Modified Newton Methods for Solving Semismooth Reformulations of Monotone Complementarity Problems

by

Nobuo Yamashita[†] and Masao Fukushima[‡]

Graduate School of Information Science,
Nara Institute of Science and Technology,
Ikoma, Nara 630-01, Japan.

May 31, 1995

Abstract. In this paper, we propose a Newton-type method for solving semismooth reformulations of monotone complementarity problems. In this method, a direction-finding subproblem, which is a system of linear equations, is uniquely solvable at each iteration. Moreover, the obtained search direction always affords a direction of sufficient decrease for the merit function defined as the squared residual for the semismooth equation equivalent to the complementarity problem. We show that the algorithm is globally convergent under some mild assumptions. Next, by slightly modifying the direction-finding problem, we propose another Newton-type method, which may be considered a restricted version of the first algorithm. We show that this algorithm has a superlinear, or possibly quadratic, rate of convergence under suitable assumptions. Finally, some numerical results are presented.

[†]The work of this author was supported by the Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists.

[‡] The work of this author was supported in part by the Scientific Research Grant-in-Aid from the Ministry of Education, Science and Culture, Japan.

1. Introduction

The nonlinear complementarity problem [2, 9, 12, 18] is to find an $x \in R^n$ such that

$$\text{NCP}(F): \qquad x \geq 0,\ F(x) \geq 0,\ x^T F(x) = 0,$$

where F is a continuously differentiable function from R^n into itself. An approach that has recently drawn much attention for solving NCP(F) is to transform it equivalently into a system of nonlinear equations or a minimization problem [13, 15]. For that purpose, a function $\psi : R^2 \to R$ satisfying

$$\psi(a, b) = 0 \Longleftrightarrow a \geq 0,\ b \geq 0,\ ab = 0 \tag{1}$$

has turned out to be useful. By using this function, we can form the following system of equations equivalent to NCP(F):

$$\Psi(x) = \begin{pmatrix} \psi(x_1, F_1(x)) \\ \psi(x_2, F_2(x)) \\ \vdots \\ \psi(x_n, F_n(x)) \end{pmatrix} = 0,$$

which may alternatively be reformulated as the minimization problem

$$\min_{x \in R^n} \quad \frac{1}{2}\|\Psi(x)\|^2.$$

Functions satisfying (1) are often called NCP-functions [7, 14]. Up to the present, various NCP-functions have been proposed [5, 13–15] and used to derive algorithms for solving NCP(F). For example, Subramanian [24] proposed a Levenberg-Marquardt-type algorithm for solving the system of nonlinear equations derived from the NCP-function proposed by Mangasarian [15]. This method has favorable properties that the direction-finding subproblem is solvable at an arbitrary point and that it has a quadratic rate of convergence. However, the condition under which a stationary point of the merit function, to which the generated sequence is expected to converge, is a solution of NCP(F) is somewhat restrictive, and the condition number of the coefficient matrix in the direction-finding subproblem is larger than that of Newton-type methods mentioned below. Pang and Gabriel [19] proposed an algorithm called NE/SQP method, which is based on the NCP-function $\psi(a, b) = \min(a, b)$. Quadratical convergence was proved under b-regularity and s-regularity of the solution. However, this method requires solving a quadratic program, rather than a system of linear equations, as a subproblem to determine a search direction at each iteration.

Yet another NCP-function useful in solving NCP(F) is the function $\phi : R^2 \to R$ given by

$$\phi(a, b) = \sqrt{a^2 + b^2} - a - b. \tag{2}$$

Using this function, we define the function $\Phi : R^n \to R^n$ to form the following system of equations equivalent to NCP(F):

$$\Phi(x) = \begin{pmatrix} \phi(x_1, F_1(x)) \\ \phi(x_2, F_2(x)) \\ \vdots \\ \phi(x_n, F_n(x)) \end{pmatrix} = 0. \tag{3}$$

We also define the merit function $\theta : R^n \to R$ by

$$\theta(x) = \frac{1}{2}\|\Phi(x)\|^2. \tag{4}$$

It is easily seen that θ is differentiable everywhere, though Φ is not differentiable at each x such that $x_i = F_i(x) = 0$ for some i. Recently, a number of algorithms based on the functions Φ and θ have been proposed [3, 4, 6, 8, 13, 21]. Among others, Fischer [6] proposed a Newton-type method based on the function Φ. Because of the nonsmoothness of Φ , this method is a kind of generalized Newton method which uses the generalized Jacobian $\partial\Phi(x)$. To cope with the possible singularity of $H \in \partial\Phi(x)$, a search direction is determined by solving a system of linear equations involving a nonsingular matrix formed by perturbing $H \in \partial\Phi(x)$. So the direction-finding subproblem is a system of linear equations which is always solvable uniquely. This method is quadratically convergent under the nondegeneracy or b-regularity of a solution. Note, however, that this method deals only with the linear complementarity problem. De Luca et al. [3] also proposed a Newton-type method based on Φ. Their method first attempts to use the Newton direction for Φ, but if the Newton equation is unsolvable or the Newton direction is not a direction of sufficient decrease for θ, then it switches to the steepest descent direction for θ. This algorithm was shown to be globally convergent as well as locally quadratically convergent to a stationary point of θ under BD-regularity with respect to Φ. However, since the solvability of the Newton equations is not guaranteed, this algorithm might waste a time, if not very often, in solving unsolvable equations.

In this paper, taking into consideration the preceding work mentioned above, we propose a Newton-type method based on Φ for the monotone NCP(F). In our method, a direction-finding subproblem, which is a system of linear equations, is always solvable uniquely at each iteration. Moreover, the obtained search direction always affords a direction of sufficient decrease for the merit function θ. We show that the algorithm is globally convergent under some mild assumptions. Then by slightly modifying the direction-finding problem, we propose another Newton-type method, which may be considered a restricted version of the first algorithm. For this algorithm, we establish superlinear and quadratical convergence results under appropriate assumptions.

We shall use the following notation. $\|\cdot\|$ denote the Euclidean norm. For vector $x \in R^n$ and an index set $\alpha \subseteq \{1, 2, \ldots, n\}$, x_α represents the vector with elements $x_i, i \in \alpha$. Similarly, F_α denotes the function with component functions $F_i, i \in \alpha$. For an $n \times n$ matrix M with elements

$M_{ij}, i, j = 1, \ldots, n$, and index sets $\alpha, \beta \subseteq \{1, 2, \ldots, n\}$, we denote by $M_{\alpha\beta}$ the $|\alpha| \times |\beta|$ submatrix of M consisting of elements $M_{ij}, i \in \alpha, j \in \beta$. We denote the Jacobian of F at x by $F'(x)$, whereas $\nabla F(x) = F'(x)^T$. For $\alpha, \beta \subseteq \{1, 2, \ldots, n\}$, we denote $\nabla_\alpha F_\beta(x) = M_{\alpha\beta}$, where $M = \nabla F(x)$ and ∇_α stands for the differential operator with respect to x_α. For vector a, $D_a = diag(a)$ denotes the diagonal matrix whose ith diagonal element is a_i. For a set X, we denote the power set of X by $\mathcal{P}(X)$.

2. Preliminaries

In this section, we review some definitions and basic results which will be used in the paper.

Monotonicity and related concepts for vector-valued functions play an important role in nonlinear complementarity problems.

Definition 2.1 **(a)** *$F : R^n \to R^n$ is monotone on a set $S \subseteq R^n$ if*

$$(x - y)^T (F(x) - F(y)) \geq 0 \quad \forall x, y \in S.$$

(b) *$F : R^n \to R^n$ is a P_0-function on S if there exists an index i such that*

$$(x_i - y_i)(F_i(x) - F_i(y)) \geq 0 \quad \forall x, y \in S.$$

(c) *$F : R^n \to R^n$ is a P-function on S if there exists an index i such that*

$$(x_i - y_i)(F_i(x) - F_i(y)) > 0 \quad \forall x, y \in S, x \neq y.$$

(d) *$F : R^n \to R^n$ is a uniform P-function on S if, for some $\mu > 0$, there exists an index i such that*

$$(x_i - y_i)(F_i(x) - F_i(y)) \geq \mu \|x - y\|^2 \quad \forall x, y \in S.$$

Definition 2.2 **(a)** *An $n \times n$ matrix M is a P_0-matrix if every principal minor of M is nonnegative.*

(b) *An $n \times n$ matrix M is a P-matrix if every principal minor of M is positive.*

It is known [16, Theorem 5.8] that if a P_0-function is differentiable, then its Jacobian is a P_0-matrix.

The results stated in the following proposition will be useful in our analysis. Recall that, for a vector a, D_a denotes the diagonal matrix with diagonal elements a_i.

Proposition 2.1 **(a)** *Let M be a P_0-matrix and let D_a and D_b be negative definite diagonal matrices. Then the matrix $D_a + D_b M$ is nonsingular.*

(b) *Let M be a P-matrix and let D_a and D_b be negative semidefinite diagonal matrices such that $a_i \neq 0$ or $b_i \neq 0$ for each i. Then the matrix $D_a + D_b M$ is nonsingular.*

Proof. The proof of (a) is given in [3, Lemma 5.1]. (See also [2, Lemma 5.9.8].) We prove (b). Let $v \in R^n$ be an arbitrary vector such that $(D_a + D_b M)v = 0$. Then, for all i, $a_i v_i + b_i (Mv)_i = 0$. If $a_i = 0$, then $(Mv)_i = 0$ since $b_i \neq 0$ by assumption. If $a_i \neq 0$, then we have

$$v_i = -\frac{b_i}{a_i}(Mv)_i.$$

Hence, it follows from $a_i \leq 0$ and $b_i \leq 0$ that $v_i(Mv)_i \leq 0$ for all i. Since M is a P-matrix, we have $v = 0$ by [2, Theorem 3.3.4]. Therefore the matrix $D_a + D_b M$ is nonsingular. □

Next, we recall some concepts related to the continuity of point-to-set mappings. For more detail, see [11].

Definition 2.3 *Let X be a subset of R^n and Θ be a point-to-set mapping from X into $\mathcal{P}(X)$.*

(a) *Θ is uniformly compact near $\bar{x} \in X$ if there exists a neighborhood N of $\bar{x}$ such that the closure of $\bigcup_{x \in N} \Theta(x)$ is compact.*

(b) *Θ is closed at $\bar{x}$ if $x^k \to \bar{x}$, $y^k \in \Theta(x^k)$ for all k and $y^k \to \bar{y}$ imply $\bar{y} \in \Theta(\bar{x})$.*

(c) *Θ is upper semi-continuous at $\bar{x}$ if Θ is uniformly compact near $\bar{x}$ and closed at $\bar{x}$.*

The following concepts of generalized differentiability have been introduced in [1, 20]. Recall that any locally Lipschitz continuous function is differentiable almost everywhere in R^n.

Definition 2.4 *Let $\Phi : R^n \to R^n$ be locally Lipschitz continuous and let $\mathcal{D}_\Phi \subseteq R^n$ be the set of points where Φ is differentiable.*

(a) *The B-subdifferential of Φ at x is the set*

$$\partial_B \Phi(x) = \{y \in R^n \mid y = \lim_{\substack{x^k \in \mathcal{D}_\Phi \\ x^k \to x}} \Phi'(x^k)\}.$$

(b) *The Clarke generalized Jacobian of Φ at x is the set*

$$\partial \Phi(x) = \mathrm{co}\partial_B \Phi(x),$$

where co *denotes the convex hull of a set.*

It is known [1] that $\partial_B \Phi(x)$ and $\partial \Phi(x)$ are nonempty and compact for each x. Moreover, it is easily seen that $\partial_B \Phi : R^n \to \mathcal{P}(R^n)$ and $\partial \Phi : R^n \to \mathcal{P}(R^n)$ are upper semi-continuous point-to-set mappings (see, e.g., [1]).

Semismoothness and strong semismoothness of vector functions were introduced in [22].

Definition 2.5 *Let $\Phi : R^n \to R^n$ be locally Lipschitz continuous.*

(a) *The function Φ is said to be semismooth at x if*

$$\lim_{\substack{H \in \partial\Phi(x+tv') \\ v' \to v, t \downarrow 0}} Hv'$$

exists for all $v \in R^n$.

(b) *Φ is said to be strongly semismooth at x if Φ is semismooth at x and, for any $d \to 0$ and for any $H \in \partial\Phi(x+d)$,*

$$Hd - F'(x;d) = O(\|d\|^2).$$

Semismoothness plays a fundamental role in the convergence analysis of Newton-type methods for nonsmooth equations. In particular, Qi [20] showed that a generalized Newton method has a superlinear rate of convergence if the function involved is semismooth and if every element of its B-subdifferential is nonsingular at a solution.

For a solution x^* of NCP(F), we introduce the following three index sets:

$$\alpha := \{i \mid x_i^* > 0,\ F_i(x^*) = 0\}, \quad \beta := \{i \mid x_i^* = 0 = F_i(x^*)\}, \quad \gamma := \{i \mid x_i^* = 0,\ F_i(x^*) > 0\}.$$

Using these sets, we may define various regularity concepts of a solution of NCP(F).

Definition 2.6 *Let x^* be a solution of NCP(F).*

(a) *A solution x^* is said to be nondegenerate if $\beta = \emptyset$.*

(b) *x^* is said to be b-regular if, for any index set δ such that $\alpha \subseteq \delta \subseteq \alpha \cup \beta$, $\nabla_\delta F_\delta(x^*)$ is nonsingular.*

(c) *x^* is said to be R-regular if $\nabla_\alpha F_\alpha(x^*)$ is nonsingular and its Schur complement in*

$$\begin{pmatrix} \nabla_\alpha F_\alpha(x^*) & \nabla_\alpha F_\beta(x^*) \\ \nabla_\beta F_\alpha(x^*) & \nabla_\beta F_\beta(x^*) \end{pmatrix}$$

is a P-matrix.

(d) *x^* is said to be BD-regular with respect to Φ if any $H \in \partial_B\Phi(x^*)$ is nonsingular.*

Note that BD-regularity is a concept associated with the particular function Φ, rather than the index sets α, β, γ. The last three regularities in Definition 2.6 were introduced in [19], [23] and [20], respectively.

Next, we review some properties of the functions Φ and θ defined by (3) and (4), respectively. It is readily seen from the definition that Φ is not differentiable at each x such that $x_i = 0 = F_i(x)$ for

some i. However, since Φ is locally Lipschitz continuous, the B-subdifferential $\partial_B\Phi(x)$ is nonempty at any $x \in R^n$. Unfortunately, it is not necessarily easy to specify the set $\partial_B\Phi(x)$ exactly at x where $\nabla\Phi(x)$ does not exist. As a remedy for this intractability, we construct two point-to-set mappings $\tilde{\mathcal{H}}$ and $\hat{\mathcal{H}}$ which approximate $\partial_B\Phi$ and are easy to evaluate in practice. First, we define the mapping $\tilde{\mathcal{H}} : R^n \to \mathcal{P}(R^{n\times n})$ by

$$\tilde{\mathcal{H}}(x) = \{\tilde{H} \in R^{n\times n} \mid \tilde{H} = D_{\tilde{a}} + D_{\tilde{b}}F'(x), (\tilde{a}, \tilde{b}) \in \tilde{\Omega}(x)\}, \tag{5}$$

where $\tilde{\Omega} : R^n \to \mathcal{P}(R^{2n})$ is given by

$$\tilde{\Omega}(x) = \{(\tilde{a}, \tilde{b}) \in R^{2n} \mid (\tilde{a}_i, \tilde{b}_i) \in \tilde{\Omega}_i(x), \quad i = 1, 2, \ldots, n\} \tag{6}$$

with $\tilde{\Omega}_i : R^n \to \mathcal{P}(R^2), i = 1, 2, \ldots, n$, given by

$$\tilde{\Omega}_i(x) = \begin{cases} \left\{(\tilde{a}_i, \tilde{b}_i) \in R^2 \;\middle|\; (\tilde{a}_i + 1)^2 + (\tilde{b}_i + 1)^2 = 1\right\} & \text{if} \quad x_i = 0 = F_i(x), \\ \left\{(\tilde{a}_i, \tilde{b}_i) \in R^2 \;\middle|\; \tilde{a}_i = \frac{x_i}{\sqrt{x_i^2 + F_i(x)^2}} - 1, \;\; \tilde{b}_i = \frac{F_i(x)}{\sqrt{x_i^2 + F_i(x)^2}} - 1\right\} & \text{otherwise.} \end{cases} \tag{7}$$

Next, define the mapping $\hat{\mathcal{H}} : R^n \to \mathcal{P}(R^{n\times n})$ by

$$\hat{\mathcal{H}}(x) = \{\hat{H} \in R^{n\times n} \mid D_{\hat{a}} + D_{\hat{b}}F'(x), (\hat{a}, \hat{b}) \in \hat{\Omega}(x)\}, \tag{8}$$

where $\hat{\Omega} : R^n \to \mathcal{P}(R^{2n})$ is given by

$$\hat{\Omega}(x) \;=\; \bigcup_{z\in Z(x)} \left\{ (\hat{a}, \hat{b}) \in R^{2n} \;\middle|\; (\hat{a}_i, \hat{b}_i) = \begin{cases} \left(\frac{z_i}{\sqrt{z_i^2 + (\nabla F_i(x)^T z)^2}} - 1, \frac{\nabla F_i(x)^T z}{\sqrt{z_i^2 + (\nabla F_i(x)^T z)^2}} - 1\right) \\ \qquad \text{if} \quad x_i = 0 = F_i(x), \\ \left(\frac{x_i}{\sqrt{x_i^2 + F_i(x)^2}} - 1, \;\; \frac{F_i(x)}{\sqrt{x_i^2 + F_i(x)^2}} - 1\right) \\ \qquad \text{otherwise} \end{cases} \right\} \tag{9}$$

with $Z : R^n \to \mathcal{P}(R^n)$ given by

$$Z(x) = \{z \in R^n \mid z_i \neq 0 \;\; \text{if} \;\; i \in \beta\}.$$

The idea of using $\hat{\Omega}(x)$ to find a particular element of $\partial_B\Phi(x)$ has been used by De Luca et al. [3]. We should point out that, for every $x \in R^n$, the elements $(\tilde{a}_i, \tilde{b}_i)$ of any vector $(\tilde{a}, \tilde{b}) \in \tilde{\Omega}(x)$ satisfy $-2 \leq \tilde{a}_i, \tilde{b}_i \leq 0$ and $\tilde{a}_i$ and $\tilde{b}_i$ do not vanish simultaneously. Likewise, for every $x \in R^n$, the elements $(\hat{a}_i, \hat{b}_i)$ of any vector $(\hat{a}, \hat{b}) \in \hat{\Omega}(x)$ satisfy $-2 \leq \hat{a}_i, \hat{b}_i \leq 0$ and $\hat{a}_i$ and $\hat{b}_i$ do not vanish simultaneously.

The point-to-set mappings $\tilde{\mathcal{H}}$ and $\hat{\mathcal{H}}$ have the following properties.

Proposition 2.2 (a) *For every $x \in R^n$, we have*

$$\hat{\mathcal{H}}(x) \subseteq \partial_B\Phi(x) \subseteq \tilde{\mathcal{H}}(x).$$

(b) *For any $x \in R^n$, $\tilde{\mathcal{H}}$ is upper semi-continuous at x.*

Proof. The left inclusion in (a) is shown in [3, Theorem 7.1]. We show the right inclusion in (a). Suppose that Φ is not differentiable at x. Then each element of $\partial_B \Phi(x)$ is given as a limit $\lim_{k\to\infty} \Phi'(x^k) = \lim_{k\to\infty} D_{\tilde{a}^k} + D_{\tilde{b}^k} F'(x^k)$, where $\{x^k\}$ is an arbitrary sequence in $\mathcal{D}_\Phi$ such that $x^k \to x$ and $\{\Phi'(x^k)\}$ is convergent. Since Φ is differentiable at each x^k, we have $\partial_B \Phi(x^k) = \tilde{\mathcal{H}}(x^k) = \{\Phi'(x^k)\}$ and

$$\Phi'(x^k) = D_{\tilde{a}^k} + D_{\tilde{b}^k} F'(x^k),$$

where

$$(\tilde{a}_i^k, \tilde{b}_i^k) = \left(\frac{x_i^k}{\sqrt{(x_i^k)^2 + F_i(x^k)^2}} - 1, \quad \frac{F_i(x^k)}{\sqrt{(x_i^k)^2 + F_i(x^k)^2}} - 1 \right).$$

Therefore the equality

$$(\tilde{a}_i^k + 1)^2 + (\tilde{b}_i^k + 1)^2 = 1$$

is satisfied for each k, and hence the same equality is satisfied in the limit. Since F' is continuous, this implies $\partial_B \Phi(x) \subseteq \tilde{\mathcal{H}}(x)$.

Next we prove (b). Since $\tilde{\Omega}_i(x), i = 1, \ldots, n$, are contained in a sphere of radius 1 and since F' is continuous, $\tilde{\mathcal{H}}$ is uniformly compact near every x. The closedness of $\tilde{\mathcal{H}}$ can be shown by a reasoning similar to the one used to prove (a). It then follows that $\tilde{\mathcal{H}}$ is upper semi-continuous. □

An inclusion similar to the right inclusion in Proposition 2.2 (a) has been shown in [4], where $(\tilde{a}_i + 1)^2 + (\tilde{b}_i + 1)^2 = 1$ in (7) is replaced with the inequality $(\tilde{a}_i + 1)^2 + (\tilde{b}_i + 1)^2 \leq 1$ and $\partial_B \Phi$ is replaced with the generalized Jacobian $\partial \Phi$.

In this paper, we shall present two algorithms. The first one, which will be presented in Section 3, uses matrices obtained by modifying $\tilde{H} \in \tilde{\mathcal{H}}(x)$. We will establish its global convergence. In Section 4, we will present another algorithm that uses matrices obtained by modifying $\hat{H} \in \hat{\mathcal{H}}(x)$. As it turns out, the second algorithm is a restricted version of the first one, so that it retains a global convergence property. We will in particular show that this algorithm has a superlinear rate of convergence.

The next proposition summarizes some important properties of the function Φ.

Proposition 2.3 **(a)** *If F is Lipschitz continuous near x, so is Φ. If F is semismooth at x, so is Φ.*

(b) *If F is twice differentiable with Lipschitz continuous Hessian at x, then Φ is strongly semismooth at x.*

(c) *If a solution x^* of NCP(F) is R-regular, then every $H \in \partial \Phi(x^*)$ is nonsingular, and hence x^* is BD-regular with respect to Φ.*

(d) *Let x^* be a solution of NCP(F) and denote $\bar{\gamma} = \alpha \cup \beta$. If $\nabla_{\bar{\gamma}} F_{\bar{\gamma}}(x^*)$ is a P-matrix, then every $H \in \tilde{\mathcal{H}}(x^*)$ is nonsingular, and hence x^* is BD-regular with respect to Φ.*

Proof. Parts (a) and (c) are proven in [4], and (b) is proven in [3]. We prove (d), by showing that for any $\tilde{H} \in \tilde{\mathcal{H}}(x^*)$, $v = 0$ is the unique solution of $\tilde{H}v = 0$. Since $\partial_B \Phi(x^*) \subseteq \tilde{\mathcal{H}}(x^*)$ by Proposition 2.2 (a), the last part then follows from the first part. To simplify the notation, we denote $F'(x^*)$ by M. By the definition (5) of mapping $\tilde{\mathcal{H}}$, we can write $\tilde{H} = D_{\tilde{a}} + D_{\tilde{b}} M$ for some $(\tilde{a}, \tilde{b}) \in \tilde{\Omega}(x^*)$. So, we have $\tilde{H}v = (D_{\tilde{a}} + D_{\tilde{b}} M)v = 0$. This implies that for every i, $\tilde{a}_i v_i + \tilde{b}_i (Mv)_i = 0$. For each $i \in \gamma$, we have $(\tilde{a}_i, \tilde{b}_i) = (-1, 0)$, and hence $v_i = 0$. It remains to show $v_{\bar{\gamma}} = 0$. Since $(D_{\tilde{a}} + D_{\tilde{b}} M)v = 0$ and $v_\gamma = 0$ imply

$$(D_{\tilde{a}_{\bar{\gamma}}} + D_{\tilde{b}_{\bar{\gamma}}} M_{\bar{\gamma}\bar{\gamma}}) v_{\bar{\gamma}} = 0,$$

it is sufficient to show that $D_{\tilde{a}_{\bar{\gamma}}} + D_{\tilde{b}_{\bar{\gamma}}} M_{\bar{\gamma}\bar{\gamma}}$ is nonsingular. By the assumption, $M_{\bar{\gamma}\bar{\gamma}}$ is a P-matrix. Moreover, by (7), we must have $\tilde{a}_i, \tilde{b}_i \leq 0$ with $\tilde{a}_i \neq 0$ or $\tilde{b}_i \neq 0$ for each i. The desired result then follows from Proposition 2.1 (b). □

Note that, for the monotone NCP(F), the condition in Proposition 2.3 (d) is satisfied if x^* is b-regular.

The next proposition shows that $\|\Phi(x)\|$ provides a local error bound for NCP(F).

Proposition 2.4 *Let x^* be a solution of NCP(F). Then, for a sufficiently small $\omega > 0$, there exists a constant $c_1 > 0$ such that $\|x - x^*\| \leq \omega$ implies*

$$\|\Phi(x)\| \leq c_1 \|x - x^*\|. \tag{10}$$

Moreover, if x^* is BD-regular with respect to Φ, then there exists a constant $c_2 > 0$ such that $\|x - x^*\| \leq \omega$ implies

$$c_2 \|x - x^*\| \leq \|\Phi(x)\|. \tag{11}$$

Proof. First, we prove (10). By the differentiability, F is Lipschitz continuous near x^*, so that Φ is also Lipschitz continuous near x^* by Proposition 2.3 (a). Let c_1 be a Lipschitz constant of Φ near x^*. Then we have

$$\|\Phi(x)\| = \|\Phi(x) - \Phi(x^*)\| \leq c_1 \|x - x^*\|,$$

provided that $\|x - x^*\|$ is sufficiently small. This shows (10). Next, let us define an affine function $\hat{F} : R^n \to R^n$ by

$$\hat{F}(x) = F(x^*) + F'(x^*)(x - x^*).$$

Since F is continuously differentiable, we have

$$\|F(x) - \hat{F}(x)\| = o(\|x - x^*\|).$$

From this, it is not difficult to see

$$\|\Phi(x) - \hat{\Phi}(x)\| \leq o(\|x - x^*\|),$$

where $\hat{\Phi}(x) = \Big(\phi(x_1, \hat{F}_1(x)), \ldots, \phi(x_n, \hat{F}_n(x))\Big)^T$. Moreover, it follows from the definition of Φ that x^* is also BD-regular with respect to $\hat{\Phi}$. Hence, by [20, Proposition 2.5], x^* is a locally unique solution of NCP($\hat{F}$), too. By [6, Lemma 4.2], for sufficiently small ω, there exists a constant $c' > 0$ such that

$$c'\|x - x^*\| \leq \|\hat{\Phi}(x)\|,$$

whenever $\|x - x^*\| \leq \omega$. Therefore, we have

$$\begin{aligned} \|\Phi(x)\| &\geq \|\hat{\Phi}(x)\| - \|\Phi(x) - \hat{\Phi}(x)\| \\ &\geq c'\|x - x^*\| - o(\|x - x^*\|). \end{aligned}$$

This shows the existence of a constant $c_2 > 0$ satisfying (11). □

Next, we state some properties of the function θ.

Proposition 2.5 *[4, 13]*

(a) *If F is differentiable, then θ is also differentiable and, for any $\tilde{H} \in \tilde{\mathcal{H}}(x)$,*

$$\nabla\theta(x) = \tilde{H}^T\Phi(x). \tag{12}$$

(b) *$\theta(x) \geq 0$ for all $x \in R^n$, and $\theta(x) = 0$ if and only if x is a solution of NCP(F).*

(c) *If F is a P_0-function, then any stationary point of θ is a solution of NCP(F).*

(d) *If F is a uniform P-function, then the level sets of θ are compact.* □

Proposition 2.5 (a) shows if F is differentiable, θ is everywhere differentiable, in spite of the fact that Φ is nondifferentiable at some points. In [4], (12) is shown with $\tilde{H}$ being replaced by $H \in \partial\Phi(x)$. It is easy to show that (12) is also valid for any $\tilde{H} \in \tilde{\mathcal{H}}(x)$, since $\Phi_i(x) = 0$ if $x_i = 0 = F_i(x)$, and since the ith column of $\tilde{H} \in \tilde{\mathcal{H}}(x)$ differs from that of $H \in \partial_B\Phi(x)$ only if $\Phi_i(x) = 0$. By Proposition 2.5 (c), if F is monotone, then any stationary point of θ is a solution of NCP(F), since the monotonicity of F implies that F is a P_0-function. Recently, it has been shown [3] that under a weaker assumption than that of (c), any stationary point of θ is a solution of NCP(F).

3. Globally convergent Newton-type algorithm

In the remainder of this paper, we shall assume that F is monotone. In this section, we propose a globally convergent Newton-type algorithm for the monotone NCP(F).

By analogy with Newton method, it is reasonable to consider the following system of linear equations as a direction-finding subproblem at each x^k:

$$\tilde{H}d = -\Phi(x^k),$$

where $\tilde{H}$ is a matrix chosen from $\tilde{\mathcal{H}}(x^k)$ defined by (5). Unfortunately, the monotonicity assumption on F does not necessarily imply that $\tilde{H}$ is nonsingular. So we slightly perturb $\tilde{H}$ to obtain a nonsingular matrix $\tilde{G}$, and determine a search direction by solving the system of linear equations

$$\tilde{G}d = -\Phi(x^k).$$

Specifically the matrix $\tilde{G}$ is constructed from $\tilde{H}$ as follows: First, we define point-to-set mappings $\Lambda_i : R^{n+2} \to \mathcal{P}(R^2), i = 1, 2, \ldots, n$, by

$$\begin{aligned}
&\Lambda_i(x, a_i, b_i) \\
&= \begin{cases}
\left\{(\bar{a}_i, \bar{b}_i) \in R^2 \;\middle|\; \bar{a}_i = \frac{\sigma(\theta(x))}{b_i},\ \bar{b}_i = 0\right\} & \text{if} \quad -\varepsilon < a_i \text{ and } b_i \le -\varepsilon, \\
\left\{(\bar{a}_i, \bar{b}_i) \in R^2 \;\middle|\; \begin{array}{l} \bar{a}_i = \tau\frac{\sigma(\theta(x))}{b_i}, \\ \bar{b}_i = (1-\tau)\frac{\sigma(\theta(x))}{a_i}, \end{array} \ \tau \in [0,1]\right\} & \text{if} \quad a_i \le -\varepsilon \text{ and } b_i \le -\varepsilon, \\
\left\{(\bar{a}_i, \bar{b}_i) \in R^2 \;\middle|\; \bar{a}_i = 0,\ \bar{b}_i = \frac{\sigma(\theta(x))}{a_i}\right\} & \text{if} \quad a_i \le -\varepsilon \text{ and } -\varepsilon < b_i,
\end{cases}
\end{aligned} \tag{13}$$

where $\varepsilon \in (0, 1 - \frac{\sqrt{2}}{2})$ and $\sigma : [0, +\infty) \to [0, +\infty)$ is a nondecreasing continuous function such that $\sigma(0) = 0$ and $\sigma(t) > 0$ for all $t > 0$. By easy calculations, we can show that for $(a, b) \in \bar{\Omega}(x)$, the case where $-\varepsilon < a_i$ and $-\varepsilon < b_i$ cannot happen whenever $\varepsilon \in (0, 1 - \frac{\sqrt{2}}{2})$. So, this case has been ignored in (13).

By using the mappings Λ_i and $\tilde{\Omega}$ defined by (13) and (6), respectively, we form a matrix $\tilde{G}$ as

$$\tilde{G} = D_{\tilde{p}} + D_{\tilde{q}}F'(x), \tag{14}$$

where $\tilde{p}$ and $\tilde{q}$ are vectors such that

$$(\tilde{p}_i, \tilde{q}_i) = (\tilde{a}_i + \bar{a}_i, \tilde{b}_i + \bar{b}_i), \quad i = 1, 2, \ldots, n, \tag{15}$$

with $(\tilde{a}, \tilde{b}) \in \tilde{\Omega}(x)$ and $(\bar{a}_i, \bar{b}_i) \in \Lambda_i(x, \tilde{a}_i, \tilde{b}_i), i = 1, 2, \ldots, n$. If $\theta(x) > 0$, (13) and (15) imply that both $D_{\tilde{p}}$ and $D_{\tilde{q}}$ are negative definite matrices. Define a point-to-set mapping $\tilde{\mathcal{G}} : R^n \to \mathcal{P}(R^{n\times n})$ by

$$\begin{aligned}
\tilde{\mathcal{G}}(x) = \{\tilde{G} \in R^{n\times n} \mid \tilde{G} = D_{\tilde{p}} + D_{\tilde{q}}F'(x), (\tilde{p}, \tilde{q}) \text{ is given by (15) with } (\tilde{a}, \tilde{b}) \in \tilde{\Omega}(x) \\
\text{and } (\bar{a}_i, \bar{b}_i) \in \Lambda_i(x, \tilde{a}_i, \tilde{b}_i) \text{ for } i = 1, 2, \ldots n\}.
\end{aligned} \tag{16}$$

Then (14) means that $\tilde{G}$ is chosen from the set $\tilde{\mathcal{G}}(x)$. In the following, for $\tilde{G} = D_{\tilde{p}} + D_{\tilde{q}}F'(x) \in \tilde{\mathcal{G}}(x)$ and $\tilde{H} = D_{\tilde{a}} + D_{\tilde{b}}F'(x) \in \tilde{\mathcal{H}}(x)$, we say that $\tilde{G}$ corresponds to $\tilde{H}$, if $(\tilde{a}, \tilde{b})$ is related to $(\tilde{p}, \tilde{q})$ by (13) and (15).

The mapping $\tilde{\mathcal{G}}$ has the following property which is fundamental to our algorithm.

Proposition 3.1 *If x is not a solution of NCP(F), i.e., $\theta(x) > 0$, then every $\tilde{G} \in \tilde{\mathcal{G}}(x)$ is nonsingular.*

Proof. By the definition of $\tilde{G}$, if $\theta(x) > 0$, then $D_{\tilde{p}}$ and $D_{\tilde{q}}$ are negative definite diagonal matrices. Since F is monotone, $F'(x^k)$ is positive semidefinite, implying that $F'(x^k)$ is a P_0-matrix. Hence, by Proposition 2.1 (a), $\tilde{G}$ is nonsingular. □

Using the mapping $\tilde{\mathcal{G}}$, we define a point-to-set mapping $\tilde{\Delta} : R^n \to \mathcal{P}(R^n)$ by

$$\tilde{\Delta}(x) = \{d \in R^n \mid \tilde{G}d = -\Phi(x), \;\; \tilde{G} \in \tilde{\mathcal{G}}(x)\}. \tag{17}$$

By Proposition 3.1, $\tilde{\Delta}(x)$ is nonempty for each x such that $\theta(x) > 0$.

Now we are ready to state our first algorithm.

Algorithm 1.

Step 1: Select an initial point $x^0 \in R^n$ and parameters $\rho \in (0, \frac{1}{2})$ and $\beta \in (0, 1)$. Set $k := 0$.

Step 2: If $\theta(x^k) = 0$, then stop. Otherwise, go to Step 3.

Step 3: Find a vector $d^k \in \tilde{\Delta}(x^k)$.

Step 4: Let m be the smallest nonnegative integer such that

$$\theta(x^k + \beta^m d^k) - \theta(x^k) \leq \beta^m \rho \nabla\theta(x^k)^T d^k$$

and set $x^{k+1} := x^k + t_k d^k$, where $t_k := \beta^m$. Return to Step 2 with k replaced by $k + 1$.

The rest of this section is devoted to show that Algorithm 1 is globally convergent to a solution of NCP(F). First, we show that every $d \in \tilde{\Delta}(x)$ is a descent direction of θ at x. To this end, we need the following lemma.

Lemma 3.1 *If $\tilde{H} \in \tilde{\mathcal{H}}(x)$ corresponds to $\tilde{G} \in \tilde{\mathcal{G}}(x)$, then $(\tilde{G} - \tilde{H})^T \tilde{H}$ is positive semidefinite.*

Proof. To simplify the notation, let us denote $M = F'(x)$ $\sigma = \sigma(\theta(x))$. Suppose that $(\bar{a}, \bar{b})$ is a vector which makes $\tilde{G} \in \tilde{\mathcal{G}}(x)$ correspond to $\tilde{H} \in \tilde{\mathcal{H}}(x)$ via (13) and (15). Then, since $(\bar{a}_i, \bar{b}_i) \in \Lambda_i(x, \tilde{a}_i, \tilde{b}_i)$ and $\tilde{G} - \tilde{H} = D_{\bar{a}} + D_{\bar{b}}M$, we have

$$\begin{aligned} (\tilde{G} - \tilde{H})^T \tilde{H} &= (D_{\bar{a}} + D_{\bar{b}}M)^T (D_{\tilde{a}} + D_{\tilde{b}}M) \\ &= D_{\bar{a}}D_{\tilde{a}} + D_{\bar{a}}D_{\tilde{b}}M + M^T D_{\bar{b}}D_{\tilde{a}} + M^T D_{\bar{b}}D_{\tilde{b}}M. \end{aligned}$$

Since $\tilde{a}_i \leq 0, \tilde{b}_i \leq 0, \bar{a}_i \leq 0$ and $\bar{b}_i \leq 0$ for every i, both $D_{\bar{a}}D_{\tilde{a}}$ and $M^T D_{\bar{b}} D_{\tilde{b}} M$ are positive semidefinite. Moreover, we have

$$v^T(D_{\bar{a}}D_{\tilde{b}}M + M^T D_{\bar{b}}D_{\tilde{a}})v = v^T(D_{\bar{a}}D_{\tilde{b}}M + D_{\tilde{a}}D_{\bar{b}}M)v,$$

for $v \in R^n$. Therefore, it is sufficient to show that $(D_{\bar{a}}D_{\tilde{b}} + D_{\tilde{a}}D_{\bar{b}})M$ is positive semidefinite. Since $(\bar{a}_i, \bar{b}_i) \in \Lambda_i(x, \tilde{a}_i, \tilde{b}_i)$, it follows from (13) that

$$(\bar{a}_i\tilde{b}_i, \tilde{a}_i\bar{b}_i) = \begin{cases} (\sigma, 0) & \text{if} \quad -\varepsilon < a_i \text{ and } b_i \leq -\varepsilon, \\ (\tau_i\sigma, (1-\tau_i)\sigma) & \text{if} \quad a_i \leq -\varepsilon \text{ and } b_i \leq -\varepsilon, \\ (0, \sigma) & \text{if} \quad a_i \leq -\varepsilon \text{ and } -\varepsilon < b_i, \end{cases}$$

where $\tau_i \in [0, 1]$. Hence, we have $\bar{a}_i\tilde{b}_i + \tilde{a}_i\bar{b}_i = \sigma$ for every i, and hence

$$D_{\bar{a}}D_{\tilde{b}} + D_{\tilde{a}}D_{\bar{b}} = \sigma I.$$

It then follows from $\sigma \geq 0$ and the positive semidefiniteness of M that $(D_{\bar{a}}D_{\tilde{b}} + D_{\tilde{a}}D_{\bar{b}})M$ is positive semidefinite. This completes the proof. □

The following lemma establishes a descent property of vectors $d \in \tilde{\Delta}(x)$.

Lemma 3.2 *If x is not a solution of NCP(F), then every $d \in \tilde{\Delta}(x)$ satisfies the descent condition for θ*

$$\nabla\theta(x)^T d < 0.$$

Proof. To simplify the notation, we denote $M = F'(x)$, $\Phi = \Phi(x)$, $\nabla\theta = \nabla\theta(x)$ and $F_i = F_i(x)$. First, we prove $\nabla\theta^T d \leq 0$ for every $d \in \tilde{\Delta}(x)$. Choose an arbitrary $d \in \tilde{\Delta}(x)$ and let $\tilde{G} \in \tilde{\mathcal{G}}(x)$ be a matrix such that $\tilde{G}d = -\Phi$. By Proposition 2.5 (a), $\nabla\theta = \tilde{H}^T\Phi$ for any $\tilde{H} \in \tilde{\mathcal{H}}(x)$. Hence we have

$$\begin{aligned} \nabla\theta^T d &= \Phi^T\tilde{H}d \\ &= -d^T\tilde{G}^T\tilde{H}d \\ &= -d^T\tilde{H}^T\tilde{H}d - d^T(\tilde{G} - \tilde{H})^T\tilde{H}d. \end{aligned} \tag{18}$$

Since $\tilde{H}$ is an arbitrarily element of $\tilde{\mathcal{H}}(x)$, we may let $\tilde{H}$ correspond to $\tilde{G}$. It then follows from Lemma 3.1 and (18) that $\nabla\theta^T d \leq 0$.

To complete the proof, we next show that if $\nabla\theta^T d = 0$ holds for some $d \in \tilde{\Delta}(x)$, then x is a solution of NCP(F). Suppose to the contrary that $\nabla\theta^T d = 0$ for some $d \in \tilde{\Delta}(x)$ but x is not a solution of NCP(F). Note that $d \neq 0$, because $\Phi \neq 0$ and, by the definition of d, there exists $\tilde{G} \in \tilde{\mathcal{G}}(x)$ such that $d = -\tilde{G}^{-1}\Phi$. Let us define the index set I_s by

$$I_s = \{i \mid \quad x_i \geq 0,\ F_i \geq 0,\ x_iF_i = 0\} = \{i \mid \quad \phi(x_i, F_i) = 0\}.$$

Furthermore, we define vectors $\check{a}$ and $\check{b}$ by

$$(\check{a}_i, \check{b}_i) = \begin{cases} (\tilde{p}_i, \tilde{q}_i) & \text{if} \quad i \in I_s, \\ (\tilde{a}_i, \tilde{b}_i) & \text{if} \quad i \notin I_s, \end{cases}$$

where $(\tilde{a}, \tilde{b})$ and $(\tilde{p}, \tilde{q})$ are defined by (6), (13) and (15). By the definition, $\tilde{p}_i$ and $\tilde{q}_i$ are negative for each i. For $i \notin I_s$, $\tilde{a}_i$ and $\tilde{b}_i$ are also negative by (6) and (7). It follows that the diagonal matrices $D_{\check{a}}$ and $D_{\check{b}}$ are negative definite. Using $D_{\check{a}}$ and $D_{\check{b}}$, let $\check{H}$ be defined by

$$\check{H} = D_{\check{a}} + D_{\check{b}} M.$$

The matrix $\check{H}$ is nonsingular by Proposition 2.1 (a).

Now let us evaluate $\check{H}d$. First, we consider the ith element $(\check{H}d)_i$ of $\check{H}d$ for $i \in I_s$. Since $\tilde{G}d = -\Phi$, and since $(\check{H}d)_i = (\tilde{G}d)_i$ and $\Phi_i = 0$ for $i \in I_s$, we have

$$(\check{H}d)_i = 0, \qquad i \in I_s. \tag{19}$$

Next, we consider $(\check{H}d)_i$ for $i \notin I_s$. By the positive semidefiniteness of $(\tilde{G} - \tilde{H})^T \tilde{H}$, it follows from (18) that $\nabla\theta^T d = 0$ implies $\tilde{H}d = 0$. Therefore, since $(\check{H}d)_i = (\tilde{H}d)_i$ for $i \notin I_s$, we have

$$(\check{H}d)_i = 0, \qquad i \notin I_s. \tag{20}$$

By (19) and (20), we have

$$\check{H}d = 0.$$

Since $\check{H}$ is nonsingular as shown above, we have $d = 0$. This contradict $d \neq 0$. The proof is complete. □

Next we show that the point-to-set mapping $\tilde{\mathcal{G}}$ defined by (16) is upper semi-continuous at any $x \in R^n$ and the point-to-set mapping $\tilde{\Delta}$ defined by (17) is upper semi-continuous at x which is not a solution of NCP(F).

Lemma 3.3 *The point-to-set mapping $\tilde{\mathcal{G}}$ is upper semi-continuous at any $x \in R^n$.*

Proof. To begin with, we show that $\tilde{\mathcal{G}}$ is uniformly compact near any $x \in R^n$. If not, there exists a sequence $\{\tilde{G}_k\}$ such that for a sequence $\{y^k\}$ converging to x, $\tilde{G}_k \in \tilde{\mathcal{G}}(y^k)$ and $\|\tilde{G}_k\| \to \infty$. By the definition of $\tilde{G}_k$, however, we then have $\sigma(\theta(y^k)) \to \infty$ or $\|F'(y^k)\| \to \infty$. This contradicts the assumption that σ, θ and F' are continuous.

Next we show that $\tilde{\mathcal{G}}$ is closed at any $x \in R^n$. By the definition (16) of $\tilde{\mathcal{G}}$, it is sufficient to show that the point-to-set mapping $\bigcup_{(\tilde{a}_i, \tilde{b}_i) \in \tilde{\Omega}_i(\cdot)} \Lambda_i(\cdot, \tilde{a}_i, \tilde{b}_i) : R^n \to \mathcal{P}(R^2)$ is closed. It is easy to see that the mapping $\tilde{\Omega}_i$ defined by (7) is upper semi-continuous at any x, and Λ_i defined by (13) is closed at any $(x, a_i, b_i) \in R^{n+2}$. The closedness of the mapping $\bigcup_{(a_i, b_i) \in \Omega_i(\cdot)} \Lambda_i(\cdot, a_i, b_i)$ can then be deduced from [11, Proposition 7]. □

Lemma 3.4 *Let x be an arbitrary point such that $\theta(x) > 0$. Then, the point-to-set mapping $\tilde{\Delta}$ is upper semi-continuous at x.*

Proof. To show that $\tilde{\Delta}$ is uniformly compact near x such that $\theta(x) > 0$, we assume to the contrary that there exists a sequence $\{d^k\}$ such that for a sequence $\{y^k\}$ converging to x, $d^k \in \tilde{\Delta}(y^k)$ and $\|d^k\| \to \infty$. By the definition of $\tilde{\Delta}$, there exists $\tilde{G}_k \in \tilde{\mathcal{G}}(y^k)$ such that $\tilde{G}_k d^k = -\Phi(y^k)$. Since Φ is continuous and $\|d^k\| \to \infty$, we must have $\|\tilde{G}_k^{-1}\| \to \infty$. Furthermore, since $\tilde{\mathcal{G}}$ is upper semi-continuous by Lemma 3.3, any accumulation point $\tilde{G}_*$ of $\{\tilde{G}_k\}$ belongs to $\hat{\mathcal{G}}(x)$. Since $\|\tilde{G}_k^{-1}\| \to \infty$, $\tilde{G}_*$ must be singular. On the other hand, by $\theta(x) > 0$ and Proposition 3.1, every $\tilde{G} \in \tilde{\mathcal{G}}(x)$ is nonsingular. This is a contradiction.

Next we show that $\tilde{\Delta}$ is closed at x. Let $\{y^k\}$ and $\{d^k\}$ be arbitrary sequences such that $y^k \to x$, $d^k \in \tilde{\Delta}(y^k)$ and $d^k \to d^*$. We need to show $d^* \in \tilde{\Delta}(x)$. By the definition of $\tilde{\Delta}$, there exists $\tilde{G}_k \in \tilde{\mathcal{G}}(y^k)$ such that $\tilde{G}_k d^k = -\Phi(y^k)$. By Lemma 3.3, $\tilde{\mathcal{G}}$ is upper semi-continuous at x, which implies that $\{\tilde{G}_k\}$ is bounded. Let $\tilde{G}_*$ be an accumulation point of $\{\tilde{G}_k\}$. Then, again by the upper semi-continuity of $\tilde{\mathcal{G}}$, we have $\tilde{G}_* \in \tilde{\mathcal{G}}(x)$, and by Proposition 3.1, $\tilde{G}_*$ is nonsingular. Since $\Phi(y^k) \to \Phi(x)$, we have $d^* = -\tilde{G}_*^{-1}\Phi(x) \in \tilde{\Delta}(x)$. □

Using the preceding results, we can prove the following global convergence theorem.

Theorem 3.1 *Any accumulation point of the sequence $\{x^k\}$ generated by Algorithm 1 is a solution of NCP(F).*

Proof. Since the sequence $\{\theta(x^k)\}$ is nonnegative and decreases monotonically, it converges to some $\theta^* \geq 0$. We assume $\theta^* > 0$. Let x^* be an accumulation point of $\{x^k\}$ and $\{x^k\}_{k\in K}$ be a subsequence converging to x^*. Taking a further subsequence if necessary, we can assume without loss of generality that $d^k \to d^*$, because $\tilde{\Delta}$ is uniformly compact near x^* by Lemma 3.4. Furthermore by the closedness of $\tilde{\Delta}$, we have

$$d^* \in \tilde{\Delta}(x^*). \tag{21}$$

Since

$$\theta(x^k + t_k d^k) - \theta(x^k) \leq t_k \rho \nabla\theta(x^k)^T d^k,$$

$\{t_k \nabla\theta(x^k)^T d^k\}$ converges to 0. In order to prove $\nabla\theta(x^k)^T d^k \to 0$, we show that $\{t_k\}$ is bounded away from 0. Now suppose that there exists a subsequence such that $t_k \to 0$. By the line search rule, we have

$$\frac{\theta(x^k + t'_k d^k) - \theta(x^k)}{t'_k} > \rho \nabla\theta(x^k)^T d^k, \tag{22}$$

where $t'_k = t_k/\beta$. Since $t'_k \to 0$, taking the limit of both sides of (22) yields

$$\nabla\theta(x^*)^T d^* > \rho \nabla\theta(x^*)^T d^*. \tag{23}$$

Since $\theta(x^*) = \theta^* > 0$ by assumption, it follows from (21) and Lemma 3.2 that $\nabla\theta(x^*)^T d^* < 0$. Since $\rho < 1$, this contradicts (23). This implies that $\{t_k\}$ is bounded away from 0, and hence, $\{\nabla\theta(x^k)^T d^k\}$ converges to 0. That is,

$$\lim_{\substack{k\to\infty \\ k\in K}} \nabla\theta(x^k)^T d^k = \nabla\theta(x^*)^T d^* = 0.$$

It then follows from (21) and Lemma 3.2 that $\theta(x^*) = 0$. This is contradictory to $\theta^* > 0$. Therefore, we must have $\theta(x^k) \to 0$, which implies that any accumulation point x^* of $\{x^k\}$ satisfies $\theta(x^*) = 0$ and hence is a solution of NCP(F). The proof is complete. □

4. Securing superlinear convergence

In the previous section, we established global convergence of Algorithm 1, which includes a generalized Newton method as a special case in the sense that it determines search direction based on $\tilde{\mathcal{H}}$ which contains the generalized Jacobian $\partial_B\Phi$ (cf. Proposition 2.2). To secure rapid convergence theoretically, however, we need to restrict the choice of search directions. In this section, we propose a restricted version of Algorithm 1, which determines search direction based on the mapping $\hat{\mathcal{H}}$ defined by (8), and establish its superlinear and quadratic convergence under suitable conditions.

First, recall that the point-to-set mapping $\hat{\mathcal{H}}$ and Λ_i are defined by (8) and (13). Again we have a trouble that a matrix $\hat{H} \in \hat{\mathcal{H}}(x)$ is not necessarily nonsingular. So we shall modify $\hat{H}$ by using a technique similar to that used for $\tilde{H}$. Let a matrix $\hat{G}$ be given by

$$\hat{G} = D_{\hat{p}} + D_{\hat{q}} F'(x), \tag{24}$$

where $\hat{p}$ and $\hat{q}$ are vectors such that

$$(\hat{p}_i, \hat{q}_i) = (\hat{a}_i + \bar{a}_i, \hat{b}_i + \bar{b}_i), \quad i = 1, 2, \ldots, n, \tag{25}$$

with $(\hat{a}, \hat{b}) \in \hat{\Omega}(x)$ and $(\bar{a}_i, \bar{b}_i) \in \Lambda_i(x, \hat{a}_i, \hat{b}_i), i = 1, 2, \ldots, n$. If $\theta(x) > 0$, (13) and (25) imply that both $D_{\hat{p}}$ and $D_{\hat{q}}$ are negative definite diagonal matrices. We define a point-to-set mapping $\hat{\mathcal{G}} : R^n \to \mathcal{P}(R^{n\times n})$ by

$$\begin{aligned}\hat{\mathcal{G}}(x) = \{\hat{G} \in R^{n\times n} \mid \hat{G} = D_{\hat{p}} + D_{\hat{q}} F'(x), (\hat{p}, \hat{q}) \text{ is given by (25) with } (\hat{a}, \hat{b}) \in \hat{\Omega}(x) \\ \text{and } (\bar{a}_i, \bar{b}_i) \in \Lambda_i(x, \hat{a}_i, \hat{b}_i) \text{ for } i = 1, 2, \ldots n\}.\end{aligned} \tag{26}$$

Namely, (24) means that $\hat{G}$ is chosen from the set $\hat{\mathcal{G}}(x)$ arbitrarily. As in the previous section, if $\hat{G} = D_{\hat{p}} + D_{\hat{q}} F'(x) \in \hat{\mathcal{G}}(x)$ and $\hat{H} = D_{\hat{a}} + D_{\hat{b}} F'(x) \in \hat{\mathcal{H}}(x)$, with $(\hat{a}, \hat{b})$ being related to $(\hat{p}, \hat{q})$ by (13) and (25), then we say that $\hat{G}$ corresponds to $\hat{H}$. Since $\hat{\mathcal{H}}(x)$ is a subset of $\tilde{\mathcal{H}}(x)$ for any x by Proposition 2.2 (a), $\hat{\mathcal{G}}(x)$ is a subset of $\tilde{\mathcal{G}}(x)$ for any x. So it follows from Proposition 3.1 that every $\hat{G} \in \hat{\mathcal{G}}(x)$ is nonsingular if $\theta(x) > 0$.

Now, we define a point-to-set mapping $\hat{\Delta} : R^n \to \mathcal{P}(R^n)$ by

$$\hat{\Delta}(x) = \{d \in R^n \mid \hat{G}d = -\Phi(x), \ \ \hat{G} \in \hat{\mathcal{G}}(x)\}. \tag{27}$$

Since $\hat{\mathcal{G}}(x)$ is a subset of $\tilde{\mathcal{G}}(x)$ for any x, $\hat{\Delta}(x)$ is a subset of $\tilde{\Delta}(x)$ for any x. Moreover, $\hat{\Delta}(x)$ is nonempty for x such that $\theta(x) > 0$.

The following algorithm determines search directions from $\hat{\Delta}(x)$ and, in fact, differs from Algorithm 1 only in the direction finding step.

Algorithm 2.

Step 1: Select an initial point $x^0 \in R^n$ and parameters $\rho \in (0, \frac{1}{2})$ and $\beta \in (0, 1)$. Set $k := 0$.

Step 2: If $\theta(x^k) = 0$, then stop. Otherwise, go to Step 3.

Step 3: Find a vector $d^k \in \hat{\Delta}(x^k)$.

Step 4: Let m be the smallest nonnegative integer such that

$$\theta(x^k + \beta^m d^k) - \theta(x^k) \leq \beta^m \rho \nabla\theta(x^k)^T d^k$$

and set $x^{k+1} := x^k + t_k d^k$, where $t_k := \beta^m$. Return to Step 2 with k replaced by $k + 1$.

Note that, since $\hat{\Delta}(x^k)$ is a subset of $\tilde{\Delta}(x^k)$ at each x^k as mentioned above, global convergence of Algorithm 2 is readily deduced from that of Algorithm 1. We state the following theorem without proof.

Theorem 4.1 *Any accumulation point of the sequence $\{x^k\}$ generated by Algorithm 2 is a solution of NCP(F).* □

In what follows, we focus our attention to the convergence rate of Algorithm 2. To begin with, we give a condition under which the sequence $\{x^k\}$ generated by Algorithm 2 has a unique limit point.

Lemma 4.1 *If an accumulation point x^* of $\{x^k\}$ is BD-regular with respect to Φ, then $x^k \to x^*$.*

Proof. It can be proven in a way similar to the proof of [3, Theorem 3.1 b, c]. □

In the remainder of this section, we assume that an accumulation point x^* of $\{x^k\}$ is BD-regular with respect to Φ. Lemma 4.1 then ensures that the whole sequence $\{x^k\}$ converges to x^*.

The following property of $\hat{G} \in \hat{\mathcal{G}}(x)$ is particularly important in the subsequent analysis.

Lemma 4.2 *There exists a constant $c > 0$ such that*

$$\|\hat{G}_k^{-1}\| \leq c$$

for all k sufficiently large, where $\hat{G}_k \in \hat{\mathcal{G}}(x^k)$.

Proof. Consider a matrix $\hat{H}_k \in \hat{\mathcal{H}}(x^k)$ corresponding to $\hat{G}_k$ for each k. Since x^* is BD-regular with respect to Φ and since $\hat{\mathcal{H}}(x^k) \subseteq \partial_B \Phi(x^k)$ for each k by Proposition 2.2, $\|\hat{H}_k^{-1}\|$ is bounded for all k sufficiently large by [20, Lemma 2.6]. Since $\sigma(\theta(x^k)) \to 0$, it follows from (9),(13) and (24) that $\|\hat{G}_k - \hat{H}_k\|$ are sufficiently small for all k large enough. Then, the desired result is easily obtained from Perturbation Lemma [17, Theorem 2.3.2]. □

Lemma 4.3 *We have*

$$\|x^k + d^k - x^*\| = o\left(\|x^k - x^*\|\right).$$

Proof. For each k, we have

$$\begin{aligned}
&\|x^k + d^k - x^*\| \\
&= \|x^k - \hat{G}_k^{-1}\Phi(x^k) - x^*\| \\
&= \|\hat{G}_k^{-1}\left(\Phi(x^*) - \Phi(x^k) + \hat{H}_k(x^k - x^*) + (\hat{G}_k - \hat{H}_k)(x^k - x^*)\right)\| \\
&\leq \|\hat{G}_k^{-1}\|\left(\|\Phi(x^*) - \Phi(x^k) + \hat{H}_k(x^k - x^*)\| + \|\hat{G}_k - \hat{H}_k\|\|(x^k - x^*)\|\right), \qquad (28)
\end{aligned}$$

where $\hat{H}_k \in \hat{\mathcal{H}}(x^k)$ is the matrix corresponding to $\hat{G}_k$. Since Φ is semismooth and $\hat{\mathcal{H}}(x^k) \subseteq \partial\Phi_B(x^k)$ for each k by Proposition 2.2, we have

$$\|\Phi(x^*) - \Phi(x^k) + \hat{H}_k(x^k - x^*)\| = o(\|x^k - x^*\|) \qquad (29)$$

(see [20, Theorem 3.1]). Moreover, by (13) and (25), we have

$$\|\hat{G}_k - \hat{H}_k\| = O\left(\sigma(\theta(x^k))\right). \qquad (30)$$

Consequently, it follows from (28)-(30) that

$$\|x^k + d^k - x^*\| \leq \|\hat{G}_k^{-1}\|\left(o(\|x^k - x^*\|) + O\left(\sigma(\theta(x^k))\right)\|x^k - x^*\|\right).$$

Since $\sigma(\theta(x^k)) \to 0$ and $\{\|\hat{G}_k^{-1}\|\}$ is bounded by Lemma 4.2, we have the desired result. □

Lemma 4.4 *Let $\delta > 0$ be arbitrary but sufficiently small. Then we have*

$$\theta(x^k + d^k) \leq \left(\frac{\delta C}{1 - \delta}\right)^2 \theta(x^k)$$

for all k large enough, where $C > 0$ is a constant independent of δ.

Proof. Since $d^k = -\hat{G}_k^{-1}\Phi(x^k)$, it follows from Lemma 4.2 that there exists a constant $c > 0$ such that

$$\begin{aligned}
\|x^k - x^*\| &\leq \|x^k + d^k - x^*\| + \|d^k\| \\
&\leq \|x^k + d^k - x^*\| + \|\hat{G}_k^{-1}\|\|\Phi(x^k)\| \\
&\leq \|x^k + d^k - x^*\| + c\|\Phi(x^k)\|
\end{aligned}$$

for all k large enough. By Lemma 4.3, for all k large enough, we have $\|x^k + d^k - x^*\| \leq \delta \|x^k - x^*\|$ and hence

$$\|x^k + d^k - x^*\| \leq \frac{\delta c}{1-\delta} \|\Phi(x^k)\|. \tag{31}$$

Since F is continuously differentiable, F is Lipschitz continuous near x^*, which in turn implies, by Proposition 2.3 (a), that Φ is also Lipschitz continuous near x^*. Let $L > 0$ be a Lipschitz constant of Φ near x^*. Then, by (31), we have

$$\|\Phi(x^k + d^k)\| = \|\Phi(x^k + d^k) - \Phi(x^*)\| \leq L\|x^k + d^k - x^*\| \leq \frac{\delta c L}{1-\delta} \|\Phi(x^k)\|$$

for all k sufficiently large. Letting $C = cL$ and squaring these inequalities, we obtain the desired inequality. □

Lemma 4.5 *For any $\eta > 0$, we have*

$$-\nabla\theta(x^k)^T d^k \leq 2(1+\eta)\theta(x^k)$$

for all k large enough.

Proof. By Lemma 4.2, there exists a positive constant c such that $\|\hat{G}_k^{-1}\| \leq c$ for all k sufficiently large. Let $\hat{G}_k \in \hat{\mathcal{G}}(x^k)$ correspond to $\hat{H}_k \in \hat{\mathcal{H}}(x^k)$. First note that Proposition 2.5 (a) implies $\nabla\theta(x^k) = \hat{H}_k^T \Phi(x^k)$, since $\hat{\mathcal{H}}(x^k) \subseteq \tilde{\mathcal{H}}(x^k)$ by Proposition 2.2. Then, since $d^k = -\hat{G}_k^{-1}\Phi(x^k)$, we have

$$\begin{aligned}
-\nabla\theta(x^k)^T d^k &= \Phi(x^k)^T \hat{H}_k \hat{G}_k^{-1} \Phi(x^k) \\
&= \Phi(x^k)^T \hat{G}_k \hat{G}_k^{-1} \Phi(x^k) + \Phi(x^k)^T (\hat{H}_k - \hat{G}_k) \hat{G}_k^{-1} \Phi(x^k) \\
&\leq 2\theta(x^k) + 2\|\hat{G}_k^{-1}\| \|\hat{H}_k - \hat{G}_k\| \theta(x^k) \\
&\leq 2\left(1 + c\|\hat{H}_k - \hat{G}_k\|\right) \theta(x^k)
\end{aligned}$$

for all k large enough. By (13), (25) and $\sigma(\theta(x^k)) \to 0$, it follows that $\|\hat{H}_k - \hat{G}_k\| \to 0$. Hence, for all k large enough, we have $c\|\hat{H}_k - \hat{G}_k\| \leq \eta$ and hence the desired inequality. □

By using these lemmas, we show that the algorithm eventually accepts the unit stepsize.

Lemma 4.6 *For all k sufficiently large, the stepsize $t_k = 1$ is chosen in Step 3 of Algorithm 2.*

Proof. Since $0 < \rho < \frac{1}{2}$, we can take $\eta > 0$ sufficiently small to satisfy

$$2(1+\eta)\rho < 1.$$

Moreover we can choose a sufficiently small $\delta > 0$ satisfying

$$1 - \left(\frac{\delta C}{1-\delta}\right)^2 > 2(1+\eta)\rho,$$

where C is a constant indicated in Lemma 4.4. It then follows from Lemmas 4.4 and 4.5 that

$$\begin{aligned}\theta(x^k+d^k)-\theta(x^k) &\leq -\left(1-\left(\frac{\delta C}{1-\delta}\right)^2\right)\theta(x^k)\\ &\leq -2(1+\eta)\rho\theta(x^k) \leq \rho\nabla\theta(x^k)^T d^k\end{aligned}$$

for all k sufficiently large. This completes the proof. □

Now we state a rate of convergence result for Algorithm 2.

Theorem 4.2 *Algorithm 2 has a superlinear rate of convergence.*

Proof. By Lemma 4.6, we have

$$x^{k+1} = x^k + d^k$$

for all k sufficiently large. It then follows from Lemma 4.3 that

$$\lim_{k\to\infty} \frac{\|x^{k+1}-x^*\|}{\|x^k-x^*\|} = 0.$$

The proof is complete. □

Next we shall give conditions under which Algorithm 2 is quadratically convergent.

Theorem 4.3 *Suppose that F is twice differentiable with Lipschitz continuous Hessian at x^*, and the function σ which constitutes the mappings Λ_i defined by (13) satisfies $\sigma(s) = O(\sqrt{s})$. Then Algorithm 2 has a quadratic rate of convergence.*

Proof. Under the given assumptions on F, Φ is strongly semismooth at x^* by Proposition 2.3 (b). Hence, taking account of Proposition 2.2 (a), we can prove

$$\|\Phi(x^*)-\Phi(x^k)+\hat{H}_k(x^k-x^*)\| = O(\|x^k-x^*\|^2) \tag{32}$$

in a way similar to the proof of [20, Theorem 3.1]. Furthermore, by the given assumption on σ,

$$\begin{aligned}\sigma(\theta(x^k)) &= O\left(\sqrt{\theta(x^k)}\right)\\ &= O\left(\|\Phi(x^k)\|\right)\\ &= O\left(\|x^k-x^*\|\right)\end{aligned} \tag{33}$$

for all k sufficiently large, where the last equality follows from Proposition 2.4. Since $t_k = 1$ for all k large enough by Lemma 4.6, it follows from Lemma 4.2, (28), (32) and (33) that for all k sufficiently large,

$$\begin{aligned}\|x^{k+1}-x^*\| &= \|x^k+d^k-x^*\|\\ &\leq \|\hat{G}_k^{-1}\|\left(\|\Phi(x^*)-\Phi(x^k)+\hat{H}_k(x^k-x^*)\|+O\left(\sigma(\theta(x^k))\right)\|x^k-x^*\|\right)\\ &= c\left(O(\|x^k-x^*\|^2)+O(\|x^k-x^*\|)\|x^k-x^*\|\right),\end{aligned}$$

where $c>0$ is a constant given in Lemma 4.2. This completes the proof. □

5. Numerical results

In this section, we report some numerical results with Algorithm 2 described in the previous section.

Throughout the computational experiments, the parameters β and ρ used in Algorithm 2 were set as $\beta = 0.5$ and $\rho = 10^{-4}$, respectively, and the parameter ε and the function σ which constitute the mappings Λ_i defined by (13) were set as $\varepsilon = 0.05$ and

$$\sigma(t) = 0.1 \min\{1, t\},$$

respectively. In a manner similar to [3], we chose a vector z to identify an element of $\partial_B \Phi(x^k)$ as

$$z_i = \begin{cases} 1 & \text{if } x_i^k = F_i(x^k) = 0, \\ 0 & \text{otherwise,} \end{cases}$$

for $i = 1, 2, \ldots, n$ (cf. (6)). The stopping criterion for the algorithm was $\theta(x^k) \leq 10^{-8}$. All computer programs were coded in FORTRAN 77 and the run was made in double precision on SUN SPARC 10.

We report the results for the following four test problems.

Example 1: F is the affine mapping

$$F(x) = Mx + q,$$

where

$$M = \begin{pmatrix} 4 & 2 & 2 & 1 \\ 2 & 4 & 0 & 1 \\ 2 & 0 & 2 & 2 \\ -1 & -1 & -2 & 0 \end{pmatrix}$$

and $q = (-8, -6, -4, 3)^T$. This problem is derived from the problem 35 in [10] via the Karush-Kurn-Tucker optimality condition. Note that M is positive semi-definite, but not positive definite. This problem has the unique solution $x = (4/3, 7/9, 4/9, 2/9)^T$. The results for this problem are shown in Table 1. From Table 1, we see that the algorithm is convergent rapidly for each starting point.

Example 2: F is the linear mapping

$$F(x) = Mx,$$

where

$$M = \begin{pmatrix} 4 & -1 & 0 & 0 \\ -1 & 4 & -1 & 0 \\ 0 & -1 & 4 & -1 \\ 0 & 0 & -1 & 4 \end{pmatrix}.$$

Since M is positive definite, this problem has the unique solution $x = 0$. Furthermore since $F(0) = 0$ at the solution, the solution is degenerate. The results are shown in Table 2, which indicates that the algorithm is superlinearly convergent, despite the fact that the solution is degenerate.

Example 3: F is the constant mapping

$$F(x) = q,$$

where $q = (1, 0, 1, 0)^T$. Any point $x = (0, x_2, 0, x_4)^T$ such that $x_2 \geq 0$ and $x_4 \geq 0$ is a solution of this problem, but it is not BD-regular. Moreover, every element of $\tilde{\mathcal{H}}(x)$ is singular whenever $x_2 > 0$ and $x_4 > 0$. The results for this problem are shown in Table 3. From Table 3, we see that the algorithm is convergent superlinearly, despite the fact that the solution is not BD-regular.

Example 4: F is the strictly monotone mapping

$$F(x) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} x_1^3 \\ x_2^3 \\ 2x_3^3 \\ 2x_4^3 \end{pmatrix} + \begin{pmatrix} -8 \\ 3 \\ -3 \\ 0 \end{pmatrix}.$$

This problem has the unique but degenerate solution $x = (2, 0, 1, 0)^T$. The results are shown in Table 4. From the table, we see that the algorithm requires many iterations to converge to a solution for some starting points. Nevertheless, it can be seen that, at the final stage of the iterations, the algorithm is convergent superlinearly for each starting point.

6. Concluding remark

In this paper, we have proposed Newton-type algorithms for solving the monotone NCP(F). The modified Newton equations used in these algorithms can always be solved uniquely. We have established global convergence and superlinear convergence theorems under appropriate assumptions.

The proposed algorithms may behave in a similar manner to the algorithm proposed by De Luca et al. [3], particularly when applied to a problem in which the solution is BD-regular and $\partial_B \Phi(x)$ contains only nonsingular matrices for all x. However, if the solution is not BD-regular or elements of $\partial_B \Phi(x)$ fail to be nonsingular in a wide region, the behavior of the proposed algorithms may be quite different from that of [3]. For example, in Example 3 given in the previous section, for all x such that $x_2 > 0$ and $x_4 > 0$, we have $\partial_B \Phi(x) = \{D_a\}$, where a is given by

$$a = \begin{pmatrix} \frac{x_1}{\sqrt{x_1^2+1}} - 1 \\ 0 \\ \frac{x_3}{\sqrt{x_3^2+1}} - 1 \\ 0 \end{pmatrix}.$$

Hence, the element of $\partial_B \Phi(x)$ is singular, so that the algorithm of [3] will reduce to the steepest descent method, in which the search direction d is given by

$$d = - \begin{pmatrix} \left(\frac{x_1}{\sqrt{x_1^2+1}} - 1 \right) (\sqrt{x_1^2+1} - x_1 - 1) \\ 0 \\ \left(\frac{x_3}{\sqrt{x_3^2+1}} - 1 \right) (\sqrt{x_3^2+1} - x_3 - 1) \\ 0 \end{pmatrix}.$$

As shown in Table 3, the proposed algorithm converged very rapidly even for this algorithm. This is because the proposed algorithm has a similar feature to the algorithm proposed by Fisher [6], which has a superlinear rate of convergence in the case where F is affine and each solution is nondegenerate.

References

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

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

[3] T. De Luca, F. Facchinei and C. Kanzow, "A semismooth equation approach to the solution of nonlinear complementarity problems," Technical Report, Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza," Roma, Italy, 1995.

[4] F. Facchinei and J. Soares, "A new merit function for nonlinear complementarity problems and a related algorithm," Technical Report, Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza," Roma, Italy, 1994.

[5] A. Fischer, "A special Newton-type optimization method," *Optimization* 24 (1992) 269–284.

[6] A. Fischer, " On the local superlinear convergence of a Newton-type method for LCP under weak conditions," Technical Report, Institute for Numerical Mathematics, Technical University of Dresden, Dresden, Germany, 1994.

[7] A. Fischer, "An NCP-function and its use for the solution of complementarity problems," to appear in: D.-Z. Du, L. Qi and R.S. Wormersley, eds., *Recent Advances in Nonsmooth Optimization* (World Scientific Publishers, Singapore, 1995).

[8] C. Geiger and C. Kanzow, "On the resolution of monotone complementarity problems," Preprint 82, Institut für Angewandte Mathematik, Universität Hamburg, Hamburg, Germany, 1994.

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

[10] W. Hock and K. Schittkowski, *Test Examples for Nonlinear Programming Codes* (Springer-Verlag, Berlin, 1981).

[11] W.W. Hogan, "Point-to-set maps in mathematical programming," *SIAM Review* 15 (1973) 591–603.

[12] G. Isac, *Complementarity Problems* (Springer-Verlag, Berlin, 1992).

[13] C. Kanzow, "Nonlinear complementarity as unconstrained optimization," to appear in *Journal of Optimization Theory and Applications*.

[14] C. Kanzow and H. Kleinmichel, "A class of Newton-type methods for equality and inequality constrained optimization," Preprint 61, Institut für Angewandte Mathematik, Universität Hamburg, Hamburg, Germany, 1992.

[15] O.L. Mangasarian, "Equivalence of the complementarity problem to a system of nonlinear equations," *SIAM Journal on Applied Mathematics* 31 (1976) 89–92.

[16] J.J. Moré and W.C. Rheinboldt, "On P- and S-functions and related class of n-dimensional nonlinear mappings," *Linear Algebra and its Applications* 6 (1973) 45–68.

[17] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables* (Academic Press, New York, 1970).

[18] J.-S. Pang, "Complementarity problems," in: R. Horst and P. Pardalos, eds., *Handbook of Global Optimization* (Kluwer Academic Publishers, Boston, 1994).

[19] J.-S. Pang and S.A. Gabriel, "NE/SQP: A robust algorithm for the nonlinear complementarity problem," *Mathematical Programming* 60 (1993) 295–337.

[20] L. Qi, "Convergence analysis of some algorithms for solving nonsmooth equations," *Mathematics of Operations Research* 18 (1993) 227–244.

[21] L. Qi and H. Jiang, "Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations," Applied Mathematics Technical Report AMR 94/5, School of Mathematics, University of New South Wales, Sydney, Australia, 1994.

[22] L. Qi and J. Sun, "A nonsmooth version of Newton's method," *Mathematical Programming* 58 (1993) 353–367.

[23] S.M. Robinson, "Strongly regular generalized equations," *Mathematics of Operations Research* 5 (1980) 43–62.

[24] P.K. Subramanian, "Gauss-Newton methods for the complementarity problems," *Journal of Optimization Theory and Applications* 77 (1993) 467–482.

Table 1: Results for Example 1

Starting point	#iterations	θ-value (last 3 iterations)
(−1,−1,−1,−1)	5	(2.35D−03, 1.43D−05, 1.87D−09)
(0, 0, 0, 0)	5	(2.85D−04, 6.66D−08, 3.52D−14)
(1, 1, 1, 1)	4	(1.30D−02, 2.96D−05, 2.13D−10)
(10, 10, 10, 10)	12	(1.16D−03, 7.35D−06, 5.19D−10)

Table 2: Results for Example 2.

Starting point	#iterations	θ-value (last 3 iterations)
(−1,−1,−1,−1)	2	(- - - - - - -, 2.20D−02, 3.72D−09)
(1, 1, 1, 1)	4	(6.58D−03, 3.06D−07, 1.37D−20)
(10, 10, 10, 10)	6	(9.35D−02, 3.74D−04, 3.59D−11)

Table 3: Results for Example 3.

Starting point	#iterations	θ-value (last 3 iterations)
(−1,−1,−1,−1)	4	(5.66D−04, 7.29D−08, 1.33D−15)
(1, 1, 1, 1)	4	(2.69D−04, 1.70D−08, 7.23D−17)
(10, 10, 10, 10)	5	(3.33D−04, 2.57D−08, 1.66D−16)

Table 4: Results for Example 4.

Starting point	#iterations	θ-value (last 3 iterations)
(−1,−1,−1,−1)	7	(1.05D−02, 1.87D−05, 9.72D−11)
(0, 0, 0, 0)	44	(2.63D−03, 5.26D−07, 2.40D−14)
(1, 1, 1, 1)	5	(2.34D−01, 4.70D−04, 6.59D−09)
(10, 10, 10, 10)	38	(3.16D−03, 2.85D−06, 2.49D−15)