Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational and Applied Mathematics 212 (2008) 75–85
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

A smoothing Newton-type method for generalized nonlinear complementarity problem☆

Xinzhen Zhang[a,*], Hefeng Jiang[b], Yiju Wang[c,d]

[a]*Department of Mathematics, Qufu Normal University, Qufu, Shandong 273165, China*
[b]*Department of Mathematics, Taiyuan Teacher's College, Taiyuan, Shanxi 030012, China*
[c]*School of Operations Research and Management Sciences, Qufu Normal University, Rizhao Shandong 276800, China*
[d]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong*

Received 19 April 2005

Abstract

By using a new type of smoothing function, we first reformulate the generalized nonlinear complementarity problem over a polyhedral cone as a smoothing system of equations, and then develop a smoothing Newton-type method for solving it. For the proposed method, we obtain its global convergence under milder conditions, and we further establish its local superlinear (quadratic) convergence rate under the BD-regular assumption. Preliminary numerical experiments are also reported in this paper.

MSC: 90C33

Keywords: GNCP; Smoothing method; Superlinear convergence

1. Introduction

Let F, G be continuously differentiable mappings from R^n to R^m, $\mathcal{K}$ be a nonempty closed convex cone in R^m and $\mathcal{K}^\circ$ denote its polar cone. The generalized nonlinear complementarity problem, denoted by GNCP($F, G, \mathcal{K}$), is to find a vector $x^* \in R^n$ such that

$$F(x^*) \in \mathcal{K}, \quad G(x^*) \in \mathcal{K}^\circ, \quad F(x^*)^\top G(x^*) = 0.$$

This problem has many interesting applications in such as engineering and economics, and is a wide class of problems that contains the classical nonlinear complementarity problem, abbreviated as NCP, as a special case, see, e.g. [1,6,10] and references therein. To solve it, one usually reformulates it as a minimization problem over a simple set or an unconstrained optimization problem, see [17] for the case that $\mathcal{K}$ is a general cone, and see [9,10] for the case that $\mathcal{K} = R^n_+$. The conditions under which a stationary point of the reformulated optimization is a solution of the GNCP($F, G, \mathcal{K}$) were also provided in the literature.

☆ This project was supported by Shandong Provincial NSF (Grant No. Y2003A02) and NSF of Qufu Normal University, and the Postdoctoral Fellowship of The Hong Kong Polytechnic University.
* Corresponding author.
E-mail addresses: zxz7709@yahoo.com.cn (X. Zhang), yiju66@hotmail.com (Y. Wang).

0377-0427/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2006.03.042

In this paper, we consider the GNCP(F, G, $\mathscr{K}$) for the case that $m = n$, and $\mathscr{K}$ is a polyhedral cone in R^n, i.e., there exist $A \in R^{s\times n}$, $B \in R^{t\times n}$ such that

$$\mathscr{K} = \{v \in R^n | Av \geqslant 0,\ \ Bv = 0\}.$$

It is easy to verify that its polar cone $\mathscr{K}^\circ$ assumes the following representation:

$$\mathscr{K}^\circ = \{u \in R^n | u = A^\top \lambda_1 + B^\top \lambda_2,\ \lambda_1 \geqslant 0,\ \lambda_1 \in R^s,\ \lambda_2 \in R^t\}.$$

From now on, the GNCP(F, G, $\mathscr{K}$) is specialized over a polyhedral cone, and in the subsequent analysis we abbreviate it as GNCP for simplicity. In [1], Andreani et al. reformulated the problem as a smooth optimization problem with simple constraints and presented the sufficient conditions under which a stationary point of the optimization problem is a solution of the concerned problem. Later, Wang et al. [18] reformulated the problem as a system of nonlinear and nonsmooth equations, and proposed a nonsmooth Levenberg–Marquardt method for solving it.

It is well known that the smoothing Newton-type method received much attention in solving such as NCP and minimization problem due to its high efficiency [2,3,5,14,16]. It seems reasonable to ask if this kind of method can be applied to the GNCP, and this actually constitutes the main motivation of this paper. In the rest of this paper, we will first present a new reformulation of the GNCP by using a new type of smoothing function, and then develop a smoothing Newton-type method for solving it which guarantees the monotonicity of the generated sequence of the objective function. Under milder conditions, we show that any accumulation point of the generated sequence is a solution of the GNCP, and we also establish the local superlinear (quadratic) convergence rate of the proposed algorithm under the BD-regular assumption. Preliminary numerical experiments show the efficiency of the proposed algorithm.

To end this section, we give some standard notions used in this paper: for a continuously differentiable function $\Gamma : R^n \to R^m$, we denote the Jacobian of Γ at $x \in R^n$ by $\Gamma'(x) \in R^{m\times n}$, whereas the transposed Jacobian is denoted as $\nabla\Gamma(x)$. In particular, if $m = 1$, $\nabla\Gamma(x)$ is a column vector. We use $x^\top y$ to denote the inner product of vectors x, $y \in R^n$, and use $[a]_i$ or a_i to denote the ith component of the vector $a \in R^n$. The null space of a matrix B is denoted by $\mathscr{N}(B)$.

2. Preliminaries

In [18], the authors reformulated the GNCP as a system of nonlinear equations based on the following Fischer function [8]:

$$\phi_{\mathrm{F}}(a, b) = \sqrt{a^2 + b^2} - a - b \quad \text{for } a, b \in R$$

as is seen from the following conclusion.

Lemma 2.1. *$x^* \in R^n$ is a solution of the GNCP if and only if there exist $\lambda_1^* \in R^s$ and $\lambda_2^* \in R^t$, such that*

$$\begin{cases} \Phi_{\mathrm{F}}(AF(x^*), \lambda_1^*) = 0, \\ BF(x^*) = 0, \\ G(x^*) - A^\top \lambda_1^* - B^\top \lambda_2^* = 0, \end{cases}$$

where $\Phi_{\mathrm{F}}(a, b) = (\phi_{\mathrm{F}}(a_1, b_1), \phi_{\mathrm{F}}(a_2, b_2), \ldots, \phi_{\mathrm{F}}(a_s, b_s))^\top$ for $a, b \in R^s$.

Now, we will establish a new type of smoothing reformulation of the GNCP based on the following smoothing approximation function to the Fischer function:

$$\phi(\varepsilon, a, b) = \sqrt{a^2 + b^2 + \alpha\varepsilon^2} - a - b, \quad a, b, \varepsilon \in R,$$

where $\alpha > 0$ is a constant.

Based on the relation between $\phi_{\mathrm{F}}(\cdot,\cdot)$ and $\phi(\cdot,\cdot,\cdot)$, we can establish the following smoothing function to the GNCP:

$$H(\varepsilon, y) := \begin{pmatrix} \theta(\varepsilon) \\ \Psi(\varepsilon, x, \lambda_1, \lambda_2) \end{pmatrix},$$

where $\theta(\varepsilon) = \sqrt{\alpha s}((\varepsilon+1)^2 - 1)$ and

$$\Psi(\varepsilon, x, \lambda_1, \lambda_2) = \begin{pmatrix} \Phi(\varepsilon, AF(x), \lambda_1) \\ BF(x) \\ G(x) - A^\top\lambda_1 - B^\top\lambda_2 \end{pmatrix},$$

$$\Phi(\varepsilon, AF(x), \lambda_1) = \begin{pmatrix} \phi(\varepsilon, [AF(x)]_1, [\lambda_1]_1) \\ \vdots \\ \phi(\varepsilon, [AF(x)]_s, [\lambda_1]_s) \end{pmatrix}.$$

In subsequent development of the paper, we denote $y = (x, \lambda_1, \lambda_2)$, $z = (\varepsilon, y)$ and define

$$f(\varepsilon, y) = \tfrac{1}{2}\|\Psi(z)\|^2, \quad T(z) = \tfrac{1}{2}\|H(z)\|^2 = \tfrac{1}{2}\theta^2(\varepsilon) + f(z).$$

From Lemma 2.1, we can see that x^* is a solution of the GNCP if and only if there exist $\lambda_1^* \in R^s$, $\lambda_2^* \in R^t$ such that $z^* = (0, x^*, \lambda_1^*, \lambda_2^*)$ is a global minimizer with zero objective function value of the unconstrained optimization problem

$$\min_{z \in R^{1+n+s+t}} T(z). \tag{2.1}$$

Obviously, the smoothing function $\phi(\varepsilon, a, b)$ is not smooth everywhere. However, it is differentiable almost everywhere, and therefore it has a nonempty generalized Jacobian in the sense of Clarke [4].

To proceed our analysis, we need to review some concepts related to nonsmooth analysis. For a locally Lipschitzian mapping $\Theta : R^n \to R^m$, we denote by $\partial\Theta(x)$ the Clarke's generalized Jacobian of $\Theta(x)$ at $x \in R^n$ which can be expressed as the convex hull of the set $\partial_B\Theta(x)$ [13], where

$$\partial_B\Theta(x) = \left\{ V \in R^{n\times n} | V = \lim_{x_k \to x} \Theta'(x_k), \Theta(x) \text{ is differentiable at } x_k \text{ for all } k \right\}.$$

The following definitions are due to Qi and Sun [15].

Definition 2.1. A locally Lipschitz continuous vector-valued function $\Theta(x) : R^n \to R^m$ is said to be semismooth at $x \in R^n$, if the limit

$$\lim_{\substack{V \in \partial\Theta(x+th') \\ h' \to h, t \downarrow 0}} \{Vh'\}$$

exists for any $h \in R^n$.

Definition 2.2. A function $\Theta(x) : R^n \to R^m$ is said to be strongly semismooth at $x \in R^n$ if $\Theta(x)$ is semismooth at x and for any $V \in \partial\Theta(x+h)$ and $h \to 0$, it holds that

$$\Theta(x+h) - \Theta(x) - Vh = \mathrm{O}(||h||^2).$$

For simplicity, we denote the Clarke's generalized Jacobian of $\Phi(\varepsilon, AF(x), \lambda_1)$ with respect to $(\varepsilon, x, \lambda_1) \in R^{1+n+s}$ by $\partial\Phi(\varepsilon, AF(x), \lambda_1)$. Similar discussion to [7, Proposition 3.1] gives the following result (also see [18, Lemmas 2.3, 2.4]).

Proposition 2.1. (1) *The function $\Phi(\varepsilon, AF(x), \lambda_1)$ is continuously differentiable for $\varepsilon \neq 0$, and for this function, it holds that*

$$\partial\Phi(\varepsilon, AF(x), \lambda_1) \subseteq (c \;\; D_a AF'(x) \;\; D_b),$$

where $c = (c_1, \ldots, c_s)^\top$, $D_a = \mathrm{diag}(a_1, \ldots, a_s)$ *and* $D_b = \mathrm{diag}(b_1, \ldots, b_s)$ *such that*

$$c_i = \frac{\alpha\varepsilon}{\sqrt{[AF(x)]_i^2 + [\lambda_1]_i^2 + \alpha\varepsilon^2}},$$

$$a_i = \frac{[AF(x)]_i}{\sqrt{[AF(x)]_i^2 + [\lambda_1]_i^2 + \alpha\varepsilon^2}} - 1,$$

$$b_i = \frac{[\lambda_1]_i}{\sqrt{[AF(x)]_i^2 + [\lambda_1]_i^2 + \alpha\varepsilon^2}} - 1$$

if $\varepsilon^2 + [AF(x)]_i^2 + [\lambda_1]_i^2 \neq 0$; *and* $a_i = \zeta_i - 1$, $b_i = \eta_i - 1$ *for any* $(c_i, \zeta_i, \eta_i) \in R^3$ *satisfying* $c_i^2/\alpha + \zeta_i^2 + \eta_i^2 \leqslant 1$ *for the case that* $\varepsilon^2 + [AF(x)]_i^2 + [\lambda_1]_i^2 = 0$.

(2) $H(z)$ *is semismooth on* $R^{1+n+s+t}$, *and is strongly semismooth on* $R^{1+n+s+t}$ *if* $F'(x)$ *and* $G'(x)$ *are both Lipschitz continuous on* R^n.

(3) $T(z)$ *is continuously differentiable on* $R^{1+n+s+t}$ *with* $\nabla T(z) = V^\top H(z)$ *for any* $V \in \partial H(z)$, *and* $f(\varepsilon, y)$ *is continuously differentiable with* $\nabla f(0, y) = V^\top \Psi(0, y)$ *for any* $V \in \partial\Psi(0, y)$.

In the end of this section, we give the definition of BD-regular and a technical lemma which will be used in the convergence analysis of the algorithm proposed in Section 4.

Definition 2.3. A function $\Theta(x) : R^n \to R^n$ is said to be BD-regular at x if any $V \in \partial\Theta(x)$ is nonsingular.

Lemma 2.2. *For any fixed* $y = (x, \lambda_1, \lambda_2) \in R^{n+s+t}$, *the function* $T(z)$ *is monotonically increasing with respect to* $\varepsilon > 0$.

Proof. Since $T(z)$ is differentiable for all $\varepsilon \neq 0$, a direct computation yields

$$\begin{aligned} T_\varepsilon'(z) &= \varepsilon \Bigg[2\alpha s(\varepsilon+1)(\varepsilon+2) \\ &\quad + \sum_{i=1}^{s} \frac{\alpha}{\sqrt{[AF(x)]_i^2 + [\lambda_1]_i^2 + \alpha\varepsilon^2}} \left(\sqrt{[AF(x)]_i^2 + [\lambda_1]_i^2 + \alpha\varepsilon^2} - [AF(x)]_i - [\lambda_1]_i \right) \Bigg] \\ &= \varepsilon \left[2\alpha s(\varepsilon+1)(\varepsilon+2) + \alpha s - \sum_{i=1}^{s} \frac{\alpha[AF(x)]_i + \alpha[\lambda_1]_i}{\sqrt{[AF(x)]_i^2 + [\lambda_1]_i^2 + \alpha\varepsilon^2}} \right] \\ &\geqslant \varepsilon\,[2\alpha s(\varepsilon+1)(\varepsilon+2) - \alpha s] > 0, \end{aligned}$$

which implies that $T_\varepsilon'(z) > 0$ for all $\varepsilon > 0$, the desired result follows. □

3. Stationary point and nonsingularity conditions

Generally, for an optimization problem, one can obtains its a stationary point when he uses the existing optimization methods to solve it. So it is necessary to establish conditions which guarantee that every stationary point of (2.1) is a solution of the GNCP.

Theorem 3.1. *Let* $z^* = (\varepsilon^*, y^*) = (\varepsilon^*, x^*, \lambda_1^*, \lambda_2^*)$ *be a stationary point of* (2.1). *If* $\varepsilon^* \geqslant 0$ *and* $\nabla F(x^*)^{-1}\nabla G(x^*)$ *is positive definite in* $\mathcal{N}(\mathscr{B})$, *then* x^* *is a solution of the GNCP*.

Proof. Define

$$U^* = \Phi(\varepsilon^*, AF(x^*), \lambda_1^*),$$

$$V^* = BF(x^*),$$

$$W^* = G(x^*) - A^\top \lambda_1^* - B^\top \lambda_2^*.$$

Since z^* is a stationary point of (2.1), $\nabla T(z^*) = 0$, i.e.,

$$\begin{cases} 2\alpha s\varepsilon^*(\varepsilon^* + 1)(\varepsilon^* + 2) + c^\top U^* = 0, & (3.1)\\ \nabla F(x^*)A^\top D_a^* U^* + \nabla F(x^*)B^\top V^* + \nabla G(x^*)W^* = 0, & (3.2)\\ D_b^* U^* - AW^* = 0, & (3.3)\\ -BW^* = 0, & (3.4) \end{cases}$$

where $c \in R^s$ is defined in Proposition 2.1.

From (3.4), one has $W^* \in \mathcal{N}(B)$. Pre-multiplying (3.2) by $W^{*\top}\nabla F(x^*)^{-1}$ and using (3.3) and (3.4), we have

$$U^{*\top}(D_b^*)^\top D_a^* U^* + W^{*\top}\nabla F(x^*)^{-1}\nabla G(x^*)W^* = 0. \tag{3.5}$$

Now, we show that $\varepsilon^* = 0$ by reductio ad absurdum. For a contradiction purpose, we assume that $\varepsilon^* > 0$. Then, from the definitions of D_a and D_b, we know $(D_b^*)^\top D_a^*$ is positive definite. Taking into account that $\nabla F(x^*)^{-1}\nabla G(x^*)$ is positive definite in the null space of B, we conclude that $W^* = 0$ and $U^* = 0$. However, this contradicts (3.1). Hence $\varepsilon^* = 0$.

Substituting $\varepsilon^* = 0$ into (3.5) and recalling [10, Lemma 3.1], we have

$$W^* = 0, \quad U^* = \Phi(0, AF(x^*), \lambda_1^*) = 0.$$

Pre-multiplying (3.2) by $F(x^*)\nabla F(x^*)^{-1}$ yields

$$BF(x^*) = 0.$$

The proof is completed. □

To establish the superlinear (quadratic) convergence rate of our algorithm proposed in the next section, we need to explore the conditions under which any element of the generalized Jacobian of the objective function of (2.1) is nonsingular at a stationary point.

Theorem 3.2. *If $z^* = (\varepsilon^*, x^*, \lambda_1^*, \lambda_2^*)$ is a stationary point of* (2.1), *both $F'(x^*)$ and $G'(x^*)$ are nonsingular, and $\varepsilon^* \geqslant 0$, $A\nabla G(x^*)^{-1}\nabla F(x^*)A^\top$ is a P-matrix, then V is nonsingular for any $V \in \partial H(z^*)$.*

Proof. By Proposition 2.1, we know that any element $V \in \partial H(z^*)$ can be written as

$$V^\top = \begin{pmatrix} 2\sqrt{\alpha s}(\varepsilon^* + 1) & c^\top & 0 & 0\\ 0 & \nabla F(x^*)A^\top D_a & \nabla F(x^*)B^\top & \nabla G(x^*)\\ 0 & D_b & 0 & -A\\ 0 & 0 & 0 & -B \end{pmatrix},$$

where D_a and D_b are defined in Proposition 2.1. For convenience, we denote

$$D = \begin{pmatrix} \nabla F(x^*)A^\top D_a & \nabla F(x^*)B^\top & \nabla G(x^*)\\ D_b & 0 & -A\\ 0 & 0 & -B \end{pmatrix}.$$

Since $\varepsilon^* \geqslant 0$, hence, V is nonsingular if and only if D is nonsingular. Similar argument to that in the proof of Theorem 4.2 in [18] may show that the matrix D is nonsingular. □

4. Algorithm and convergence

In this section, We first give a description of our proposed smoothing Newton-type method for solving the equation $H(z) = 0$ and then analyze its convergence.

Algorithm 4.1.

S.0 Choose $\delta \in (0, 1)$, $\varepsilon_0 > 0$, $\beta \in (0, +\infty)$, $\sigma \in (0, \frac{1}{2})$, $\mu \geqslant 0$, and take $y_0 \in R^{n+s+t}$ as an initial point. Let $k := 0$ and $z_0 = (\varepsilon_0, y_0)$, L be a positive integer.

S.1 If $\|\nabla T(z_k)\| \leqslant \mu$, stop; otherwise, go to S.2.

S.2 If the following linear system:

$$H(z_k) + H'(z_k)\Delta z = 0 \tag{4.1}$$

is solvable and there exists m_k such that m_k is a smallest nonnegative integer satisfying

$$m_k \leqslant L, \quad T(z_k + \delta^{m_k}\Delta z_k) \leqslant (1 - 2\sigma\delta^{m_k})T(z_k), \tag{4.2}$$

then let

$$z_{k+1} := z_k + \delta^{m_k}\Delta z, \tag{4.3}$$

$k := k + 1$, and go to S.1; otherwise, go to S.3.

S.3 Let $\Delta y_k = -\nabla_y f(\varepsilon_k, y_k)$ and l_k be the smallest nonnegative integer such that

$$f(\varepsilon_k, y_k + \delta^{l_k}\Delta y_k) \leqslant f(\varepsilon_k, y_k) + \sigma\delta^{l_k}(\Delta y_k)^\top \nabla_y f(\varepsilon_k, y_k). \tag{4.4}$$

Let

$$y_{k+1} := y_k + \delta^{l_k}\Delta y_k,$$

and adjust the parameter ε as follows:

$$\varepsilon_{k+1} := \begin{cases} \frac{\varepsilon_k}{2} & \text{if } \|\nabla_y f(\varepsilon_k, y_{k+1})\| \leqslant \beta\varepsilon_k, \quad (4.5) \\ \varepsilon_k & \text{otherwise.} \quad (4.6) \end{cases}$$

Let $z_{k+1} := (\varepsilon_{k+1}, y_{k+1})$ and $k := k + 1$, go to S.1.

Now, we come to the convergence analysis of Algorithm 4.1.

Certainly, if the smoothing parameter $\varepsilon_k = 0$ at certain step, then the proposed method reduces to the generalized Newton method [18]. However, the following conclusion tells us that the positiveness of parameter ε_k would be kept throughout the computation.

Lemma 4.1. *The sequence* $\{\varepsilon_k\}$ *generated by Algorithm* 4.1 *is nonincreasing, and* $\varepsilon_k > 0$ *for all* k.

Proof. We only need to prove the assertion for the case in Step 2, since it is obvious for the case in Step 3.

From the definition of $H(z)$, a direct computation leads to that $\Delta\varepsilon_k = -(\varepsilon_k^2 + 2\varepsilon_k)/(2\varepsilon_k + 2)$, so it is easy to see that

$$-\varepsilon_k < \Delta\varepsilon_k < 0,$$

and hence,

$$0 < \varepsilon_{k+1} = \varepsilon_k + \delta^{m_k}\Delta\varepsilon_k = \left(1 - \delta^{m_k}\frac{\varepsilon_k + 2}{2\varepsilon_k + 2}\right)\varepsilon_k < \varepsilon_k. \qquad \square$$

From Lemmas 2.2 and 4.1, we know that

$$T(\varepsilon_{k-1}, y_i) \geqslant T(\varepsilon_k, y_i), \quad \forall\, k, i \in N. \tag{4.7}$$

As a consequence, we obtain the following result.

Lemma 4.2. *Suppose* $\{z_k\}$ *is a sequence generated by Algorithm* 4.1, *then the sequence* $\{T(z_k)\}$ *is monotonically decreasing*.

Proof. Obviously, $T(z_k) \leqslant T(z_{k-1})$ if z_k is generated by S.2, so it suffices to show the assertion holds for the case that z_k is generated by S.3.

If ε_{k+1} is generated by (4.6), then

$$f(\varepsilon_{k+1}, y_{k+1}) = f(\varepsilon_k, y_{k+1}) \leqslant f(\varepsilon_k, y_k).$$

Since $T(z) = \frac{1}{2}\theta^2(\varepsilon) + f(\varepsilon, y)$, we conclude that $T(z_{k+1}) \leqslant T(z_k)$.

If ε_{k+1} is generated by (4.5), then $\varepsilon_{k+1} = \varepsilon_k/2 < \varepsilon_k$. Let $\bar{z}_{k+1} = (\varepsilon_k, y_{k+1})$, then $T(\bar{z}_{k+1}) \leqslant T(z_k)$. Combining this with (4.7), we have

$$T(z_{k+1}) \leqslant T(\bar{z}_{k+1}) \leqslant T(z_k). \qquad \square$$

Theorem 4.1. *For the sequence* $\{z_k\}$ *generated by Algorithm* 4.1, *if the index set* $K = \{k \in N | z_k$ *is generated by* S.2$\}$ *is an infinite set*, *then any accumulation point of* $\{z_k\}$ *is a solution of* (2.1).

Proof. Obviously,

$$T(z_k) \leqslant (1 - 2\sigma\delta^L) T(z_{k-1}), \quad \forall k \in K.$$

Since K is an infinite index set, by Lemma 4.2, one has

$$\lim_{\substack{k\to\infty \\ k\in K}} T(z_k) = 0.$$

The desired result follows. $\square$

In the following analysis, we assume that precision $\mu = 0$ and the algorithm generates an infinite sequence.

Theorem 4.2. *Any accumulation point* z^* *of the sequence* $\{z_k\}$ *generated by Algorithm* 4.1 *is a stationary point of* (2.1).

Proof. If the index set K, defined in Theorem 4.1, is infinite, then the result follows from Theorem 4.1. So in the following analysis, we assume that the whole sequence $\{z_k\}$ is generated by Step 3.

Let $z^* = (\varepsilon^*, y^*)$ be an arbitrary accumulation point of $\{z_k\}$, then there exists an infinite subsequence $K_1 \subseteq \{1, 2, \ldots\}$ such that $\{z_k\}_{K_1} \to z^*$ as $k \in K_1$ and $k \to \infty$.

Now, we claim that $\varepsilon_k \to \varepsilon^* = 0$ as $k \to \infty$, otherwise, without loss of generality, we can assume that the whole sequence $\{z_k\}$ is generated by S.3 and ε_{k+1} is generated by (4.6), i.e., $\varepsilon_k = \varepsilon^*$ for all k. By the descent property of the algorithm, we know that any accumulation point of the sequence $\{y_k\}$ generated by Algorithm 4.1 must be a stationary point of the following optimization problem:

$$\min_{y\in R^{n+s+t}} f(\varepsilon^*, y),$$

i.e.,

$$\lim_{\substack{k\in K_1 \\ k\to\infty}} \nabla_y f(\varepsilon^*, y_k) = 0.$$

However, by (4.5) and (4.6), for all $k \in K_1$, it holds that

$$\|\nabla_y f(\varepsilon^*, y_k)\| > \beta\varepsilon^*,$$

and we arrive at a contradiction. So $\varepsilon^* = 0$, which implies that $T'_\varepsilon(z^*) = 0$.

To show $T'_y(z^*) = 0$, we consider the following two cases.

First, if the sequence $\{l_k | k \in K_1\}$ has an upper bound L^*, then by Algorithm 4.1, we know that

$$f(\varepsilon_k, y_{k+1}) \leqslant f(\varepsilon_k, y_k) - \sigma\delta^{L^*}\|\nabla_y f(\varepsilon_k, y_k)\|^2,$$

i.e.,

$$f(\varepsilon_k, y_k) - f(\varepsilon_k, y_{k+1}) \geqslant \sigma\delta^{L^*}\|\nabla_y f(\varepsilon_k, y_k)\|^2.$$

Let $\bar{z}_k = (\varepsilon_k, y_{k+1})$. Then from $\varepsilon_k \geqslant \varepsilon_{k+1}$, we have

$$T(\bar{z}_k) \geqslant T(z_{k+1}).$$

Therefore,

$$T(z_k) - T(z_{k+1}) \geqslant \sigma\delta^{L^*}\|\nabla_y f(\varepsilon_k, y_k)\|^2.$$

Since $\{T(z_k)\}$ is nonincreasing and bounded from below, we have

$$\nabla_y f(\varepsilon_k, y_k) \to 0, \quad k \in K_1, k \to \infty,$$

i.e.,

$$\nabla_y f(0, y^*) = 0.$$

Second, if the sequence $\{l_k | k \in K_1\}$ has no bound, then there exists a subsequence $K_2 \subseteq K_1$ such that $l_k \to \infty$ as $k \in K_2$ and $k \to \infty$. From the stepsize rule, we know that

$$f(\varepsilon_k, y_k + \delta^{l_k-1}\Delta y_k) > f(\varepsilon_k, y_k) - \sigma\delta^{l_k-1}\|\nabla_y f(\varepsilon_k, y_k)\|^2.$$

By the Mean-Value theorem, there exists $\bar{y}_k = y_k + \mu_1\delta^{l_k-1}\Delta y_k$ with $\mu_1 \in (0, 1)$ such that

$$\nabla_y f(\varepsilon_k, \bar{y}_k)^\top \nabla_y f(\varepsilon_k, y_k) < \sigma\|\nabla_y f(\varepsilon_k, y_k)\|^2.$$

Thus,

$$\langle \nabla_y f(\varepsilon_k, y_k) - \nabla_y f(\varepsilon_k, \bar{y}_k), \nabla_y f(\varepsilon_k, y_k)\rangle > (1-\sigma)\|\nabla_y f(\varepsilon_k, y_k)\|^2.$$

Using Cauchy–Schwartz inequality, we have

$$\|\nabla_y f(\varepsilon_k, \bar{y}_k) - \nabla_y f(\varepsilon_k, y_k)\| \geqslant (1-\sigma)\|\nabla_y f(\varepsilon_k, y_k)\|. \tag{4.8}$$

From the fact that $z_k \to z^*$ and $l_k \to \infty$ as $k \in K_2, k \to \infty$, we know that $\lim_{\substack{k\in K_2\\ k\to\infty}} y_k = y^*$. Consequently, there exists a closed neighborhood $\mathcal{N}(y^*, \delta^*)$ of y^* such that for sufficiently large $k \in K_2$, $y_k, \bar{y}_k \in \mathcal{N}(y^*, \delta^*)$. From the uniform continuity of function $\nabla_y f(\varepsilon_k, y)$, we have

$$\lim_{\substack{k\in K_2\\ k\to\infty}} \|\nabla_y f(\varepsilon_k, \bar{y}_k) - \nabla_y f(\varepsilon_k, y_k)\| = 0.$$

Recalling (4.8), one has

$$\lim_{\substack{k\in K_2\\ k\to\infty}} \nabla_y f(\varepsilon_k, y_k) = 0.$$

By the definition of $f(\varepsilon, y)$ and $T(z)$, we have

$$\nabla_y T(z^*) = \nabla_y f(z^*) = 0.$$

Combining this result with $\varepsilon^* = 0$, we conclude that z^* is a stationary point of (2.1). □

Theorem 4.3. *Let* $\{z_k\}$ *be generated by Algorithm* 4.1. *Suppose* z^* *is an accumulation point of* $\{z_k\}$ *and a BD-regular solution of* $H(z) = 0$, *then*

(1) *the point* x^* *is a solution of the GNCP*;
(2) *the sequence* $\{\varepsilon_k\}$ *converges to* 0 *quadratically*;
(3) *the sequence* $\{z_k\}$ *converges to* z^* *superlinearly. In particular, if* F' *and* G' *are locally Lipschitz continuous at* z^*, *then* $\{z_k\}$ *converges to* z^* *Q-quadratically*.

Proof. The first statement is obvious and we omit the proof.

We first prove (3). Since H is semismooth at z^* according to Proposition 2.1, and z^* is a BD-regular solution from the assumption, we can conclude from [12, Proposition 3] that z^* is a unique solution of $H(z) = 0$ and therefore it is an isolated solution of the GNCP. That is, z^* is an isolated accumulated point of $\{z_k\}$. On the other hand, from Theorem 4.2, we know that $\lim_{k\to\infty}\|z_k - z_{k-1}\| = 0$. From [11, Proposition 6.1] we can obtain the global convergence of the generated sequence $\{z_k\}$.

Using [13, Theorem 4.3], for sufficiently large k, we have $m_k = 0$ i.e., $\delta^{m_k} = 1$. By Theorem 3.1 [13], we conclude that the sequence $\{z_k\}$ converges to z^* superlinearly (quadratically). Hence (3) holds.

To prove (2), by (3), we know for sufficiently large k, z_k is generated by (4.3), and hence from (4.1), we know that

$$\varepsilon_{k+1} = \varepsilon_k + \Delta\varepsilon_k = \varepsilon_k - \frac{(\varepsilon_k + 1)^2 - 1}{2(\varepsilon_k + 1)} = \frac{(\varepsilon_k)^2}{2(\varepsilon_k + 1)},$$

from which we obtain that

$$\lim_{k\to\infty} \frac{\varepsilon_{k+1}}{(\varepsilon_k)^2} = \lim_{k\to\infty} \frac{1}{2(\varepsilon_k + 1)} = \frac{1}{2}.$$

This completes the proof. □

5. Computational experiments

In this section, we give two sets of numerical experiments, and throughout our computational experiments, the parameters in Algorithm 4.1 are set as

$$\alpha = 0.01, \quad L = 10, \quad \beta = 0.8, \quad \sigma = 0.4, \quad \delta = 0.5.$$

Example 5.1. This example is an implicit complementarity problems [1,9] with the following form: find $y \in R^n$ such that

$$y - m(y) \geqslant 0, \quad F(y) \geqslant 0, \quad F(y)^\top (y - m(y)) = 0,$$

where $m_i : R^n \to R$, $i = 1, \ldots, n$, and

$$F(y) = Ay + b = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix} y + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

$m(y) = \phi(Ay + b)$ with $\phi : R^n \to R^n$ being twice continuously differentiable. The following choices of function ϕ define our test problems:

$$\text{POZ1}: \ \phi_i(x) = -0.5 - x_i, \quad i = 1, 2, \ldots, n,$$

$$\text{POZ2}: \ \phi_i(x) = -1.5x_i + 0.25x_i^2, \quad i = 1, 2, \ldots, n.$$

Table 1
Numerical results of Example 1

ϕ	ST	Iter	f^*	CPU	ε_{-2}	ε_{-1}
(1)	(a)	7	7.8886×10^{-31}	0.1870	8.8886×10^{-10}	3.9504×10^{-19}
(2)	(a)	8	7.7429×10^{-46}	0.2030	7.3518×10^{-12}	2.7026×10^{-23}
(1)	(b)	7	1.7256×10^{-31}	0.1720	8.8886×10^{-10}	3.9504×10^{-19}
(2)	(b)	7	109769×10^{-39}	0.1720	8.8886×10^{-10}	3.9504×10^{-19}
(1)	(c)	7	2.4649×10^{-32}	0.1880	8.8886×10^{-10}	3.9504×10^{-19}
(2)	(c)	8	2.7733×10^{-31}	0.1880	8.5546×10^{-13}	3.6625×10^{-25}
(1)	(d)	7	1.9720×10^{-31}	0.1880	8.8886×10^{-10}	3.9504×10^{-19}
(2)	(d)	8	1.7964×10^{-33}	0.1880	1.0842×10^{-8}	5.8886×10^{-17}

Table 2
Numerical results of Example 2

s	r	Family	SP	Iter	Inner	CPU
3	0.1	(1)	0.333	6.6	10.2	0.2190
		(2)	1	8.67	17.1	0.1875
	1	(1)	0.417	5.4	7	0.2500
		(2)	1	5.2	5.6	0.2250
	10	(1)	0.333	6.5	9.667	0.2317
		(2)	1	6.6	7.8	0.2280
5	0.1	(1)	0.4	7.75	14.25	0.2382
		(2)	1	6.83	13.67	0.2345
	1	(1)	0.5	6.83	10.67	0.2055
		(2)	1	7.8	10.6	0.1996
	10	(1)	0.3	8.67	17.67	0.1976
		(2)	1	9.2	14.25	0.1938
9	0.1	(1)	0.3	11	16.6	0.3040
		(2)	1	9.8	17.6	0.2528
	1	(1)	0.5	7.67	17.2	0.2590
		(2)	1	7.4	10.6	0.2190
	10	(1)	0.5	9.75	16.5	0.2070
		(2)	1	9.71	15.16	0.2090
12	0.1	(1)	0.3	11.33	29	0.2760
		(2)	1	12.6	33.4	0.2132
	1	(1)	0.462	8.83	15	0.2215
		(2)	1	11.71	22.85	0.1928
	10	(1)	0.5	11.11	19	0.1873
		(2)	1	10.8	19.2	0.1876

In Table 1, Iter denotes the iterative number, f^* is the final value of f when the algorithm terminates, and CPU denotes the computing time in the computer, ε_{-2} and ε_{-1} denote the values of ε_k at the last two iterates, respectively. Table 1 gives the numerical experiment of this example with termination parameter $\mu = 10^{-14}$ and the following starting points with $\varepsilon = 5$, denoted by ST:

(a) $(0, 0, \ldots, 0)^\top$, (b) $(-0.5, -0.5, \ldots, -0.5)^\top$,
(c) $(-1, -1, \ldots, -1)^\top$, (d) $(0.5, 0.5, \ldots, 0.5)^\top$.

From Table 1, we can see that the parameter ε does not arrive at zero before the algorithm terminates and the algorithm is very efficient in solving the problem.

Example 5.2. Consider the problem of finding $x^* \in R^n$ such that

$$\begin{cases} x \in \mathcal{K} = \{v \in R^n | Av \geqslant 0\}, \\ Nx + d \in \mathcal{K}^\circ = \{v \in R^n | v = A^\top \lambda, \lambda \in R^s_+\}, \\ x^\top (Nx + d) = 0. \end{cases}$$

The matrix A and vector d will be generated following the pattern similar to [18, Example 6.2]. For the matrix N, different from the choice in [18], we always let it be symmetric, i.e., $N = QD_NQ$, where Q is an Householder matrix $Q = I - 2uu^\top/\|u\|^2$ and the components u are generated randomly from the interval $(-1, 1)$, $D_N \in R^{n\times n}$ is the diagonal matrix whose diagonal elements are generated randomly from $(-10, 10)$.

For this problem, we only consider the following two cases: (1) N is indefinite and (2) N is positive definite. Similar to the notation used in [18], we call a case successful if the value T is less than 10^{-10} in 1 minute, SP denotes the successful rate. For all successful cases, the following data are included: Iter denotes the average number of iterations, Inner denotes the average number of the inner iterations, CPU denotes the working time of the computer excluding input/output time.

The numerical results are reported in Table 2 with $z = (0, 0, \ldots, 0)$ and $\varepsilon = 2$ being the starting point. From the numerical results we can see our algorithm perform well when it is applied to the problem in symmetric and positive definite case.

Acknowledgments

The authors wish to give sincere thanks to the anonymous referees for their valuable suggestions and helpful comments on the paper.

References

[1] R. Andreani, A. Friedlander, S.A. Santos, On the resolution of the generalized nonlinear complementarity problem, SIAM J. Optim. 12 (2001) 303–321.
[2] B.T. Chen, P.T. Harker, Smooth approximations to nonlinear complementarity problem, SIAM J. Optim. 7 (2) (1997) 403–420.
[3] X.J. Chen, Smoothing methods for complementarity problems and their applications: a survey, J. Oper. Res. Soc. Japan 43 (1) (2000) 32–47.
[4] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, NY, 1983.
[5] S. Engelke, C. Kanzow, On the solution of linear programs by Jacobian smoothing methods, Ann. Oper. Res. 103 (2001) 49–70.
[6] F. Facchinei, J.S. Pang, Finite-dimensional Variational Inequalities and Complementarity Problems, Springer, New York, NY, 2003.
[7] F. Facchinei, J. Soares, A new merit function for nonlinear complementarity problems and a related algorithm, SIAM J. Optim. 7 (1997) 225–247.
[8] A. Fischer, A special Newton-type optimization method, Optimization 24 (1992) 269–284.
[9] H. Jiang, M. Fukushima, L. Qi, D. Sun, A trust region method for solving generalized complementarity problems, SIAM J. Optim. 8 (1998) 140–157.
[10] C. Kanzow, M. Fukushima, Equivalence of the generalized complementarity problem to differentiable unconstrained minimization, J. Optim. Theory Appl. 90 (1996) 581–603.
[11] C. Kanzow, H. Pieper, Jacobian smoothing methods for general nonlinear complementarity problems, SIAM J. Optim. 9 (1999) 342–372.
[12] J.S. Pang, L. Qi, Nonsmooth equations: motivation and algorithms, SIAM J. Optim. 3 (1993) 443–465.
[13] L. Qi, Convergence analysis of some algorithms for solving nonsmooth equations, Math. Oper. Res. 18 (1993) 227–244.
[14] L. Qi, D. Sun, G.L. Zhou, A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities, Math. Programming 87 (1) (2000) 1–35.
[15] L. Qi, J. Sun, A nonsmooth version of Newton's method, Math. Programming 58 (1993) 353–367.
[16] L. Qi, G.L. Zhou, A smoothing Newton method for minimizing a sum of Euclidean norms, SIAM J. Optim. 11 (2) (2000) 389–410.
[17] P. Tseng, N. Yamashita, M. Fukushima, Equivalence of complementarity problems to differentiable minimization: a unified approach, SIAM J. Optim. 6 (1996) 446–460.
[18] Y.J. Wang, F.M. Ma, J.Z. Zhang, A nonsmooth L-M method for solving the generalized nonlinear complementarity problem, Appl. Math. Optim. 52 (2005) 73–92.