JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 113, No. 1, pp. 121–147, April 2002 (©2002)

Smoothing Functions and Smoothing Newton Method for Complementarity and Variational Inequality Problems[1]

L. QI[2] AND D. SUN[3]

Communicated by O. L. Mangasarian

Abstract. This paper provides for the first time some computable smoothing functions for variational inequality problems with general constraints. This paper proposes also a new version of the smoothing Newton method and establishes its global and superlinear (quadratic) convergence under conditions weaker than those previously used in the literature. These are achieved by introducing a general definition for smoothing functions, which include almost all the existing smoothing functions as special cases.

Key Words. Variational inequality problems, computable smoothing functions, smoothing Newton methods, quadratic convergence.

1. Introduction

Consider the equation

$$H(x) = 0, \tag{1}$$

where $H: \Re^n \to \Re^n$ is locally Lipschitz continuous but not necessarily smooth (continuously differentiable). By the Rademacher theorem, H is differentiable almost everywhere. So, for any $x \in \Re^n$, the Clarke generalized Jacobian $\partial H(x)$ is well defined (Ref. 1). Such nonsmooth equations arise from nonlinear complementarity problems, variational inequality problems,

[1]This work was supported by the Australian Research Council and the Research Grant Council of Hong Kong.

[2]Professor, Department of Applied Mathematics, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.
[3]Assistant Professor, Department of Mathematics, National University of Singapore, Singapore, Republic of Singapore.

maximal monotone operator problems (Ref. 2), and interpolation problems (Ref. 3). Various Newton-type methods have been proposed to solve (1). Among them, smoothing methods have received an increasing interest in the literature for solving (1) in connection with nonlinear complementarity problems and variational inequality problems with simple constraints; e.g. see Ref. 4 for a review.

The feature of a smoothing method is to construct a smoothing function $G_\epsilon : \Re^n \to \Re^n$ of H such that, for any $\epsilon > 0$, G_ϵ is continuously differentiable on $\Re^n$ and, for any $x \in \Re^n$, it satisfies

$$\|H(z) - G_\epsilon(z)\| \to 0, \qquad \text{as } \epsilon \downarrow 0,\ z \to x, \tag{2}$$

and then to find a solution of (1) by solving approximately the following problems for a given positive sequence $\{\epsilon^k\}$, $k = 0, 1, 2, \ldots,$:

$$G_{\epsilon^k}(x^k) = 0. \tag{3}$$

Equation (2) provides a generalized definition for a smoothing function, which includes almost all the existing smoothing functions as special cases. Smoothing functions of locally Lipschitz functions satisfying (2) can be obtained via convolution (Ref. 5). Usually, a multivariate integral is involved in computing these smoothing functions via convolution, which makes them uncomputable in practice. In this paper, we focus on studying computable smoothing functions for those nonsmooth functions arising from complementarity problems and variational inequality problems. We shall develop first some new properties of existing smoothing functions, which are essential in designing high-order convergent methods and then discuss several new smoothing functions. In particular, we shall show for the first time a way to get a class of computable smoothing functions for variational inequality problems with general constraints. Another aim of this paper is to establish globally and locally superlinearly (quadratically) convergent methods for solving (1) based on smoothing functions of H. The methods introduced in Refs. 6 and 7 require the inequality

$$\|H(x) - G_\epsilon(x)\| \le \mu\epsilon \tag{4}$$

to hold for some known $\mu > 0$ and all $x \in \Re^n$. In this paper, (4) is replaced by (2), which is a weaker condition, if not the weakest one.

The variational inequality problem (VIP) is to find $x^* \in X$ such that

$$(x - x^*)^T F(x^*) \ge 0, \qquad \text{for all } x \in X, \tag{5}$$

where X is a nonempty closed convex subset of $\Re^n$ and F: $D \to \Re^n$ is continuously differentiable on an open set D which contains X. When $X = \Re^n_+$, the

VIP reduces to the nonlinear complementarity problem (NCP): Find $x^* \in \Re^n_+$ such that

$$F(x^*) \in \Re^n_+ \quad \text{and} \quad F(x^*)^T x^* = 0. \tag{6}$$

It is well known (see e.g. Refs. 8 and 9) that solving (5) is equivalent to finding a root of the following equation:

$$H(x) := x - \Pi_X[x - F(x)] = 0, \tag{7}$$

where (for any $x \in \Re^n$) $\Pi_X(x)$ is the Euclidean projection of x onto X. For the NCP, (7) becomes

$$H(x) = x - \max\{0, x - F(x)\} = \min\{x, F(x)\} = 0, \tag{8}$$

where max and min are componentwise operators. Also, to solve the NCP is equivalent to solve

$$H_i(x) := \phi(x_i, F_i(x)) = 0, \qquad i = 1, 2, \ldots, n, \tag{9}$$

where $\phi : \Re^2 \to \Re$ is the Fischer–Burmeister function (Ref. 10),

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

It is also well known that solving the VIP is equivalent to solving the following normal equation:

$$H(y) := F(\Pi_X(y)) + y - \Pi_X(y) = 0, \tag{10}$$

in the sense that, if $y^* \in \Re^n$ is a solution of (10), then $x^* := \Pi_X(y^*)$ is a solution of (5); conversely, if x^* is a solution of (5), then $y^* := x^* - F(x^*)$ is a solution of (10); see Ref. 11.

In Section 2, we give some preliminaries. Section 3 discusses the properties of smoothing functions for three simple one-dimensional nonsmooth functions, which will form a base for discussing smoothing functions for complicated nonsmooth functions. In Section 4, we show how to compute a class of smoothing functions for the variational inequality problem under the condition that the constraint set X has a nonempty interior. These smoothing functions will play essential roles in designing smoothing Newton methods when the variational inequality problem is defined only on X. In Section 5, we design a new algorithm to solve nonsmooth equations based on smoothing functions and give its convergence analysis in Section 6 with mild assumptions on the smoothing functions involved. Finally, we make some remarks in Section 7.

2. Preliminaries

In this section, we give some basic concepts and preliminary results used in our analysis.

2.1. Smoothing Functions. A function $G_\epsilon: \Re^n \to \Re^m$ is called a smoothing function of a nonsmooth function $H: \Re^n \to \Re^m$ if, for any $\epsilon > 0$, $G_\epsilon(\cdot)$ is continuously differentiable and, for any $x \in \Re^n$,

$$\lim_{\epsilon \downarrow 0,\, z \to x} G_\epsilon(z) = H(x). \tag{11}$$

Definition 2.1. Let $H: \Re^n \to \Re^m$ be a locally Lipschitz continuous function.

(i) $G_\epsilon: \Re^n \to \Re^m$ is called a regular smoothing function of H if, for any $\epsilon > 0$, G_ϵ is continuously differentiable and, for any compact set $D \subseteq \Re^n$ and $\bar{\epsilon} > 0$, there exists a constant $L > 0$ such that, for any $x \in D$ and $\epsilon \in (0, \bar{\epsilon}]$,

$$\|G_\epsilon(x) - H(x)\| \leq L\epsilon. \tag{12}$$

(ii) G_ϵ is said to approximate H at x superlinearly if, for any $y \to x$ and $\epsilon \downarrow 0$, we have

$$G_\epsilon(y) - H(x) - G'_\epsilon(y)(y - x) = o(\|y - x\|) + O(\epsilon). \tag{13}$$

(iii) G_ϵ is said to approximate H at x quadratically if, for any $y \to x$ and $\epsilon \downarrow 0$, we have

$$G_\epsilon(y) - H(x) - G'_\epsilon(y)(y - x) = O(\|y - x\|^2) + O(\epsilon). \tag{14}$$

It is clear that a regular smoothing function of H is a smoothing function of H.

2.2. Semismoothness. In order to establish the superlinear convergence of generalized Newton methods for nonsmooth equations, we need the concept of semismoothness. Semismoothness was introduced originally by Mifflin (Ref. 12) 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 (Ref. 12). Semismooth functionals play an important role in the global convergence theory of nonsmooth optimization; see Polak (Ref. 13). In Ref. 14, Qi and Sun extended the definition of semismooth functions to vector-valued functions. Suppose that $H: \Re^n \to \Re^m$ is locally Lipschitz continuous. H is said to be semismooth at $x \in \Re^n$ if the following limit exists for any $h \in \Re^n$:

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

It has been proved in Ref. 14 that H is semismooth at x if and only if all its component functions are semismooth. Also, the directional derivative $H'(x; h)$ of H at x in the direction h exists for any $h \in \Re^n$ if H is semismooth at x.

Theorem 2.1. See Ref. 14. Suppose that $H: \Re^n \to \Re^m$ is a locally Lipschitzian and semismooth function at $x \in \Re^n$.

(i) For any $V \in \partial H(x+h), \quad h \to 0,$

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

(ii) For any $h \to 0$,

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

The following result is extracted from Theorem 2.3 of Ref. 14.

Theorem 2.2. Suppose that $H: \Re^n \to \Re^m$ is a locally Lipschitzian function. Then, the following two statements are equivalent:

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

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

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

$$H(x+h) - H(x) - Vh = O(\|h\|^2).$$

A function H is said to be a strongly semismooth function if it is strongly semismooth everywhere.

2.3. Weakly Univalent Functions. A function $H: X \subseteq \Re^n \to \Re^n$ is called weakly univalent if it is continuous and there exists a sequence of continuous injective functions $H^j: X \to \Re^n$ such that $\{H^j\}$ converges to H uniformly on any bounded subset of X. The following results on weakly univalent functions were obtained by Gowda and Sznajder (Ref. 15, Theorem 2) and by Ravindran and Gowda (Ref. 16, Theorem 1). These results will be used in the analysis of bounded level sets of smoothing functions.

Theorem 2.3. Let $H: \Re^n \to \Re^n$ be a weakly univalent function. Assume that the inverse image $H^{-1}(0)$ is nonempty and bounded. Then:

(i) $H^{-1}(0)$ is connected;
(ii) there exists a bounded open set D containing $H^{-1}(0)$ such that $\deg(H, D) = 1$;
(iii) there exists a scalar $\delta > 0$ such that the level set $\{x \in \Re^n \,|\, \|H(x)\| \le \delta\}$ is bounded.

3. Smoothing Functions for Simple Nonsmooth Functions

A function $\rho: \Re \to \Re_+$ is called a kernel function if it is integrable (in the sense of Lebesgue) and

$$\int_{\Re} \rho(s)\, ds = 1.$$

Suppose that ρ is a kernel function. Define $\Theta: \Re_{++} \times \Re^m \to \Re_+$ by

$$\Theta(\epsilon, x) := \epsilon^{-m} \Phi(\epsilon^{-1} x),$$

where $(\epsilon, x) \in \Re_{++} \times \Re^m$ and

$$\Phi(z) := \prod_{i=1}^{m} \rho(z_i), \qquad z \in \Re^m.$$

Then, a smoothing approximation of a nonsmooth function $F: \Re^m \to \Re^p$ via convolution can be described by

$$\begin{aligned} F_\epsilon(x) &:= \int_{\Re^n} F(x-y)\Theta(\epsilon, y)\, dy \\ &= \int_{\Re^n} F(x - \epsilon y)\Phi(y)\, dy \\ &= \int_{\Re^n} F(y)\Theta(\epsilon, x-y)\, dy, \end{aligned} \tag{15}$$

where $(\epsilon, x) \in \Re_{++} \times \Re^m$. Such a smoothing function $F_\epsilon(\cdot)$ has many good properties (Refs. 17 and 5). However, in general, $F_\epsilon(\cdot)$ is uncomputable, since a multivariate integral is involved. Nevertheless, when F has a special structure, F_ϵ can be expressed explicitly.

3.1. Plus Function. One of the simplest but very useful nonsmooth functions is the plus function $p\colon \Re \to \Re_+$, defined by

$$p(t) := \max\{0, t\},$$

for any $t \in \Re$.

Suppose that $\rho\colon \Re \to \Re_+$ is any kernel function with

$$\kappa := \int_{\Re} |s| \rho(s)\, ds < +\infty. \tag{16}$$

Chen and Mangasarian (Ref. 18) discussed the following smoothing function for p:

$$P(\epsilon, t) := \int_{\Re} p(t - \epsilon s) \rho(s)\, ds, \qquad (\epsilon, t) \in \Re_{++} \times \Re. \tag{17}$$

For convenience, we define always

$$P(0, t) := p(t) \quad \text{and} \quad P(-|\epsilon|, t) := P(|\epsilon|, t), (\epsilon, t) \in \Re^2.$$

Define

$$\operatorname{supp}(\rho) := \{t | \rho(t) > 0\}.$$

Proposition 3.1.

(i) For any $\epsilon > 0$ and $t \in \Re$,

$$|P(\epsilon, t) - p(t)| \le \kappa \epsilon.$$

(ii) For any $\epsilon > 0$, $P(\epsilon, \cdot)$ is continuously differentiable on $\Re$ and

$$P'_t(\epsilon, t) \in [0, 1], \qquad t \in \Re,$$

where $P'(\epsilon, \cdot)$ is the derivative function of $P(\epsilon, \cdot)$. If $\operatorname{supp}(\rho) = \Re$, then

$$P'_t(\epsilon, t) \in (0, 1), \qquad (\epsilon, t) \in \Re_{++} \times \Re.$$

(iii) P is globally Lipschitz continuous on $\Re^2$.

(iv) P is continuously differentiable on $\Re_{++} \times \Re$ and, for $(\epsilon, t) \in \Re_{++} \times \Re$, we have

$$\nabla P(\epsilon, t) = \begin{bmatrix} -\displaystyle\int_{-\infty}^{t/\epsilon} s\rho(s)\, ds \\ \displaystyle\int_{-\infty}^{t/\epsilon} \rho(s)\, ds \end{bmatrix}.$$

(v) The directional derivative of P at $(0, t)$ exists; for any $h = (h_\epsilon, h_t) \in \Re^2$ with $h_\epsilon \neq 0$, we have

$$P'((0, t); h) = \begin{cases} h_t - |h_\epsilon| \int_{\Re} s\rho(s)\, ds, & \text{if } t > 0, \\ h_t \int_{-\infty}^{h_t/|h_\epsilon|} \rho(s)\, ds - |h_\epsilon| \int_{-\infty}^{h_t/|h_\epsilon|} s\rho(s)\, ds, & \text{if } t = 0, \\ 0, & \text{if } t < 0, \end{cases}$$

and when $h_\epsilon = 0$, we have

$$P'((0, t); h) = \begin{cases} h_t, & \text{if } t > 0, \\ p(h_t), & \text{if } t = 0, \\ 0, & \text{if } t < 0. \end{cases}$$

(vi) For any $\epsilon \downarrow 0$ and $\Delta t \to 0$, we have

$$P(\epsilon, t + \Delta t) - P(0, t) - P'(\epsilon, t + \Delta t) \begin{bmatrix} \epsilon \\ \Delta t \end{bmatrix} = \begin{cases} 0, & \text{if } t = 0, \\ o(\epsilon), & \text{if } t \neq 0. \end{cases}$$

If $\text{supp}(\rho)$ is bounded, then for all $\epsilon > 0$ and $|(\epsilon, \Delta t)|$ sufficiently small, we have

$$P(\epsilon, t + \Delta t) - P(0, t) - P'(\epsilon, t + \Delta t) \begin{bmatrix} \epsilon \\ \Delta t \end{bmatrix} = 0.$$

If $\text{supp}(\rho)$ is unbounded, but there exists a number $\alpha > 2$ such that

$$\limsup_{s \to \infty} \rho(s) |s|^\alpha < \infty,$$

then for any $\epsilon \downarrow 0$ and $\Delta t \to 0$,

$$P(\epsilon, t + \Delta t) - P(0, t) - P'(\epsilon, t + \Delta t) \begin{bmatrix} \epsilon \\ \Delta t \end{bmatrix} = \begin{cases} 0, & \text{if } t = 0, \\ O(\epsilon^{(\alpha - 1)}), & \text{if } t \neq 0. \end{cases}$$

(vii) P is semismooth on $\Re^2$. If $\text{supp}(\rho)$ is bounded or if

$$\limsup_{s \to \infty} \rho(s) |s|^3 < \infty, \tag{18}$$

then P is strongly semismooth on $\Re^2$.

Proof. Parts (i) and (ii) are proved in Ref. 18 and Part (iii) is proved in Ref. 19. Then, we need only to prove Parts (iv)–(vii). By direct computation, we obtain (iv)–(vi). Next, we prove (vii). By (iv), we can prove that $P'(\cdot)$ is locally Lipschitz continuous around any $(\epsilon, t)\in\Re^2$ with $\epsilon\neq 0$. Thus, P is strongly semismooth at $(\epsilon, t)\in\Re^2$ with $\epsilon\neq 0$. Then, we need only to consider the points $(0, t)\in\Re^2$. By (iii), (v), Theorem 2.1, and Theorem 2.2, in order to prove that P is semismooth at $(0, t)$, we need only to prove that, for any $(\epsilon, \Delta t)\in\Re^2$ with $(\epsilon, \Delta t)\to 0$ and all $V\in\partial P(\epsilon, t+\Delta t)$, we have

$$P(\epsilon, t+\Delta t) - P(0, t) - V\begin{bmatrix}\epsilon \\ \Delta t\end{bmatrix} = o(\|(\epsilon, t)\|),$$

which according to (vi) holds for all $\epsilon\neq 0$. When $\epsilon = 0$, by (iv) we can verify easily that, for any $V\in\partial P(0, t+\Delta t)$, there exists a $W\in\partial p(t+\Delta t)$ such that

$$P(\epsilon, t+\Delta t) - P(0, t) - V\begin{bmatrix}\epsilon \\ \Delta t\end{bmatrix} = p(t+\Delta t) - p(t) - W\Delta t = 0,$$

for all Δt sufficiently small. Then, we have proved that P is semismooth on $\Re^2$. Under our further assumptions, this implies that P is strongly semismooth on $\Re^2$. □

The following are three well-known smoothing functions for the plus function p:

(a) neural network function,

$$P(\epsilon, t) = t + \epsilon\log(1 + e^{-t/\epsilon}), \tag{19a}$$

$$\rho(s) = e^{-s}/(1 + e^{-s})^2; \tag{19b}$$

(b) Chen–Harker–Kanzow–Smale (CHKS) function (Refs. 20–22),

$$P(\epsilon, t) = (\sqrt{4\epsilon^2 + t^2} + t)/2, \tag{20a}$$

$$\rho(s) = 2/(s^2 + 4)^{3/2}; \tag{20b}$$

(c) uniform smoothing function (Ref. 23),

$$P(\epsilon, t) = \begin{cases} t, & \text{if } t \geq \epsilon/2, \\ (1/2\epsilon)(t + \epsilon/2)^2, & \text{if } -\epsilon/2 < t < \epsilon/2, \\ 0, & \text{if } t \leq -\epsilon/2, \end{cases} \tag{21a}$$

$$\rho(s) = \begin{cases} 1, & \text{if } -(1/2) \leq s \leq (1/2), \\ 0, & \text{otherwise}, \end{cases} \tag{21b}$$

where $(\epsilon, t)\in\Re_{++}\times\Re$.

3.2. Absolute Value Function. The absolute value function $q: \Re \to \Re$ is defined by

$$q(t) = |t|, \qquad t \in \Re.$$

Then, because

$$q(t) = p(t) + p(-t),$$

the smoothing function of q via convolution can be written as

$$Q(\epsilon, t) := \int_{\Re} |t - \epsilon s| \rho(s)\, ds = P(\epsilon, t) + P(\epsilon, -t), \qquad (\epsilon, t) \in \Re_{++} \times \Re, \tag{22}$$

where P is defined by (17). We define also

$$Q(0, t) := |t| \quad \text{and} \quad Q(-|\epsilon|, t) := Q(|\epsilon|, t), (\epsilon, t) \in \Re^2.$$

Apparently,

$$Q(\epsilon, t) = Q(\epsilon, -t) = Q(\epsilon, |t|) = Q(\epsilon, \sqrt{t^2}), \qquad (\epsilon, t) \in \Re^2.$$

Many properties of Q inherit those from P. For example, by Proposition 3.1, Q is continuously differentiable on $\Re^2$ except on the line $(0, t)$, $t \in \Re$.

Analogously to (19)–(21), we have the following smoothing functions for q:

$$Q(\epsilon, t) = \epsilon[\log(1 + e^{-t/\epsilon}) + \log(1 + e^{t/\epsilon})], \tag{23}$$

$$Q(\epsilon, t) = \sqrt{4\epsilon^2 + t^2}, \tag{24}$$

$$Q(\epsilon, t) = \begin{cases} t, & \text{if } t \geq \epsilon/2, \\ (t^2/\epsilon) + (\epsilon/4), & \text{if } -\epsilon/2 < t < \epsilon/2, \\ -t, & \text{if} t \leq -\epsilon/2, \end{cases} \tag{25}$$

where $(\epsilon, t) \in \Re_{++} \times \Re$.

3.3 One-Dimensional Projection Function. If $F(t) = \Pi_D(t)$, $t \in \Re$, where $D = [l, u]$ and $l \leq u$, then the smoothing function of F via convolution is also computable. Gabriel and Moré (Ref. 24) discussed smoothing functions of Π_D as a generalization of (17). Analogously, we can parallelize the results in Proposition 3.1 to this class of smoothing functions. Furthermore, we have

$$\Pi_D(t) \equiv p(t - l) - p(t - u) + l,$$

for all $t \in \Re$. Hence, we may develop formulas of smoothing functions for Π_D by the formulas of P. We omit the details here.

4. Smoothing Functions for Variational Inequality Problems

In this section, we shall discuss a smoothing approximation of Π_X. It has been discussed in Ref. 5 that Π_X can be approximated well by smoothing functions obtained via convolution. If X is a rectangle, these smoothing functions are computable. However, in general a multivariate integral is involved in the convolution. This makes the convolution approach impractical for computing smoothing functions of Π_X. Here, we will study a class of computable smoothing functions of Π_X when X can be expressed explicitly as

$$X := \{x \in \Re^n | g_i(x) \leq 0, i = 1, 2, \ldots, m\}, \tag{26}$$

where each g_i is a twice continuously differentiable convex function. Suppose that the Slater constraint qualification holds; i.e., there exists a point $\bar{x}$ such that

$$g_i(\bar{x}) < 0, \qquad \text{for all } i \in \{1, 2, \ldots, m\}.$$

Then, for any $x \in \Re^n$, there exists a vector $\lambda \in \Re^m_+$ such that

$$y - x + \sum_{i=1}^{m} \lambda_i \nabla g_i(y) = 0, \tag{27}$$

$$\lambda - p(\lambda + g(y)) = 0, \tag{28}$$

where p is the plus function. Suppose that $P(\epsilon, t)$ is the CHKS smoothing function of p, given by

$$P(\epsilon, t) := (\sqrt{4\epsilon^2 + t^2} + t)/2, \qquad (\epsilon, t) \in \Re^2.$$

From the analysis given below, we can see that the CHKS smoothing function can be replaced by other smoothing functions, e.g., the neural network smoothing function. For ease of discussion, we use only the CHKS smoothing function.

Define $A: \Re \times \Re^m \to \Re^m$ by

$$A_i(\epsilon, z) := P(\epsilon, z_i), \qquad (\epsilon, z) \in \Re \times \Re^m, \qquad i = 1, 2, \ldots, m.$$

Consider the perturbed system of (27) and (28),

$$D((y, \lambda), (\epsilon, x)) := \begin{bmatrix} y - x + \sum_{i=1}^{m} \lambda_i \nabla g_i(y) \\ \lambda - A(\epsilon, \lambda + g(y)) \end{bmatrix} = 0, \tag{29}$$

where $(y, \lambda) \in \Re^n \times \Re^m$ are variables and $(\epsilon, x) \in \Re \times \Re^n$ are parameters. For any fixed $(\epsilon, x) \in \Re_{++} \times \Re^n$, the system (29) has a unique solution. In fact, let $y(\epsilon, x)$ be the unique solution of the following problem:

$$\begin{aligned} \min \quad & (1/2)\|y-x\|^2 - \epsilon^2 \sum_{i=1}^{m} \log(-g_i(y)), \\ \text{s.t.} \quad & g(y) < 0. \end{aligned} \tag{30}$$

Let

$$\lambda_i(\epsilon, x) = -\epsilon^2 (g_i(y(\epsilon, x)))^{-1}, \qquad i \in \{1, 2, \ldots, m\}.$$

Then, for each i,

$$-g_i(y(\epsilon, x)) > 0, \lambda_i(\epsilon, x) > 0 \quad \text{and} \quad \lambda_i(\epsilon, x)[-g_i(y(\epsilon, x))] = \epsilon^2.$$

On the other hand, suppose that $(z(\epsilon, x), \mu(\epsilon, x)) \in \Re^n \times \Re^m$ is a solution of

$$D((y, \lambda), (\epsilon, x)) = 0.$$

Then, $z(\epsilon, x)$ is a solution of (30). Thus,

$$z(\epsilon, x) = y(\epsilon, x).$$

Together with (29), this implies that

$$\mu(\epsilon, x) = \lambda(\epsilon, x).$$

We use $(y(\epsilon, x), \lambda(\epsilon, x))$ to denote the unique solution of (29). Since $D'_{(y,\lambda)}((y, \lambda), (\epsilon, x))$ is nonsingular for all $\lambda \geq 0$ and since $(\epsilon, x) \in \Re_{++} \times \Re^n$, by Ref. 25 $(y(\epsilon, x), \lambda(\epsilon, x))$ is continuously differentiable on $\Re_{++} \times \Re^n$.

Proposition 4.1. Suppose that the Slater constraint qualification holds. Then, the following statements hold:

(i) $y(\cdot, \cdot)$ is continuously differentiable on $\Re_{++} \times \Re^n$ and, for any $x \in \Re^n$ and $\epsilon > 0$, $y'_x(\epsilon, x)$ is symmetric, positive semidefinite and

$$\|y'_x(\epsilon, x)\| \leq 1.$$

Moreover, for any $\epsilon > 0$ and $z, x \in \Re^n$, we have

$$(y(\epsilon, z) - y(\epsilon, x))^T (z - x) \geq \|y(\epsilon, z) - y(\epsilon, x)\|^2. \tag{31}$$

(ii) $\displaystyle\lim_{\epsilon \downarrow 0, x \to x^0} y(\epsilon, x) = \Pi_X(x^0).$

Proof.

(i) By the arguments before this proposition, we know that $y(\cdot)$ is continuously differentiable on $\Re_{++} \times \Re^n$. Now, we consider $y'_x(\epsilon, x)$ for some $x \in \Re^n$ and $\epsilon > 0$. Since $(y(\epsilon, x), \lambda(\epsilon, x))$ satisfies (29), by direct computation we have

$$\left[I + \sum_{j=1}^{m} \lambda_i(\epsilon, x)\nabla^2 g_i(y(\epsilon, x)) + \nabla g(y(\epsilon, x)) S \nabla g(y(\epsilon, x))^T\right] y'_x(\epsilon, x) = I, \tag{32}$$

where S is a diagonal matrix with

$$S_{ii} = -\lambda_i(\epsilon, x)/g_i(y(\epsilon, x)), \qquad i = 1, 2, \ldots, m.$$

Thus, from (32), we know that $y'_x(\epsilon, x)$ is symmetric, positive semidefinite and

$$\|y'_x(\epsilon, x)\| \leq 1.$$

Moreover, (31) can be verified easily in terms of the properties of $y'_x(\epsilon, x)$.

. (ii) For $x \in \Re^n$, define $H_x : \Re^{n+m} \to \Re^{n+m}$ by

$$H_x(y, \lambda) = \begin{bmatrix} y - x + \sum_{i=1}^{m} \lambda_i \nabla g_i(y) \\ -g(y) \end{bmatrix}.$$

By the arguments before this proposition, there exists a (x^0, y^0, λ^0) such that $\lambda^0 > 0, -g(y^0) > 0$, and

$$y^0 - x^0 + \sum_{i=1}^{m} \lambda_i^0 \nabla g_i(y^0) = 0.$$

Then, because all g_i are convex functions, we have

$$\begin{aligned} &\{-g(y(\epsilon, x)) + g(y^0)\}^T (\lambda(\epsilon, x) - \lambda^0) \\ &= \{H_{x^0}(y(\epsilon, x), \lambda(\epsilon, x)) - H_{x^0}(y^0, \lambda^0)\}^T \begin{bmatrix} y(\epsilon, x) - y^0 \\ \lambda(\epsilon, x) - \lambda^0 \end{bmatrix} \\ &\quad + (x^0 - x)^T (y(\epsilon, x) - y^0) \\ &\geq \|y(\epsilon, x) - y^0\|^2 + (x^0 - x)^T (y(\epsilon, x) - y^0), \end{aligned}$$

which together with

$$[-g_i(y(\epsilon, x))]\lambda_i(\epsilon, x) = \epsilon^2, \qquad \text{for all } i,$$

implies that

$$\begin{aligned} m\epsilon^2 + (-g(y^0))^T \lambda^0 \geq &-g(y(\epsilon, x))^T \lambda^0 - g(y^0)^T \lambda(\epsilon, x) \\ &+ \|y(\epsilon, x) - y^0\|^2 + (x^0 - x)^T (y(\epsilon, x) - y^0). \end{aligned}$$

This shows that, for any $\bar{\epsilon} > 0$ and $\delta > 0$,

$$\{(y(\epsilon, x), \lambda(\epsilon, x)) | 0 < \epsilon \leq \bar{\epsilon}, \|x - x^0\| \leq \delta\}$$

is bounded. Suppose that (y^*, λ^*) is any accumulation point of $(y(\epsilon, x), \lambda(\epsilon, x))$ as $\epsilon \to 0$ and $x \to x^0$. Then,

$$D((y^*, \lambda^*), (0, x^0)) = 0,$$

which implies that

$$y^* = \Pi_X(x^0).$$

Since $\Pi_X(x^0)$ is single-valued,

$$\lim_{\epsilon \downarrow 0,\, x \to x^0} y(\epsilon, x) = \Pi_X(x^0). \qquad \square$$

Proposition 4.1 shows that $y(\epsilon, x)$ is a computable smoothing function of $\Pi_X(x)$ if X is represented by (26). In order to get $y(\epsilon, x)$, one needs to solve (30), which itself is in general a nonlinear optimization problem and thus is difficult to solve. However, to solve (30) is equivalent to solving (29), and thus it is no more difficult than to compute $\Pi_X(x)$ because in (29), for any $\epsilon > 0$, $A(\epsilon, \cdot)$ is continuously differentiable. In practical applications, some VIPs are defined only on X (Ref. 26) and thus the computation of $\Pi_X(x)$ or its approximation is unavoidable. It can be seen later that not only $y(\epsilon, x)$ approximates $\Pi_X(x)$, but can be used also together with (10) to design a class of Newton-type methods with the VIPs to be defined on X only. This explains clearly why we need to introduce $y(\epsilon, x)$.

Theorem 4.1. Suppose that, at some point $x \in \Re^n$, the vectors

$$\{\nabla g_i(\Pi_X(x))\}, \qquad i \in I(x) := \{j | g_j(\Pi_X(x)) = 0, j = 1, 2, \ldots, m\}$$

are linearly independent. Then, we have that:

(i) $\left\{ \lim_{\epsilon \downarrow 0,\, z \to x} y'_x(\epsilon, z) \right\} \subseteq \partial \Pi_X(x).$

(ii) $y(\cdot)$ is Lipschitz continuous near $(0, x)$ and, for any $z \to x$ and $\epsilon \downarrow 0$,

$$\begin{bmatrix} y(\epsilon, z) \\ \lambda(\epsilon, z) \end{bmatrix} - \begin{bmatrix} y(0, x) \\ \lambda(0, x) \end{bmatrix} - \begin{bmatrix} y'(\epsilon, z) \\ \lambda'(\epsilon, z) \end{bmatrix} \begin{bmatrix} \epsilon \\ z - x \end{bmatrix} = o(\|(\epsilon, z - x)\|). \tag{33}$$

(iii) If $\nabla^2 g_i$ are Lipschitz continuous near $\Pi_X(x)$, then for any $z \to x$ and $\epsilon \downarrow 0$,

$$\begin{bmatrix} y(\epsilon, z) \\ \lambda(\epsilon, z) \end{bmatrix} - \begin{bmatrix} y(0, x) \\ \lambda(0, x) \end{bmatrix} - \begin{bmatrix} y'(\epsilon, z) \\ \lambda'(\epsilon, z) \end{bmatrix} \begin{bmatrix} \epsilon \\ z - x \end{bmatrix} = O(\|(\epsilon, z - x)\|^2). \tag{34}$$

Proof.

(i) First, since the vectors $\{\nabla g_i(\Pi_X(x))\}$, $i \in I(x)$ are linearly independent, there exists a unique vector λ^* such that (27) and (28) hold with $y = \Pi_X(x)$ and $\lambda = \lambda^*$. Since P is Lipschitz continuous (Proposition 3.1), D is locally Lipschitz continuous on $\Re^{2n+m+1}$. Let

$$\mathscr{B} := \{B \in \Re^{(n+m)\times(n+m)} \mid \text{there exists a matrix } C \in \Re^{(n+m)\times(n+1)} \text{ such that } (B\ C) \in \partial D((\Pi(x), \lambda^*), (0, x))\}.$$

It is not difficult to verify that all $B \in \mathscr{B}$ are nonsingular (see e.g. Ref. 27). Then, by the implicit function theorem (Ref. 1, Section 7.1), there exist a neighborhood $\mathscr{N}$ of $(0, x)$ and Lipschitz functions y: $\mathscr{N} \to \Re^n$ and λ: $\mathscr{N} \to \Re^m$ such that

$$D((y(\epsilon, z), \lambda(\epsilon, z)), (\epsilon, z)) = 0, \qquad \text{for any } (\epsilon, z) \in \mathscr{N}.$$

For any $(\epsilon, z) \in \mathscr{N}$ and $\epsilon \neq 0$, (y, λ) is continuously differentiable around (ϵ, z) and satisfies

$$\begin{bmatrix} I + \sum_{i=1}^{m} \lambda_i(\epsilon, z) \nabla^2 g_i(y(\epsilon, z)) & \nabla g(y(\epsilon, z)) \\ S[-\nabla g(y(\epsilon, z))]^T & I - S \end{bmatrix} \begin{bmatrix} y'_x(\epsilon, z) \\ \lambda'_x(\epsilon, z) \end{bmatrix} = \begin{bmatrix} I \\ 0 \end{bmatrix}, \tag{35}$$

where S is a diagonal matrix with

$$S_{ii} = \lambda_i(\epsilon, z)/[\lambda_i(\epsilon, z) - g_i(\epsilon, z)], \qquad i = 1, 2, \ldots, m.$$

Suppose that (U^*, V^*) is a limit point of $(y'_x(\epsilon, z), \lambda'_x(\epsilon, z))$ for $(\epsilon, z) \in \mathscr{N}$ with $\epsilon \neq 0$ and $(\epsilon, z) \to (0, X)$. Then, from (35), there exists a diagonal matrix S^* with

$$S^*_{ii} = \begin{cases} 0, & \text{if } g_i(y(0, x)) < 0, \\ \alpha \in [0, 1], & \text{if } g_i(y(0, x)) = 0 \text{ and } \lambda_i(0, x) = 0, \\ 1, & \text{if } \lambda_i(0, x) > 0, \end{cases}$$

for $i = 1, 2, \ldots, m$, such that

$$\begin{bmatrix} I + \sum_{i=1}^{m} \lambda_i^* \nabla^2 g_i(\Pi_X(x)) & \nabla g(\Pi_X(x)) \\ S^*[-\nabla g^T \Pi_X(x)] & I - S^* \end{bmatrix} \begin{bmatrix} U^* \\ V^* \end{bmatrix} = \begin{bmatrix} I \\ 0 \end{bmatrix}, \tag{36}$$

which by Ref. 28 implies that

$$U^* \in \partial \Pi_X(x).$$

So, (i) is proved.

(ii) Since P is semismooth at $(0, x)$ by Proposition 3.1, D is semismooth at $(y(0, x), \lambda(0, x), 0, x)$. Thus, for any $(y, \lambda, \epsilon, z) \to (y(0, x), \lambda(0, x), 0, x)$ and $W \in \partial D((y, \lambda), (\epsilon, z))$, we have

$$D((y, \lambda), (\epsilon, z)) - D((y(0, x), \lambda(0, x)), (0, x)) - W \begin{bmatrix} y - y(0, x) \\ \lambda - \lambda(0, x) \\ \epsilon - 0 \\ z - x \end{bmatrix}$$

$$= o(\|(y, \lambda, \epsilon, z) - (y(0, x), \lambda(0, x), 0, x)\|).$$

In particular, for $\epsilon \neq 0$ and $(\epsilon, z) \to (0, x)$, we have

$$D((y(\epsilon, z), \lambda(\epsilon, z)), (\epsilon, z)) - D((y(0, x), \lambda(0, x)), (0, x))$$

$$- D'((y(\epsilon, z), \lambda(\epsilon, z)), (\epsilon, z)) \begin{bmatrix} y(\epsilon, z) - y(0, x) \\ \lambda(\epsilon, z) - \lambda(0, x) \\ \epsilon - 0 \\ z - x \end{bmatrix}$$

$$= o(\|(y(\epsilon, z), \lambda(\epsilon, z), \epsilon, z) - (y(0, x), \lambda(0, x), 0, x)\|),$$

which together with the facts that

$$D((y(\epsilon, z), \lambda(\epsilon, z)), (\epsilon, z)) = 0,$$

$$D((y(0, x), \lambda(0, x)), (0, x)) = 0,$$

implies

$$D'((y(\epsilon, z), \lambda(\epsilon, z)), (\epsilon, z)) \begin{bmatrix} y(\epsilon, z) - y(0, x) \\ \lambda(\epsilon, z) - \lambda(0, x) \\ \epsilon - 0 \\ z - x \end{bmatrix}$$

$$= o(\|(y(\epsilon, z), \lambda(\epsilon, z), \epsilon, z) - (y(0, x), \lambda(0, x), 0, x)\|).$$

Since $(y(\cdot), \lambda(\cdot))$ is Lipschitz continuous on a neighborhood of $(0, x)$ and since, for $\epsilon \neq 0$,

$$D'_{(y,\lambda)}((y(\epsilon, z), \lambda(\epsilon, z)), (\epsilon, z)) \begin{bmatrix} y'(\epsilon, z) \\ \lambda'(\epsilon, z) \end{bmatrix}$$

$$+ D'_{(\epsilon,z)}((y(\epsilon, z), \lambda(\epsilon, z)), (\epsilon, z)) = 0,$$

for $\epsilon \neq 0$ and $(\epsilon, z) \to (0, x)$, we have

$$D'_{(y,\lambda)}((y(\epsilon, z), \lambda(\epsilon, z)), (\epsilon, z)) \left[\begin{bmatrix} y(\epsilon, z) \\ \lambda(\epsilon, z) \end{bmatrix} - \begin{bmatrix} y(0, x) \\ \lambda(0, x) \end{bmatrix} - \begin{bmatrix} y'(\epsilon, z) \\ \lambda'(\epsilon, z) \end{bmatrix} \right.$$

$$\left. \times [y(\epsilon, z) - y(0, x)\lambda'(\epsilon, z) - \lambda(0, x)] \right]$$

$$= o(\|(\epsilon, z - x)\|).$$

Since $\partial D(\cdot)$ is upper semicontinuous (Ref. 1), by the proof of (i), for any $\epsilon \neq 0$ and $(\epsilon, z) \to 0$,

$$\|[D'_{(y,\lambda)}((y(\epsilon, z), \lambda(\epsilon, z)), (\epsilon, z))]^{-1}\|$$

is uniformly bounded. So, (33) is proved.

(iii) If $\nabla^2 g_i$ are Lipschitz continuous near $\Pi_X(x)$, then D is strongly semismooth at $(y(0, x), \lambda(0, x), 0, x)$. By following the arguments in (ii), we can prove (34). □

Now, we are ready to describe smoothing functions of H defined by (7). For any $\epsilon > 0$ and $\alpha \geq 0$, define $G_\epsilon : \Re^n \to \Re^n$ by

$$G_\epsilon(x) := x - y(\epsilon, x - H_\epsilon(x)), \qquad x \in \Re^n, \tag{37}$$

where

$$H_\epsilon(x) := F(x) + \alpha \epsilon x.$$

When $\alpha > 0$, (37) is a Tikhonov-type regularized smoothing function for variational inequality problems. For any $c \geq 0$ and $\epsilon \geq 0$, define

$$L_{\epsilon,c} = \{x \in \Re^n \mid \|G_\epsilon(x)\| \leq c\}.$$

Theorem 4.2. Suppose that the Slater constraint qualification holds. If F is continuous and H_ϵ is strongly monotone on $\Re^n$ for some $\epsilon > 0$, then:

(i) there exists a unique x^* such that $G_\epsilon(x^*) = 0$;
(ii) for any $c \geq 0$, $L_{\epsilon,c}$ is bounded.

Proof.

(i) For any $t \in [0, 1]$, define

$$G_{\epsilon,t}(x) := x - y(\epsilon, x - [tH_\epsilon(x) + (1-t)x]), \qquad x \in \Re^n.$$

Let SOL(t) denote the solution set of $G_{\epsilon,t}(x) = 0$. We will first show that the set

$$S := \bigcup_{t \in [0,1]} \mathrm{SOL}(t)$$

is bounded.

By contradiction, suppose that S is not bounded. Then, there exist sequences $\{t_k\}$ and $\{x^k\}$ such that $t_k \in [0, 1]$, $x^k \in \mathrm{SOL}(t_k)$, and $\{x^k\}$ is unbounded. By (31) in Proposition 4.1, we have

$$(w^* - x^k)^T[t_k H_\epsilon(x^k) + (1-t_k)x^k] \geq (w^* - x^k)^T w^*,$$

where $w^* := y(\epsilon, 0)$. Thus, there exists a constant c_1 such that

$$t_k(H_\epsilon(x^k) - H_\epsilon(w^*))^T(x^k - x^*) + (1-t_k)(x^k - w^*)^T(x^k - w^*) \leq c_1 \|x^k - w^*\|,$$

which contradicts our assumption that H_ϵ is a strongly monotone function. This contradiction shows that the set S is bounded. By the homotopy invariance property of degree theory (Refs. 25 and 29), we know that the solution set SOL(1) of $G_\epsilon(x) = 0$ is nonempty.

Suppose that there are two distinct points $x^1, x^2 \in \mathrm{SOL}(1)$. By (31) in Proposition 4.1, we have

$$(x^2 - x^1)^T(H_\epsilon(x^2) - H_\epsilon(x^1)) \leq 0,$$

which contradicts our assumption that H_ϵ is strongly monotone. This contradiction shows that $G_\epsilon(x) = 0$ has a unique solution.

(ii) By contradiction, suppose that $L_{\epsilon,c}$ is not bounded. Then, there exists a sequence $\{x^k\} \in L_{\epsilon,c}$ and $\{x^k\}$ is unbounded. That is,

$$y(\epsilon, x^k - H_\epsilon(x^k)) = x^k - G_\epsilon(x^k),$$
$$\|G_\epsilon(x^k)\| \leq c.$$

By (31) in Proposition 4.1, we have

$$H_\epsilon(x^k)^T[w^* - (x^k - G_\epsilon(x^k))] \geq G_\epsilon(x^k)^T[w^* - (x^k - G_\epsilon(x^k))],$$

where $w^* := y(\epsilon, 0)$. Thus,

$$(H_\epsilon(x^k) - H_\epsilon(s^k))^T(s^k - x^k) \geq (G_\epsilon(x^k) + H_\epsilon(s^k))^T(x^k - s^k), \tag{38}$$

where

$$s^k := w^* + G_\epsilon(x^k).$$

Since $\{s^k\}$, $\{G_\epsilon(x^k)\}$, $\{H_\epsilon(s^k)\}$ are all bounded, (38) contradicts our assumption that H_ϵ is strongly monotone. This contradiction shows that $L_{\epsilon,c}$ is bounded. □

For any $(\epsilon, x) \in \Re \times \Re^n$, define

$$G_0(x) = x - \Pi_X[x - F(x)] \quad \text{and} \quad G_{-|\epsilon|}(x) = G_{|\epsilon|}(x).$$

Theorem 4.3. Suppose that the Slater constraint qualification holds. Suppose that F is a continuous monotone function of $\Re^n$. If the solution set of the VIP is nonempty and bounded, then there exists a $\delta > 0$ such that the following set is bounded:

$$L_\delta := \{(\epsilon, x) \in \Re \times \Re^n \mid |\epsilon| + \|G_\epsilon(x)\| \leq \delta\}.$$

Proof. Define $T: \Re^{n+1} \to \Re^{n+1}$ by

$$T(\epsilon, x) := \begin{bmatrix} \epsilon \\ G_\epsilon(x) \end{bmatrix}, \qquad (\epsilon, x) \in \Re \times \Re^n.$$

Then, from Proposition 4.1(ii), T is a continuous function. For any $j > 0$, define

$$G^j_\epsilon(x) := x - y(\epsilon, x - (H_\epsilon(x) + j^{-1}x))$$

$$T^j(\epsilon, x) := \begin{bmatrix} \epsilon \\ G^j_\epsilon(x) \end{bmatrix}, \qquad (\epsilon, x) \in \Re \times \Re^n.$$

Since F is monotone, by (31) in Proposition 4.1, it is easy to verify that, for any $j > 0$, T^j is a univalent function. Thus, T is a weakly univalent function. By the assumption that the solution set of the VIP is nonempty and bounded, the inverse image $T^{-1}(0)$ is nonempty and bounded. Therefore, by Theorem 2.3, there exists a $\delta > 0$ such that L_δ is bounded. □

When X is of a separable structure, in Theorem 4.3 the assumption that F is a monotone function can be replaced by that F is a generalized P_0-function. See Ref. 30 for the definition of generalized P_0-functions.

The smoothing function of H defined by (10) can be described by

$$G_\epsilon(x) := F(y(\epsilon, x)) + \alpha y(\epsilon, x) + x - y(\epsilon, x), \tag{39}$$

where $(\epsilon, x) \in \Re_{++} \times \Re^n$ and $\alpha \geq 0$. It is not difficult to give results similar to Theorems 4.2 and 4.3 for the smoothing function (39). Here, we omit the details.

To the best of our knowledge, (37) and (39) are the first classes of computable smoothing functions for the variational inequality problem when the constraint set X is not a rectangle. More interestingly, the smoothing function (39) requires F to be defined only on X. Actually, this motivates us to study computable smoothing functions for the projection operator directly since, if F is defined only on X, then any numerical method for the VIP must work on X directly, which needs to compute $\Pi_X(x)$ or its approximation.

5. New Version of the Smoothing Newton Method

In this section, we suppose that $H: \Re^n \rightarrow \Re^n$ is a locally Lipschitz function and that $G_\epsilon: \Re^n \rightarrow \Re^n$ is a smoothing function of H.

Let

$$\theta(x) := (1/2)\|H(x)\|^2, \qquad x \in \Re^n,$$

$$\theta_\epsilon(x) = (1/2)\|G_\epsilon(x)\|^2, \qquad (\epsilon, x) \in \Re \times \Re^n.$$

Algorithm 5.1.

Step 0. Choose constants $\delta \in (0, 1)$, $\beta \in (0, \infty)$, $\sigma \in (0, 1/2)$, $\rho_1 \in (0, +\infty)$, and $\rho_2 \in (2, +\infty)$. Let $x^0 \in \Re^n$ be an arbitrary point; let $k := 0$ and $y^0 := x^0$.

Step 1. Let $d^k \in \Re^n$ satisfy

$$G_{\epsilon^k}(y^k) + G'_{\epsilon^k}(y^k)\, d = 0. \tag{40}$$

If (40) is not solvable or if

$$-(d^k)^T \nabla \theta_{\epsilon^k}(y^k) \geq \rho_1 \|d^k\|^{\rho_2} \tag{41}$$

does not hold, let

$$d^k = -\nabla \theta_{\epsilon^k}(y^k).$$

Step 2. Let l_k be the smallest nonnegative integer l satisfying

$$\theta_{\epsilon^k}(y^k + \delta^l d^k) \leq \theta_{\epsilon^k}(y^k) + \sigma \delta^l \nabla \theta_{\epsilon^k}(y^k)^T d^k. \tag{42}$$

If

$$\|G_{\epsilon^k}(y^k + \delta^{l_k} d^k)\| \leq \epsilon^k \beta, \tag{43}$$

or if

$$\|H(y^k + \delta^{l_k} d^k)\| \leq (1/2)\|H(x^k)\|, \tag{44}$$

let

$$y^{k+1} := y^k + \delta^{l_k} d^k, \qquad x^{k+1} := y^{k+1},$$

and

$$0 < \epsilon^{k+1} \leq \min\{(1/2)\epsilon_k, \theta(x^{k+1})\}. \tag{45}$$

Otherwise, let

$$y^k := y^k + \delta^{l_k} d^k,$$

and go to Step 1.

Step 3. Replace k by $k+1$ and go to Step 1.

Remark 5.1.

(i) Compared to the methods proposed in Ref. 6, two fundamental conditions used in Ref. 6 are weakened. In Ref. 6, the inequality (4) is required to hold for all $x \in \Re^n$ and the following condition is assumed to hold:

$$\operatorname{dist}(G'_{\epsilon^k}(x^k), \partial_C H(x^k)) \to 0, \qquad \text{as } k \to \infty. \tag{46}$$

Here, (4) is replaced by (11) and (46) is not required at all.

(ii) In (44)–(45), the constant 1/2 can be replaced by any positive constant in (0, 1) and $\beta\epsilon$ in (43) can be replaced by any continuous forcing function $\gamma(\epsilon)$ satisfying

$$\gamma(\epsilon) \geq 0, \qquad \forall \epsilon \in [0, +\infty),$$

and $\gamma(\epsilon) = 0$ if and only if $\epsilon = 0$.

6. Convergence Analysis

Assumption 6.1.

(i) There exists a constant $\bar{\epsilon} > 0$ such that

$$D_1 := \{x \in \Re^n \mid \|G_\epsilon(x)\| \leq \beta\epsilon, 0 < \epsilon \leq \bar{\epsilon}\}$$

is bounded.

(ii) For any $\epsilon > 0$ and $\delta > 0$, the following set:

$$L_{\epsilon,\delta} := \{x \in \Re^n \mid \theta_\epsilon(x) \leq \delta\}$$

is bounded and $\nabla\theta_\epsilon(x) = 0$ for any $\epsilon > 0$ and $x \in \Re^n$ imply that $\theta_\epsilon(x) = 0$.

(iii) There exists a constant $c>0$ such that

$$D_2 := \{x \in \Re^n | \theta(x) \leq c\}$$

is bounded.

Remark 6.1. In the previous section, we have discussed several cases such that Assumption 6.1 holds. Related discussions to Assumption 6.1(i) can be found in Refs. 31–34. See Refs. 7 and 16 for a condition to guarantee Assumption 6.1(iii) to hold when nonlinear complementarity problems are concerned.

Theorem 6.1. Suppose that $G_\epsilon : \Re^n \to \Re^n$ is a smoothing function of H and that Assumption 6.1 holds. Then, an infinite bounded sequence $\{x^k\}$ is generated by Algorithm 5.1 and any accumulation point of $\{x^k\}$ is a solution of $H(x)=0$.

Proof. Without loss of generality, we suppose that

$$G_{\epsilon^k}(x^k) \neq 0, \qquad \text{for all } k \geq 0.$$

If only a finite sequence $x^0, x^1, \ldots, x^k$ is generated by Algorithm 5.1, then our algorithm is applied to solve

$$\min_{y \in \Re^n} \theta_{\epsilon^k}(y),$$

starting with $y^k = x^k$. Thus, an infinite sequence $\{y^{k+j}, j=0,1,\ldots\}$ is generated. By Assumption 6.1, $\{y^{k+j}, j=0,1,\ldots\}$ must be bounded. Then, $\{y^{k+j}, j=0,1,\ldots\}$ has at least one accumulation point, say y^*. Then,

$$\nabla \theta_{\epsilon^k}(y^*) = 0;$$

according to Assumption 6.1, this implies that $\theta_{\epsilon^k}(y^*) = 0$. This means that

$$\lim_{j \to \infty} \theta_{\epsilon^k}(y^{k+j}) = 0,$$

which shows that (43) must be satisfied for some $j \geq 0$. Thus, an infinite sequence $\{x^k\}$ is generated and $\epsilon^k \to 0$ as $k \to \infty$. Define

$$K := \{k | \text{inequality (43) is satisfied for } k\}.$$

Suppose that

$$K = \{k_0, k_1, \ldots, k_j, \ldots\}.$$

Then, for all $k \in K$ sufficiently large,

$$x^{k+1} \in D_1.$$

If there are infinitely many elements in K, then by the fact that $\lim_{k\to\infty}\epsilon^k = 0$, we have

$$\lim_{k\in K, k\to\infty} G_{\epsilon^k}(x^{k+1}) = 0.$$

Since G_ϵ is a smoothing function of H, D_1 is bounded, and since $\epsilon^k \to 0$ as $k\to 0$, we have

$$\|G_{\epsilon^k}(x^{k+1}) - H(x^{k+1})\| \to 0, \qquad k\to\infty, k\in K. \tag{47}$$

For any $k\in(k_j, k_{j+1}-1]$, we have

$$\|H(x^{k+1})\| \leq (1/2)\|H(x^{k_j+1})\| \leq (1/2)[\beta\epsilon^{k_j} + \|H(x^{k_j+1}) - G_{\epsilon^{k_j}}(x^{k_j+1})\|],$$

which together with (47) implies that

$$\lim_{k\to\infty, k\notin K} \|H(x^{k+1})\| = 0.$$

By Assumption 6.1(iii) this proves that $\{x^{k+1}, k\notin K\}$ is bounded.

If there are only finitely many elements in K, then there exists some $\bar{k}\geq 0$ such that, for all $k\geq\bar{k}$,

$$\|H(x^{k+1})\| \leq (1/2^{k-\bar{k}})\|H(x^{\bar{k}})\|,$$

which implies that

$$\lim_{k\to\infty} H(x^k) = 0;$$

by Assumption 6.1(iii), this implies that $\{x^k\}$ is bounded.

Overall, we have proved that $\{x^k\}$ is bounded. It is easy to see that any accumulation point of $\{x^k\}$ is a solution of $H(x)=0$. □

For variational inequality problems, we have the following result.

Corollary 6.1. Suppose that H is defined by (7), X is given by (26), and F is a continuously differentiable monotone function. Suppose that the Slater constraint qualification holds and that G_ϵ is defined by (37) with $\alpha > 0$. If the solution set of (5) is nonempty and bounded, then a bounded infinite sequence $\{x^k\}$ is generated by Algorithm 5.1 and any accumulation point of $\{x^k\}$ is a solution of (5).

Proof. By Proposition 4.1, Theorem 4.2, and Theorem 4.3, we know that Assumption 6.1 holds. Then, from Theorem 6.1, we get the results of this corollary. □

Theorem 6.2. Suppose that $G_\epsilon : \Re^n \to \Re^n$ is a smoothing function of H and that Assumption 6.1 holds. Suppose that $\bar{x}$ is an accumulation point of $\{x^k\}$ generated by Algorithm 5.1, that all $V \in \partial H(\bar{x})$ are nonsingular, and that $\{\|(G'_{\epsilon^k}(x^k))^{-1}\|\}$ is uniformly bounded for all x^k sufficiently close to $\bar{x}$. If G_ϵ approximates H at $\bar{x}$ superlinearly [respectively quadratically], then the whole sequence $\{x^k\}$ converges to $\bar{x}$ superlinearly [respectively quadratically].

Proof. By Theorem 6.1, $\bar{x}$ is a solution of $H(x)=0$. Suppose that $\{x^{k_j}\}$ is a subsequence of $\{x^k\}$ that converges to $\bar{x}$. Since $\{\|G'_{\epsilon^k}(x^k))^{-1}\|\}$ is uniformly bounded for all x^k sufficiently close to $\bar{x}$, (41) is satisfied for all k_j sufficiently large. So, when k_j is sufficiently large, we have

$$d^{k_j} = -(G'_{\epsilon^{k_j}}(x^{k_j}))^{-1} G_{\epsilon^{k_j}}(x^{k_j}).$$

From Definition 2.1 and (40), for $k_j \to \infty$, we have

$$\begin{aligned}\|x^{k_j}+d^{k_j}-\bar{x}\| &= O(\|G_{\epsilon^{k_j}}(x^{k_j}) - H(\bar{x}) - G'_{\epsilon^{k_j}}(x^{k_j})(x^{k_j}-\bar{x})\|)\\ &= o(\|x^{k_j}-\bar{x}\|) + O(\epsilon^{k_j}).\end{aligned}$$

Since all $V \in \partial H(\bar{x})$ are nonsingular, from Refs. 14 and 35 there exist $c_1, c_2 > 0$ and a neighborhood $\mathcal{N}(\bar{x})$ of $\bar{x}$ such that, for all $x \in \mathcal{N}(\bar{x})$,

$$c_1\|x-\bar{x}\| \le \|H(x)\| \le c_2\|x-\bar{x}\|.$$

Thus,

$$\|x^{k_j}+d^{k_j}-\bar{x}\| = o(\|x^{k_j}-\bar{x}\|),$$

which implies that, for all k_j sufficiently large, (44) is satisfied and

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

By repeating the above process, we can prove that $\{x^k\}$ converges to $\bar{x}$ and

$$\|x^{k+1}-\bar{x}\| = o(\|x^k-\bar{x}\|).$$

If G_ϵ approximates H at $\bar{x}$ quadratically, by following the above arguments, we can prove that $\{x^k\}$ converges to $\bar{x}$ quadratically. □

Corollary 6.2. Suppose that all the conditions in Corollary 6.1 hold. Then, $\{x^k\}$ generated by Algorithm 5.1 has at least one accumulation point $\bar{x}$ such that $H(\bar{x})=0$. If all

$$V \in \{I - \partial\Pi_X(\bar{x}-F(\bar{x}))[I-F'(\bar{x})]\}$$

are nonsingular and if the vectors

$$\{\nabla g_i(\Pi_X(\bar{x}))\}, \qquad i \in I(\bar{x}) := \{j \,|\, g_j(\Pi_X(\bar{x}) = 0\},$$

are linearly independent, then $\{x^k\}$ converges to $\bar{x}$ superlinearly. Furthermore, if all g'_i are Lipschitz continuous around $\Pi_X(\bar{x})$ and if F' is Lipschitz continuous around $\bar{x}$, then $\{x^k\}$ converges to $\bar{x}$ quadratically.

Proof. By Theorem 4.1, Theorem 6.2, and Corollary 6.1, we get the results of this corollary. □

In Corollary 6.2, we have assumed the condition that all

$$V \in \{I - \partial\Pi_X(\bar{x} - F(\bar{x}))[I - F'(\bar{x})]\}$$

are nonsingular. In Ref. 5, this condition has been discussed. In particular, it was shown in Ref. 5 that, if $F'(\bar{x})$ is a positive-definite matrix, then this condition holds.

7. Final Remarks

For the first time, we have provided some computable smoothing functions for variational inequality problems with general constraints. This class of smoothing functions is essential in solving the variational inequality problem when it is not well defined outside X. Certainly, more smoothing functions can be obtained via our discussion. When some special problems are studied, stronger results can be obtained.

References

1. Clarke, F. H., *Optimization and Nonsmooth Analysis*, Wiley, New York, NY, 1983.
2. Pang, J. S., and Qi, L., *Nonsmooth Equations: Motivation and Algorithms*, SIAM Journal on Optimization, Vol. 3, pp. 443–465, 1993.
3. Dontchev, A. L., Qi, H. D., and Qi, L., *Convergence of Newton's Method for Best Convex Interpolation*, Numerische Mathematik, Vol. 87, pp. 435–456, 2001.
4. Qi, L., and Sun, D., *A Survey of Some Nonsmooth Equations and Smoothing Newton Methods*, Progress in Optimization: Contributions from Australasia, Edited by A. Eberhard, B. Glover, R. Hill, and D. Ralph, Kluwer Academic Publishers, Boston, Massachusetts, pp. 121–146, 1999.
5. Sun, D., and Qi, L., *Solving Variational Inequality Problems via Smoothing-Nonsmooth Reformulations*, Journal of Computational and Applied Mathematics, Vol. 129, pp. 37–62, 2001.
6. Chen, X., Qi, L., and Sun, D., *Global and Superlinear Convergence of the Smoothing Newton Method and Its Application to General Box-Constrained Variational Inequalities*, Mathematics of Computation, Vol. 67, pp. 519–540, 1998.

7. Chen, X., and Ye, Y., *On Homotopy-Smoothing Methods for Variational Inequalities*, SIAM Journal on Control and Optimization, Vol. 37, pp. 589–616, 1999.
8. Eaves, B. C., *On the Basic Theorem of Complementarity*, Mathematical Programming, Vol. 1, pp. 68–75, 1971.
9. Harker, P. T., and Pang, J. S., *Finite-Dimensional Variational Inequality and Nonlinear Complementarity Problems: A Survey of Theory, Algorithms, and Applications*, Mathematical Programming, Vol. 48, pp. 161–220, 1990.
10. Fischer, A., *A Special Newton-Type Optimization Method*, Optimization, Vol. 24, pp. 269–284, 1992.
11. Robinson, S. M., *Normal Maps Induced by Linear Transformations*, Mathematics of Operations Research, Vol. 17, pp. 691–714, 1992.
12. Mifflin, R., *Semismooth and Semiconvex Functions in Constrained Optimization*, SIAM Journal on Control and Optimization, Vol. 15, pp. 957–972, 1977.
13. Polak, E., *Optimization: Algorithms and Consistent Approximations*, Springer, New York, NY, 1997.
14. Qi, L., and Sun, J., *A Nonsmooth Version of Newton's Method*, Mathematical Programming, Vol. 58, pp. 353–367, 1993.
15. Gowda, M. S., and Sznajder, R., *Weak Univalence and Connectedness of Inverse Images of Continuous Functions*, Mathematics of Operations Research, Vol. 24, pp. 255–261, 1999.
16. Ravindran, G., and Gowda, M. S., *Regularization of P_0-Functions in Box-Constrained Variational Inequality Problems*, SIAM Journal on Optimization, Vol. 11, pp. 748–760, 2000.
17. Qi, L., and Chen, X., *A Globally Convergent Successive Approximation Method for Severely Nonsmooth Equations*, SIAM Journal on Control and Optimization, Vol. 33, pp. 402–418, 1995.
18. Chen, C., and Mangasarian, O. L., *A Class of Smoothing Functions for Nonlinear and Mixed Complementarity Problems*, Computational Optimization and Applications, Vol. 5, pp. 97–138, 1996.
19. Qi, L., Sun, D., and Zhou, G., *A New Look at Smoothing Newton Methods for Nonlinear Complementarity Problems and Box-Constrained Variational Inequalities*, Mathematical Programming, Vol. 87, pp. 1–35, 2000.
20. Chen, B., and Harker, P. T., *A Noninterior-Point Continuation Method for Linear Complementarity Problems*, SIAM Journal on Matrix Analysis and Applications, Vol. 14, pp. 1168–1190, 1993.
21. Kanzow, C., *Some Noninterior Continuation Methods for Linear Complementarity Problems*, SIAM Journal on Matrix Analysis and Applications, Vol. 17, pp. 851–868, 1996.
22. Smale, S., *Algorithms for Solving Equations*, Proceedings of the International Congress of Mathematicians, Berkeley, California, pp. 172–195, 1986.
23. Zang, I., *A Smoothing-Out Technique for Min–Max Optimization*, Mathematical Programming, Vol. 19, pp. 61–71, 1980.
24. Gabriel, S. A., and Moré, J. J., *Smoothing of Mixed Complementarity Problems*, Complementarity and Variational Problems: State of the Art, Edited by

M. C. Ferris and J. S. Pang, SIAM, Philadelphia, Pennsylvania, pp. 105–116, 1997.

25. Ortega, J. M., and Rheinboldt, W. C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, NY, 1970.
26. Ferris, M. C., and Pang, J. S., *Engineering and Economic Applications of Complementarity Problems*, SIAM Review, Vol. 39, pp. 669–713, 1997.
27. Kanzow, C., and Jiang, H., *A Continuation Method for (Strongly) Monotone Variational Inequalities*, Mathematical Programming, Vol. 81, pp. 140–157, 1998.
28. Kuntz, L., and Scholtes, S., *Structural Analysis of Nonsmooth Mappings, Inverse Functions, and Metric Projections*, Journal of Mathematical Analysis and Applications, Vol. 188, pp. 346–386, 1994.
29. Lloyd, N. G., *Degree Theory*, Cambridge University Press, Cambridge, England, 1978.
30. Facchinei, F., and Pang, J. S., *Total Stability of Variational Inequalities*, Preprint, Department of Mathematical Sciences, Johns Hopkins University, Baltimore, Maryland, 1998.
31. Burke, J., and Xu, S., *The Global Linear Convergence of a Noninterior Path-Following Algorithm for Linear Complementarity Problem*, Mathematics of Operations Research, Vol. 23, pp. 719–734, 1998.
32. Chen, B., and Chen, X., *A Global and Local Superlinear Continuation-Smoothing Method for $P_0 + R_0$ and Monotone NCP*, SIAM Journal on Optimization, Vol. 9, pp. 624–645, 1999.
33. Chen, B., and Harker, P. T., *Smooth Approximations to Nonlinear Complementarity Problems*, SIAM Journal on Optimization, Vol. 7, pp. 403–420, 1997.
34. Chen, B., and Xiu, N., *A Global Linear and Local Quadratic Noninterior Continuation Method for Nonlinear Complementarity Problems Based on the Chen–Mangasarian Smoothing Function*, SIAM Journal on Optimization, Vol. 9, pp. 605–623, 1999.
35. Qi, L., *Convergence Analysis of Some Algorithms for Solving Nonsmooth Equations*, Mathematics of Operations Research, Vol. 18, pp. 227–244, 1993.