GLOBAL AND SUPERLINEAR CONVERGENCE OF THE
SMOOTHING NEWTON METHOD AND ITS APPLICATION
TO GENERAL BOX CONSTRUANDED VARIATIONAL
INEQUALITIES

X. CHEN, L. QI, AND D. SUN

ABSTRACT. The smoothing Newton method for solving a system of nonsmooth equations $F(x) = 0$, which may arise from the nonlinear complementarity problem, the variational inequality problem or other problems, can be regarded as a variant of the smoothing method. At the $k$th step, the nonsmooth function $F$ is approximated by a smooth function $f(\cdot, \varepsilon_k)$, and the derivative of $f(\cdot, \varepsilon_k)$ at $x_k$ is used as the Newton iterative matrix. The merits of smoothing methods and smoothing Newton methods are global convergence and convenience in handling. In this paper, we show that the smoothing Newton method is also superlinearly convergent if $F$ is semismooth at the solution and $f$ satisfies a Jacobian consistency property. We show that most common smooth functions, such as the Gabriel-Moré function, have this property. As an application, we show that for box constrained variational inequalities if the involved function is $P$-uniform, the iteration sequence generated by the smoothing Newton method will converge to the unique solution of the problem globally and superlinearly (quadratically).

1. INTRODUCTION

Let $p, q : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be two smooth (continuously differentiable) mappings and $X$ be a closed convex set in $\mathbb{R}^n$. The general variational inequality problem, GVI($X, p, q$) for short, is to find a vector $x \in \mathbb{R}^n$ such that

$$q(x) \in X, \quad (y - q(x))^T p(x) \geq 0 \quad \text{for all } y \in X.$$  \hspace{1cm} (1.1)

It is known (for example, see [10] for a proof in the case that $q(x) = x$) that GVI($X, p, q$) is equivalent to finding a zero of the following nonsmooth equation

$$q(x) - \Pi_X [q(x) - p(x)] = 0,$$  \hspace{1cm} (1.2)

where $\Pi_X$ is the projection operator onto $X$ under the Euclidean norm. Equation (1.2) is called the generalized normal equation in [39]. GVI($X, p, q$) is a generalization of variational inequalities and general complementarity problems. The variational inequality problem is to find an $x \in X$ such that

$$(y - x)^T p(x) \geq 0 \quad \text{for all } y \in X.$$  \hspace{1cm} (1.3)

Received by the editor June 17, 1996 and, in revised form, January 9, 1997.

1991 Mathematics Subject Classification. Primary 90C33, 90C30, 65H10.

Key words and phrases. Variational inequalities, nonsmooth equations, smoothing approximation, smoothing Newton method, convergence.

This work is supported by the Australian Research Council.

©1998 American Mathematical Society
The general complementarity problem, GCP($p, q$) for short, is to find an $x \in \mathbb{R}^n$ such that
\begin{equation}
q(x) \geq 0, \quad p(x) \geq 0, \quad q(x)^T p(x) = 0.
\end{equation}
When $q(x) = x$, GCP($p, q$) reduces to the nonlinear complementarity problem of finding an $x \in \mathbb{R}^n$ such that
\begin{equation}
x \geq 0, \quad p(x) \geq 0, \quad x^T p(x) = 0.
\end{equation}
A lot of effort has been spent on complementarity problems and variational inequalities, for a comprehensive survey see [20], [37].

In this paper we focus on general box constrained variational inequalities, i.e., we assume that in (1.1) $X$ has the following box form:
\begin{equation}
X = \{ x \in \mathbb{R}^n \mid l \leq x \leq u \},
\end{equation}
where $l \in \{ \mathbb{R} \cup \{-\infty\} \}^n$, $u \in \{ \mathbb{R} \cup \{+\infty\} \}^n$ and $l < u$. In this case GVI($X, p, q$) will be denoted by GVI($l, u, p, q$). GVI($l, u, p, q$) includes two very useful models: general complementarity problems and box constrained variational inequalities while the latter is actually equivalent to what is called mixed complementarity problems in some papers [2], [4], [9], [18]. Furthermore, GVI($l, u, p, q$) models many important problems in engineering, management and economics [20], [37].

When $X$ is of the structure (1.6), problem (1.2) is equivalent to
\begin{equation}
q(x) - \text{mid}(l, u, q(x) - p(x)) = 0.
\end{equation}
Here mid($\cdot$) is the median operator, i.e., for three vectors $a, b, c \in \{ \mathbb{R} \cup \{ \pm \infty \} \}^n$ and $a \leq b$,
\begin{equation}
\text{mid}(a, b, c)_i = \text{mid}(a_i, b_i, c_i) = \begin{cases} 
a_i & \text{if } c_i < a_i, 
b_i & \text{if } b_i < c_i, 
c_i & \text{if } a_i \leq c_i \leq b_i, 
\end{cases}
i = 1, \ldots, n.
\end{equation}
If $q(x) = x$, (1.7) reduces to
\begin{equation}
x - \text{mid}(l, u, x - p(x)) = 0.
\end{equation}
Since the median operator is piecewise smooth, (1.7) and (1.8) are systems of nonsmooth equations.

A considerable number of generalizations of Newton-type methods [15], [21], [23], [29], [35], [36], [44], [47] have been developed for solving nonsmooth equation
\begin{equation}
F(x) = 0,
\end{equation}
where $F : \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz continuous but not differentiable. Some of these methods solve a linear complementarity problem or a linear variational inequality problem at each step. A natural extension of the classical Newton method for solving (1.9) is
\begin{equation}
x^{k+1} = x^k - V_k^{-1} F(x^k),
\end{equation}
where $V_k$ is an $n \times n$ matrix in a generalized Jacobian of $F$ at $x^k$. For this method, only a system of linear equations needs to be solved at each step. There are several possible definitions of generalized Jacobians. We will discuss them in the next section. It was proved in [40], [43] for the generalized Jacobians in the sense of [7] and [40] that the sequence generated by (1.10) superlinearly (quadratically) converges in a neighbourhood of a solution $x^*$ of (1.9) if $F$ is (strongly) semismooth at $x^*$ and all matrices in the generalized Jacobian of $F$ at $x^*$ are nonsingular. Since
most common nonsmooth functions, such as convex functions, piecewise smooth
functions, the Burmeister-Fischer function which is useful for nonsmooth equations
arising from the nonlinear complementarity problem [16], and their compositions
are semismooth functions, this extends the superlinear convergence theory of the
classical Newton method to the nonsmooth case. The function
\[
F(x) = q(x) - \text{mid}(l, u, q(x) - p(x))
\]
is (strongly) semismooth if \(p \) and \(q \) are (twice) smooth. Globally and superlinearly
convergent methods for solving (1.9) can be constructed by combining this theory
with some global convergence techniques. Applications of this theory to the nonlin-
erar constraints and variational inequality problems can be found in [8], [11], [12],
[13], [14], [25], [27], [30], [31], [33], [34], [38], [48], [51]. For a general
survey on this development, see [26].
Another approach for solving (1.9) is the smoothing method [2], [3], [4], [18],
[28]. The feature of smoothing methods is to construct a smoothing approximation
functions \(f: \mathbb{R}^n \times \mathbb{R}_+^+ \rightarrow \mathbb{R}^n \) of \(F \) such that for any \(\varepsilon > 0\), \(f(\cdot, \varepsilon) \) is continuously
differentiable and
\[
\|F(x) - f(x, \varepsilon)\| \rightarrow 0 \quad \text{as} \quad \varepsilon \to 0^+, \quad \text{for all} \quad x \in \mathbb{R}^n,
\]
and then to find a solution of (1.9) by (inexactly) solving the following problems
for a given positive sequence \(\{\varepsilon_k\}, k = 0, 1, 2, \ldots,\)
\[
f(x, \varepsilon_k) = 0.
\]
In [4], Chen and Mangasarian introduced a class of smoothing approximation func-
tions for nonlinear complementarity problems. Gabriel and Moré [18] extended
Chen-Mangasarian’s smoothing approach to box constrained variational inequali-
ties (1.8). Another class of smoothing approximation functions for general com-
plementarity problems (1.4) was given in [5]. Problems (1.8) and (1.4) are special
cases of general box constrained variational inequalities GVI \((l, u, p, q)\). The mer-
its of the smoothing method are global convergence and convenience in handling
smooth functions instead of nonsmooth functions. However, (1.12), which needs to
be solved at each step, is nonlinear in general.
The smoothing Newton method can be regarded as a variant of the smoothing
method. It uses the derivative of \(f \) with respect to the first variable in the Newton
method, namely
\[
x^{k+1} = x^k - t_k f_x(x^k, \varepsilon_k)^{-1} F(x^k),
\]
where \(\varepsilon_k > 0, f_x(x^k, \varepsilon_k) \) denotes the derivative of \(f \) with respect to the first variable
at \((x^k, \varepsilon_k)\) and \(t_k > 0\) is the steplsize. The smoothing Newton method (1.13) for
solving nonsmooth equation (1.9) has been studied for decades in different areas [1],
[5], [6], [22], [32], [42], [46], [50]. In some previous papers, method (1.13) is called a
splitting method because \(F(\cdot) \) is split into a smooth part \(f(\cdot, \varepsilon)\) and a nonsmooth
part \(F(\cdot) - f(\cdot, \varepsilon)\). The global and linear convergence of (1.13) has been discussed in
[42], but so far no superlinear convergence result has been obtained. In this paper
we will address this problem by investigating the relation between the derivative
\(f_x(x, \varepsilon)\) and the generalized Jacobian of \(F \) at \(x\). We define a Jacobian consistency
property and show that the smoothing approximation functions in [4], [5], [18],
[42] have this property. Under mild conditions, we prove that the sequence \(\{x^k\}\)
generated by the smoothing Newton method is bounded and each accumulation
point is a solution of (1.9). Furthermore, the convergence rate is superlinear if \(F \) is

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use
semismooth at the solution and the smoothing approximation function satisfies this
Jacobian consistency property. Moreover, the convergence rate is quadratic if \( F \) is
strongly semismooth at the solution. In particular, for box constrained variational
inequalities if \( p \) is a uniform \( P \)-function, then the smoothing Newton method has
three advantages:

- Solving a linear system of equations at each step;
- Guaranteeing that \( \{x^k\} \) is bounded and converges to the unique solution;
- Having superlinear convergence rate.

There are a wide variety of algorithms for the solution of variational inequalities
with box constraints \([4], [9], [13], [18], [27], [45], [48]\). As said before, some of them
have nonlinear subproblems. The algorithm proposed in \([8]\) for nonlinear comple-
mentarity problem (1.5), which is a special case of the box constrained variational
inequality problem, has the above three properties. The algorithms proposed in
\([27], [48]\) based on a differentiable merit function (for a survey on merit functions,
see \([17]\)) for the box constrained variational inequality problem have the above
three properties if \( p \) is a strongly monotone function, which is a stronger condition
than that of a uniform \( P \)-function. Hence, as an application, we present a method
for solving the general box constrained variational inequality problem with better
convergence properties.

This paper is organized as follows. In section 2, we define the Jacobian consistency
property. In section 3, we present the smoothing Newton method in detail and
prove that the method is globally and superlinearly convergent. In section 4,
we discuss the application of the smoothing Newton method to GVI(\( l, u, p, q \))
and verify various assumptions. In section 5, we give some final remarks and point out
the possible availability of the smoothing Newton method to the order complemen-
tarity problem and the variational inequality problem (1.3).

We let \( \| \cdot \| \) denote the Euclidean norm of \( \mathbb{R}^n \) and let
\[
\mathbb{R}_+ = \{ \varepsilon \mid \varepsilon \geq 0, \varepsilon \in \mathbb{R} \}
\]
and
\[
\mathbb{R}_{++} = \{ \varepsilon \mid \varepsilon > 0, \varepsilon \in \mathbb{R} \}.
\]
We denote the set of all nonnegative integers by \( N = \{0, 1, \ldots\} \).

2. JACOBIAN CONSISTENCY PROPERTY

Let \( H : \mathbb{R}^n \to \mathbb{R}^m \) be locally Lipschitz continuous. According to Rademacher’s
theorem, \( H \) is differentiable almost everywhere. Let \( D_H \) be the set where \( H \) is
differentiable. There are several definitions of generalized Jacobians of \( H \), which
can be used in the generalized Newton method. The B-differential of \( H \) \([40]\) is
defined by
\[
\partial_B H(x) = \left\{ \lim_{x^k \to x \atop x^k \in D_H} H'(x^k) \right\}.
\]
The generalized Jacobian of \( H \) at \( x \) in the sense of Clarke \([7]\) is
\[
\partial H(x) = \text{conv} \partial_B H(x).
\]
The superlinear convergence of (1.10) is established for these two kinds of general-
ized Jacobians in \([40], [43]\). Some other variants of Jacobians and their perturba-
tions are used in the literature \([48], [49], [51]\). A general range of different kinds
of generalized Jacobians, which are associated with superlinear convergence of (1.10),
is discussed in [41]. In this paper, for the function $F$, we use a kind of generalized Jacobian, denoted by $\partial C F$ and defined as

$$\partial C F(x) = \partial F_1(x) \times \partial F_2(x) \times \ldots \times \partial F_n(x).$$

This definition can be seen as a special case of the C-differential operator discussed in [41] and is more suitable to the discussion in this paper.

We are now able to define the Jacobian consistency property.

**Definition 2.1.** Let $F$ be a Lipschitz continuous function in $\mathbb{R}^n$. We call $f : \mathbb{R}^n \times \mathbb{R}_{++} \to \mathbb{R}^n$ a smoothing approximation function of $F$ if $f$ is continuously differentiable with respect to the first variable and there is a constant $\mu > 0$ such that for any $x \in \mathbb{R}^n$ and $\varepsilon \in \mathbb{R}_{++}$,

$$\|f(x, \varepsilon) - F(x)\| \leq \mu \varepsilon. \quad (2.1)$$

Furthermore, if for any $x \in \mathbb{R}^n$,

$$\lim_{\varepsilon \downarrow 0} \text{dist}((\nabla_x f(x, \varepsilon))^T, \partial C F(x)) = 0, \quad (2.2)$$

then we say $f$ satisfies the Jacobian consistency property.

**Remark 2.1.** In condition (2.1), $\mu \varepsilon$ may be replaced by any nondecreasing function $\sigma : \mathbb{R}_{++} \to \mathbb{R}_{++}$ such that $\sigma(0) = 0$ and $\sigma(t) > 0$ for $t > 0$. In this paper, however, we will restrict our discussion to (2.1), because it makes the analysis significantly simple.

For simplicity, in the remainder of this paper we denote

$$f_x(x, \varepsilon) \equiv (\nabla_x f(x, \varepsilon))^T.$$

It was proved in [42] that for any continuous function $F$ by using convolution we can construct a smoothing approximation function $f$ of $F$. We now investigate the cases in which $f$ has the Jacobian consistency property.

Chen and Mangasarian [4] introduced a class of smoothing functions for the nonlinear complementarity problem (1.5). Gabriel and More [18] extended Chen-Mangasarian’s smoothing approach to the box constrained variational inequality problem (1.8). The result in [18] may be easily generalized to the function $F$ defined in (1.11). Let $\rho : \mathbb{R} \to \mathbb{R}_+$ be a density function with a bounded absolute mean, that is

$$\kappa := \int_{-\infty}^{\infty} |s| \rho(s)ds < \infty. \quad (2.3)$$

Define the smoothing approximation $h(x, \varepsilon) = (h_i(x, \varepsilon))$ to the mid function in (1.11) by

$$h_i(x, \varepsilon) = \int_{-\infty}^{\infty} \text{mid}(l_i, u_i, q_i(x) - p_i(x) - \varepsilon s) \rho(s)ds.$$

Let

$$f(x, \varepsilon) = q(x) - h(x, \varepsilon). \quad (2.4)$$

Following Lemma 2.3 and Theorem 3.3 in [18], we can show that $f_i$ is continuously differentiable with respect to $x$ and satisfies (2.1) with $\mu = \sqrt{n} \kappa$. Hence $f$ is a smoothing approximation function of $F$. 

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use
Hence for any fixed \(2.6\)
\[
\|Suppose that \(x\) follows Lemma 2.3 in [18] again, we have \(2.7\)
\[
Based on the Gabriel-Moré function, such\(2.8\)
\[
Jacobian consistency property.
\]
\[
Lim it (2.2) implies that for any \(f(x)\)
\[
\text{dist}(f(x, \varepsilon), \partial_C F(x)) \leq \delta.
\]
Based on the Gabriel-Moré function, such \(2.9\) for GVI\(l, u, p, q\) can be chosen as follows.

Let
\[
\gamma(x) = \min_{1 \leq i < j \leq n} \{|q_i(x) - p_i(x) - l_i|, |q_j(x) - p_j(x) - u_j| : q_i(x) - p_i(x) \neq l_i, q_j(x) - p_j(x) \neq u_j\}.
\]
(2.7)
\[
\text{Since } \int_{-\infty}^{\infty} s^{-1} ds = \infty, (2.3) implies that the density function } \rho \text{ satisfies}
\[
\lim_{s \to \infty} |s|^2 \rho(s) = 0.
\]
Thus we may choose a positive constant \(\tau \leq 1\) such that for any \(\nu \) with \(|\nu| \in (0, \tau]\),
\[
\rho(\frac{1}{|\nu|}) < \nu^2.
\]
This implies that for any \(\varepsilon \in (0, \tau]\)
\[
2.8\)
\[
\{ \int_{-\infty}^{-1/\varepsilon} \rho(s) ds, \int_{1/\varepsilon}^{\infty} \rho(s) ds \} \leq \int_{1/\varepsilon}^{\infty} \frac{1}{s^2} ds = \varepsilon.
\]
If \(||q'(x) - p'(x)|| = 0\), then by (2.6) from any \(\varepsilon > 0, f(x, \varepsilon) = q'(x) \in \partial_C F(x)\).
Suppose that \(||q'(x) - p'(x)|| \neq 0\). Let
\[
\varepsilon(x, \delta) = \min \{ \gamma(x), \frac{\gamma(x) \delta}{2\sqrt{n} ||q'(x) - p'(x)||} \}.
\]
If \( q_i(x) - p_i(x) \not\in [l_i, u_i] \), then \( F_i \) is differentiable at \( x \) and \( F_i'(x) = q_i'(x) \). Hence in this case for any \( \varepsilon \in (0, \varepsilon(x, \delta)] \),

\[
\|(f_x(x, \varepsilon) - F'(x))_i\| \\
= \left( \int_{(q_i(x)-p_i(x)-l_i)/\varepsilon}^{(q_i(x)-p_i(x)-u_i)/\varepsilon} \rho(s)ds \right) \|q_i'(x) - p_i'(x)\| \\
\leq \left( \int_{\gamma(x)/\varepsilon}^{\infty} \frac{1}{s^2}ds \right) \|q_i'(x) - p_i'(x)\| \\
\leq \frac{\varepsilon}{\gamma(x)} \|q_i'(x) - p_i'(x)\| \\
\leq \frac{\delta}{2\sqrt{n}},
\]

where the first inequality follows from (2.8) and \( \tau \geq \varepsilon/\gamma(x) \).

If \( q_i(x) - p_i(x) \in (l_i, u_i) \), then \( F_i \) is differentiable at \( x \) and \( F_i'(x) = p_i'(x) \). Hence in this case for any \( \varepsilon \in (0, \varepsilon(x, \delta)] \)

\[
\|(f_x(x, \varepsilon) - F'(x))_i\| \\
= \left( 1 - \int_{(q_i(x)-p_i(x)-l_i)/\varepsilon}^{(q_i(x)-p_i(x)-u_i)/\varepsilon} \rho(s)ds \right) \|q_i'(x) - p_i'(x)\| \\
= \left( \int_{-\infty}^{(q_i(x)-p_i(x)-u_i)/\varepsilon} \rho(s)ds + \int_{(q_i(x)-p_i(x)-l_i)/\varepsilon}^{\infty} \rho(s)ds \right) \|q_i'(x) - p_i'(x)\| \\
\leq \frac{2\varepsilon}{\gamma(x)} \|q_i'(x) - p_i'(x)\| \\
\leq \delta/\sqrt{n}.
\]

If \( q_i(x) - p_i(x) = l_i \) or \( q_i(x) - p_i(x) = u_i \), then

\[
(f_x(x, \varepsilon))_i = q_i'(x) - \lambda(x)(q_i(x) - p_i(x)),
\]

where \( \lambda(x) \in [0, 1] \). Thus in this case \( (f_x(x, \varepsilon))_i \in \partial F_i(x) \) for any \( \varepsilon > 0 \).

Hence, in any case, we have

\[
\text{dist}((f_x(x, \varepsilon))_i, \partial F_i(x)) \leq \delta/\sqrt{n}.
\]

Thus

\[
\text{dist}(f_x(x, \varepsilon), \partial C F(x)) \leq \delta.
\]

By choosing a special density function, for any fixed \( x \in \mathbb{R}^n, \) we even can ask that

\[
f_x(x, \varepsilon) \in \partial C F(x)
\]

for any \( \varepsilon > 0 \) sufficiently small. For instance, we consider the uniform density function

\[
\rho(s) = \begin{cases} \frac{1}{b-a} & \text{if } s \in [a, b], \\ 0 & \text{otherwise}, \end{cases}
\]

where \( a \) and \( b \) are two finite real numbers and \( a < b \). Notice that the uniform density function has

\[
\int_{-\infty}^{\varepsilon} \rho(s)ds = 0, \quad \int_{\varepsilon}^{\infty} \rho(s)ds = 0 \quad \text{and} \quad \int_{0}^{d} \rho(s)ds = 1
\]
for any $d \geq b$ and $c \leq a$. Then it is easy to verify that for any $\varepsilon \in (0, \frac{\gamma(x)}{\max|a|, |b|})$, 
\[
\int_{(q_i(x) - p_i(x) - l_i)/\varepsilon}^{(q_i(x) - p_i(x) - u_i)/\varepsilon} \rho(s)ds = \begin{cases} 
0 & \text{if } q_i(x) - p_i(x) \notin [l_i, u_i], \\
1 & \text{if } q_i(x) - p_i(x) \in (l_i, u_i).
\end{cases}
\]
By (2.9) and (2.10), we obtain (2.11).

To see a smoothing approximation function with a concrete density function, we consider
\[
\rho(s) = \begin{cases} 
1 & \text{if } -\frac{1}{2} \leq s \leq \frac{1}{2}, \\
0 & \text{otherwise}.
\end{cases}
\]
By a straightforward calculation we have that
\[
f_i(x, \varepsilon) = q_i(x) - \int_{-1/2}^{1/2} \mid d l_i, u_i, q_i(x) - p_i(x) - \varepsilon s \mid ds 
= \begin{cases} 
q_i(x) - \frac{1}{\varepsilon}(q_i(x) - p_i(x))(u_i - l_i) - \frac{1}{2}(l_i + u_i) - \frac{1}{\varepsilon^2}(l_i^2 - u_i^2) & \text{if } |q_i(x) - p_i(x) - l_i| \leq \frac{\varepsilon}{2}, |q_i(x) - p_i(x) - u_i| \leq \frac{\varepsilon}{2}, \\
\frac{1}{\varepsilon}(q_i(x) + p_i(x)) - \frac{1}{\varepsilon^2}(q_i(x) - p_i(x) - l_i)^2 - \frac{\varepsilon}{2} - \frac{\varepsilon}{2} & \text{if } |q_i(x) - p_i(x) - l_i| \leq \frac{\varepsilon}{2}, q_i(x) - p_i(x) - u_i < -\frac{\varepsilon}{2}, \\
\frac{1}{\varepsilon}(q_i(x) + p_i(x)) + \frac{1}{\varepsilon^2}(q_i(x) - p_i(x) - u_i)^2 - \frac{\varepsilon}{2} + \frac{\varepsilon}{2} & \text{if } |q_i(x) - p_i(x) - u_i| \leq \frac{\varepsilon}{2}, q_i(x) - p_i(x) - l_i > \frac{\varepsilon}{2}, \\
F_i(x) & \text{otherwise}.
\end{cases}
\]
We can simplify the definition of $f(x, \varepsilon)$ for special $\varepsilon$. For example, let
\[
\varepsilon = \min_{1 \leq i \leq n} \{u_i - l_i\}.
\]
For $i = 1, 2, ..., n$ and $\varepsilon \in (0, \bar{\varepsilon}]$, the smoothing approximation function (2.4) with density function (2.12) reduces to
\[
f_i(x, \varepsilon) = \begin{cases} 
\frac{1}{2}(p_i(x) + q_i(x)) + \frac{1}{\varepsilon^2}(p_i(x) - q_i(x) + u_i)^2 + \frac{\varepsilon}{2} - \frac{\varepsilon}{2} & \text{if } |q_i(x) - p_i(x) - u_i| \leq \frac{\varepsilon}{2}, \\
\frac{1}{2}(p_i(x) + q_i(x)) - \frac{1}{\varepsilon^2}(p_i(x) - q_i(x) + l_i)^2 - \frac{\varepsilon}{2} - \frac{\varepsilon}{2} & \text{if } |q_i(x) - p_i(x) - u_i| \leq \frac{\varepsilon}{2}, q_i(x) - p_i(x) - l_i > \frac{\varepsilon}{2}, \\
F_i(x) & \text{otherwise}.
\end{cases}
\]
Let
\[
\varepsilon(x) = \min\{\bar{\varepsilon}, \gamma(x)\},
\]
where $\bar{\varepsilon}$ is defined by (2.13) and $\gamma(x)$ is defined by (2.7). Then for any $\varepsilon \in (0, \varepsilon(x)]$, 
\[
(f_x(x, \varepsilon))_i = \begin{cases} 
\frac{1}{2}(p_i(x) + q_i(x)) & \text{if } q_i(x) - p_i(x) = u_i \text{ or } q_i(x) - p_i(x) = l_i, \\
F_i(x) & \text{otherwise}.
\end{cases}
\]
By (2.5), this implies that for any $\varepsilon \in (0, \varepsilon(x)]$,
\[
\text{dist}(f_x(x, \varepsilon), \partial_C F(x)) = 0.
\]
For the general complementarity problem GCP($p, q$), $f(x, \varepsilon)$ defined by (2.14) reduces to
\[
f_i(x, \varepsilon) = \begin{cases} 
\frac{1}{2}(p_i(x) + q_i(x)) - \frac{1}{\varepsilon^2}(p_i(x) - q_i(x))^2 - \frac{\varepsilon}{2} & \text{if } |q_i(x) - p_i(x)| \leq \frac{\varepsilon}{2}, \\
F_i(x) & \text{otherwise}.
\end{cases}
\]
The function defined by (2.15) has been studied in [5], [42].
Some smoothing approximation functions are not derivable from the integration of probability densities, for example, the Auto-scaling interior point smooth function [3], but also satisfy the Jacobian consistency property.

3. A SMOOTHING NEWTON METHOD

In this section we assume that \( f \) has the Jacobian consistency property, and present a smoothing Newton method with a line search based on \( f \). We prove that this method converges globally and superlinearly.

We denote

\[
\Theta(x) = \frac{1}{2} \| F(x) \|^2
\]

and

\[
\theta_k(x) = \frac{1}{2} \| f(x, \varepsilon_k) \|^2.
\]

**Algorithm 3.1.** Given \( \rho, \alpha, \eta \in (0, 1), \gamma \in (0, +\infty) \) and a starting point \( x^0 \in \mathbb{R}^n \). Choose \( \sigma \in (0, 1/2(1 - \alpha)) \), and \( \mu > 0 \) satisfying (2.1).

**Initial step.** Let \( \beta_0 = \| F(x^0) \| \) and \( \varepsilon_0 = \frac{\alpha}{2\mu} \beta_0 \). For \( k \geq 0 \),

1. Solve

\[ F(x^k) + f_x(x^k, \varepsilon_k) d^k = 0. \]

Let \( d^k \) be the solution of (3.1).

2. Let \( m_k \) be the smallest nonnegative integer \( m \) such that

\[ \theta_k(x^k + \rho^m d^k) - \theta_k(x^k) \leq -2\sigma \rho^m \Theta(x^k). \]

Set \( t_k = \rho^m \) and \( x^{k+1} = x^k + t_k d^k \).

3.1 If \( \| F(x^{k+1}) \| = 0 \), terminate.

3.2 If \( \| F(x^{k+1}) \| > 0 \) and

\[ \| F(x^{k+1}) \| \leq \max\{ \eta \beta_k, \alpha^{-1} \| F(x^{k+1}) - f(x^{k+1}, \varepsilon_k) \| \}, \]

we let

\[ \beta_{k+1} = \| F(x^{k+1}) \| \]

and choose an \( \varepsilon_{k+1} \) satisfying

\[ 0 < \varepsilon_{k+1} \leq \min\{ \frac{\alpha}{2\mu} \beta_{k+1}, \frac{\varepsilon_k}{2} \} \]

and

\[ \text{dist}(f_x(x^{k+1}, \varepsilon_{k+1}), \partial C F(x^{k+1})) \leq \gamma \beta_{k+1}. \]

3.3 If \( \| F(x^{k+1}) \| > 0 \) but (3.3) does not hold, we let \( \beta_{k+1} = \beta_k \) and \( \varepsilon_{k+1} = \varepsilon_k \).

Without loss of generality, we assume that \( \| F(x^k) \| \neq 0 \) for all \( k \) in the following convergence analysis.

**Remark 3.1.** Condition (3.5) is the crucial condition for superlinear convergence of Algorithm 3.1. To guarantee global convergence, one only needs to choose a smoothing approximation function \( f \) and ignore (3.5). If \( f \) has the Jacobian consistency property, we can find an \( \varepsilon_{k+1} > 0 \) such that (3.4) and (3.5) hold by Definition 2.1. Moreover, we have shown, in section 2, how to choose an \( \varepsilon_{k+1} \) satisfying (3.4)
Suppose that Lemma 3.1. and (3.7) may not hold for those \( K \) such that (3.3) fail to hold. Notice that (3.7) holds.

\[ \|F(x^k) - f(x^k, \varepsilon_k)\| \leq \alpha \|F(x^k)\| \quad \text{for } k \geq 0 \]

and
\[ \text{dist}(f_x(x^k, \varepsilon_k), \partial_C F(x^k)) \leq \gamma \|F(x^k)\| \quad \text{for } k \in K, \ k \geq 1, \]

where
\[ K = \{0\} \cup \{k \mid \|F(x^k)\| \leq \max\{\eta \beta_{k-1}, \alpha^{-1} \|f(x^k, \varepsilon_{k-1}) - F(x^k)\|\}, \ k \in N\}. \]

Notice that (3.7) may not hold for those \( k \) such that (3.3) fail to hold.

**Lemma 3.1.** Suppose that \( f_x(x^k, \varepsilon_k) \) is nonsingular. Then there exists a finite nonnegative integer \( m_k \) such that (3.2) holds.

**Proof.** The continuous differentiability of \( f(\cdot, \varepsilon_k) \) implies that \( \theta_k \) is continuously differentiable and \( \theta_k(x^k) = f(x^k, \varepsilon_k)^T f_x(x^k, \varepsilon_k) \). By the construction of Algorithm 3.1, \( f_x(x^k, \varepsilon_k) d^k = -F(x^k) \). Then, from (3.6) we have
\[
\begin{align*}
\theta_k(x^k + td^k) - \theta_k(x^k) &= t \theta_k'(x^k) d^k + o(t) \\
&= -tF(x^k)^T f(x^k, \varepsilon_k) + o(t) \\
&= -2t \Theta(x^k) + tF(x^k)^T (F(x^k) - f(x^k, \varepsilon_k)) + o(t) \\
&\leq -2t \Theta(x^k) + 2t \Theta(x^k) + o(t) \\
&= -2t(1 - \alpha) \Theta(x^k) + o(t).
\end{align*}
\]

Since \( \sigma < \frac{1}{2}(1 - \alpha) < 1 - \alpha \), there exists a finite nonnegative integer \( m_k \) such that (3.2) holds.

**Assumption 1.** The level set
\[ D_0 = \{x \in \mathbb{R}^n : \Theta(x) \leq (1 + \alpha)^2 \Theta(x^0)\} \]
is bounded.

**Assumption 2.** For any \( \varepsilon \in \mathbb{R}_+ \) and \( x \in D_0 \), \( f_x(x, \varepsilon) \) is nonsingular.

**Theorem 3.1.** Suppose that Assumptions 1 and 2 hold. Then Algorithm 3.1 is well defined and the generated sequence \( \{x^k\} \) remains in \( D_0 \) and satisfies
\[ \lim_{k \to 0} F(x^k) = 0. \]

**Proof.** Let us denote
\[ K_1 = \{k \in K \mid \eta \beta_{k-1} \geq \alpha^{-1} \|f(x^k, \varepsilon_{k-1}) - F(x^k)\|\} \]
and
\[ K_2 = \{k \in K \mid \eta \beta_{k-1} < \alpha^{-1} \|f(x^k, \varepsilon_{k-1}) - F(x^k)\|\}. \]
Then \( K_1 \cup K_2 \cup \{0\} = K \), which is defined in (3.8). Assume that \( K \) consists of \( k_0 < k_1 < k_2 < \ldots \). Let \( k \) be an arbitrary nonnegative integer. Let \( k_j \) be the largest number in \( K \) such that \( k_j \leq k \). Then
\[ \varepsilon_k = \varepsilon_{k_j} \quad \text{and} \quad \beta_k = \beta_{k_j}. \]
Notice that \( f \) is a smoothing approximation function. By the line search rule (3.2),
\[
\|f(x^k, \varepsilon_{k_j})\| \leq \|f(x^{k_j}, \varepsilon_{k_j})\|.
\]
Then by (2.1), for \( j \geq 0 \),
\[
\|F(x^k)\| \leq \|f(x^k, \varepsilon_k)\| + \|F(x^k) - f(x^k, \varepsilon_k)\|
\leq \|f(x^k, \varepsilon_k)\| + \|F(x^k) - f(x^{k_j}, \varepsilon_{k_j})\|
\]
(3.10)
\[
\leq \|f(x^{k_j}, \varepsilon_{k_j})\| + \mu \varepsilon_{k_j}
\leq \|F(x^{k_j})\| + \mu \varepsilon_{k_j} + \mu \varepsilon_{k_j}
= \beta_{k_j} + 2\mu \varepsilon_{k_j}.
\]
If \( j = 0, \beta_{k_j} = \beta_0, \varepsilon_{k_j} = \varepsilon_0 \) and
\[
\|F(x^k)\| \leq \beta_0 + 2\mu \varepsilon_0 \leq (1 + \alpha)\|F(x^0)\|.
\]
If \( j \geq 1 \), by step 3 of Algorithm 3.1,
\[
\varepsilon_{k_j} \leq \frac{1}{2} \varepsilon_{k_{j-1}} = \frac{1}{2} \varepsilon_{k_{j-1}}
\]
and
\[
\beta_{k_j} \leq \eta \beta_{k_{j-1}} = \eta \beta_{k_{j-1}}, \quad \text{if} \ k_j \in K_1,
\]
or
\[
\beta_{k_j} \leq \frac{\alpha}{2} \|f(x^{k_j}, \varepsilon_{k_j-1}) - F(x^{k_j})\| \leq \frac{\mu}{\alpha} \varepsilon_{k_j-1} = \frac{\mu}{\alpha} \varepsilon_{k_j-1} \leq \frac{1}{2} \beta_{k_{j-1}}, \quad \text{if} \ k_j \in K_2.
\]
Let
\[
r = \max\{\frac{1}{2}, \eta\}.
\]
Then by the definitions of \( \varepsilon_0 \) and \( \beta_0 \), for \( j \geq 1 \),
\[
\varepsilon_{k_j} \leq \frac{1}{2(2j-1)} \varepsilon_0 = \frac{1}{2j} \frac{\alpha}{\mu} \|F(x^0)\|
\]
(3.11)
and
\[
\beta_{k_j} \leq r^{j-1} \beta_0 = r^{j-1} \|F(x^0)\|.
\]
(3.12)
Hence by (3.10), for \( j \geq 1 \)
\[
\|F(x^k)\| \leq (r^{j-1} + \frac{\alpha}{2j}) \|F(x^0)\|
\]
(3.13)
\[
\leq r^{j-1} (1 + \alpha) \|F(x^0)\|
\]
where the last inequality follows from the fact that \( \frac{1}{2} \leq r \).
Therefore in any case
\[
\|F(x^k)\| \leq (1 + \alpha) \|F(x^0)\|.
\]
This implies that the sequence \( \{x^k\} \) remains in the level set \( D_0 \).
Now we prove (3.9). If \( K \) is infinite, by (3.13),
\[
\lim_{k \to \infty} \|F(x^k)\| \leq \lim_{j \to \infty} r^{j-1} (1 + \alpha) \|F(x^0)\| = 0.
\]
Hence to prove (3.9), it suffices to prove that \( K \) is infinite. Suppose that \( K \) is finite. This means that both \( K_1 \) and \( K_2 \) are finite. Let \( \bar{k} \) be the largest number in \( K \). Then for all \( k > \bar{k}, \)
\[
\varepsilon_k = \varepsilon_{\bar{k}}, \quad \beta_k = \beta_{\bar{k}} = \|F(x^{\bar{k}})\|,
\]
(3.14)
(3.15) \[ \|F(x^k)\| > \eta \beta_k = \eta \|F(x^k)\| > 0 \]

and

(3.16) \[ \alpha \|F(x^k)\| > \|f(x^k, \varepsilon_k) - F(x^k)\|. \]

By (3.15), for all \( k > \hat{k} \),

(3.17) \[ \Theta(x^k) \geq \eta^2 \Theta(x^k). \]

Let

\[ \hat{\varepsilon} = \varepsilon_k, \]

\[ \hat{f}(x) = f(x, \hat{\varepsilon}) \]

and

\[ \hat{\theta}(x) = \frac{1}{2} \|\hat{f}(x)\|^2. \]

Notice that for all \( k > \hat{k} \),

\[ f(x^k, \varepsilon_k) = \hat{f}(x^k) \text{ and } \theta_k(x^k) = \hat{\theta}(x^k). \]

By Assumptions 1 and 2 there is an \( M > 0 \) such that for all \( x \in D_0, \|f_x(x, \hat{\varepsilon})^{-1}\| \leq M \). Then for all \( k > \hat{k} \),

\[
\|d^k\| = \|f_x(x^k, \hat{\varepsilon})^{-1}F(x^k)\| \\
\leq M \|F(x^k)\| \\
\leq M (1 + \alpha) \|F(x^k)\| \\
=: L.
\]

If \( \inf_k t_k = t^* > 0 \), then from (3.17) and the line search rule (3.2), for all \( k \geq 0 \),

\[ \hat{\theta}(x^{k+1}) - \hat{\theta}(x^k) \leq -2\sigma t_k \Theta(x^k) \leq -2\sigma t^* \eta^2 \Theta(x^k) < 0. \]

This, together with the monotonicity of \( \{\hat{\theta}(x^k)\}_{k \geq \hat{k}} \), implies that \( \hat{\theta}(x^k) \to -\infty \) as \( k \to \infty \). This contradicts the fact that \( \hat{\theta}(x^k) \geq 0 \) for all \( k \geq 0 \). Hence \( K \) cannot be finite. Thus (3.9) holds.

Now we consider the case that \( \inf_k t_k = 0 \). Let \( K_0 \) be a subsequence of \( N \) such that \( \{t_k\}_{k \in K_0} \) converges to zero. Since \( \{x^k\} \) is bounded, without loss of generality, we assume that \( \{x^k\}_{k \in K_0} \) converges to \( x^* \).

By the line search rule (3.2) for all \( k \geq \hat{k} \),

(3.18) \[ -2\sigma \rho^{m_k-1} \Theta(x^k) < \hat{\theta}(x^k + \rho^{m_k-1}d^k) - \hat{\theta}(x^k). \]

Dividing both sides by \( \rho^{m_k-1} \), we obtain

\[
-2\sigma \Theta(x^k) < \frac{\hat{\theta}(x^k + \rho^{m_k-1}d^k) - \hat{\theta}(x^k)}{\rho^{m_k-1}} \\
= \hat{\theta}'(x^k)d^k + \int_0^1 (\hat{\theta}'(x^k + \lambda \rho^{m_k-1}d^k) - \hat{\theta}'(x^k))d\lambda.
\]

Notice that

\[
\hat{\theta}'(x^k)d^k = -F(x^k)^T \hat{f}(x^k) \\
= -2 \Theta(x^k) + F(x^k)(\hat{f}(x^k) - F(x^k)) \\
\leq -2 \Theta(x^k) + 2\alpha \Theta(x^k),
\]

where the last inequality follows from (3.16).
By the continuity of $\hat{\theta}'$, the boundedness of $\{d_k\}$ and $\lim_{k\to\infty} m_k = \infty$, we have

$$\lim_{k \to \infty} \int_0^1 ((\hat{\theta}'(x^k + \lambda m_k^{-1} d^k) - \hat{\theta}'(x^k)) d^k) d\lambda = 0.$$  

By taking the limit in (3.18) on the subsequence $k \in K_0$, we obtain

$$-2\sigma \Theta(x^*) \leq -2(1 - \alpha) \Theta(x^*) < 0.$$  

This implies $\sigma \geq 0$, which contradicts the fact that $\sigma < (1 - \alpha)/2$. Hence $K$ cannot be finite. Thus (3.9) holds.

To show the superlinear convergence rate, we give the following lemma.

**Lemma 3.2.** If there exists a scalar

$$\lambda \in \left[ \frac{1}{2} - \frac{(1 - \alpha - 2\sigma)^2}{2(2 + \alpha)^2}, \frac{1}{2} \right] \tag{3.19}$$

such that for some $k \in K$,

$$\Theta(y) - \Theta(x^k) \leq -2\lambda \Theta(x^k), \tag{3.20}$$

then it holds that

$$\theta_k(y) - \theta_k(x^k) \leq -2\sigma \Theta(x^k). \tag{3.21}$$

**Proof.** By the definition of $K$, we have

$$0 < \varepsilon_k \leq \frac{\alpha}{2\mu} \|F(x^k)\|, \quad k \in K.$$  

Hence, from (2.1), for any $y \in \mathbb{R}^n$, $k \in K$,

$$\|f(y, \varepsilon_k)\| \leq \|F(y)\| + \frac{\alpha}{2} \|F(x^k)\|$$

and

$$\|f(x^k, \varepsilon_k)\| \geq \|F(x^k)\| - \frac{\alpha}{2} \|F(x^k)\|.$$  

Using these two inequalities and (3.20), we obtain

$$\theta_k(y) - \theta_k(x^k) = \frac{1}{2} \|f(y, \varepsilon_k)\|^2 - \frac{1}{2} \|f(x^k, \varepsilon_k)\|^2 \leq \frac{1}{2} (\|F(y)\| + \frac{\alpha}{2} \|F(x^k)\|)^2 - \frac{1}{2} (1 - \alpha)^2 \|F(x^k)\|^2 \leq \Theta(y) + \frac{1}{2} \alpha \|F(y)\| \|F(x^k)\| + \frac{\alpha^2}{4} \Theta(x^k) - (1 - \alpha + \frac{\alpha^2}{4}) \Theta(x^k) \leq \Theta(y) + \frac{1}{2} \alpha \|F(y)\| \|F(x^k)\| - (1 - \alpha) \Theta(x^k) \leq \Theta(y) + \alpha \varepsilon_k - (1 - \alpha) \Theta(x^k) \leq \Theta(y) - \Theta(x^k) + \alpha(1 + \sqrt{1 - 2\lambda}) \Theta(x^k) \leq \frac{1}{2} \alpha(1 + \sqrt{1 - 2\lambda}),$$

where the second and last inequalities follow from (3.20).

Let us denote

$$\phi(\lambda) = \lambda - \frac{1}{2} \alpha(1 + \sqrt{1 - 2\lambda}).$$
To prove (3.21), it suffices to show
\begin{equation}
\phi(\lambda) \geq \sigma \quad \text{for} \quad \lambda \in \left[ 1, \frac{(1 - \alpha - 2\sigma)^2}{2(2 + \alpha)^2} \right].
\end{equation}
Since $0 < (1 - \alpha - 2\sigma)/(2 + \alpha) < 1$,
\begin{equation}
\frac{(1 - \alpha - 2\sigma)^2}{(2 + \alpha)^2} \leq 1 - \alpha - 2\sigma.
\end{equation}
Notice that $\phi$ is monotone increasing in $[0, \frac{1}{2}]$. We only need to show (3.22) at $\lambda = \frac{1}{2} - \frac{(1 - \alpha - 2\sigma)^2}{2(2 + \alpha)^2}$. By the definition of $\phi$ and (3.23),
\begin{align*}
\phi(\lambda) &= \frac{1}{2} - \frac{(1 - \alpha - 2\sigma)^2}{2(2 + \alpha)^2} - \frac{1}{2} \alpha \left( 1 + \frac{1 - \alpha - 2\sigma}{2 + \alpha} \right) \\
&\geq \frac{1}{2} - \frac{1 - \alpha - 2\sigma}{2 + \alpha} - \frac{\alpha}{2} \left( 1 + \frac{1 - \alpha - 2\sigma}{2 + \alpha} \right) \\
&= \frac{1}{2} - \frac{\alpha}{2} \\
&\geq \sigma.
\end{align*}
This completes the proof.

\begin{theorem}
Suppose that Assumptions 1 and 2 hold. Suppose that for an accumulation point $x^*$ of the sequence $\{x^k\}$, all $V \in \partial_C F(x^*)$ are nonsingular and that $F$ is semismooth at $x^*$. Then $x^*$ is a solution of $F(x) = 0$ and the sequence $\{x^k\}$ generated by Algorithm 3.1 converges to $x^*$ superlinearly. Moreover, if $F$ is strongly semismooth at $x^*$, then $\{x^k\}$ converges to $x^*$ quadratically.
\end{theorem}

\begin{proof}
By Theorem 3.1, $x^*$ is a solution of $F(x) = 0$. Notice that $\partial B F(x^*) \subseteq \partial C F(x^*)$. By Proposition 2.5 in [40] there is a neighbourhood of $x^*$ such that $x^*$ is the unique solution in this neighbourhood.

By Theorem 3.1, the set $K$ defined by (3.8) is infinite, and there is a subsequence $K_0$ of $K$ such that $\{x^k\}_{k \in K_0}$ converges to $x^*$. Now we consider the convergence behaviour of the subsequence $\{x^k\}_{k \in K_0}$.

Notice that for any $x \in \mathbb{R}^n$, $\partial_C F(x)$ is a compact set. Let $V_k \in \partial_C F(x^k)$ be such that
\begin{equation}
\text{dist}(f_k(x^k, \varepsilon_k), \partial_C F(x^k)) = \|f_k(x^k, \varepsilon_k) - V_k\|.
\end{equation}
By construction of Algorithm 3.1,
\begin{equation}
\|f_k(x^k, \varepsilon_k) - V_k\| \leq \gamma / \beta_k, \quad k \in K_0.
\end{equation}
By Theorem 3.1, $\beta_k \to 0$ as $k \to \infty$. This, together with the compactness of $\partial_C F(x^*)$, the nonsingularity of all $V \in \partial_C F(x^*)$ and the upper semicontinuity of $\partial_C F(\cdot)$ at $x^*$, implies that there exist $M > 0$ and $\hat{k} \geq 0$ such that for all $k \geq \hat{k}$ and $k \in K_0$, $\|f_k(x^k, \varepsilon_k)^{-1}\| \leq M$. Therefore, by the construction of Algorithm 3.1, for all $k \geq \hat{k}$ and $k \in K_0$,
\begin{align}
\|x^k + d_k - x^*\| &= \|x^k - x^* - f_k(x^k, \varepsilon_k)^{-1} F(x^k)\| \\
&= \|f_k(x^k, \varepsilon_k)^{-1}(f_k(x^k, \varepsilon_k)(x^k - x^*) - F(x^k) + F(x^*))\| \\
&\leq \|f_k(x^k, \varepsilon_k)^{-1}\| (\|f_k(x^k, \varepsilon_k) - V_k\|(x^k - x^*)\| + \|V_k(x^k - x^*) - F(x^k) + F(x^*)\|) \\
&\leq M (\gamma / \beta_k) \|x^k - x^*\| + \|V_k(x^k - x^*) - F(x^k) + F(x^*)\|.
\end{align}
Since $F$ is semismooth at $x^*$ if and only if each $F_i$ is semismooth at $x^*$ [43], by Theorem 3.2 in [43],

$$
\|V_k(x^k - x^*) - F(x^k) + F(x^*)\|
\leq \sqrt{\sum_{i=1}^{n} \|V^i_k(x^k - x^*) - F_i(x^k) + F_i(x^*)\|^2}
= o(\|x^k - x^*\|) \quad \text{as } k \to \infty, \quad k \in K_0,
$$

(3.25)

where $V^i_k$ denotes the $i$th row of $V_k$. Hence

$$
|\|x^k + d^k - x^*\| = o(\|x^k - x^*\|) \quad \text{as } k \to \infty, \quad k \in K_0.
$$

Furthermore, following the proof of Theorem 3.1 in [40]

$$
F(x^k + d^k) = o(F(x^k)) \quad \text{as } k \to \infty, \quad k \in K_0.
$$

(3.27)

Let $\lambda = \max \left\{ \frac{1}{2} - \frac{(1-\alpha-2\sigma)^2}{2(2+\alpha)^2}, \frac{1-\eta}{2} \right\}$. Then (3.27) implies that there is $\hat{k} \geq \bar{k}$ such that $\hat{k} \in K_0$ and for any $k \geq \hat{k}$ and $k \in K_0$,

$$
\Theta(x^k + d^k) - \Theta(x^k) \leq -2\lambda \Theta(x^k).
$$

(3.28)

By Lemma 3.2, for any $k \geq \hat{k}$ and $k \in K_0$,

$$
\theta_k(x^k + d^k) - \theta_k(x^k) \leq -2\sigma \Theta(x^k),
$$

that is, $t_k \equiv 1$ and $x^{k+1} = x^k + d^k$ for all $k \geq \hat{k}$ and $k \in K_0$. In particular, $x^{k+1} = x^k + d^k$ and from (3.28),

$$
\|F(x^{k+1})\| \leq \sqrt{1 - 2\lambda \|F(x^k)\|} \leq \eta \|F(x^k)\| = \eta \beta_k,
$$

which implies that $x^{k+1} \in K_0$. Repeating the above process we may prove that for all $k \geq \hat{k}$,

$$
k \in K_0
$$

and

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

Then by using (3.26) we have proved that $\{x^k\}$ converges to $x^*$ superlinearly.

If $F$ is strongly semismooth at $x^*$, then each $F_i$ is also strongly semismooth at $x^*$. By (3.25) and Lemma 2.3 in [40],

$$
\|V_k(x^k - x^*) - F(x^k) + F(x^*)\| = O(\|x^k - x^*\|^2).
$$

By the Lipschitz continuity of $F$, for all $k \in K_0$,

$$
\beta_k = \|F(x^k)\| = O(\|x^k - x^*\|).
$$

Hence the quadratic convergence follows easily from (3.24) and the above proof. □

4. Application

In this section we discuss an application of Algorithm 3.1 to general box constrained variational inequalities GVI($l, u, p, q$). Conditions used in the convergence analysis of Algorithm 3.1 are Assumptions 1 and 2, and the condition that $F$ is semismooth at a solution point $x^*$ and all $V \in \partial_{C}F(x^*)$ are nonsingular. Now we consider when these conditions hold for the smoothing approximation function $f$ defined by (2.4) and the nonsmooth function $F$ defined by (1.11).

Let $A$ be an $n \times n$ matrix. $A$ is called a $P_0$-matrix, if its principal minors are all nonnegative, and $A$ is called a $P$-matrix, if its principal minors are all positive.
We call \( p : \mathbb{R}^n \rightarrow \mathbb{R}^n \) a uniform \( P \)-function with respect to \( q \) if there is a positive constant \( \kappa \) such that

\[
\max_{1 \leq i \leq n} (q_i(x) - q_i(y))(p_i(x) - p_i(y)) \geq \kappa \|x - y\|^2.
\]

If \( q \) is the identity map and (4.1) holds, then we call \( p \) a uniform \( P \)-function directly.

The concepts of a \( P \)-matrix and a uniform \( P \)-function have been frequently used in complementarity and variational inequalities areas.

**Proposition 4.1.** Suppose that \( q \) is norm coercive, i.e., \( \|q(x)\| \rightarrow \infty \) if and only if \( \|x\| \rightarrow \infty \). Then the level sets

\[
D(\Gamma) = \{x \in \mathbb{R}^n : \|F(x)\| \leq \Gamma\}
\]

are bounded for all positive numbers \( \Gamma \) if one of the following two conditions is satisfied:

(i) \( l \) and \( u \) are both bounded;

(ii) \( p \) is a uniform \( P \)-function with respect to \( q \) and \( q \) is surjective and Lipschitz continuous.

**Proof.** Since \( q \) is norm coercive, the boundedness of \( D(\Gamma) \) under assumption (i) follows easily.

Next we prove the boundedness of \( D(\Gamma) \) under assumption (ii). It is not difficult to verify that for \( a, b \in \mathbb{R}^n \),

\[
|a_i - \text{mid}(l_i, u_i, a_i - b_i)| \rightarrow \infty \quad \text{as} \quad |a_i|, |b_i| \rightarrow \infty.
\]

Suppose that there exists one \( \Gamma > 0 \) such that \( D(\Gamma) \) is unbounded, i.e., there exists a sequence \( \{x_k\} \subseteq D(\Gamma) \) such that \( \|x_k\| \rightarrow \infty \). Since \( q \) is norm coercive, \( \{\|q(x_k)\|\} \) is unbounded. Define the index set \( J \) by \( J := \{i | q_i(x_k) \} \) is unbounded, \( i = 1, 2, \ldots, n \}. \) Then \( J \neq \emptyset \). Since \( q \) is surjective, we can choose \( y^k \in \mathbb{R}^n \) such that

\[
q_i(y^k) = \begin{cases} 
q_i(x_k) & \text{if } i \notin J,
0 & \text{if } i \in J.
\end{cases}
\]

Then \( \{\|q(y^k)\|\} \), and so \( \{\|y^k\|\} \), is bounded. Since \( p \) is a uniform \( P \)-function with respect to \( q \), there is a positive number \( \kappa \) such that

\[
\kappa \|x^k - y^k\|^2 \leq \max_{1 \leq i \leq n} (q_i(x^k) - q_i(y^k))(p_i(x^k) - p_i(y^k)) \leq \max_{1 \leq i \leq n} |q_i(x^k) - q_i(y^k)||p_i(x^k) - p_i(y^k)|,
\]
where, together with the Lipschitz continuity of \( q \), implies that
\[
L^{-2} \kappa \sum_{i \in J} (q_i(x^k))^2 = L^{-2} \kappa \|q(y^k) - q(x^k)\|^2 \\
\leq \kappa \|x^k - y^k\|^2 \\
\leq \max_{1 \leq i \leq n} |q_i(y^k) - q_i(x^k)||p_i(x^k) - p_i(y^k)| \\
= \max_{i \in J} |q_i(x^k)||p_i(x^k) - p_i(y^k)| \\
\leq \sqrt{\sum_{i \in J} (q_i(x^k))^2} \max_{i \in J} |p_i(x^k) - p_i(y^k)|,
\]
where \( L \) is the Lipschitz constant of \( q \). Then \( \max_{i \in J} |p_i(x^k) - p_i(y^k)| \to \infty \) as \( k \to \infty \). Since \( \{p_i(y^k)\} \) is bounded, for each \( k \) there exists at least one \( i_k \in J \) such that
\[
|p_{i_k}(x^k)| \to \infty.
\]
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
\[
|p_i(x^k)| \to \infty.
\]
Then we have proved that there exists at least one \( i \in J \) such that
\[
|q_i(x^k)|, |p_i(x^k)| \to \infty,
\]
which, together with (4.2), implies that for such \( i \), \( \{|F_i(x^k)|\} \) is unbounded. This is a contradiction. So for every \( \Gamma > 0 \), the level set \( D(\Gamma) \) is bounded. \( \square \)

Remark 4.1. Let \( I = \{i \mid l_i = -\infty \text{ or } u_i = \infty \text{, } i = 1, \ldots, n\} \). Notice that the index set \( J \) in the proof of Proposition 4.1 is a subset of \( I \). Conditions (i) and (ii) in Proposition 4.1 can be replaced by the following condition
\[
(4.3) \quad \kappa \sum_{i \in I} (q_i(x) - q_i(y))^2 \leq \max_{1 \leq i \leq n} (q_i(x) - q_i(y))(p_i(x) - p_i(y)).
\]
Moreover, if condition (i) holds, then \( I = \emptyset \), so (4.3) holds directly. If condition (ii) holds, then from
\[
\sum_{i \in I} (q_i(x) - q_i(y))^2 \leq \|q(x) - q(y)\|^2,
\]
we have (4.3).

**Proposition 4.2.** Suppose that \( q'(x) \) is nonsingular at \( x \in \mathbb{R}^n \).

(i) If \( p'(x)q'(x)^{-1} \) is a \( P \)-matrix, then for any \( \varepsilon > 0 \), \( f_\varepsilon(x, \varepsilon) \) and all elements in \( \partial \mathcal{C} F(x) \) are nonsingular;

(ii) If \( p'(x)q'(x)^{-1} \) is a \( P_0 \)-matrix and the support of the density function \( \rho \),
\[
\text{supp}(\rho) = \{s : \rho(s) > 0\}
\]
is the whole real line. Then for any \( \varepsilon > 0 \), \( f_\varepsilon(x, \varepsilon) \) is nonsingular.
Proof. (i) By (2.6),
\begin{equation}
4.4 \quad f_x(x, \varepsilon) = q'(x) - D(x)(q'(x) - p'(x)),
\end{equation}
where \( D(x) = \text{diag}(d_i(x)) \) has \( d_i(x) \in [0, 1] \). Let
\[
\mathcal{P}(x) := q'(x) - \begin{pmatrix} [0, 1] \\ \vdots \\ [0, 1] \end{pmatrix} (q'(x) - p'(x)).
\]
By Lemma 5.1 in [39], all elements in \( \mathcal{P}(x) \) are nonsingular. Obviously, \( f_x(x, \varepsilon) \in \mathcal{P}(x) \) and \( \partial_C F(x) \subseteq \mathcal{P}(x) \). Hence we obtain (i).

(ii) \( \text{Supp}(\rho) = \mathbb{R} \) implies that the diagonal matrix \( D(x) \) in (4.4) has \( d_i(x) \in [0, 1] \). By Theorem 4.2 in [18],
\[
I - D(x)(I - p'(x)q'(x)^{-1})
\]
is nonsingular. Thus
\[
f_x(x, \varepsilon) = (I - D(x)(I - p'(x)q'(x)^{-1}))q'(x)
\]
is nonsingular. \( \square \)

The condition in Proposition 4.1 that \( p'(x)q'(x)^{-1} \) is a \( P \)-matrix can be weakened to the condition that \( [p'(x)q'(x)^{-1}]_I \) is nonsingular and its Schur complement in \( [p'(x)q'(x)^{-1}]_{I \cup B} \),
\[
[p'(x)q'(x)^{-1}]_{I \cup B}/[p'(x)q'(x)^{-1}]_I,
\]
is a \( P \)-matrix, where
\[
I = \{i : l_i < q_i(x) - p_i(x) < u_i\},
\]
\[
B = \{i : q_i(x) - p_i(x) = l_i \} \cup \{i : q_i(x) - p_i(x) = u_i\}.
\]

Corollary 4.1. Suppose that \( p, q : \mathbb{R}^n \rightarrow \mathbb{R}^n \) are continuously differentiable, \( p \) is a uniform \( P \)-function with respect to \( q \), and \( q \) is norm coercive, surjective and Lipschitz continuous. Then the iteration sequence \( \{x^k\} \) generated by Algorithm 3.1 for \( F \) given by (1.11) and \( f \) defined by (2.4) is well defined and converges to the unique solution \( x^* \) of \( F(x) = 0 \) superlinearly. Furthermore, if \( p' \) and \( q' \) are locally Lipschitz continuous around \( x^* \), the convergence is quadratic.

Proof. By the assumption that \( p \) is a uniform \( P \)-function with respect to \( q \), there exists a \( \kappa > 0 \) such that (4.1) holds. Since \( p, q \) are continuously differentiable, from (4.1) for any \( x \in \mathbb{R}^n \) we have
\[
\max_{1 \leq i \leq n} (q'(x)z)_i (p'(x)z)_i \geq \kappa z^T z \quad \text{for all } z \in \mathbb{R}^n.
\]
This, from Lemma 5.1 of [39], implies that both \( q'(x) \) and \( p'(x) \) are nonsingular and \( p'(x)q'(x)^{-1} \) is a \( P \)-matrix. Then from Proposition 4.2, for any \( x \in \mathbb{R}^n \) and \( \varepsilon > 0 \), \( f_x(x, \varepsilon) \) is nonsingular and all elements in \( \partial_C F(x) \) are nonsingular. The boundedness of \( D_0 \) follows from Proposition 4.1. The uniqueness of the solution follows from the nonsingularity of all elements in \( \partial_C F(x) \). Finally, \( F \) is piecewise smooth, hence semismooth everywhere, from Theorem 3.2 we may conclude that \( \{x^k\} \) is well defined and converges to the unique solution \( x^* \) superlinearly.

The quadratic convergence follows from the fact that when \( p' \) and \( q' \) are Lipschitz continuous around \( x^* \), \( F \) is strongly semismooth at \( x^* \). \( \square \)
Remark 4.2. When $q$ is the identity map, Corollary 4.1 says that if $p$ is a continuously differentiable uniform $P$-function, then the sequence $\{x^k\}$ generated by Algorithm 3.1 is well defined and converges to the unique solution of box constrained variational inequalities superlinearly. Such a result was only obtained for the nonlinear complementarity problem in [8]. In [27], [48], a similar result for box constrained variational inequalities was obtained by assuming that $p$ is a strongly monotone function, which is a stronger condition than that of a uniform $P$-function.

5. Final remarks

In this paper we have shown that the smoothing Newton method for solving nonsmooth equations can have a convergence rate better than the linear rate. We have established global and superlinear convergence of the smoothing Newton method based on the Jacobian consistency property. Furthermore, we have investigated our conditions used in the convergence analysis for general box constrained variational inequalities $\text{GVI}(l, u, p, q)$. We have shown that all assumptions hold if $p$ is a uniform $P$-function with respect to $q$ and $q$ is norm coercive, surjective and Lipschitz continuous. In contrast with other methods for $\text{GVI}(l, u, p, q)$, the smoothing Newton method has three advantages: solution of a linear system at each iteration, the boundedness of the iterates and superlinear convergence rate.

Algorithm 3.1 and the convergence analysis in section 3 are not restricted to $\text{GVI}(l, u, p, q)$. More applications of the smoothing Newton method are possible. For example, we can use the smoothing Newton method to solve the order complementarity problem. The order complementarity problem is to find an $x \in \mathbb{R}^n$ such that

\begin{equation}
q^i(x) \geq 0, \quad i = 1, \ldots, m, \quad \prod_{i=1}^{m} q^j_i(x) = 0, \quad j = 1, \ldots, n,
\end{equation}

where all $q^i : \mathbb{R}^n \to \mathbb{R}^n$ are continuously differentiable. Problem (5.1) has received as increasing amount of interest recently [19], [24]. Let “min” be the component minimum operator. Problem (5.1) is equivalent to finding a zero of the following nonsmooth equation

$$\min(q^i(x) : i = 1, \ldots, m) = 0.$$ 

For simplicity, let us consider

\begin{equation}
F(x) = \min(r(x), q(x), p(x)).
\end{equation}

By using the results of Zang [52], (5.2) is equivalent to

$$F(x) = r(x) - \max(r(x) - q(x) + \max(q(x) - p(x), 0), 0).$$

A smoothing approximation function $f$, which has the Jacobian consistency property, can be given as

$$f_i(x, \varepsilon) = r_i(x) - \int_{-\infty}^{\infty} \max \left( r_i(x) - q_i(x) - \varepsilon t + \int_{-\infty}^{\infty} \max(q(x) - p(x) - s\varepsilon, 0) \rho(s)ds, 0 \right) \rho(t)dt,$$

where $\rho : \mathbb{R} \to \mathbb{R}_+$ is a density function with a bounded absolute mean.

Another possible application of the smoothing Newton method is for the variational inequality problem (1.3) with $X$ given by

$$X = \{x \in \mathbb{R}^n | g(x) \leq 0, \quad h(x) = 0, \quad l \leq x \leq u\},$$
where $g: \mathbb{R}^n \rightarrow \mathbb{R}^{m_1}$ and $h: \mathbb{R}^n \rightarrow \mathbb{R}^{m_2}$ are assumed to be twice continuously differentiable. The Karush-Kuhn-Tucker conditions of problem (1.3) can be written as [49]

$$
\begin{pmatrix}
    x - \Pi_{[l,u]}[x - L(x, \lambda, \mu)] \\
    \lambda - \Pi_{\mathbb{R}^n_+} [\lambda - (-g(x))] \\
    -h(x)
\end{pmatrix} = 0,
$$

(5.3)

where

$$
L(x, \lambda, \mu) = p(x) + \sum_{i=1}^{m_1} \nabla g_i(x) \lambda_i + \sum_{j=1}^{m_2} \nabla h_j(x) \mu_j.
$$

Problem (5.3) is a special box constrained variational inequality problem, so the smoothing Newton method can be used directly to find a solution of (5.3), and in turn to find a solution of (1.3) under some constraint qualification conditions [20]. However for problem (5.3), the boundedness assumption on the level sets may not hold even if $p$ is strongly monotone. We will leave this and the comparison of different smoothing approximation functions in the smoothing Newton methods as further research topics.

ACKNOWLEDGMENT

The authors would like to thank H. Jiang, C. Kanzow, S.M. Robinson and two referees for their helpful comments.

REFERENCES


SMOOTHING NEWTON METHOD


License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use


School of Mathematics, The University of New South Wales, Sydney 2052, Australia

E-mail address: X.Chen@unsw.edu.au

E-mail address: L.Qi@unsw.edu.au

E-mail address: sun@alpha.maths.unsw.edu.au

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use