The proximal point algorithm with genuine superlinear convergence for the monotone complementarity problem

by

Nobuo Yamashita and Masao Fukushima

Department of Applied Mathematics and Physics,
Graduate School of Informatics, Kyoto University,
Kyoto 606-8501, Japan

May 31, 1999

Abstract. In this paper, we consider a proximal point algorithm (PPA) for solving monotone nonlinear complementarity problems (NCP). PPA generates a sequence by solving subproblems that are regularizations of the original problem. It is known that PPA has global and superlinear convergence property under appropriate criteria for approximate solutions of subproblems. However, it is not always easy to solve subproblems or to check those criteria. In this paper, we adopt the generalized Newton method proposed by De Luca, Facchinei and Kanzow to solve subproblems and some NCP functions to check the criteria. Then we show that the PPA converges globally provided that the solution set of the problem is nonempty. Moreover, without assuming the local uniqueness of the solution, we show that the rate of convergence is superlinear in a genuine sense, provided that the limit point satisfies the strict complementarity condition.

Key words: Nonlinear complememtarity problem, proximal point algrithm, genuine superlinear convergence.

AMS: 47H05, 90C33

1 Introduction

The nonlinear complementarity problem (NCP) [8] is to find a vector $x \in R^n$ such that

$$\text{NCP}(F): \qquad F(x) \geq 0, \quad x \geq 0, \quad \langle x, F(x) \rangle = 0,$$

where F is a mapping from R^n into R^n and $\langle \cdot, \cdot \rangle$ denotes the inner product in R^n. Throughout this paper we assume that F is continuously differentiable and monotone.

Until now, a variety of methods for solving NCP have been proposed and investigated. Among them, the proximal point algorithm (PPA) proposed by Martinet [7] and further studied by Rockafellar [10] is known for its theoretically nice convergence properties. PPA is originally designed to find a vector x satisfying $0 \in T(x)$, where T is a maximal monotone operator. Hence it is applicable to a wide class of problems such as convex programming problems, monotone variational inequality problems and monotone complementarity problems. In this paper, we focus on PPA for solving monotone complementarity problems. PPA generates a sequence $\{x^k\}$ by solving subproblems that are regularizations of the original problem. For NCP(F), given the current point x^k, PPA obtains the next point x^{k+1} by approximately solving the subproblem

$$F^k(x) \geq 0, \quad x \geq 0, \quad \langle x, F^k(x) \rangle = 0, \tag{1}$$

where $F^k : R^n \to R^n$ is defined by

$$F^k(x) := F(x) + c_k(x - x^k) \tag{2}$$

and $c_k > 0$. The mapping F^k is strongly monotone when F is monotone. Hence subproblem (1) is expected to be more tractable than the original problem. With appropriate criteria for approximate solutions of subproblems (1), PPA has global and superlinear convergence property under mild conditions [6, 10] However, it is not always easy to check those criteria for general monotone operator problems. In this paper, we will show that, for monotone complementarity problems, some NCP functions turn out to be useful in constructing practical approximation criteria. Another implementation issue is how to solve subproblems efficiently. In the PPA proposed in this paper, we will use the generalized Newton method proposed by De Luca, Facchinei and Kanzow [2] to solve subproblems (1). Since F^k is strongly monotone, we can show that the approximation criteria for each subproblem are attained finitely. The PPA then converges globally provided that the solution set of NCP(F) is nonempty. Moreover, without assuming the local uniqueness of the solution, we can show that the rate of convergence is superlinear. From the practical viewpoint, it is important to estimate computational costs for solving a subproblem at each iteration. We give conditions under which the approximation criteria for the subproblem are eventually fulfilled by a single Newton iteration.

The paper is organized as follows. In Section 2, we review some concepts and preliminary results that will be used in the subsequent analysis. In Section 3, we describe the proposed PPA for NCP(F). In Section 4 we show its convergence properties.

Throughout we adopt the following notation. For $a \in R$, $(a)_+$ denotes $\max\{0, a\}$, and for $x \in R^n$, $[x]_+$ denote the projection of x onto R^n_+, the nonnegative orthant of R^n. For two vectors x and y, $\min\{x, y\}$ denotes the vector whose ith element is $\min\{x_i, y_i\}$.

2 Preliminaries

In this section, we first review some mathematical concepts and basic properties of PPA that will be used in the subsequent analysis. We then discuss reformulations of NCP and related results concerning error bounds. Finally, we briefly mention the generalized Newton method for NCP proposed in [2], which will be used to solve subproblems in PPA.

2.1 Mathematical concepts

First we recall some definitions concerning the monotonicity of a mapping from R^n into itself.

Definition 2.1. *The mapping $F : R^n \to R^n$ is called a*

(a) *monotone function if*

$$\langle x - y, F(x) - F(y) \rangle \geq 0 \quad \text{for all } x, y \in R^n, \tag{3}$$

(b) *strongly monotone function with modulus* $\mu > 0$ *if*

$$\langle x - y, F(x) - F(y) \rangle \geq \mu \|x - y\|^2 \quad \text{for all } x, y \in R^n. \tag{4}$$

From the definition, it is clear that a strongly monotone function is monotone. Moreover, if F is a differentiable monotone function, then $\nabla F(x)$ is positive semidefinite for all $x \in R^n$.

Definition 2.2. *Let* $H : R^n \to R^n$ *be locally Lipschitz continuous. Then the B subdifferential of* H *at* x *is the set of* $n \times n$ *matrices defined by*

$$\partial_B H(x) := \left\{ \lim_{\substack{x^i \in D_H \\ x^i \to x}} \nabla H(x^i)^T \right\},$$

where $D_H \subseteq R^n$ *is the set where* H *is differentiable.*

Note that the Clarke subdifferential of H is defined by

$$\partial H(x) := \text{co } \partial_B H(x),$$

where co denotes the convex hull of a set [1].

Next we recall the notion of semismoothness, which lies in between the differentiability and the directional differentiability.

Definition 2.3. *Let* $H : R^n \to R^n$ *be locally Lipschitz continuous. We say that* H *is semismooth at* x *if*

$$\lim_{\substack{V \in \partial H(x+td') \\ d' \to d, t \downarrow 0}} V d' \tag{5}$$

exists for all d. *Moreover, we say that* H *is strongly semismooth at* x *if for any* $d \to 0$ *and for any* $V \in \partial H(x + d)$,

$$Vd - H'(x; d) = O(\|d\|^2),$$

where $H'(x; d)$ *denotes the directional derivative of* H *at* x *along direction* d.

Note that, when H is semismooth at x, the limit (5) is equal to the directional derivative $H'(x; d)$.

2.2 Proximal point algorithm

NCP(F) is equivalent to the problem of finding a point x such that

$$0 \in T(x), \tag{6}$$

where $T : R^n \to 2^{R^n}$ is the maximal monotone mapping defined by

$$T(x) := F(x) + N(x), \tag{7}$$

with $N : R^n \to 2^{R^n}$ being the normal cone mapping for R^n_+ defined by

$$N(x) := \begin{cases} \{y \in R^n \mid \langle x - z,\ y \rangle \geq 0,\ \ \forall z \geq 0\} & \text{if } x \geq 0, \\ \emptyset & \text{otherwise.} \end{cases}$$

With an arbitrary initial point x^0, PPA generates a sequence $\{x^k\}$ converging to a solution of (6) by the iterative scheme:

$$x^{k+1} \approx P_k(x^k),$$

where $P_k : R^n \to R^n$ is the mapping defined by $P_k := (I + \frac{1}{c_k} T)^{-1}$, $\{c_k\}$ is a positive sequence, and $x^{k+1} \approx P_k(x^k)$ means that x^{k+1} is an approximation to $P_k(x^k)$. For NCP(F), this procedure amounts to approximately solving the following subproblem NCP(F^k): Find $x \in R^n$ such that

$$F^k(x) \geq 0,\ x \geq 0,\ \langle x, F^k(x) \rangle = 0, \tag{8}$$

where F^k is defined by (2). Note that when c_k is small, the subproblem is close to the original one. On the other hand, when c_k is large, a solution of the subproblem is expected to lie near x^k, and hence the subproblem is presumably easy to solve.

To ensure convergence of PPA, x^{k+1} has to be located sufficiently near the solution $P_k(x^k)$ of subproblem (1). There have been proposed a number of criteria for the approximate solution of the subproblem. Among others, Rockefeller [10] proposed the following two criteria.

Criterion 1. $\|x^{k+1} - P_k(x^k)\| \leq \varepsilon_k, \quad \sum_{k=0}^{\infty} \varepsilon_k < \infty.$

Criterion 2. $\|x^{k+1} - P_k(x^k)\| \leq \eta_k \|x^{k+1} - x^k\|, \quad \sum_{k=0}^{\infty} \eta_k < \infty.$

Note that Criterion 1 guarantees global convergence, while Criterion 2, which is rather restrictive, ensures superlinear convergence of PPA.

Theorem 2.1 *([10, Theorem 1]) Suppose that the sequence $\{x^k\}$ is generated by PPA with Criterion 1 and that $\{c_k\}$ is bounded. If NCP(F) has at least one solution, then $\{x^k\}$ converges to a solution x^* of NCP(F).* □

Note that it is not necessary to let $\{c_k\}$ converge to 0 for the global convergence. Therefore, we may keep F^k uniformly strongly monotone, so that subproblems (1) are numerically well-conditioned.

On the other hand, if we let $\{c_k\}$ converge to 0, we can expect rapid convergence of PPA. Luque [6, Theorem 2] showed superlinear convergence without assuming the local uniqueness of the solution of NCP(F).

Theorem 2.2 *([6, Theorem 2]) Suppose that $\{x^k\}$ is generated by PPA with Criteria 1 and 2 and that $c^k \to 0$. If there exist positive constants δ and C such that*

$$\mathrm{dist}(x, \bar{X}) \leq C\|w\| \text{ whenever } x \in T^{-1}(w) \text{ and } \|w\| \leq \delta,$$

where $\mathrm{dist}(x, \bar{X})$ denotes the distance from point x to the solution set $\bar{X}$ of NCP(F), then the sequence $\{\mathrm{dist}(x^k, \bar{X})\}$ converges to 0 superlinearly. □

2.3 Reformulations of NCP

NCP can be reformulated as a system of equations in various ways. In this subsection, we review basic properties of two reformulations of NCP that will play a crucial role in solving subproblems of PPA. In the remainder of this section, we deal with the problem NCP($\hat{F}$), where $\hat{F} : R^n \to R^n$ is a certain mapping.

First we consider the function $\phi_{FB} : R^2 \to R$ defined by

$$\phi_{FB}(a, b) := a + b - \sqrt{a^2 + b^2}. \tag{9}$$

This function is called Fischer-Burmeister function and has the following property:

$$\phi_{FB}(a, b) = 0 \iff a \geq 0, b \geq 0, ab = 0.$$

Any function with this property is often called an NCP function. Using the function ϕ_{FB}, we define the mapping $H : R^n \to R^n$ by

$$H(x) := \begin{pmatrix} \phi_{FB}(x_1, \hat{F}_1(x)) \\ \vdots \\ \phi_{FB}(x_n, \hat{F}_n(x)) \end{pmatrix}. \tag{10}$$

Then it is straightforward to see that NCP($\hat{F}$) is equivalent to the system of equations

$$H(x) = 0. \tag{11}$$

The mapping H is not differentiable at a point x such that $x_i = \hat{F}_i(x) = 0$ for some i. However, when $\hat{F}$ is continuously differentiable, H is locally Lipschitz, and hence it has the B subdifferential everywhere. Though it is not necessarily easy to calculate the B subdifferential of a general locally Lipschitz mapping, De Luca et al. [2] show that, for the mapping H, an element V of $\partial_B H(x)$ is expressed as

$$V = D_a + \nabla \hat{F}(x)^T D_b, \tag{12}$$

where D_a, D_b are diagonal matrices defined by

$$((D_a)_{ii},(D_b)_{ii}) = \begin{cases} \left(1-\frac{x_i}{\sqrt{x_i^2+\hat{F}_i(x)^2}}, 1-\frac{\hat{F}_i(x)}{\sqrt{x_i^2+\hat{F}_i(x)^2}}\right), & \text{if } (x_i,\hat{F}_i(x)) \neq (0,0) \\ (1-\eta, 1-\xi), & \text{otherwise} \end{cases} \tag{13}$$

and (η,ξ) is a vector satisfying $\eta^2+\xi^2=1$. De Luca et al. [2] also discuss how to calculate (η,ξ) when $(x_i,\hat{F}_i(x))=(0,0)$.

The next proposition will be useful in the analysis of the generalized Newton method for solving subproblems of PPA.

Proposition 2.1 *Let M be a positive definite matrix and μ be a positive constant such that*

$$\langle v, Mv\rangle \geq \mu\|v\|^2, \ \ \forall v \in R^n. \tag{14}$$

Let $D_a = \mathrm{diag}(a_i)$ and $D_b = \mathrm{diag}(b_i)$ be diagonal matrices whose diagonal elements are nonnegative and satisfy $a_i + b_i \geq d$ for all i, where d is a positive constant. Then we have

$$\inf_{\|v\|=1} \|(D_a + M^T D_b)v\| \geq \bar{B}\mu,$$

where $\bar{B} = d/(n\max\{1,\|M\|\})$. Moreover, the following inequality holds:

$$\|(D_a + M^T D_b)^{-1}\| \leq \frac{1}{\bar{B}\mu}.$$

Proof. Since any square matrix satisfies

$$\inf_{\|v\|=1}\|Av\| = \inf_{\|v\|=1}\|A^T v\|,$$

to prove the first part of the lemma it suffices to show that

$$\inf_{\|v\|=1}\|(D_a + D_b M)v\| \geq \bar{B}\mu.$$

Let v be an arbitrary vector such that $\|v\|=1$. Then, since

$$\langle v, Mv\rangle \geq \mu$$

holds by (14), there exists an index i such that

$$v_i(Mv)_i \geq \frac{\mu}{n}. \tag{15}$$

Since

$$v_i(Mv)_i \leq |v_i|\|M\|,$$

it follows from (15) that

$$|v_i| \geq \frac{\mu}{n\|M\|}. \tag{16}$$

Moreover, (15) implies that v_i has the same sign as $(Mv)_i$. Hence, by (15) and (16), we have

$$\begin{aligned}
|((D_a + D_b M)v)_i| &= a_i|v_i| + b_i|(Mv)_i| \\
&\geq \frac{\mu}{n\|M\|}a_i + \frac{\mu}{n|v_i|}b_i \\
&\geq \frac{\mu}{n\|M\|}a_i + \frac{\mu}{n}b_i \\
&\geq \frac{(a_i+b_i)\mu}{n\max\{1,\|M\|\}} \\
&\geq \bar{B}\mu.
\end{aligned}$$

Consequently, we have

$$\|(D_a + D_b M)v\| \geq \bar{B}\mu.$$

Next we show the last part of the lemma. Note that, under the given assumptions, $D_a + M^T D_b$ is nonsingular [2, Lemma 5.1]. Since

$$\|(D_a + M^T D_b)^{-1}\| = \frac{1}{\inf_{\|v\|=1}\|(D_a + M^T D_b)v\|},$$

it follows that

$$\|(D_a + M^T D_b)^{-1}\| \leq \frac{1}{\bar{B}\mu}.$$

□

As a direct consequence of this proposition, we have the following corollary.

Corollary 2.1 *Suppose that $\hat{F}$ is strongly monotone with modulus μ. Let D_a and D_b be defined by (13). Then we have*

$$\|(D_a + \nabla\hat{F}(x)^T D_b)^{-1}\| \leq \frac{1}{B_1\mu},$$

where $B_1 = (2-\sqrt{2})/(n\max\{1,\|\nabla\hat{F}(x)\|\})$. □

Now we define the function $\Phi_{FB} : R^n \to R$ by

$$\Phi_{FB}(x) := \frac{1}{2}\|H(x)\|^2, \tag{17}$$

where H is given by (10). We note that Φ_{FB} attains its global minimum at a solution of NCP($\hat{F}$), because NCP($\hat{F}$) is equivalent to (11).

Lemma 2.1 *The mapping $H : R^n \to R^n$ defined by (10) has the following properties:*

(a) *If $\hat{F}$ is differentiable, then H is semismooth.*

(b) *If $\nabla\hat{F}$ is locally Lipschitz continuous, then H is strongly semismooth.*

(c) *If $\nabla \hat{F}(x)$ is positive definite, then every $V \in \partial_B H(x)$ is nonsingular.*

Proof. Items (a) and (c) are shown in [3]. Item (b) is shown in [11]. □

Lemma 2.2 *The function $\Phi_{FB} : R^n \to R$ defined by (17) has the following properties:*

(a) *If $\hat{F}$ is differentiable, then Φ_{FB} is differentiable.*

(b) *If $\hat{F}$ is monotone, then any stationary point of Φ_{FB} is a solution of NCP($\hat{F}$).*

(c) *If $\hat{F}$ is strongly monotone with modulus μ and Lipschitz continuous with constant L, then $\sqrt{\Phi_{FB}(x)}$ provides a global error bound for NCP($\hat{F}$), that is,*

$$\|x - \hat{x}\| \leq \frac{B_2(L+1)}{\mu}\sqrt{\Phi_{FB}(x)} \quad \text{for all } x \in R^n,$$

where $\hat{x}$ is the unique solution of NCP($\hat{F}$) and B_2 is a positive constant independent of $\hat{F}$.

(d) *If $\hat{F}$ is affine and NCP($\hat{F}$) has a solution, then $\sqrt{\Phi_{FB}(x)}$ provides a local error bound for NCP($\hat{F}$), that is, there exist positive constants B_3 and B_4 such that*

$$\mathrm{dist}(x, \hat{X}) \leq B_3\sqrt{\Phi_{FB}(x)} \quad \text{for all } x \in \{y \in R^n \mid \Phi_{FB}(y) \leq B_4\},$$

where $\hat{X}$ denotes the solution set of NCP($\hat{F}$).

Proof. Items (a) and (b) are shown in [5]. Items (c) and (d) are shown in [4]. □

In the PPA to be presented in the next section, we will also utilize the following function $\Psi : R^n \to R$, which has a more favorable error bound property than Φ_{FB}:

$$\Psi(x) := \sum_{i=1}^{n} \psi(x_i, \hat{F}_i(x)),$$

where $\psi : R^2 \to R$ is defined by

$$\psi(a, b) := |ab| + |\min\{a, b\}|.$$

Note that ψ is also an NCP function. It is clear that $\Psi(x) \geq 0$ for all x, and $\Psi(x) = 0$ if and only if x is a solution of NCP($\hat{F}$).

The next lemma shows an interesting error bound result for the function Ψ, which will play an important role in Section 4. Note that this error bound is valid only on the set R^n_+.

Lemma 2.3 *Suppose that $\hat{F}$ is strongly monotone with modulus μ. Then we have*

$$\|x - \hat{x}\| \leq 2\max\{1, \|x\|\}\sqrt{\frac{\Psi(x)}{\mu}} \quad \text{for all } x \in \{y \in R^n_+ \mid \Psi(y) \leq \frac{\mu}{4}\},$$

where $\hat{x}$ is the unique solution of NCP($\hat{F}$).

Proof. Let $x \in R^n_+$ be arbitrary. Since $\hat{F}$ is strongly monotone with modulus μ, we have

$$\begin{aligned}
\mu\|x-\hat{x}\|^2 &\le \langle x-\hat{x}, \hat{F}(x)-\hat{F}(\hat{x})\rangle \\
&= \langle x, \hat{F}(x)\rangle + \langle \hat{x}, -\hat{F}(x)\rangle + \langle \hat{F}(\hat{x}), -x\rangle \\
&\le \sum_{i=1}^n |x_i\hat{F}_i(x)| + \sum_{i=1}^n |\hat{x}_i||(-\hat{F}_i(x))_+| + \sum_{i=1}^n |\hat{F}_i(\hat{x})||(-x_i)_+| \\
&= \sum_{i=1}^n |x_i\hat{F}_i(x)| + \sum_{i=1}^n |\hat{x}_i||(-\hat{F}_i(x))_+|.
\end{aligned}$$

Since

$$(-b)_+ \le |\min\{a,b\}| \quad \text{for all } (a,b)\in R^2,$$

it follows that

$$\begin{aligned}
\mu\|x-\hat{x}\|^2 &\le \sum_{i=1}^n |x_i\hat{F}_i(x)| + \sum_{i=1}^n |\hat{x}_i||(-\hat{F}_i(x))_+| \\
&\le \sum_{i=1}^n \left\{|x_i\hat{F}_i(x)| + |\hat{x}_i||\min\{x_i,\hat{F}_i(x)\}|\right\} \\
&\le \max\{1,\|\hat{x}\|_\infty\}\Psi(x) \\
&\le \max\{1,\|\hat{x}\|\}\Psi(x).
\end{aligned}$$

Hence we have

$$\begin{aligned}
\|x-\hat{x}\| &\le \sqrt{\frac{\max\{1,\|\hat{x}\|\}}{\mu}\Psi(x)} \\
&\le \max\{1,\|\hat{x}\|\}\sqrt{\frac{\Psi(x)}{\mu}} \\
&\le \max\{1,\|\hat{x}-x\|+\|x\|\}\sqrt{\frac{\Psi(x)}{\mu}}.
\end{aligned}$$

Therefore, if $\|\hat{x}-x\|+\|x\|\le 1$, then the desired inequality holds. If $\|\hat{x}-x\|+\|x\|>1$, then

$$\left(1-\sqrt{\frac{\Psi(x)}{\mu}}\right)\|x-\hat{x}\| \le \|x\|\sqrt{\frac{\Psi(x)}{\mu}}.$$

Since $1-\sqrt{\Psi(x)/\mu}\ge\frac{1}{2}$ whenever $\Psi(x)\le\frac{\mu}{4}$, we also have the desired inequality. □

We note that, unlike Lemma 2.1 (c), Lemma 2.3 does not assume the Lipschitz continuity of $\hat{F}$. Moreover, unlike Lemma 2.1 (d), the error bound result shown in Lemma 2.3 is explicitly represented in terms of the modulus of strong monotonicity of $\hat{F}$.

2.4 Generalized Newton method

In this section, we review the generalized Newton method for solving NCP proposed by De Luca, Facchinei and Kanzow [2]. The PPA to be presented in the next section will use this method to solve subproblems.

Procedure 1. (Generalized Newton method for NCP($\hat{F}$))

Step 1: Choose a constant $\beta \in (0, \frac{1}{2})$. Let x^0 be an initial point and set $j := 0$.

Step 2: If x^j satisfies a termination criterion, then stop.

Step 3: Choose $V_j \in \partial_B H(x^j)$ and get d^j satisfying

$$V_j d^j = -H(x^j). \tag{18}$$

Step 4: If $x^j + d^j$ satisfies the termination criterion, then stop. Otherwise, find the smallest nonnegative integer i_j such that

$$\Phi_{FB}(x^j + 2^{-i_j} d^j) \leq (1 - \beta 2^{-i_j}) \Phi_{FB}(x^j).$$

Step 5: Set $x^{j+1} := x^j + 2^{-i_j} d^j$ and $j := j + 1$, and go to Step 2.

Note that Procedure 1 is a slight simplification of the algorithm in [2]. Within the framework of the present paper, however, there is essentially no difference between them, because we only consider the case where F is strongly monotone.

For Procedure 1 with the termination criterion ignored, the following convergence result holds.

Proposition 2.2 *[2] Suppose that $\hat{F}$ is differentiable and strongly monotone and that $\nabla \hat{F}$ is Lipschitz continuous around the unique solution $\hat{x}$ of NCP($\hat{F}$). Then Procedure 1 globally converges to $\hat{x}$ and the rate of convergence is quadratic.* □

Since the mappings F^k involved in the subproblems generated by PPA are strongly monotone, Procedure 1 can be applied to these problems effectively.

3 Algorithm and its convergence properties

In this section we describe PPA for NCP(F) and study its convergence properties.

Algorithm 1

Step 1: Choose parameters $\alpha \in (0, 1)$, $c_0 \in (0, 1)$ and an initial point $x^0 \in R^n$. Set $k := 0$.

Step 2: If x^k satisfies $\Psi_{FB}(x^k) = 0$, then stop.

Step 3: Let $F^k : R^n \to R^n$ be defined by (2), and apply Procedure 1 to obtain an approximate solution $\tilde{x}^{k+1}$ of NCP(F^k) that satisfies the conditions

$$\Psi^k([\tilde{x}^{k+1}]_+) \leq \frac{c_k^3}{4\max\{1, \|[\tilde{x}^{k+1}]_+\|\}^2} \tag{19}$$

and

$$\sqrt{\Phi_{FB}^k(\tilde{x}^{k+1})} \leq c_k^4 \|x^k - \tilde{x}^{k+1}\|, \tag{20}$$

where

$$\Psi^k(x) := \sum_{i=1}^n \psi(x_i, F_i^k(x))$$

and

$$\Phi_{FB}^k(x) := \sum_{i=1}^n \phi_{FB}(x_i, F_i^k(x))^2.$$

Step 4: Set $x^{k+1} := [\tilde{x}^{k+1}]_+$, $c_{k+1} := \alpha c_k$, and $k := k+1$. Go to Step 2.

Remark 3.1 *The condition (19) in Step 3 corresponds to Criterion 1 of PPA, while the condition (20) corresponds to Criterion 2.*

Remark 3.2 *If x^k is a solution of NCP(F), then the algorithm stops at Step 2. Otherwise, since $F^k(x^k) = F(x^k)$, x^k is not a solution of NCP(F^k) at Step 3. Moreoversince F^k is strongly monotone, Theorem 2.2 ensures that Procedure 1 can finitely find $\tilde{x}^{k+1}$ satisfying (19) and (20).*

First we show that Algorithm 1 has a global convergence property.

Theorem 3.1 *Suppose that NCP(F) has at least one solution. Then the sequence $\{x^k\}$ generated by Algorithm 1 converges to a solution x^* of NCP(F).*

Proof. It suffices to show that $\{x^k\}$ satisfies the assumption of Theorem 2.1, that is, $\{x^k\}$ satisfies Criterion 1. Since $x^{k+1} = [\tilde{x}^{k+1}]_+$ in Step 4 and $0 < c_k < 1$, we have, by (19) in Step 3,

$$\Psi^k(x^{k+1}) \leq \frac{c_k^3}{4\max\{1, \|x^{k+1}\|\}^2} \tag{21}$$

$$\leq \frac{c_k}{4}. \tag{22}$$

Since F^k is strongly monotone with modulus c_k, it then follows from Lemma 2.3 and (22) that

$$\|x^{k+1} - P_k(x^k)\| \leq 2\max\{1, \|x^{k+1}\|\}\sqrt{\frac{\Psi^k(x^{k+1})}{c_k}}, \tag{23}$$

where $P_k(x^k)$ is the unique solution of NCP(F^k). By (21) and (23), we have

$$\|x^{k+1} - P_k(x^k)\| \leq c_k.$$

Since $\Sigma_{k=1}^{\infty} c_k < \infty$, it follows from Theorem 2.1 that $\{x^k\}$ converges to a solution of NCP(F). □

Next we give conditions for Algorithm 1 to converge superlinearly. For this purpose, we first show that the inverse of the maximal monotone operator T defined by (7) is locally Lipschitz near the solution set of NCP(F) under the following assumption:

Assumption 1: $\|\min\{x, F(x)\}\|$ provides a local error bound for NCP(F).

Note that when F is affine, Assumption 1 holds by Lemma 2.2 (d) On the other hand, when $\nabla F(x)$ is positive definite at any solution x of NCP(F), Assumption 1 holds by Lemma 2.1 (c) and [9, Propostion 3].

Proposition 3.1 *Let T be the maximal monotone mapping defined by (7). If Assumption 1 holds and the solution set $\bar{X}$ of NCP(F) is nonempty, then there exist positive constants C and δ such that*

$$\mathrm{dist}(x, \bar{X}) \le C\|w\| \quad \forall x \in T^{-1}(w), \ \ \forall w \text{ with } \|w\| \le \delta.$$

Proof. The mapping T defined by (7) is expressed as

$$T(x) = T_1(x) \times \cdots \times T_n(x),$$

where $T_i(x) \subseteq R$ is given by

$$T_i(x) = \begin{cases} \{F_i(x)\} & \text{if } x_i > 0 \\ \{F_i(x) + v_i \mid v_i \in (-\infty, 0]\} & \text{if } x_i = 0 \\ \emptyset & \text{otherwise} \end{cases}$$

for $i = 1, \cdots, n$.

Consider a pair (x, w) such that $w \in T(x)$. If $x_i > 0$, we have

$$|w_i| = |F_i(x)| \ge |\min\{x_i, F_i(x)\}|.$$

If $x_i = 0$ and $F_i(x) > 0$, it is clear that

$$|w_i| \ge 0 = |\min\{x_i, F_i(x)\}|.$$

If $x_i = 0$ and $F_i(x) \le 0$, then there exists $v_i \le 0$ such that

$$|w_i| = |F_i(x) + v_i|.$$

Hence we have

$$|w_i| = |F_i(x) + v_i| \ge |F_i(x)| = |\min\{x_i, F_i(x)\}|.$$

Consequently we have

$$\|w\| \ge \|\min\{x, F(x)\}\|.$$

It then follows from Assumption 1 that the desired property holds. □

By using Proposition 3.1, we show that Algorithm 1 has superlinear rate of convergence.

Theorem 3.2 *Suppose that Assumption 1 holds. Let $\{x^k\}$ be generated by Algorithm 1. Then the sequence $\{\text{dist}(x^k, \bar{X})\}$ converges to 0 superlinearly.*

Proof. By Theorem 3.1, $\{x^k\}$ is bounded. Hence we may suppose that F^k is uniformly Lipschitz continuous with modulus L on a bounded set containing $\{x^k\}$. It then follows from Lemma 2.2 (c) that there exists a positive constant B_2 such that

$$\|\tilde{x}^{k+1} - P_k(x^k)\| \le \frac{B_2(L+1)}{c_k}\sqrt{\Phi^k_{FB}(\tilde{x}^{k+1})}.$$

Hence by (20) in Step 3, we have

$$\|\tilde{x}^{k+1} - P_k(x^k)\| \le B_2(L+1)c_k^3\|x^k - \tilde{x}^{k+1}\|.$$

Since $\Sigma_{k=1}^{\infty} c_k^3 < \infty$, the last inequality implies that $\{\tilde{x}^k\}$ satisfies Criterion 2. Thereforeby Proposition 3.1 and [6, Theorem 2.1], there exists a constant $C > 0$ such that for sufficiently large k

$$\text{dist}(\tilde{x}^{k+1}, \bar{X}) \le \frac{C}{(C^2 + (1/c_k)^2)^{\frac{1}{2}}}\text{dist}(x^k, \bar{X}).$$

Noting that $\text{dist}(x^{k+1}, \bar{X}) \le \text{dist}(\tilde{x}^{k+1}, \bar{X})$, we then have

$$\text{dist}(x^{k+1}, \bar{X}) \le \frac{C}{(C^2 + (1/c_k)^2)^{\frac{1}{2}}}\text{dist}(x^k, \bar{X}).$$

Since $c_k \to 0$, $\{\text{dist}(x^k, \bar{X})\}$ converges to 0 superlinearly. □

Theorem 3.2 says that the sequence $\{x^k\}$ generated by Algorithm 1 converges to the solution set $\bar{X}$ superlinearly under mild conditions. However, this does not necessarily mean that Algorithm 1 is practically efficient, because it says nothing about computational costs to solve a subproblem at each iteration. So it is important to estimate the number of iterations Procedure 1 spends at each iteration of Algorithm 1. Moreover, it is particularly interesting to see under what conditions Procedure 1 requires just a single iteration. In the next section, we answer this question.

4 Genuine superlinear convergence

In this section we give conditions under which a single Newton step of Procedure 1 for NCP(F^k) attains (19) and (20) in Step 3 of Algorithm 1, thereby genuine superlinear convergence of Algorithm 1 is ensured.

First we show that (19) is implied by (20) for sufficiently large k.

Lemma 4.1 *When k is sufficiently large, if*

$$\sqrt{\Phi^k_{FB}(\tilde{x}^{k+1})} \le c_k^4\|x^k - \tilde{x}^{k+1}\|$$

holds, then

$$\Psi^k([\tilde{x}^{k+1}]_+) \le \frac{c_k^3}{4\max\{1, \|[\tilde{x}^{k+1}]_+\|\}^2}$$

also holds

Proof. Since Ψ^k is uniformly locally Lipschitz continuous, there exists $L > 0$ such that

$$\Psi^k([\tilde{x}^{k+1}]_+) \le L\|[\tilde{x}^{k+1}]_+ - P_k(x^k)\| \le L\|\tilde{x}^{k+1} - P_k(x^k)\|, \tag{24}$$

for all k sufficiently large Moreoversince $\sqrt{\Phi_{FB}^k(x)}$ provides a global error bound for NCP(F^k) by Lemma 2.2 (c), there exists $\tau > 0$ such that

$$\|\tilde{x}^{k+1} - P_k(x^k)\| \le \frac{\tau}{c_k}\sqrt{\Phi_{FB}^k(\tilde{x}^{k+1})}. \tag{25}$$

It then follows from (20)(24) and (25) that there exists a positive constant τ' such that

$$\Psi^k([\tilde{x}^{k+1}]_+) \le \tau' c_k^3 \|\tilde{x}^{k+1} - x^k\|.$$

Since $\{\|\tilde{x}^{k+1} - x^k\|\}$ converges to 0 and since $\{\|[\tilde{x}^{k+1}]_+\|\}$ is bounded, (19) holds for sufficiently large k. □

This lemma says that (20) implies (19) for all k sufficiently large. Therefore, in the remainder of this section, we only consider conditions under which (20) is satisfied after a single Newton step for NCP(F^k).

The following lemma indicates the relation between $\|x^k - P_k(x^k)\|$ and $\text{dist}(x^k, \bar{X})$.

Lemma 4.2 *For sufficiently large k, there exists a constant $B_5 > 0$ such that*

$$\|x^k - P_k(x^k)\| \le \frac{B_5}{c_k}\text{dist}(x^k, \bar{X}).$$

Proof. Let $\bar{x}^k$ be the nearest point in $\bar{X}$ from x^k. Since $\{x^k\}$ is bounded, so is $\{\bar{x}^k\}$. Thus the function $\sqrt{\Phi_{FB}(x)}$ is Lipschitz continuous on a bounded set containing $\{x^k\}$ and $\{\bar{x}^k\}$. Moreover $\sqrt{\Phi_{FB}^k(x)}$ is also uniformly Lipschitz continuous on the same set. Let $L_1 > 0$ and $L_2 > 0$ be Lipschitz constants of $\sqrt{\Phi_{FB}(x)}$ and $\sqrt{\Phi_{FB}^k(x)}$, respectively. Then we have

$$\begin{aligned} \sqrt{\Phi_{FB}(x^k)} &= \left|\sqrt{\Phi_{FB}(x^k)} - \sqrt{\Phi_{FB}(\bar{x}^k)}\right| \\ &\le L_1\|x^k - \bar{x}^k\| \\ &= L_1\text{dist}(x^k, \bar{X}). \end{aligned}$$

It follows from Lemma 2.2 (c) that

$$\begin{aligned} \|x^k - P_k(x^k)\| &\le \frac{B_2(L_2+1)}{c_k}\sqrt{\Phi_{FB}^k(x^k)} \\ &= \frac{B_2(L_2+1)}{c_k}\sqrt{\Phi_{FB}(x^k)}. \end{aligned}$$

Combining the above inequalities and letting $B_5 = B_2(L_2+1)L_1$ yield the desired inequality. □

Next we assume that the strict complementarity is satisfied at the limit point of the generated sequence. The assumption ensures the twice differentiability of H.

Assumption 2:

(a) The limit point x^* of the sequence $\{x^k\}$ generated by Algorithm 1 is nondegenerate, that is, $x_i^* + F_i(x^*) > 0$ holds for all i.

(b) F is twice continuously differentiable.

For the sake of convenience, we define the mapping $H^k : R^n \to R^n$ by

$$H^k(x) := \begin{pmatrix} \phi_{FB}(x_1, F_1^k(x)) \\ \vdots \\ \phi_{FB}(x_n, F_n^k(x)) \end{pmatrix}.$$

Lemma 4.3 *Suppose that Assumptions 1 and 2 hold. Then H^k is twice continuously differentiable in a neighborhood of x^k for sufficiently large k, and there exists a positive constant B_6 such that*

$$\|\nabla H^k(x^k)^T(x^k - P_k(x^k)) - H^k(x^k) + H^k(P_k(x^k))\| \le B_6 \|x^k - P_k(x^k)\|^2.$$

Proof. Let $\bar{F} : R^{2n+1} \to R^n$ and $\bar{H} : R^{2n+1} \to R^n$ be defined by

$$\begin{aligned} \bar{F}(x,y,\mu) &:= F(x) + \mu(x-y), \\ \bar{H}(x,y,\mu) &:= \begin{pmatrix} \phi_{FB}(x_1, \bar{F}_1(x,y,\mu)) \\ \vdots \\ \phi_{FB}(x_n, \bar{F}_n(x,y,\mu)) \end{pmatrix}, \end{aligned}$$

respectively. Suppose that x^* is the limit point of the sequence $\{x^k\}$. Thenby Assumption 2, $\bar{H}$ is twice continuously differentiable in a neighborhood N of $(x^*, x^*, 0)$. Hence, there exists a positive constant B_6 such that

$$\left\| \begin{pmatrix} \nabla_x \bar{H}(x,y,\mu)^T & \nabla_y \bar{H}(x,y,\mu)^T & \nabla_\mu \bar{H}(x,y,\mu)^T \end{pmatrix} \begin{pmatrix} x - x' \\ y - y' \\ \mu - \mu' \end{pmatrix} - \bar{H}(x,y,\mu) + \bar{H}(x',y',\mu') \right\|$$
$$\le B_6(\|x-x'\|^2 + \|y-y'\|^2 + \|\mu-\mu'\|^2), \quad \forall (x,y,\mu), (x',y',\mu') \in N. \tag{26}$$

We also note that H^k is twice continuously differentiable near x^k when k is sufficiently large. Since

$$\nabla_x \bar{H}(x, x^k, c_k) = \nabla H^k(x)$$

and since $(x^k, x^k, c_k), (P_k(x^k), x^k, c_k) \in N$ for sufficiently large k, substituting $(x,y,\mu) = (x^k, x^k, c_k)$ and $(x',y',\mu') = (P_k(x^k), x^k, c_k)$ into (26) yields the desired inequality. □

Now let us denote

$$x_N^k := x^k - V_k^{-1} H^k(x^k), \quad V_k \in \partial_B H^k(x^k). \tag{27}$$

Note that x_N^k is a point produced by a single Newton iteration of Procedure 1 for NCP(F^k) with the initial point x^k.

By using Corollary 2.1 and Lemma 4.3, we can show the following key lemma.

Lemma 4.4 *Suppose that Assumptions 1 and 2 hold. Then there exists $B_7 > 0$ such that*

$$\|x_N^k - P_k(x^k)\| \leq \frac{B_7 \|x^k - P_k(x^k)\|^2}{c_k}$$

for sufficiently large k.

Proof. First note that, by Lemma 4.3, $\nabla H^k(x^k)$ exists and hence $V^k = \nabla H^k(x^k)$ for all k sufficiently large. By (12), $\nabla H^k(x^k)$ is expressed as $\nabla H^k(x^k) = D_a + \nabla F^k(x^k)^T D_b$. Moreover$\{\|\nabla F^k(x^k)\|\}$ is bounded. Thereforeby Corollary 2.1 and Lemma 4.3, there exist $B_1 > 0$ and $B_6 > 0$ such that

$$\begin{aligned}
\|x_N^k - P_k(x^k)\| &= \|x^k - P_k(x^k) - \nabla H^k(x^k)^{-1}(H^k(x^k) - H^k(P_k(x^k)))\| \\
&\leq \|\nabla H^k(x^k)^{-1}\| \|\nabla H^k(x^k)(x^k - P_k(x^k)) - H^k(x^k) + H^k(P_k(x^k))\| \\
&\leq \frac{B_6 \|x^k - P_k(x^k)\|^2}{B_1 c_k}
\end{aligned}$$

for sufficiently large k. Consequently, letting $B_7 = B_6/B_1$ shows the lemma. □

Now we are in a position to establish the main result of this section.

Theorem 4.1 *Suppose that Assumptions 1 and 2 hold. Let x_N^k be given by (27). Then for sufficiently large k, x_N^k satisfies the condition (20) in Step 3 of Algorithm 1, that is*

$$\sqrt{\Phi_{FB}^k(x_N^k)} \leq c_k^4 \|x_N^k - x^k\|.$$

Proof. Let $\gamma > 0$ be arbitrary. Since $\{\mathrm{dist}(x^k, \bar{X})\}$ converges to 0 superlinearly by Theorem 3.2, we have for sufficiently large k

$$\mathrm{dist}(x^k, \bar{X}) \leq \gamma c_k^6.$$

It follows from Lemma 4.2 that

$$\|x^k - P_k(x^k)\| \leq \gamma B_5 c_k^5.$$

Then by Lemma 4.4, we have

$$\begin{aligned}
\|x_N^k - P_k(x^k)\| &\leq \frac{B_7}{c_k} \|x^k - P_k(x^k)\|^2 \\
&\leq \gamma B_5 B_7 c_k^4 \|x^k - P_k(x^k)\|.
\end{aligned}$$

By the triangle inequality, the last inequality yields

$$\frac{1 - \gamma B_5 B_7 c_k^4}{\gamma B_5 B_7} \|x_N^k - P_k(x^k)\| \leq c_k^4 \|x_N^k - x^k\|. \tag{28}$$

On the other hand, since $\sqrt{\Phi_{FB}^k(x)}$ is uniformly locally Lipschitz continuous, there exists $L_2 > 0$ such that

$$\sqrt{\Phi_{FB}^k(x_N^k)} \leq L_2 \|x_N^k - P_k(x^k)\|.$$

Hence, by (28) it suffices to show

$$L_2 \leq \frac{1 - \gamma B_5 B_7 c_k^4}{\gamma B_5 B_7}.$$

Since γ is arbitrary, choosing γ sufficiently small yields the last inequality. □

This theorem along with Theorem 3.2 ensures that Algorithm 1 converges superlinearly in a genuine sense, provided that the limit of the generated sequence $\{x^k\}$ is nondegenerate.

References

[1] F.H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, New York, 1983.

[2] T. De Luca, F. Facchinei and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems,* Mathematical Programming, 75 (1996), pp. 407–439.

[3] F. Facchinei and J. Soares, *A new merit function for nonlinear complementarity problems and a related algorithm,* SIAM Journal on Optimization, 7 (1997), pp. 225–247.

[4] A. Fischer, *An NCP-function and its use for the solution of complementarity problems,* Recent Advances in Nonsmooth Optimization, eds., D.-Z. Du, L. Qi and R.S. Wormersley, World Scientific Publishers, Singapore, pp. 88–105, 1995.

[5] C. Geiger and C. Kanzow, *On the resolution of monotone complementarity problems,* Computational Optimization and Application, 5 (1996), pp. 155–173.

[6] F.J. Luque, *Asymptotic convergence analysis of the proximal point algorithm,* SIAM Journal on Control and Optimization, 22 (1984), pp. 277–293.

[7] B. Martinet, *Perturbation des méthodes d'opimisation*, R.A.I.R.O., Analyse Numérique, 12 (1978), pp. 153–171.

[8] J.-S. Pang, *Complementarity problems,* Handbook of Global Optimization, eds., R. Horst and P. Pardalos, Kluwer Academic Publishers, Boston, Massachusetts, pp. 271–338, 1994.

[9] J.-S. Pang, and L. Qi, *Nonsmooth equations: motivation and algorithms,* SIAM Journal on Optimization, 3 (1993), pp. 443–465.

[10] R.T. Rockafellar, *Monotone operators and the proximal point algorithm,* SIAM Journal on Control and Optimization, 14 (1976), pp. 877–898.

[11] N. Yamashita and M. Fukushima, *Modified Newton methods for solving a semismooth reformulation of monotone complementarity problems,* Mathematical Programming, 76 (1997), pp. 469–491.