SIAM J. OPTIM.
Vol. 9, No. 2, pp. 388–413

A NEW UNCONSTRAINED DIFFERENTIABLE MERIT FUNCTION FOR BOX CONSTRAINED VARIATIONAL INEQUALITY PROBLEMS AND A DAMPED GAUSS–NEWTON METHOD*

DEFENG SUN† AND ROBERT S. WOMERSLEY†

Abstract. In this paper we propose a new unconstrained differentiable merit function f for box constrained variational inequality problems VIP(l, u, F). We study various desirable properties of this new merit function f and propose a Gauss–Newton method in which each step requires only the solution of a system of linear equations. Global and superlinear convergence results for VIP(l, u, F) are obtained. Key results are the boundedness of the level sets of the merit function for any uniform P-function and the superlinear convergence of the algorithm without a nondegeneracy assumption. Numerical experiments confirm the good theoretical properties of the method.

Key words. variational inequality problems, box constraints, merit functions, Gauss–Newton method, superlinear convergence

AMS subject classifications. 90C33, 90C30, 65H10

PII. S1052623496314173

1. Introduction.

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable mapping and S be a nonempty closed convex set in $\mathbb{R}^n$. The variational inequality problem, denoted by VIP(S, F), is to find a vector $x \in S$ such that

$$F(x)^T (y - x) \geq 0 \quad \text{for all } \ y \in S. \tag{1.1}$$

A box constrained variational inequality problem, denoted VIP(l, u, F), has

$$S = \{x \in \mathbb{R}^n \mid \, l \leq x \leq u\}, \tag{1.2}$$

where $l_i \in \mathbb{R} \cup \{-\infty\}$, $u_i \in \mathbb{R} \cup \{+\infty\}$, and $l_i < u_i$, $i = 1, \ldots, n$. In some papers, e.g., [2, 7], VIP(l, u, F) is called the mixed complementarity problem. Further, if $S = \mathbb{R}^n_+$, VIP(S, F) reduces to the nonlinear complementarity problem, denoted NCP(F), which is to find $x \in \mathbb{R}^n$ such that

$$x \geq 0, \quad F(x) \geq 0, \quad x^T F(x) = 0. \tag{1.3}$$

Two comprehensive surveys of variational inequality problems and nonlinear complementarity problems are [23] and [34].

Recently much effort has been made to derive *merit functions* for VIP(S, F) and then to use these functions to develop solution methods. Formally, we say that a function $h : X \to [0, \infty)$ is a merit function for VIP(S, F) on a set X (typically $X = \mathbb{R}^n$ or $X = S$) provided $h(x) \geq 0$ for all $x \in X$ and $x \in X$ satisfies (1.1) if and only if $h(x) = 0$. Then, we may reformulate VIP(S, F) as the minimization problem

$$\min_{x \in X} \quad h(x). \tag{1.4}$$

*Received by the editors December 27, 1996; accepted for publication (in revised form) April 17, 1998; published electronically March 17, 1999. This research was supported by the Australian Research Council.

http://www.siam.org/journals/siopt/9-2/31417.html

†School of Mathematics, University of New South Wales, Sydney, NSW 2052, Australia (sun@maths.unsw.edu.au, R.Womersley@unsw.edu.au).

Recent developments of this area are summarized in [20].

It is well known [9] that $x \in \mathbb{R}^n$ solves VIP(S, F) if and only if x is a solution of the equation

$$H(x) := x - \Pi_S[x - \alpha^{-1}F(x)] = 0 \tag{1.5}$$

for an arbitrary positive constant α. Here Π_S is the orthogonal projection operator onto S. An obvious merit function for (1.1) is

$$h(x) := \frac{1}{2}\|H(x)\|^2. \tag{1.6}$$

We can find a solution of (1.1) by solving (1.4) with $X = \mathbb{R}^n$ or $X = S$. Unfortunately the function h defined by (1.5) and (1.6) is not continuously differentiable, so gradient-based methods cannot be used directly. Nevertheless global and superlinear convergence properties have been obtained under some regularity conditions [32, 33, 35, 36]. Another approach based on nonsmooth equations and nonsmooth merit functions is Ralph's path search method [41, 7].

Recent interests have focused on (unconstrained) differentiable merit functions. Early differentiable merit functions such as the regularized gap function [19] are constrained ones. By applying the Moreau–Yosida regularization to some gap functions, Yamashita and Fukushima [48] proposed unconstrained differentiable merit functions for (1.1). These functions possess nice theoretical properties but are not easy to evaluate in general. Peng [37] showed that the difference of two regularized gap functions constitutes an unconstrained differentiable merit function for VIP(S, F). Later, Yamashita, Taji, and Fukushima [49] extended the idea of Peng [37] and investigated some important properties related to this merit function. Specifically, the latter authors considered the function $h_{\alpha\beta} : \mathbb{R}^n \to \mathbb{R}$ defined by

$$h_{\alpha\beta}(x) := f_\alpha(x) - f_\beta(x), \tag{1.7}$$

where α and β are arbitrary positive parameters such that $\alpha < \beta$ and f_α is the regularized gap function

$$f_\alpha(x) := \max_{y \in S} \left\{ F(x)^T(x - y) - \frac{\alpha}{2}\|x - y\|^2 \right\}. \tag{1.8}$$

(The function f_β is defined similarly with α replaced by β.) In the special case $\beta = \alpha^{-1}$ and $\alpha < 1$ in (1.7), the function $h_{\alpha\beta}$ reduces to the merit function studied by Peng [37]. The function $h_{\alpha\beta}$ defined by (1.7) is called the D-gap function. Based on this merit function, globally and superlinearly convergent Newton-type methods for solving VIP(S, F) have been proposed under the assumption that F is a strongly monotone function [44]. It was pointed out by Peng and Yuan [38] that when $S = \mathbb{R}^n_+$, $\beta = \alpha^{-1}$, and $0 < \alpha < 1$, the function $h_{\alpha\beta}$ is actually the implicit Lagrangian function

$$\begin{aligned} m_\alpha(x) := x^T F(x) + \frac{\alpha}{2}\big(&\|[x - \alpha^{-1}F(x)]_+\|^2 - \|x\|^2 \\ &+ \|[F(x) - \alpha^{-1}x]_+\|^2 - \|F(x)\|^2\big) \end{aligned} \tag{1.9}$$

introduced by Mangasarian and Solodov [30] for the nonlinear complementarity problem (1.3). Here $[z]_+$ denotes the vector with components $\max\{z_i, 0\}, i = 1, \ldots, n$. The function $m_\alpha(x)$ is one of the many unconstrained differentiable merit functions for NCP(F) and its various properties were further studied in [11, 24, 28, 37, 46, 47, 49].

Another well-studied unconstrained differentiable merit function for NCP(F) has the form

$$\theta(x) := \frac{1}{2}\sum_{i=1}^{n} \phi(x_i, F_i(x))^2, \tag{1.10}$$

where $\phi : \mathbb{R}^2 \to \mathbb{R}$ is the function

$$\phi(a, b) := \sqrt{a^2 + b^2} - (a + b) \tag{1.11}$$

introduced by Fischer [16] but attributed to Burmeister and called the Fischer–Burmeister function. This merit function θ has been much studied and used in solving nonlinear complementarity problems [6, 13, 14, 18, 21, 24, 25, 26, 45] (see [17] for a survey). In particular, based on this merit function, globally and superlinearly convergent Newton-type methods for NCP(F) were given in [6] under the assumption that F is a uniform P-function, which is a weaker condition than the assumption that F is a strongly monotone function. Unlike the implicit Lagrangian function m_α, the nice properties of the merit function θ based on the Fischer–Burmeister function cannot be naturally generalized to VIP(S, F).

In this paper we study new unconstrained merit functions for the box constrained variational inequality problem VIP(l, u, F) where S is of the form (1.2). Despite its special structure, VIP(l, u, F) has many applications in engineering, economics, and sciences. An available unconstrained differentiable merit function for VIP(l, u, F) is the D-gap function $h_{\alpha\beta}$. However, when reduced to NCP(F), the D-gap function $h_{\alpha\beta}$ with $\beta = \alpha^{-1}$ and $\alpha \in (0, 1)$ becomes the implicit Lagrangian function m_α. This merit function suffers from the drawback that it needs more restrictive assumptions to get globally and superlinearly convergent methods for NCP(F) than the merit function $\theta(x)$ based on the Fischer–Burmeister function does. This motivates the investigation of other unconstrained differentiable merit functions which need less restrictive assumptions. Throughout this paper we adopt the convention that $\pm\infty \times 0 = 0$. Then it is easy to see that VIP(l, u, F) is equivalent to its Karush–Kuhn–Tucker (KKT) system

$$\begin{array}{c} v - w = F(x), \\ x_i - l_i \geq 0, \quad v_i \geq 0, \quad (x_i - l_i)v_i = 0, \quad i = 1, \ldots, n, \\ u_i - x_i \geq 0, \quad w_i \geq 0, \quad (u_i - x_i)w_i = 0, \quad i = 1, \ldots, n. \end{array} \tag{1.12}$$

If $x \in \mathbb{R}^n$ solves VIP(l, u, F), then $(x, v, w) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$ with $v = [F(x)]_+$ and $w = [-F(x)]_+$ solves the KKT system (1.12). Conversely, if $(x, v, w) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$ solves the KKT system (1.12), then x solves VIP(l, u, F). Define $E : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{3n}$ as

$$E(x, v, w) := \begin{pmatrix} v - w - F(x) \\ \phi(x_i - l_i, v_i),\ i = 1, \ldots, n \\ \phi(u_i - x_i, w_i),\ i = 1, \ldots, n \end{pmatrix}.$$

Then an obvious unconstrained differentiable merit function for the KKT system (1.12) is

$$\xi(x, v, w) := \frac{1}{2}\|E(x, v, w)\|^2.$$

The merit function ξ for VIP(l,u,F) has many good properties; for example, see [10]. However, it also suffers from several drawbacks. A disadvantage of this merit function is that the level sets $L_c(\xi)$ of ξ are in general not bounded for all nonnegative numbers c. Here the level sets $L_c(g)$ of $g : \mathbb{R}^m \to \mathbb{R}$ are

$$L_c(g) := \{z \in \mathbb{R}^m |\ g(z) \le c\}.$$

This can easily be shown by taking $l_i = -1$, $u_i = 1$, and $v_i = w_i \to \infty$, $i = 1, \dots, n$. The unboundedness of the level sets could allow the sequence of iterates to diverge to infinity. This unfavorable property is caused by introducing the variables v and w. So it appears better to consider VIP(l,u,F) in its original space instead of in the larger-dimensional space.

We propose a new merit function which has bounded level sets for any uniform P-function (see section 3) and establish superlinear convergence of a damped Gauss–Newton algorithm without a nondegeneracy assumption (see section 7). First define $\psi : \mathbb{R}^2 \to \mathbb{R}_+$ as

$$\psi(a,b) := ([-\phi(a,b)]_+)^2 + ([-a]_+)^2, \tag{1.13}$$

where $\phi(a,b)$ is the Fischer–Burmeister function defined in (1.11).

The following proposition is simple but is essential to the discussion of this paper.

PROPOSITION 1.1. *The function ψ defined by* (1.13) *is continuously differentiable on the whole space of $\mathbb{R}^2$ and has the property*

$$\psi(a,b) = 0 \Longleftrightarrow a \ge 0, \quad ab_+ = 0. \tag{1.14}$$

Proof. Since both $([-\phi(a,b)]_+)^2$ and $([-a]_+)^2$ are continuously differentiable, ψ is continuously differentiable. By considering the fact that

$$a \ge 0,\ \sqrt{a^2+b^2} - (a+b) \ge 0 \Longleftrightarrow a \ge 0,\ ab_+ = 0,$$

we get (1.14) easily. □

After simple computation we can see that the function ψ can be rewritten as

$$\psi(a,b) = \varphi(a,b)^2, \tag{1.15}$$

where

$$\begin{aligned} \varphi(a,b) &:= \begin{cases} [-\phi(a,b)]_+ & \text{if } a \ge 0, \\ a & \text{otherwise}, \end{cases} \\ &= \begin{cases} -\phi(a,[b]_+) & \text{if } a \ge 0, \\ a & \text{otherwise}, \end{cases} \\ &= \min\{[-\phi(a,b)]_+, a\}. \end{aligned} \tag{1.16}$$

Such equivalent expressions for ψ and φ will be useful in the following discussions. Note that φ is not differentiable at $(0,b)$ for any $b \le 0$ and at $(a,0)$ for any $a \ge 0$, but ψ is continuously differentiable.

Since for any $b \in \mathbb{R}$

$$\lim_{a\to\infty} \phi(a,b) = -b,$$

it is natural to define

$$\phi(+\infty, b) = -b.$$

Thus, for any $b \in \mathbb{R}$ we define

$$\psi(+\infty, b) = ([b]_+)^2.$$

Based on ψ, we define $f : \mathbb{R}^n \to \mathbb{R}$ as

$$f(x) := \frac{1}{2}\left[\sum_{i=1}^{n} \psi(x_i - l_i, F_i(x)) + \sum_{i=1}^{n} \psi(u_i - x_i, -F_i(x))\right]. \tag{1.17}$$

This function is an unconstrained differentiable merit function for VIP(l, u, F) (see Theorem 2.2) and has many good properties. When VIP(l, u, F) reduces to NCP(F), i.e., $l_i = 0$ and $u_i = +\infty$, $i = 1, \ldots, n$, the function (1.17) becomes

$$f(x) = \frac{1}{2}\sum_{i=1}^{n} \eta(x_i, F_i(x)), \tag{1.18}$$

where for any $(a, b) \in \mathbb{R}^2$

$$\eta(a, b) := \begin{cases} ((a+b) - \sqrt{a^2 + b^2})^2 & \text{if } a \geq 0, b \geq 0, \\ b^2 & \text{if } a \geq 0, b < 0, \\ a^2 & \text{if } a < 0, b \geq 0, \\ a^2 + b^2 & \text{if } a < 0, b < 0. \end{cases} \tag{1.19}$$

The organization of this paper is as follows. In the next section we study some preliminary properties of the new merit function. In section 3 we study the conditions under which the level sets of f are bounded. In section 4 we give conditions which ensure that a stationary point of f is a solution of VIP(l, u, F). Section 5 is devoted to the nonsingularity of the iteration matrices. In section 6 we state the algorithm. We analyze the convergence properties of the algorithm in section 7 and give numerical results in section 8. Some concluding remarks are given in section 9.

For a continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}^n$, we denote the Jacobian of F at $x \in \mathbb{R}^n$ by $F'(x)$, whereas the transposed Jacobian is $\nabla F(x)$. Throughout $\|\cdot\|$ denotes the Euclidean norm. If $\mathcal{J}$ and $\mathcal{K}$ are index sets such that $\mathcal{J}, \mathcal{K} \subseteq \{1, \ldots, m\}$, we denote by $W_{\mathcal{JK}}$ the $|\mathcal{J}| \times |\mathcal{K}|$ submatrix of W consisting of entries W_{jk}, $j \in \mathcal{J}$, $k \in \mathcal{K}$. If $W_{\mathcal{JJ}}$ is nonsingular, we denote by $W/W_{\mathcal{JJ}}$ the Schur complement of $W_{\mathcal{JJ}}$ in W, i.e., $W/W_{\mathcal{JJ}} := W_{\mathcal{KK}} - W_{\mathcal{KJ}}W_{\mathcal{JJ}}^{-1}W_{\mathcal{JK}}$, where $\mathcal{K} = \{1, \ldots, m\}\backslash\mathcal{J}$. If w is an m vector, we denote by $w_{\mathcal{J}}$ the subvector with components $j \in \mathcal{J}$.

2. Some preliminaries. By noting that $x \in \mathbb{R}^n$ solves VIP(l, u, F) if and only if $H(x) = 0$ and that $\pm\infty \times 0 = 0$, we have the following results directly.

LEMMA 2.1. *A vector $x \in \mathbb{R}^n$ solves VIP(l, u, F) if and only if it satisfies*

$$l_i \leq x_i \leq u_i,\ (x_i - l_i)[F_i(x)]_+ = 0,\ (u_i - x_i)[-F_i(x)]_+ = 0,\ i = 1, \ldots, n. \tag{2.1}$$

THEOREM 2.2. *The function $f(x)$ defined by* (1.17) *is nonnegative on $\mathbb{R}^n$, and $f(x) = 0$ if and only if $x \in \mathbb{R}^n$ solves VIP(l, u, F). In addition, if F is continuously differentiable, then f is also continuously differentiable.*

Proof. Since $\psi(a, b) \geq 0$ for all $(a, b) \in \mathbb{R}^2$, $f(x)$ is nonnegative on $\mathbb{R}^n$. From Proposition 1.1 and Lemma 2.1, $x \in \mathbb{R}^n$ solves VIP(l, u, F) if and only if it satisfies

$$\psi(x_i - l_i, F_i(x)) = 0, \quad \psi(u_i - x_i, -F_i(x)) = 0, \quad i = 1, \ldots, n.$$

Thus $f(x) = 0$ if and only if $x \in \mathbb{R}^n$ solves VIP(l, u, F). Moreover, it is easy to see that if F is continuously differentiable, then so is F, as ψ is continuously differentiable by Proposition 1.1. ◻

Note that although f is continuously differentiable, it is not twice continuously differentiable and its gradient ∇f may not be locally Lipschitz continuous. For such an example we refer to the one-dimensional function given in section 1 of [44]. So a direct use of Newton's method for minimizing $f(x)$ may fail. However, we can still expect to obtain globally and superlinearly convergent Newton-type methods for minimizing $f(x)$. The tool used here is semismoothness.

Semismoothness was originally introduced by Mifflin [31] 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 (see [31]). In [40] Qi and Sun extended the definition of semismooth functions to $G : \mathbb{R}^n \to \mathbb{R}^m$. A locally Lipschitz continuous vector valued function $G : \mathbb{R}^n \to \mathbb{R}^m$ has a generalized Jacobian $\partial G(x)$ as in Clarke [5]. G is said to be *semismooth* at $x \in \mathbb{R}^n$ if

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

exists for any $h \in R^n$. It has been proved in [40] that G is semismooth at x if and only if all its component functions are. Also $G'(x; h)$, the directional derivative of G at x in the direction h, exists for any $h \in \mathbb{R}^n$ and is equal to the above limit if G is semismooth at x.

LEMMA 2.3 (see [40]). *Suppose that $G : \mathbb{R}^n \to \mathbb{R}^m$ is a locally Lipschitzian function and is semismooth at x. Then*

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

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

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

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

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

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

(Note that in [40] and [39] different names for strong semismoothness are used.) A function G is said to be a (strongly) semismooth function if it is (strongly) semismooth everywhere.

In [39] Qi defined the generalized Jacobian

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

This concept will be used in the design of our algorithm.

Let $\varphi : \mathbb{R}^2 \to \mathbb{R}$ be the function defined by (1.16) and define $\Psi,\ \Phi : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\Psi_i(x) := \varphi(x_i - l_i, F_i(x)) = \min\{[-\phi(x_i - l_i, F_i(x))]_+, x_i - l_i\}$$

and

$$\Phi_i(x) := \varphi(u_i - x_i, -F_i(x)) = \min\{[-\phi(u_i - x_i, -F_i(x))]_+, u_i - x_i\}$$

for $i = 1, \ldots, n$. Define $G : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\begin{aligned} G_i(x) \quad &:= \sqrt{\Psi_i(x)^2 + \Phi_i(x)^2} \\ &= \begin{cases} (l_i - x_i) & \text{if } x_i < l_i \ \&\ F_i(x) \geq 0, \\ -\phi(x_i - l_i, F_i(x)) & \text{if } l_i \leq x_i \leq u_i \ \&\ F_i(x) \geq 0, \\ \sqrt{(u_i - x_i)^2 + \phi(x_i - l_i, F_i(x))^2} & \text{if } x_i > u_i \ \&\ F_i(x) \geq 0, \\ \sqrt{(x_i - l_i)^2 + \phi(u_i - x_i, -F_i(x))^2} & \text{if } x_i < l_i \ \&\ F_i(x) < 0, \\ -\phi(u_i - x_i, -F_i(x)) & \text{if } l_i \leq x_i \leq u_i \ \&\ F_i(x) < 0, \\ (x_i - u_i) & \text{if } x_i > u_i \ \&\ F_i(x) < 0 \end{cases} \end{aligned} \tag{2.2}$$

for $i = 1, \ldots, n$ and where $\phi(\cdot)$ is the Fischer–Burmeister function defined in (1.11). Then the merit function $f(x)$ defined by (1.17) can be rewritten as

$$f(x) = \frac{1}{2}\|G(x)\|^2. \tag{2.3}$$

PROPOSITION 2.4. *Suppose that F is continuously differentiable at $x \in \mathbb{R}^n$. Then G is semismooth at x. Moreover if F' is locally Lipschitz continuous around x, then G is strongly semismooth at x.*

Proof. We need only to prove that for each i, G_i is (strongly) semismooth at x under the assumptions. First note that $\phi(\cdot)$ is a strongly semismooth function [18, Lemma 20] and $[\cdot]_+ : \mathbb{R} \to \mathbb{R}_+$ is strongly semismooth everywhere. Then by Theorem 19 in Fischer [18], which states that the composition of strongly semismooth functions is a strongly semismooth function, we know that $[-\phi(\cdot)]_+$ is strongly semismooth everywhere. It is easy to see that $\min\{\cdot,\cdot\} : \mathbb{R}^2 \to \mathbb{R}$ is strongly semismooth everywhere. Thus, by using Theorem 19 in Fischer [18] again, $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is a strongly semismooth function. Then by Theorem 5 in Mifflin [31], which states that the composition of semismooth functions is semismooth, we know that Ψ_i and Φ_i are semismooth at x, and because $\sqrt{\alpha^2 + \beta^2}$ is a strongly semismooth function of α and β, G_i is semismooth at x. If F' is locally Lipschitz continuous, then $(y_i - l_i, F_i(y))$ and $(u_i - y_i, -F_i(y))$ are strongly semismooth at x. Thus, by Theorem 19 in Fischer [18] we know that Ψ_i and Φ_i are strongly semismooth at x. So, G_i is strongly semismooth at x. □

We need the following definitions concerning matrices and functions.

DEFINITION 2.5. *A matrix $W \in \mathbb{R}^{n\times n}$ is called a*

- *P_0-matrix if each of its principal minors is nonnegative;*
- *P-matrix if each of its principal minors is positive.*

Obviously a positive semidefinite matrix is a P_0-matrix and a positive definite matrix is a P-matrix.

DEFINITION 2.6. *A function $F : \mathbb{R}^n \to \mathbb{R}^n$*

- *is a P_0-function if for every x and y in $\mathbb{R}^n$ with $x \neq y$ there is an index i such that*

$$x_i \neq y_i, \quad (x_i - y_i)(F_i(x) - F_i(y)) \geq 0;$$

- *is a P-function if for every x and y in $\mathbb{R}^n$ with $x \neq y$ there is an index i such that*

$$x_i \neq y_i, \quad (x_i - y_i)(F_i(x) - F_i(y)) > 0;$$

- *is a uniform P-function if there exists a positive constant μ such that for every x and y in $\mathbb{R}^n$ there is an index i such that*

$$(x_i - y_i)(F_i(x) - F_i(y)) \geq \mu \|x - y\|^2;$$

- *is a monotone function if for every x and y in $\mathbb{R}^n$*

$$(x - y)^T (F(x) - F(y)) \geq 0;$$

- *is a strongly monotone function if there exists a positive constant μ such that for every x and y in $\mathbb{R}^n$*

$$(x - y)^T (F(x) - F(y)) \geq \mu \|x - y\|^2.$$

It is known that every strongly monotone function is a uniform P-function and every monotone function is a P_0-function. Furthermore, the Jacobian of a continuously differentiable P_0-function (uniform P-function) is a P_0-matrix (P-matrix).

3. Bounded level sets. In this section we study the conditions under which the level sets of the merit function f are bounded. Since for any $c \in \mathbb{R} \cup \{-\infty\}$, $d \in \mathbb{R} \cup \{\infty\}$ with $c < d$ and $a \in \mathbb{R}$, $\Pi_{[c,d]\cap\mathbb{R}}(a) = \Pi_{[c,d]}(a)$, for the sake of simplicity we use $\Pi_{[c,d]}(a)$ instead of $\Pi_{[c,d]\cap\mathbb{R}}(a)$ to represent the orthogonal projection of a onto $[c,d] \cap \mathbb{R}$. Boundedness results for NCP(F) with uniform P-functions have been established by Jiang [24], Facchinei and Soares [14], and De Luca, Facchinei, and Kanzow [6]. Here these results are extended to VIP(l, u, F).

The following lemma is essential to develop conditions which ensure bounded level sets.

LEMMA 3.1. *For four given numbers $a, b \in \mathbb{R}$, $c \in \mathbb{R} \cup \{-\infty\}$, and $d \in \mathbb{R} \cup \{\infty\}$ with $c < d$, we have*

$$\gamma_1 \left| a - \Pi_{[c,d]}[a - b] \right|^2 \leq \psi(a - c, b) + \psi(d - a, -b) \leq \gamma_2 \left| a - \Pi_{[c,d]}[a - b] \right|^2 \tag{3.1}$$

with $\gamma_1 = 1/(6 + 4\sqrt{2})$ and $\gamma_2 = 12 + 8\sqrt{2}$.

Proof. First, from Tseng [45], for any two numbers $v, w \in \mathbb{R}$ we have

$$\frac{1}{2 + \sqrt{2}} |\min\{v, w\}| \leq |\phi(v, w)| \leq (2 + \sqrt{2}) |\min\{v, w\}|. \tag{3.2}$$

Then by using the second equality of (1.16), if $v \geq 0$ we have

$$\frac{1}{2 + \sqrt{2}} |\min\{v, w_+\}| \leq |\varphi(v, w)| \leq (2 + \sqrt{2}) |\min\{v, w_+\}| \tag{3.3}$$

and if $v < 0$ we have

$$|\varphi(v,w)| = |v| = |\min\{v, w_+\}|. \tag{3.4}$$

Thus, for all $(v,w) \in \mathbb{R}^2$ we have

$$\frac{1}{6+4\sqrt{2}}|\min\{v,w_+\}|^2 \le \varphi(v,w)^2 \le (6+4\sqrt{2})|\min\{v,w_+\}|^2.$$

Let

$$\begin{aligned} t &:= |\min\{a-c, b_+\}|^2 + |\min\{d-a, [-b]_+\}|^2 \\ &= \begin{cases} |\min\{a-c,b\}|^2 + (d-a)^2 & \text{if } b \ge 0 \;\&\; a \ge d, \\ |\min\{a-c,b\}|^2 & \text{if } b \ge 0 \;\&\; a < d, \\ |\min\{d-a,-b\}|^2 & \text{if } b < 0 \;\&\; a \ge c, \\ (a-c)^2 + |\min\{d-a,-b\}|^2 & \text{if } b < 0 \;\&\; a < c. \end{cases} \end{aligned} \tag{3.5}$$

Then

$$\frac{1}{6+4\sqrt{2}}t \le \psi(a-c,b) + \psi(d-a,-b) \le (6+4\sqrt{2})t. \tag{3.6}$$

Denote

$$r := |a - \Pi_{[c,d]}[a-b]|^2 = \begin{cases} (a-c)^2 & \text{if } a-b \le c, \\ b^2 & \text{if } c < a-b < d, \\ (a-d)^2 & \text{if } a-b \ge d. \end{cases}$$

Next we prove that

$$r \le t \le 2r. \tag{3.7}$$

First, if either $b \ge 0$ and $a < d$ or $b < 0$ and $a > c$, then we can directly verify that $r = t$. Next, we consider the other two cases.

Case 1. $b \ge 0$ and $a \ge d$. Then

$$t = |\min\{a-c,b\}|^2 + (d-a)^2 = (d-a)^2 + \begin{cases} (a-c)^2 & \text{if } b \ge a-c, \\ b^2 & \text{if } b < a-c. \end{cases}$$

After simple computation we get

$$r \le t \le 2r.$$

Case 2. $b < 0$ and $a \le c$. Then

$$t = (a-c)^2 + |\min\{d-a,-b\}|^2 = (a-c)^2 + \begin{cases} (d-a)^2 & \text{if } d-a \le -b, \\ b^2 & \text{if } d-a > -b. \end{cases}$$

Again, after simple computation we get

$$r \le t \le 2r.$$

Overall we have proved (3.7). By combining (3.6) and (3.7), we get (3.1). $\square$

THEOREM 3.2. *Suppose that for any sequence $\{x^k\}$ with $\|x^k\| \to \infty$ there exists an index $i \in \{1, \dots, n\}$ independent of k such that*

$$|x_i^k - \Pi_{[l_i,u_i]}[x_i^k - F_i(x^k)]| \to \infty. \tag{3.8}$$

Then for any $c \geq 0$, $L_c(f)$ is bounded. In particular, if S is bounded or if F is a uniform P-function, then $L_c(f)$ is bounded.

Proof. Suppose that for some given $c \geq 0$, $L_c(f)$ is unbounded. Then there exists a sequence $\{x^k\}$ diverging to infinity and satisfying

$$f(x^k) \leq c.$$

But, on the other hand, from Lemma 3.1 and the assumption that there exists an index i, independent of k, such that (3.8) holds, we have

$$f(x^k) \geq \frac{1}{2}\gamma_1 |x_i^k - \Pi_{[l_i,u_i]}[x_i^k - F_i(x^k)]|^2 \to \infty,$$

where $\gamma_1 = 1/(6 + 4\sqrt{2})$. This is a contradiction. So for any $c \geq 0$, $L_c(f)$ is bounded if (3.8) holds.

By noting that if S is bounded then (3.8) holds automatically, we can conclude that for any given $c \geq 0$, $L_c(f)$ is bounded.

If F is a uniform P-function, then by [14] for any sequence $\{x^k\}$ with $\|x^k\| \to \infty$ there exists an index $i \in \{1, \dots, n\}$ independent of k such that

$$|x_i^k| \to \infty, \quad |F_i(x^k)| \to \infty,$$

which, in turn, implies (3.8). This completes the proof. $\square$

4. Stationary point conditions. In general a stationary point of a merit function may not be a solution of the underlying problem. Many people [6, 11, 14, 21, 24, 25, 26, 29, 47] have studied the conditions under which a stationary point is a solution of NCP(F). In this section we study the conditions under which a stationary point of (1.17) is a solution of VIP(l, u, F). Similar work has been done in [10, 27] for box constrained variational inequality problems.

First let us study the structure of $\partial_B G_i(x)$, where $G_i(\cdot)$, $i = 1, \dots, n$ are defined in (2.2). Denote by e_i the ith unit row vector of $\mathbb{R}^n$, $i = 1, \dots, n$. For any $x \in \mathbb{R}^n$ we discuss five cases, each of which includes three subcases.

Case 1. $x_i < l_i$.

Case 1.1 $F_i(x) > 0$. Then $G_i(x) = l_i - x_i$ and $\partial_B G_i(x) = \{-e_i\}$.

Case 1.2. $F_i(x) < 0$. Then

$$G_i(x) = \sqrt{(x_i - l_i)^2 + \phi(u_i - x_i, -F_i(x))^2},$$
$$\partial_B G_i(x) = \{\alpha_i(-e_i) + \beta_i(-F_i'(x))\},$$

where

$$\alpha_i = \frac{l_i - x_i}{G_i(x)} + \frac{\phi(u_i - x_i, -F_i(x))}{G_i(x)} \left(\frac{u_i - x_i}{\sqrt{(u_i - x_i)^2 + (-F_i(x))^2}} - 1 \right),$$

$$\beta_i = \frac{\phi(u_i - x_i, -F_i(x))}{G_i(x)} \left(\frac{-F_i(x)}{\sqrt{(u_i - x_i)^2 + (-F_i(x))^2}} - 1 \right).$$

Case 1.3. $F_i(x) = 0$. Then $G_i(x) = l_i - x_i$ and $\partial_B G_i(x) = \{-e_i\}$.

Case 2. $x_i > u_i$.

Case 2.1. $F_i(x) > 0$. Then

$$G_i(x) = \sqrt{(u_i - x_i)^2 + \phi(x_i - l_i, F_i(x))^2},$$
$$\partial_B G_i(x) = \{\alpha_i e_i + \beta_i F_i'(x)\},$$

where

$$\alpha_i = \frac{x_i - u_i}{G_i(x)} + \frac{\phi(x_i - l_i, F_i(x))}{G_i(x)} \left(\frac{x_i - l_i}{\sqrt{(x_i - l_i)^2 + F_i(x)^2}} - 1 \right),$$
$$\beta_i = \frac{\phi(x_i - l_i, F_i(x))}{G_i(x)} \left(\frac{F_i(x)}{\sqrt{(x_i - l_i)^2 + F_i(x)^2}} - 1 \right).$$

Case 2.2. $F_i(x) < 0$. Then $G_i(x) = x_i - u_i$ and $\partial_B G_i(x) = \{e_i\}$.

Case 2.3. $F_i(x) = 0$. Then $G_i(x) = x_i - u_i$ and $\partial_B G_i(x) = \{e_i\}$.

Case 3. $l_i < x_i < u_i$.

Case 3.1. $F_i(x) > 0$. Then

$$G_i(x) = -\phi(x_i - l_i, F_i(x)),$$
$$\partial_B G_i(x) = \{\alpha_i e_i + \beta_i F_i'(x)\},$$

where

$$\alpha_i = 1 - \frac{x_i - l_i}{\sqrt{(x_i - l_i)^2 + F_i(x)^2}} \text{ and } \beta_i = 1 - \frac{F_i(x)}{\sqrt{(x_i - l_i)^2 + F_i(x)^2}}.$$

Case 3.2. $F_i(x) < 0$. Then

$$G_i(x) = -\phi(u_i - x_i, -F_i(x)),$$
$$\partial_B G_i(x) = \{\alpha_i(-e_i) + \beta_i(-F_i'(x))\},$$

where

$$\alpha_i = 1 - \frac{u_i - x_i}{\sqrt{(u_i - x_i)^2 + F_i(x)^2}} \text{ and } \beta_i = 1 - \frac{-F_i(x)}{\sqrt{(u_i - x_i)^2 + F_i(x)^2}}.$$

Case 3.3. $F_i(x) = 0$. Then $G_i(x) = 0$ and $\partial_B G_i(x) \subseteq \{F_i'(x), -F_i'(x)\}$.

Case 4. $x_i = l_i$.

Case 4.1. $F_i(x) > 0$. Then $G_i(x) = 0$ and $\partial_B G_i(x) \subseteq \{e_i, -e_i\}$.

Case 4.2. $F_i(x) < 0$. Then

$$G_i(x) = -\phi(u_i - l_i, -F_i(x)),$$
$$\partial_B G_i(x) = \{\alpha_i(-e_i) + \beta_i(-F_i'(x))\},$$

where

$$\alpha_i = 1 - \frac{u_i - l_i}{\sqrt{(u_i - l_i)^2 + F_i(x)^2}} \text{ and } \beta_i = 1 - \frac{-F_i(x)}{\sqrt{(u_i - l_i)^2 + F_i(x)^2}}.$$

Case 4.3. $F_i(x) = 0$. Then $G_i(x) = 0$ and

$$\partial_B G_i(x) \subseteq \{\alpha_i e_i + \beta_i F_i'(x)\} \cup \{\bar{\alpha}_i(-e_i) + \bar{\beta}_i(-F_i'(x))\},$$

where $\alpha_i, \beta_i \in [0,1]$ satisfy $(\alpha_i - 1)^2 + (\beta_i - 1)^2 = 1$ and $\bar{\alpha}_i, \bar{\beta}_i \in [0,1]$ satisfy $\bar{\alpha}_i^2 + \bar{\beta}_i^2 = 1$.

Case 5. $x_i = u_i$.

Case 5.1. $F_i(x) > 0$. Then

$$G_i(x) = -\phi(u_i - l_i, F_i(x)),$$
$$\partial_B G_i(x) = \{\alpha_i e_i + \beta_i F_i'(x)\},$$

where

$$\alpha_i = 1 - \frac{u_i - l_i}{\sqrt{(u_i - l_i)^2 + F_i(x)^2}} \text{ and } \beta_i = 1 - \frac{F_i(x)}{\sqrt{(u_i - l_i)^2 + F_i(x)^2}}.$$

Case 5.2. $F_i(x) < 0$. Then $G_i(x) = 0$ and $\partial_B G_i(x) \subseteq \{e_i, -e_i\}$.

Case 5.3. $F_i(x) = 0$. Then $G_i(x) = 0$ and

$$\partial_B G_i(x) \subseteq \{\alpha_i(-e_i) + \beta_i(-F_i'(x))\} \cup \{\bar{\alpha}_i e_i + \bar{\beta}_i F_i'(x)\},$$

where $\alpha_i, \beta_i \in [0,1]$ satisfy $(\alpha_i - 1)^2 + (\beta_i - 1)^2 = 1$ and $\bar{\alpha}_i, \bar{\beta}_i \in [0,1]$ satisfy $\bar{\alpha}_i^2 + \bar{\beta}_i^2 =1$.

For any $x \in \mathbb{R}^n$ define the index sets $\mathcal{A}_{jk}(x)$ by

$$\mathcal{A}_{jk}(x) := \{i| \text{ Case } j.k \text{ occurs at } x_i,\ i = 1, \dots, n\},\ j = 1, \dots, 5,\ k = 1, \dots, 3.$$

For example, some $i \in \mathcal{A}_{42}(x)$ means that Case 4.2 occurs at x_i, i.e., $x_i = l_i$ and $F_i(x) < 0$. Furthermore let

$$\mathcal{A}_{31}^{-\infty}(x) := \{i|\ i \in \mathcal{A}_{31}(x) \text{ and } l_i = -\infty,\ i = 1, \dots, n\},$$
$$\mathcal{A}_{32}^{\infty}(x) := \{i|\ i \in \mathcal{A}_{32}(x) \text{ and } u_i = \infty,\ i = 1, \dots, n\},$$
$$\mathcal{A}_{42}^{\infty}(x) := \{i|\ i \in \mathcal{A}_{42}(x) \text{ and } u_i = \infty,\ i = 1, \dots, n\},$$
$$\mathcal{A}_{51}^{-\infty}(x) := \{i|\ i \in \mathcal{A}_{51}(x) \text{ and } l_i = -\infty,\ i = 1, \dots, n\}.$$

For convenience we define the four additional index sets

$$\mathcal{O}(x) := \mathcal{A}_{11}(x) \cup \mathcal{A}_{13}(x) \cup \mathcal{A}_{22}(x) \cup \mathcal{A}_{23}(x),$$
$$\mathcal{P}(x) := \mathcal{A}_{31}^{-\infty} \cup \mathcal{A}_{32}^{\infty} \cup \mathcal{A}_{42}^{\infty} \cup \mathcal{A}_{51}^{-\infty},$$
$$\mathcal{Q}(x) := \mathcal{A}_{33}(x) \cup \mathcal{A}_{41}(x) \cup \mathcal{A}_{43}(x) \cup \mathcal{A}_{52}(x) \cup \mathcal{A}_{53}(x),$$
$$\mathcal{R}(x) := \{1, \dots, n\} \backslash \{\mathcal{O}(x) \cup \mathcal{P}(x) \cup \mathcal{Q}(x)\}.$$

LEMMA 4.1. *For any $x \in \mathbb{R}^n$, each $i \in \{1, \dots, n\}$, and any $W \in \partial_B G_i(x)$ we have that*

i) *if $i \in \mathcal{O}(x)$, then either $W^T G_i(x) = G_i(x) e_i^T$ or $W^T G_i(x) = -G_i(x) e_i^T$;*

ii) *if $i \in \mathcal{P}(x)$, then either $W^T G_i(x) = G_i(x) \nabla F_i(x)$ or $W^T G_i(x) = -G_i(x) \nabla F_i(x)$;*

iii) *if $i \in \mathcal{Q}(x)$, then $W^T G_i(x) = 0$;*

iv) *if $i \in \mathcal{R}$, then there exist c_i and d_i such that $W^T G_i(x) = c_i e_i^T + d_i \nabla F_i(x)$ and $c_i d_i > 0$.*

Proof. Parts i)–iii) can be easily verified. For part iv) we need only to note that if $i \in \mathcal{R}$, then $G_i(x) \neq 0$ and there exist positive numbers α_i and β_i such that $W = \alpha_i e_i + \beta_i F_i'(x)$ or $W = \alpha_i(-e_i) + \beta_i(-F_i'(x))$. ∎

Without causing any confusion we will use $\mathcal{O}$, $\mathcal{P}$, $\mathcal{Q}$, and $\mathcal{R}$ to represent $\mathcal{O}(x)$, $\mathcal{P}(x)$, $\mathcal{Q}(x)$, and $\mathcal{R}(x)$, respectively. Without loss of generality, assume that $\nabla F(x)$ is partitioned in the form

$$\nabla F(x) = \begin{pmatrix} \nabla F(x)_{\mathcal{O}\mathcal{O}} & \nabla F(x)_{\mathcal{O}\mathcal{P}} & \nabla F(x)_{\mathcal{O}\mathcal{Q}} & \nabla F(x)_{\mathcal{O}\mathcal{R}} \\ \nabla F(x)_{\mathcal{P}\mathcal{O}} & \nabla F(x)_{\mathcal{P}\mathcal{P}} & \nabla F(x)_{\mathcal{P}\mathcal{Q}} & \nabla F(x)_{\mathcal{P}\mathcal{R}} \\ \nabla F(x)_{\mathcal{Q}\mathcal{O}} & \nabla F(x)_{\mathcal{Q}\mathcal{P}} & \nabla F(x)_{\mathcal{Q}\mathcal{Q}} & \nabla F(x)_{\mathcal{Q}\mathcal{R}} \\ \nabla F(x)_{\mathcal{R}\mathcal{O}} & \nabla F(x)_{\mathcal{R}\mathcal{P}} & \nabla F(x)_{\mathcal{R}\mathcal{Q}} & \nabla F(x)_{\mathcal{R}\mathcal{R}} \end{pmatrix}.$$

Now we are ready to give the main result of this section.

THEOREM 4.2. *Suppose that $x \in \mathbb{R}^n$ is a stationary point of f, i.e., $\nabla f(x) = 0$, and that $\nabla F(x)_{\mathcal{P}\mathcal{P}}$ is nonsingular and its Schur complement in*

$$\begin{pmatrix} \nabla F(x)_{\mathcal{P}\mathcal{P}} & \nabla F(x)_{\mathcal{P}\mathcal{R}} \\ \\ \nabla F(x)_{\mathcal{R}\mathcal{P}} & \nabla F(x)_{\mathcal{R}\mathcal{R}} \end{pmatrix}$$

is a P_0-matrix. Then x is a solution of VIP(l, u, F).

Proof. Since f is continuously differentiable and G is locally Lipschitz continuous, by Clarke [5] we have that for any $y \in \mathbb{R}^n$ and any $V \in \partial G(y)$

$$\nabla f(y) = V^T G(y).$$

Let V be an element of $\partial_B G(x) (\subseteq \partial G(x))$. Then for $i = 1, \dots, n$ there exist matrices $W_i \in \partial_B G_i(x)$ such that

$$V = W_1 \times W_2 \times \cdots \times W_n.$$

Thus

$$\nabla f(x) = \sum_{i=1}^{n} W_i^T G_i(x) = 0.$$

By considering parts i) and ii) of Lemma 4.1, without loss of generality we assume that

$$W_i^T G_i(x) = G_i(x) e_i^T \text{ for } i \in \mathcal{O} \qquad \text{and} \qquad W_i^T G_i(x) = G_i(x) \nabla F_i(x) \text{ for } i \in \mathcal{P}.$$

Thus from Lemma 4.1,

$$\sum_{i\in\mathcal{O}} G_i(x) e_i^T + \sum_{i\in\mathcal{P}} G_i(x) \nabla F_i(x) + \sum_{i\in\mathcal{R}} (M_i e_i^T + N_i \nabla F_i(x)) = 0, \tag{4.1}$$

where

$$M_i := c_i G_i(x), \quad N_i := d_i G_i(x), \quad i \in \mathcal{R},$$

and c_i and d_i are numbers defined in part iv) of Lemma 4.1. Equation (4.1) can be rewritten as

$$\begin{aligned} G_{\mathcal{O}}(x) + \nabla F(x)_{\mathcal{O}\mathcal{P}} G_{\mathcal{P}}(x) + \nabla F(x)_{\mathcal{O}\mathcal{R}} N_{\mathcal{R}} &= 0, \\ \nabla F(x)_{\mathcal{P}\mathcal{P}} G_{\mathcal{P}}(x) + \nabla F(x)_{\mathcal{P}\mathcal{R}} N_{\mathcal{R}} &= 0, \\ \nabla F(x)_{\mathcal{Q}\mathcal{P}} G_{\mathcal{P}}(x) + \nabla F(x)_{\mathcal{Q}\mathcal{R}} N_{\mathcal{R}} &= 0, \\ \nabla F(x)_{\mathcal{R}\mathcal{P}} G_{\mathcal{P}}(x) + M_{\mathcal{R}} + \nabla F(x)_{\mathcal{R}\mathcal{R}} N_{\mathcal{R}} &= 0. \end{aligned} \tag{4.2}$$

From the second equality of (4.2) we have

$$G_{\mathcal{P}}(x) = -\left(\nabla F(x)_{\mathcal{P}\mathcal{P}}\right)^{-1} \nabla F(x)_{\mathcal{P}\mathcal{R}} N_{\mathcal{R}}.$$

This and the fourth equality of (4.2) give

$$M_{\mathcal{R}} + [\nabla F(x)_{\mathcal{R}\mathcal{R}} - \nabla F(x)_{\mathcal{R}\mathcal{P}}(\nabla F(x)_{\mathcal{P}\mathcal{P}})^{-1}\nabla F(x)_{\mathcal{P}\mathcal{R}}]N_{\mathcal{R}} = 0. \tag{4.3}$$

Since $\nabla F(x)_{\mathcal{R}\mathcal{R}} - \nabla F(x)_{\mathcal{R}\mathcal{P}}(\nabla F(x)_{\mathcal{P}\mathcal{P}})^{-1}\nabla F(x)_{\mathcal{P}\mathcal{R}}$ is a P_0-matrix, there exists an index $j \in \{1, \ldots, |\mathcal{R}|\}$ such that

$$(N_{\mathcal{R}})_j\{[\nabla F(x)_{\mathcal{R}\mathcal{R}} - \nabla F(x)_{\mathcal{R}\mathcal{P}}(\nabla F(x)_{\mathcal{P}\mathcal{P}})^{-1}\nabla F(x)_{\mathcal{P}\mathcal{R}}]N_{\mathcal{R}}\}_j \geq 0,$$

which, with (4.3), gives

$$(M_{\mathcal{R}})_j(N_{\mathcal{R}})_j \leq 0.$$

This contradicts part iv) of Lemma 4.1. So, we have

$$\mathcal{R} = \emptyset.$$

Then from the second and the first equalities of (4.2) we get

$$G_{\mathcal{P}}(x) = G_{\mathcal{O}}(x) = 0.$$

Thus, by Lemma 4.1 we have proved that $G(x) = 0$, so x is a solution of VIP(l, u, F) by Theorem 2.2. $\square$

The conditions used in Theorem 4.2 are quite mild. In particular, if all $l_i = -\infty$ and all $u_i = \infty$, i.e., VIP(l, u, F) reduces to the nonlinear system of equations $F(x) = 0$, we require only that $\nabla F(x)$ is nonsingular. Also, if all l_i and u_i are bounded we require only that $\nabla F(x)_{\mathcal{R}\mathcal{R}}$ is a P_0-matrix, which is implied by assuming that F is a P_0-function.

5. Nonsingularity conditions. In this section we study the conditions under which the elements of a generalized Jacobian are nonsingular at a solution point $x^* \in \mathbb{R}^n$ of VIP(l, u, F). The basic idea follows from Facchinei and Soares [14]. Since x^* is a solution of VIP(l, u, F),

$$\mathcal{O} = \mathcal{P} = \mathcal{R} = \emptyset, \quad \mathcal{Q} = \{1, \ldots, n\},$$

where $\mathcal{O}$, $\mathcal{P}$, $\mathcal{Q}$, and $\mathcal{R}$ are abbreviations of $\mathcal{O}(x^*)$, $\mathcal{P}(x^*)$, $\mathcal{Q}(x^*)$, and $\mathcal{R}(x^*)$, respectively. For notational convenience let

$$\begin{aligned}
\mathcal{I} &:= \mathcal{A}_{33}(x^*) = \{i \in 1, \ldots, n \mid l_i < x_i^* < u_i \text{ and } F_i(x^*) = 0\},\\
\mathcal{J} &:= \mathcal{A}_{43}(x^*) \cup \mathcal{A}_{53}(x^*)\\
&\;= \{i \in 1, \ldots, n \mid x_i^* = l_i \text{ and } F_i(x^*) = 0\}\\
&\quad\; \cup \{i \in 1, \ldots, n \mid x_i^* = u_i \text{ and } F_i(x^*) = 0\},\\
\mathcal{K} &:= \mathcal{A}_{41}(x^*) \cup \mathcal{A}_{52}(x^*)\\
&\;= \{i \in 1, \ldots, n \mid x_i^* = l_i \text{ and } F_i(x^*) > 0\}\\
&\quad\; \cup \{i \in 1, \ldots, n \mid x_i^* = u_i \text{ and } F_i(x^*) < 0\}.
\end{aligned}$$

Then

$$\mathcal{I} \cup \mathcal{J} \cup \mathcal{K} = \{1, \ldots, n\}.$$

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

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

VIP(l, u, F) is said to be R-regular at x^* if $F'(x^*)_{\mathcal{I}\mathcal{I}}$ is nonsingular and its Schur complement in the matrix

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

is a P-matrix; see [14]. R-regularity coincides with the notion of regularity introduced in [42].

PROPOSITION 5.1. *Suppose that VIP(l, u, F) is R-regular at x^*. Then all $V \in \partial_B G(x^*)$ are nonsingular.*

Proof. Since

$$\partial_B G(x^*) \subseteq \partial_C G(x^*) := \partial_B G_1(x^*) \times \partial_B G_2(x^*) \times \cdots \times \partial_B G_n(x^*),$$

it is sufficient to prove the conclusion by showing that all $U \in \partial_C G(x^*)$ are nonsingular. Let U be an arbitrary element of $\partial_C G(x^*)$. By the discussion of section 4 on the structure of $\partial_B G_i(x^*)$ and as x^* is a solution of VIP(l, u, F), we have

$$U_i = \begin{cases} F_i'(x^*) \quad \text{or} \quad -F_i'(x^*) & \text{if } i \in \mathcal{I}, \\ \alpha_i e_i + \beta_i F_i'(x^*) \quad \text{or} \quad \alpha_i(-e_i) + \beta_i(-F_i'(x^*)) & \text{if } i \in \mathcal{J}, \\ e_i \quad \text{or} \quad -e_i & \text{if } i \in \mathcal{K}, \end{cases} \tag{5.1}$$

where in (5.1) α_i and β_i are nonnegative numbers satisfying $(\alpha_i - 1)^2 + (\beta_i - 1)^2 = 1$ or $(\alpha_i)^2 + (\beta_i)^2 = 1$ for $i \in \mathcal{J}$. By using standard analysis (see, for example, [14, Proposition 3.2]) we can prove that U is nonsingular under the assumptions and, so, complete the proof. ◻

THEOREM 5.2. *Suppose that VIP(l, u, F) is R-regular at x^*. Then there exist a neighborhood $N(x^*)$ of x^* and a constant c such that for any $x \in N(x^*)$ and any $V \in \partial_B G(x)$, V is nonsingular and satisfies*

$$\|V^{-1}\| \leq c.$$

Proof. This follows directly from Proposition 5.1 and [39, Lemma 2.6]. ◻

COROLLARY 5.3. *If $F'(x^*)$ is a P-matrix, then the conclusion of Theorem* 5.2 *holds.*

Proof. This corollary is established by noting that if $F'(x^*)$ is a P-matrix, then VIP(l, u, F) is R-regular at x^*. ◻

We note that Sun, Fukushima, and Qi [44, Theorem 3.2] proved a similar result to Corollary 5.3 for the D-gap function $h_{\alpha\beta}(x)$ defined by (1.7). For VIP(l, u, F), their condition becomes

$$\lambda_{\min}(F'(x^*) + F'(x^*)^T) \geq \alpha + \beta^{-1}\|\nabla F(x^*)\|^2, \tag{5.2}$$

where $0 < \alpha < \beta$ and $\lambda_{\min}(W)$ denotes the smallest eigenvalue of the symmetric matrix W. Condition (5.2) implies that $F'(x^*)$ must be a positive definite matrix and hence a P-matrix. It may not be satisfied if $F'(x^*)$ is only a P-matrix. For example, let $n = 2$, $S = \mathbb{R}^2_+$, and $F(x) = Wx + q$ with

$$W = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}, \quad q = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Then $x^* = (0,0)^T$, $F'(x^*) = W$ is a P-matrix, but (5.2) fails to hold because $\lambda_{\min}(F'(x^*) + F'(x^*)^T) = 0$. Thus our assumption is weaker. In fact, one of our main motivations of this paper is to pursue a simple and differentiable merit function for VIP(l, u, F) such that the iteration matrix is nonsingular if $F'(x^*)$ is only a P-matrix.

6. A damped Gauss–Newton method. In this section a damped Gauss–Newton method for solving VIP(l, u, F) is outlined. It is similar to that in Facchinei and Kanzow [12], except that a negative gradient direction is not used. The motivation for using damped Gauss–Newton methods for solving semismooth equations is discussed in [12].

Let $I \in \mathbb{R}^{n\times n}$ be the identity matrix. An outline of a damped Gauss–Newton method is as follows.

Step 0. Choose $x^0 \in \mathbb{R}^n$, $\rho \in (0, 1)$, $p_1, p_2 > 0$, and $\sigma \in (0, 1/2)$. Set $k := 0$.

Step 1. If $\|\nabla f(x^k)\| = 0$, stop.

Step 2. Select an element $V_k \in \partial_B G(x^k)$. Let d^k be the solution of the linear system

$$\left(V_k^T V_k + p_1 \|G(x^k)\|^{p_2} I\right) d = -\nabla f(x^k). \tag{6.1}$$

Step 3. Let m_k be the smallest nonnegative integer m such that

$$f(x^k + \rho^m d^k) \le f(x^k) + \sigma\rho^m \nabla f(x^k)^T d^k. \tag{6.2}$$

Set $x^{k+1} := x^k + \rho^{m_k} d^k$, $k := k + 1$ and go to Step 1.

The above method is different from the classical damped Gauss–Newton method for solving nonlinear least squares problems in that G is not continuously differentiable. Note that if in (6.1) p_1 is set to zero, the solution of (6.1) is exactly the solution of the linear least squares problem

$$\min_{d\in\mathbb{R}^n} \frac{1}{2}\|V_k d + G(x^k)\|^2$$

as $f(x)$ is continuously differentiable and $\nabla f(x^k) = V_k^T G(x^k)$ [5]. In (6.1), the term $p_1\|G(x^k)\|^{p_2} I$ is used to make sure that $V_k^T V_k + p_1\|G(x^k)\|^{p_2} I$ is positive definite. If x^k is not a solution of VIP(l, u, F), then $\nabla f(x^k)^T d^k < 0$, which means that the above algorithm is well defined at the kth iteration. If V_k is nonsingular, then the term $V_k^T V_k$ is positive definite and with $p_1 = 0$ the solution of (6.1) reduces to solving the linear system

$$V_k d = -G(x^k)$$

to get a generalized Newton direction.

7. Convergence analysis. In this section we analyze the convergence properties of the damped Gauss–Newton method described in section 6, establishing superlinear convergence without any nondegeneracy assumption. The analysis builds on the work of [6, 12] for NCP(F).

First we state a global convergence theorem.

THEOREM 7.1. *Suppose that $\{x^k\}$ is a sequence generated by the damped Gauss–Newton method. Then each accumulation point x^* of $\{x^k\}$ is a stationary point of f.*

Proof. The proof is similar to that of [12, Theorem 15]. We omit the detail. ◻

Now we are ready to prove the superlinear (quadratic) convergence of the damped Gauss–Newton method. We proceed along the lines of the proof of [12, Theorem 17], except that for superlinear convergence we do not assume that F' is Lipschitz continuous and for quadratic convergence we do not assume that F' is continuously differentiable.

THEOREM 7.2. *Suppose that $\{x^k\}$ is a sequence generated by the damped Gauss–Newton method and x^*, an accumulation point of $\{x^k\}$, is a solution of VIP(l, u, F). If VIP(l, u, F) is R-regular at x^*, then the whole sequence $\{x^k\}$ converges to x^* Q-superlinearly. Furthermore, if F' is Lipschitz continuous around x^* and $p_2 \geq 1$, then the convergence is Q-quadratic.*

Proof. From Lemma 2.3, Proposition 2.4, and Theorem 5.2, for all x^k sufficiently close to x^* we have

$$\begin{aligned}
&\|x^k + d^k - x^*\| \\
&\quad = \|x^k - \left(V_k^T V_k + p_1\|G(x^k)\|^{p_2} I\right)^{-1} \nabla f(x^k) - x^*\| \\
&\quad \leq \| \left(V_k^T V_k + p_1\|G(x^k)\|^{p_2} I\right)^{-1} \| \, \|\nabla f(x^k) - \left(V_k^T V_k + p_1\|G(x^k)\|^{p_2} I\right)(x^k - x^*)\| \\
&\quad = O(1)\|V_k^T G(x^k) - V_k^T V_k(x^k - x^*) - p_1\|G(x^k)\|^{p_2}(x^k - x^*)\| \\
&\quad \leq O(1)\left[\|V_k^T\|\|G(x^k) - G(x^*) - V_k(x^k - x^*)\| + p_1\|G(x^k)\|^{p_2}\|x^k - x^*\|\right] \\
&\quad \leq O(1)\|G(x^k) - G(x^*) - V_k(x^k - x^*)\| + O\left(\|G(x^k)\|^{p_2}\right)\|x^k - x^*\| \\
&\quad \leq o(\|x^k - x^*\|).
\end{aligned} \tag{7.1}$$

Then, for all x^k sufficiently close to x^*,

$$\|d^k\| = \|x^k - x^*\| + o(\|x^k - x^*\|),$$

and so

$$\begin{aligned}
f(x^k + d^k) &= \frac{1}{2}\|G(x^k + d^k)\|^2 \\
&= \frac{1}{2}\|G(x^k + d^k) - G(x^*)\|^2 \\
&= O(\|x^k + d^k - x^*\|^2) \\
&= o(\|d^k\|^2).
\end{aligned}$$

Thus from Lemma 2.3 and Theorem 5.2, for all x^k sufficiently close to x^*,

$$\begin{aligned}
&f(x^k + d^k) - f(x^k) - \sigma \nabla f(x^k)^T d^k \\
&= o(\|d^k\|^2) - \frac{1}{2}\|G(x^k)\|^2 + \sigma (d^k)^T (V_k^T V_k + p_1 \|G(x^k)\|^{p_2} I) d^k \\
&= -\frac{1}{2}\|G(x^k) - G(x^*)\|^2 + \sigma (d^k)^T (V_k^T V_k) d^k + o(\|d^k\|^2) \\
&= -\frac{1}{2}(\|V_k(x^k - x^*)\| + o(\|x^k - x^*\|))^2 + \sigma (d^k)^T (V_k^T V_k) d^k + o(\|d^k\|^2) \\
&= -\frac{1}{2}\|V_k(-d_k + x^k + d^k - x^*)\|^2 + \sigma (d^k)^T (V_k^T V_k) d^k + o(\|d^k\|^2) \\
&= -\frac{1}{2}(d^k)^T (V_k^T V_k) d^k + \sigma (d^k)^T (V_k^T V_k) d^k + o(\|d^k\|^2) \\
&= \left(\sigma - \frac{1}{2}\right)(d^k)^T (V_k^T V_k) d^k + o(\|d^k\|^2) \\
&< 0.
\end{aligned}$$

Then we can deduce that for all x^k sufficiently close to x^*,

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

Thus from (7.1) we have proved that $\{x^k\}$ converges to x^* Q-superlinearly.

Finally, if F' is locally Lipschitz continuous around x^* and $p_2 \geq 1$, we can easily modify the above arguments to get the Q-quadratic convergence of $\{x^k\}$. □

COROLLARY 7.3. *If F is a uniform P-function, then the sequence $\{x^k\}$ generated by the damped Gauss–Newton method is bounded and converges to the unique solution x^* of VIP(l, u, F) Q-superlinearly. Furthermore, if F' is locally Lipschitz continuous around x^* and $p_2 \geq 1$, then the convergence is Q-quadratic.*

Proof. From Theorem 3.2 the level set $L_{f(x^0)}(f)$ is bounded. Then the sequence $\{x^k\}$ generated by the damped Gauss–Newton method is bounded and hence has at least one accumulation point, say, $\bar{x}$. According to Theorem 7.1, $\bar{x}$ is a stationary point of f. From Theorem 4.2, this stationary point $\bar{x}$ must be a solution of VIP(l, u, F) because $F'(\bar{x})$ is a P-matrix under the assumption that F is a uniform P-function. Since F is a uniform P-function, VIP(l, u, F) has a unique solution x^* (see, for example, [23, Theorem 3.9]). This means that $\bar{x} = x^*$. The conclusions of this corollary follow from Theorem 7.2 and the fact that if $F'(x^*)$ is a P-matrix, then VIP(l, u, F) is R-regular at x^*. □

Corollary 7.3 says that if F is a continuously differentiable uniform P-function, then the sequence $\{x^k\}$ generated by the damped Gauss–Newton method based on the new merit function f is well defined and converges to the unique solution of VIP(l, u, F) superlinearly. Such a result was only obtained for the nonlinear complementarity problem based on the merit function $\theta(x)$ (see, for example, [6]). In [27, 44] a similar result based on the D-gap function $h_{\alpha\beta}$ was obtained by assuming that F is a strongly monotone function, which is a stronger condition than that of a uniform P-function. Additionally, by technically choosing a sequence of smooth mappings $H_\varepsilon(x)$, $\varepsilon \to 0^+$ to approximate the nonsmooth mapping $H(x)$, Chen, Qi, and Sun [4] gave a similar result to Corollary 7.3 based on the so-called Jacobian consistency property. Here we directly construct a continuously differentiable merit function to obtain Corollary 7.3 instead of constructing a series of smooth approximating functions.

8. Numerical results. In this section we present some numerical experiments for the algorithm proposed in section 6 using the whole set of test problems from GAMS and MCP libraries (GAMSLIB and MCPLIB) [2, 8, 15]. The algorithm was implemented in MATLAB and run on a Sun SPARC Server 3002. Instead of a monotone linesearch we used a nonmonotone version as described in [10], which was originally due to Grippo, Lampariello, and Lucidi [22] and can be stated as follows. Let $\ell \geq 1$ be a prespecified constant and $\ell_k \geq 1$ be an integer which is adjusted at each iteration k. Calculate a steplength $t_k > 0$ satisfying the nonmonotone Armijo-rule

$$f(x^k + t_k d^k) \leq \mathcal{W}_k + \sigma t_k \nabla f(x^k)^T d^k, \tag{8.1}$$

where $\mathcal{W}_k := \max\{f(x^j) | j = k+1-\ell_k, \ldots, k\}$ denotes the maximal function value of f over the last ℓ_k iterations. Note that $\ell_k = 1$ corresponds to the monotone Armijo-rule. In the implementation, we used the following adjustment of ℓ_k:

1. Set $\ell_k = 1$ for $k = 0, 1, 2, 3, 4$, i.e., start the algorithm using the monotone Armijo-rule for the first four steps.
2. $\ell_{k+1} = \min\{\ell_k + 1, \ell\}$ at all remaining iterations ($\ell = 5$ in our implementation).

Throughout the computational experiments the starting points are provided by GAMSLIB or MCPLIB. The parameters used in the algorithm were $\rho = 0.5$, $p_1 = 5.0 \times 10^{-7}/\sqrt{n}$ $(n < 100), 10^{-6}/n$ $(n \geq 100)$, $p_2 = 1$, and $\sigma = 10^{-4}$. We replaced the term $p_1\|G(x^k)\|^{p_2}$ in the algorithm by $\min\{p_0, p_1\|G(x^k)\|^{p_2}\}$ with $p_0 = 10^{-4}$. If $n > 2500$, instead of using a Gauss–Newton direction, we simply used a pure Gauss–Newton direction $d^k = -(V_k^T V_k)^{-1}\nabla f(x^k) = -V_k^{-1}G(x^k)$, which is actually a (generalized) Newton direction. The iteration of the algorithm is stopped if either

$$f(x^k)/n \leq 10^{-12} \qquad \text{or} \qquad \|\nabla f(x^k)\|/\sqrt{n} \leq 10^{-10}$$

or if either

—the number of iterations exceeds 300, or

—the number of linesearch steps exceeds 40, giving a stepsize $t_k < 9.09 \times 10^{-13}$.

Finally we note that in our algorithm we assume that F is well defined everywhere, whereas there are a few examples in the GAMSLIB and MCPLIB where the function F may not be defined outside of S or even on the boundary of S. To partially avoid this problem our implementation used the following heuristic technique introduced in [10]. Let t denote a stepsize for which inequality (8.1) shall be tested. Before testing check whether $F(x^k + td^k)$ is well defined. If $F(x^k + td^k)$ is not well defined, then set $t := t/2$ and check again. Repeat this process until F is well defined or the limit of 40 linesearch steps is exceeded. In the first case continue with the nonmonotone Armijo linesearch. Otherwise the algorithm stops. This is equivalent to taking $f(x) = \infty$ for all points x where $F(x)$ is not defined.

The numerical results are summarized in Table 8.1 for the GAMSLIB problems and Tables 8.2–8.4 for the MCPLIB problems. In these tables the first column gives the name of the problem; n is the number of the variables in the problem; Nit denotes the number of iterations (LSF means the maximum number of linesearch steps was exceeded); NF denotes the number of evaluations of the function F; f_0 and f_F denote the value of f/n at the starting point and the final iterate, respectively; $\|\nabla f_F\|$ denotes the value of $\|\nabla f\|/\sqrt{n}$ at the final iterate; and CPU denotes the CPU time in seconds for the MATLAB implementation. Nit is equal to the number of evaluations of the Jacobian $F'(x)$ and the number of subproblems (6.1) or systems of linear equations solved. In the "Problem" column of Tables 8.2–8.4, the number after each problem specifies which starting point from the library is used. In the "f_0" column of Tables

Table 8.1
Numerical results for the problems from GAMSLIB.

Problem	n	f_0	Nit	NF	f_F	$\|\nabla f_F\|$	CPU
cafemge	101	$1.6\times10^{+1}$	7	12	7.2×10^{-16}	2.0×10^{-6}	0.6
cammcp	242	$1.6\times10^{+2}$	7	10	3.6×10^{-16}	5.5×10^{-6}	1.4
cammge	128	4.0×10^{-15}	0	1	4.0×10^{-15}	2.8×10^{-5}	0.2
cirimge	9	$1.1\times10^{+3}$	5	7	4.0×10^{-14}	3.7×10^{-5}	0.2
co2mge	208	3.1×10^{-15}	0	1	3.1×10^{-14}	1.1×10^{-5}	0.2
dmcmge	170	7.3	89	514	1.4×10^{-13}	$1..4\times10^{-3}$	23.2
ers82mcp	232	7.0	7	9	8.5×10^{-20}	4.2×10^{-8}	2.9
etamge	114	$1.0\times10^{+1}$	15	25	2.4×10^{-21}	5.8×10^{-8}	1.1
finmge	153	1.4×10^{-16}	0	1	1.4×10^{-16}	6.1×10^{-7}	0..2
gemmcp	262	9.6×10^{-14}	0	1	9.6×10^{-14}	1.1×10^{-5}	0.1
gemmge	178	1.4×10^{-13}	0	1	1.4×10^{-13}	2.3×10^{-6}	0.2
hansmcp	43	4.1	36	87	3.2×10^{-19}	2.3×10^{-9}	1.8
hansmge	43	3.7	14	41	3.2×10^{-25}	2.3×10^{-12}	1.0
harkmcp	32	$3.6\times10^{+1}$	23	44	1.6×10^{-13}	8.6×10^{-7}	0.7
harmge	11	$2.9\times10^{+2}$	24	57	1.3×10^{-17}	6.7×10^{-8}	0.7
kehomge	9	$1.9\times10^{+1}$	12	17	4.2×10^{-16}	1.2×10^{-6}	0.3
kormcp	78	$7.3\times10^{+2}$	5	6	6.3×10^{-13}	1.2×10^{-3}	0.3
mr5mcp	350	$2.9\times10^{+2}$	9	11	3.5×10^{-15}	2.3×10^{-5}	1.7
nsmge	212	$1.6\times10^{+1}$	15	25	2.1×10^{-20}	2.0×10^{-9}	3.8
oligomcp	6	$8..8\times10^{+2}$	6	9	2.7×10^{-21}	9.2×10^{-10}	0.2
sammge	23	0	0	1	0	0	0.1
scarfmcp	18	$1.2\times10^{+1}$	7	10	1.8×10^{-16}	6.9×10^{-7}	0.2
scarfmge	18	6.5	11	18	5.1×10^{-16}	3.6×10^{-7}	0.4
shovmge	51	9.4×10^{-9}	1	2	1.9×10^{-16}	9.6×10^{-7}	0.2
threemge	9	0	0	1	0	0	0.1
transmcp	11	$1.2\times10^{+4}$	6	21	2.6×10^{-5}	6.8×10^{-11}	0.3
two3mcp	6	$2.0\times10^{+2}$	7	10	2.2×10^{-17}	2.6×10^{-7}	0.2
unstmge	5	5.5×10^{-2}	8	10	7.6×10^{-19}	1.6×10^{-9}	0.2
vonthmcp	125	$2.1\times10^{+4}$	> 300	-	3.5	$2.4\times10^{+8}$	-
vonthmge	80	$3.3\times10^{+4}$	18(LSF)	-	$3.5\times10^{+2}$	$1.9\times10^{+6}$	-
wallmcp	6	1.2	4	5	8.8×10^{-26}	2.9×10^{-12}	0.2

8.1–8.4, DomainV means that the starting point is not in the domain of function or Jacobian.

Tables 8.1–8.4 show that the algorithm was able to solve most problems in GAMSLIB and MCPLIB. More precisely, for the GAMSLIB, for the problem `transmcp` our algorithm converged to a local minimum of $f(x)$ with $f_F = 2.6 \times 10^{-5}$ and $\|\nabla f_F\| = 6.8 \times 10^{-11}$. This is not strange because our algorithm can be used only to find local solutions of f, which may not be solutions of $\mathrm{VIP}(l, u, F)$. We also have failures on the problems `vonthmcp` and `vonthmge`. These are two von Thünen problems which are known to be very hard. By choosing different parameters we can solve `transmcp` and `vonthmge` with high precision but still fail on `vonthmcp`. On problems from the MCPLIB we have more failures. However, by using different parameters than those reported here, we can also solve all these failed problems except for `billups` and `pgvon105` with the second starting point, which violates the domain of Jacobian evaluation. The `billups` problem was constructed by Billups [1] in order to make almost all state-of-the-art methods fail on this problem. Note that the function F in the

TABLE 8.2
Numerical results for the problems from MCPLIB.

Problem	n	f_0	Nit	NF	f_F	$\|\nabla f_F\|$	CPU
bertsekas(1)	15	1.8×10^{-2}	30	115	1.5×10^{-21}	5.0×10^{-9}	1.1
bertsekas(2)	15	1.0×10^{-2}	30	107	1.6×10^{-21}	5.1×10^{-9}	1.1
bertsekas(3)	15	$4.3\times10^{+3}$	32	126	3.8×10^{-14}	2.5×10^{-5}	1.2
billups	1	5.0×10^{-5}	132	3494	1.0×10^{-5}	1.0×10^{-12}	4.6
bert_oc	5000	2.5×10^{-2}	4	6	2.5×10^{-31}	1.1×10^{-15}	49.2
bratu	5625	2.3×10^{-3}	12	50	1.9×10^{-17}	2.0×10^{-8}	141.0
choi	13	2.2×10^{-3}	4	5	3.2×10^{-17}	5.5×10^{-9}	0.6
colvdual(1)	20	$2.0\times10^{+1}$	> 300	-	2.5×10^{-4}	1.0×10^{-1}	-
colvdual(2)	20	$3.3\times10^{+2}$	> 300	-	2.5×10^{-4}	1.1×10^{-1}	-
colvnlp(1)	15	$2.7\times10^{+1}$	14	36	9.2×10^{-26}	3.6×10^{-11}	0.5
colvnlp(2)	15	$4.4\times10^{+1}$	11	20	1.1×10^{-13}	6.6×10^{-5}	0.4
cycle	1	4.4×10^{-1}	3	5	3.3×10^{-14}	2.6×10^{-7}	0.2
ehl_k40	41	$3.1\times10^{+3}$	38	158	1.6×10^{-18}	2.8×10^{-6}	5.1
ehl_k60	61	$9.2\times10^{+3}$	> 300	-	$3.4\times10^{+1}$	$1.3\times10^{+6}$	-
ehl_k80	81	$2.0\times10^{+4}$	134	1050	3.8×10^{-14}	1.0×10^{-3}	93.7
ehl_kost	101	$3.4\times10^{+4}$	> 300	-	$2.7\times10^{+1}$	$1.1\times10^{+5}$	-
explcp	16	5.0×10^{-1}	22	68	4.7×10^{-18}	3.1×10^{-9}	0.7
freebert(1)	15	1.8×10^{-2}	27	110	1.6×10^{-21}	5.1×10^{-9}	1.1
freebert(2)	15	$5.2\times10^{+6}$	59	108	1.7×10^{-21}	5.2×10^{-9}	1.2
freebert(3)	15	1.8×10^{-2}	26	106	1.6×10^{-21}	5.1×10^{-9}	1.1
freebert(4)	15	1.8×10^{-2}	30	115	1.5×10^{-21}	5.0×10^{-9}	1.2
freebert(5)	15	$5.2\times10^{+6}$	271	374	1.6×10^{-21}	5.1×10^{-9}	4.1
freebert(6)	15	1.8×10^{-2}	27	110	5.7×10^{-14}	3.0×10^{-5}	1.0
gafni(1)	5	5.3×10^{-2}	11	20	1.2×10^{-15}	5.7×10^{-6}	0.3
gafni(2)	5	1.4×10^{-2}	12	30	7.9×10^{-13}	1.5×10^{-4}	0.4
gafni(3)	5	5.5×10^{-2}	29	40	2.1×10^{-16}	2.4×10^{-6}	0.5
hanskoop(1)	14	3.8×10^{-1}	17	41	1.1×10^{-18}	7.7×10^{-9}	0.6
hanskoop(2)	14	1.3	18	45	2.9×10^{-15}	8.9×10^{-7}	0.6
hanskoop(3)	14	1.8×10^{-1}	57	172	2.6×10^{-19}	5.7×10^{-9}	1.9
hanskoop(4)	14	1.1×10^{-1}	22	116	2.8×10^{-19}	6.3×10^{-9}	1.2
hanskoop(5)	14	$2.1\times10^{+2}$	116	235	6.8×10^{-21}	1.4×10^{-9}	2.6
hydroc06	29	2.2×10^{-1}	5	7	2.3×10^{-17}	1.5×10^{-7}	0.2
hydroc20	99	1.6×10^{-1}	16	21	7.1×10^{-14}	1.8×10^{-6}	0.8
jel	6	$2.0\times10^{+2}$	7	10	2.2×10^{-17}	2.6×10^{-7}	0.2
josephy(1)	4	6.3	6	10	1.1×10^{-14}	1.4×10^{-6}	0.2
josephy(2)	4	4.3×10^{-1}	6	9	2.1×10^{-19}	5.9×10^{-9}	0.2
josephy(3)	4	$5.0\times10^{+3}$	> 300	-	3.8×10^{-2}	1.2	-

`billups` problem is pseudomonotone at a solution, which is exactly what is needed for some globally convergent methods [43]. To solve this problem, we can first use the method in [43] to make the iterates approximate the solution to some extent and then switch to the above algorithm. In fact, by using the method in [43], after 76 iterations and 147 function evaluations we get a final x with $|\min\{x, F(x)\}| = 3.8 \times 10^{-8}$. Note that for `pgvon106` we have $\|\nabla f_F\| = 2.1 \times 10^{+1}$ while f_F is very small. This also confirms that `pgvon106` is really a hard problem. The main focus of this paper is problems with both lower bounds and upper bounds on the variables. Some of the larger examples are `bratu` with $n = 5625$, `opt_cont127`, `opt_cont255`, and `opt_cont511`

Table 8.3
Numerical results for the problems from MCPLIB (continued).

Problem	n	f_0	Nit	NF	f_F	$\|\nabla f_F\|$	CPU
josephy(4)	4	6.0×10^{-1}	5	6	1.0×10^{-21}	4.1×10^{-10}	0.2
josephy(5)	4	1.6	4	5	4.1×10^{-22}	2.6×10^{-10}	0.2
josephy(6)	4	1.3	6	9	3.6×10^{-21}	7.6×10^{-10}	0.2
kojshin(1)	4	$1.6\times10^{+1}$	> 300	-	1.7×10^{-1}	3.8×10^{-1}	-
kojshin(2)	4	4.3×10^{-1}	7	15	1.3×10^{-13}	1.9×10^{-6}	0.2
kojshin(3)	4	$5.0\times10^{+3}$	10	14	1.6×10^{-14}	1.6×10^{-6}	0.2
kojshin(4)	4	2.5	2	3	1.2×10^{-20}	1.9×10^{-9}	0.1
kojshin(5)	4	6.1	4	5	6.4×10^{-22}	5.0×10^{-10}	0.2
kojshin(6)	4	4.4	7	9	1.5×10^{-24}	1.5×10^{-11}	0.2
mathinum(1)	3	2.9×10^{-1}	18	35	4.7×10^{-13}	1.9×10^{-6}	0.4
mathinum(2)	3	2.9×10^{-1}	18	35	4.7×10^{-13}	1.9×10^{-6}	0.4
mathinum(3)	3	9.7	25	62	3.2×10^{-13}	1.9×10^{-6}	0.6
mathinum(4)	3	2.1	6	7	1.8×10^{-20}	5.1×10^{-10}	0.9
mathisum(1)	4	2.0×10^{-1}	4	6	3.3×10^{-13}	1.9×10^{-6}	0.2
mathisum(2)	4	1.5	6	7	1.0×10^{-21}	1.1×10^{-10}	0.2
mathisum(3)	4	$3.8\times10^{+1}$	5	7	1.3×10^{-20}	1.0×10^{-10}	0.2
mathisum(4)	4	8.1×10^{-1}	5	6	2.7×10^{-19}	1.8×10^{-9}	0.2
methan08	31	1.1	4	5	2.4×10^{-21}	2.3×10^{-9}	0.2
nash(1)	10	$1.0\times10^{+4}$	6	7	1.9×10^{-17}	1.8×10^{-7}	0.2
nash(2)	10	$4.0\times10^{+1}$	9	20	6.5×10^{-17}	8.6×10^{-7}	0.3
obstacle(1)	2500	1.4×10^{-4}	10	11	$2..6\times10^{-23}$	3.2×10^{-11}	12.9
obstacle(2)	2500	1.4×10^{-2}	12	15	2.6×10^{-23}	3.2×10^{-11}	17.0
opt_cont31	1024	4.0×10^{-2}	9	16	4.4×10^{-24}	5.4×10^{-13}	10.6
opt_cont127	4096	5.8×10^{-3}	7	19	1.1×10^{-13}	1.3×10^{-6}	171.6
opt_cont255	8193	2.2×10^{-3}	10	34	1.4×10^{-15}	1.5×10^{-7}	894.7
opt_cont511	16384	8.3×10^{-4}	11	45	3.9×10^{-13}	2.5×10^{-6}	6003.8
pgvon105(1)	105	$2.5\times10^{+1}$	31	100	8.7×10^{-20}	2.2×10^{-6}	4.9
pgvon105(2)	105	DomainV	-	-	-	-	-
pgvon105(3)	105	5.3×10^{-1}	37(LSF)	-	3.6×10^{-3}	$2.0\times10^{+2}$	-
pgvon106	106	$2.6\times10^{+1}$	57	233	7.9×10^{-13}	$2.1\times10^{+1}$	10..9
pies	42	$1.7\times10^{+4}$	10	11	1.2×10^{-19}	3.4×10^{-8}	0.3
powell(1)	16	5.6×10^{-1}	4	5	1.0×10^{-20}	1.4×10^{-9}	0.2
powell(2)	16	$1.2\times10^{+1}$	10	18	2.7×10^{-25}	1.3×10^{-11}	0.4
powell(3)	16	$2.9\times10^{+3}$	10	11	1.4×10^{-21}	4.5×10^{-10}	0.3
powell(4)	16	$1.5\times10^{+2}$	9	10	2.5×10^{-17}	6.7×10^{-8}	0.3
powell_mcp(1)	8	$2.9\times10^{+1}$	5	6	3.4×10^{-13}	$4..9\times10^{-6}$	0.2
powell_mcp(2)	8	$1.4\times10^{+2}$	6	7	1.1×10^{-13}	2.8×10^{-6}	0.2
powell_mcp(3)	8	$1.8\times10^{+4}$	8	9	1.4×10^{-16}	9.9×10^{-8}	0.2
powell_mcp(4)	8	$1.1\times10^{+3}$	7	8	1.1×10^{-15}	2.7×10^{-7}	0.2

with 4096, 8193, and 16,394 variables, respectively. All of these problems were solved to high accuracy within 12 iterations and 50 function evaluations.

The MATLAB implementation used for these numerical experiments is not very sophisticated. The Jacobian $F'(x^k)$ and the element V_k of the generalized Jacobian are stored as a sparse matrix, but then for small problems ($n \leq 2500$) the matrix $V_k^T V_k$ is formed directly, resulting in considerable fill-in. For large problems we simply calculate the generalized Newton direction using MATLAB's direct sparse linear

TABLE 8.4
Numerical results for the problems from MCPLIB (continued).

Problem	n	f_0	Nit	NF	f_F	$\|\nabla f_F\|$	CPU
scarfanum(1)	13	3.4	8	16	2.6×10^{-20}	5.6×10^{-9}	0.3
scarfanum(2)	13	4.5	12	33	2.6×10^{-20}	5.6×10^{-9}	0.5
scarfanum(3)	13	3.0	9	12	2.5×10^{-20}	5.5×10^{-9}	0.3
scarfasum(1)	14	5.4×10^{-1}	4	6	7.3×10^{-18}	1.7×10^{-7}	0.2
scarfasum(2)	14	4.1×10^{-1}	8	19	3.5×10^{-19}	7.4×10^{-8}	0.3
scarfasum(3)	14	2.4	11	24	4.7×10^{-19}	8.5×10^{-8}	0.4
scarfbnum(1)	39	$1.0\times10^{+2}$	46	94	2.0×10^{-13}	6.0×10^{-5}	1.7
scarfbnum(2)	39	$1.1\times10^{+2}$	13	18	3.8×10^{-15}	3.2×10^{-5}	0.5
scarfbsum(1)	40	$6.1\times10^{+1}$	9	21	1.8×10^{-16}	2.5×10^{-6}	0.5
scarfbsum(2)	40	$6.8\times10^{+1}$	32(LSF)	-	4.0×10^{-1}	$5.8\times10^{+1}$	-
sppe(1)	27	$1.1\times10^{+2}$	11	19	6.8×10^{-24}	1.2×10^{-11}	0.4
sppe(2)	27	$4.8\times10^{+1}$	7	8	8.5×10^{-20}	1.4×10^{-9}	0.2
tobin(1)	42	$1.8\times10^{+2}$	9	15	1.8×10^{-16}	4.3×10^{-8}	0..4
tobin(2)	42	$1.8\times10^{+2}$	12	15	2.0×10^{-22}	6.8×10^{-11}	0.4

equation solver. In particular, note from the formulae in section 4 that the generalized Jacobian V_k has rows which consist of either the unit vector e_i, the corresponding row $F_i'(x^k)$ of the Jacobian of F, or a linear combination of these terms. Thus V_k has at least the sparsity structure of $F'(x^k)$ and often considerably more when a row $F_i'(x^k)$ is replaced by the (scaled) unit vector e_i. Thus there is considerable potential to exploit the sparsity of V_k, for example, by reordering the columns to produce more efficient matrix factorizations. In particular, if the sparsity of $F'(x)$ does not change, then this reordering could be done once rather than on every iteration.

The numerical experiments in this paper are simply meant to demonstrate the viability of the proposed merit function $f(x)$ for solving $VIP(l,u,F)$. Further work is needed to produce robust, efficient software.

9. Final remarks. In this paper we presented a new differentiable merit function for solving a box constrained variational inequality problem. This new merit function has many desirable properties over the existing ones. The key idea is to use the fact that

$$\psi(a,b)=0 \Longleftrightarrow a\geq 0, \quad ab_+=0$$

to reformulate $\mathrm{VIP}(l,u,F)$ as the minimization of an unconstrained differentiable merit function. This reformulation allows us to construct a globally and superlinearly convergent damped Gauss–Newton method for solving $\mathrm{VIP}(l,u,F)$. One of the most important features of the damped Gauss–Newton method introduced here is that at each iteration we need only to solve a linear system of equations. Besides the formula introduced in this paper, there are other possible functions $\psi(a,b)$. For example we can let

$$\psi^{\text{new}}(a,b) := ([-\phi_R(a,b)]_+)^2 + ([-a]_+)^2, \tag{9.1}$$

where $\phi_R:\mathbb{R}^2\to\mathbb{R}$ is defined by

$$\phi_R(a,b) := \phi(a,b) - a_+b_+ = \sqrt{a^2+b^2}-(a+b)-a_+b_+ \tag{9.2}$$

and can be regarded as a regularized Fischer–Burmeister function. It is not difficult to verify that $\phi_R(\cdot)^2$ and $\psi^{\text{new}}(\cdot)$ are continuously differentiable functions on $\mathbb{R}^2$ (for any $b \in \mathbb{R}$ we define $\phi_R(+\infty, b) = -b$). This modification may enhance the boundedness results of the corresponding merit function. Chen, Chen, and Kanzow [3] reported some interesting properties of the modified Fischer–Burmeister function

$$\begin{aligned} \phi_{CCK}(a,b) &= -[\lambda\phi(a,b) - (1-\lambda)a_+b_+] \\ &= -\lambda\left(\phi(a,b) - \tfrac{1-\lambda}{\lambda}a_+b_+\right), \quad \lambda \in (0,1). \end{aligned} \tag{9.3}$$

By letting $\alpha = \frac{1-\lambda}{\lambda}$ and ignoring the outside $-\lambda$ parameter, the function ϕ_{CCK} defined in (9.3) takes the form

$$\phi_{CCK}(a,b) = \phi(a,b) - \alpha a_+b_+, \quad \alpha \in (0,\infty). \tag{9.4}$$

Note that $a, b \geq 0$ and $ab = 0$ is equivalent to $\alpha a, \alpha b \geq 0$ and $(\alpha a)(\alpha b) = 0$ for any $\alpha > 0$. The function ϕ_{CCK} defined in (9.4) can be treated as a scaled form of ϕ_R. Numerically this scaling can play an important role in the behavior of the corresponding algorithm. The application of ϕ_R or ϕ_{CCK} to box constrained variational inequality problems needs further investigation.

In this paper we introduced the merit function f for $\text{VIP}(l,u,F)$ without considering whether l_i, u_i, $i = 1,\ldots,n$ are finite or not. However, if some l_i and/or u_i are infinite, we can modify our merit function. For example we can define

$$\begin{aligned} f^{\text{new}}(x) := \frac{1}{2}\Bigg[&\sum_{i\in I_l}\phi(x_i - l_i, F_i(x))^2 + \sum_{i\in I_u}\phi(u_i - x_i, -F_i(x))^2 \\ &+ \sum_{i\in I_{lu}}\psi(x_i - l_i, F_i(x)) + \sum_{i\in I_{lu}}\psi(u_i - x_i, -F_i(x))\Bigg], \end{aligned} \tag{9.5}$$

where

$$\begin{aligned} I_l &:= \{i|\ l_i > -\infty,\ u_i = \infty,\ i = 1,\ldots,n\}, \\ I_u &:= \{i|\ l_i = -\infty,\ u_i < \infty,\ i = 1,\ldots,n\}, \\ I_{lu} &:= \{1,\ldots,n\}\backslash\{I_l \cup I_u\}. \end{aligned}$$

The function $f^{\text{new}}(x)$ has similar properties to $f(x)$ and possibly has only slightly different stationary point conditions. Roughly speaking, the stationary point conditions for f need a stronger nonsingularity condition on ∇F, while f^{new} needs a stronger P_0 property on ∇F (i.e., the set $\mathcal{P}$ in Theorem 4.2 may contain fewer elements). Note that in (9.5) the functions $\phi(\cdot)$ and $\psi(\cdot)$ can be replaced by $\phi_R(\cdot)$ and $\psi^{\text{new}}(\cdot)$, respectively.

Acknowledgments. The authors would like to thank the two anonymous referees and the associate editor for their helpful comments, which improved the presentation of this paper. Thanks also to Houduo Qi for his comments on (9.1) and (9.2).

REFERENCES

[1] S. C. Billups, *Algorithms for complementarity problems and generalized equations*, Ph.D. thesis, Computer Sciences Department, University of Wisconsin, Madison, WI, 1995.

[2] S. C. Billups, S. P. Dirkse, and M. C. Ferris, *A comparison of algorithms for large scale mixed complementarity problems*, Comput. Optim. Appl., 7 (1997), pp. 3–25.

[3] B. Chen, X. Chen, and C. Kanzow, *A Modified Fischer-Burmeister NCP-function*, Talk presented at the 1997 International Symposium on Mathematical Programming, EPFL, Lausanne, Switzerland, 1997.

[4] X. Chen, L. Qi, and D. Sun, *Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities*, Math. Comp., 67 (1998), pp. 519–540.

[5] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley, New York, 1983; reprinted by SIAM, Philadelphia, 1990.

[6] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Programming, 75 (1996), pp. 407–439.

[7] S. P. Dirkse and M. C. Ferris, *The PATH solver: A non-monotone stabilization scheme for mixed complementarity problems*, Optimization Methods and Software, 5 (1995), pp. 123–156.

[8] S. P. Dirkse and M. C. Ferris, *MCPLIB: A collection of nonlinear mixed complementarity problems*, Optim. Methods Softw., 5 (1995), pp. 319–345.

[9] B. C. Eaves, *On the basic theorem of complementarity*, Math. Programming, 1 (1971), pp. 68–75.

[10] F. Facchinei, A. Fischer, and C. Kanzow, *A semsimooth Newton method for variational inequalities: The case of box constraints*, in Complementarity and Variational Problems: State of the Art, M. C. Ferris and J. S. Pang, eds., SIAM, Philadelphia, PA, 1997, pp. 76–90.

[11] F. Facchinei and C. Kanzow, *On unconstrained and constrained stationary points of the implicit Lagrangian*, J. Optim. Theory Appl., 92 (1997), pp. 99–115.

[12] F. Facchinei and C. Kanzow, *A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems*, Math. Programming, 76 (1997), pp. 493–512.

[13] F. Facchinei and J. Soares, *Testing a new class of algorithms for nonlinear complementarity problems*, in Variational Inequalities and Network Equilibrium Problems, F. Giannessi, ed., Plenum Press, New York, 1995, pp. 69–83.

[14] F. Facchinei and J. Soares, *A new merit function for nonlinear complementarity problems and a related algorithm*, SIAM J. Optim., 7 (1997), pp. 225–247.

[15] M. C. Ferris and T. F. Rutherford, *Accessing realistic mixed complementarity problems within MATLAB*, in Nonlinear Optimization and Applications, G. Di Pillo and F. Giannessi, eds., Plenum Press, New York, 1996, pp. 141–153.

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

[17] A. Fischer, *An NCP-function and its use for the solution of complementarity problems*, in Recent Advances in Nonsmooth Optimization, D. Du, L. Qi, and R. Womersley, eds., World Scientific, River Edge, NJ, 1995, pp. 88–105.

[18] A. Fischer, *Solution of monotone complementarity problems with locally Lipschitzian functions*, Math. Programming, 76 (1997), pp. 513–532.

[19] M. Fukushima, *Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems*, Math. Programming, 53 (1992), pp. 99–110.

[20] M. Fukushima, *Merit functions for variational inequality and complementarity problems*, in Nonlinear Optimization and Applications, G. Di Pillo and F. Giannessi, eds., Plenum Press, New York, 1996, pp. 155–170.

[21] C. Geiger and C. Kanzow, *On the resolution of monotone complementarity problems*, Comput. Optim. Appl., 5 (1996), pp. 155–173.

[22] L. Grippo, F. L. Lampariello, and S. Lucidi, *A nonmonotone linesearch technique for Newton's method*, SIAM J. Numer. Anal., 23 (1986), pp. 707–716.

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

[24] H. Jiang, *Unconstrained minimization approaches to nonlinear complementarity problems*, J. Global Optim., 9 (1996), pp. 169–181.

[25] H. Jiang and L. Qi, *A new nonsmooth equations approach to nonlinear complementarity problems*, SIAM J. Control Optim., 35 (1997), pp. 178–193.

[26] C. Kanzow, *Nonlinear complementarity as unconstrained optimization*, J. Optim. Theory Appl., 88 (1996), pp. 139–155.

[27] C. Kanzow and M. Fukushima, *Theoretical and numerical investigation of the D-gap function for box constrained variational inequalities*, Math. Programming, 83 (1998), pp. 55–87.

[28] Z. -Q. Luo, O. L. Mangasarian, J. Ren, and M. V. Solodov, *New error bounds for the linear complementarity problem*, Math. Oper. Res., 19 (1994), pp. 880–892.

[29] Z. -Q. Luo and P. Tseng, *A new class of merit functions for the nonlinear complementarity problem*, in Complementarity and Variational Problems: State of the Art, M. C. Ferris and J. S. Pang, eds., SIAM, Philadelphia, PA, 1997, pp. 204–225.

[30] O. L. Mangasarian and M. V. Solodov, *Nonlinear complementarity as unconstrained and constrained minimization*, Math. Programming, 62 (1993), pp. 277–297.

[31] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15 (1977), pp. 959–972.

[32] J. S. Pang, *Newton's method for B-differentiable equations*, Math. Oper. Res., 15 (1990), pp. 331–341.

[33] J. S. Pang, *A B-differentiable equation based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems*, Math. Programming, 51 (1991), pp. 101–131.

[34] J. S. Pang, *Complementarity problems*, in Handbook of Global Optimization, R. Horst and P. Pardalos, eds., Kluwer Academic Publishers, Norwell, MA, 1995, pp. 271–338.

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

[36] J. S. Pang and L. Qi, *Nonsmooth equations: Motivation and algorithms*, SIAM J. Optim., 3 (1993), pp. 443–465.

[37] J. M. Peng, *Equivalence of variational inequality problems to unconstrained optimization*, Math. Programming, 78 (1997), pp. 347–355.

[38] J. M. Peng and Y. Yuan, *Unconstrained methods for generalized complementarity problems*, J. Comput. Math., 15 (1997), pp. 253–264.

[39] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), pp. 227–244.

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

[41] D. Ralph, *Global convergence of damped Newton's method for nonsmooth equations via the path search*, Math. Oper. Res., 19 (1994), pp. 352–389.

[42] S. M. Robinson, *Generalized equations*, in Mathematical Programming: The State of the Art, A. Bachem, M. Grötschel, and B. Korte, eds., Springer-Verlag, Berlin, 1983, pp. 346–367.

[43] D. Sun, *A class of iterative methods for solving nonlinear projection equations*, J. Optim. Theory Appl., 91 (1996), pp. 123–140.

[44] D. Sun, M. Fukushima, and L. Qi, *A computable generalized Hessian of the D-gap function and Newton-type methods for variational inequality problems*, in Complementarity and Variational Problems: State of the Art, M. C. Ferris and J. S. Pang, eds., SIAM, Philadelphia, PA, 1997, pp. 452–473.

[45] P. Tseng, *Global behaviour of a class of merit functions for the nonlinear complementarity problem*, J. Optim. Theory Appl., 89 (1996), pp. 17–37.

[46] P. Tseng, N. Yamashita, and M. Fukushima, *Equivalence of complementarity problems to differentiable minimization: A unified approach*, SIAM J. Optim., 6 (1996), pp. 446–460.

[47] N. Yamashita and M. Fukushima, *On stationary points of implicit Lagrangian for nonlinear complementarity problems*, J. Optim. Theory Appl., 84 (1995), pp. 653–663.

[48] N. Yamashita and M. Fukushima, *Equivalent unconstrained minimization and global error bounds for variational inequality problems*, SIAM J. Control Optim., 35 (1997), pp. 273–284.

[49] N. Yamashita, K. Taji, and M. Fukushima, *Unconstrained optimization reformulations of variational inequality problems*, J. Optim. Theory Appl., 92 (1997), pp. 439–456.