Nonlinear Optimization and Applications 2, pp. 1-000
G. Di Pillo and F. Giannessi, Editors
©1998 Kluwer Academic Publishers B.V.

Error Bounds and Superlinear Convergence Analysis of Some Newton-Type Methods in Optimization

Paul Tseng (tseng@math.washington.edu)
Department of Mathematics, Box 354350
University of Washington, Seattle, WA 98195-4350, U.S.A.

Abstract

We show that, for some Newton-type methods such as primal-dual interior-point path following methods and Chen-Mangasarian smoothing methods, local superlinear convergence can be shown without assuming the solutions are isolated. The analysis is based on local error bounds on the distance from the iterates to the solution set.

Keywords: Error bound, superlinear convergence, Newton method, continuation/smoothing method, interior-point method.

This research is supported by National Science Foundation Grant CCR-9731273.

1 Introduction

We consider the complementarity problem (CP) 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 continuous mapping from $\Re^n$ (or $\Re^n_+$) to $\Re^n$ that is continuously differentiable on $\Re^n$ (or $\Re^n_{++}$) [10, 12, 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. An important special case is when F is affine, giving rise to a linear CP (LCP) [11, 20]. Another important special case is when

$$x = \begin{bmatrix} u \\ v \end{bmatrix}, \quad F(x) = \begin{bmatrix} \nabla f_0(u) + \sum_{i=1}^m v_i \nabla f_i(u) \\ -f_i(u)_{i=1}^m \end{bmatrix},$$

where $f_0, f_1, ..., f_m$ are twice continuously differentiable functions defined on $\Re^l$ (or $\Re^l_{++}$), $m \geq 0, l \geq 1$. In this case, (1) is equivalent to the Karush-Kuhn-Tucker condition associated with the nonlinear program

$$\text{minimize } f_0(u) \text{ subject to } u \geq 0,\ f_i(u) \leq 0,\ i = 1, ..., m.$$

Moreover, if $f_0, f_1, ..., f_m$ are convex (respectively, quadratic) on their respective domains, then this F is monotone (respectively, affine) and continuous on $\Re^{l+m}_+$.

Among the various approaches for solving CP, one that has received much attention concerns Newton-type methods based on continuation/smoothing. In this approach, one chooses a continuously differentiable mapping H_μ from $\Re^{2n}$ (or $\Re^{2n}_{++}$) to $\Re^{2n}$, parameterized by $\mu \in \Re_{++}$, having the property that

$$H_\mu(x, y) \to 0,\ (x, y, \mu) \to (\bar{x}, \bar{y}, 0) \qquad \Longrightarrow \qquad (\bar{x}, \bar{y}) \in S.$$

Then, from an initial μ and an initial guess (x, y) of the solution, (x, y) is updated by moving it along a Newton direction for the equaton $H_\mu(x, y) = 0$ and afterwards μ is decreased. Termination occurs when μ is sufficiently small. One choice of the Newton direction is the solution $(u, v) \in \Re^{2n}$ of the Newton equation

$$H'_\mu(x, y) \begin{bmatrix} u \\ v \end{bmatrix} + H_\mu(x, y) = 0, \tag{2}$$

which is useful in achieving global convergence. A second choice, useful for practical efficiency as well as achieving local superlinear convergence, is the solution (u, v) of the modified Newton equation

$$H'_\mu(x, y) \begin{bmatrix} u \\ v \end{bmatrix} + H_0(x, y) = 0, \tag{3}$$

where $H_0(x, y) = \lim_{\mu \to 0} H_\mu(x, y)$, assuming this point-wise limit exists. The two directions (2) and (3) may be used in series (i.e., use (2) to update (x, y) and then use (3) to further update (x, y)) or in parallel (i.e., use a convex combination of (2) and (3) to update (x, y)). Primal-dual interior-point path following methods correspond to H_μ given by

$$H_\mu(x, y) = \begin{bmatrix} (x_i y_i - \mu)_{i=1}^n \\ F(x) - y \end{bmatrix}, \qquad x > 0, y > 0. \tag{4}$$

The use of both directions (2) and (3) give rise to the so-called predictor-corrector methods. The literature on such methods is vast and, in the context of CP, include [15, 18, 19, 20, 21, 25, 28, 29, 30, 31, 33, 34, 39, 42]. The smoothing method of Chen and Mangasarian [7, 8] and its extensions [3, 4, 6, 9, 13, 40] correspond to H_μ given by

$$H_\mu(x, y) = \begin{bmatrix} (x_i - \mu g((x_i - y_i)/\mu))_{i=1}^n \\ F(x) - y \end{bmatrix}, \tag{5}$$

where $g : \Re \to \Re$ is any convex continuously differentiable function with the properties that $g(-\tau) \to 0$ and $g(\tau) - \tau \to 0$ as $\tau \to \infty$ and $0 < g'(\tau) < 1$ for all $\tau \in \Re$. We will denote the class of such functions g by $\mathcal{CM}$. [In the case where g is twice continuously differentiable, $\mu g(\tau/\mu)$ may be written as the integral $\int_{-\infty}^{\tau} \int_{-\infty}^{\alpha} \frac{1}{\mu} g''(\xi/\mu) d\xi d\alpha$, where g'' is interpreted as a probability density function. This is the form presented in [8] and used by others. The form (5) is from [40] and is easier to use for our analysis.] Examples of the smoothing function g include one proposed independently by Chen and Harker [5] , Kanzow [16] , and Smale [32] (see [1, 2, 14, 43] and references therein for methods based on this g):

$$g(\tau) = (\sqrt{\tau^2 + 4} + \tau)/2, \tag{6}$$

and one obtained by integrating the sigmoid function $\tau \mapsto 1/(1 + e^{-\tau})$ used in neural networks [7, 8]:

$$g(\tau) = \ln(e^{\tau} + 1). \tag{7}$$

See [7, 8] for other examples of g. A new example of g that is useful for our analysis is

$$g(\tau) = \begin{cases} (1/2)^{\gamma+2}(1/2 - \tau)^{-\gamma}/\gamma & \text{if } \tau \geq 0 \\ (1/2)^{\gamma+2}(1/2 + \tau)^{-\gamma}/\gamma + \tau & \text{if } \tau < 0 \end{cases}, \tag{8}$$

with $\gamma > 0$. For H_μ given by (5), it is easily seen that $H_0(x, y) = (\min\{x, y\}, F(x) - y)$.

Our focus is on the local superlinear convergence analysis of the preceding Newton-type methods and, for simplicity, we will restrict our attention to H_μ given by (4) and (5). In the typical superlinear convergence analysis, the solutions need to be isolated. An exception to this is the primal-dual interior-point predictor-corrector methods, for which local superlinear convergence can be shown for linear programs and, more generally, monotone LCP having a strictly complementary solution (see [37, 41] and references therein). Further extensions of this analysis to monotone (nonlinear) CP and variational inequalities have been obtained by Sun and Zhao [34], assuming (in addition to existence of strictly complementary solution and some mild assumptions) constant range space and column rank of certain matrices based on $F'(x)$ [34, Assumptions 3-4]. The related analysis of Ralph and Wright [29, 30] allows for nonunique multipliers, but when specialized CP, it still requires uniqueness of solution. Monteiro and Zhou [24] improved the analysis of [29] for the case of linearly constrained convex program, assuming constant null space of certain matrices based on $F'(x)$. This assumption is similar to [34, Assumptions 3-4] and allows for multiple primal solutions. In this paper, we consider a new approach to local superlinear convergence analysis without assuming isolated solutions, based on error bound of the form:

$$\text{dist}(x, X^*) \leq \kappa \| \min\{x, F(x)\} \| \quad \text{whenever } \| \min\{x, F(x)\} \| \leq \delta, \tag{9}$$

where $X^* = \{x : (x, y) \in S \text{ for some } y\}$, and $\kappa > 0, \delta > 0$. In the case of primal-dual interior-point methods for monotone CP, we show that this error bound and

strict complementarity, together with some mild assumptions, imply that the Newton direction (u, v) solving (3) is asymptotically in the order of μ (see Lemmas 1, 2), from which a standard argument yields that (x, y) can be moved along (u, v) with a stepsize approaching 1 and with μ decreased at a quadratic rate (see Prop. 1). The same approach applies to the Chen-Mangasarian smoothing method for monotone LCP (see Lemmas 4, 5 and Prop. 2). This gives a new superlinear convergence result for this method without assuming isolated solutions. Although the error bound (9) does not hold for general F, it is known to hold if (i) F is affine or (ii) $F(x) = A^T G(Ax) + b$, with $A \in \Re^{m\times n}$, $b \in \Re^n$, and $G : \Re^m \mapsto \Re^m$ strongly monotone and Lipschitz continuous [22, 36, 27], or (iii) F is Lipschitz continuous and a uniformly P-function on $\Re^n$ [17, Cor. 7.3], [27, p. 320]. In cases (i) and (ii), the elements of S and X^* need not be isolated. In case (iii), S and X^* are singletons. For a good survey of this error bound and its applications, see [27]. It is an open question how this error bound relates to [34, Assumptions 3-4] or [24, Assumption 4].

This paper is organized as follows: In Sec. 2, we describe the Newton-type method based on continuation/smoothing. In Sec. 3, we analyze the local superlinear convergence of this method with H_μ given by (4), assuming that F is monotone, that (9) holds, and that the iterates asymptotically satisfy strict complementarity. In Sec. 4, we do an analogous analysis for H_μ given by (5) with $g \in \mathcal{CM}$. In Sec. 5, we discuss possible extensions of the results.

In our notation, all vectors are column vectors, ^{T} denotes transpose, $\Re^n$ denotes the space of n–dimensional real column vectors, and $\Re^n_+$ (respectively, $\Re^n_{++}$) denotes the nonnegative (respectively, positive) orthant in $\Re^n$. For convenience, we also write $\begin{bmatrix} x \\ y \end{bmatrix}$ as "(x, y)" for any two vectors x and y. We denote $N = \{1, ..., n\}$. 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 denote by $\|x\|$ and $\|x\|_\infty$ the 2- and ∞-norm of x, by X the diagonal matrix with diagonal entries $x_1, ..., x_n$, and by e a vector of 1s. We denote by $\Re^{n\times n}$ the space of $n \times n$ real matrices. 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 $\|M\| := \max_{x\in\Re^n:\|x\|=1} \|Mx\|$. For any continuously differentiable function $H = (H_1, ..., H_m)^T : \Re^m \mapsto \Re^m$, we denote its Jacobian by $H' = (\nabla H_1, ..., \nabla H_m)^T$, where ∇H_i denotes the gradient of H_i for $i = 1, ..., m$.

2 A Newton-Type Continuation/Smoothing Method

In order to be amenable to a global convergence analysis, continuation/smoothing methods maintain the iterates (x, y, μ) to lie in a "neighborhood" of the solutions to $H_\mu(x, y) = 0$. There are many choices for the neighborhood. For H_μ given by (4), a

reasonable choice is

$$\mathcal{N}_{\beta_1,\beta_2} = \{(x,y,\mu) \in \Re^{2n+1}_{++} : \mu/\beta_1 \leq x_i y_i \leq \beta_1 \mu \ \forall i, \|F(x) - y\| \leq \beta_2 \mu\}, \tag{10}$$

with $\beta_1 > 1, \beta_2 \geq 0$. Alternative choices, discussed in [15, 20, 29, 30, 33, 34, 39, 42], can possibly also be used. For H_μ given by (5), a reasonable choice is

$$\mathcal{N}_{\beta_1,\beta_2} = \{(x,y,\mu) \in \Re^{2n} \times \Re_{++} : |x_i - \mu g((x_i - y_i)/\mu)| \leq \beta_1 \mu \ \forall i, \|F(x) - y\| \leq \beta_2 \mu\}, \tag{11}$$

with $\beta_1 > g(0), \beta_2 \geq 0$. Alternative choices, discussed in [1, 2, 3, 4, 6, 40, 43], can possibly also be used. We describe the general method, parameterized by H_μ and $\mathcal{N}_{\beta_1,\beta_2}$, formally below.

Algorithm 1 *Choose suitable* $\beta^0 \in (0, \beta_1]$ *and* $(x^0, y^0, \mu^0) \in \mathcal{N}_{\beta^0,\beta_2}$. *For* $t = 0, 1, ...,$ *generate* $(x^{t+1}, y^{t+1}, \mu^{t+1})$ *from* (x^t, y^t, μ^t) *according to the following iteration: Let* (u^t, v^t) *be the solution of* (3) *with* $(x, y, \mu) = (x^t, y^t, \mu^t)$. *Let*

$$\theta^t \in \arg\min_{\theta \in [0,1]} m(x^t + \theta u^t, y^t + \theta v^t). \tag{12}$$

where we define $m(x,y) = \min\{\mu \in \Re_+ : (x,y,\mu) \in \mathcal{N}_{\beta_1,\beta_2}\}$, *and let*

$$(\hat{x}^t, \hat{y}^t) = (x^t + \theta^t u^t, y^t + \theta^t v^t), \quad \hat{\mu}^t = m(\hat{x}^t, \hat{y}^t). \tag{13}$$

Generate a $\beta^{t+1} \in (0, \beta_1]$ *and* $(x^{t+1}, y^{t+1}, \mu^{t+1}) \in \mathcal{N}_{\beta^{t+1},\beta_2}$ *satisfying* $\mu^{t+1} \leq \hat{\mu}^t$.

Note 1. The minimization in (12) and the evaluation of $m(\cdot)$ can be done inexactly (and inexpensively) without affecting our superlinear convergence results, as can be seen from Lemmas 3 and 6.

Note 2. We have left unspecified how $(x^{t+1}, y^{t+1}, \mu^{t+1})$ and β^{t+1} are to be generated so that our results may be applied to different continuation/smoothing methods. One popular choice is

$$(x^{t+1}, y^{t+1}, \mu^{t+1}) = (\hat{x}^t + \hat{\theta}^t \hat{u}^t, \hat{y}^t + \hat{\theta}^t \hat{v}^t, (1 - \sigma\hat{\theta}^t)\hat{\mu}^t), \tag{14}$$

where $(\hat{u}^t, \hat{v}^t)$ is the solution of (2) with $(x, y, \mu) = (\hat{x}^t, \hat{y}^t, \hat{\mu}^t)$, and $\hat{\theta}^t$ is the largest element of $\{1, \psi, \psi^2, ...\}$ such that this triple remains in $\mathcal{N}_{\hat{\beta}_1,\beta_2}$, with $\psi \in (0,1)$ and $\sigma \in (0,1), \hat{\beta}_1 \in (0, \beta_1)$ suitably chosen (see [1, 2, 3, 6, 29, 39, 40, 42, 43]). Instead of a single Newton step, one can also take multiple Newton steps.

Note 3. In the context of interior-point methods, i.e., H_μ given by (4), Algorithm 1 is reminiscent of the predictor-corrector method of Mizuno, Todd, and Ye for linear programming [23]. Also, Algorithm 1 is an infeasible method in the sense that it does not maintain $F(x) = y$ satisfied at all iterations. If we assume that an initial $(x^0, y^0, \mu^0) \in \mathcal{N}_{\beta^0,\beta_2}$ satisfying $F(x^0) = y^0$ is available, we can alternatively consider a feasible method whereby $y^t + \theta v^t$ in (12) is replaced by $F(x^t + \theta u^t)$ and analogous changes are made in (13) and (14). Such a feasible method has the advantage that "polynomial convergence" can be shown assuming that F satisfies a scaled Lipschitz condition (see [15, 34, 35] and references therein). However, the assumption of a feasible interior solution limits the practicality of such a method.

3 Superlinear convergence analysis of interior-point method

In this section we analyze the local superlinear convergence of Algorithm 1 with H_μ and $\mathcal{N}_{\beta_1,\beta_2}$ given by (4) and (10), respectively. We begin with two technical lemmas that together show that the Newton direction (u,v) given by (3) is in the order of μ, assuming that F is (locally) monotone and Lipschitz continuous, that (9) holds, and that $x+y$ is bounded away from zero componentwise.

Lemma 1 *Assume F is monotone and F' is Lipschitz continuous with constant $\lambda \geq 0$ on $\Re^n_+$. Then, for any $(x,y,\mu) \in \Re^{2n+1}_{++}$ and any $(x^*,y^*) \in S$, we have*

$$\begin{array}{rl} \|u\| \leq & \|X\|(\lambda\|X\|\|x-x^*\|^2/2 + \|x-x^*\|\|y-y^*\|)/(\min_i x_i y_i) + \|x-x^*\|, \\ \|v\| \leq & \|Y\|(\lambda\|X\|\|x^*-x\|^2/2 + 2\|x-x^*\|\|y-y^*\|)/(\min_i x_i y_i) + \|y-y^*\|, \end{array} \tag{15}$$

where (u,v) solves (3) and H_μ is given by (4).

Proof. Using (4), (3) may be written as

$$F'(x)u - v + F(x) - y = 0, \qquad Yu + Xv = -Xy.$$

Using $F(x^*) - y^* = 0$ and $X^*y^* = 0$, this can be rewritten as

$$\begin{array}{rl} F'(x)(x-x^*+u) - (y-y^*+v) & = F(x^*) - F(x) + F'(x)(x-x^*), \\ Y(x-x^*+u) + X(y-y^*+v) & = (X-X^*)(y-y^*). \end{array} \tag{16}$$

Since $\|F(x^*) - F(x) + F'(x)(x-x^*)\| = \|\int_0^1 (F'(x+t(x^*-x)) - F'(x))(x^*-x)dt\| \leq \int_0^1 \lambda t\|x^*-x\|^2 dt = \lambda\|x^*-x\|^2/2$, multiplying the first equation on the left by $(x-x^*+u)^T$ and using the positive semidefiniteness of $F'(x)$ yields

$$\begin{array}{rcl} -(x-x^*+u)^T(y-y^*+v) & \leq & (x-x^*+u)^T(F(x^*) - F(x) + F'(x)(x-x^*)) \\ & \leq & \lambda\|X\|\|X^{-1}(x-x^*+u)\|\|x^*-x\|^2/2. \end{array}$$

Then, multiplying the second equation on the left by $(x-x^*+u)^T X^{-1}$ and rearranging terms yield

$$\begin{array}{rl} & (\min_i x_i y_i)\|X^{-1}(x-x^*+u)\|^2 \\ \leq & (x-x^*+u)^T Y X^{-1}(x-x^*+u) \\ = & -(x-x^*+u)^T(y-y^*+v) + (x-x^*+u)^T X^{-1}(X-X^*)(y-y^*) \\ \leq & \lambda\|X\|\|X^{-1}(x-x^*+u)\|\|x^*-x\|^2/2 + \|X^{-1}(x-x^*+u)\|\|(X-X^*)(y-y^*)\|. \end{array}$$

Thus

$$(\min_i x_i y_i)\|X^{-1}(x-x^*+u)\| \leq \lambda\|X\|\|x^*-x\|^2/2 + \|(X-X^*)(y-y^*)\|. \tag{17}$$

Since $\|X^{-1}(x-x^*+u)\| \geq \|x-x^*+u\|/\|X\| \geq (\|u\|-\|x-x^*\|)/\|X\|$, this yields the first inequality in (15). Finally, multiplying the second equation in (16) by X^{-1} yields

$$\begin{aligned}\|y-y^*+v\| &= \|X^{-1}(X-X^*)(y-y^*) - YX^{-1}(x-x^*+u)\| \\ &\leq \|(XY)^{-1}\|\|Y\|\|(X-X^*)(y-y^*)\| + \|Y\|\|X^{-1}(x-x^*+u)\| \\ &\leq \|Y\|\|(X-X^*)(y-y^*)\|/(\min_i x_i y_i) + \|Y\|(\lambda\|X\|\|x^*-x\|^2/2 \\ &+ \|(X-X^*)(y-y^*)\|)/(\min_i x_i y_i),\end{aligned}$$

where the last inequality also uses (17). Rearranging terms yields the second inequality in (15). ∎

Lemma 2 *Assume F is Lipschitz continuous on $\Re^n_+$ with constant $L > 0$ and (9) holds for some $\kappa > 0, \delta > 0$. Let H_μ and $\mathcal{N}_{\beta,\beta_2}$ be given by (4) and (10) with $\beta > 0, \beta_2 \geq 0$. Then, for any $(x,y,\mu) \in \mathcal{N}_{\beta,\beta_2}$ satisfying $(x_I, y_J) \geq \rho e$ and $\beta\mu \leq \rho^2, \beta_2\sqrt{n}\mu \leq \delta\rho$, where $\rho > 0$ and $I \subset N, J = N\backslash I$, there exists an $(x^*, y^*) \in S$ satisfying*

$$\|x-x^*\| \leq (\kappa\beta\sqrt{n}/\rho)\mu, \qquad \|y-y^*\| \leq (L\kappa\beta\sqrt{n}/\rho + \beta_2)\mu.$$

Proof. Since $(x,y,\mu) \in \mathcal{N}_{\beta,\beta_2}$ so that $Xy \leq \beta\mu e$, then $(x_I, y_J) \geq \rho e$ implies

$$0 < (y_I, x_J) \leq \beta\mu/\rho e. \tag{18}$$

Then, $(x_I, y_J) \geq \rho e$ implies, since $\beta\mu \leq \rho^2$, that $\min\{(x_I, x_J), (y_I, y_J)\} = (y_I, x_J)$, so if in addition $\beta\sqrt{n}\mu \leq \delta\rho$, then (9) and (18) would imply the existence of $(x^*, y^*) \in S$ such that

$$\|x^*-x\| \leq \kappa\|(y_I, x_J)\| \leq \kappa\beta\sqrt{n}\mu/\rho.$$

We also have

$$\|y^*-y\| = \|F(x^*)-F(x)+F(x)-y\| \leq L\|x^*-x\| + \beta_2\mu \leq (L\kappa\beta\sqrt{n}/\rho + \beta_2)\mu.$$

∎

The next lemma shows that, by moving (x,y) in the direction (u,v) given by (3), it is possible to decrease μ by a factor of $\|(u,v)\|$ while remaining in $\mathcal{N}_{\beta_1,\beta_2}$.

Lemma 3 *Assume F' is Lipschitz continuous on $\Re^n_+$ with constant $\lambda \geq 0$. Let H_μ and $\mathcal{N}_{\beta_1,\beta_2}$ be given by (4) and (10), with $\beta_1 > 0, \beta_2 \geq 0$. Then, for any $(x,y,\mu) \in \mathcal{N}_{\beta,\beta_2}$ with $\beta < \beta_1$ and (u,v) satisfying (3), we have $(x+\theta u, y+\theta v, \mu^+) \in \mathcal{N}_{\beta_1,\beta_2}$ whenever*

$$1-\theta \geq \frac{4/c}{(1+\sqrt{1+4/c})^2}, \tag{19}$$

$$((1-\theta)(\min_i x_i y_i) + \theta^2(\min_i u_i v_i))\beta_1 \geq \mu^+ \geq \max\left\{\begin{array}{c}((1-\theta)\|Xy\|_\infty + \theta^2\|Uv\|_\infty)/\beta_1, \\ (1-\theta)\mu + \theta^2\lambda\|u\|^2/(2\beta_2)\end{array}\right\}, \tag{20}$$

where $c = (\beta_1/\beta - 1)\mu/(\|Uv\|_\infty\beta_1 + \max\{\|Uv\|_\infty/\beta_1, \lambda\|u\|^2/(2\beta_2)\})$.

Proof. By (3) and (4), $Yu + Xv + Xy = 0$, so

$$(X + \theta U)(y + \theta v) = Xy + \theta Yu + \theta Xv + \theta^2 Uv = (1-\theta)Xy + \theta^2 Uv.$$

Then, a sufficient condition for $(\mu^+/\beta_1)e \leq (X+\theta U)(y+\theta v) \leq \beta_1 \mu^+ e$ is

$$\mu^+/\beta_1 \leq (1-\theta)(\min_i x_i y_i) + \theta^2(\min_i u_i v_i), \quad \beta_1 \mu^+ \geq (1-\theta)\|Xy\|_\infty + \theta^2 \|Uv\|_\infty.$$

Also, by (3) and (4), $F(x) + F'(x)u - (y+v) = 0$, so

$$\begin{aligned}
\|F(x+\theta u) - (y+\theta v)\| &= \|(1-\theta)(F(x)-y) + F(x+\theta u) - (F(x) + \theta F'(x)u)\| \\
&= \|(1-\theta)(F(x)-y) + \int_0^1 F'(x+t\theta u) - F'(x))\theta u \; dt\| \\
&\leq (1-\theta)\|F(x)-y\| + \lambda \int_0^1 t\|\theta u\|^2 dt \\
&\leq (1-\theta)\beta_2 \mu + \lambda \|u\|^2 \theta^2/2,
\end{aligned}$$

where the last inequality uses $(x,y,\mu) \in \mathcal{N}_{\beta,\beta_2}$. Thus, $\|F(x+\theta u) - (y+\theta v)\| \leq \beta_2 \mu^+$ whenever $\mu^+ \geq (1-\theta)\mu + \lambda\|u\|^2\theta^2/(2\beta_2)$. Thus $(x+\theta u, y+\theta v, \mu^+) \in \mathcal{N}_{\beta_1,\beta_2}$ whenever μ^+ satisfies (20). For such μ^+ to exist, it is sufficient that

$$((1-\theta)(\min_i x_i y_i) + \theta^2(\min_i u_i v_i))\beta_1 \geq \max \left\{ \begin{array}{c} ((1-\theta)\|Xy\|_\infty + \theta^2\|Uv\|_\infty)/\beta_1, \\ (1-\theta)\mu + \theta^2 \lambda \|u\|^2/(2\beta_2) \end{array} \right\}.$$

Since $\mu/\beta \leq x_i y_i \leq \beta\mu$ for all i and $\beta < \beta_1$, a sufficient condition for this inequality to hold is

$$((1-\theta)\mu/\beta - \theta^2\|Uv\|_\infty)\beta_1 \geq ((1-\theta)\mu + \theta^2 \max\{\|Uv\|_\infty/\beta_1, \lambda\|u\|^2/(2\beta_2)\}.$$

This is a quadratic inequality of the form $(1-\theta)c - \theta^2 \geq 0$. Solving it gives $\theta \leq 2/(1+\sqrt{1+4/c})$ and (19) follows. ∎

Combining Lemmas 1, 2 and 3, we obtain the following main result of this section.

Proposition 1 *Assume F is monotone and Lipschitz continuous on $\Re^n_+$. Also, assume F' is Lipschitz continuous on $\Re^n_+$ and (9) holds for some $\kappa > 0, \delta > 0$. Let H_μ and $\mathcal{N}_{\beta_1,\beta_2}$ be given by (4) and (10). Let $\{(x^t, y^t, \mu^t, u^t, v^t, \hat{\mu}^t, \beta^t)\}_{t=0,1,\ldots}$ be generated by Algorithm 1. Assume $\{(x^t, y^t)\}$ is bounded, $\mu^t \to 0$, $\lim_{t\to\infty} \inf x^t + y^t > 0$, and $\lim_{t\to\infty} \inf \beta_1/\beta^t > 1$. Then, there exist $\bar{t}$ and $C_1 > 0, C_2 > 0$ such that $\|(u^t, v^t)\| \leq C_1 \mu^t$ and $\mu^{t+1} \leq C_2(\mu^t)^2$ for all $t \geq \bar{t}$.*

Proof. Since $\mu^t \to 0$ and $\lim_{t\to\infty} \inf x^t + y^t > 0$, Lemma 2 implies that, for each t exceeding some $\bar{t}$, there is an $(\bar{x}^t, \bar{y}^t) \in S$ with

$$\|x^t - \bar{x}^t\| = O(\mu^t), \qquad \|y^t - \bar{y}^t\| = O(\mu^t).$$

[Here $b = O(a)$ means $b \leq Ca$, with C a constant independent of t.] Then Lemma 1, together with the fact $\min_i x_i^t y_i^t \geq \mu^t/\beta^t \geq \mu^t/\beta_1$, yields

$$\|u^t\| = O(\|X^t\|(\|X^t\|+1)\mu^t + \mu^t), \qquad \|v^t\| = O(\|Y^t\|(\|X^t\|+1)\mu^t + \mu^t).$$

Since $\{(x^t, y^t)\}$ is bounded, this implies $\|(u^t, v^t)\| \leq C_1\mu^t$ for suitable constant $C_1 > 0$. By Lemma 3, we have $(x^t + \tilde{\theta}^t u^t, y^t + \tilde{\theta}^t v^t, \tilde{\mu}^t) \in \mathcal{N}_{\beta_1,\beta_2}$, where

$$\begin{aligned}
c^t &= (\beta_1/\beta^t - 1)\mu^t/(\|U^t v^t\|_\infty \beta_1 + \max\{\|U^t v^t\|_\infty/\beta_1, \lambda\|u^t\|^2/(2\beta_2)\}),\\
\tilde{\theta}^t &= 1 - \frac{4/c^t}{(1+\sqrt{1+4/c^t})^2},\\
\tilde{\mu}^t &= \max\{\,((1-\tilde{\theta}^t)\|X^t y^t\|_\infty + (\tilde{\theta}^t)^2\|U^t v^t\|_\infty)/\beta_1, (1-\tilde{\theta}^t)\mu^t + (\tilde{\theta}^t)^2\lambda\|u^t\|^2/(2\beta_2)\,\}.
\end{aligned}$$

Since $\lim_{t\to\infty} \inf \beta_1/\beta^t > 1$ and $\|(u^t, v^t)\| \leq C_1\mu^t$, we see that $1/c^t = O(\mu^t)$ for t large. Thus, $1 - \tilde{\theta}^t = O(\mu^t)$ and hence

$$\mu^{t+1} \leq \hat{\mu}^t \leq \tilde{\mu}^t = O((\mu^t)^2).$$

■

We remark that the assumptions that $\mu^t \to 0$, $\lim_{t\to\infty} \inf x^t + y^t > 0$, and $\lim_{t\to\infty} \inf \beta_1/\beta^t > 1$ are reasonable and, in particular, hold when $(x^{t+1}, y^{t+1}, \mu^{t+1})$ is generated by (14), assuming in addition F is affine or uniformly nondegenerate [39, Lemma 2.1(c),(d)]. The same reference gives conditions on F that ensure boundedness of $\{(x^t, y^t)\}$. For the feasible variant, as discussed in Note 3, these assumptions hold if F satisfies a scaled Lipschitz condition [15, 34, 35]. Also, Prop. 1 still holds if the assumptions on F hold only locally around the cluster points of $\{x^t\}$.

4 Superlinear convergence analysis of Chen-Mangasarian smoothing method

In this section we analyze the local superlinear convergence of Algorithm 1 with H_μ and $\mathcal{N}_{\beta_1,\beta_2}$ given by (5) and (11), respectively. We begin with two technical lemmas that are analogous to Lemmas 1 and 2.

Lemma 4 *Assume F is monotone and F' is Lipschitz continuous with constant $\lambda \geq 0$ on $\Re^n$. Let H_μ and $\mathcal{N}_{\beta,\beta_2}$ be given by (5) and (11) with $\beta > 0, \beta_2 \geq 0$, and $g \in \mathcal{CM}$. Then, for any $(x, y, \mu) \in \mathcal{N}_{\beta,\beta_2}$ and any $(x^*, y^*) \in S$ with $(x_I, y_J) \geq \rho e$, $(\beta + g(0))\mu < \rho$ and $(y_I^*, x_J^*) = 0$, where $\rho > 0$ and $I \subset N, J = N\backslash I$, we have*

$$\begin{aligned}
\|u\| &\leq (\lambda\|x^* - x\|^2/2 + \max\{\|I - D_{II}\|, \|D_{JJ}\|\}(\|x - x^*\| + \|y - y^*\|))/\min_i D_{ii}(1 - D_{ii})\\
&\quad + \|x - x^*\|,\\
\|v\| &\leq (\lambda\|x^* - x\|^2/2 + 2\max\{\|I - D_{II}\|, \|D_{JJ}\|\}(\|x - x^*\| + \|y - y^*\|))/\min_i D_{ii}(1 - D_{ii})\\
&\quad + \|y - y^*\|,
\end{aligned} \tag{21}$$

where (u, v) solves (3), and $D \in \Re^{n\times n}$ is diagonal with diagonal entries $D_{ii} = g'((x_i - y_i)/\mu)$, $i = 1, ..., n$. In addition,

$$\min\{(x_I, x_J), (y_I, y_J)\} = (y_I, x_J), \qquad \|(y_I, x_J)\|_\infty \leq (\beta + g(0))\mu. \tag{22}$$

Proof. Using (4), (3) may be written as

$$F'(x)u - v + F(x) - y = 0, \qquad (I - D)u + Dv = -\min\{x, y\}.$$

Using $F(x^*) - y^* = 0$, this can be rewritten as

$$\begin{array}{rl} F'(x)(x - x^* + u) - (y - y^* + v) & = F(x^*) - F(x) + F'(x)(x - x^*), \\ (I - D)(x - x^* + u) + D(y - y^* + v) & = r, \end{array} \tag{23}$$

where we let $r = (I - D)(x - x^*) + D(y - y^*) - \min\{x, y\}$. Since $\|F(x^*) - F(x) + F'(x)(x - x^*)\| = \|\int_0^1 (F'(x + t(x^* - x)) - F'(x))(x^* - x)dt\| \leq \int_0^1 \lambda t\|x^* - x\|^2 dt = \lambda\|x^* - x\|^2/2$, multiplying the first equation on the left by $(x - x^* + u)^T$ and using the positive semidefiniteness of $F'(x)$ yields

$$\begin{array}{rcl} -(x - x^* + u)^T(y - y^* + v) & \leq & (x - x^* + u)^T(F(x^*) - F(x) + F'(x)(x - x^*)) \\ & \leq & \lambda\|D\|\|D^{-1}(x - x^* + u)\|\|x^* - x\|^2/2. \end{array}$$

Then, multiplying the second equation on the left by $(x-x^*+u)^T D^{-1}$ and rearranging terms yield

$$\begin{array}{rcl} & \min\limits_i D_{ii}(1 - D_{ii})\|D^{-1}(x - x^* + u)\|^2 & \\ \leq & (x - x^* + u)^T(I - D)D^{-1}(x - x^* + u) & \\ = & -(x - x^* + u)^T(y - y^* + v) + (x - x^* + u)^T D^{-1} r & \\ \leq & \lambda\|D\|\|D^{-1}(x - x^* + u)\|\|x^* - x\|^2/2 + \|D^{-1}(x - x^* + u)\|\|r\|. & (24) \end{array}$$

We have from a consequence of [40, Lemma 3.1] that $0 < \mu g(t/\mu) - \max\{0, t\} \leq g(0)\mu$ for all $\mu > 0$ and $t \in \Re$. Thus, for each $i \in N$,

$$\begin{array}{rcl} |\min\{x_i, y_i\}| & = & |x_i - \max\{0, x_i - y_i\}| \\ & = & |x_i - \mu g((x_i - y_i)/\mu)| + |\mu g((x_i - y_i)/\mu) - \max\{0, x_i - y_i\}| \\ & \leq & \beta\mu + g(0)\mu, \end{array}$$

where the last inequality also uses $(x, y, \mu) \in \mathcal{N}_{\beta,\beta_2}$. Since $(\beta + g(0))\mu < \rho$ and $(x_I, y_J) \geq \rho e$, this yields (22). Rearranging coordinates if necessary, we can assume without loss of generality that $x = (x_I, x_J), y = (y_I, y_J)$. Then (22) and $(y_I^*, x_J^*) = 0$ yield

$$\begin{array}{rcl} r & = & (I - D)(x - x^*) + D(y - y^*) - \min\{x, y\} \\ & = & (I - D)\begin{bmatrix} x_I - x_I^* \\ x_J \end{bmatrix} + D\begin{bmatrix} y_I \\ y_J - y_J^* \end{bmatrix} - \begin{bmatrix} y_I \\ x_J \end{bmatrix} \\ & = & \begin{bmatrix} (I - D_{II})(x_I - x_I^* - y_I + y_I^*) \\ D_{JJ}(y_J - y_J^* - x_J + x_J^*) \end{bmatrix}. \end{array}$$

This and (24) imply

$$\begin{aligned} & \min_i D_{ii}(1-D_{ii})\|D^{-1}(x-x^*+u)\| \\ \leq\ & \lambda\|D\|\|x^*-x\|^2/2+\max\{\|I-D_{II}\|,\|D_{JJ}\|\}(\|x-x^*\|+\|y-y^*\|). \quad (25) \end{aligned}$$

Since $\|D^{-1}(x-x^*+u)\| \geq \|x-x^*+u\|/\|D\| \geq (\|u\|-\|x-x^*\|)/\|D\|$ and $\|D\| \leq 1$, this yields

$$\|u\| \leq \|x-x^*\|+(\lambda\|x^*-x\|^2/2+\max\{\|I-D_{II}\|,\|D_{JJ}\|\}(\|x-x^*\|+\|y-y^*\|))/\min_i D_{ii}(1-D_{ii}),$$

so the first inequality in (21) is shown. Finally, multiplying the second equation in (23) by D^{-1} yields

$$\begin{aligned} & \|y-y^*+v\| \\ =\ & \|D^{-1}r-(I-D)D^{-1}(x-x^*+u)\| \\ \leq\ & \|(D(I-D))^{-1}\|\|I-D\|\|r\|+\|(I-D)\|\|D^{-1}(x-x^*+u)\| \\ \leq\ & (\max\{\|I-D_{II}\|,\|D_{JJ}\|\}(\|x-x^*\|+\|y-y^*\|))/\min_i D_{ii}(1-D_{ii}) \\ +\ & (\lambda\|x^*-x\|^2/2+\max\{\|I-D_{II}\|,\|D_{JJ}\|\}(\|x-x^*\|+\|y-y^*\|))/\min_i D_{ii}(1-D_{ii}), \end{aligned}$$

where the last inequality also uses (25) and $\|D\| \leq 1, \|I-D\| \leq 1$. Rearranging terms yields the second inequality in (21). ∎

Lemma 5 *Assume F is Lipschitz continuous on $\Re^n$ with constant $L > 0$ and (9) holds for some $\kappa > 0, \delta > 0$. Let H_μ and $\mathcal{N}_{\beta,\beta_2}$ be given by (5) and (11) with $\beta > 0, \beta_2 \geq 0$, and $g \in \mathcal{CM}$. Then, for any $(x,y,\mu) \in \mathcal{N}_{\beta,\beta_2}$ with $(x_I, y_J) \geq \rho e$ and $\beta_2\sqrt{n}\mu \leq \delta\rho$, $(\max\{1, L\kappa\sqrt{n}\}(\beta+g(0))+\beta)\mu < \rho$, where $\rho > 0$ and $I \subset N, J = N\backslash I$, there exists an $(x^*,y^*) \in S$ satisfying $(y^*_I, x^*_J) = 0$ and*

$$\|x-x^*\| \leq \kappa(\beta+g(0))\sqrt{n}\mu, \quad \|y-y^*\| \leq (L\kappa(\beta+g(0))\sqrt{n}+\beta)\mu.$$

Proof. By Lemma 4, we have that (22) holds, so (9) implies the existence of $(x^*,y^*) \in S$ such that

$$\|x^*-x\| \leq \kappa\|(y_I, x_J)\| \leq \kappa(\beta+g(0))\sqrt{n}\mu.$$

We also have

$$\|y^*-y\| = \|F(x^*)-F(x)+F(x)-y\| \leq L\|x^*-x\|+\beta\mu \leq (L\kappa(\beta+g(0))\sqrt{n}+\beta)\mu.$$

Since $(x_I, y_J) \geq \rho e$ and $(L\kappa(\beta+g(0))\sqrt{n}+\beta)\mu < \rho$, this implies $(x^*_I, y^*_J) > 0$, so $(y^*_I, x^*_J) = 0$. ∎

The next lemma is analogous to Lemma 3.

Lemma 6 *Assume F' is Lipschitz continuous on $\Re^n_+$ with constant $\lambda \geq 0$. Let H_μ and $\mathcal{N}_{\beta,\beta_2}$ be given by (5) and (11), with $\beta > 0, \beta_2 \geq 0$ and $g \in \mathcal{CM}$ satisfying $g'(-\tau) \leq \kappa_1/\tau$ and $1 - g'(\tau) \leq \kappa_1/\tau$ for all $\tau > 0$ ($\kappa_1 > 0$). Then, for any $(x, y, \mu) \in \mathcal{N}_{\beta,\beta_2}$ with $(x_I, y_J) \geq \rho e$ and $(\beta + g(0))\mu \leq \rho/2$ and (u, v) satisfying (3) with $\|u\|_\infty + \|v\|_\infty < \rho/2$, where $\rho > 0$ and $I \subset N, J = N \backslash I$, we have $(x + u, y + v, \mu^+) \in \mathcal{N}_{\beta_1,\beta_2}$ whenever $\beta_1 > g(0)$ and*

$$\mu^+ \geq \max\{2\kappa_1\mu(\|u\|_\infty + \|v\|_\infty)/(\rho(\beta_1 - g(0))), \lambda\|u\|^2/(2\beta_2)\}. \tag{26}$$

Proof. Since $(x_I, y_J) \geq \rho e$ and $(\beta + g(0))\mu < \rho$, Lemma 4 implies (22) holds. Also, we have from (3) and (5) that

$$(I - D)u + Dv = -\min\{x, y\}, \tag{27}$$

where $D \in \Re^{n\times n}$ is diagonal with diagonal entries $D_{ii} = g'((x_i - y_i)/\mu)$, $i = 1, ..., n$. Since $0 \leq \tau - g(\tau) \leq g(0)$ for all $\tau > 0$ and $\beta_1 > g(0)$, there exists a $\eta > 0$ such that $|\tau - g(\tau - s)| \leq \beta_1$ for all $\tau > 0$ and $|s| \leq \eta$. For each $i \in I$, we have from (22) and (27) that

$$(1 - D_{ii})(u_i - v_i) = -(y_i + v_i).$$

Since $x_i \geq \rho$ and $|y_i| \leq (\beta + g(0))\mu \leq \rho/2$, we have $x_i - y_i \geq \rho/2$ and hence $1 - D_{ii} \leq \kappa_1\mu/(x_i - y_i) \leq 2\kappa_1\mu/\rho$. This implies $|y_i + v_i| = |1 - D_{ii}||u_i - v_i| \leq 2\kappa_1\mu(\|u\|_\infty + \|v\|_\infty)/\rho$. Thus, for $\mu^+ \geq 2\kappa_1\mu(\|u\|_\infty + \|v\|_\infty)/(\rho\eta)$, we have $|y_i + v_i|/\mu^+ \leq \eta$ as well as $x_i + u_i > 0$ (since $\|u\|_\infty < \rho \leq x_i$), implying

$$|(x_i + u_i)/\mu^+ - g(((x_i + u_i) - (y_i + v_i))/\mu^+)| \leq \beta_1.$$

For each $i \in J$, we have from (22) and (27) that

$$D_{ii}(u_i - v_i) = x_i + u_i.$$

Since $y_i \geq \rho$ and $|x_i| \leq (\beta + g(0))\mu \leq \rho/2$, we have $x_i - y_i \leq -\rho/2$ and hence $D_{ii} \leq \kappa_1\mu/|x_i - y_i| \leq 2\kappa_1\mu/\rho$. This implies $|x_i + u_i| = |D_{ii}||u_i - v_i| \leq 2\kappa_1\mu(\|u\|_\infty + \|v\|_\infty)/\rho$. Since $0 \leq g(\tau) \leq g(0)$ for $t < 0$ and $(x_i + u_i) - (y_i + v_i) < 0$ (since $\|u\|_\infty + \|v\|_\infty < \rho/2 \leq y_i - x_i$), this implies

$$|(x_i + u_i)/\mu^+ - g(((x_i + u_i) - (y_i + v_i))/\mu^+)| \leq |(x_i + u_i)/\mu^+| + g(0) \leq \beta_1$$

whenever $2\kappa_1\mu(\|u\|_\infty + \|v\|_\infty)/\rho \leq (\beta_1 - g(0))\mu^+$. Lastly, as in the proof of Lemma 3, we have $\|F(x + u) - (y + v)\| \leq \beta_2\mu^+$ whenever $\mu^+ \geq \lambda\|u\|^2/(2\beta_2)$. Thus, $(x + u, y + v, \mu^+) \in \mathcal{N}_{\beta_1,\beta_2}$ whenever μ^+ satisfies (26). ∎

Combining Lemmas 4, 5 and 6, we obtain the following main result of this section.

Proposition 2 *Assume F is monotone and affine. Let H_μ and $\mathcal{N}_{\beta_1,\beta_2}$ be given by (5) and (11), with $\beta_1 > 0, \beta_2 \geq 0$, $g \in \mathcal{CM}$ satisfying $g(0) < \beta_1$, $g'(-\tau) \leq \kappa_1/\tau$ and*

$1 - g'(\tau) \leq \kappa_1/\tau$ *for all* $\tau > 0$ *(*$\kappa_1 > 0$*), and, for each* $\kappa_2 \geq 1$*, there exist* $\kappa_3 > 0$ *such that* $g'(-\tau)/g'(-s) \leq \kappa_3$ *and* $(1 - g'(\tau))/(1 - g'(s)) \leq \kappa_3$ *whenever* $0 < \tau < s \leq \kappa_2\tau$*. Let* $\{(x^t, y^t, \mu^t, u^t, v^t, \hat{\mu}^t, \beta^t)\}_{t=0,1,...}$ *be generated by Algorithm 1. Assume* $\{(x^t, y^t)\}$ *is bounded,* $\mu^t \to 0$ *and* $\liminf_{t\to\infty} x^t + y^t > 0$*. Then there exist* $\bar{t}$ *and* $C_1 > 0, C_2 > 0$ *such that* $\|(u^t, v^t)\| \leq C_1\mu^t$ *and* $\mu^{t+1} \leq C_2(\mu^t)^2$ *for all* $t \geq \bar{t}$*.*

Proof. Since F is affine, then F' is Lipschitz continuous on $\Re^n_+$ with constant $\lambda = 0$ and, by the remark following (9), (9) holds for some $\kappa > 0, \delta > 0$. Since $\mu^t \to 0$ and $\liminf_{t\to\infty} x^t + y^t > 0$, Lemma 5 yields there exist $\bar{t}$ and $\rho > 0$ such that $t \geq \bar{t}$ implies $(x^t_{I^t}, y^t_{J^t}) \geq \rho e$ for some $I^t \subset N, J^t = N \backslash I^t$, and the existence of $(\bar{x}^t, \bar{y}^t) \in S$ satisfying

$$(\bar{y}^t_{I^t}, \bar{x}^t_{J^t}) = 0, \quad \|x^t - \bar{x}^t\| = O(\mu^t), \quad \|y^t - \bar{y}^t\| = O(\mu^t). \tag{28}$$

[Here $b = O(a)$ means $b \leq Ca$, with C a constant independent of t.] Then Lemma 4 with $\lambda = 0$ yields

$$\|(u^t, v^t)\| = O(\mu^t \max\{\|I - D^t_{I^tI^t}\|, \|D^t_{J^tJ^t}\|\}/\min_i D^t_{ii}(1 - D^t_{ii}) + \mu^t),$$

where $D^t \in \Re^{n\times n}$ is the diagonal matrix with diagonal entries $D^t_{ii} = g'((x^t_i - y^t_i)/\mu^t)$, $i = 1, ..., n$. Since $\mu^t \to 0$, (28) implies $(y^t_{I^t}, x^t_{J^t}) \to 0$ so, taking $\bar{t}$ larger if necessary, we can assume $\|(y^t_{I^t}, x^t_{J^t})\|_\infty \leq \rho/2$ for all $t \geq \bar{t}$. Then, $(x^t_{I^t}, y^t_{J^t}) \geq \rho e$ implies $x^t_{I^t} - y^t_{I^t} \geq \rho e/2$ and $x^t_{J^t} - y^t_{J^t} \leq -\rho e/2$. Since $\{(x^t, y^t)\}$ is bounded, this and our last assumption on g' imply

$$\max\{\|I - D^t_{I^tI^t}\|, \|D^t_{J^tJ^t}\|\}/\min_i D^t_{ii}(1 - D^t_{ii}) \text{ is bounded}, \tag{29}$$

so that $\|(u^t, v^t)\| = O(\mu^t)$, i.e., $\|(u^t, v^t)\| \leq C_1\mu^t$ for suitable constant $C_1 > 0$. By Lemma 6 and $\lambda = 0$, we have $(x^t + u^t, y^t + v^t, \tilde{\mu}^t) \in \mathcal{N}_{\beta_1,\beta_2}$, where

$$\tilde{\mu}^t = 2\kappa_1\mu^t(\|u^t\|_\infty + \|v^t\|_\infty)/(\rho(\beta_1 - g(0))).$$

Thus,

$$\mu^{t+1} \leq \hat{\mu}^t \leq \tilde{\mu}^t = O((\mu^t)^2).$$

■

We remark that the assumption of $\mu^t \to 0$ is reasonable and, in particular, holds when $(x^{t+1}, y^{t+1}, \mu^{t+1})$ is generated by (14) [3, 6, 40, 43]. The same references discuss conditions on F that ensures boundedness of $\{(x^t, y^t)\}$. The reasonableness of the assumption $\liminf_{t\to\infty} x^t + y^t > 0$ is unclear and remains to be determined. Notice that g given by (6) or (8) satisfies the assumptions of Prop. 2, whereas g given by (7) does not satisfy these assumptions. Also, like Prop. 1, Prop. 2 still holds if the assumptions on F hold only locally around the cluster points of $\{x^t\}$. Unlike Prop. 1, Prop. 2 does not assume $\liminf_{t\to\infty} \beta_1/\beta^t > 1$ but it does assume F is affine. The latter is needed to establish (29). Otherwise we would need to assume instead that

$g'(-\tau) \geq \bar{\kappa}_1/\tau$ and $1-g'(\tau) \geq \bar{\kappa}_1/\tau$ for some $\bar{\kappa}_1 > 0$, which would imply $g(\tau)$ behaves like $-\ln(-\tau)$ for $\tau < 0$ and like $\tau - \ln(\tau)$ for $\tau > 0$. An example is

$$g(\tau) = \begin{cases} -\ln(1/2-\tau)/4 & \text{if } \tau \leq 0 \\ \tau - \ln(1/2+\tau)/4 & \text{if } \tau > 0 \end{cases}.$$

Such a g unfortunately *cannot* belong to $\mathcal{CM}$ due to $g(-\tau) \not\to 0, \tau - g(\tau) \not\to 0$ as $\tau \to \infty$. It is an open question whether Prop. 2 can be suitably extended to such a g.

5 Conclusion and Extensions

We have shown that error bounds can be useful in the superlinear convergence analysis of Newton-type methods, without assuming the solutions are isolated. Although we have restricted our analysis to a specific type of continuation/smoothing methods and to the context of CP, the underlying idea can be applied to other continuation/smoothing methods and to other contexts. For example, both Lemma 1 and Prop. 1 can be extended to monotone semidefinite linear complementarity problem where, instead of $\Re^n$, we work in the space of $n \times n$ symmetric real matrices and componentwise nonnegativity is replaced by nonnegativity of the eigenvalues [38]. It may be possible to extend our analysis to H_μ based on the Kanzow smoothing function $\sqrt{x_i^2 + y_i^2 + 2\mu}$ [16], as well as to nonsmooth Newton methods, but further study would be required.

References

[1] Burke, J. and Xu, S. (1997), "A non-interior predictor-corrector path following algorithm for the monotone linear complementarity problem," Preprint, Department of Mathematics, University of Washington, Seattle.

[2] Burke, J. and Xu, S. (1997), "A non-interior predictor-corrector path-following method for LCP," Preprint, Department of Mathematics, University of Washington, Seattle.

[3] Chen, B. and Chen, X. (1997), "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.

[4] Chen, B. and Chen, X. (1997), "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; *SIAM J. Optim.*, to appear.

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

[6] Chen, B. and Xiu, N. (1997), "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.

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

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

[9] Chen, X., Qi, L., and Sun, D. (1996), "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; *Math. Comp.*, to appear.

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

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

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

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

[14] Hotta, K., Inaba, M., and Yoshise, A. (1998), "A complexity analysis of a smoothing method using CHKS-functions for monotone linear complementarity problems," Discussion paper, Institute of Policy and Planning Sciences, University of Tsukuba, Tsukuba.

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

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

[17] Kanzow, C. and Fukushima, M. (1996), "Equivalence of the generalized complementarity problem to differentiable unconstrained minimization," *J. Optim. Theory Appl.*, 90, 581–603.

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

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

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

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

[22] Luo, Z.-Q. and Tseng, P. (1993), "Error bounds and convergence analysis of feasible descent methods: a general approach," *Ann. Oper. Res.*, 46, 157–178.

[23] Mizuno, S., Todd, M., and Ye, Y. (1993), "On adaptive-step primal-dual interior-point algorithms for linear programming," *Math. Oper. Res.*, 18, 964-981.

[24] Monteiro, R. D. C. and Zhou, F. (1997), "On superlinear convergence of infeasible-interior-point algorithms for linearly constrained convex programs," *Comput. Optim. Appl.*, 8, 245-262.

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

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

[27] Pang, J.-S. (1997), "Error bounds in mathematial programming," *Math. Programming*, 79, 299–332.

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

[29] Ralph, D. and Wright, S. J. (1997), "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, Philadelphia, 345–385.

[30] Ralph, D. and Wright, S. J. (1996; revised 1998), "Superlinear convergence of an interior-point method despite dependent constraints," preprint, Mathematics and Computer Science Division, Argonne Naional Laboratory, Argonne.

[31] Simantiraki, E. M. and Shanno, D. F. (1997), "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, Philadelphia, 386–404.

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

[33] Sun, J. and Zhao, G. (1996), "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; *SIAM J. Optim.*, to appear.

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

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

[36] Tseng, P. (1995), "On linear convergence of iterative methods for the variational inequality problem," *J. Comp. Appl. Math.*, 60, 237–252.

[37] Tseng, P. (1995), "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, 423-434.

[38] Tseng, P. (1996), "Analysis of infeasible path-following methods using the Alizadeh-Haeberly-Overton direction for the monotone semi-definite LCP," Report, Department of Mathematics, University of Washington, Seattle.

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

[40] Tseng, P. (1998), "Analysis of a non-interior continuation method based on Chen-Mangasarian smoothing functions for complementarity problems," in *Reformulation – Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods* edited by M. Fukushima and L. Qi, Kluwer Academic Publishers, Boston, 381–404.

[41] Wright, S. J. (1997), *Primal-Dual Interior-Point Methods*, SIAM, Philadelphia.

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

[43] Xu, S. (1997), "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.