Math. Program., Ser. A 87: 1–35 (2000)

Digital Object Identifier (DOI) 10.1007/s101079900127

Liqun Qi · Defeng Sun · Guanglu Zhou

A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities*

Received June 23, 1997 / Revised version received July 29, 1999
Published online December 15, 1999

Abstract. In this paper we take a new look at smoothing Newton methods for solving the nonlinear complementarity problem (NCP) and the box constrained variational inequalities (BVI). Instead of using an infinite sequence of smoothing approximation functions, we use a single smoothing approximation function and Robinson's normal equation to reformulate NCP and BVI as an equivalent nonsmooth equation $H(u, x) = 0$, where $H : \Re^{2n} \to \Re^{2n}$, $u \in \Re^n$ is a parameter variable and $x \in \Re^n$ is the original variable. The central idea of our smoothing Newton methods is that we construct a sequence $\{z^k = (u^k, x^k)\}$ such that the mapping $H(\cdot)$ is continuously differentiable at each z^k and may be non-differentiable at the limiting point of $\{z^k\}$. We prove that three most often used Gabriel-Moré smoothing functions can generate strongly semismooth functions, which play a fundamental role in establishing superlinear and quadratic convergence of our new smoothing Newton methods. We do not require any function value of F or its derivative value outside the feasible region while at each step we only solve a linear system of equations and if we choose a certain smoothing function only a reduced form needs to be solved. Preliminary numerical results show that the proposed methods for particularly chosen smoothing functions are very promising.

Key words. variational inequalities – nonsmooth equations – smoothing approximation – smoothing Newton method – convergence

1. Introduction

Consider the variational inequalities (VI for abbreviation): Find $y^* \in X$ such that

$$(y - y^*)^T F(y^*) \geq 0 \quad \text{for all } y \in X, \tag{1}$$

where X is a nonempty closed subset of $\Re^n$ and $F : D \to \Re^n$ is continuously differentiable on some open set D, which contains X. In this paper, unless otherwise stated, we assume that

$$X := \{y \in \Re^n \,|\, a \leq y \leq b\}, \tag{2}$$

where $a \in \{\Re \cup \{-\infty\}\}^n$, $b \in \{\Re \cup \{\infty\}\}^n$ and $a < b$. Then (1) becomes the box constrained variational inequalities (BVI for abbreviation). This assumption is not restrictive because if in (1) the set X is not of the form (2) but is represented by several

L. Qi, D. Sun, G. Zhou: School of Mathematics, The University of New South Wales, Sydney 2052, Australia
e-mail: L.Qi@unsw.edu.au, sun@maths.unsw.edu.au, zhou@maths.unsw.edu.au

Mathematics Subject Classification (1991): 90C33, 90C30, 65H10

* This work is supported by the Australian Research Council.

equalities and inequalities, then under standard constraint qualifications [25] we can equivalently transform (1) into a new VI with the constraint set of form (2), possibly with increased dimension (see, e.g., [54]). When $X = \Re^n_+$, VI reduces to the nonlinear complementarity problem (NCP for abbreviation): Find $y^* \in \Re^n_+$ such that

$$F(y^*) \in \Re^n_+ \quad \text{and} \quad F(y^*)^T y^* = 0. \tag{3}$$

It is well known (see, e.g., [25]) that solving (1) is equivalent to finding a root of the following equation:

$$W(y) := y - \Pi_X(y - F(y)) = 0, \tag{4}$$

where for any $x \in \Re^n$, $\Pi_X(x)$ is the Euclidean projection of x onto X and X is a nonempty closed convex subset of $\Re^n$, which is not necessarily of the form (2). It is also well known that if X is a closed convex subset of $\Re^n$, then solving VI is equivalent to solving the following Robinson's normal equation

$$E(x) := F(\Pi_X(x)) + x - \Pi_X(x) = 0 \tag{5}$$

in the sense that if $x^* \in \Re^n$ is a solution of (5) then $y^* := \Pi_X(x^*)$ is a solution of (1), and conversely if y^* is a solution of (1) then $x^* := y^* - F(y^*)$ is a solution of (5) [49]. Both (4) and (5) are nonsmooth equations and have led to various generalized Newton's methods. See [25], [40], [22] and [18] for a review of these methods.

By using the Gabriel-Moré smoothing function for $\Pi_X(\cdot)$, we can construct approximations for $E(\cdot)$:

$$G(u, x) := F(p(u, x)) + x - p(u, x), \quad (u, x) \in \Re^n \times \Re^n, \tag{6}$$

where for each $i \in N := \{1, 2, ..., n\}$, $p_i(u, x) = q(u_i, a_i, b_i, x_i)$ and for any $(\mu, c, d, w) \in \Re \times \Re \cup \{-\infty\} \times \Re \cup \{\infty\} \times \Re$ with $c \leq d$, $q(\mu, c, d, w)$ is defined by

$$q(\mu, c, d, w) = \begin{cases} \phi(|\mu|, c, d, w) & \text{if } \mu \neq 0 \\ \Pi_{[c,d]\cap\Re}(w) & \text{if } \mu = 0 \end{cases} \tag{7}$$

and $\phi(\mu, c, d, w)$, $(\mu, w) \in \Re_{++} \times \Re$ is a Gabriel-Moré smoothing approximation function [23], also, see Sect. 2 for the definition of $\phi(\cdot)$. For example, for NCP we can take the Chen-Harker-Kanzow-Smale smoothing NCP function [4,31,51]

$$\phi(\mu, 0, \infty, w) = \frac{\sqrt{w^2 + 4\mu^2} + w}{2}, \quad (\mu, w) \in \Re_{++} \times \Re,$$

which is a special Gabriel-Moré smoothing function. In this paper, unless otherwise stated, we always assume that $c \in \Re \cup \{-\infty\}$, $d \in \Re \cup \{\infty\}$ and $c \leq d$. By Lemma 2.2

of [23], for any $(\mu, w) \in \Re_{++} \times \Re$,

$$\phi(\mu, c, d, w) \in [c, d] \cap \Re,$$

and so, for any $(u, x) \in \Re^n \times \Re^n$,

$$p(u, x) \in X. \tag{8}$$

Then the mapping $G(\cdot)$ is defined on $\Re^{2n}$ while $F(\cdot)$ is only required to have definition on X, the feasible region. It is noted that we can also use $p(\cdot)$ to construct a class of approximation functions for $W(\cdot)$ defined in (4):

$$V(u, y) := y - p(u, y - F(y)), \quad (u, y) \in \Re^n \times \Re^n. \tag{9}$$

However, in order to make $V(\cdot)$ to have definition on the whole space $\Re^{2n}$ one has to assume that $F(\cdot)$ is well defined on the whole space $\Re^n$. This requirement for F is not satisfied for many NCPs and BVI transformed from economic equilibrium problems [18]. Moreover, even if F has definition on the whole space $\Re^n$, some important properties of F, like monotonicity, which holds on X, may not hold outside X. These observations lead us to focus on the approximation functions defined by (6) rather than (9). However, the techniques used here can be applied to (9) too.

For the sake of convenience, let $\phi_{cd} : \Re_{++} \times \Re \to \Re$ be defined by

$$\phi_{cd}(\mu, w) := \phi(\mu, c, d, w), \quad (\mu, w) \in \Re_{++} \times \Re \tag{10}$$

and for any given $\mu \in \Re_{++}$, let $\phi_{\mu cd} : \Re \to \Re$ be defined by

$$\phi_{\mu cd}(w) := \phi(\mu, c, d, w), \quad w \in \Re. \tag{11}$$

Then, for any given $\mu \in \Re_{++}$, $\phi_{\mu cd}(\cdot)$ is continuously differentiable at any $w \in \Re$ [23]. Moreover, for several most often used Gabriel-Moré smoothing functions it can be verified that $\phi_{cd}(\cdot)$ is also continuously differentiable at any $(\mu, w) \in \Re_{++} \times \Re$. In this paper, we are interested in smoothing functions with this property, which we make it as an assumption.

Assumption 1. *The function $\phi_{cd}(\cdot)$ is continuously differentiable at any $(\mu, w) \in \Re_{++} \times \Re$.*

Let $z := (u, x) \in \Re^n \times \Re^n$ and define $H : \Re^{2n} \to \Re^{2n}$ by

$$H(z) := \begin{pmatrix} u \\ G(z) \end{pmatrix}. \tag{12}$$

Then it is easy to see that H is continuously differentiable at any $z \in \Re^n_{++} \times \Re^n$ if Assumption 1 is satisfied.

Recently, smoothing Newton methods have attracted a lot of attention in the literature partially due to their superior numerical performance [2], e.g., see [3–12, 26, 33, 47, 46, 58] and references therein. Among them the first globally and superlinearly (quadratically) convergent smoothing Newton method was proposed by Chen, Qi and Sun in [11], where the authors exploited a Jacobian consistency property and applied this property to an infinite sequence of smoothing approximation functions to get high-order convergent methods. They dealt with general box constrained variational inequalities. However, even for NCP they had to assume that F had definition on the whole space $\Re^n$. The result of [11] has been further investigated by Chen and Ye [12] still with the same requirement. On the contrary, here we avoid this requirement by making use of the mapping $H(\cdot)$ and most importantly, we use only one smoothing approximation function instead of using an infinite sequence of those functions. In this way we make our smoothing Newton methods much simpler. It is deserved to point out that the Jacobian consistency property may not hold for the smoothing function (6) if F is not globally Lipschitz continuous, and there is no high-order convergent methods based on (6). To treat the smoothing parameter u as a free variable may restrict the updating rules for choosing u. However, by doing so we can provide a globally and superlinearly convergent method for solving $H(z) = 0$. The idea of using (6) for solving the NCP was suggested by Chen, Harker and Pınar in [7]. Here we first study the smooth and semismoothness propertiies of (6) about u and x jointly and then use these properties to get globally and superlinearly convergent results based on (12). Chen, Harker and Pınar also pointed out that by choosing smooth functions with finite-support, the resulting Newton equation has a reduced dimension. This property carries to the generalized form (6).

There are few globally and superlinearly convergent methods in the literature dealing with NCP and BVI with requiring F defined on X only while at each step only solving a linear system of equations. In [24], by combining a modified extragradient method [52] and a generalized Newton method, Han and Sun gave such an algorithm for solving pseudomonotone variational inequalities with X being a nonempty closed convex subset represented by several twice continuously differentiable inequalities. Very recently, Kanzow and Qi [34] designed a QP-free constrained Newton-type method for BVI, with assumption $b_i = \infty, i \in N$, by combining an updated ε-active projected gradient direction and a modified Gauss-Newton direction. The result of Kanzow and Qi is based on the Fischer-Burmeister function [19], which recently has received a lot of attention in the fields of NCP and BVI, e.g., see [14–16, 20, 22, 28, 29, 35, 55] and references therein. However, it is believed that the Newton-type direction is much better than either the extragradient direction or the projected gradient direction. In this paper, instead of resorting to some hybrid techniques, at each step we use one minor modified Newton direction. This modification is crucial to the design of our algorithms.

The organization of this paper is as follows. In the next section we study some preliminary properties of smoothing functions. In Sect. 3 we prove that $\phi_{cd}(\cdot)$ is strongly semismooth with several particularly chosen $\phi(\cdot)$. In Sect. 4 we state the algorithm and prove several propositions related to the algorithm. In Sect. 5 we establish the global convergence of the algorithm. In Sect. 6 we study under what conditions the level sets of the merit function $\psi(\cdot) = \|H(\cdot)\|^2$ are bounded. We analyze the superlinear and quadratic convergence properties of the algorithm in Sect. 7 and give preliminary numerical results in Sect. 8. Final conclusions are given in Sect. 9.

A word about our notation is in order. For a continuously differentiable function $\Phi : \Re^m \to \Re^m$, we denote the Jacobian of Φ at $x \in \Re^m$ by $\Phi'(x)$, whereas the transposed Jacobian as $\nabla\Phi(x)$. $\|\cdot\|$ denotes the Euclidean norm. If W is an $m \times m$ matrix with entries W_{jk}, $j, k = 1, \ldots, m$, and $\mathcal{J}$ and $\mathcal{K}$ are index sets such that $\mathcal{J}, \mathcal{K} \subseteq \{1, \ldots, m\}$, we denote by $W_{\mathcal{J}\mathcal{K}}$ the $|\mathcal{J}| \times |\mathcal{K}|$ sub-matrix of W consisting of entries W_{jk}, $j \in \mathcal{J}$, $k \in \mathcal{K}$. If $W_{\mathcal{J}\mathcal{J}}$ is nonsingular, we denote by $W/W_{\mathcal{J}\mathcal{J}}$ the Shur-complement of $W_{\mathcal{J}\mathcal{J}}$ in W, i.e., $W/W_{\mathcal{J}\mathcal{J}} := W_{\mathcal{K}\mathcal{K}} - W_{\mathcal{K}\mathcal{J}} W_{\mathcal{J}\mathcal{J}}^{-1} W_{\mathcal{J}\mathcal{K}}$, where $\mathcal{K} = \{1, \ldots, m\} \backslash \mathcal{J}$. If w is an m vector, we denote by $w_{\mathcal{J}}$ the sub-vector with components $j \in \mathcal{J}$.

2. Some preliminaries

In this section we give some properties related to smoothing functions. In [10], Chen and Mangasarian introduced a class of smoothing approximation functions for $\max\{0, w\}$, $w \in \Re$. Gabriel and Moré [23] extended Chen-Mangasarian's smoothing approach to $\Pi_{[c,d]\cap\Re}(w)$, $w \in \Re$. Let $\rho : \Re \to \Re_+$ be a density function, i.e., $\rho(s) \geq 0$ and $\int_{-\infty}^{\infty} \rho(s)ds = 1$, with a bounded absolute mean, that is

$$\kappa := \int_{-\infty}^{\infty} |s|\rho(s)ds < \infty. \tag{13}$$

Recall that for any three numbers $c \in \Re \cup \{-\infty\}$, $d \in \Re \cup \{\infty\}$ with $c \leq d$ and $e \in \Re$, the median function $\text{mid}(\cdot)$ is defined by

$$\text{mid}(c, d, e) = \Pi_{[c,d]\cap\Re}(e) = \begin{cases} c & \text{if } e < c \\ e & \text{if } c \leq e \leq d \\ d & \text{if } d < e \end{cases}.$$

Then the Gabriel-Moré smoothing function $\phi(\mu, c, d, w)$ for $\Pi_{[c,d]\cap\Re}(w)$ [23] is defined by

$$\phi(\mu, c, d, w) = \int_{-\infty}^{\infty} \text{mid}(c, d, w - \mu s)\rho(s)ds, \quad (\mu, w) \in \Re_{++} \times \Re. \tag{14}$$

If $c = -\infty$ and/or $d = \infty$, the value of ϕ takes the limit of ϕ as $c \to -\infty$ and/or $d \to \infty$, correspondingly. For example, if c is finite and $d = \infty$, then

$$\phi(\mu, c, \infty, w) = \lim_{d' \to \infty} \phi(u, c, d', w), \quad (u, w) \in \Re_{++} \times \Re.$$

Let

$$\text{supp}(\rho) = \{s \in \Re : \rho(s) > 0\}.$$

Lemma 1. *[23, Lemma 2.3] For any given $\mu > 0$, the mapping $\phi_{\mu cd}(\cdot)$ is continuously differentiable with*

$$\phi'_{\mu cd}(w) = \int_{(w-d)/\mu}^{(w-c)/\mu} \rho(s)ds,$$

where $\phi_{\mu cd}(\cdot)$ is defined by (11). In particular, $\phi'_{\mu cd}(w) \in [0, 1]$. Furthermore, if $\text{supp}(\rho) = \Re$ *and at least one of c and d is finite, then $\phi'_{\mu cd}(w) \in (0, 1)$.*

Let $q_{cd} : \Re^2 \to \Re$ be defined by

$$q_{cd}(\mu, w) = q(\mu, c, d, w), \quad (\mu, w) \in \Re^2, \tag{15}$$

where $q(\mu, c, d, w)$ is defined by (7). Then we have the following lemma.

Lemma 2. *The mapping $q_{cd}(\cdot)$ defined by (15) is Lipschitz continuous on $\Re^2$ with Lipschitz constant $L := 2\max\{1, \kappa\}$.*

Proof. Suppose that (μ_1, w_1) and (μ_2, w_2) are two arbitrary points of $\Re^2$. Then, since the mapping $\mathrm{mid}(c, d, \cdot)$ is non-expansive, we have

$$\begin{aligned}
&|q_{cd}(\mu_1, w_1) - q_{cd}(\mu_2, w_2)| \\
&= \left| \int_{-\infty}^{\infty} \mathrm{mid}(c, d, w_1 - |\mu_1|s)\rho(s)ds - \int_{-\infty}^{\infty} \mathrm{mid}(c, d, w_2 - |\mu_2|s)\rho(s)ds \right| \\
&\le \int_{-\infty}^{\infty} |\mathrm{mid}(c, d, w_1 - |\mu_1|s) - \mathrm{mid}(c, d, w_2 - |\mu_2|s)|\rho(s)ds \\
&\le \int_{-\infty}^{\infty} |(w_1 - |\mu_1|s) - (w_2 - |\mu_2|s)|\rho(s)ds \\
&\le \int_{-\infty}^{\infty} |w_1 - w_2|\rho(s)ds + \int_{-\infty}^{\infty} |\mu_1 - \mu_2||s|\rho(s)ds \\
&= |w_1 - w_2| + \kappa|\mu_1 - \mu_2| \\
&\le 2\max\{1, \kappa\}\|(\mu_1, w_1) - (u_2, w_2)\|,
\end{aligned}$$

which completes the proof of this lemma.

□

The following examples are three most often used Gabriel-Moré smoothing functions in the literature.

Example 1. Neural Networks Smoothing Function
The density function is

$$\rho(s) = \frac{e^{-s}}{(1 + e^{-s})^2}.$$

We have $\kappa = \log 2$, $\mathrm{supp}(\rho) = \Re$ and the smoothing function

$$\phi(\mu, c, d, w) = d + \mu \ln\left\{1 + e^{(c-w)/\mu}\right\} - \mu \ln\left\{1 + e^{(d-w)/\mu}\right\}, \quad (\mu, w) \in \Re_{++} \times \Re. \tag{16}$$

Then it is easy to see that $\phi_{cd}(\cdot)$ is continuously differentiable at any $(\mu, w) \in \Re_{++} \times \Re$, i.e., Assumption 1 is satisfied for this smoothing function. If $c = 0$ and $d = \infty$, then the smoothing function in (16) reduces to the neural networks smoothing plus function [9]:

$$\phi(\mu, 0, \infty, w) = w + \mu \ln(1 + e^{-w/\mu}), \quad (\mu, w) \in \Re_{++} \times \Re. \tag{17}$$

The latter has been shown to have superior smoothing properties in global optimization work of Moré and Wu [39].

Example 2. Chen-Harker-Kanzow-Smale Smoothing Function
The density function is

$$\rho(s) = \frac{2}{(s^2+4)^{3/2}}.$$

We have $\kappa = 1$, $\text{supp}(\rho) = \Re$ and the smoothing function

$$\phi(\mu, c, d, w) = \frac{c + \sqrt{(c-w)^2 + 4\mu^2}}{2} + \frac{d - \sqrt{(d-w)^2 + 4\mu^2}}{2}, \quad (\mu, w) \in \Re_{++} \times \Re. \tag{18}$$

Apparently, $\phi_{cd}(\cdot)$ is continuously differentiable at any $(\mu, w) \in \Re_{++} \times \Re$, i.e., Assumption 1 is satisfied for this smoothing function. If $c = 0$ and $d = \infty$, then the smoothing function in (18) reduces to the Chen-Harker-Kanzow-Smale smoothing NCP function:

$$\phi(\mu, 0, \infty, w) = \frac{\sqrt{w^2 + 4\mu^2} + w}{2}, \quad (\mu, w) \in \Re_{++} \times \Re. \tag{19}$$

Example 3. Uniform Smoothing Function
The density function is

$$\rho(s) = \begin{cases} 1 & \text{if } -\frac{1}{2} \le s \le \frac{1}{2} \\ 0 & \text{otherwise} \end{cases}.$$

We have $\kappa = \frac{1}{8}$, $\text{supp}(\rho) = [-\frac{1}{2}, \frac{1}{2}]$ and for any $(\mu, w) \in \Re_{++} \times \Re$, the smoothing function

$$\phi(\mu, c, d, w) = \begin{cases} \frac{w}{\mu}(d-c) + \frac{1}{2}(d+c)^2 + \frac{1}{2\mu}(c^2 - d^2) & \text{if } |w-c| < \mu/2,\ |w-d| < \mu/2 \\ \frac{1}{2}[w - \mu/4 + d - (w-d)^2/\mu] & \text{if } |w-d| < \mu/2,\ w-c > \mu/2 \\ \frac{1}{2}[w + \mu/4 + c + (w-c)^2/\mu] & \text{if } |w-c| < \mu/2,\ w-d < -\mu/2 \\ \text{mid}(c, d, w) & \text{otherwise} \end{cases}. \tag{20}$$

By direct computation, we can see that $\phi_{cd}(\cdot)$ is continuously differentiable at any $(\mu, w) \in \Re_{++} \times \Re$, i.e., Assumption 1 is satisfied for this smoothing function. If $0 < \mu \le d - c$, then the smoothing function $\phi(\cdot)$ has the following simple form:

$$\phi(\mu, c, d, w) = \begin{cases} \frac{1}{2}[w - \mu/4 + d - (w-d)^2/\mu] & \text{if } |w-d| < \mu/2 \\ \frac{1}{2}[w + \mu/4 + c + (w-c)^2/\mu] & \text{if } |w-c| < \mu/2 \\ \text{mid}(c, d, w) & \text{otherwise} \end{cases}. \tag{21}$$

If $c = 0$ and $d = \infty$, the function in (20) is the Zang smoothing plus function [59]:

$$\phi(\mu, 0, \infty, w) = \begin{cases} 0 & \text{if } w \le -\mu/2 \\ \dfrac{1}{2\mu}(w + \mu/2)^2 & \text{if } |w| < \mu/2 \\ w & \text{if } w \ge \mu/2 \end{cases}, \quad (\mu, w) \in \Re_{++} \times \Re. \tag{22}$$

Similar functions to (22) can be found in [42] and [46].

Theorem 1. *Suppose that Assumption 1 holds for a chosen smoothing function $\phi(\mu, c, d, w)$, $(\mu, w) \in \Re_{++} \times \Re$. Then*

(i) The mapping $H(\cdot)$ is continuously differentiable at any $z = (u, x) \in \Re^n_{++} \times \Re^n$ and

$$H'(z) = \begin{pmatrix} I & 0 \\ (F'(p(z)) - I)D(u) & F'(p(z))C(x) + I - C(x) \end{pmatrix}, \tag{23}$$

where $D(u) = \mathrm{diag}\{d_i(u),\ i \in N\}$, $C(x) = \mathrm{diag}\{c_i(x),\ i \in N\}$ and $d_i(u) = \partial p_i(u, x)/\partial u_i$, $c_i(x) = \partial p_i(u, x)/\partial x_i$ and $c_i(x) \in [0, 1]$, $i \in N$.

(ii) Suppose that for some $z \in \Re^n_{++} \times \Re^n$, $F'(p(z))$ is a P_0-matrix, i.e., its every principal minor is nonnegative. Then $H'(z)$ is nonsingular if $\mathrm{supp}(\rho) = \Re$ and for each $i \in N$, at least one of a_i and b_i is finite.

(iii) If for some $z \in \Re^n_{++} \times \Re^n$, $F'(p(z))$ is a P-matrix, i.e., its every principal minor is positive, then $H'(z)$ is nonsingular.

Proof. (i) Since Assumption 1 is satisfied for $\phi(\cdot)$, from the definition, we know that $H(\cdot)$ is continuously differentiable at any $z = (u, x) \in \Re^n_{++} \times \Re^n$. By direct computation we have (23). From Lemma 1 and the definition of $p_i(\cdot)$, $c_i(x) \in [0, 1]$, $i \in N$.
(ii) Under the assumptions, from Lemma 1, $c_i(x) \in (0, 1)$, $i \in N$. Then, it is easy to see that $F'(p(z))C(x) + I - C(x)$ is nonsingular under the assumption that $F'(p(z))$ is a P_0-matrix, see, e.g., [4, Theorem 3.3]. It then follows from (23) that $H'(z)$ is also nonsingular.
(iii) The assumption that $F'(p(z))$ is a P-matrix and the fact that $c_i(x) \in [0, 1]$, $i \in N$ ensure that $F'(p(z))C(x) + I - C(x)$ is nonsingular, e.g., see [7, Lemma 2]. So, $H'(z)$ is nonsingular.

□

Since the smoothing functions defined by (16), (18) and (20) all satisfy Assumption 1, from Theorem 1 we have the following result.

Theorem 2. *Suppose that the smoothing function $\phi(\mu, c, d, w)$, $(\mu, w) \in \Re_{++} \times \Re$ is defined by either (16) or (18) or (20). Then*

(i) The mapping H is continuously differentiable at any $z = (u, x) \in \Re^n_{++} \times \Re^n$ and if $F'(p(z))$ is a P-matrix, then $H'(z)$ is nonsingular.

(ii) Suppose that $\phi(\mu, c, d, w)$, $(\mu, w) \in \Re_{++} \times \Re$ is defined by either (16) or (18) (in each case $\mathrm{supp}(\rho) = \Re$) and for each $i \in N$, at least one of a_i and b_i is finite. Then $H'(z)$ is nonsingular if $F'(p(z))$ is a P_0-matrix at $z = (u, x) \in \Re^n_{++} \times \Re^n$.

3. Semismoothness properties

In order to design high-order convergent Newton methods we need the concept of semismoothness. Semismoothness was originally introduced by Mifflin [37] for functionals. Convex functions, smooth functions, and piecewise linear functions are examples of semismooth functions. The composition of semismooth functions is still a semismooth function [37]. In [48], Qi and Sun extended the definition of semismooth functions to $\Phi : \Re^{m_1} \to \Re^{m_2}$. A locally Lipschitz continuous vector valued function $\Phi : \Re^{m_1} \to \Re^{m_2}$ has a generalized Jacobian $\partial\Phi(x)$ as in Clarke [13]. Φ is said to be *semismooth* at $x \in \Re^{m_1}$, if

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

exists for any $h \in \Re^{m_1}$. It has been proved in [48] that Φ is semismooth at x if and only if all its component functions are. Also, $\Phi'(x; h)$, the directional derivative of Φ at x in the direction h, exists for any $h \in \Re^{m_1}$ if Φ is semismooth at x.

Lemma 3. *[48] Suppose that $\Phi : \Re^{m_1} \to \Re^{m_2}$ is a locally Lipschitzian function and semismooth at x. Then*

(i) *for any $V \in \partial\Phi(x+h)$, $h \to 0$,*

$$Vh - \Phi'(x; h) = o(\|h\|);$$

(ii) *for any $h \to 0$,*

$$\Phi(x+h) - \Phi(x) - \Phi'(x; h) = o(\|h\|).$$

The following lemma is extracted from Theorem 2.3 of [48].

Lemma 4. *Suppose that $\Phi : \Re^{m_1} \to \Re^{m_2}$ is a locally Lipschitzian function. Then the following two statements are equivalent:*

(i) *$\Phi(\cdot)$ is semismooth at x.*
(ii) *For any $V \in \partial\Phi(x+h)$, $h \to 0$,*

$$Vh - \Phi'(x; h) = o(\|h\|).$$

A stronger notion than semismoothness is strong semismoothness. $\Phi(\cdot)$ is said to be *strongly semismooth* at x if Φ is semismooth at x and for any $V \in \partial\Phi(x+h)$, $h \to 0$,

$$Vh - \Phi'(x; h) = O(\|h\|^2).$$

(Note that in [48] and [45] different names for strong semismoothness are used.) A function Φ is said to be a (strongly) semismooth function if it is (strongly) semismooth everywhere.

Recall that from Lemma 2 the function $q_{cd}(\cdot)$ defined by (15) is globally Lipschitz continuous on $\Re^2$. Then, from Lemma 4 and the definition of strong semismoothness, we can prove that $q_{cd}(\cdot)$ is strongly semismooth at $x \in \Re^2$ by verifying that for any $V \in \partial q_{cd}(x+h)$, $h \to 0$,

$$Vh - q'_{cd}(x; h) = O(\|h\|^2). \tag{24}$$

The following three propositions are about the strong semismoothness of q_{cd} resulted from Examples 1–3, respectively. Their proofs can be found in Appendix A.

Proposition 1. *Suppose that the smoothing function $\phi(\cdot)$ is the neural networks smoothing function defined by (16). Then the corresponding function $q_{cd} : \Re^2 \to \Re$ defined by (15) is a strongly semismooth function.*

Proposition 2. *Suppose that the smoothing function $\phi(\cdot)$ is the Chen-Harker-Kanzow-Smale smoothing function defined by (18). Then the corresponding function $q_{cd} : \Re^2 \to \Re$ defined by (15) is a strongly semismooth function.*

Proposition 3. *Suppose that the smoothing function $\phi(\cdot)$ is the uniform smoothing function defined by (20). Then the corresponding function $q_{cd} : \Re^2 \to \Re$ defined by (15) is a strongly semismooth function.*

Theorem 3. *Suppose that the smoothing function $\phi(\mu, c, d, w)$, $(\mu, w) \in \Re_{++} \times \Re$ is defined by either (16) or (18) or (20). Then*

(i) H is semismooth at any $z \in \Re^{2n}$, and
(ii) if for some point $z \in \Re^{2n}$, F' is Lipschitz continuous around $p(z) \in \Re^n$, then H is strongly semismooth at z.

Proof. (i) Since p is strongly semismooth at z if and only if its component functions p_i, $i \in N$ are, and the composition of strongly semismooth functions is a strongly semismooth function [21, Theorem 19], from Propositions 1–3, it follows that p is a strongly semismooth function. Hence, by making use of the proposition that the composition of two semismooth functions is semismooth [37] that for each $i = 1, 2, ..., n$, G_i is a semismooth function. Hence, G, and so H, is a semismooth function.
(ii) It is noted that F' is Lipschitz continuous around $p(z) \in \Re^n$ implies that F is strongly semismooth at $p(z)$. Hence, by [21, Theorem 19], $F(p(\cdot))$ is strongly semismooth at z because p is strongly semismooth at z too. Then G, and so H, is strongly semismooth at z.

□

4. Smoothing Newton methods

Throughout the rest of this paper, unless otherwise stated, we assume that the smoothing function $\phi(\cdot)$ satisfies Assumption 1.

Choose $\bar{u} \in \Re^n_{++}$ and $\gamma \in (0, 1)$ such that $\gamma\|\bar{u}\| < 1$. Let $\bar{z} := (\bar{u}, 0) \in \Re^n \times \Re^n$. Define the merit function $\psi : \Re^{2n} \to \Re_+$ by

$$\psi(z) := \|H(z)\|^2$$

and define $\beta : \Re^{2n} \to \Re_+$ by

$$\beta(z) := \gamma \min\{1, \psi(z)\}.$$

Let

$$\Omega := \{z = (u, x) \in \Re^n \times \Re^n \,|\, u \geq \beta(z)\bar{u}\}.$$

Then, because for any $z \in \Re^{2n}$, $\beta(z) \leq \gamma < 1$, it follows that for any $x \in \Re^n$,

$$(\bar{u}, x) \in \Omega.$$

Proposition 4. *The following relations hold:*

$$H(z) = 0 \Longleftrightarrow \beta(z) = 0 \Longleftrightarrow H(z) = \beta(z)\bar{z}.$$

Proof. It follows from the definitions of $H(\cdot)$ and $\beta(\cdot)$ that

$$H(z) = 0 \Longleftrightarrow \beta(z) = 0 \quad \text{and} \quad \beta(z) = 0 \Longrightarrow H(z) = \beta(z)\bar{z}.$$

Then we only need to prove

$$H(z) = \beta(z)\bar{z} \Longrightarrow \beta(z) = 0.$$

However, this is an easy task because from $H(z) = \beta(z)\bar{z}$ we have

$$u = \beta(z)\bar{u} \quad \text{and} \quad G(z) = 0.$$

Hence, from the definitions of $\psi(\cdot)$ and $\beta(\cdot)$, and the fact that $\gamma\|\bar{u}\| < 1$, we get

$$\psi(z) = \|u\|^2 + \|G(z)\|^2 = \|u\|^2 = \beta(z)^2\|\bar{u}\|^2 \leq \gamma^2\|\bar{u}\|^2 < 1.$$

Therefore,

$$\beta(z) = \gamma\psi(z) = \gamma\beta(z)^2\|\bar{u}\|^2. \tag{25}$$

If $\beta(z) \neq 0$, it follows from (25) and the fact $\beta(z) \leq \gamma$ that

$$1 = \gamma\beta(z)\|\bar{u}\|^2 \leq \gamma^2\|\bar{u}\|^2,$$

which contradicts the fact that $\gamma\|\bar{u}\| < 1$. This contradiction completes our proof.

□

Algorithm 1.

Step 0. Choose constants $\delta \in (0, 1)$ and $\sigma \in (0, 1/2)$. Let $u^0 := \bar{u}$, $x^0 \in \Re^n$ be an arbitrary point and $k := 0$.

Step 1. If $H(z^k) = 0$ then stop. Otherwise, let $\beta_k := \beta(z^k)$.

Step 2. Compute $\Delta z^k := (\Delta u^k, \Delta x^k) \in \Re^n \times \Re^n$ by

$$H(z^k) + H'(z^k)\Delta z^k = \beta_k\bar{z}. \tag{26}$$

Step 3. Let l_k be the smallest nonnegative integer l satisfying

$$\psi(z^k + \delta^l\Delta z^k) \leq [1 - 2\sigma(1 - \gamma\|\bar{u}\|)\delta^l]\psi(z^k). \tag{27}$$

Define $z^{k+1} := z^k + \delta^{l_k}\Delta z^k$.

Step 4. Replace k by $k + 1$ and go to Step 1.

Remark 1. (i) Since we have assumed that Assumption 1 is satisfied for the smoothing function ϕ used in the algorithm, $H(\cdot)$ is continuously differentiable at any $z^k \in \Re^n_{++} \times \Re^n$.

(ii) From Theorem 1, for $z^k \in \Re^n_{++} \times \Re^n$ if $F'(p(z^k))$ is a P-matrix, then $H'(z^k)$ is nonsingular, and if $\text{supp}(\rho) = \Re$ and for each $i \in N$, at least one of a_i and b_i is finite, then the condition that $F'(p(z^k))$ is a P_0-matrix is sufficient to guarantee that $H'(z^k)$ is nonsingular.

(iii) We can solve equation (26) in the following way: Let $\Delta u^k = -u^k + \beta_k \bar{u}$. Solve

$$\left[F'(p(z^k))C(x^k) + I - C(x^k)\right]\Delta x^k = -G(z^k) - \left[(F'(p(z^k)) - I)D(u^k)\right]\Delta u^k \tag{28}$$

to get Δx^k. Then $\Delta z^k = (\Delta u^k, \Delta x^k)$. Equation (28) is an n-dimensional linear system. If we choose the uniform smoothing function $\phi(\cdot)$ defined by (20), then solving (26) can be further simplified because in this case some or all of the diagonal entries of the diagonal matrix $C(x^k)$ are zeros. For example, for NCP, if $w^k \le -u^k/2$, then from Lemma 1, $c_i(x^k) = 0$, $i \in N$, and so, $C(x^k) = 0$. Hence, in this case

$$\Delta x^k = -G(z^k) - \left[(F'(p(z^k)) - I)D(u^k)\right]\Delta u^k.$$

Solving a form reduced linear equation is very favourable. However, this is not without a price because by choosing the smoothing function $\phi(\cdot)$ defined by (20) we need stronger conditions to ensure the nonsingularity of $H'(z^k)$ than by choosing the smoothing function $\phi(\cdot)$ defined by either (16) or (18).

(iv) From the design of our algorithm we can see that the paremeter u may not change until $\psi(z) < 1$. To make the parameter u change at each step we can let the steplength $\delta^{l_k} < 1$ in Step 3 for all k such that $\psi(z^k) > 1$ though this is not recommended in practiacl computation.

Lemma 5. *Suppose that Assumption 1 holds. For any* $\tilde{z} := (\tilde{u}, \tilde{x}) \in \Re^n_{++} \times \Re^n$ *and* $H'(\tilde{z})$ *is nonsingular, then there exist a closed neighbourhood* $\mathcal{N}(\tilde{z})$ *of* $\tilde{z}$ *and a positive number* $\bar{\alpha} \in (0, 1]$ *such that for any* $z = (u, x) \in \mathcal{N}(\tilde{z})$ *and all* $\alpha \in [0, \bar{\alpha}]$ *we have* $u \in \Re^n_{++}$, $H'(z)$ *is invertible and*

$$\psi(z + \alpha\Delta z) \le [1 - 2\sigma(1 - \gamma\|\bar{u}\|)\alpha]\psi(z). \tag{29}$$

Proof. Since $H'(\tilde{z})$ is invertible and $\tilde{u} \in \Re^n_{++}$, there exists a closed neighbourhood $\mathcal{N}(\tilde{z})$ of $\tilde{z}$ such that for any $z = (u, x) \in \mathcal{N}(\tilde{z})$ we have $u \in \Re^n_{++}$ and that $H'(z)$ is invertible. For any $z \in \mathcal{N}(\tilde{z})$, let $\Delta z = (\Delta u, \Delta x) \in \Re^n \times \Re^n$ be the unique solution of the following equation:

$$H(z) + H'(z)\Delta z = \beta(z)\bar{z} \tag{30}$$

and for any $\alpha \in [0, 1]$, define

$$g_z(\alpha) = G(z + \alpha\Delta z) - G(z) - \alpha G'(z)\Delta z.$$

From (30), for any $z \in \mathcal{N}(\tilde{z})$,

$$\Delta u = -u + \beta(z)\bar{u}.$$

Then for all $\alpha \in [0, 1]$ and all $z \in \mathcal{N}(\tilde{z})$,

$$u + \alpha\Delta u = (1 - \alpha)u + \alpha\beta(z)\bar{u} \in \Re^n_{++}. \tag{31}$$

It follows from the Mean Value Theorem that

$$g_z(\alpha) = \alpha \int_0^1 [G'(z + \theta\alpha\Delta z) - G'(z)]\Delta z d\theta.$$

Since $G'(\cdot)$ is uniformly continuous on $\mathcal{N}(\tilde{z})$ and $\Delta z \to \Delta\tilde{z}$ as $z \to \tilde{z}$, for all $z \in \mathcal{N}(\tilde{z})$,

$$\lim_{\alpha\downarrow 0} \|g_z(\alpha)\|/\alpha = 0.$$

Then, from (31), (30) and the fact that $\beta(z) \leq \gamma\psi(z)^{1/2}$, for all $\alpha \in [0, 1]$ and all $z \in \mathcal{N}(\tilde{z})$, we have

$$\begin{aligned}
&\|u + \alpha\Delta u\|^2 \\
&\quad = \|(1-\alpha)u + \alpha\beta(z)\bar{u}\|^2 \\
&\quad = (1-\alpha)^2\|u\|^2 + 2(1-\alpha)\alpha\beta(z)u^T\bar{u} + \alpha^2\beta(z)^2\|\bar{u}\|^2 \\
&\quad \leq (1-\alpha)^2\|u\|^2 + 2(1-\alpha)\alpha\beta(z)\|u\|\|\bar{u}\| + \alpha^2\beta(z)^2\|\bar{u}\|^2 \\
&\quad \leq (1-\alpha)^2\|u\|^2 + 2\alpha\beta(z)\|u\|\|\bar{u}\| + O(\alpha^2) \\
&\quad \leq (1-\alpha)^2\|u\|^2 + 2\alpha\gamma\psi(z)^{1/2}\|H(z)\|\|\bar{u}\| + O(\alpha^2) \\
&\quad = (1-\alpha)^2\|u\|^2 + 2\alpha\gamma\|\bar{u}\|\psi(z) + O(\alpha^2)
\end{aligned} \tag{32}$$

and

$$\begin{aligned}
&\|G(z + \alpha\Delta z)\|^2 \\
&\quad = \|G(z) + \alpha G'(z)\Delta z + g_z(\alpha)\|^2 \\
&\quad = \|(1-\alpha)G(z) + g_z(\alpha)\|^2 \\
&\quad = (1-\alpha)^2\|G(z)\|^2 + 2(1-\alpha)G(z)^T g_z(\alpha) + \|g_z(\alpha)\|^2 \\
&\quad = (1-\alpha)^2\|G(z)\|^2 + o(\alpha).
\end{aligned} \tag{33}$$

It then follows from (32) and (33) that for all $\alpha \in [0, 1]$ and all $z \in \mathcal{N}(\tilde{z})$, we have

$$\begin{aligned}
&\psi(z + \alpha\Delta z) \\
&\quad = \|H(z + \alpha\Delta z)\|^2 \\
&\quad = \|u + \alpha\Delta u\|^2 + \|G(z + \alpha\Delta z)\|^2 \\
&\quad \leq (1-\alpha)^2\|u\|^2 + 2\alpha\gamma\|\bar{u}\|\psi(z) + (1-\alpha)^2\|G(z)\|^2 + o(\alpha) + O(\alpha^2) \\
&\quad = (1-\alpha)^2\psi(z) + 2\alpha\gamma\|\bar{u}\|\psi(z) + o(\alpha) \\
&\quad = (1-2\alpha)\psi(z) + 2\alpha\gamma\|\bar{u}\|\psi(z) + o(\alpha) \\
&\quad = [1 - 2(1-\gamma\|\bar{u}\|)\alpha]\psi(z) + o(\alpha).
\end{aligned} \tag{34}$$

Then from inequality (34) we can find a positive number $\bar{\alpha} \in (0, 1]$ such that for all $\alpha \in [0, \bar{\alpha}]$ and all $z \in \mathcal{N}(\tilde{z})$, (29) holds.

□

We can get the following result directly from Lemma 5.

Proposition 5. *Suppose that Assumption 1 holds. For any $k \geq 0$, if $z^k \in \Re^n_{++} \times \Re^n$ and $H'(z^k)$ is nonsingular, then Algorithm 1 is well defined at kth iteration and $z^{k+1} \in \Re^n_{++} \times \Re^n$.*

Proposition 6. *Suppose that Assumption 1 holds. For each fixed $k \geq 0$, if $u^k \in \Re^n_{++}$, $z^k \in \Omega$ and $H'(z^k)$ is nonsingular, then for any $\alpha \in [0, 1]$ such that*

$$\psi(z^k + \alpha\Delta z^k) \leq [1 - 2\sigma(1 - \gamma\|\bar{u}\|)\alpha]\psi(z^k), \tag{35}$$

it holds that $z^k + \alpha\Delta z^k \in \Omega$.

Proof. We prove this proposition by considering the following two cases:
(i) If $\psi(z^k) > 1$. Then, $\beta_k = \gamma$. It therefore follows from $z^k \in \Omega$ and $\beta(z) = \gamma \min\{1, \psi(z)\} \leq \gamma$ for any $z \in \Re^{2n}$ that for all $\alpha \in [0, 1]$, we have

$$\begin{aligned}
&u^k + \alpha\Delta u^k - \beta(z^k + \alpha\Delta z^k)\bar{u} \\
&\quad\geq (1 - \alpha)u^k + \alpha\beta_k\bar{u} - \gamma\bar{u} \\
&\quad\geq (1 - \alpha)\beta_k\bar{u} + \alpha\beta_k\bar{u} - \gamma\bar{u} \\
&\quad= (1 - \alpha)\gamma\bar{u} + \alpha\gamma\bar{u} - \gamma\bar{u} \\
&\quad= 0.
\end{aligned} \tag{36}$$

(ii) If $\psi(z^k) \leq 1$. Then, for any $\alpha \in [0, 1]$ satisfying (35), we have

$$\psi(z^k + \alpha\Delta z^k) \leq [1 - 2\sigma(1 - \gamma\|\bar{u}\|)\alpha]\psi(z^k) \leq 1. \tag{37}$$

So, for any $\alpha \in [0, 1]$ satisfying (35),

$$\beta(z^k + \alpha\Delta z^k) = \gamma\psi(z^k + \alpha\Delta z^k).$$

Hence, again because $z^k \in \Omega$, by using the first inequality in (37), for any $\alpha \in [0, 1]$ satisfying (35) we have

$$\begin{aligned}
&u^k + \alpha\Delta u^k - \beta(z^k + \alpha\Delta z^k)\bar{u} \\
&\quad= (1 - \alpha)u^k + \alpha\beta_k\bar{u} - \gamma\psi(z^k + \alpha\Delta z^k)\bar{u} \\
&\quad\geq (1 - \alpha)\beta_k\bar{u} + \alpha\beta_k\bar{u} - \gamma[1 - 2\sigma(1 - \gamma\|\bar{u}\|)\alpha]\psi(z^k)\bar{u} \\
&\quad= \beta_k\bar{u} - \gamma[1 - 2\sigma(1 - \gamma\|\bar{u}\|)\alpha]\psi(z^k)\bar{u} \\
&\quad= \gamma\psi(z^k)\bar{u} - \gamma[1 - 2\sigma(1 - \gamma\|\bar{u}\|)\alpha]\psi(z^k)\bar{u} \\
&\quad= [2\gamma\sigma(1 - \gamma\|\bar{u}\|)]\alpha\psi(z^k)\bar{u} \\
&\quad\geq 0.
\end{aligned} \tag{38}$$

Thus, by combining (36) and (38), we have proved that for all $\alpha \in [0, 1]$ satisfying (35),

$$z^k + \alpha\Delta z^k \in \Omega.$$

This completes our proof.

□

By combining Propositions 5 and 6, we have

Proposition 7. *Suppose that Assumption 1 holds. For each fixed $k \geq 0$, if $u^k \in \Re^n_{++}$, $z^k \in \Omega$ and $H'(z^k)$ is invertible, then*

$$u^{k+1} \in \Re^n_{++} \quad \text{and} \quad z^{k+1} \in \Omega.$$

Proposition 8. *Suppose that Assumption 1 holds and that for every $k \geq 0$ with $u^k \in \Re^n_{++}$ and $z^k \in \Omega$ we have that $H'(z^k)$ is invertible. Then an infinite sequence $\{z^k\}$ is generated by Algorithm 1, $u^k \in \Re^n_{++}$ and $\{z^k\} \in \Omega$.*

Proof. First, because $z^0 = (\bar{u}, x^0) \in \Omega$, we have from Proposition 7 that z^1 is well defined, $u^1 \in \Re^n_{++}$ and $z^1 \in \Omega$. Then, by repeatedly resorting to Proposition 7 we can prove that an infinite sequence $\{z^k\}$ is generated, $u^k \in \Re^n_{++}$ and $z^k \in \Omega$.

□

5. Global convergence

In order to discuss the global convergence of Algorithm 1 we need the following assumption.

Assumption 2. *(i) For every $k \geq 0$, if $u^k \in \Re^n_{++}$ and $z^k \in \Omega$, then $H'(z^k)$ is nonsingular; and*
(ii) for any accumulation point $z^* = (u^*, x^*)$ of $\{z^k\}$ if $u^* \in \Re^n_{++}$ and $z^* \in \Omega$, then $H'(z^*)$ is nonsingular.

Theorem 4. *Suppose that Assumptions 1 and 2 are satisfied. Then an infinite sequence $\{z^k\}$ is generated by Algorithm 1 and each accumulation point $\tilde{z}$ of $\{z^k\}$ is a solution of $H(z) = 0$.*

Proof. It follows from Proposition 8 and Assumption 2 that an infinite sequence $\{z^k\}$ is generated such that $\{z^k\} \in \Omega$. From the design of Algorithm 1, $\psi(z^{k+1}) < \psi(z^k)$ for all $k \geq 0$. Hence the two sequences $\{\psi(z^k)\}$ and $\{\beta(z^k)\}$ are monotonically decreasing. Since $\psi(z^k), \beta(z^k) \geq 0$ $(k \geq 0)$, there exist $\tilde{\psi}, \tilde{\beta} \geq 0$ such that $\psi(z^k) \to \tilde{\psi}$ and $\beta(z^k) \to \tilde{\beta}$ as $k \to \infty$. If $\tilde{\psi} = 0$ and $\{z^k\}$ has an accumulation point $\tilde{z}$, then from the continuity of $\psi(\cdot)$ and $\beta(\cdot)$ we obtain $\psi(\tilde{z}) = 0$ and $\beta(\tilde{z}) = 0$. Then we obtain the desired result. Suppose that $\tilde{\psi} > 0$ and $\tilde{z} = (\tilde{u}, \tilde{x}) \in \Re^n \times \Re^n$ is an accumulation point of $\{z^k\}$. By taking a subsequence if necessary, we may assume that $\{z^k\}$ converges to $\tilde{z}$. It is easy to see that $\tilde{\psi} = \psi(\tilde{z})$, $\beta(\tilde{z}) = \tilde{\beta}$ and $\tilde{z} \in \Omega$. Thus, from $\beta(\tilde{z}) = \gamma \min\{1, \psi(\tilde{z})\} > 0$ and $\tilde{z} \in \Omega$, we see that $\tilde{u} \in \Re^n_{++}$. Then, from (ii) of Assumption 2, $H'(\tilde{z})$ exists and is invertible. Hence, from Lemma 5 there exist a closed neighbourhood $\mathcal{N}(\tilde{z})$ of $\tilde{z}$ and a positive number $\bar{\alpha} \in (0, 1]$ such that for any $z = (u, x) \in \mathcal{N}(\tilde{z})$ and all $\alpha \in [0, \bar{\alpha}]$ we have $u \in \Re^n_{++}$, $H'(z)$ is invertible and (29) holds. Therefore, for a nonnegative integer l such that $\delta^l \in (0, \bar{\alpha}]$, we have

$$\psi(z^k + \delta^l \Delta z^k) \leq \left[1 - 2\sigma(1 - \gamma\|\bar{u}\|)\delta^l\right]\psi(z)$$

for all sufficiently large k. Then, for every sufficiently large k, we see that $l^k \leq l$ and hence $\delta^{l_k} \geq \delta^l$. Then

$$\psi(z^{k+1}) \leq [1 - 2\sigma(1 - \gamma\|\bar{u}\|)\delta^{l_k}]\psi(z^k) \leq [1 - 2\sigma(1 - \gamma\|\bar{u}\|)\delta^l]\psi(z^k)$$

for all sufficiently large k. This contradicts the fact that the sequence $\{\psi(z^k)\}$ converges to $\tilde{\psi} > 0$. So, we complete our proof.

□

Theorem 5. *Suppose that the smoothing function $\phi(\cdot)$ is defined by either (16) or (18) or (20). Then an infinite sequence $\{z^k\}$ is generated by Algorithm 1 and each accumulation point $\tilde{z}$ of $\{z^k\}$ is a solution of $H(z) = 0$, under one of the following two conditions,*

(i) for each $z = (u, x) \in \Omega$ with $u \in \Re^n_{++}$, $F'(p(z))$ is a P-matrix;
(ii) if $\phi(\cdot)$ is defined by either (16) or (18), for each $i \in N$, at least one of a_i and b_i is finite, and for each $z = (u, x) \in \Omega$ with $u \in \Re^n_{++}$, $F'(p(z))$ is a P_0-matrix.

Proof. By using Theorems 2 and 4 directly, we get the results of this theorem.

□

6. Bounded level sets

In Sect. 5 we proved, under the assumptions of Theorem 4, that any accumulation point of $\{z^k\}$ generated by Algorithm 1, if it exists, is a solution of $H(z) = 0$. An important question remained unanswered is whether such an accumulation point exists or not. In this section we answer this question by investigating under what conditions the level sets of $\psi(z) = \|H(z)\|^2$ are bounded. For this, let

$$L(z^0) = \left\{z \in \Re^{2n} |\ \psi(z) \leq \psi(z^0)\right\}.$$

Theorem 6. *If X is bounded, then $L(z^0)$ is bounded.*

Proof. Since X is bounded, it follows from (8) that $\|p(z)\|$ is bounded for any $z \in \Re^{2n}$. For the sake of contradiction, suppose that there exists a sequence $\{z^k = (u^k, x^k) \in \Re^n \times \Re^n\}$ such that $z^k \in L(z^0)$ and $\|z^k\| \to \infty$. Apparently, since $z^k \in L(z^0)$, $\|u^k\| \leq \|H(z^k)\| \leq \|H(z^0)\|$. So, $\|x^k\| \to \infty$. Hence, by using the fact that $\|p(u^k, x^k)\|$ is bounded, we have

$$\left\|G(u^k, x^k)\right\| = \left\|F(p(u^k, x^k)) + x^k - p(u^k, x^k)\right\| \to \infty,$$

which contradicts that $z^k \in L(z^0)$ because $\|H(z^k)\| \geq \|G(z^k)\|$.

□

Theorem 7. *Suppose that F is a uniform P-function on X, i.e., there exists a positive number $\nu > 0$ such that*

$$\max_{i \in N} \left(y_i^1 - y_i^2\right)\left(F_i(y^1) - F_i(y^2)\right) \geq \nu\left\|y^1 - y^2\right\|^2 \quad \forall\ y^1, y^2 \in X. \tag{39}$$

Then $L(z^0)$ is bounded.

Proof. For the sake of contradiction, suppose that there exists a sequence $\{z^k = (u^k, x^k) \in \Re^n \times \Re^n\}$ such that $z^k \in L(z^0)$ and $\|z^k\| \to \infty$. Since, apparently, $\|u^k\|$ is bounded, $\|x^k\| \to \infty$. It is easy to prove that

$$\left|\mathrm{mid}(a_i, b_i, x_i^k)\right| \to \infty \Longrightarrow \left|x_i^k\right| \to \infty \text{ and } \left|x_i^k - \mathrm{mid}(a_i, b_i, x_i^k)\right| \to 0, \quad i \in N. \tag{40}$$

From Lemma 2 and the definition of $p(\cdot)$, there exists a constant $L' > 0$ such that

$$\left|p_i(u^k, x^k) - \mathrm{mid}(a_i, b_i, x_i^k)\right| \le L'\left|u_i^k\right|, \quad i \in N. \tag{41}$$

Define the index set J by $J := \{i | \{p_i(u^k, x^k)\}$ is unbounded, $i \in N\}$. Then it follows that $J \neq \emptyset$ because otherwise $\|G(z^k)\| = \|F(p(z^k)) + x^k - p(z^k)\| \to \infty$. Let $\bar{z}^k = (\bar{u}^k, \bar{x}^k) \in \Re^n \times \Re^n$ be defined by

$$\bar{u}_i^k = \begin{cases} u_i^k & \text{if } i \notin J \\ 0 & \text{if } i \in J \end{cases}$$

and

$$\bar{x}_i^k = \begin{cases} x_i^k & \text{if } i \notin J \\ 0 & \text{if } i \in J \end{cases}, \quad i \in N.$$

Then

$$p_i(\bar{z}^k) = \begin{cases} p_i(z^k) & \text{if } i \notin J \\ \mathrm{mid}(a_i, b_i, 0) & \text{if } i \in J \end{cases}, \quad i \in N.$$

Hence $\{\|p(\bar{z}^k)\|\}$ is bounded. Therefore, from (39), we have

$$\begin{aligned} \nu \sum_{i \in J} (p_i(z^k) - p_i(\bar{z}^k))^2 &= \nu \|p(z^k) - p(\bar{z}^k)\|^2 \\ &\le \max_{i \in N} (p_i(z^k) - p_i(\bar{z}^k))(F_i(p(z^k)) - F_i(p(\bar{z}^k))) \\ &\le \max_{i \in N} |p_i(z^k) - p_i(\bar{z}^k)||F_i(p(z^k)) - F_i(p(\bar{z}^k))| \\ &= \max_{i \in J} |p_i(z^k) - p_i(\bar{z}^k)||F_i(p(z^k)) - F_i(p(\bar{z}^k))| \\ &\le \sqrt{\sum_{i \in J} (p_i(z^k) - p_i(\bar{z}^k))^2} \max_{i \in J} |F_i(p(z^k)) - F_i(p(\bar{z}^k))|. \end{aligned}$$

Then $\max_{i \in J} |F_i(p(z^k)) - F_i(p(\bar{z}^k))| \to \infty$ as $k \to \infty$. Since $\{\|F(p(\bar{z}^k))\|\}$ is bounded, for each k there exists at least one $i_k \in J$ such that

$$\left|F_{i_k}(p(z^k))\right| \to \infty.$$

Since J has only a finite number of elements, by taking a subsequence if necessary, we may assume that there exists an $i \in J$ such that

$$\left|F_i(p(z^k))\right| \to \infty.$$

Then, in view of (41), the definition of J and the boundedness of $\{\|u^k\|\}$, we have proved that there exists at least one $i \in J$ such that

$$\left|F_i(p(z^k))\right|, \left|p_i(z^k)\right|, \left|\mathrm{mid}(a_i, b_i, x_i^k)\right| \to \infty.$$

Hence, by (40), (41) and that $\{\|u^k\|\}$ is bounded, for such $i \in J$, $|x_i^k - p_i(z^k)|$ is bounded. It then follows that for such $i \in J$, $\{|G_i(z^k)|\}$ is unbounded. This is a contradiction because $\|H(z^k)\| \geq \|G(z^k)\|$. This contradiction shows $L(z^0)$ is bounded.

□

There are several papers in the literature dealing with the bounded level sets issue for different merit functions by assuming that F is a uniform P-function on $\Re^n$, i.e., (39) holds for all $y^1, y^2 \in \Re^n$, see, e.g., [16,28,14,11,32,55]. Here we only require (39) to hold on X.

7. Superlinear and quadratic convergence

Theorem 8. *Suppose that Assumptions 1 and 2 are satisfied and z^* is an accumulation point of the infinite sequence $\{z^k\}$ generated by Algorithm 1. Suppose that H is semismooth at z^* and that all $V \in \partial H(z^*)$ are nonsingular. Then the whole sequence $\{z^k\}$ converges to z^*,*

$$\left\|z^{k+1} - z^*\right\| = o(\|z^k - z^*\|) \tag{42}$$

and

$$u_i^{k+1} = o\big(u_i^k\big), \quad i \in N. \tag{43}$$

Furthermore, if H is strongly semismooth at z^*, then

$$\left\|z^{k+1} - z^*\right\| = O\big(\|z^k - z^*\|^2\big) \tag{44}$$

and

$$u_i^{k+1} = O\big(u_i^k\big)^2, \quad i \in N. \tag{45}$$

Proof. First, from Theorem 4 that z^* is a solution of $H(z) = 0$. Then, from [48, Proposition 3.1], for all z^k sufficiently close to z^*,

$$\left\|H'(z^k)^{-1}\right\| = O(1).$$

Hence, under the assumption that H is semismooth (strongly semismooth, respectively) at z^*, from Lemma 3, for z^k sufficiently close to z^*, we have

$$\begin{aligned}
&\left\|z^k + \Delta z^k - z^*\right\| \\
&\quad = \left\|z^k + H'(z^k)^{-1}\big[-H(z^k) + \beta_k \bar{z}\big] - z^*\right\| \\
&\quad = O\big(\|H(z^k) - H(z^*) - H'(z^k)(z^k - z^*)\| + \beta_k \|\bar{u}\|\big) \\
&\quad = o(\|z^k - z^*\|) + O(\psi(z^k)) \quad \big(= O\big(\|z^k - z^*\|^2\big) + O(\psi(z^k))\big).
\end{aligned} \tag{46}$$

Then, because H is semismooth at z^*, H is locally Lipschitz continuous near z^*, for all z^k close to z^*,

$$\psi(z^k) = \big\|H(z^k)\big\|^2 = O\big(\|z^k - z^*\|^2\big). \tag{47}$$

Therefore, from (46) and (47), if H is semismooth (strongly semismooth, respectively) at z^*, for all z^k sufficiently close to z^*,

$$\|z^k + \Delta z^k - z^*\| = o\big(\|z^k - z^*\|\big) \quad \big(= O\big(\|z^k - z^*\|^2\big)\big). \tag{48}$$

By following the proof of Theorem 3.1 of [45], for all z^k sufficiently close to z^*, we have

$$\|z^k - z^*\| = O\big(\|H(z^k) - H(z^*)\|\big). \tag{49}$$

Hence, if H is semismooth (strongly semismooth, respectively) at z^*, for all z^k sufficiently close to z^*, we have

$$\begin{aligned}
&\psi(z^k + \Delta z^k)\\
&\quad = \|H(z^k + \Delta z^k)\|^2\\
&\quad = O\big(\|z^k + \Delta z^k - z^*\|^2\big)\\
&\quad = o(\|z^k - z^*\|^2) \quad \big(= O\big(\|z^k - z^*\|^4\big)\big)\\
&\quad = o\big(\|H(z^k) - H(z^*)\|^2\big) \quad \big(= O\big(\|H(z^k) - H(z^*)\|^4\big)\big)\\
&\quad = o(\psi(z^k)) \quad \big(= O\big(\psi(z^k)^2\big)\big).
\end{aligned} \tag{50}$$

Therefore, for all z^k sufficiently close to z^* we have

$$z^{k+1} = z^k + \Delta z^k,$$

which, together with (48), proves (42), and if H is strongly semismooth at z^*, proves (44).

Next, from the definition of β_k and the fact that $z^k \to z^*$ as $k \to \infty$, for all k sufficiently large,

$$\beta_k = \gamma\psi(z^k) = \gamma\|H(z^k)\|^2.$$

Also, because for all k sufficiently large, $z^{k+1} = z^k + \Delta z^k$, we have for all k sufficiently large that

$$u^{k+1} = u^k + \Delta u^k = \beta_k \bar{u}.$$

Hence, for all k sufficiently large,

$$u^{k+1} = \gamma\|H(z^k)\|^2\bar{u},$$

which, together with (42), (47) and (49), gives

$$\lim_{k\to\infty} \frac{u_i^{k+1}}{u_i^k} = \lim_{k\to\infty} \frac{\|H(z^k)\|^2}{\|H(z^{k-1})\|^2} = \lim_{k\to\infty} \frac{\|H(z^k) - H(z^*)\|^2}{\|H(z^{k-1}) - H(z^*)\|^2} = 0, \quad i \in N.$$

This proves (43). If H is strongly semismooth at z^*, then from the above argument we can easily get (45). So, we complete our proof.

□

Next, we study under what conditions all the matrices $V \in \partial H(z^*)$ are nonsingular at a solution point $z^* = (u^*, x^*) \in \Re^n \times \Re^n$ of $H(z) = 0$. Apparently, $u^* = 0$ and x^* is a solution of $E(x) = 0$, where E is defined in (5). For convenience of handling notation we denote

$$\mathcal{I} := \left\{ i \mid a_i < x_i^* < b_i \ \& \ F_i(\Pi_X(x^*)) = 0,\ i \in N \right\},$$

$$\mathcal{J} := \left\{ i \mid x_i^* = a_i \ \& \ F_i(\Pi_X(x^*)) = 0,\ i \in N \right\} \cup \left\{ i \mid x_i^* = b_i \ \& \ F_i(\Pi_X(x^*)) = 0,\ i \in N \right\}$$

and

$$\mathcal{K} := \left\{ i \mid x_i^* < a_i \ \& \ F_i(\Pi_X(x^*)) > 0,\ i \in N \right\} \cup \left\{ i \mid x_i^* > b_i \ \& \ F_i(\Pi_X(x^*)) < 0,\ i \in N \right\}.$$

Then

$$\mathcal{I} \cup \mathcal{J} \cup \mathcal{K} = N.$$

By rearrangement we assume that $\nabla F(\Pi_X(x^*))$ can be rewritten as

$$\nabla F(\Pi_X(x^*)) = \begin{pmatrix} \nabla F(\Pi_X(x^*))_{\mathcal{I}\mathcal{I}} & \nabla F(\Pi_X(x^*))_{\mathcal{I}\mathcal{J}} & \nabla F(\Pi_X(x^*))_{\mathcal{I}\mathcal{K}} \\ \nabla F(\Pi_X(x^*))_{\mathcal{J}\mathcal{I}} & \nabla F(\Pi_X(x^*))_{\mathcal{J}\mathcal{J}} & \nabla F(\Pi_X(x^*))_{\mathcal{J}\mathcal{K}} \\ \nabla F(\Pi_X(x^*))_{\mathcal{K}\mathcal{I}} & \nabla F(\Pi_X(x^*))_{\mathcal{K}\mathcal{J}} & \nabla F(\Pi_X(x^*))_{\mathcal{K}\mathcal{K}} \end{pmatrix}.$$

BVI is said to be R-regular at x^* if $\nabla F(\Pi_X(x^*))_{\mathcal{I}\mathcal{I}}$ is nonsingular and its Shur-complement in the matrix

$$\begin{pmatrix} \nabla F(\Pi_X(x^*))_{\mathcal{I}\mathcal{I}} & \nabla F(\Pi_X(x^*))_{\mathcal{I}\mathcal{J}} \\ \nabla F(\Pi_X(x^*))_{\mathcal{J}\mathcal{I}} & \nabla F(\Pi_X(x^*))_{\mathcal{J}\mathcal{J}} \end{pmatrix}$$

is a P-matrix, see [50].

Proposition 9. *Suppose that $z^* = (u^*, x^*) \in \Re^n \times \Re^n$ is a solution of $H(z) = 0$. If BVI is R-regular at x^*, then all $V \in \partial H(z^*)$ are nonsingular.*

Proof. It is easy to see that for any $V \in \partial H(z^*)$ there exists a $W = (W_u, W_x) \in \partial G(z^*)$ with $W_u, W_x \in \Re^{n\times n}$ such that

$$V = \begin{pmatrix} I & 0 \\ W_u & W_x \end{pmatrix}.$$

Hence, proving V is nonsingular is equivalent to proving W_x is nonsingular. Recall that $G(u, x) = F(p(u, x)) + x - p(u, x)$. Then, for any $W = (W_u, W_x) \in \partial G(z^*)$ with $W_u, W_x \in \Re^{n\times n}$ there exists a $U = (U_u, U_x) \in \partial p(z^*)$ such that

$$W_x = F'(p(z^*))U_x + I - U_x.$$

By the definition of p, we have

$$\partial p_1(z^*) \times \partial p_2(z^*) \times \cdots \times \partial p_n(z^*) = \partial p(z^*).$$

Then for each $i \in N$, the ith row of U, $U_i \in \partial p_i(z^*)$. Apparently, from the definition of p and Lemma 1,

$$U_x = \mathrm{diag}\{(u_x)_i,\ i \in N\},$$

where

$$\begin{cases} (u_x)_i = 1 & \text{if } i \in \mathcal{I} \\ (u_x)_i \in [0,1] & \text{if } i \in \mathcal{J} \\ (u_x)_i = 0 & \text{if } i \in \mathcal{K} \end{cases}.$$

Hence, for each W_x and each $i \in \mathcal{J}$ there exists $\lambda_i \in [0, 1]$ such that

$$\left(W_x^T\right)_i = \begin{cases} \nabla F(p(z^*))_i & \text{if } i \in \mathcal{I} \\ \lambda_i \nabla F(p(z^*))_i + (1-\lambda_i)e_i & \text{if } i \in \mathcal{J} \\ e_i & \text{if } i \in \mathcal{K} \end{cases},$$

where e_i is the ith unit row vector of $\Re^n$, $i \in N$. Then, by using standard analysis (see, e.g., [16, Proposition 3.2]), we can prove that W_x^T, and so W_x, is nonsingular under the assumption of R-regularity (note that $p(z^*) = \Pi_X(x^*)$). Then, any $V \in \partial H(z^*)$ is nonsingular. So, we complete our proof.

□

The following result follows from Theorem 8 and Proposition 9 directly.

Theorem 9. *Suppose that Assumptions 1 and 2 are satisfied and* $z^* = (u^*, x^*) \in \Re^n \times \Re^n$ *is an accumulation point of the infinite sequence* $\{z^k\}$ *generated by Algorithm 1. Suppose that* H *is semismooth at* z^* *and that BVI is R-regular at* x^**. Then (42) and (43) in Theorem 8 hold. Furthermore, if* H *is strongly semismooth at* z^**, then (44) and (45) in Theorem 8 hold.*

By combining Theorems 3, 5 and 9 we can directly obtain the following result.

Theorem 10. *Suppose that the smoothing function* $\phi(\cdot)$ *is defined by either (16) or (18) or (20). Then an infinite sequence* $\{z^k\}$ *is generated by Algorithm 1 and each accumulation point* $z^* = (u^*, x^*) \in \Re^n \times \Re^n$ *of* $\{z^k\}$ *is a solution of* $H(z) = 0$*, under one of the following two conditions,*

(i) for each $z = (u, x) \in \Omega$ *with* $u \in \Re^n_{++}$, $F'(p(z))$ *is a* P*-matrix;*
(ii) if $\phi(\cdot)$ *is defined by either (16) or (18), for each* $i \in N$*, at least one of* a_i *and* b_i *is finite, and for each* $z = (u, x) \in \Omega$ *with* $u \in \Re^n_{++}$, $F'(p(z))$ *is a* P_0*-matrix.*

Further, if the R-regularity holds at x^**, then the whole sequence* $\{z^k\}$ *converges to* z^**, and (42) and (43) in Theorem 8 hold. Moreover, if* F' *is Lipschitz continuous near* $\Pi_X(x^*)$*, then (44) and (45) in Theorem 8 hold.*

Corollary 1. *Suppose that the smoothing function* $\phi(\cdot)$ *is defined by either (16) or (18) or (20). If* F *is a uniform* P*-function on* X*, then*

(i) a bounded infinite sequence $\{z^k\}$ *is generated by Algorithm 1 and the whole sequence* $\{z^k\}$ *converges to the unique solution* $z^* = (u^*, x^*) \in \Re^n \times \Re^n$ *of* $H(z) = 0$*;*

(ii) (42) and (43) in Theorem 8 hold, and
(iii) if F' is Lipschitz continuous near $\Pi_X(x^*)$, then (44) and (45) in Theorem 8 hold.

Proof. It follows from (39) in Theorem 7 that for any $y \in X$ and all $h \in \Re^n$,

$$\max_{i \in N} h_i (F'(y)h)_i \geq \nu \|h\|^2,$$

which, according to [38, Lemma 3.6], implies that $F'(y)$ is a P-matrix. Then, for any $z \in \Re^{2n}$, because $p(z) \in X$, $F'(p(z))$ is a P-matrix. Therefore, from Theorem 10, an infinite sequence $\{z^k\}$ is generated by Algorithm 1 and each accumulation point z^* of $\{z^k\}$ is a solution of $H(z) = 0$. Also, since F is a uniform P-function on X, from Theorem 7, $L(z^0)$, and so $\{z^k\}$, is bounded. Hence, there exists at least one accumulation point $z^* = (u^*, x^*) \in \Re^n \times \Re^n$ of $\{z^k\}$ such that $H(z^*) = 0$. Since $F'(\Pi_X(x^*))$, and so $\nabla F(\Pi_X(x^*))$, is a P-matrix, R-regularity holds at x^*. Hence, we obtain from Theorem 10 that the bounded sequence $\{z^k\}$ converges to z^* and (ii) and (iii) hold. Finally, since F is a uniform P-function, BVI has a unique solution $y^* \in X$ (see, e.g., [25, Theorem 3.9]). Hence, the equation $E(x) = 0$ has a unique solution $x^* = y^* - F(y^*)$, and so, $H(z) = 0$ has a unique solution $z^* = (0, x^*)$. So, we complete our proof.

□

8. Preliminary numerical results

In this section we present some numerical experiments for the nonmonotone line search version of Algorithm 1: *Step 3* is replaced by

Step 3′ Let l_k be the smallest nonnegative integer l satisfying

$$z^k + \delta^l \Delta z^k \in \Omega \tag{51}$$

and

$$\psi(z^k + \delta^l \Delta z^k) \leq \mathcal{W} - 2\sigma(1 - \gamma\|\bar{u}\|)\delta^l \psi(z^k), \tag{52}$$

where $\mathcal{W}$ is any value satisfying

$$\psi(z^k) \leq \mathcal{W} \leq \max_{j=0,1,\ldots,M^k} \psi(z^{k-j})$$

and M^k are nonnegative integers bounded above for all k such that the occurrence of nonnegative indices does not happen. Define $z^{k+1} := z^k + \delta^{l_k} \Delta z^k$.

Remark 2. (i) We choose a nonmonotone line search here is because in most cases it increases the stability of algorithms.
(ii) The requirement (51) is for guaranteeing the global convergence of the algorithm. This requirement automatically holds for our algorithm with a monotone line search, see Proposition 7. The consistency between (51) and (52) can be seen clearly from Propositions 5 and 6.

In the implementation we choose $\mathcal{W}$ as follows:

(1) Set $\mathcal{W} = \psi(z^0)$ at the beginning of the algorithm.
(2) Keep the value of $\mathcal{W}$ fixed as long as

$$\psi(z^k) \leq \min_{j=0,1,\dots,5} \psi(z^{k-j}). \tag{53}$$

(3) If (53) is not satisfied at kth iteration, set $\mathcal{W} = \psi(z^k)$.

For a detailed description of the above nonmonotone line search technique and its motivation, see [14].

The above algorithm was implemented in Matlab and run on a DEC Alpha Server 8200. Throughout the computational experiments, the parameters used in the algorithm were $\delta = 0.5$, $\sigma = 0.5 \times 10^{-4}$, $\bar{u} = (0.1, 0.1, \dots, 0.1)$, and $\gamma = 0.2 \times \min\{1, 1/\|\bar{u}\|\}$. We used $\psi(z) \leq 10^{-12}$ as the stopping rule. The numerical results are summarized in Tables 1–3 for different smoothing functions and different tested problems. In Tables 1–3, **Dim** denotes the number of the variables in the problem, **Start. points** denote the starting points, **Iter** denotes the number of iterations, which is also equal to the number of Jacobian evaluations for the function F, **NF** denotes the number of function evaluations for the function F, and **FF** denotes the value of ψ at the final iterate. In the following, we give a brief description of the tested problems, where **0** is the vector of all zeros and **e** is the vector of all ones. The source reported for the problem is not necessarily the original one.

Problem 1. This is the Kojima-Shindo problem, see [41]. $F(y)$ is not a P_0-function. This problem has two solutions: $y^1 = (\sqrt{6}/2, 0, 0, 0.5)$ and $y^2 = (1, 0, 3, 0)$.
Starting points: (a) **0**, (b)$-$**e**, (c) $\mathbf{e} - F(\mathbf{e})$.

Problem 2. This is a linear complementarity problem. See the first example of Jiang and Qi [28].
Starting points: (a) **0**, (b) **e**.

Problem 3. This is a linear complementarity problem. See the second example of Jiang and Qi [28].
Starting points: (a) **0**, (b) **e**.

Problem 4. This is the fourth example of Watson [56]. This problem represents the KKT conditions for a convex programming problem involving exponentials. The resulting F is monotone on the positive orthant but not even P_0 on R^n.
Starting points: (a) **0**, (b) **e**.

Problem 5. This is a modification of the Mathiesen example of a Walrasian equilibrium model as suggested in [30]. F is not defined everywhere and does not belong to any known class of functions.
Starting points: (a) $\mathbf{0} - F(\mathbf{0})$, (b) $\mathbf{e} - F(\mathbf{e})$, (c) **e**

Problem 6. This is the Nash-Cournot production problem [41]. F is not twice continuously differentiable. F is a P-function on the strictly positive orthant.
Starting points: (a) **0**, (b) **e**, (c) 10**e**.

Problem 7. This is a Mathiesen equilibrium problem [36,41], in which F is not defined everywhere. Two set of constants were used: $(\alpha, b_2, b_3) = (0.75, 1, 0.5)$ and $(\alpha, b_2, b_3) = (0.9, 5, 3)$. We use Problem 7a and Problem 7b to represent this problem with these two set of constants, respectively.
Starting points: (a) $\mathbf{e}$, (b) $\mathbf{e}/2$,

Problem 8. This is the Kojima-Josephy problem, see [14]. $F(x)$ is not a P_0-function. The problem has a unique solution which is not R-regular.
Starting points: (a) $-\mathbf{e}$, (b) $\mathbf{e} - F(\mathbf{e})$, (c) $\mathbf{0}$.

Problem 9. This is a problem arising from a spatial equilibrium model, see [41]. F is a P-function and the unique solution is R-regular.
Starting points: (a) $\mathbf{0}$, (b) $\mathbf{e}$.

Problem 10. This is a traffic equilibrium problem with elastic demand, see [41].
Starting points: (a) All the components are 0 except x_1, x_2, x_3, x_{10}, x_{11}, x_{20}, x_{21}, x_{22}, x_{29}, x_{30}, x_{40}, x_{45} which are 1, x_{39}, x_{42}, x_{43}, x_{46} which are 7, x_{41}, x_{47}, x_{48}, x_{50} which are 6, and x_{44} and x_{49} which are 10, (b) $\mathbf{0}$.

Problem 11. See Problem 9 of [57]. This is a linear complementarity problem for which Lemke's algorithm is known to run in exponential time.
Starting points: (a) $\mathbf{0}$.

Problem 12. This is the third problem of Watson [56], which is a linear complementarity problem with $F(x) = Mx + q$. M is not even semimonotone and none of the standard algebraic techniques can solve it. Let q be the vector with -1 in the 8th coordinate and zeros elsewhere. The continuation method of [56] fails on this problem.
Starting points: (a) $\mathbf{0}$.

Problem 13. See [1]. This is a linear variational inequality problem with lower and upper bounds. Here $a = (0, \ldots, 0)$, $b = (1, \ldots, 1)$.
Starting points: (a) $\mathbf{e}$, (b) $-2\mathbf{e}$.

Problem 14. This problem is transformed from Problem 11 by adding lower and upper bounds to the constraint set. The resulting problem is a linear variational inequality problem with box constraints. Here we choose $a = (-10, \ldots, -10)$, $b = (0, \ldots, 0)$.
Starting points: (a) $\mathbf{0}$, (b) $\mathbf{e}$.

Problem 15. This problem is transformed from Problem 1 by adding lower and upper bounds to the constraint set. The resulting problem is a nonlinear variational inequality problem with box constraints. Here we choose $a = (-10, \ldots, -10)$, $b = (10, \ldots, 10)$.
Starting points: (a) $\mathbf{0}$, (b) $\mathbf{e}$, (c) $\mathbf{0} - F(\mathbf{0})$.

Problem 16. This is a nonlinear variational inequality problem with box constraints [52]. The mapping F is a polynomial operator. Here $a = (0, \ldots, 0)$, $b = (1, \ldots, 1)$.
Starting points: (a) $\mathbf{0}$, (b) $\mathbf{e}$.

Problem 17. This problem is transformed from a linear complementarity problem in [17] by adding lower and upper bounds to the constraint set. The resulting problem is a linear variational inequality problem with box constraints. Here we choose $a = (-10, \ldots, -10)$, $b = (-5, \ldots, -5)$.
Starting points: (a) $\mathbf{e}$.

Table 1. Numerical results for the algorithm with the neural networks smoothing function (16)

Problem	Dim.	Start. point	Iter	NF	FF
Problem 1	4	a	5	8	4.7×10^{-20}
	4	b	5	10	1.3×10^{-19}
	4	c	4	6	2.7×10^{-20}
Problem 2	10,000	a	5	6	1.1×10^{-21}
	10,000	b	5	6	1.1×10^{-21}
Problem 3	10,000	a	5	6	1.1×10^{-21}
	10,000	b	5	6	1.1×10^{-21}
Problem 4	5	a	19	20	1.4×10^{-23}
	5	b	15	16	1.9×10^{-14}
Problem 5	4	a	4	5	1.8×10^{-13}
	4	b	5	6	3.4×10^{-14}
	4	c	7	8	2.7×10^{-20}
Problem 6	10	a	9	10	1.4×10^{-14}
	10	b	7	8	1.3×10^{-20}
	10	c	7	8	1.6×10^{-21}
Problem 7a	4	a	7	9	9.2×10^{-20}
	4	b	7	9	1.8×10^{-15}
Problem 7b	4	a	fail		
	4	b	7	8	1.9×10^{-16}
Problem 8	4	a	6	9	1.1×10^{-25}
	4	b	5	7	9.0×10^{-13}
	4	c	> 50		
Problem 9	42	a	11	22	6.7×10^{-18}
	42	b	10	14	9.0×10^{-19}
Problem 10	50	a	12	26	8.3×10^{-18}
	50	b	16	42	1.4×10^{-14}
Problem 11	1000	a	10	11	1.7×10^{-20}
Problem 12	10	a	6	10	4.2×10^{-24}
Problem 13	10,000	a	6	7	1.4×10^{-21}
	10,000	b	5	6	1.1×10^{-21}
Problem 14	1000	a	6	7	1.1×10^{-21}
	1000	b	5	6	1.1×10^{-21}
Problem 15	4	a	fail		
	4	b	4	5	1.8×10^{-13}
	4	c	6	7	3.9×10^{-13}
Problem 16	10,000	a	6	8	1.1×10^{-21}
	10,000	b	7	8	2.2×10^{-23}
Problem 17	400	a	5	6	1.1×10^{-21}

The numerical results reported in Tables 1–3 show that the algorithms proposed in this paper for the three chosen smoothing functions work quite well for both nonlinear complementarity problems and box constrained variational inequalities. It is observed during our numerical experiment that the algorithms based on the neural networks smoothing function (16) and the Chen-Harker-Kanzow-Smale smoothing function (18),

Table 2. Numerical results for the algorithm with the Chen-Harker-Kanzow-Smale smoothing function (18)

Problem	Dim.	Start. point	Iter	NF	FF
Problem 1	4	a	6	9	5.6×10^{-18}
	4	b	6	11	3.3×10^{-23}
	4	c	5	7	3.9×10^{-23}
Problem 2	10,000	a	5	6	1.1×10^{-21}
	10,000	b	5	6	1.1×10^{-21}
Problem 3	10,000	a	5	6	1.1×10^{-21}
	10,000	b	5	6	1.1×10^{-21}
Problem 4	5	a	19	20	7.6×10^{-26}
	5	b	16	17	2.8×10^{-24}
Problem 5	4	a	5	7	8.8×10^{-15}
	4	b	6	7	1.8×10^{-22}
	4	c	8	9	2.7×10^{-20}
Problem 6	10	a	10	11	5.0×10^{-15}
	10	b	8	9	1.4×10^{-22}
	10	c	7	8	1.6×10^{-13}
Problem 7a	4	a	8	11	3.4×10^{-23}
	4	b	7	10	4.2×10^{-13}
Problem 7b	4	a	5	7	2.7×10^{-21}
	4	b	4	6	1.1×10^{-15}
Problem 8	4	a	6	9	4.2×10^{-16}
	4	b	6	8	3.2×10^{-19}
	4	c	> 50		
Problem 9	42	a	15	28	8.2×10^{-26}
	42	b	14	24	8.3×10^{-20}
Problem 10	50	a	14	28	1.2×10^{-19}
	50	b	17	41	5.5×10^{-15}
Problem 11	1000	a	13	15	3.6×10^{-18}
Problem 12	10	a	6	10	9.9×10^{-20}
Problem 13	10,000	a	11	12	5.5×10^{-17}
	10,000	b	13	14	9.7×10^{-26}
Problem 14	1000	a	11	13	7.1×10^{-19}
	1000	b	9	11	6.6×10^{-20}
Problem 15	4	a	6	33	8.4×10^{-13}
	4	b	4	5	5.0×10^{-13}
	4	c	6	7	1.4×10^{-16}
Problem 16	10,000	a	6	8	4.7×10^{-21}
	10,000	b	7	9	4.8×10^{-17}
Problem 17	400	a	5	6	1.1×10^{-21}

in particular the latter one, have stronger numerical stability. However, when using the uniform smoothing function (20) we only solve a form reduced linear equation per iteration. A possible way is to use the neural networks smoothing function (16) or the Chen-Harker-Kanzow-Smale smoothing function (18) at the first several iterations and then to use the uniform smoothing function (20). We leave this as a future research topic.

Table 3. Numerical results for the algorithm with the uniform smoothing function (20)

Problem	Dim.	Start. point	Iter	NF	FF
Problem 1	4	a	> 50		
	4	b	6	10	9.8×10^{-20}
	4	c	4	6	2.7×10^{-20}
Problem 2	10,000	a	5	6	1.1×10^{-21}
	10,000	b	5	6	1.1×10^{-21}
Problem 3	10,000	a	5	6	1.1×10^{-21}
	10,000	b	4	5	1.1×10^{-21}
Problem 4	5	a	19	20	2.5×10^{-13}
	5	b	15	16	9.9×10^{-17}
Problem 5	4	a	4	5	1.3×10^{-17}
	4	b	4	5	3.7×10^{-16}
	4	c	9	58	2.7×10^{-20}
Problem 6	10	a	12	13	1.9×10^{-25}
	10	b	7	8	1.3×10^{-20}
	10	c	7	8	1.6×10^{-21}
Problem 7a	4	a	7	10	4.0×10^{-24}
	4	b	6	9	4.0×10^{-14}
Problem 7b	4	a	5	7	5.0×10^{-16}
	4	b	4	6	3.8×10^{-24}
Problem 8	4	a	5	8	4.0×10^{-13}
	4	b	5	7	9.0×10^{-13}
	4	c	> 50		
Problem 9	42	a	10	18	9.9×10^{-14}
	42	b	10	14	5.7×10^{-19}
Problem 10	50	a	11	25	5.8×10^{-16}
	50	b	16	41	6.5×10^{-13}
Problem 11	1000	a	11	12	1.1×10^{-21}
Problem 12	10	a	6	15	4.7×10^{-19}
Problem 13	10,000	a	8	10	9.7×10^{-26}
	10,000	b	10	12	9.7×10^{-26}
Problem 14	1000	a	11	12	1.1×10^{-21}
	1000	b	4	5	1.1×10^{-21}
Problem 15	4	a	fail		
	4	b	4	5	5.0×10^{-13}
	4	c	6	7	8.9×10^{-17}
Problem 16	10,000	a	5	7	1.1×10^{-21}
	10,000	b	7	9	2.5×10^{-15}
Problem 17	400	a	4	5	1.1×10^{-21}

9. Conclusions

In this paper we constructed a new class of smoothing Newton methods for solving nonlinear complementarity problems and box constrained variational inequalities by making use of the facts that the most often used smoothing functions are continuously differentiable jointly with μ and w at any $(\mu, w) \in \Re_{++} \times \Re$ and they lead to strongly semismooth functions on $\Re^2$. The techniques provided in this paper can be applied to smoothing Newton methods based on the mapping $W(\cdot)$ defined in (4). Numerical results showed that our algorithms worked very satisfactorily for the tested problems. We expect that these algorithms can be used to solve practical large-scale problems efficiently.

In [27] Jiang proposed a smoothing method for solving nonlinear complementarity problems with P_0-functions. Jiang's approach shares some similarities with ours in the sense that the smoothing parameter is treated as a variable. For any $a, b, \varepsilon \in \Re$, define

$$\phi_{FB}(a, b, \varepsilon) = \sqrt{a^2 + b^2 + \varepsilon^2} - (a + b). \tag{54}$$

This function is a smoothed form of Fischer-Burmeister function and was first defined by Kanzow [31]. Jiang [27] proves that $\psi_{FB}(\cdot, \cdot, \cdot) := \phi_{FB}(\cdot, \cdot, \cdot)^2$ is continuously differentiable on $\Re^3$. Define $G : \Re^{n+1} \to \Re^n$ by

$$G_i(\varepsilon, x) := \phi_{FB}(x_i, F_i(x), \varepsilon), \quad i = 1, 2, ..., n. \tag{55}$$

Jiang [27] provided a different form of H, which was defined by

$$H(\varepsilon, x) := \begin{pmatrix} e^\varepsilon - 1 \\ G(\varepsilon, x) \end{pmatrix}. \tag{56}$$

An interesting property of such defined H is that for any $\varepsilon > 0$ and any $\lambda \in (0, 1]$,

$$\varepsilon + \lambda \Delta\varepsilon > 0$$

and

$$\varepsilon + \lambda \Delta\varepsilon < \varepsilon,$$

where $d := (\Delta\varepsilon, \Delta x) \in \Re \times \Re^n$ is a solution of

$$H(\varepsilon, x) + H'(\varepsilon, x)d = 0.$$

Based on this observation, Jiang [27] designed a smoothing Newton method for solving the NCP with the assumption that F is a P_0-function. By using the continuous differentiability of $\|H\|^2$ and the assumption that the search directions are bounded, which can be satisfied by assuming that F is a uniform P-function, Jiang proved global and local superlinearly (quadratically) convergent results of his method. It is noted that Jiang's idea may be used to any smoothing function H where a smoothing parameter is involved if the following two conditions are satisfied:

(i) The square $\|H\|^2$ is continuously differentiable;
(ii) The search directions are bounded.

Meanwhile, our approach needs neither condition (i) nor condition (ii). The key point is that we can control the smoothing parameter in such a way that it converges neither too fast nor too slow by using a particularly designed Newton equation and a line search model. To see the benefit of our approach clearly, let's consider the NCP with a P_0-function. For this problem we proved that any accumulation point of our iteration sequence is a solution while Jiang proved the same result under the additional condition (ii). Moreover, we require F to be a P_0-function on $\Re^n_+$ only instead of on $\Re^n$.

After the announcement of this paper, our method has soon been used to solve various regularized smoothing equations to get stronger global convergent results [53, 43, 60] and to solve extended vertical linear complementarity problems [44].

In this paper we treated the smoothing parameter u the same as the original variable x except that u is always kept positive. It is an interesting question to know that if the updating rule for u used in this paper can be relaxed or more interestingly if it is possible not to treat the parameter u as a free variable while superlinearly convergent results can still be obtained for the problems considered here or in [53, 43, 60]. We leave these as further reaserch topics.

Acknowledgements. The authors would like to thank Houduo Qi and the two referees for their helpful comments.

Appendix A

Proof of Proposition 1. From (16) and the definition of $q_{cd}(\cdot)$, we can see that $q_{cd}(\cdot)$ is at least twice continuously differentiable at any $x \in \Re^2$ with $x_1 \neq 0$. So, we only need to prove that (24) holds at any $x = (0, x_2)$ with $x_2 \in \Re$. After simple computations, for any $(\mu, w) \in \Re^2$ with $\mu > 0$, we have

$$\nabla q_{cd}(\mu, w) = \begin{pmatrix} t(\mu, w) \\ \dfrac{1}{1 + e^{(c-w)/\mu}} - \dfrac{1}{1 + e^{(d-w)/\mu}} \end{pmatrix}, \tag{A. 1}$$

where

$$t(\mu, w) = \ln\left[1 + e^{(c-w)/\mu}\right] + \frac{c - w}{\mu} \frac{1}{1 + e^{(c-w)/\mu}}$$
$$- \ln\left[1 + e^{(d-w)/\mu}\right] - \frac{d - w}{\mu} \frac{1}{1 + e^{(d-w)/\mu}} + \frac{d - c}{\mu}$$

and for any $(\mu, w) \in \Re^2$ with $\mu < 0$, we have

$$\nabla q_{cd}(\mu, w) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \nabla q_{cd}(-\mu, w). \tag{A. 2}$$

Suppose $x = (0, x_2) \in \Re^2$ and $h = (h_1, h_2) \in \Re^2$. If $h_1 \neq 0$, it then follows from the fact that q_{cd} is continuously differentiable at $x + h$,

$$\partial q_{cd}(x + h) = \left\{\nabla q_{cd}(x + h)^T\right\}.$$

We consider several cases:

(i) $x_2 < c$. Suppose that h is small enough such that for all $t \in [0, 1]$, $x_2 + th_2 < c$. Then from the definition,

$$q'_{cd}(x; h) = \lim_{t \downarrow 0} \frac{q_{cd}(x + th) - q_{cd}(x)}{t} = 0.$$

Then, for any $V \in \partial q_{cd}(x+h)$, $h \to 0$ and $h_1 \neq 0$,

$$\begin{aligned} Vh - q'_{cd}(x;h) &= \nabla q_{cd}(x+h)^T h - 0 \\ &= |h_1| \ln\left\{1 + e^{[c-(x_2+h_2)]/|h_1|}\right\} - |h_1| \ln\left\{1 + e^{[d-(x_2+h_2)]/|h_1|}\right\} \\ &\quad + \frac{c - x_2}{1 + e^{[c-(x_2+h_2)]/|h_1|}} - \frac{d - x_2}{1 + e^{[d-(x_2+h_2)]/|h_1|}} + (d - c) \\ &= O(h_1^2) \\ &= O(\|h\|^2) \end{aligned}$$

and for any $V \in \partial q_{cd}(x+h)$, $h = (0, h_2) \to 0$,

$$Vh - q'_{cd}(x;h) = V_1 * 0 + 0 * h_2 - 0 = 0$$

because in the latter case $q'_{cd}(x;h) = 0$ and for any $V \in \partial q_{cd}(0, x_2+h_2)$ and $x_2+h_2 < c$, $V_2 = 0$.
(ii) $c < x_2 < d$. Suppose that h is small enough such that for all $t \in [0,1]$, $c < x_2 + th_2 < d$. Then, from the definition,

$$q'_{cd}(x;h) = \lim_{t \downarrow 0} \frac{q_{cd}(x+th) - q_{cd}(x)}{t} = h_2.$$

Then for any $V \in \partial q_{cd}(x+h)$, $h \to 0$ and $h_1 \neq 0$,

$$\begin{aligned} Vh - q'_{cd}(x;h) &= \nabla q_{cd}(x+h)^T h - h_2 \\ &= O(h_1^2) \\ &= O(\|h\|^2) \end{aligned}$$

and for any $V \in \partial q_{cd}(x+h)$, $h = (0, h_2) \to 0$,

$$Vh - q'_{cd}(x;h) = V_1 * 0 + 1 * h_2 - h_2 = 0$$

because in the latter case for any $V \in \partial q_{cd}(0, x_2+h_2)$ and $c < x_2 + h_2 < d$, $V_2 = 1$.
(iii) $x_2 = c$. Suppose that $h \in \Re^2$ is sufficiently small such that $x_2 + h_2 < b$. Then from the definition, if $h_1 \neq 0$,

$$q'_{cd}(x;h) = \lim_{t \downarrow 0} \frac{q_{cd}(x+th) - q_{cd}(x)}{t} = |h_1| \ln\left(1 + e^{-h_2/|h_1|}\right) + h_2$$

and if $h_1 = 0$

$$q'_{cd}(x;h) = \max\{0, h_2\}.$$

Then, for any $V \in \partial q_{cd}(x+h)$, $h \to 0$ and $h_1 \neq 0$,

$$\begin{aligned}
Vh - q'_{cd}(x;h) &= V_1 * h_1 + V_2 * h_2 - \left[|h_1| \ln\left(1 + e^{-h_2/|h_1|}\right) + h_2\right] \\
&= |h_1| \ln\left(1 + e^{-h_2/|h_1|}\right) - \frac{h_2}{|h_1|} \frac{1}{1 + e^{-h_2/|h_1|}} |h_1| \\
&\quad - |h_1| \ln\left[1 + e^{(d-c-h_2)/|h_1|}\right] + \frac{h_2}{|h_1|} \frac{1}{1 + e^{(d-c-h_2)/|h_1|}} |h_1| \\
&\quad + \frac{d-c}{|h_1|} + \frac{h_2}{1 + e^{-h_2/|h_1|}} |h_1| - \frac{h_2}{1 + e^{(d-c-h_2)/|h_1|}} \\
&\quad - \left[|h_1| \ln\left(1 + e^{-h_2/|h_1|}\right) + h_2\right] \\
&= -|h_1| \ln\left[1 + e^{(d-c-h_2)/|h_1|}\right] + d - c - h_2 \\
&= O(h_1^2) \\
&= O(\|h\|^2)
\end{aligned}$$

and for any $V \in \partial q_{cd}(x+h)$, $h = (0, h_2) \to 0$,

$$Vh - q'_{cd}(x;h) = V_1 * 0 + V_2 * h_2 - \max\{0, h_2\} = 0$$

because in the latter case $q'_{cd}(x;h) = \max\{0, h_2\}$ and for any $V \in \partial q_{cd}(0, x_2 + h_2)$ and $x_2 + h_2 < d$ if $h_2 > 0$, $V_2 = 1$ and if $h_2 < 0$, $V_2 = 0$.
(iv) $d < x_2$. Suppose that h is small enough such that for all $t \in [0, 1]$, $d < x_2 + th_2$. Then, from the definition,

$$q'_{cd}(x;h) = \lim_{t \downarrow 0} \frac{q_{cd}(x+th) - q_{cd}(x)}{t} = 0.$$

Then, for any $V \in \partial q_{cd}(x+h)$, $h \to 0$ and $h_1 \neq 0$,

$$\begin{aligned}
Vh - q'_{cd}(x;h) &= \nabla q_{cd}(x+h)^T h - 0 \\
&= O(h_1^2) \\
&= O(\|h\|^2)
\end{aligned}$$

and for any $V \in \partial q_{cd}(x+h)$, $h = (0, h_2) \to 0$,

$$Vh - q'_{cd}(x;h) = V_1 * 0 + 0 * h_2 - 0 = 0.$$

because in the latter case for any $V \in \partial q_{cd}(0, x_2 + h_2)$ and $x_2 + h_2 > d$, $V_2 = 0$.
(v) $x_2 = d$. The proof of this case is similar to that of case (iii). We omit the detail here. Then we have proved that $q_{cd}(\cdot)$ is a strongly semismooth function.

□

Proof of Proposition 2. By the definition, for any $(\mu, w) \in \Re^2$,

$$q_{cd}(\mu, w) = \frac{c + \sqrt{(c-w)^2 + 4\mu^2}}{2} + \frac{d - \sqrt{(d-w)^2 + 4\mu^2}}{2}.$$

It is easy to prove that $\sqrt{\nu^2 + \mu^2}$ is a strongly semismooth function. Since the composition of strongly semismooth functions is still a strongly semismooth function [21, Theorem 19], the two functions $\frac{c+\sqrt{(c-w)^2+4\mu^2}}{2}$ and $\frac{d-\sqrt{(d-w)^2+4\mu^2}}{2}$ are strongly semismooth ones, and so, $q_{cd}(\cdot)$ is a strongly semismooth function. This completes our proof. □

Proof of Proposition 3. For the sake of simplicity we only prove the case $c = 0$ and $d = \infty$. The proof of other cases is very similar.

From (22) (note that we assume $c = 0$ and $d = \infty$) and the definition of $q_{cd}(\cdot)$, we can see that $q_{cd}(\cdot)$ is continuously differentiable at any $(\mu, w) \in \Re^2$ with $\mu \neq 0$ and if $\mu > 0$,

$$q'_{cd}(\mu, w) = \begin{cases} (0 \ \ 0) & \text{if } w \leq -\mu/2 \\ \left(\frac{1}{8} - \frac{w^2}{2\mu^2} \ \ \frac{w}{\mu} + \frac{1}{2}\right) & \text{if } |w| < \mu/2 \\ (0 \ \ 1) & \text{if } w \geq \mu/2 \end{cases}$$

and if $\mu < 0$,

$$q'_{cd}(\mu, w) = q'_{cd}(-\mu, w) \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

It is easy to verify that $q'_{cd}(\cdot)$ is locally Lipschitz continuous around any $(\mu, w) \in \Re^2$ with $\mu \neq 0$ and it is known that a function which is continuously differentiable at a certain point is strongly semismooth at this point if its derivative is locally Lipschitz continuous around this point. Hence $q'_{cd}(\cdot)$ is strongly semismooth at any $(\mu, w) \in \Re^2$ with $\mu \neq 0$.

Next, we prove that (24) holds at any $x = (0, x_2) \in \Re^2$. We consider the following several cases to prove this.

(i) $x_2 < 0$. Then for any $V \in \partial q_{cd}(x + h)$, $h \to 0$ and $h_1 \neq 0$,

$$Vh - q'_{cd}(x; h) = 0 * h_1 + 0 * h_2 - 0 = 0$$

and for any $V \in \partial q_{cd}(x + h)$, $h = (0, h_2) \to 0$,

$$Vh - q'_{cd}(x; h) = V_1 * 0 + 0 * h_2 - 0 = 0.$$

(ii) $x_2 > 0$. Then for any $V \in \partial q_{cd}(x + h)$, $h \to 0$ and $h_1 \neq 0$,

$$Vh - q'_{cd}(x; h) = 0 * h_1 + 1 * h_2 - h_2 = 0$$

and for any $V \in \partial q_{cd}(x + h)$, $h = (0, h_2) \to 0$,

$$Vh - q'_{cd}(x; h) = V_1 * 0 + 1 * h_2 - h_2 = 0.$$

(iii) $x_2 = 0$. Then for any $V \in \partial q_{cd}(x + h)$, $h \to 0$, $h_1 \neq 0$ and $h_2 \leq -|h_1|/2$,

$$Vh - q'_{cd}(x; h) = 0 * h_1 + 0 * h_2 - 0 = 0;$$

for any $V \in \partial q_{cd}(x + h)$, $h \to 0$, $h_1 \neq 0$ and $h_2 \geq |h_1|/2$,

$$Vh - q'_{cd}(x; h) = 0 * h_1 + 1 * h_2 - h_2 = 0;$$

for any $V \in \partial q_{cd}(x+h)$, $h \to 0$, $h_1 \neq 0$ and $|h_2| < |h_1|/2$,

$$Vh - q'_{cd}(x;h) = \left(\frac{1}{8} - \frac{h_2^2}{2h_1^2}\right)|h_1| + \left(\frac{h_2}{|h_1|} + \frac{1}{2}\right)h_2 - \frac{(h_2 + |h_1|/2)^2}{2|h_1|} = 0$$

and for any $V \in \partial q_{cd}(x+h)$, $h = (0, h_2) \to 0$,

$$Vh - q'_{cd}(x;h) = V_1 * 0 + V_2 * h_2 - \max\{0, h_2\} = \max\{0, h_2\} - \max\{0, h_2\} = 0$$

because in the latter case $q'_{cd}(x;h) = \max\{0, h_2\}$ and for any $V \in \partial q_{cd}(0, x_2 + h_2)$ if $h_2 > 0$, $V_2 = 1$ and if $h_2 < 0$, $V_2 = 0$.

Then we have proved that $q_{cd}(\cdot)$ is a strongly semismooth function.

□

References

1. Ahn, B.H. (1983): Iterative methods for linear complementarity problem with upperbounds and lowerbounds. Math. Program. **26**, 295–315
2. Billups, S.C., Dirkse, S.P., Ferris, M.C. (1997): A comparison of algorithms for large-scale mixed complementarity problems. Comput. Optim. Appl. **7**, 3–25
3. Chen, B., Chen, X. (1999): A global and local superlinear continuation-smoothing method for $P_0 + R_0$ and monotone NCP. SIAM J. Optim. **9**, 624–645
4. Chen, B., Harker, P.T. (1993): A non-interior-point continuation method for linear complementarity problems. SIAM J. Matrix Anal. Appl. **14**, 1168–1190
5. Chen, B., Harker, P.T. (1995): A continuation method for monotone variational inequalities. Math. Program. **69**, 237–253
6. Chen, B., Harker, P.T. (1997): Smooth approximations to nonlinear complementarity problems. SIAM J. Optim. **7**, 403–420
7. Chen, B., Harker, P.T., Pınar, M.C. (1995): Continuation methods for nonlinear complementarity problems via normal maps. Preprint, Department of Management and Systems, Washington State University, Pullman, November 1995
8. Chen, B., Xiu, N. (1999): A global linear and local quadratic non-interior continuation method for nonlinear complementarity problems based on Chen-Mangasarian smoothing function. SIAM J. Optim. **9**, 605–623
9. Chen, C., Mangasarian, O.L. (1995): Smoothing methods for convex inequalities and linear complementarity problems. Math. Program. **71**, 51–69
10. Chen, C., Mangasarian, O.L. (1996): A class of smoothing functions for nonlinear and mixed complementarity problems. Comput. Optim. Appl. **5**, 97–138
11. Chen, X., Qi, L., Sun, D. (1998): Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities. Math. Comput. **67**, 519–540
12. Chen, X., Ye, Y. (1999): On homotopy-smoothing methods for variational inequalities. SIAM J. Control Optim. **37**, 589–616
13. Clarke, F.H. (1983): Optimization and Nonsmooth Analysis. Wiley, New York
14. De Luca, T., Facchinei, F., Kanzow, C. (1996): A semismooth equation approach to the solution of nonlinear complementarity problems. Math. Program. **75**, 407–439
15. Facchinei, F., Fischer, A., Kanzow, C. (1998): Regularity properties of a semismooth equation reformulation of variational inequalities. SIAM J. Optim. **8**, 850–869
16. Facchinei, F., Soares, J. (1997): A new merit function for nonlinear complementarity problems and a related algorithm. SIAM J. Optim. **7**, 225–247
17. Fathi, Y. (1979): Computational complexity of LCPs associated with positive definite matrices. Math. Program. **17**, 335–344
18. Ferris, M.C., Pang, J.-S. (1997): Engineering and economic applications of complementarity problems. SIAM Rev. **39**, 669–713
19. Fischer, A. (1992): A special Newton-type optimization method. Optimization **24**, 269–284
20. Fischer, A. (1995): An NCP-function and its use for the solution of complementarity problems. In: Recent Advances in Nonsmooth Optimization, Du, D., Qi, L., Womersley, R., eds., World Scientific Publishers, New Jersey, pp. 88–105

21. Fischer, A. (1997): Solution of monotone complementarity problems with locally Lipschitzian functions. Math. Program. **76**, 513–532
22. Fukushima, M. (1996): Merit functions for variational inequality and complementarity problems. In: Nonlinear Optimization and Applications, Di Pillo, G., Giannessi, F., eds., Plenum Publishing Corporation, New York, pp. 155–170
23. Gabriel, S.A., Moré, J.J. (1997): Smoothing of mixed complementarity problems. In: Complementarity and Variational Problems: State of the Art, Ferris, M.C., Pang, J.S., eds., SIAM, Philadelphia, Pennsylvania, pp. 105–116
24. Han, J., Sun, D. (1998): Newton-type methods for variational inequalities. Advances in Nonlinear Programming, Yuan, Y., ed., Kluwer, Boston, pp. 105–118
25. Harker, P.T., Pang, J.-S. (1990): Finite-dimensional variational inequality and nonlinear complementarity problem: A survey of theory, algorithms and applications. Math. Program. **48** (1990), 161–220
26. Hotta, K., Yoshise, A. (1999): Global convergence of a class of non-interior-point algorithms using Chen-Harker-Kanzow-Smale functions for nonlinear complementarity problems. Math. Program. **86**, 105–133
27. Jiang, H. (1997): Smoothed Fischer-Burmeister equation methods for the nonlinear complementarity problem. Preprint, Department of Mathematics & Statistics, the University of Melbourne, Victoria 3052, Australia, June 1997
28. Jiang, H., Qi, L. (1997): A new nonsmooth equation approach to nonlinear complementarity problems. SIAM J. Control Optim. **35**, 178–193
29. Jiang, H., Qi, L., Chen, X., Sun, D. (1996): Semismoothness and superlinear convergence in nonsmooth optimization and nonsmooth equations. In: Nonlinear Optimization and Applications, Di Pillo, G., Giannessi, F., eds., Plenum Publishing Corporation, New York, pp. 197–212
30. Kanzow, C. (1994): Some equation-based methods for the nonlinear complementarity problem. Optim. Methods Software **3**, 327–340
31. Kanzow, C. (1996): Some noninterior continuation methods for linear complementarity problems. SIAM J. Matrix Anal. Appl. **17**, 851–868
32. Kanzow, C., Fukushima, M. (1998): Theoretical and numerical investigation of the D-gap function for box constrained variational inequalities. Math. Program. **83**, 55–87
33. Kanzow, C., Jiang, H. (1998): A continuation method for (strongly) monotone variational inequalities. Math. Program. **81**, 103–125
34. Kanzow, C., Qi, H.-D. (1999): A QP-free constrained Newton-type method for variational inequality problems. Math. Program. **85**, 81–106
35. Luo, Z.-Q., Tseng, P. (1997): A new class of merit functions for the nonlinear complementarity problem. In: Complementarity and Variational Problems: State of the Art, Ferris, M.C., Pang, J.S., eds., SIAM, Philadelphia, Pennsylvania, pp. 204–225
36. Mathiesen, L. (1987): An algorithm based on a sequence of linear complementarity problems applied to a Walrasian equilibrium model: An example. Math. Program. **37**, 1–18
37. Mifflin, R. (1977): Semismooth and semiconvex functions in constrained optimization. SIAM J. Control Optim. **15**, 957–972
38. Moré, J.J. (1974): Coercivity conditions in nonlinear complementarity problems. SIAM Rev. **16**, 1–16
39. Moré, J.J., Wu, Z. (1994): Smoothing techniques for macromolecular global optimization. Preprint MCS-P542-0995, Argonne National Laboratory, Argonne, Illinois
40. Pang, J.-S. (1995): Complementarity problems. In: Handbook of Global Optimization, Horst, R., Pardalos, P., eds., Kluwer Academic Publishers, Boston, pp. 271–338
41. Pang, J.-S., Gabriel, S.A. (1993): NE/SQP: A robust algorithm for the nonlinear complementarity problem. Math. Program. **60**, 295–337
42. Pınar, M.C., Zenios, S.A. (1994): On smoothing exact penalty functions for constrained optimization. SIAM J. Optim. **4**, 486–511
43. Qi, H.-D. (1997): A regularized smoothing Newton method for box constrained variational inequality problems with P_0 functions. Preprint, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, August 1997. To appear in SIAM J. Optim.
44. Qi, H.-D., Liao, L.Z. (1999): A smoothing Newton method for extended vertical linear complementarity problems. SIAM J. Matrix Anal. Appl. **21**, 45–66
45. Qi, L. (1993): Convergence analysis of some algorithms for solving nonsmooth equations. Math. Oper. Res. **18**, 227–244
46. Qi, L., Chen, X. (1995): A globally convergent successive approximation method for severely nonsmooth equations. SIAM J. Control Optim. **33**, 402–418
47. Qi, L., Sun, D. (2000): Improving the convergence of non-interior point algorithms for nonlinear complementarity problems. Math. Comput. **69**, 283–304
48. Qi, L., Sun, J. (1993): A nonsmooth version of Newton's method. Math. Program. **58**, 353–367

49. Robinson, S.M. (1992): Normal maps induced by linear transformation. Math. Oper. Res. **17**, 691–714
50. Robinson, S.M. (1983): Generalized equations. In: Mathematical Programming: The State of the Art, Bachem, A., Grötschel, M., Korte, B., eds., Springer-Verlag, Berlin, pp. 346–347
51. Smale, S. (1986): Algorithms for solving equations. In: Proceedings of the International Congress of Mathematicians. Berkeley, California, pp. 172–195
52. Sun, D. (1996): A class of iterative methods for solving nonlinear projection equations. J. Optim. Theory Appl. **91**, 123–140
53. Sun, D. (1999): A regularization Newton method for solving nonlinear complementarity problems. Appl. Math. Optim. **40**, 315–339
54. Sun, D., Han, J. (1997): Newton and quasi-Newton methods for a class of nonsmooth equations and related problems. SIAM J. Optim. **7**, 463–480
55. Sun, D., Womersley, R.S. (1999): A new unconstrained differentiable merit function for box constrained variational inequality problems and a damped Gauss-Newton method. SIAM J. Optim. **9**, 388–413
56. Watson, L.T. (1979): Solving the nonlinear complementarity problem by a homotopy method. SIAM J. Control Optim. **17**, 36–46
57. Xiao, B., Harker, P.T. (1994): A nonsmooth Newton method for variational inequalities, II: Numerical results. Math. Program. **65**, 195–216
58. Xu, S. (1980): The global linear convergence of an infeasible non-interior path-following algorithm for complementarity problems with uniform P-functions. Preprint, Department of Mathematics, University of Washington, Seattle, WA 98195, December 1996
59. Zang, I. (1980): A smoothing-out technique for min–max optimization. Math. Program. **19**, 61–71
60. Zhou, G., Sun, D., Qi, L. (1998): Numerical experiments for a class of squared smoothing Newton methods for complementarity and variational inequality problems. In: Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, Fukushima, M., Qi , L., eds., Kluwer Academic Publishers B.V., 421–441