ANALYSIS OF A NON-INTERIOR CONTINUATION METHOD BASED ON CHEN-MANGASARIAN SMOOTHING FUNCTIONS FOR COMPLEMENTARITY PROBLEMS[1]

July 14, 1997 (revised April 11, 1998)

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

ABSTRACT

Recently Chen and Mangasarian proposed a class of smoothing functions for linear/nonlinear programs and complementarity problems that unifies many previous proposals. Here we study a non-interior continuation method based on these functions in which, like interior path-following methods, the iterates are maintained to lie in a neighborhood of some path and, at each iteration, one or two Newton-type steps are taken and then the smoothing parameter is decreased. We show that the method attains global convergence and linear convergence under conditions similar to those required for other methods. We also show that these conditions are in some sense necessary. By introducing an inexpensive active-set strategy in computing one of the Newton directions, we show that the method attains local superlinear convergence under conditions milder than those for other methods. The proof of this uses a local error bound on the distance from an iterate to a solution in terms of the smoothing parameter. **Keywords**: Complementarity problem, smoothing, continuation method, global linear convergence, local superlinear convergence.

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 $F = (F_1, ..., F_n)$ is a given continuously differentiable function from $\Re^n$ to $\Re^n$ [12, 14, 30]. In the case where F is affine, CP is called the linear complementarity problem or LCP [13, 26]. 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 and, of these, one that has generated much interest recently (and is the focus of this paper) is a non-interior smoothing

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

approach proposed by Chen and Mangasarian [8, 9], based on an earlier work of Kreimer and Rubinstein. In this approach, we rewrite (1) as a set of nonsmooth equations $H_0(x,y)=0$, where

$$H_0(x,y) := \begin{bmatrix} x - \max\{0, x-y\} \\ F(x) - y \end{bmatrix}, \tag{2}$$

and the max is taken componentwise, and then we approximate the nondifferentiable convex function $\max\{0,\tau\}$ by a smooth function of the form

$$\mu g(\tau/\mu), \tag{3}$$

where $\mu \in \Re_{++}$ and $g : \Re \to \Re$ is any convex continuously differentiable function with the properties that $\lim_{\tau\to-\infty} g(\tau) = 0$ and $\lim_{\tau\to\infty} g(\tau) - \tau = 0$ and $0 < g'(\tau) < 1$ for all $\tau \in \Re$. [We denote the class of such functions g by $\mathcal{CM}$. In the case where g is twice continuously differentiable, (3) may be written as the integral $\int_{-\infty}^{\tau}\int_{-\infty}^{\alpha} \frac{1}{\mu} d(\xi/\mu)d\xi d\alpha$, where $d := g''$ is interpreted as a probability density function. This is the form presented in [9] and used by others. We prefer the form (3) for its simplicity of analysis.] This yields a set of nonlinear equations of the form $H_\mu(x,y)=0$, parameterized by the smoothing parameter μ, where

$$H_\mu(x,y) := \begin{bmatrix} x - \mu G((x-y)/\mu) \\ F(x) - y \end{bmatrix}, \qquad G(\tau_1, ..., \tau_n) := \begin{bmatrix} g(\tau_1) \\ \vdots \\ g(\tau_n) \end{bmatrix}. \tag{4}$$

[Thus, H_μ converges to H_0 pointwise as $\mu \to 0$.] Then we can use a continuation/homotopy solution approach whereby, starting with any $\mu \in \Re_{++}$ and $z \in \Re^{2n}$, we fix μ and apply a few Newton-type steps for $H_\mu(z)=0$ to update z, and then we decrease μ and re-iterate. This approach has been widely used [2, 3, 7, 10, 11, 18, 20, 38] and is the one that we will consider. A related approach involves having a separate μ for each component and viewing it as a variable so that one instead works with $3n$ equations in $3n$ variables [18, 31] (in [32], $H_0(x,y)=0$ is reformulated, say via $y := F(x)$ and $u := x-y$, as the normal equation $u + F(\max\{0,u\}) - \max\{0,u\} = 0$, to which smoothing is applied). Also, instead of applying Newton steps to $H_\mu(z)=0$, one can fix μ and minimize $\|H_\mu(z)\|$, possibly inexactly, using standard methods for unconstrained minimization and then decrease μ [6, 8, 9, 15].

Examples of the smoothing function g include one proposed independently by Chen and Harker [4, 5], Kanzow [20, 21], and Smale [33]:

$$g(\tau) = (\sqrt{\tau^2 + 4} + \tau)/2, \tag{5}$$

and one obtained by integrating the sigmoid function $\tau \mapsto 1/(1+e^{-\tau})$ used in neural networks [8, 9]:

$$g(\tau) = \ln(e^\tau + 1). \tag{6}$$

See [8, 9] for other examples of g, and see [15] for extensions of (3) to the mixed complementarity problem. [The graph of $\zeta = g(\tau)$ given by (5) has the equation

$-\tau = 1/\zeta - \zeta$. Other such g can be obtained from the general equation $-\tau = h(\zeta) - \zeta$, with $h : \Re_{++} \mapsto \Re_{++}$ any convex twice continuously differentiable function satisfying $\lim_{\zeta \to 0} h(\zeta) = \infty$ and $\lim_{\zeta \to \infty} h(\zeta) = 0$.] Also, the function (3) has been further generalized to the form $p(\tau, \mu)$ and properties of solution to the equation $H_\mu(z) = 0$ has been studied similarly [6]. Finally, we note the related works in [1, 19, 37], which analyze the global and local convergence rate of continuation methods and generalized Newton methods based on approximating $a - \max\{0, a - b\}$ by a smoothing function $a + b - \sqrt{a^2 + b^2 + 2\mu}$ proposed by Kanzow [20]. [In particular, [1] gave the first global linear convergence analysis of a non-interior continuation method for LCP.] This function cannot be written in the Chen-Mangasarian form of $a - \mu g((a-b)/\mu)$ for some $g \in \mathcal{CM}$, as neither can the function $ab - \mu$ used in interior-point methods [24, 25, 27, 35, 36].

As is mentioned earlier, we are interested in continuation methods and, in particular, ones that maintain (z, μ) to lie in a neighborhood of the "path" defined by $H_\mu(z) = 0$. As in the case of infeasible interior-point methods [35], such a path-following approach appears to be needed for achieving both global linear and local superlinear convergence on nonlinear problems [2, 7, 38]. We will use the following choice of neighborhood:

$$\mathcal{N}_\beta := \{\ (z, \mu) \in \Re^{2n} \times \Re_{++} : \|H_\mu(z)\| \leq \beta\mu\ \}, \tag{7}$$

where $\beta \in \Re_{++}$ is fixed. This choice, an analog of that used in [35] for an interior path-following method, was also used in [2] (see [7] for a similar choice and see [3] for a weighted version involving different μ for different components). Our method iteratively moves (z, μ) 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 $(\hat{z}, \pi(\mu))$, where π is a superlinear function and $\hat{z}$ is any "guess" of a solution (see Algorithm 1). We show that the method converges globally and linearly under conditions similar to those required by other methods [2, 7] (see Proposition 1). Moreover, we show that these conditions are in some sense necessary (see Corollaries 1 and 2) and we give a condition for $(z, \mu) \in \mathcal{N}_\beta$ to be bounded when F is monotone that is simpler than existing conditions (see Lemma 4). Our proofs are also quite compact. Lastly, we propose a choice of $\hat{z}$ that generalizes $z - H'_\mu(z)^{-1} H_0(z)$ used in [2, 3, 7, 10] via an active-set strategy, and we show that the associated method attains local superlinear convergence under conditions milder than those required for other 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 μ (see Lemma 5). Some preliminary computational experience with the above method is reported in Section 5.

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, $\Re_{++}$ denotes the positive reals, and $N := \{1, ..., n\}$. For any vector $x \in \Re^n$, we denote by x_i the ith component of x and, for any $I \subseteq N$, by x_I the vector obtained after removing from x those x_i with $i \notin I$. We also denote by $\|x\|_1$, $\|x\|$, and $\|x\|_\infty$ the 1-, 2- and ∞-norm of x, so $\|x\| = \sqrt{x^T x}$, etc. 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 $\beta \in \Re_{++}$ and $g \in CM$. This method is patterned after an interior path-following method studied in [35] and, except for the choice of the Newton direction (8) and the neighborhood $\mathcal{N}_\beta$ (and the fact that an Armijo-Goldstein line search is used), is quite different from other non-interior path-following methods [2, 7, 38].

Algorithm 1 *Choose any* $(z^0, \mu^0) \in \mathcal{N}_\beta$, *any* ψ, σ *in* $(0, 1)$ *satisfying* $\sqrt{n}g(0)\sigma < (1-\sigma)\beta$, *and any continuous* $\pi : \Re_{++} \mapsto \Re$ *satisfying* $0 < \pi(\mu) \leq (1-\sigma)\mu$ *for all* $\mu \in \Re_{++}$, *and* $\pi(\mu)/\mu \to 0$ *as* $\mu \to 0$. *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*

$$H'_{\mu^t}(z^t)w^t = -H_{\mu^t}(z^t). \tag{8}$$

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* $\hat{z}^t \in \Re^{2n}$ *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*

$$\Big((z^t + \theta^t w^t)(1-\alpha) + \hat{z}^t\alpha, (1-\sigma\theta^t)\mu^t(1-\alpha) + \pi(\mu^t)\alpha\Big) \in \mathcal{N}_\beta, \tag{9}$$

and let

$$z^{t+1} := (z^t + \theta^t w^t)(1-\alpha^t) + \hat{z}^t\alpha^t, \quad \mu^{t+1} := (1-\sigma\theta^t)\mu^t(1-\alpha^t) + \pi(\mu^t)\alpha^t. \tag{10}$$

Roughly speaking, at iteration t of Algorithm 1, we first compute a Newton direction w^t by solving the linear equations (8), 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 $(\hat{z}^t, \pi(\mu^t))$ as possible while remaining in $\mathcal{N}_\beta$. [If $\alpha^t = 1$, then $(z^{t+1}, \mu^{t+1}) = (\hat{z}^t, \pi(\mu^t))$. 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 (note that, by choice of π, $\pi(\mu^t)$ is always below $1-\sigma\theta^t$ and tends to zero superlinearly in μ^t) 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 reasons of practical efficiency and improved local convergence rate.

In Section 4 we will consider a choice of $\hat{z}^t$ that is relatively cheap to compute and yields local superlinear convergence under assumptions milder than those for other methods. 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 (9)), hence (z^{t+1}, μ^{t+1}) is well defined and, by $\sigma \in (0,1)$ and $\pi(\mu^t) > 0$, μ^{t+1} is positive.

3 Global (Linear) Convergence Analysis

In this section we show 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 bounded and F and g are sufficiently smooth, then linear convergence is also achieved. The proof, which uses ideas similar to those used for other path-following methods, is quite compact. Moreover, we show that existing sufficient condition for $H'_\mu(z)$ to be nonsingular (namely, F being a P_0-function) is in some sense also necessary, and similarly for the boundedness of $\|H'_\mu(z)^{-1}\|$ (see Corollaries 1(b), 2(b)). We also refine existing results on boundedness of $\|z\|$ when F is monotone and a feasible interior point exists (see Lemma 4).

First, we derive some basic properties of any $g \in \mathcal{CM}$. Since $g'(\tau) > 0$ for all $\tau \in \Re$ and $\lim_{\tau \to -\infty} g(\tau) = 0$, it must be that $g(\tau)$ is positive and strictly increasing with τ. Similarly, since $g'(\tau) - 1 < 0$ for all τ and $\lim_{\tau \to \infty} g(\tau) - \tau = 0$, it must be that $g(\tau) - \tau$ is positive and strictly decreasing with τ (so $g(\tau) > \max\{0, \tau\}$ for all $\tau \in \Re$). Using these properties, we have the following lemma showing that $\mu g(\tau/\mu)$ is Lipschitz continuous in μ with Lipschitz constant independent of τ. This result has been shown by Kanzow [20, Lemma 3.7] in the case g given by (5) and by Chen and Xiu [7, Proposition 1] in the case g is twice continuously differentiable and g'' is an even function with $\int_0^\infty \tau g''(\tau) d\tau < \infty$. Our proof seems simpler than that given in [7].

Lemma 1 *Fix any $g \in \mathcal{CM}$. For any $\tau \in \Re$ and any scalars $\mu > \nu > 0$, we have $0 < \nu g(\tau/\nu) - \mu g(\tau/\mu) \le g(0)(\mu - \nu)$.*

Proof. Fix any $\tau \in \Re$ and define $h : \Re_{++} \mapsto \Re$ by $h(\mu) := \mu g(\tau/\mu)$ for all $\mu \in \Re_{++}$. Then, h is continuously differentiable and $h'(\mu) = g(\tau/\mu) - g'(\tau/\mu)(\tau/\mu)$. Letting $a := \tau/\mu$ and using the properties that $g(a) > \max\{0, a\}$ and $0 < g'(a) < 1$, we see that $g(a) - g'(a)a > 0$. Also, since g is convex, we have $g(0) \ge g(a) - g'(a)a$. Thus, $0 < h'(\mu) \le g(0)$ for all $\mu \in \Re_{++}$ so, for any scalars $\mu > \nu > 0$, we have from the mean-value theorem that

$$0 < h(\mu) - h(\nu) = h'(\bar{\mu})(\mu - \nu) \le g(0)(\mu - \nu)$$

for some $\bar{\mu} \in [\nu, \mu]$. ■

Lemma 1 also holds if $\nu = 0$, as was shown by Chen and Mangasarian, in the case of g given by (6) [8, Lemma 1.1]. Using this lemma, we obtain the following proposition, establishing the global (linear) convergence of Algorithm 1 under conditions similar to those used for related methods [2, 7]. The proof is, in part, adapted from that used in [35] for an interior path-following method.

Proposition 1 *Fix any $\beta \in \Re_{++}$ and $g \in \mathcal{CM}$. Assume $H'_\mu(z)$ is nonsingular for all $(z,\mu) \in \mathcal{N}_\beta$ with $\mu \le \mu^0 \in \Re_{++}$. Then $\{(z^t,\mu^t,\theta^t)\}_{t=0,1...}$ 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 exist $\kappa, \lambda_1, \lambda_2 \in \Re_{++}$ and $\gamma > 1$ such that $\|H'_{\mu^t}(z^t)^{-1}\| \le \kappa$ for all t and $\|G(r+s) - G(r) - G'(r)s\| \le \lambda_1\|s\|^2$ and $\|F(x+u) - F(x) - F'(x)u\| \le \lambda_2\|u\|^\gamma$ for all $r,s,x,u \in \Re^n$, then $\mu^{t+1} \le c\mu^t$ for all t, where $c \in (0,1)$ depends on $\beta, \psi, \sigma, \sqrt{n}g(0), \lambda_1\kappa^2, \lambda_2\kappa^\gamma(\mu^0)^{\gamma-1}, \gamma$ only.

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 (10). Moreover, our choice (9) of α^t ensures that $(z^{t+1},\mu^{t+1}) \in \mathcal{N}_\beta$ and the property $\pi(\mu) \le (1-\sigma)\mu$ for all $\mu \in \Re_{++}$ ensures that $\mu^{t+1} \le (1-\sigma\theta^t)\mu^t$.

(a) By assumption, there is 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 $t \in T$ and, by properties of g, $\lim_{\mu\to 0,\tau\to\alpha} \mu g(\tau/\mu) = \max\{0,\alpha\}$ for any $\alpha \in \Re$, then if $\mu^\infty = 0$, we would have in the limit $x^\infty - \max\{0, x^\infty - y^\infty\}, F(x^\infty) - y^\infty = 0$, implying $z^\infty \in S$ (see (2)). Thus, it remains to consider the case $\mu^\infty > 0$. For any $(z,\mu) \in \mathcal{N}_\beta$ and $\theta \in [0,1]$, $w := -H'_\mu(z)^{-1}H_\mu(z)$ and $z^+ := z + \theta w$ satisfy

$$\begin{aligned}
\|H_\mu(z^+)\| &= \|H_\mu(z+\theta w) - H_\mu(z) - \theta H'_\mu(z)w + (1-\theta)H_\mu(z)\| \\
&\le \|H_\mu(z+\theta w) - H_\mu(z) - \theta H'_\mu(z)w\| + (1-\theta)\|H_\mu(z)\| \\
&\le r(z,\mu,\theta w) + (1-\theta)\beta\mu,
\end{aligned}$$

where we denote the 1st-order remainder $r(z,\mu,a) := \|H_\mu(z+a) - H_\mu(z) - H'_\mu(z)a\|$. Thus, for $\mu^+ := (1-\sigma\theta)\mu$ and writing $z^+ = (x^+, y^+)$, we have

$$\begin{aligned}
\|H_{\mu^+}(z^+)\| &\le \|H_\mu(z^+)\| + \|H_\mu(z^+) - H_{\mu^+}(z^+)\| \\
&= \|H_\mu(z^+)\| + \|\mu G((x^+ - y^+)/\mu) - \mu^+ G((x^+ - y^+)/\mu^+)\| \\
&\le r(z,\mu,\theta w) + (1-\theta)\beta\mu + \sqrt{n}g(0)(\mu - \mu^+),
\end{aligned}$$

where the equality uses (4) and the last inequality also uses Lemma 1. Since the right-hand side is below $\beta\mu^+$ whenever

$$r(z,\mu,\theta w)/\theta + \sqrt{n}g(0)\sigma\mu \le (1-\sigma)\beta\mu,$$

which, by our choice of σ and the fact $r(z,\mu,\theta w)/\theta \to 0$ 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

$$r(z^t,\mu^t,(\theta^t/\psi)w^t)/(\theta^t/\psi) > ((1-\sigma)\beta - \sqrt{n}g(0)\sigma)\mu^t. \tag{11}$$

Since $\pi(\mu^t) \le (1-\sigma)\mu^t$ so that, by (10), $\mu^{t+1} \le (1-\sigma\theta^t)\mu^t$ for all t, we see from $\{\mu^t\} \to \mu^\infty > 0$ that $\{\theta^t\} \to 0$. Also, since $\{H'_{\mu^t}(z^t)\}_{t\in T}$ converges entrywise to $H'_{\mu^\infty}(z^\infty)$, which is nonsingular, and $\{H_{\mu^t}(z^t)\}_{t\in T}$ converges, (8) implies $\{w^t\}_{t\in T}$ is bounded. These two observations, together with $\{z^t\}_{t\in T} \to z^\infty$ and $\{\mu^t\} \to \mu^\infty > 0$ (and g, F being continuously differentiable), implies the left-hand side of (11) tends to zero as $t \to \infty$, $t \in T$. On the other hand, by our choice of σ and $\mu^t \ge \mu^\infty > 0$, the right-hand side of (11) is bounded away from zero for all t, a contradiction.

(b) By (4) and our assumptions on G and F,

$$\begin{aligned} r(z,\mu,a) &= \left\| \begin{bmatrix} -\mu(G(r+s) - G(r) - G'(r)s) \\ F(x+u) - F(x) - F'(x)u \end{bmatrix} \right\| \\ &\le \mu\|G(r+s) - G(r) - G'(r)s\| + \|F(x+u) - F(x) - F'(x)u\| \\ &\le \mu\lambda_1\|s\|^2 + \lambda_2\|u\|^\gamma \\ &\le 4\lambda_1\|a\|^2/\mu + \lambda_2\|a\|^\gamma, \end{aligned} \tag{12}$$

where for simplicity we write $z := (x,y), a := (u,v)$, $r := (x-y)/\mu, s := (u-v)/\mu$. Also, by (8) and $(z^t,\mu^t) \in \mathcal{N}_\beta$, we have

$$\|w^t\| = \|H'_{\mu^t}(z^t)^{-1}H_{\mu^t}(z^t)\| \le \|H'_{\mu^t}(z^t)^{-1}\|\|H_{\mu^t}(z^t)\| \le \kappa\beta\mu^t.$$

This together with (11) and (12) yields

$$\begin{aligned} 0 < (1-\sigma)\beta - \sqrt{n}g(0)\sigma &< r(z^t,\mu^t,(\theta^t/\psi)w^t)/(\mu^t\theta^t/\psi) \\ &\le 4\lambda_1(\theta^t/\psi)\|w^t\|^2/(\mu^t)^2 + \lambda_2(\theta^t/\psi)^{\gamma-1}\|w^t\|^\gamma/\mu^t \\ &\le 4\lambda_1(\kappa\beta)^2(\theta^t/\psi) + \lambda_2(\kappa\beta)^\gamma(\theta^t/\psi)^{\gamma-1}(\mu^t)^{\gamma-1}, \end{aligned}$$

from which we obtain that θ^t is bounded below by a positive constant depending on $\beta, \psi, \sigma, \sqrt{n}g(0), \lambda_1\kappa^2, \lambda_2\kappa^\gamma(\mu^0)^{\gamma-1}, \gamma$ only. Since $\mu^{t+1} \le (1-\sigma\theta^t)\mu^t$ for all t, the global linear convergence of $\{\mu^t\}$ follows. ■

If in addition $\|\hat{z}^t - z^t\|$ is in the order of μ^t (such as when $\hat{z}^t$ is given by (18)–(19) 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, note that the convergence ratio c depends on $\lambda_1, \lambda_2, \kappa, \mu^0$ through their respective products only and, in the case where F is affine (so $\lambda_2 = 0$), c does not depend on μ^0.

We next have a lemma giving a characterization of P_0-matrices 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 [15, page 112], was first given by Sandberg and Willson and was rediscovered by Chen and Harker [4, Theorem 3.3] and De Luca et al. The sufficient part of the corollary is not new (see [4, Proposition 3.6], [21, Theorem 3.5], [20, Theorem 3.5] for the case of g given by (5) and see [6, Proposition 14], [15, page 112] for the general case) but the necessary part appears to be new (see [4, Proposition 3.6] for a related result). It is known [29, Corollary 5.3 and Theorem 5.8] that $F'(x)$ being a P_0-matrix for all x in an open box $X \subset \Re^n$ is equivalent to F being a P_0-function on X (i.e., for every $x, y \in X$ 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 2 *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 $g \in \mathcal{CM}$ and 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 (4), we have for every $\mu \in \Re_{++}$ and $y \in \Re^n$ that

$$\begin{aligned}
\det\left[H'_\mu(x,y)\right] &= \det\begin{bmatrix} I-E & E \\ M & -I \end{bmatrix} \qquad (13)\\
&= \det\begin{bmatrix} I & -E \\ 0 & I \end{bmatrix} \cdot \det\begin{bmatrix} I-E+EM & 0 \\ M & -I \end{bmatrix}\\
&= (-1)^n \det\left[I - E + EM\right]\\
&= (-1)^n \det[E] \det\left[E^{-1} - I + M\right], \qquad (14)
\end{aligned}$$

where $E := G'((x-y)/\mu)$ and $M := F'(x)$ (so E is a diagonal matrix with diagonal entries in $(0,1)$).

Assume M is a P_0-matrix. Since the diagonal entries of E are between 0 and 1, then $E^{-1} - I$ is a diagonal matrix with positive diagonal entries, so, by Lemma 2, $E^{-1} - I + M$ is nonsingular (in fact, a P-matrix) and hence, by (14), $H'_\mu(x,y)$ is nonsingular.

Assume M is not a P_0-matrix. By Lemma 2, 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 $G'((x-y)/\mu) = (I+D)^{-1}$, i.e., $y_i = x_i - \mu \cdot (g')^{-1}(1/(1+D_{ii}))$ for $i \in N$. For this choice of y, (14) yields that $\det[H'_\mu(x,y)] = 0$. ∎

Next, we have the following lemma and an accompanying corollary showing that, in the case where F is a P_0-function, $\det[F'(\bar{x})_{II}]$ being positive for all $\{i \in N : \bar{x}_i > \bar{y}_i\} \subset I \subset \{i \in N : \bar{x}_i \geq \bar{y}_i\}$ is sufficient and in some sense necessary for $\|H'_\mu(x,y)^{-1}\|$ to be uniformly bounded (with an explicit bound given) near $(\bar{x}, \bar{y})$. The sufficient part can also be inferred from the proof of [2, Proposition 5] while the explicit bound and the necessary part appears to be new.

Lemma 3 *Fix any $\omega \in (0,1)$ and any $M \in \Re^{n\times n}$.*
(a) If $\min_{J\subset N} \det[M_{JJ}] \geq 0$, then for every nonempty $I \subset N$ and every diagonal $E \in \Re^{n\times n}$ with diagonal entries in $[0,1]$ and satisfying $I = \{i \in N : E_{ii} \geq \omega\}$, we have $\left|\det \begin{bmatrix} I-E & E \\ M & -I \end{bmatrix}\right| \geq \det[M_{II}] \min\{\omega, 1-\omega\}^n$.
(b) If $\det[M_{II}] = 0$ for some nonempty $I \subset N$, then for the diagonal $E \in \Re^{n\times n}$ with $E_{II} = I$ and $E_{I^cI^c} = 0$, we have $\det \begin{bmatrix} I-E & E \\ M & -I \end{bmatrix} = 0$.

Proof. For any $M \in \Re^{n\times n}$ and any diagonal $E \in \Re^{n\times n}$ with diagonal entries in $[0,1]$, we have that

$$\begin{aligned} \left|\det \begin{bmatrix} I-E & E \\ M & -I \end{bmatrix}\right| &= \left|\det \begin{bmatrix} I & -E \\ 0 & I \end{bmatrix} \cdot \det \begin{bmatrix} I-E+EM & 0 \\ M & -I \end{bmatrix}\right| \\ &= |\det[I-E+EM]| \\ &= \left|\sum_{J\subset N} \det[I-E_{J^cJ^c}]\det[E_{JJ}]\det[M_{JJ}]\right|, \end{aligned} \tag{15}$$

where the third equality uses formula (2.2.1) in [13].

(a) Since $\det[E_{JJ}], \det[I-E_{J^cJ^c}]$ and $\det[M_{JJ}]$ are nonnegative for all $J \subset N$ and, for $J = I$, we have $\det[E_{JJ}] \geq \omega^{|I|}$ and $\det[I-E_{J^cJ^c}] \geq (1-\omega)^{|I^c|}$, then the right-hand side of (15) is bounded below by $\det[E_{II}]\omega^{|I|}(1-\omega)^{|I^c|}$.

(b) For our particular choice of I and E, the formula above (15) yields that

$$\left|\det \begin{bmatrix} I-E & E \\ M & -I \end{bmatrix}\right| = |\det[I-E_{II}+E_{II}M_{II}]| = |\det[M_{II}]| = 0.$$

■

Corollary 2 *Fix any $g \in \mathcal{CM}$ and any $\bar{z} = (\bar{x}, \bar{y}) \in \Re^{2n}$. Let $\bar{I} := \{i \in N : \bar{x}_i > \bar{y}_i\}$, $\bar{J} := \{i \in N : \bar{x}_i < \bar{y}_i\}$.*
(a) If $\rho := \min_{\bar{I}\subset I\subset \bar{J}^c} \det[F'(\bar{x})_{II}] > 0$ and $\min_{J\subset N} \det[F'(x)_{JJ}] \geq 0$ for all x in some open set containing $\bar{x}$, then there exists $\delta \in \Re_{++}$ such that for every $z \in \Re^{2n}$ with $\|z-\bar{z}\| \leq \delta$ and every $\mu \in \Re_{++}$, we have

$$\left|\det\left[H'_\mu(z)\right]\right| \geq \rho\varpi/2, \qquad \|H'_\mu(z)^{-1}\| \leq 2\|\mathrm{adj}\, H'_\mu(z)\|/(\rho\varpi),$$

where $\varpi := \min\{g'(0), 1-g'(0)\}^n$ and $\mathrm{adj}A$ denotes the adjoint of $A \in \Re^{2n\times 2n}$.
(b) If $\det[F'(\bar{x})_{II}] = 0$ for some $\bar{I} \subset I \subset \bar{J}^c$, then

$$\liminf_{(z,\mu)\to(\bar{z},0)} \det\left[H'_\mu(z)\right] = 0, \qquad \limsup_{(z,\mu)\to(\bar{z},0)} \|H'_\mu(z)^{-1}\| = \infty.$$

Proof. (a) Fix any $\delta \in \Re_{++}$ small enough so that for every $z = (x,y) \in \Re^{2n}$ with $\|z-\bar{z}\| \leq \delta$, we have $x_{\bar{I}} > y_{\bar{I}}$ and $x_{\bar{J}} < y_{\bar{J}}$ and $\min_{\bar{I}\subset I\subset \bar{J}^c} \det[F'(x)_{II}] \geq \rho/2$

and $\min_{J\subset N}\det[F'(x)_{JJ}] \geq 0$. Then, for any $\mu \in \Re_{++}$, by using (13) with $E := G'((x-y)/\mu)$ and $M := F'(x)$ and observing that E is a diagonal matrix with diagonal entries in $(0,1)$ and $I := \{i \in N : x_i \geq y_i\} = \{i \in N : E_{ii} \geq g'(0)\}$ satisfies $\bar{I} \subset I \subset \bar{J}^c$, it readily follows from Lemma 3(a) that

$$\left|\det\left[H'_\mu(z)\right]\right| = \left|\begin{bmatrix} I-E & E \\ M & -I \end{bmatrix}\right| \geq \det[M_{II}]\cdot\varpi \geq \rho\varpi/2.$$

Using the adjoint formula for matrix inverse [17, page 20], i.e., $A^{-1} = \mathrm{adj}A/\det[A]$ for any $A \in \Re^{2n\times 2n}$, completes the proof.

(b) For each $\mu \in \Re_{++}$, let $x_i := \bar{x}_i + \sqrt{\mu}$ and $y_i := \bar{y}_i$ for all $i \in I$ and let $x_i := \bar{x}_i$ and $y_i := \bar{y}_i + \sqrt{\mu}$ for all $i \in I^c$. Define E and M as above. Then, as $\mu \to 0$ (so $E_{II} \to I$ and $E_{I^cI^c} \to 0$), we have from (13) and Lemma 3(b) (and using the continuity property of determinant) that

$$\det[H'_\mu(x,y)] \;=\; \det\begin{bmatrix} I-E & E \\ M & -I \end{bmatrix} \;\to\; 0.$$

Since the entries of $H'_\mu(x,y)$ are bounded as $\mu \to 0$, using the aforementioned adjoint formula for matrix inverse completes the proof. ■

Corollary 2(a), together with Proposition 1(b) and Corollary 1, shows that Algorithm 1 attains linear convergence provided that, in addition to F being a P_0-function, the ratio $\bar{\kappa} := \|F'(\bar{x})\|/\min_{\bar{I}\subset I\subset \bar{J}^c}\det[F'(\bar{x})_{II}]$ is uniformly bounded at all cluster points $\bar{x}$ of the iterate sequence, where $\bar{I}$ and $\bar{J}$ are as defined in Corollary 2, and F and g are sufficiently smooth (see [2, Propositions 4 and 5] for a similar result). Moreover, with some algebra, an explicit formula for the convergence ratio c in terms of the parameters in Proposition 1(b) and $\bar{\kappa}$, $g'(0)$ (replacing κ) can be derived. Corollary 2(b), together with Corollary 1, suggests that, as long as our linear convergence analyses rely on showing $\|H'_\mu(z)^{-1}\|$ to be bounded along the iterate sequence, we cannot hope to significantly improve upon the above sufficient condition. In particular, for any $M \in \Re^{n\times n}$ satisfying $\det[M_{II}] = 0$ for some $I \subset N$, we can choose any $(x^*, y^*) \in \Re^{2n}$ satisfying $x^*_I > y^*_I = 0$ and $y^*_{I^c} > x^*_{I^c} = 0$, and let $q := y^* - Mx^*$. Then, for $F(x) := Mx + q$, we have that $(x^*, y^*) \in S$ while, as $\mu \to 0$, $H_\mu(x^*, y^*) \to 0$ and, by Corollary 2(b), either $H'_\mu(x^*, y^*)$ is singular or $\|H'_\mu(x^*, y^*)^{-1}\| \to \infty$.

To ensure that the iterates z^t generated by Algorithm 1 remain bounded, we need to show that $(z,\mu) \in \mathcal{N}_\beta$ is bounded whenever μ is bounded. It is known that this holds if F is an R_0-function [7, Proposition 4] (for earlier results, see [4, Lemma 3.8], [21, Theorem 3.9], [20, Theorem 3.8] in the case of g given by (5), and see [9, Proposition 3.2] in the case where F is strongly monotone and Lipschitz continuous) or if F is monotone and β and μ are sufficiently small in some sense [2, Proposition 8] (see [38, Lemma 2.4] for a similar result in the case F is monotone affine and g is given by (5)). An example by Sun cited in [38, Example 2.5] ($n = 1$, $F(x) = 1$, $\beta > 1$, g given by (5)) suggests that, for monotone F, β and/or μ being sufficiently small is also necessary (see [23] for an analogous example, where boundedness of

$\{x \in \Re^n : \|H_0(x, F(x))\| \leq \gamma\}$ requires γ to be sufficiently small). The analysis in [2] assumes g is twice continuously differentiable and g'' is an even function and $g''(\tau)$ decays to zero faster in some sense than $1/\tau^2$ as $|\tau| \to \infty$. We show in the lemma below that this assumption can be removed and the threshold on β and μ can be more simply estimated.

Lemma 4 *Assume F is monotone, i.e., $(x-y)^T(F(x)-F(y)) \geq 0$ for all $x, y \in \Re^n$. Also, assume there exists $(\bar{x}, \bar{y}) \in \Re^{2n}$ satisfying $\bar{x} > 0, \bar{y} > 0, \bar{y} = F(\bar{x})$. Then, for any $\beta \in \Re_{++}$ and $\mu^0 \in \Re_{++}$ satisfying $2\beta\mu^0 < \min_{i \in N} \bar{y}_i$ and $\beta\mu^0 < \min_{i \in N} \bar{x}_i$ and any $g \in \mathcal{CM}$, the set $\{(z, \mu) \in \mathcal{N}_\beta : 0 < \mu \leq \mu^0\}$ is bounded.*

Proof. Fix any $\epsilon \in \Re_{++}$ satisfying $(2\beta + \epsilon)\mu^0 < \min_i \bar{y}_i$ and $(\beta + \epsilon)\mu^0 < \min_i \bar{x}_i$. Since $g(\tau)$ and $g(\tau) - \tau$ monotonically decrease to 0 as $\tau \to -\infty$ and $\tau \to \infty$ respectively, there exist unique τ_1 and τ_2 satisfying $g(\tau_1) = \epsilon$ and $g(\tau_2) - \tau_2 = \epsilon$ with $0 < g(\tau) < \epsilon$ for all $\tau < \tau_1$ and $0 < g(\tau) - \tau < \epsilon$ for all $\tau > \tau_2$. Consider any $a, b \in \Re$ satisfying $|a - g(a-b)| \leq \beta$. If $a - b < \tau_1$ so that $|g(a-b)| < \epsilon$, then we have $|a| < \beta + \epsilon$. If $a - b > \tau_2$ so that $|a - b - g(a-b)| < \epsilon$, then we have $|b| < \beta + \epsilon$. Otherwise $\tau_1 \leq a - b \leq \tau_2$, implying $|a| \leq |a - g(a-b)| + |g(a-b)| \leq \beta + g(\tau_2)$ and $|b| \leq |a - g(a-b)| + |a - b - g(a-b)| \leq \beta + g(\tau_1) - \tau_1$. Also, $|a - g(a-b)| \leq \beta$ implies (since $g(a-b) > 0$) $a > -\beta$ and (since $g(a-b) - (a-b) > 0$) $b > -\beta$.

Fix any $(x, y, \mu) \in \mathcal{N}_\beta$ with $0 < \mu \leq \mu^0$. The monotonicity of F implies

$$\begin{aligned} 0 &\leq (\bar{x} - x)^T(F(\bar{x}) - F(x)) \\ &= (\bar{x} - x)^T(\bar{y} - y + y - F(x)) \\ &= \bar{x}^T\bar{y} - x^T\bar{y} - \bar{x}^Ty + x^Ty + (\bar{x} - x)^T(y - F(x)) \\ &\leq \bar{x}^T\bar{y} - x^T\bar{y} - \bar{x}^Ty + x^Ty + \|\bar{x} - x\|_1\beta\mu, \end{aligned}$$

where the last inequality uses (4) and (7). Thus,

$$x^T\bar{y} + \bar{x}^Ty - x^Ty - \|\bar{x} - x\|_1\beta\mu \leq \bar{x}^T\bar{y}. \tag{16}$$

Let $I := \{i \in N : (x_i - y_i)/\mu > \tau_2\}$, $J := \{i \in N : (x_i - y_i)/\mu < \tau_1\}$. Since $(x, y, \mu) \in \mathcal{N}_\beta$ so that $|x_i/\mu - g((x_i - y_i)/\mu)| \leq \beta$ for all $i \in N$, it follows from the above facts that $|y_i|/\mu < \beta + \epsilon$ for $i \in I$, $|x_i|/\mu < \beta + \epsilon$ for $i \in J$, and $|x_i|/\mu \leq \beta + g(\tau_2)$, $|y_i|/\mu \leq \beta + g(\tau_1) - \tau_1$ for $i \in (I \cup J)^c$. This, together with (16) and $\mu \leq \mu^0$, implies

$$\sum_{i \in I} x_i\bar{y}_i + \sum_{i \in J} \bar{x}_iy_i - \sum_{i \in I} x_iy_i - \sum_{i \in J} x_iy_i - \sum_{i \in I} |\bar{x}_i - x_i|\beta\mu^0 \leq \eta$$

for some constant η. Rewriting this as

$$\sum_{i \in I} x_i(\bar{y}_i - y_i - \beta\mu^0) + \sum_{i \in J} (\bar{x}_i - x_i)y_i \leq \eta$$

and noting that, by the above facts and our choice of ϵ, $\bar{y}_i - y_i - \beta\mu^0 > \bar{y}_i - (\beta + \epsilon)\mu^0 - \beta\mu^0 > 0$ and $x_i > -\beta$ for all $i \in I$ and $\bar{x}_i - x_i > \bar{x}_i - (\beta + \epsilon)\mu^0 > 0$ and

$y_i > -\beta$ for all $i \in J$, we see that x_i, $i \in I$, and y_i, $i \in J$, are bounded above by some constant. ■

As was noted by one referee, Lemma 4 is related to [11, Prop. 2.4]. In particular, Lemma 4 equivalently says that the set $\{(x,y) \in \Re^{2n} : \|H_\mu(x,y)\| \le \gamma, 0 < \mu \le \mu^0\}$ is bounded for any $\mu^0 > 0$ and any $0 < \gamma < \sup_{\bar{x}>0, F(\bar{x})>0} \min_{i\in N}\{\bar{x}_i, F_i(\bar{x})/2\}$, and Proposition 2.4 in [11] equivalently says that the set $\{x \in \Re^n : \|H_0(x, F(x))\| \le \gamma\}$ is bounded for any $0 < \gamma < \sup_{\bar{x}>0, F(\bar{x})>0} \min_{i\in N}\{\bar{x}_i, F_i(\bar{x})\}$. Thus the two results are analogous to each other, with one for $H_\mu(x,y)$ $(\mu > 0)$ and the other for $H_0(x, F(x)) = \min\{x, F(x)\}$.

It is readily seen from the proof that Lemma 4 can be refined to show that, under its assumptions, the set

$$\{(x,y,\mu) \in \Re^{2n} \times (0,\mu^0] : \|x - \mu G((x-y)/\mu)\|_\infty \le \beta_1\mu, \|F(x) - y\|_\infty \le \beta_2\mu\}, \quad (17)$$

where $\beta_1, \beta_2, \mu^0 \in \Re_{++}$ and $g \in \mathcal{CM}$, is bounded whenever $(\beta_1 + \beta_2)\mu^0 < \min_{i\in N} \bar{y}_i$ and $\beta_1 < \min_{i\in N} \bar{x}_i$. In the case where F is affine, say $F(x) = Mx + q$ for some $M \in \Re^{n\times n}$ and $q \in \Re^n$, this sufficient condition is also close to being necessary. In particular, assume there exists $(\bar{x}, \bar{y}) \in \Re^{2n}$ satisfying $\min_{i\in N} |\bar{x}_i| > 0, \min_{i\in N} |\bar{y}_i| > 0, \bar{y} = M\bar{x} + q$ and assume M is not an R_0-matrix (otherwise the set (17) is bounded for any $\beta_1, \beta_2, \mu^0 \in \Re_{++}$), so there exists nonzero $(u,v) \in \Re^{2n}$ satisfying $u \ge 0, v = Mu \ge 0, u^T v = 0$. Let $I := \{i \in N : u_i > 0\}$ and $J := \{i \in N : v_i > 0\}$. Then, for any $\beta_1, \beta_2, \mu^0 \in \Re_{++}$ satisfying $\beta_1 > g(0)$ and

$$\beta_1\mu^0 > \max\left\{\|\bar{y}_I\|_\infty, \|\bar{x}_J\|_\infty\right\}, \quad (\beta_1 - g(0))\mu^0 \ge \max_{i\in(I\cup J)^c} \left|\min\{\bar{x}_i, \bar{y}_i\}\right|,$$

the set (17) is not bounded. To see this, for each $\zeta \in \Re_{++}$, let $x := \bar{x} + \zeta u$ and $y := \bar{y} + \zeta v$. Then $F(x) - y = 0$ and, by $\lim_{\tau\to\infty} g(\tau) - \tau = 0$, we have for all $i \in I$ that

$$\left|\frac{x_i}{\mu^0} - g\left(\frac{x_i - y_i}{\mu^0}\right)\right| = \left|\frac{\bar{x}_i + \zeta u_i}{\mu^0} - g\left(\frac{\bar{x}_i + \zeta u_i}{\mu^0} - \frac{\bar{y}_i}{\mu^0}\right)\right| \to \left|\frac{\bar{y}_i}{\mu^0}\right| < \beta_1 \text{ as } \zeta \to \infty$$

and, by $\lim_{\tau\to-\infty} g(\tau) = 0$, we have for all $i \in J$ that

$$\left|\frac{x_i}{\mu^0} - g\left(\frac{x_i - y_i}{\mu^0}\right)\right| = \left|\frac{\bar{x}_i}{\mu^0} - g\left(\frac{\bar{x}_i - \bar{y}_i - \zeta v_i}{\mu^0}\right)\right| \to \left|\frac{\bar{x}_i}{\mu^0}\right| < \beta_1 \text{ as } \zeta \to \infty$$

and for all $i \in (I \cup J)^c$ with $\bar{x}_i \ge \bar{y}_i$ that

$$\left|\frac{x_i}{\mu^0} - g\left(\frac{x_i - y_i}{\mu^0}\right)\right| = \left|\frac{\bar{x}_i}{\mu^0} - g\left(\frac{\bar{x}_i - \bar{y}_i}{\mu^0}\right)\right| \le \left|\frac{\bar{y}_i}{\mu^0}\right| + \left|\frac{\bar{x}_i - \bar{y}_i}{\mu^0} - g\left(\frac{\bar{x}_i - \bar{y}_i}{\mu^0}\right)\right| \le \left|\frac{\bar{y}_i}{\mu^0}\right| + g(0) \le \beta_1$$

and for all $i \in (I \cup J)^c$ with $\bar{x}_i < \bar{y}_i$ that

$$\left|\frac{x_i}{\mu^0} - g\left(\frac{x_i - y_i}{\mu^0}\right)\right| = \left|\frac{\bar{x}_i}{\mu} - g\left(\frac{\bar{x}_i - \bar{y}_i}{\mu^0}\right)\right| \le \left|\frac{\bar{x}_i}{\mu^0}\right| + \left|g\left(\frac{\bar{x}_i - \bar{y}_i}{\mu^0}\right)\right| \le \left|\frac{\bar{x}_i}{\mu^0}\right| + g(0) \le \beta_1.$$

[The above argument does not depend on F being monotone.]

4 Local Superlinear Convergence Analysis

In this section we consider a special choice of $\hat{z}^t$ 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 non-interior continuation methods either require solving an extra $n \times n$ system of linear equations with different matrix coefficients [2, 3, 11] or assume the limit point is strictly complementary and nondegenerate [2, 3, 7]. The choice of $\hat{z}^t$ we consider is one that satisfies the following reduced system of linear equations:

$$H'_{\mu^t}(z^t)_{K^t K^t}(\hat{z}^t_{K^t} - z^t_{K^t}) = -H_0(z^t_{K^t}, 0)_{K^t}, \qquad \hat{z}^t_{(K^t)^c} = 0, \tag{18}$$

where

$$I^t := \{i \in N : x^t_i > \sqrt{\mu^t}\}, \quad J^t := \{i \in N : y^t_i > \sqrt{\mu^t}\}, \quad K^t := I^t \cup J^t. \tag{19}$$

[Here, 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 $n + K$, and by z_K the vector obtained by removing all components not indexed by K nor $n + K$.] The index sets I^t, J^t, K^t estimate, respectively,

$$I^* := \{i \in N : x^*_i > 0, y^*_i = 0\}, \; J^* := \{i \in N : x^*_i = 0, y^*_i > 0\}, \; K^* := I^* \cup J^*, \tag{20}$$

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 μ^t. The choice of $\hat{z}^t$ given by (18) may be viewed as a generalization of that used in [2, 3, 7, 10] which corresponds to $K^t = N$. The idea of setting the smoothing parameter μ to zero on the right-hand side to achieve superlinear convergence can be traced to the interior predictor-corrector methods. Lastly, instead of $\sqrt{\mu}$, any sublinear function of μ can be used in (19).

An important feature of $\hat{z}^t$ given by (18)–(19) is that it can be computed with little extra effort. For example, if we compute w^t given by (8) 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 $\hat{z}^t$ 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 $\hat{z}^t$ in this manner is in the order of $|K^t|^2$. Alternatively, we can compute the inverse of $H'_{\mu^t}(z^t)_{K^t K^t}$, represented in some suitable form, and use this inverse to compute the inverse of $H'_{\mu^t}(z^t)$ using an appropriate formula such as ones given in [13, page 76], [17, page 18].

We will show that if $\hat{z}^t$ is chosen according to (18)–(19) and z^t is sufficiently near an $z^* = (x^*, y^*) \in S$ that is nondegenerate in the sense that $F'(x^*)_{I^* I^*}$ is nonsingular and

$$\Big\{F'(x^*)_{II} u_I = 0, F'(x^*)_{(I^c \setminus J^*)I} u_I \geq 0, u_{I \setminus I^*} \geq 0 \implies u_I = 0\Big\} \quad \forall I^* \subset I \subset (J^*)^c, \tag{21}$$

where I^*, J^*, K^* are given by (20), then $z^t \to z^*$ and $\mu^t \to 0$ superlinearly (and "almost" quadratically if F and g are sufficiently smooth). Thus, our condition for local superlinear convergence is less restrictive than the widely used condition of z^* being a strictly complementary nondegenerate solution [2, 3, 7, 35, 36], which further requires $K^* = N$. Condition (21) is also less restrictive than the nonsingularity condition used in [2, Proposition 4], [3, 11] for hybrid non-interior methods, which further requires $F'(x^*)_{II}$ to be nonsingular for all $I^* \subset I \subset (J^*)^c$. In particular, for the example

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

we have $I^* = J^* = \emptyset$ and $F'(x) = M$ is a P_0-matrix for all $x \in \Re^2$. Clearly $M_{I^*I^*}$ is nonsingular and it can be seen that (21) holds. However, M_{II} is singular for $I = \{1\}$.

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 (20) implicitly. [The index sets (20) 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 (21) is in the order of μ, for all $(z, \mu) \in \mathcal{N}_\beta$ near $(z^*, 0)$. This local error bound is reminiscent of one used in the superlinear convergence analysis of an interior path-following method for monotone LCP [34, Lemma 4.1].

Lemma 5 *Fix any $\beta \in \Re_{++}$, $g \in \mathcal{CM}$, and $z^* = (x^*, y^*) \in S$ satisfying (21), where I^*, J^*, K^* are given by (20). There exists $\delta_1 \in \Re_{++}$ such that for all $(z, \mu) = (x, y, \mu) \in \mathcal{N}_\beta$ satisfying $\|z - z^*\| + \mu \le \delta_1$, we have $\|z - z^*\| \le \mu/\delta_1$.*

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 $\|z - z^*\| > \mu/\delta$. Since $(z, \mu) \in \mathcal{N}_\beta$, we have from (4) and (7) that $\beta \ge |x_i/\mu - g((x_i - y_i)/\mu)|$ for all $i \in N$. Thus, for each $i \in I := \{i \in N : x_i \ge y_i\}$,

$$\beta \ge |y_i/\mu| - |(x_i - y_i)/\mu - g((x_i - y_i)/\mu)| \ge |y_i/\mu| - g(0), \tag{22}$$

where the last inequality follows from $0 < g(\tau) - \tau \le g(0)$ for all $\tau \ge 0$. Similarly, for each $i \in J := \{i \in N : x_i < y_i\} = I^c$,

$$\beta \ge |x_i/\mu| - |g((x_i - y_i)/\mu)| \ge |x_i/\mu| - g(0), \tag{23}$$

where the last inequality follows from $0 < g(\tau) < g(0)$ for all $\tau < 0$. Also, we have $\|F(x) - y\| \le \beta\mu$ and $F(x^*) - y^* = 0$, implying

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

where we denote $M := \int_0^1 F'(x^*+\tau(x-x^*))d\tau$. As $\delta \to 0$, we obtain from $\|z-z^*\|+\mu \leq \delta$ and $\|z-z^*\| > \mu/\delta$ 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 $x_i - x_i^* \geq y_i - y_i^*$ for all $i \in I \backslash I^*$, $x_i - x_i^* < y_i - y_i^*$ for all $i \in J \backslash J^*$) and $\mu/\|z-z^*\| \to 0$ (so that, by (22), $y_I/\|z-z^*\| \to 0$ and, by (23), $x_J/\|z-z^*\| \to 0$). By passing to a subsequence if necessary, we assume I and J are fixed. Then, the above inequality yields in the limit

$$0 \geq \left\| \begin{bmatrix} F'(x^*)_{II} & 0 \\ F'(x^*)_{JI} & -I \end{bmatrix} \begin{bmatrix} u_I \\ v_J \end{bmatrix} \right\|, \quad v_I = 0,\ u_J = 0,\ u_{I \backslash I^*} \geq v_{I \backslash I^*},\ u_{J \backslash J^*} \leq v_{J \backslash J^*}, \quad (24)$$

where (u,v) is any cluster point of $(x-x^*, y-y^*)/\|z-z^*\|$. Thus, (24) yields $F'(x^*)_{II} u_I = 0$ and $F'(x^*)_{(J \backslash J^*)I} u_I = v_{J \backslash J^*} \geq 0$ and $u_{I \backslash I^*} \geq 0$. Moreover, we have $(u,v) \neq 0$, so either $u_I \neq 0$ or $v_J \neq 0$. Since, by (24), $F'(x^*)_{JI} u_I = v_J$ so that $v_J \neq 0$ implies $u_I \neq 0$, it must be that $u_I \neq 0$. Since $J = I^c$, this contradicts (21). ■

In the case where F is affine, i.e., $F(x) = Mx + q$ for some $M \in \Re^{n \times n}$ and $q \in \Re^n$, and $g(0) \leq \beta/\sqrt{n}$, the converse of Lemma 5 also holds. In particular, for any $z^* = (x^*, y^*) \in S$ not satisfying (21), so that there exist $I^* \subset I \subset (J^*)^c$ and nonzero $u_I \in \Re^{|I|}$ satisfying $M_{II} u_I = 0$ and $M_{(I^c \backslash J^*)I} u_I \geq 0$ and $u_{I \backslash I^*} \geq 0$, by letting $z := (x,y)$ with

$$x_I := x_I^* + \sqrt{\mu} u_I, \quad x_{I^c} := x_{I^c}^*, \quad y_I := y_I^*, \quad y_{I^c} := y_{I^c}^* + \sqrt{\mu} M_{I^c I} u_I,$$

we obtain $F(x) - y = 0$ and, as $\mu \to 0$, that $\|z-z^*\| \to 0$, $\|z-z^*\|/\mu \to \infty$ and $\|x/\mu - G((x-y)/\mu)\| \leq \beta$. Thus, for any $\delta > 0$, we have $(z,\mu) \in \mathcal{N}_\beta$ and $\|z-z^*\| + \mu \leq \delta$ but $\|z-z^*\| > \mu/\delta$, for all μ sufficiently small. [Instead of $\sqrt{\mu}$, we can use any function of μ that tends to 0 slower than μ, thus showing that $\|z-z^*\|$ cannot be bounded by any positive power of μ either.]

The next lemma shows that $\hat{z}^t$ given by (18) approaches an $z^* \in S$ superlinearly provided that $\|z^t - z^*\|$ is in the order of μ^t and $K^t = K^*$ (so the degenerate components are identified), and $F'(x^*)_{I^* I^*}$ is nonsingular, where I^* and K^* are given by (20).

Lemma 6 *Fix any $\beta \in \Re_{++}$, $g \in \mathcal{CM}$, 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 $\|z-z^*\| + \mu \leq \delta_1$ we have $\|z-z^*\| \leq \mu/\delta_1$, where I^*, J^*, K^* are given by (20). There exist $\delta_2 \in \Re_{++}$ and $\chi : \Re_{++} \mapsto \Re_{++}$ with $\lim_{\mu \to 0} \chi(\mu)/\mu = 0$ such that for all $(z,\mu) \in \mathcal{N}_\beta$ with $\|z-z^*\| + \mu \leq \delta_2$, we have that $H'_\mu(z)_{KK}$ is nonsingular and $\|\hat{z} - z^*\| \leq \chi(\mu)$, where $\hat{z}$ is given by*

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

with $K = K^*$. If in addition $\limsup_{\tau \to \infty} \tau(1-g'(\tau)) < \infty$, $\liminf_{\tau \to -\infty} \tau g'(\tau) > -\infty$, and there exists $\lambda \in \Re_{++}$ such that $\|F(x+u) - F(x) - F'(x)u\| \leq \lambda \|u\|^2$ whenever $\|x-x^*\|$ and $\|u\|$ are sufficiently small, then we can further choose χ such that $\limsup_{\mu \to 0} \chi(\mu)/\mu^2 < \infty$.

Proof. For simplicity, denote $I := I^*, J := J^*$ (so $K = I \cup J$). Consider the continuously differentiable mapping $\hat{H} : \Re^{2|K|} \mapsto \Re^{2|K|}$ given by

$$\hat{H}(z_K) = \hat{H}(x_K, y_K) := (y_I, x_J, F(x_K, 0)_K - y_K).$$

By taking δ_1 smaller if necessary, we can further assume that $x_I - y_I > x_I^*/2$ and $x_J - y_J < -y_J^*/2$ whenever $\|z - z^*\| \leq \delta_1$, in which case $\hat{H}(z_K) = H_0(z_K, 0)_K$. Now, for all $(z, \mu) \in \mathcal{N}_\beta$ satisfying $\|z - z^*\| + \mu \leq \delta_1$ we have

$$H'_\mu(z)_{KK} - \hat{H}'(z_K) = \begin{bmatrix} I - E_{II} & 0 & E_{II} & 0 \\ 0 & I - E_{JJ} & 0 & E_{JJ} \\ M_{II} & M_{IJ} & -I & 0 \\ M_{JI} & M_{JJ} & 0 & -I \end{bmatrix} - \begin{bmatrix} 0 & 0 & I & 0 \\ 0 & I & 0 & 0 \\ \hat{M}_{II} & \hat{M}_{IJ} & -I & 0 \\ \hat{M}_{JI} & \hat{M}_{JJ} & 0 & -I \end{bmatrix}, \tag{26}$$

where $E = G'((x - y)/\mu)$, $M = F'(x)$ and $\hat{M} = F'(x_K, 0)$. Moreover, for $i \in I$, we have $0 < 1 - E_{ii} = 1 - g'((x_i - y_i)/\mu) < 1 - g'(x_i^*/(2\mu)) \to 0$ as $\mu \to 0$. Similarly, for $i \in J$, we have $0 < E_{ii} = g'((x_i - y_i)/\mu) < g'(-y_i^*/(2\mu)) \to 0$ as $\mu \to 0$. Also, as $\mu \to 0$ we have $x_{K^c} \to x^*_{K^c} = 0$ (since $\|z - z^*\| \leq \mu/\delta_1$) so that $\|M_{II} - \hat{M}_{II}\| = \|F'(x)_{II} - F'(x_K, 0)_{II}\| \to 0$. A similar argument shows that $\|M_{IJ} - \hat{M}_{IJ}\|$, $\|M_{JI} - \hat{M}_{JI}\|$, and $\|M_{JJ} - \hat{M}_{JJ}\|$ also tend to zero. Thus, the matrix difference (26) tends to zero entrywise as $\mu \to 0$. Since $\hat{H}(z_K^*) = 0$ and $\hat{H}(z_K) = H_0(z_K, 0)_K$, (25) implies

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

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

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

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 $\|\hat{z}_K - z_K^*\|/\mu \to 0$ uniformly in z as $\mu \to 0$. Also $\hat{z}_{K^c} = z^*_{K^c} = 0$ trivially.

If in addition g and F have the stated properties, then we obtain that $\|\hat{r}(z_K)\| = O(\|z_K - z_K^*\|^2) = O(\mu^2)$, $1 - E_{ii} = O(\mu)$ for all $i \in I$, $E_{ii} = O(\mu)$ for all $i \in J$, and $\|M_{KK} - \hat{M}_{KK}\| = O(\mu)$ for all z sufficiently near z^* and all $\mu \in \Re_{++}$ sufficiently small, from which we deduce that $\|\hat{z}_K - z_K^*\| = O(\mu^2)$. [Here $b = O(a)$ means $b \leq \kappa a$ for some constant $\kappa \in \Re_{++}$ independent of (z, μ).] ∎

We note that g and F satisfy the additional assumptions in Lemma 6 when g is chosen by (5) or (6) and F is twice continuously differentiable on $\Re^n$. Also, the fact that (26) goes to zero entrywise superlinearly in $\|z_K - z_K^*\|$, which we used in the proof, is known in the case $K = N$ and g is twice continuously differentiable and satisfies $\int_{-\infty}^{\infty} \tau g''(\tau) d\tau < \infty$ [2, 7, 10]. The next lemma shows that if $\|z - z^*\|$ is in the order of μ and coincides with z^* in the degenerate components, then $(z, \mu) \in \mathcal{N}_\beta$.

Lemma 7 *Fix any $\beta \in \Re_{++}$ and $g \in \mathcal{CM}$ with $g(0) \leq \beta/\sqrt{2n}$, and any $z^* = (x^*, y^*) \in S$. Let I^*, J^*, K^* be given by (20) and $\kappa := .9/\max\{\sqrt{2n}, \sqrt{2}(\|F'(x^*)\| + 1)\}$. There exists $\delta_3 \in \Re_{++}$ such that for all $z = (x, y) \in \Re^{2n}$ and $\mu \in (0, \delta_3]$ satisfying $\|z - z^*\| \leq \kappa\beta\mu$ and $x_{(K^*)^c} = y_{(K^*)^c} = 0$, we have $(z, \mu) \in \mathcal{N}_\beta$.*

Proof. Let $\hat{\beta} := \beta/\sqrt{2n}$. First, since $g(\tau) - \tau \to 0$ monotonically as $\tau \to \infty$, there exists $\tau_1 \in \Re$ such that $0 < g(\tau) - \tau \leq .1\hat{\beta}$ for all $\tau \geq \tau_1 - .9\hat{\beta}$. Then, for any $a \geq \tau_1$ and $|b| \leq .9\hat{\beta}$, we have $0 < g(a-b) - (a-b) \leq .1\hat{\beta}$, implying in the case $b > 0$ that $-.9\hat{\beta} \leq -b < g(a-b) - a \leq .1\hat{\beta} - b \leq .1\hat{\beta}$ and in the case $b \leq 0$ that $0 \leq -b \leq g(a-b) - a \leq .1\hat{\beta} - b \leq \hat{\beta}$. Thus, in either case $|a - g(a-b)| \leq \hat{\beta}$. Second, since $g(\tau) \to 0$ monotonically as $\tau \to -\infty$, there exists $\tau_2 \in \Re$ such that $0 < g(-\tau) \leq .1\hat{\beta}$ for all $\tau \geq \tau_2 - .9\hat{\beta}$. Then, for any $b \geq \tau_2$ and $|a| \leq .9\hat{\beta}$, we have $0 < g(a-b) \leq .1\hat{\beta}$, implying in the case $a > 0$ that $-.9\hat{\beta} \leq -a < g(a-b) - a \leq .1\hat{\beta} - a \leq .1\hat{\beta}$ and in the case $a \leq 0$ that $0 \leq -a < g(a-b) - a \leq .1\hat{\beta} - a \leq \hat{\beta}$. Thus, in either case $|a - g(a-b)| \leq \hat{\beta}$.

Choose any $\delta_3 \in \Re_{++}$ sufficiently small so that for all $z = (x, y) \in \Re^{2n}$ and $\mu \in (0, \delta_3]$ satisfying $\|z - z^*\| \leq \kappa\beta\mu$, we have $x_i/\mu \geq \tau_1$ for all $i \in I^*$ and $y_i/\mu \geq \tau_2$ for all $i \in J^*$. Then for $i \in I^*$, we have $x_i/\mu \geq \tau_1$ and $|y_i/\mu| = |y_i - y_i^*|/\mu \leq \kappa\beta \leq .9\hat{\beta}$ so the first fact above yields

$$|x_i/\mu - g((x_i - y_i)/\mu)| \leq \hat{\beta},$$

and, for $i \in J^*$, we have $y_i/\mu \geq \tau_2$ and $|x_i/\mu| = |x_i - x_i^*|/\mu \leq \kappa\beta \leq .9\hat{\beta}$ so the second fact above yields the same. Suppose in addition that $x_{(K^*)^c} = y_{(K^*)^c} = 0$. Then for $i \in (K^*)^c$, we have $|x_i/\mu - g((x_i - y_i)/\mu)| = |g(0)| \leq \hat{\beta}$. Thus $\|x - G((x-y)/\mu)\| \leq \sqrt{n}\hat{\beta} = \beta/\sqrt{2}$. Finally, we have

$$\begin{aligned}
\|F(x) - y\| &= \|F(x) - F(x^*) + y^* - y\| \\
&= \|M(x - x^*) + y^* - y\| \\
&\leq \|M\|\|x - x^*\| + \|y^* - y\| \\
&\leq (\|M\| + 1)\kappa\beta\mu \\
&\leq .9(\|M\| + 1)\beta\mu/(\sqrt{2}(\|F'(x^*)\| + 1)),
\end{aligned}$$

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 δ_3 smaller if necessary, we can assume that $.9(\|M\| + 1) \leq \|F'(x^*)\| + 1$. Thus $\|F(x) - y\|/\mu \leq \beta/\sqrt{2}$. Then, by (4), $\|H_\mu(z)\|/\mu \leq \beta$. ∎

By combining Lemmas 5, 6 and 7, we obtain the following local superlinear convergence result for Algorithm 1 with $\hat{z}^t$ chosen according to (18)–(19).

Proposition 2 *Fix any $\beta \in \Re_{++}$ and $g \in \mathcal{CM}$ with $g(0) \leq \beta/\sqrt{2n}$, and any $z^* = (x^*, y^*) \in S$ such that $F'(x^*)_{I^*I^*}$ is nonsingular and (21) holds, where I^*, J^*, K^* are given by (20). Assume $H'_\mu(z)$ is nonsingular for all $(z, \mu) \in \mathcal{N}_\beta$ with $\mu \leq \mu^0 \in \Re_{++}$. Let $\{(z^t, \mu^t)\}_{t=0,1\ldots}$ be generated by Algorithm 1 with $\hat{z}^t$ chosen by (18)–(19) and π chosen such that $\chi(\mu)/\pi(\mu) \leq \kappa\beta$ for all $\mu \in \Re_{++}$ sufficiently small, where χ is the function given in Lemma 6 (via Lemma 5) and κ is the constant defined in Lemma 7. 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. Then, by (18) and Lemmas 5 and 6, there exists $\delta_2 \in \Re_{++}$ which, together with χ, has the property that

$$\|z^t - z^*\| \leq \mu^t/\delta_2,\ \|\hat{z}^t - z^*\| \leq \chi(\mu^t) \ \text{ whenever } \|z^t - z^*\| + \mu^t \leq \delta_2,\ K^t = K^*. \tag{28}$$

Also, by Lemma 7 and $\hat{z}^t_{(K^t)^c} = 0$, there exists $\delta_3 \in \Re_{++}$ such that

$$(\hat{z}^t, \pi(\mu^t)) \in \mathcal{N}_\beta \ \text{ whenever } \|\hat{z}^t - z^*\| \leq \kappa\beta\pi(\mu^t),\ \pi(\mu^t) \leq \delta_3,\ K^t = K^*.$$

Since $\chi(\mu) \leq \kappa\beta\pi(\mu) \leq \kappa\beta\delta_3$ for all μ sufficiently small, the above two facts imply that there exists $\delta \in (0, \delta_2]$ such that

$$\chi(\mu^t) \leq \kappa\beta\pi(\mu^t),\ (\hat{z}^t, \pi(\mu^t)) \in \mathcal{N}_\beta \ \text{ whenever } \|z^t - z^*\| + \mu^t \leq \delta,\ K^t = K^*, \tag{29}$$

in which case our choice of α^t in Algorithm 1 yields $\alpha^t = 1$, $z^{t+1} = \hat{z}^t$ and $\mu^{t+1} \leq \pi(\mu^t)$. By taking δ smaller if necessary (depending on $\min_{i\in I^*} x^*_i$, $\min_{i\in J^*} y^*_i$ and δ_2 only), we can also assume that whenever $\|z^t - z^*\| + \mu^t \leq \delta$, we have $x^t_i > \sqrt{\mu^t}$ for all $i \in I^*$, $y^t_i > \sqrt{\mu^t}$ for all $i \in J^*$, and from $\|z^t - z^*\| \leq \mu^t/\delta_2$ (see (28)) that $|x^t_i| \leq \mu^t/\delta_2 < \sqrt{\mu^t}$ and $|y^t_i| \leq \mu^t/\delta_2 < \sqrt{\mu^t}$ for all $i \in (K^*)^c$, in which case (19) yields $K^t = K^*$. Then (28) and (29) and $\delta \leq \delta_2$ yield

$$\|z^{t+1} - z^*\| \leq \kappa\beta\pi(\mu^t),\ \mu^{t+1} \leq \pi(\mu^t) \ \text{ whenever } \|z^t - z^*\| + \mu^t \leq \delta. \tag{30}$$

By taking δ smaller if necessary, we can assume that $\max\{\kappa\beta, 1\}\pi(\mu)/\mu < 1/2$ for all $\mu \in (0, \delta]$, so $\kappa\beta\pi(\mu) < \mu/2 \leq \delta/2$. Then, if $\|z^{\bar{t}} - z^*\| \leq \delta/2$ and $\mu^{\bar{t}} \leq \delta/2$, a simple induction argument using (30) yields $\|z^t - z^*\| \leq \kappa\beta\pi(\mu^t) \leq \delta/2$ and $\mu^{t+1} \leq \pi(\mu^t) < \mu^t/2$ for all $t \geq \bar{t}$. 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 $\pi(\mu) \to 0$ as $\mu \to 0$. If g and F satisfy the additional assumptions in Lemma 6 so that χ can be chosen such that $\limsup_{\mu\to 0} \chi(\mu)/\mu^2 < \infty$, 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.

5 Preliminary Computational Experience

To gain some understanding of the numerical behavior/performance of Algorithm 1 with $\hat{z}^t$ 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 set, for a given $x^0 \in \Re^n$, $y^0 = F(x^0)$, $\mu^0 = \|H_0(z^0)\|/4$ and $\beta = 1.5\|H_{\mu^0}(z^0)\|/\mu^0$. Also, we choose $\pi(\mu) = \min\{(1-\sigma)\mu, \mu^{1.8}\}$, $\psi = .7$, and $\sigma = \min\{.1, \hat{\sigma}\}$, where $\hat{\sigma} := \beta/(\beta + \sqrt{n}g(0))$. Finally, we choose $\hat{z}^t$ by (18)–(19) whenever $\mu^t < 0.1$ (otherwise we set $\alpha^t = 0$) and choose $\nu^t = 0$ (corresponding to α^t being either 0 or 1). The above parameter choices, though reasonable, were made without much fine-tuning and can conceivably be improved. To further decrease μ^t, we replace σ by $\hat{\sigma}$ in (10) whenever (9) with this replacement is satisfied by $\alpha = 0$.

In our tests, we choose x^0 to be the vector of 1s multiplied by either .1 or 1 or 10, and we terminate the method whenever $\mu^t \leq 10^{-7}$. We choose g given by either (5) or (6). For g given by (6), due to Matlab treating $g'(\tau)$ as zero for large $|\tau|$, we replace $g(\tau)$ by its linear approximation $g(30) + g'(30)(\tau - 30)$ for $\tau > 30$ and analogously for $\tau < -30$. The performance of the method on a set of six test problems is summarized in Table 1. The test problems are (see [16, 28]):

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

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 $j < i$) for all i and j (so M is positive semidefinite), and every entry of q equals -1.

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 $A \in \Re^{n\times 20}$ 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).

Problem			g of (5)	g of (6)
Name	n	x^0	**iter/nf/nfast/resid**	**iter/nf/nfast/resid**
Kojima-Shindo	4	(.1,...,.1)	$10/16/3/1{\cdot}10^{-20}$	$10/16/3/9{\cdot}10^{-16}$
		(1,...,1)	$9/14/4/4{\cdot}10^{-16}$	$9/17/4/9{\cdot}10^{-30}$
		(10,...,10)	$23/96/4/6{\cdot}10^{-29}$	$26/136/3/9{\cdot}10^{-16}$
Nash-Cournot	10	(.1,...,.1)	$31/68/4/1{\cdot}10^{-14}$	$24/40/4/4{\cdot}10^{-14}$
		(1,...,1)	$24/39/4/4{\cdot}10^{-14}$	$18/28/3/2{\cdot}10^{-14}$
		(10,...,10)	$26/102/4/4{\cdot}10^{-14}$	$25/116/3/1{\cdot}10^{-14}$
Lemke	30	(.1,...,.1)	$8/12/4/9{\cdot}10^{-22}$	$8/12/4/3{\cdot}10^{-41}$
		(1,...,1)	$14/18/4/6{\cdot}10^{-22}$	$11/15/4/8{\cdot}10^{-41}$
		(10,...,10)	$21/25/4/3{\cdot}10^{-21}$	$15/18/3/7{\cdot}10^{-29}$
Harker-Pang	30	(.1,...,.1)	$29/40/2/9{\cdot}10^{-13}$	$28/44/3/1{\cdot}10^{-12}$
		(1,...,1)	$28/40/2/9{\cdot}10^{-13}$	$26/37/3/6{\cdot}10^{-13}$
		(10,...,10)	$25/38/4/1{\cdot}10^{-12}$	$23/36/3/1{\cdot}10^{-12}$
SkewSymLCP	30	(.1,...,.1)	$49/109/0/2{\cdot}10^{-8}$	$47/139/0/6{\cdot}10^{-9}$
		(1,...,1)	$48/115/0/2{\cdot}10^{-8}$	$45/112/0/2{\cdot}10^{-8}$
		(10,...,10)	$43/111/0/2{\cdot}10^{-8}$	$43/156/0/2{\cdot}10^{-8}$
SymPSDLCP	30	(.1,...,.1)	$23/30/2/1{\cdot}10^{-7}$	$27/38/1/1{\cdot}10^{-7}$
		(1,...,1)	$22/29/2/1{\cdot}10^{-7}$	$22/29/2/7{\cdot}10^{-8}$
		(10,...,10)	$17/24/2/1{\cdot}10^{-7}$	$17/24/2/5{\cdot}10^{-8}$

Table 1: Performance of Algorithm 1 on six 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 $\|H_0(x, F(x))\|$ upon termination (resid).

As can be seen from Table 1, the performance varies, depending on the problem and the starting point x^0, but is relatively insensitive to the choice of g given by either (5) or (6). The number of iterations, between 8 and 49, is acceptable for the level of accuracy achieved. In [28], a similar set of test problems were solved using a non-continuation 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.] In general, we found the use of $\hat{z}^t$ 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 caution that the effectiveness of the acceleration depends on the nonsingularity of $F'(x^*)_{I^*I^*}$ and (21). This acceleration mechanism was ineffective (i.e., nfast=0) on the fifth problem, for which (21) likely fails, and the number of iterations is correspondingly higher. The same occurs on the the sixth problem if the rank of M is changed from 20 to 10, with a resulting increase in the number of iterations of at least threefold.

6 Conclusion and Extensions

In this paper we have studied the global (linear) convergence and local superlinear convergence of a non-interior continuation method, based on the Chen-Mangasarian class of smoothing functions, for complementarity problems. The method uses an inexpensive active-set strategy to attain local superlinear convergence under conditions milder than those for other methods.

There is a number of directions in which our work can be extended. One direction is extension to smoothing functions outside the Chen-Mangasarian class, such as one suggested by Kanzow mentioned in Section 1. Another direction is extension to interior-point path-following methods, which we are presently studying. A third direction is a more comprehensive computational study that considers detailed implementation issues such as problem preprocessing, non-monotone line search, and uses a broader set of test problems such as those from the MCPLIB and NETLIB libraries. We hope to study these and other extensions presently or in the future.

References

[1] Burke, J. and Xu, S., The global linear convergence of a non-interior path-following algorithm for linear complementarity problems, Preprint, Department of Mathematics, University of Washington, Seattle, Washington (December 1996); *Math. Oper. Res.*, to appear.

[2] 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).

[3] Chen, B. and Chen, X., A global and local superlinear continuation-smoothing method for $P_0 + R_0$ and monotone NCP, Report, Department of Management and Systems, Washington State University, Pullman, Washington (May 1997); *SIAM J. Optim.*, to appear.

[4] Chen, B. and Harker, P. T., A non-interior-point continuation method for linear complementarity problems, *SIAM J. Matrix Anal. Appl.*, 14 (1993), 1168-1190.

[5] Chen, B. and Harker, P. T., A continuation method for monotone variational inequalities, *Math. Programming*, 69 (1995), 237-253.

[6] Chen, B. and Harker, P. T., Smooth approximations to nonlinear complementarity problems, *SIAM J. Optim.*, 7 (1997), 403-420.

[7] 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); *SIAM J. Optim.*, to appear.

[8] Chen, C. and Mangasarian, O. L., Smoothing methods for convex inequalities and linear complementarity problems, *Math. Programming*, 71 (1995), 51-69.

[9] Chen, C. and Mangasarian, O. L., A class of smoothing functions for nonlinear and mixed complementarity problems, *Comput. Optim. Appl.*, 5 (1996), 97-138.

[10] 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, *Math. Comput.*, 67 (1998), 519-540.

[11] Chen, X. and Ye, Y., On homotopy-smoothing methods for variational inequalities, Report, School of Mathematics, University of New South Wales, Sydney, Australia (December 1996).

[12] 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.

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

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

[15] 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.

[16] 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.

[17] Horn, R. A. and Johnson, C. R., *Matrix Analysis*, Cambridge University Press, Cambridge, United Kingdom, 1985.

[18] Hotta, K. and Yoshise, A., Global convergence of a class of non-interior-point algorithms using Chen-Harker-Kanzow functions for nonlinear complementarity problems, Discussion Paper 708, Institute of Policy and Planning Sciences, University of Tsukuba, Tsukuba, Japan (December 1996).

[19] Jiang, H., Smoothed Fischer-Burmeister equation methods for the complementarity problem, Report, Department of Mathematics, University of Melbourne, Parkville, Australia (June 1997).

[20] Kanzow, C., Some noninterior continuation methods for linear complementarity problems, *SIAM J. Matrix Anal. Appl.*, 17 (1996), 851-868.

[21] Kanzow, C., A new approach to continuation methods for complementarity problems with uniform P-functions, *Oper. Res. Letters*, 20 (1997), 85-92.

[22] Kanzow, C. and Jiang, H., A continuation method for (strongly) monotone variational inequalities, Preprint, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany (October 1994); *Math. Programming*, to appear.

[23] Kanzow, C., Yamashita, N., and Fukushima, M., New NCP-functions and their properties, *J. Optim. Theory Appl.*, 94 (1997), 115-135.

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

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

[26] 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.

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

[28] 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.

[29] 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.

[30] 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.

[31] Qi, L. and Sun, D., Improving the convergence of non-interior point algorithms for nonlinear complementarity problems, Preprint, School of Mathematics, University of New South Wales, Sydney, Australia (May 1997); *Math. Comput.*, to appear.

[32] Qi, L., Sun, D. and Zhou, G., A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities, Preprint, School of Mathematics, University of New South Wales, Sydney, Australia (June 1997).

[33] Smale, S., Algorithms for solving equations, in *Proceedings of the International Congress of Mathematicians*, edited by A. M. Gleason, American Mathematical Society, Providence, Rhode Island (1987), 172-195.

[34] Tseng, P., Simplified analysis of an $O(nL)$-iteration infeasible predictor-corrector path following method for monotone LCP, in *Recent Trends in Optimization Theory and Applications*, edited by R. P. Agarwal, World Scientific, Singapore, 1995, 423-434.

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

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

[37] Xu, S., The global linear convergence an infeasible non-interior path-following algorithm for complementarity problems with uniform P-functions, Preprint, Department of Mathematics, University of Washington, Seattle, Washington (December 1996).

[38] Xu, S., The global linear convergence and complexity of a non-interior path-following algorithm for monotone LCP based on Chen-Harker-Kanzow-Smale smoothing functions, Preprint, Department of Mathematics, University of Washington, Seattle, Washington (February 1997).