufficient and in some sense necessary for $H'_\mu(x,y)$ to be nonsingular. The lemma, as is noted in [7, page 112], appeared in a paper of Sandberg and Willson and was rediscovered by others. The corollary is analogous to [31, Corollary 1] for the Chen-Mangasarian smoothing functions and is an analog of [13, Lemma 4.1] for the case where F is affine. [The sufficient part can be inferred from [12, Lemma 5.4] or [13, Lemma 4.1], though not the necessary part.] It is known [21, Corollary 5.3 and Theorem 5.8] that $F'(x)$ being a P_0-matrix for all $x \in \Re^n_{++}$ is equivalent to F being a P_0-function on $\Re^n_{++}$, i.e., for every $x, y \in \Re^n_{++}$ with $x \neq y$, there is an $i \in N$ with $x_i \neq y_i$ and $(x_i - y_i)(F_i(x) - F_i(y)) \geq 0$.

Lemma 1 *An $M \in \Re^{n\times n}$ is a P_0-matrix if and only if $D + M$ is nonsingular for all diagonal $D \in \Re^{n\times n}$ with positive diagonal entries.*

Corollary 1 *Fix any $x \in \Re^n_{++}$. If $F'(x)$ is a P_0-matrix, then $H'_\mu(x,y)$ is nonsingular for all $\mu \in \Re_{++}$ and $y \in \Re^n_{++}$. If $F'(x)$ is not a P_0-matrix, then for every $\mu \in \Re_{++}$ there exists $y \in \Re^n_{++}$ such that $H'_\mu(x,y)$ is singular.*

Proof. Using (2), we have for every $\mu \in \Re_{++}$ and $y \in \Re^n_{++}$ that

$$\begin{aligned} \det\left[H'_\mu(x,y)\right] &= \det\begin{bmatrix} Y & X \\ M & -I \end{bmatrix} \\ &= \det\begin{bmatrix} I & -X \\ 0 & I \end{bmatrix} \cdot \det\begin{bmatrix} Y + XM & 0 \\ M & -I \end{bmatrix} \\ &= (-1)^n\det\left[Y + XM\right] \\ &= (-1)^n\det[X]\det\left[X^{-1}Y + M\right], \end{aligned} \tag{11}$$

where $M := F'(x)$.

Assume M is a P_0-matrix. Since $X^{-1}Y$ is a diagonal matrix with positive diagonal entries, so, by Lemma 1, $X^{-1}Y + M$ is nonsingular (in fact, a P-matrix) and hence, by (11), $H'_\mu(x, y)$ is nonsingular.

Assume M is not a P_0-matrix. By Lemma 1, there exists a diagonal $D \in \Re^{n\times n}$ with positive diagonal entries such that $D + M$ is singular. Then, for each $\mu \in \Re_{++}$, choose $y \in \Re^n_{++}$ so $X^{-1}Y = D$, i.e., $y = Dx$. For this choice of y, (11) yields that $\det[H'_\mu(x, y)] = 0$. ∎

Proposition 1(a) and Corollary 1 show that if $F'(x)$ is a P_0-matrix for all $x \in \Re^n_{++}$, then boundedness of $\{z^t\}$ generated by Algorithm 1 implies every cluster point solves CP. Interestingly, it has been shown by Kojima et al. [13, page 33] that LCP with P_0-matrix is NP-complete. [Conditions for boundedness of $\{z^t\}$ will be discussed in Lemma 3.] Next, we have the following lemma and an accompanying corollary showing that $F'(\bar{x})_{\bar{J}^c\bar{J}^c}$ being a P-matrix [5, 13, 21] for a certain $\bar{J} \subset N$ is sufficient for $H'_\mu(x, y)^{-1}$ to be in the order of $1/\sqrt{\mu}$ in the weak sense of (7) (with an explicit estimate of κ given) for all (x, y) near $(\bar{x}, \bar{y})$. In what follows, we denote

$$\mathrm{mod}[A] := \min_{u\in\Re^k:\|u\|=1} \left\{ \max_{i\in\{1,\ldots,k\}} u_i(Au)_i \right\}$$

for any $A \in \Re^{k\times k}$ and k (see [5, 33]).

Lemma 2 *Consider any $M \in \Re^{n\times n}$ and any $I \subset N$ such that M_{II} is a P-matrix, i.e., $\mathrm{mod}[M_{II}] \geq \rho$ for some $\rho \in \Re_{++}$. Fix any $\beta_1 \in (0, 1)$ and $\eta \geq \max\{\|M_{II^c}\|, \|M_{I^cI}\|, \sqrt{\|M_{I^cI^c}\|}\}$ and let $\varrho := \sqrt{\beta_1\rho/(\rho + 2)}$. Then for every $(x, y, \mu) \in \Re^{2n}_{++} \times \Re_{++}$ satisfying $Xy \geq \beta_1\mu e$ and $\{i \in N : x_i > \sqrt{\mu}\varrho/\eta\} = I$ and every $(r, s) \in \Re^{2n}$, we have*

$$\begin{bmatrix} Y & X \\ M & -I \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} r \\ s \end{bmatrix} \implies \max\left\{\|u\|, \|X^{-1}u\|\sqrt{\mu}\right\} \leq \kappa\left(\frac{\|r\|}{\sqrt{\mu}} + \|s\|\right), \qquad (12)$$

where $\kappa := (2\eta/\rho + 2 + \rho/\eta)\max\{1/\eta, 1/\varrho, \eta/\varrho^2\}$.

Proof. Upon eliminating v in (12), we can write the equation in (12) as

$$(M + X^{-1}Y)u = s + X^{-1}r.$$

For convenience, let $J := I^c$ (so $\|X_J\| \leq \sqrt{\mu}\varrho/\eta$). Letting $d := X^{-1}u$, we have that

$$\begin{aligned}
\|u_I\|\|s_I + X_I^{-1}r_I\| &\geq \sum_{j\in I} |u_j||s_j + r_j/x_j| \\
&\geq u_i(s_i + r_i/x_i) \\
&= u_i(M_{iI}u_I + M_{iJ}u_J + x_i^{-1}y_iu_i) \\
&\geq u_iM_{iI}u_I - |u_i||M_{iJ}u_J| \\
&\geq \rho\|u_I\|^2 - \|u_I\|\|M_{IJ}\|\|X_J\|\|X_J^{-1}u_J\| \\
&\geq \rho\|u_I\|^2 - \|u_I\|\sqrt{\mu}\varrho\|d_J\|,
\end{aligned}$$

where $i \in I$ satisfies $u_i M_{iI} u_I = \max_{j\in I} u_j M_{jI} u_I$, so

$$\rho\|u_I\| - \sqrt{\mu}\varrho\|d_J\| \le \|s_I + X_I^{-1} r_I\|. \tag{13}$$

We have similarly that

$$\begin{aligned}
\|d_J\|\|X_J s_J + r_J\| &\ge d_J^T(X_J s_J + r_J) \\
&= d_J^T(X_J M_{JI} u_I + (X_J M_{JJ} X_J + X_J Y_J) d_J) \\
&\ge -\|d_J\|(\|X_J\|\|M_{JI}\|\|u_I\| + \|X_J\|\|M_{JJ}\|\|X_J\|\|d_J\|) + d_J^T X_J Y_J d_J \\
&\ge -\|d_J\|(\sqrt{\mu}\varrho\|u_I\| + \mu\varrho^2\|d_J\|) + \beta_1\mu\|d_J\|^2,
\end{aligned}$$

so

$$-\sqrt{\mu}\varrho\|u_I\| + (\beta_1 - \varrho^2)\mu\|d_J\| \le \|X_J s_J + r_J\|. \tag{14}$$

Letting $a := \|u_I\|$ and $b := \sqrt{\mu}\|d_J\|$ and dividing (14) on both sides by $\sqrt{\mu}$, we obtain from (13) and (14) that

$$\begin{bmatrix} \rho & -\varrho \\ -\varrho & \beta_1 - \varrho^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \le \begin{bmatrix} \|s_I + X_I^{-1} r_I\| \\ \|X_J s_J + r_J\|/\sqrt{\mu} \end{bmatrix}.$$

Since $(\beta_1 - \varrho^2)\rho = 2\varrho^2$ so the left-hand matrix is diagonally dominant and hence has an inverse with positive entries, multiplying both sides by this inverse and using $\|X_I^{-1}\| \le \eta/(\sqrt{\mu}\varrho)$ and $\|X_J\| \le \sqrt{\mu}\varrho/\eta$ yields

$$\begin{bmatrix} a \\ b \end{bmatrix} \le \begin{bmatrix} 2/\rho & 1/\varrho \\ 1/\varrho & \rho/\varrho^2 \end{bmatrix} \begin{bmatrix} \|s_I\| + \|r_I\|\eta/(\sqrt{\mu}\varrho) \\ \|s_J\|\varrho/\eta + \|r_J\|/\sqrt{\mu} \end{bmatrix}.$$

Since $\|u_I\| = a$ and $\|u_J\| = \|X_J d_J\| \le \|X_J\|\|d_J\| \le b\varrho/\eta$, this yields, after some algebra, that

$$\|u\| \le \|u_I\| + \|u_J\| \le \eta^{-1}(2\eta/\rho + 2 + \rho/\eta)\|s\| + \varrho^{-1}(2\eta/\rho + 2 + \rho/\eta)\|r\|/\sqrt{\mu}.$$

Similarly, since $\|d_I\| = \|X_I^{-1} u_I\| \le \|X_I^{-1}\|\|u_I\| \le a\eta/(\sqrt{\mu}\varrho)$ and $\|d_J\| = b/\sqrt{\mu}$, the above inequality yields

$$\|d\| \le \|d_I\| + \|d_J\| \le \varrho^{-1}(2\eta/\rho + 2 + \rho/\eta)\|s\|/\sqrt{\mu} + \varrho^{-2}\eta(2\eta/\rho + 2 + \rho/\eta)\|r\|/\mu.$$

This and the definition of κ shows that (12) holds. ∎

Corollary 2 *Fix any $(\bar{x}, \bar{y}) \in \Re^{2n}_+$ and any $\beta = (\beta_1, \beta_2, \beta_3) \in (0,1) \times (1,\infty) \times \Re_{++}$. Suppose there exist $\bar{I} \subset N, \bar{J} \subset N$ and $\rho, \eta \in \Re_{++}$ such that $\operatorname{mod}[F'(\bar{x})_{II}] > \rho$ and $\max\{\|F'(\bar{x})_{II^c}\|, \|F'(\bar{x})_{I^cI}\|, \sqrt{\|F'(\bar{x})_{I^cI^c}\|}\} < \eta$ for all $\bar{I} \subset I \subset \bar{J}^c$. Also suppose that $\beta_1\omega\bar{x}_{\bar{I}} > \bar{y}_{\bar{I}}$ and $\beta_2\omega\bar{x}_{\bar{J}} < \bar{y}_{\bar{J}}$, where $\varrho := \sqrt{\beta_1\rho/(\rho+2)}$ and $\omega := \eta^2/\varrho^2$. Then there exists $\delta \in \Re_{++}$ such that, for every $(x, y, \mu) \in \mathcal{N}_\beta$ with $\|(x,y) - (\bar{x}, \bar{y})\| \le \delta$ and every $(r, s) \in \Re^{2n}$, we have that (7) holds with κ defined as in Lemma 2.*

Proof. Fix any $\delta \in \Re_{++}$ small enough so that for every $(x,y) \in \Re^{2n}_{++}$ with $\|(x,y) - (\bar{x},\bar{y})\| \le \delta$, we have $\beta_1 \omega x_{\bar{I}} > y_{\bar{I}}$ and $\beta_2 \omega x_{\bar{J}} < y_{\bar{J}}$ and $\operatorname{mod}[F'(x)_{II}] \ge \rho$ and $\max\{\|F'(x)_{II^c}\|, \|F'(x)_{I^cI}\|, \sqrt{\|F'(x)_{I^cI^c}\|}\} \le \eta$ for all $\bar{I} \subset I \subset \bar{J}^c$. Then, if in addition $(x,y,\mu) \in \mathcal{N}_\beta$ for some $\mu \in \Re_{++}$, we have that $I := \{i \in N : x_i > \sqrt{\mu/\omega}\}$ satisfies $\bar{I} \subset I \subset \bar{J}^c$ (since $i \in \bar{I}$ implies $\beta_1\mu \le x_i y_i < \beta_1\omega(x_i)^2$ so $x_i > \sqrt{\mu/\omega}$ and $i \in \bar{J}$ implies $\beta_2\mu \ge x_i y_i > \beta_2 \omega (x_i)^2$ so $x_i < \sqrt{\mu/\omega}$), it readily follows from Lemma 2 with $M := F'(x)$ and the fact $H'_\mu(x,y) = \begin{bmatrix} Y & X \\ M & -I \end{bmatrix}$ that (7) holds with the given κ for any $(r,s) \in \Re^{2n}$. ■

Corollary 2, together with Proposition 1(b) and Corollary 1, shows that Algorithm 1 attains local linear convergence provided that, in addition to F being a P_0-function and $\{z^t\}$ being bounded, the following quantity

$$\operatorname{mod}[F'(\bar{x})_{\bar{J}^c\bar{J}^c}] = \min_{I \subset \bar{J}^c} \operatorname{mod}[F'(\bar{x})_{II}]$$

is uniformly bounded away from zero at all cluster points $(\bar{x},\bar{y}) \in S$ of $\{z^t\}$, where $\bar{J} := \{i \in N : \bar{y}_i > 0\}$, and F is sufficiently smooth (see [1, 31] for related results for non-interior methods). In contrast to the local linear convergence result in [12, Section 7], the above result does not require $(\bar{x},\bar{y})$ to be strictly complementary. If F is a uniformly P-function on $\Re^n_{++}$ [21] so that $\operatorname{mod}[F'(x)_{II}]$ is uniformly bounded away from zero for all $x \in \Re^n_{++}$ and $I \subset N$, then Algorithm 1 attains global linear convergence (boundedness of $\{z^t\}$ is implied by Lemma 3 below). To our knowledge, this is the first result showing that an interior path-following method attains global linear convergence on CP with uniformly P-function. Moreover, with some algebra, an explicit formula for the convergence ratio c in terms of the parameters in Proposition 1(b), with κ expressed in terms of the above quantity, can be derived. In the case of LCP with P-matrix, an analogous result was obtained by Ye [33] for a potential reduction algorithm.

To ensure that $\{z^t\}$ generated by Algorithm 1 is bounded, we need to show that $(z,\mu) \in \mathcal{N}_\beta$ is bounded whenever μ is bounded. The following lemma shows that this holds if F is an R_0-function in the sense of [30, Lemma 4.3] (of which a uniformly P-function is a special case) or if F is monotone and a certain Slater point $\bar{x}$ exists. The proof of this lemma is nearly identical to that of [30, Lemma 4.3], with only minor changes to account for a different choice of the neighborhood $\mathcal{N}_\beta$, and is omitted for brevity.

Lemma 3 *Fix any $\beta = (\beta_1,\beta_2,\beta_3) \in (0,1) \times (1,\infty) \times \Re_{++}$ and any $\mu^0 \in \Re_{++}$. The set $\{\ (x,y) \in \Re^{2n} : (x,y,\mu) \in \mathcal{N}_\beta$ for some $\mu \in (0,\mu^0]\ \}$ is bounded if any of the following two conditions holds:*
(a). F is monotone and there exists $\bar{x} \in \Re^n_+$ satisfying $F(\bar{x}) > \beta_3\sqrt{\mu^0}e$.
(b). F is an R_0-function in the sense that, for any sequence $x^1, x^2, \ldots$ in $\Re^n_{++}$ satisfying

$$\{\|x^t\|\} \to \infty \quad \textit{and} \quad \liminf_{t\to\infty} (\min_{i\in N} F_i(x^t))/\|x^t\| \ge 0,$$

we have $\{(x^t)^T F(x^t)/\|x^t\|\} \to \infty$.

In the case where F is affine, F being an R_0-function is equivalent to $M = F'$ being an R_0-matrix, which includes P-matrix and strictly semimonotone matrix as special cases [5]. Also, related results were given in [11, Sections 4 and 5] for the set $\{(x,y) \in \Re^{2n}_+ : H_\mu(x,y) = \mu(0,b), \mu \le \mu^0\}$, where $b \in \Re^n_{++}$. For example, it is shown therein that, when F is affine, this set is bounded when $M = F'$ is an L_*-matrix (also see [34, Section 3.4] for further results of this kind).

4 Local Superlinear Convergence Analysis

In this section we consider a special choice of $(\hat{z}^t, \hat{\mu}^t)$ in Algorithm 1, based on active-set strategy, that can be computed using only one extra backsolve and for which local superlinear convergence can be shown without assuming the limit point is strictly complementary. In contrast, other superlinearly convergent interior methods for CP require the limit point to be strictly complementary and nondegenerate [12, 24, 27, 30, 32]. The nondegeneracy assumption could be relaxed to a constant rank assumption on F', but strict complementarity was still required [19, 25, 28]. The choice of $\hat{z}^t = (\hat{x}^t, \hat{y}^t)$ and $\hat{\mu}^t$ we consider, constructed from $z^t = (x^t, y^t)$ and μ^t, is:

$$\hat{\mu}^t := \min\{(1-\sigma\theta^t)\mu^t, \pi(\mu^t)\}, \tag{15}$$

$$\hat{x}^t_{I^t} := \bar{x}^t_{I^t}, \hat{y}^t_{I^t} := \hat{\mu}^t(\bar{X}^t_{I^t})^{-1}e, \hat{x}^t_{J^t} := \hat{\mu}^t(\bar{Y}^t_{J^t})^{-1}e, \hat{y}^t_{J^t} := \bar{y}^t_{J^t}, \hat{x}^t_{(K^t)^c} := \hat{y}^t_{(K^t)^c} := \sqrt{\beta_1\hat{\mu}^t}e, \tag{16}$$

provided $H'_{\mu^t}(z^t)_{K^tK^t}$ is nonsingular and $(\bar{x}^t_{I^t}, \bar{y}^t_{J^t}) > 0$ (otherwise we construct $(\hat{z}^t, \hat{\mu}^t)$ by other means), where $\pi : \Re_{++} \mapsto \Re_{++}$ satisfies $\lim_{\mu\to 0}\pi(\mu)/\mu = 0$ and $\bar{z}^t = (\bar{x}^t, \bar{y}^t)$ satisfies the following reduced system of linear equations:

$$H'_{\mu^t}(z^t)_{K^tK^t}(\bar{z}^t_{K^t} - z^t_{K^t}) = -H_0(z^t_{K^t}, 0)_{K^t}, \qquad \bar{z}^t_{(K^t)^c} = 0, \tag{17}$$

with

$$I^t := \{i \in K^t : x^t_i > y^t_i\},\ J^t := \{i \in K^t : x^t_i \le y^t_i\},\ K^t := \{i \in N : \max\{x^t_i, y^t_i\} > \sqrt[4]{\mu^t}\}. \tag{18}$$

[Here and subsequently, for any $A \in \Re^{2n\times 2n}$ and $z \in \Re^{2n}$ and $K \subset N$, we denote by A_{KK} the matrix obtained by removing all rows and columns of A not indexed by K nor $n+K$, by z_K the vector obtained by removing all components of z not indexed by K nor $n+K$, and by $(z_K, 0)$ the vector obtained by replacing all components of z not indexed by K nor $n+K$ with zero.] The index sets I^t, J^t, K^t estimate, respectively,

$$I^* := \{i \in N : x^*_i > 0, y^*_i = 0\},\ J^* := \{i \in N : x^*_i = 0, y^*_i > 0\},\ K^* := I^* \cup J^*, \tag{19}$$

when $z^t = (x^t, y^t)$ is near an $z^* = (x^*, y^*) \in S$. In fact, we would have $I^t = I^*$, $J^t = J^*$ if μ^t is small and $\|z^t - z^*\|$ is in the order of $\sqrt{\mu^t}$. The vector $\bar{z}^t$ given by (17) may be viewed as a generalization of that used in [12, 24, 27, 30, 32] which corresponds to $K^t = N$. We modify $\bar{z}^t$ to obtain a nearby vector $\hat{z}^t$ that, together with $\hat{\mu}^t$, lies in $\mathcal{N}_\beta$ (see (29)). For monotone LCP, Mizuno had proposed an infeasible interior-point method that is superlinearly convergent without assuming existence of a strictly complementary solution [17]. This method also entails estimating the index sets (19), but otherwise is quite different from ours. Lastly, instead of $\sqrt[4]{\mu}$, any postive-valued function of μ that tends to zero slower than $\sqrt{\mu}$ can be used in (18).

An important feature of $\bar{z}^t$ given by (17), introduced in [31] in the context of a non-interior method, is that it can be computed with little extra effort. For example, if we compute w^t given by (4) using Gaussian elimination on the $n \times (n+1)$ matrix $[H'_{\mu^t}(z^t)\ H_{\mu^t}(z^t)]$, then we can simultaneously compute $\bar{z}^t$ by appending to this matrix the column $H_0(z^t_{K^t}, 0)$ and ordering the row reduction so that the rows indexed by K^t are reduced first (to, say, upper triangular form). The extra effort for computing $\bar{z}^t$ in this manner is in the order of $|K^t|^2$. Alternatively, we can compute the inverse of $H'_{\mu^t}(z^t)_{K^tK^t}$, represented in some suitable form, and use this inverse to compute the inverse of $H'_{\mu^t}(z^t)$ using an appropriate formula.

We will show that if $(\hat{z}^t, \hat{\mu}^t)$ is chosen according to (15)–(18) and z^t is sufficiently near an $z^* = (x^*, y^*) \in S$ that is nondegenerate in the sense of [31], namely, $F'(x^*)_{I^*I^*}$ is nonsingular and

$$\left\{F'(x^*)_{II}u_I = 0, F'(x^*)_{(I^c\setminus J^*)I}u_I \geq 0, u_{I\setminus I^*} \geq 0 \implies u_I = 0\right\} \quad \forall I^* \subset I \subset (J^*)^c, \tag{20}$$

where I^*, J^* are given by (19), then $z^t \to z^*$ and $\mu^t \to 0$ superlinearly (and "almost" quadratically if F is sufficiently smooth and π is almost quadratic near zero). Thus, our condition for local superlinear convergence is less restrictive than the widely used condition of z^* being a strictly complementary nondegenerate solution, which further assumes $I^* \cup J^* = N$ [12, 24, 27, 30, 32]. In particular, for the following example from [31, Section 4]:

$$n = 2, \quad F(x) = Mx, \quad M = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}, \quad x^* = y^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

we have $I^* = J^* = \emptyset$ and $F'(x) = M$ is a P_0-matrix for all $x \in \Re^2$, and it can be checked that (20) holds. However, (x^*, y^*) is not strictly complementary.

The proof of our result is rather intricate since we need to show that superlinear convergence in μ^t can be achieved while maintaining $(z^t, \mu^t) \in \mathcal{N}_\beta$ for all t and eventually identifying the index sets (19) implicitly. [The index sets (19) cannot be identified explicitly since x^* cannot be known exactly.] We start with the following lemma showing that the distance from z to any $z^* = (x^*, y^*) \in S$ satisfying (20) is in the order of $\sqrt{\mu}$, for all $(z, \mu) \in \mathcal{N}_\beta$ near $(z^*, 0)$. This local error bound is analogous to one given in [31, Lemma 5] for a non-interior neighborhood and is derived similarly.

Lemma 4 *Fix any $\beta = (\beta_1, \beta_2, \beta_3) \in (0,1) \times (1, \infty) \times \Re_{++}$ and any $z^* = (x^*, y^*) \in S$ satisfying (20), where I^*, J^* are given by (19). There exists $\delta_1 \in \Re_{++}$ such that for all $(z, \mu) = (x, y, \mu) \in \mathcal{N}_\beta$ satisfying $\|z - z^*\| + \mu \leq \delta_1$, we have $\|z - z^*\| \leq \sqrt{\mu}/\delta_1$.*

Proof. Suppose our lemma is false so that, for every $\delta > 0$, there exists some $(z, \mu) = (x, y, \mu) \in \mathcal{N}_\beta$ such that $\|z - z^*\| + \mu \le \delta$ but $\|z - z^*\| > \sqrt{\mu}/\delta$. Since $(z, \mu) \in \mathcal{N}_\beta$, we have from (3) that $\beta_1\mu \le x_i y_i \le \beta_2\mu$ for all $i \in N$. Thus, for each $i \in I := \{i \in N : x_i \ge y_i\}$,

$$\beta_2\mu \ge x_i y_i \ge (y_i)^2. \tag{21}$$

Similarly, for each $i \in J := \{i \in N : x_i < y_i\} = I^c$,

$$\beta_2\mu \ge x_i y_i > (x_i)^2. \tag{22}$$

Also, we have $\|F(x) - y\| \le \beta_3\sqrt{\mu}$ and $F(x^*) - y^* = 0$, implying

$$\begin{aligned}
\beta_3\sqrt{\mu}/\|z - z^*\| &\ge \|F(x) - F(x^*) + y^* - y\|/\|z - z^*\| \\
&= \|M(x - x^*) + y^* - y\|/\|z - z^*\| \\
&= \left\| \begin{bmatrix} M_{II} & 0 \\ M_{JI} & -I \end{bmatrix} \begin{bmatrix} x_I - x_I^* \\ y_J - y_J^* \end{bmatrix} + \begin{bmatrix} M_{IJ}(x_J - x_J^*) - (y_I - y_I^*) \\ M_{JJ}(x_J - x_J^*) \end{bmatrix} \right\| /\|z - z^*\|
\end{aligned}$$

where we denote $M := \int_0^1 F'(x^* + \tau(x - x^*))d\tau$. As $\delta \to 0$, we obtain from $\|z - z^*\| + \mu \le \delta$ and $\|z - z^*\| > \sqrt{\mu}/\delta$ that $\|z - z^*\| \to 0$ (so that M converges to $F'(x^*)$ entrywise and, for all δ sufficiently small, $I^* \subset I$ and $J^* \subset J$, implying $y_I^* = 0, x_J^* = 0$ and $x_i - x_i^* \ge y_i - y_i^*$ for all $i \in I \backslash I^*$, $x_i - x_i^* < y_i - y_i^*$ for all $i \in J \backslash J^*$) and $\sqrt{\mu}/\|z - z^*\| \to 0$ (so that, by (21), $y_{I^*}/\|z - z^*\| \to 0$ and, by (22), $x_{J^*}/\|z - z^*\| \to 0$). By passing to a subsequence if necessary, we assume I and J are fixed. Then, the above inequality yields in the limit

$$0 \ge \left\| \begin{bmatrix} F'(x^*)_{II} & 0 \\ F'(x^*)_{JI} & -I \end{bmatrix} \begin{bmatrix} u_I \\ v_J \end{bmatrix} \right\|, \quad v_I = 0,\ u_J = 0,\ u_{I \backslash I^*} \ge v_{I \backslash I^*},\ u_{J \backslash J^*} \le v_{J \backslash J^*}, \tag{23}$$

where (u, v) is any cluster point of $(x - x^*, y - y^*)/\|z - z^*\|$. Thus, (23) yields $F'(x^*)_{II} u_I = 0$ and $F'(x^*)_{(J \backslash J^*)I} u_I = v_{J \backslash J^*} \ge 0$ and $u_{I \backslash I^*} \ge 0$. Moreover, we have $(u, v) \ne 0$, so either $u_I \ne 0$ or $v_J \ne 0$. Since, by (23), $F'(x^*)_{JI} u_I = v_J$ so that $v_J \ne 0$ implies $u_I \ne 0$, it must be that $u_I \ne 0$. Since $J = I^c$, this contradicts (20). ∎

The next lemma shows that $\hat{z}^t$ given by (16)–(17) is well defined and approaches an $z^* = (x^*, y^*) \in S$ superlinearly provided that $\|z^t - z^*\|$ is in the order of $\sqrt{\mu^t}$ and $I^t = I^*$, $J^t = J^*$ (so the degenerate components are identified) and $F'(x^*)_{I^*I^*}$ is nonsingular, where I^* and J^* are given by (19). This lemma is analgous to [31, Lemma 5] for the Chen-Mangasarian smoothing functions, although here the analysis is complicated by the need to modify $\bar{z}^t$ to obtain a nearby $\hat{z}^t = (\hat{x}^t, \hat{y}^t)$ that is positive and satisfies $\beta_1\hat{\mu}^t e \le \hat{X}^t \hat{y}^t \le \beta_2\hat{\mu}^t e$.

Lemma 5 *Fix any $\beta = (\beta_1, \beta_2, \beta_3) \in (0, 1) \times (1, \infty) \times \Re_{++}$, and $z^* = (x^*, y^*) \in S$ such that $F'(x^*)_{I^*I^*}$ is nonsingular and there exists $\delta_1 \in \Re_{++}$ such that for all $(z, \mu) \in \mathcal{N}_\beta$ satisfying $\|z - z^*\| + \mu \le \delta_1$ we have $\|z - z^*\| \le \sqrt{\mu}/\delta_1$, where I^*, J^* are given by (19). Then, there exist $\delta_2 \in \Re_{++}$ and $\chi : \Re_{++} \mapsto \Re_{++}$ with $\lim_{\mu \to 0} \chi(\mu)/\sqrt{\mu} = 0$ such that* *for any $(z,\mu)\in\mathcal{N}_\beta$ with $\|z-z^*\|+\mu\le\delta_2$ and any $\hat\mu\in\Re_{++}$, we have that $H'_\mu(z)_{KK}$ is nonsingular and $(\bar x_I,\bar y_J)>0$ and $\|\hat z_K-z^*_K\|\le\chi(\mu)+\hat\mu/\delta_2$, where $\hat z_K=(\hat x_K,\hat y_K)$ is given by*

$$\hat x_I:=\bar x_I,\quad \hat y_I:=\hat\mu(\bar X_I)^{-1}e,\quad \hat x_J:=\hat\mu(\bar Y_J)^{-1}e,\quad \hat y_J:=\bar y_J, \tag{24}$$

and $\bar z=(\bar x,\bar y)\in\Re^{2n}$ satisfies

$$H'_\mu(z)_{KK}(\bar z_K-z_K)=-H_0(z_K,0)_K,\qquad \bar z_{K^c}=0, \tag{25}$$

with $I=I^*,J=J^*,K=I\cup J$. If in addition there exists $\lambda\in\Re_{++}$ such that $\|F(x+u)-F(x)-F'(x)u\|\le\lambda\|u\|^2$ whenever $\|x-x^*\|$ and $\|u\|$ are sufficiently small ($x\in\Re^n_+,u\in\Re^n$), then we can further choose χ such that $\limsup_{\mu\to0}\chi(\mu)/\mu<\infty$.

Proof. Consider the continuously differentiable mapping $\bar H:\Re^{2|K|}\mapsto\Re^{2|K|}$ given by

$$\bar H(z_K):=H_0(z_K,0)_K\quad\text{or, equivalently,}\quad \bar H(x_K,y_K):=\begin{bmatrix} X_Ky_K\\ F(x_K,0)_K-y_K\end{bmatrix}.$$

[Here $(x_K,0)$ denotes the vector obtained by replacing all components of x not indexed by K with zero.] For all $(z,\mu)\in\mathcal{N}_\beta$ satisfying $\|z-z^*\|+\mu\le\delta_1$ we have

$$H'_\mu(z)_{KK}-\bar H'(z_K)=\begin{bmatrix} Y_K & X_K\\ M_{KK} & -I\end{bmatrix}-\begin{bmatrix} Y_K & X_K\\ \bar M_{KK} & -I\end{bmatrix}=\begin{bmatrix} 0 & 0\\ M_{KK}-\bar M_{KK} & 0\end{bmatrix}, \tag{26}$$

where $M=F'(x)$ and $\bar M=F'(x_K,0)$. Also, we have that $\|x_{K^c}\|=\|x_{K^c}-x^*_{K^c}\|\le\|z-z^*\|\le\sqrt{\mu}/\delta_1$ so the continuity of F' yields $\|M_{KK}-\bar M_{KK}\|=\|F'(x)_{KK}-F'(x_K,0)_{KK}\|\to0$ as $\mu\to0$. Thus, the matrix difference (26) tends to zero entrywise as $\mu\to0$. Since $\bar H(z^*_K)=0$ and $\bar H(z_K)=H_0(z_K,0)_K$, (25) implies

$$H'_\mu(z)_{KK}(\bar z_K-z^*_K)=(\bar H'(z_K)-H'_\mu(z)_{KK})(z^*_K-z_K)+\bar r(z_K), \tag{27}$$

where $\bar r(z_K):=\bar H(z^*_K)-\bar H(z_K)-\bar H'(z_K)(z^*_K-z_K)$. Since $\bar H'$ is continuous so that $\|\bar r(z_K)\|/\|z_K-z^*_K\|\to0$ as $\|z_K-z^*_K\|\to0$, it follows from (26) tending to zero entrywise and $\|z-z^*\|\le\sqrt{\mu}/\delta_1$ that the right-hand side of (27) tends to zero componentwise faster than $\sqrt{\mu}$. Also, as $(z,\mu)\to(z^*,0)$, $H'_\mu(z)_{KK}$ converges entrywise to the matrix

$$\begin{bmatrix} 0 & 0 & X^*_I & 0\\ 0 & Y^*_J & 0 & 0\\ M^*_{II} & M^*_{IJ} & -I & 0\\ M^*_{JI} & M^*_{JJ} & 0 & -I\end{bmatrix},$$

where $M^*:=F'(x^*)$ (see (26)), which is nonsingular due to M^*_{II} being nonsingular. Thus, by taking δ_1 smaller if necessary, we can further assume that $\|(H'_\mu(z)_{KK})^{-1}\|$ is bounded, so then (27) yields $\|\bar z_K-z^*_K\|/\sqrt{\mu}\to0$ uniformly in z as $\mu\to0$. In other words, there exists $\chi:\Re_{++}\mapsto\Re_{++}$ with $\lim_{\mu\to0}\chi(\mu)/\sqrt{\mu}=0$ such that

$\|\bar{z}_K - z_K^*\| \leq \chi(\mu)$ whenever $(z, \mu) \in \mathcal{N}_\beta$ and $\|z - z^*\| + \mu \leq \delta_1$. By taking δ_1 smaller if necessary, we can assume that $\bar{x}_I \geq x_I^*/2 > 0$ and $\bar{y}_J \geq y_J^*/2 > 0$. Then, (24) yields $\hat{z}_K > 0$ and $\|(\hat{x}_I, \hat{y}_J) - (x_I^*, y_J^*)\| \leq \chi(\mu)$ and $\|(\hat{y}_I, \hat{x}_J) - (y_I^*, x_J^*)\| = \hat{\mu}\|((\bar{X}_I)^{-1}e, (\bar{Y}_J)^{-1}e)\| \leq 2\hat{\mu}\|((X_I^*)^{-1}e, (Y_J^*)^{-1}e)\|$.

If in addition F' has the stated Lipschitz property, then we see that $\|\bar{r}(z_K)\| = O(\|z_K - z_K^*\|^2) = O(\mu)$ and $\|M_{KK} - \bar{M}_{KK}\| = O(\sqrt{\mu})$ for all z sufficiently near z^* and all $\mu \in \Re_{++}$ sufficiently small, from which we deduce that $\|\bar{z}_K - z_K^*\| = O(\mu)$ and the desired conclusion follows. [Here $b = O(a)$ means $b \leq \kappa a$ for some constant $\kappa \in \Re_{++}$ independent of (z, μ).] ∎

We note that F' has the additional Lipschitz property in Lemma 5 when F is twice continuously differentiable on an open set containing $\Re_+^n$. The next lemma is based on the fact that F is locally Lipschitz continuous, so that $\|F(x) - y\|$ is in the order of $\|z - z^*\|$ whenever $z = (x, y)$ is near an $z^* \in S$.

Lemma 6 *Fix any $\beta_3 \in \Re_{++}$ and any $z^* = (x^*, y^*) \in S$. Let $\kappa := .9/(\|F'(x^*)\| + 1)$. There exists $\delta_3 \in \Re_{++}$ such that for all $z = (x, y) \in \Re^{2n}$ and $\mu \in (0, \delta_3]$ satisfying $\|z - z^*\| \leq \kappa\beta_3\sqrt{\mu}$, we have $\|F(x) - y\| \leq \beta_3\sqrt{\mu}$.*

Proof. For any $z = (x, y) \in \Re^{2n}$ and $\mu \in \Re_{++}$ satisfying $\|z - z^*\| \leq \kappa\beta_3\sqrt{\mu}$, we have

$$\begin{aligned} \|F(x) - y\| &= \|F(x) - F(x^*) + y^* - y\| \\ &= \|M(x - x^*) + y^* - y\| \\ &\leq \|M\|\|x - x^*\| + \|y^* - y\| \\ &\leq (\|M\| + 1)\kappa\beta_3\sqrt{\mu} \\ &\leq .9(\|M\| + 1)\beta_3\sqrt{\mu}/(\|F'(x^*)\| + 1), \end{aligned}$$

where $M := \int_0^1 F'(x^* + \tau(x - x^*))d\tau$. Since M converges entrywise to $F'(x^*)$ as $\|z - z^*\| \to 0$, by taking μ sufficiently small, we would have $.9(\|M\|+1) \leq \|F'(x^*)\|+1$, implying $\|F(x) - y\| \leq \beta_3\sqrt{\mu}$. ∎

By combining Lemmas 4, 5 and 6, we obtain the following local superlinear convergence result for Algorithm 1 with $(\hat{z}^t, \hat{\mu}^t)$ chosen according to (15)–(18).

Proposition 2 *Fix any $z^* = (x^*, y^*) \in S$ such that $F'(x^*)_{I^*I^*}$ is nonsingular and (20) holds, where I^*, J^*, K^* are given by (19). Fix any $\beta = (\beta_1, \beta_2, \beta_3) \in (0, 1) \times (1, \infty) \times \Re_{++}$ with $\sqrt{2n\beta_1} \leq \kappa\beta_3/2$, where κ is the constant given in Lemma 6. Let χ be the function given in Lemma 5 (via Lemma 4). Assume $H'_\mu(z)$ is nonsingular for all $(z, \mu) \in \mathcal{N}_\beta$ with $\mu \leq \mu^0 \in \Re_{++}$. Let $\{(z^t, \mu^t)\}_{t=0,1...}$ be generated by Algorithm 1 with $\hat{\mu}^t$ and $\hat{z}^t = (\hat{x}^t, \hat{y}^t)$ chosen by (15)–(16) whenever $(\bar{x}_{I^t}^t, \bar{y}_{J^t}^t) > 0$, where $\bar{z}^t = (\bar{x}^t, \bar{y}^t)$ and I^t, J^t, K^t are given by (17)–(18) and π is chosen such that $\lim_{\mu \to 0} \pi(\mu)/\mu = 0$ and $\kappa\beta_3\sqrt{\pi(\mu)}/4 \geq \chi(\mu)$ for all $\mu \in \Re_{++}$ sufficiently small. Then, there exists $\delta \in \Re_{++}$ such that if $\|z^{\bar{t}} - z^*\| + \mu^{\bar{t}} \leq \delta$ for some $\bar{t}$, then $\{z^t\}$ converges to z^* and $\mu^{t+1} \leq \pi(\mu^t)$ for all $t \geq \bar{t}$.*

Proof. By Proposition 1, $\{(z^t, \mu^t)\}_{t=0,1\ldots}$ is well defined and satisfies $(z^t, \mu^t) \in \mathcal{N}_\beta$ for all t. By $\pi(\mu)/\mu \to 0$ as $\mu \to 0$, we have $\pi(\mu^t) \le (1-\sigma)\mu^t$ whenever μ^t is sufficiently small and, by Lemmas 4 and 5, we have that $H'_{\mu^t}(z^t)_{K^tK^t}$ is nonsingular and $(\bar{x}^t_{I^t}, \bar{y}^t_{J^t}) > 0$ whenever $\|z^t - z^*\| + \mu^t$ is sufficiently small and $I^t = I^*, J^t = J^*$, in which case $\hat{\mu}^t$ and $\hat{z}^t$ are given by (15) and (16). Then Lemmas 4 and 5 imply that there exists $\delta_2 \in \Re_{++}$ which, together with χ, has the property that

$$\left\{ \begin{array}{c} \|z^t - z^*\| \le \sqrt{\mu^t}/\delta_2, (\bar{x}^t_{I^t}, \bar{y}^t_{J^t}) > 0, \\ \|\hat{z}^t_{K^t} - z^*_{K^t}\| \le \chi(\mu^t) + \hat{\mu}^t/\delta_2 \end{array} \right\} \text{ whenever } \|z^t - z^*\| + \mu^t \le \delta_2, I^t = I^*, J^t = J^*. \tag{28}$$

Also, (16) and $K^t = I^t \cup J^t$ imply $\hat{X}^t_{K^t}\hat{y}^t_{K^t} = \hat{\mu}^t e$ and $\hat{X}^t_{(K^t)^c}\hat{y}^t_{(K^t)^c} = \beta_1\hat{\mu}^t e$ and $\hat{z}^t > 0$ whenever $(\bar{x}^t_{I^t}, \bar{y}^t_{J^t}) > 0$ and, if in addition $K^t = K^*$, then $\|\hat{z}^t_{(K^t)^c} - z^*_{(K^t)^c}\| = \|\hat{z}^t_{(K^t)^c}\| = \sqrt{2|(K^t)^c|\beta_1\hat{\mu}^t} \le \kappa\beta_3\sqrt{\hat{\mu}^t}/2$. So it follows from Lemma 6 and (3) that there exists $\delta_3 \in \Re_{++}$ such that

$$(\hat{z}^t, \hat{\mu}^t) \in \mathcal{N}_\beta \text{ whenever } \|\hat{z}^t_{K^t} - z^*_{K^t}\| \le \kappa\beta_3\sqrt{\hat{\mu}^t}/2, \hat{\mu}^t \le \delta_3, (\bar{x}^t_{I^t}, \bar{y}^t_{J^t}) > 0, K^t = K^*.$$

Since $\hat{\mu}^t = \pi(\mu^t)$ (see (15)) so that $\chi(\mu^t) + \hat{\mu}^t/\delta_2 \le \kappa\beta_3\sqrt{\hat{\mu}^t}/2$ and $\hat{\mu}^t \le \delta_3$ whenever μ^t is sufficiently small, the above two facts imply that there exists $\delta \in (0, \delta_2]$ such that

$$(\hat{z}^t, \hat{\mu}^t) \in \mathcal{N}_\beta, \|\hat{z}^t - z^*\| \le \kappa\beta_3\sqrt{\hat{\mu}^t} \text{ whenever } \|z^t - z^*\| + \mu^t \le \delta, I^t = I^*, J^t = J^*. \tag{29}$$

By taking δ smaller if necessary (depending on $\min_{i\in I^*} x_i^*$, $\min_{i\in J^*} y_i^*$ and δ_2), we can also assume that whenever $\|z^t - z^*\| + \mu^t \le \delta$, we have $x_i^t > \sqrt[4]{\mu^t}$ for all $i \in I^*$, $y_i^t > \sqrt[4]{\mu^t}$ for all $i \in J^*$, and from $\|z^t - z^*\| \le \sqrt{\mu^t}/\delta_2$ (see (28)) that $|x_i^t| \le \sqrt{\mu^t}/\delta_2 < \sqrt[4]{\mu^t}$ and $|y_i^t| \le \sqrt{\mu^t}/\delta_2 < \sqrt[4]{\mu^t}$ for all $i \in (K^*)^c$, in which case (18) yields $I^t = I^*$ and $J^t = J^*$. Since we also have $\hat{\mu}^t = \pi(\mu^t) \le (1-\sigma)\mu^t$, (29) and our choice of α^t in Algorithm 1 yield $\alpha^t = 1$, $z^{t+1} = \hat{z}^t$, $\mu^{t+1} = \hat{\mu}^t$ and so

$$\|z^{t+1} - z^*\| \le \kappa\beta_3\sqrt{\mu^{t+1}}, \ \mu^{t+1} = \pi(\mu^t) \le (1-\sigma)\mu^t \ \text{ whenever } \|z^t - z^*\| + \mu^t \le \delta. \tag{30}$$

By taking δ smaller if necessary, we can assume that $\delta^2 \le \delta/2$ and $\kappa\beta_3\sqrt{\pi(\mu)} \le \sqrt{\mu}/2$ for all $\mu \in (0, \delta]$. Then, if $\|z^{\bar{t}} - z^*\| \le \delta/2$ and $\mu^{\bar{t}} \le \delta^2$, a simple induction argument using (30) yields $\|z^t - z^*\| \le \kappa\beta_3\sqrt{\pi(\mu^t)} \le \delta/2$ and $\mu^{t+1} = \pi(\mu^t) \le \delta^2 \le \delta/2$ for all $t \ge \bar{t}$. This shows that $\{\mu^t\} \to 0$ superlinearly and $\{z^t\} \to z^*$. ∎

Proposition 2 shows that the rate of local superlinear convergence depends on the rate at which $\chi(\mu) \to 0$ as $\mu \to 0$. If F' has the additional Lipschitz property in Lemma 5 so that χ can be chosen such that $\limsup_{\mu\to 0}\chi(\mu)/\mu < \infty$, then any π having the additional property $\lim_{\mu\to 0}\pi(\mu)/\mu^2 = \infty$ can be used (e.g., $\pi(\mu) = \mu^2|\ln(\mu)|$). This allows Algorithm 1 to achieve a convergence rate as close to quadratic as desired. The condition $\sqrt{2n\beta_1} \le \kappa\beta_3/2$ is rather conservative and requires estimating

κ. A more practical alternative is to start with any $\beta_1 \in (0,1)$ and $\beta_3 \in \Re_{++}$ and to periodically decrease β_1 or increase β_3 when $(\hat{z}^t, \hat{\mu}^t)$ given by (15)–(18) fails to satisfy $\|F(\hat{x}^t) - \hat{y}^t\| \leq \beta_3 \sqrt{\hat{\mu}^t}$. Finally, it may be possible to relax the assumption of $F'(x^*)_{I^*I^*}$ being nonsingular to a constant rank type of assumption similar to those used in [19, 25, 28], but this would require further investigation.

5 Preliminary Computational Experience

To gain some understanding of the numerical behavior/performance of Algorithm 1 with $\hat{\mu}^t$ and $\hat{z}^t$ chosen as described in Proposition 2, we implemented this method in Matlab and ran it on a set of test problems. In this section we describe the implementation and report on our preliminary numerical experience with it.

In our Matlab implementation of Algorithm 1, we choose $\beta_1 = 1/n$, $\beta_2 = \max\{n, 4\}$, $\sigma = \psi = .7$. For any given $x^0 \in \Re^n_{++}$, we set $y^0 = \mu^0 (X^0)^{-1} e$ and choose μ^0 to minimize the ratio $\|F(x^0) - y^0\| / \sqrt{\mu^0}$. [A little calculus yields $\mu^0 = \|F(x^0)\| / \|(X^0)^{-1} e\|$.] Then β_3 is set to this minimum ratio. [Thus $(x^0, y^0, \mu^0) \in \mathcal{N}_\beta$ with $\beta = (\beta_1, \beta_2, \beta_3)$.] After some experimentation, we settled on the choice $\nu^t = 0$ (corresponding to α^t being either 0 or 1) which yielded fewer function evaluations than other choices of ν^t. To further reduce the number of function evaluations in the line search, we modified the calculation of θ^t whereby we start with $\theta = \bar{\theta}^t$ (instead of $\theta = 1$), where

$$\bar{\theta}^t := \min\{1, \theta^{t-1}/\psi, .95 \min_{i: w_i^t < 0} \{-z_i^t / w_i^t\}\}$$

and $\theta^{-1} := 1$, and then θ is successively decreased by a factor of ψ until $(z^t + \theta w^t, (1 - \sigma\theta)\mu^t)$ is in $\mathcal{N}_\beta$. Notice that if θ^{t-1} is small, then $\bar{\theta}^t$ will also be small so θ will typically need to be decreased only one or two times. Moreover, it can be checked that this modification does not affect the convergence properties of the method. Finally, we choose $\hat{\mu}^t$ and $\hat{z}^t = (\hat{x}^t, \hat{y}^t)$ by (15)–(16) whenever $(\bar{x}^t_{I^t}, \bar{y}^t_{J^t}) > 0$ (otherwise we set $\alpha^t = 0$), where $\bar{z}^t = (\bar{x}^t, \bar{y}^t)$ and I^t, J^t, K^t are given by (17)–(18) and π is chosen to be $\pi(\mu) = \mu^{1.8}$. We remark that, while the above parameter choices are reasonable, they were made without much fine-tuning and can conceivably be improved.

In our tests, we set the starting point x^0 to be e multiplied by either .1 or 1 or 10, and we terminate the method whenever $\mu^t \leq 10^{-7}$. The performance of the method on a set of seven test problems is summarized in Table 1. The test problems are (see [8, 15, 18]):

Kojima-Shindo problem: Here $n = 4$ and F is not a P_0-function on $\Re^n_+$.

Watson problem: Here $n = 5$ and F is the gradient of a strictly convex function $f(x) = exp(\sum_{i=1}^5 (x_i - i + 2)^2)$. The solution is unique but not strictly complementary.

Nash-Cournot problem: Here $n = 10$ and F is defined on $\Re^n_+$ only.

Lemke problem: Here $n = 30$ and $F(x) = Mx + q$, where the (i,j)th entry of $M \in \Re^{n \times n}$ equals 2 (respectively, 1 and 0) if $j > i$ (respectively, $j = i$ and $j < i$)

Problem		$x^0 = (.1, ..., .1)^T$	$x^0 = (1, ..., 1)^T$	$x^0 = (10, ..., 10)^T$
Name	n	**iter/nf/nfast/resid**	**iter/nf/nfast/resid**	**iter/nf/nfast/resid**
Kojima-Shindo	4	$12/22/4/6{\cdot}10^{-11}$	$11/26/4/4{\cdot}10^{-10}$	$19/46/7/3{\cdot}10^{-11}$
Watson	5	$20/30/4/2{\cdot}10^{-2}$	$20/36/6/2{\cdot}10^{-7}$	fail[1]
Nash-Cournot	10	$14/18/4/2{\cdot}10^{-14}$	$12/18/6/4{\cdot}10^{-14}$	$14/24/5/2{\cdot}10^{-14}$
Lemke	30	$9/15/2/2{\cdot}10^{-11}$	$13/24/2/6{\cdot}10^{-12}$	$17/33/2/3{\cdot}10^{-12}$
Harker-Pang	30	$22/31/3/3{\cdot}10^{-11}$	$16/22/2/4{\cdot}10^{-10}$	$19/30/2/2{\cdot}10^{-11}$
SkewSymLCP	30	$37/52/1/2{\cdot}10^{-4}$	$24/37/0/2{\cdot}10^{-4}$	$29/43/1/2{\cdot}10^{-4}$
SymPSDLCP	30	$37/70/4/3{\cdot}10^{-4}$	$20/34/6/5{\cdot}10^{-4}$	fail[2]

Table 1: Performance of Algorithm 1 on seven test problems, as indicated by the number of iterations to termination (iter), the number of F-evaluations (nf), the number of iterations in which $\alpha^t \neq 0$ (nfast), and the residual $\|\min\{x, F(x)\}\|$ upon termination (res). ([1] As $t \to \infty$, μ^t is stuck at $5.42... \cdot 10^{14}$ and $\|H'_{\mu^t}(z^t)\|$ is in the order of 10^{16}. [2] As $t \to \infty$, μ^t is stuck at .0067... and Matlab warns of $H'_{\mu^t}(z^t)$ being close to singular.)

for all i and j (so M is positive semidefinite), and every entry of q equals -1. The problem is one for which Lemke's method is known to run in exponential time.

Harker-Pang problem: Here $n = 30$ and $F(x) = Mx+q$, where $M = AA^T+B+D$ and every entry of $A \in \Re^{n\times n}$ and of skew-symmetric $B \in \Re^{n\times n}$ is uniformly generated from $(-5, 5)$ and every diagonal entry of the diagonal $B \in \Re^{n\times n}$ is uniformly generated from $(0, 0.3)$ (so M is positive definite), and every entry of $q \in \Re^n$ is uniformly generated from $(-500, 0)$.

Skew Symmetric LCP: Here $n = 30$ and $F(x) = Mx + q$, where $M = A - A^T$ and every entry of $A \in \Re^{n\times n}$ is uniformly generated from $[-5, 5]$ (so M is skew-symmetric), and $q = -M\bar{x}+\bar{y}$, where each entry of $\bar{x}$ has equal probability of being 0 or being uniformly generated from $[0, 5]$ and each entry of $\bar{y}$ is 0 if the corresponding entry of $\bar{x}$ is nonzero and otherwise has equal probability of being 0 or being uniformly generated from $[0, 5]$ (so $(\bar{x}, \bar{y})$ is typically not strictly complementary).

Symmetric Positive Semi-Definite LCP: Here $n = 30$ and $F(x) = Mx+q$, where $M = AA^T$ and every entry of $A \in \Re^{n\times 10}$ is uniformly generated from $[-1, 1]$ (so M is symmetric positive semi-definite), and $q = -M\bar{x} + \bar{y}$, where each entry of $\bar{x}$ has equal probability of being 0 or being uniformly generated from $[0, 10]$ and each entry of $\bar{y}$ is 0 if the corresponding entry of $\bar{x}$ is nonzero and otherwise has equal probability of being 0 or being uniformly generated from $[0, 10]$ (so $(\bar{x}, \bar{y})$ is typically not strictly complementary).

As can be seen from Table 1, the best performance is obtained when $x^0 = e$, suggesting that the proper scaling of x^0 is important. For $x^0 = 10 \cdot e$, failure occured on the Watson problem due to taking the exponential of a large positive number. The failure on the symmetric positive semi-definite LCP is less easily explained. One possibility is that a large x^0 leads to a large μ^0 (see the formula for μ^0 earlier), for which boundedness of the iterates $\{z^t\}$ is no longer assured by Lemma 3(a). However, this is still speculative since, while z^t had some large components, the Matlab warning made it difficult to know whether these components really were diverging or not. Regardless, it suggests that one is better off to err on the side of small x^0 than large x^0. The number of iterations, between 9 and 37, is reasonable for the level of accuracy achieved. In [15], the same set of test problems were solved using a non-interior Newton-type method in fewer iterations but with less accuracy. [This method also solves an $n \times n$ system of linear equations at each iteration, so its work per iteration is similar to that for Algorithm 1.] We also ran our implementation on an oligopolistic pricing example described in [26]. For this example, $n = 63$ and F does not appear to be a P_0-function. As with the method in [26], termination of Algorithm 1 on this example depended on the starting point x^0. For some choices of x^0, termination occured within 25 to 38 iterations. For other choices, the method made very slow progress due to $\|w^t\|$ becoming large (in the order of 10^{10}) and, accordingly, θ^t becoming very small (in the order of $1/\|w^t\|^2$ as dictated by the centering condition $Xy \geq \beta_1 \mu e$), yielding an iterate change $\theta^t w^t$ that was in the order of 10^{-10} only.

In general, we found the use of $\hat{\mu}^t$ and $\hat{z}^t$ to significantly accelerate the local convergence of Algorithm 1, enabling μ^t to decrease very rapidly in the last few iterations before termination. Moreover, in cases where $\{z^t\}$ converged to a non-strictly complementary solution $z^* = (x^*, y^*)$, the index sets I^t, J^t, K^t correctly identified I^*, J^*, K^* in the later iterations. [However, we did not check whether $F'(x^*)_{I^*I^*}$ is nonsingular or (20) holds.] The stepsize θ^t was often equal to $\bar{\theta}^t$, so it required few function evaluations to compute.

We also ran our implementation on some randomly generated knapsack problems that were formulated as LCP (with P_0-matrix) using the transformation of Kojima et al. [13, page 33]. As these are NP-complete problems, our intent was not so much to solve them but rather to see where convergence breaks down. As expected, for most choices of x^0, the iterates $\{z^t\}$ diverged, though slowly, with $\|w^t\|$ large (in the order of 10^7 to 10^8) and θ^t small (in the order of 10^{-10} to 10^{-9}). [If $\{z^t\}$ did not diverge, then Proposition 1(a) and Corollary 1 would imply that any cluster point of $\{z^t\}$ is a solution.] We also experimented with perturbing the diagonal entries of the P_0-matrix by a small positive number to make it a P-matrix, which would ensure that $\{z^t\}$ remains bounded (see Lemma 3(b)) and thus a solution would be found by Algorithm 1. However, the (unique) solution of the perturbed problem is typically far from solutions of the original problem.

References

[1] Chen, B. and Chen, X., A global linear and local quadratic continuation smoothing method for variational inequalities with box constraints, Report, Department of Management and Systems, Washington State University, Pullman, Washington (March 1997).

[2] Chen, B. and Xiu, N., A global linear and local quadratic non-interior continuation method for nonlinear complementarity problems based on Chen-Mangasarian smoothing function, Report, Department of Management and Systems, Washington State University, Pullman, Washington (February 1997).

[3] Chen, X., Qi, L., and Sun, D., Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities, Report, School of Mathematics, University of New South Wales, Sydney, Australia (May 1996; revised December 1996).

[4] Cottle, R. W., Giannessi, F., and Lions, J.-L., Eds., *Variational Inequalities and Complementarity Problems: Theory and Applications*, John Wiley & Sons, New York, New York, 1980.

[5] Cottle, R. W., Pang, J.-S., and Stone, R. E., *The Linear Complementarity Problem*, Academic Press, New York, New York, 1992.

[6] Ferris, M. C. and Pang, J.-S., editors, *Complementarity and Variational Problems: State of the Art*, SIAM Publishing, Philadelphia, Pennsylvania, 1997.

[7] Gabriel, S. A. and Moré, J. J., Smoothing of mixed complementarity problems, in *Complementarity and Variational Problems: State of the Art*, edited by M. C. Ferris and J.-S. Pang, SIAM Publishing, Philadelphia, Pennsylvania, 1997, 105-116.

[8] Harker, P., and Pang, J.-S., Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications, *Math. Programming*, 48 (1990), 161–220.

[9] Jansen, B., Roos, K., Terlaky, T., and Yoshise, A., Polynomiality of primal-dual affine scaling algorithms for nonlinear complementarity problems, *Math. Programming*, 78 (1997), 315–345.

[10] Kojima, M., Mizuno, S., and Noma, T., A new continuation method for complementary problems with uniform P-functions, *Math. Programming*, 43 (1989), 107-113.

[11] Kojima, M., Megiddo, N., and Mizuno, S. A general framework of continuation methods for complementarity problems, *Math. Oper. Res.*, 18 (1993), 945–963.

[12] Kojima, M., Megiddo, N., and Noma T., Homotopy continuation methods for nonlinear complementarity problems, *Math. Oper. Res.*, 16 (1991), 754–774.

[13] Kojima, M., Megiddo, N., Noma, T. and Yoshise, A., *A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems*, Lecture Notes in Computer Science 538, Springer-Verlag, Berlin, 1991.

[14] Kojima, M., Noma, T., and Yoshise, A., Global convergence in infeasible-interior-point algorithms, *Math. Programming*, 65 (1994), 43–72.

[15] Luo, Z.-Q. and Tseng, P., A new class of merit functions for the nonlinear complementarity problem, in *Complementarity and Variational Problems: State of the Art*, edited by M. C. Ferris and J.-S. Pang, SIAM Publishing, Philadelphia, Pennsylvania, 1997, 204–225.

[16] McLinden, L., The complementarity problem for maximal monotone multifunctions, in *Variational Inequalities and Complementarity Problems: Theory and Applications*, edited by R. W. Cottle, F. Giannessi and J.-L. Lions, John Wiley & Sons, New York, NY, 1980, 251–270.

[17] Mizuno, S., A superlinearly convergent infeasible-interior-point algorithm for geometrical LCPs without a strictly complementary condition, *Math. Oper. Res.*, 21 (1996), 382–400.

[18] Monteiro, R. D. C., Pang, J.-S. and Wang, T., A positive algorithm for the nonlinear complementarity problem, *SIAM J. Optim.*, 5 (1995), 129–148.

[19] Monteiro, R. D. C. and Zhou, F., On superlinear convergence of infeasible-interior-point algorithms for linearly constrained convex programs," working paper, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta (July 1995); to appear in *Comput. Optim. Appl.*

[20] Moré, J. J., Global methods for nonlinear complementarity problems, *Math. Oper. Res.*, 21 (1996), 589-614.

[21] Moré, J. J. and Rheinboldt, W. C., On P- and S-functions and related classes of n-dimensional nonlinear mappings, *Linear Algebra Appl.*, 6 (1973), 45-68.

[22] Nesterov, Y. E. and Nemirovskii, A. S., *Interior Point Polynomial Methods in Convex Programming: Theory and Applications*, SIAM, Philadelphia, 1993.

[23] Pang, J.-S., Complementarity problems, in *Handbook on Global Optimization*, edited by R. Horst and P. Pardalos, Kluwer Academic Publishers, Norwell, Massachusetts, 1995, 271–338.

[24] Potra, F. A. and Ye, Y., Interior point methods for nonlinear complementarity problems, *J. Optim. Theory Appl.*, 88 (1996), 617–647.

[25] Ralph, D. and Wright, S. J., Superlinear convergence of an interior-point method for monotone variational inequalities, in *Complementarity and Variational Problems: State of the Art*, edited by M. C. Ferris and J.-S. Pang, SIAM Publishing, Philadelphia, Pennsylvania, 1997, 345–385.

[26] Simantiraki, E. M. and Shanno, D. F., An infeasible-interior-point algorithm for solving mixed complementarity problems, in *Complementarity and Variational Problems: State of the Art*, edited by M. C. Ferris and J.-S. Pang, SIAM Publishing, Philadelphia, Pennsylvania, 1997, 386–404.

[27] Sun, J. and Zhao, G., Global linear and local quadratic convergence of a long-step adaptive-mode interior point method for some monotone variational inequality problems", Report, Department of Decision Sciences, National University of Singapore, Singapore (1996); to appear in *SIAM J. Optim.*

[28] Sun, J. and Zhao, G., A long-step quadratically convergent interior point method for nonlinear monotone variational inequality problems", Report, Department of Decision Sciences, National University of Singapore, Singapore, (1997).

[29] Tseng, P., Global linear convergence of a path-following algorithm for some monotone variational inequality problems, *J. Optim. Theory Appl.*, 75 (1992), 265–279.

[30] Tseng, P., An infeasible path-following method for monotone complementarity problems, *SIAM J. Optim.*, 7 (1997), 386–402.

[31] Tseng, P., Analysis of a non-interior continuation method based on Chen-Mangasarian smoothing functions for complementarity problems, Report, Department of Mathematics, University of Washington, Seattle (July 1997); available from Web page
`http://www.math.washington.edu/~tseng/personal.html`

[32] Wright, S. J. and Ralph, D., A superlinear infeasible-interior-point algorithm for monotone complementarity problems, *Math. Oper. Res.*, 21 (1996), 815-838.

[33] Ye, Y., A further result on potential reduction algorithm for the P-matrix linear complementarity problem, in *Advances in Optimization and Parallel Computing*, edited by P. M. Pardalos, North-Holland, New York, 1992, 310–316.

[34] Zhang, D. and Zhang, Y., On constructing interior-point path-following methods for certain semimonotone linear complementarity problems, Technical Report TR97-19, Department of Computational and Applied Mathematics, Rice University, Houston (July 1997).