Computational Optimization and Applications, 17, 255–277, 2000

Some Methods Based on the D-Gap Function for Solving Monotone Variational Inequalities

MICHAEL V. SOLODOV solodov@impa.br
Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ, 22460-320, Brazil

PAUL TSENG tseng@math.washington.edu
Department of Mathematics, University of Washington, Seattle, WA 98195, USA

Received April 20, 1999; Accepted August 30, 1999

Abstract. The D-gap function has been useful in developing unconstrained descent methods for solving strongly monotone variational inequality problems. We show that the D-gap function has certain properties that are useful also for monotone variational inequality problems with bounded feasible set. Accordingly, we develop two unconstrained methods based on them that are similar in spirit to a feasible method of Zhu and Marcotte based on the regularized-gap function. We further discuss a third method based on applying the D-gap function to a regularized problem. Preliminary numerical experience is also reported.

Keywords: monotone variational inequalities, implicit Lagrangian, D-gap function, stationary point, descent methods

1. Introduction

Consider the monotone variational inequality problem (VI) of finding an $x \in X$ satisfying

$$F(x)^T(y - x) \geq 0 \quad \forall y \in X,$$

where $F = (F_1, \ldots, F_n)^T : \Re^n \mapsto \Re^n$ is a given continuously differentiable monotone mapping and X is a nonempty closed convex set in $\Re^n$. This is a well-known problem in optimization (see [7] and references therein). For each $\alpha > 0$, denote

$$y_\alpha(x) := \text{Proj}_X[x - \alpha F(x)] \quad \forall x \in \Re^n, \tag{1}$$

where $\text{Proj}_X[z] = \min_{y \in X} \|y - z\|$. It is well known and easily shown that x solves VI if and only if $x = y_\alpha(x)$.

A popular approach to solving VI entails reformulating VI as an equivalent minimization problem. In particular, Auslender [3] suggested the equivalent problem of minimizing the (nondifferentiable, and possibly extended-valued) *gap* function:

$$g(x) := \max_{y \in X} F(x)^T(x - y),$$

subject to $x \in X$. To overcome the nondifferentiability of g, Fukushima [8] and Auchmuty [2] independently proposed the *regularized-gap* function:

$$g_\alpha(x) := \max_{y \in X} F(x)^T(x - y) - \frac{1}{2\alpha}\|x - y\|^2$$
$$= F(x)^T(x - y_\alpha(x)) - \frac{1}{2\alpha}\|x - y_\alpha(x)\|^2, \tag{2}$$

which is continuously differentiable. Moreover, Fukushima showed that any stationary point x^* of the problem $\min_{x \in X} g_\alpha(x)$ with $\nabla F(x^*)$ positive definite is a solution of VI. Fukushima further developed a feasible descent method, based on the (derivative-free) descent direction $y_\alpha(x) - x$, for solving VI and showed convergence when F is strongly monotone. Zhu and Marcotte [33] proposed an interesting modification of Fukushima's method whereby, given a current iterate $x \in X$, it either takes a descent step with direction $y_\alpha(x) - x$ or increases α. Convergence of this method was shown assuming X bounded (F need not be strongly monotone).

In the above reformulations, x is constrained to the feasible set X. More recently, Mangasarian and Solodov [16] proposed in the case of NCP (i.e., $X = \Re^n_+$) a reformulation involving the unconstrained minimization of an implicit Lagrangian. This was subsequently extended by Peng [20] and then Fukushima et al. [31] to the *D-gap* function:

$$h_{\alpha,\beta}(x) := g_\alpha(x) - g_\beta(x) \quad \forall x \in \Re^n, \tag{3}$$

with $\alpha > \beta > 0$. The implicit Lagrangian corresponds to the case $\alpha > 1$ and $\beta = 1/\alpha$ (see [24] for a survey of its properties and further references). It was shown [20, 30, 31] that this function is continuously differentiable and any stationary point x^* of the problem $\min_{x \in \Re^n} h_{\alpha,\beta}(x)$ with $\nabla F(x^*)$ positive definite is a solution of VI. And if X is a box, then $\nabla F(x^*)$ only needs to be a P-matrix [11]. Based on this fact, unconstrained descent methods were developed to solve VI, and convergence to a solution was shown when F is strongly monotone or, if X is a box, when F is a uniformly P-function. The reference [11] reported particularly promising numerical results with a truncated Gauss-Newton-type method. Other properties of the D-gap function such as boundedness of level sets and error bounds were also studied in these references, as well as [12, 21, 22]. [Fukushima et al. considered the more general setting where $\|x - y\|^2/2$ is replaced by a class of distance-like functions $\phi(x, y)$ studied by Wu et al. [28].] Extensions of some of the results to nonsmooth F were given by Jiang [10] and Xu [32]. Sun et al. [27] introduced a computable generalized Hessian of the D-gap function, based on which a (local) generalized Newton method and a (global) trust-region algorithm for solving VI were developed. Peng and Fukushima [21] proposed a hybrid Josephy-Newton method, with the D-gap function acting as a merit function and its negative gradient used as a safeguard descent direction. Assuming F strongly monotone, global convergence and local superlinear convergence to a solution are shown. Peng et al. [22] refined this result for the case of box constraints, with strong monotonicity replaced by F being a uniformly P-function. For the box-constrained case, Kanzow and Fukushima [12] proposed an alternative Newton method based on the B-subdifferential of the natural residual. Accordingly, global convergence (to stationary point of D-gap function), local superlinear convergence (to b-regular solution), and finite termination (at b-regular solution, assuming F is affine) are shown. Promising numerical results are reported in [12, 22]. The survey papers [9, 14] give additional references on merit functions for VI.

The following (slightly modified) example of Yamashita and Fukushima [30]:

$$n = 1, \quad F(x) = (x-1)^3 - 1, \quad X = \Re_+, \tag{4}$$

shows that computing a stationary point of the implicit Lagrangian and the D-gap function may not be suitable when F is only monotone. Here, $x = 1$ is a stationary point of $h_{\alpha,\beta}$ for all $\alpha > \beta > 0$ but is not a solution of the VI, which occurs at $x = 2$. A closer examination of this example reveals that X being unbounded is also critical. In particular, if we replace X by a bounded subset containing the solution $x = 2$, say $X = [0, u]$ with $u > 2$, then for any $\alpha \geq u$ and $0 < \beta < \alpha$, we have that $x = 2$ is the unique stationary point of $h_{\alpha,\beta}$. This suggests that, for X bounded and α sufficiently large, it may still be possible to solve the VI by computing a stationary point of $h_{\alpha,\beta}$. In this paper, we show that this intuition is essentially valid. Based on this, we develop two methods, in the spirit of the Zhu–Marcotte method but using the D-gap function, to solve monotone VI with bounded feasible set X. The key to our methods is a strategy for adjusting α and β when a stationary point of $h_{\alpha,\beta}$ is not a solution of VI. For the purposes of comparison, we also study a regularization method based on computing an approximate stationary point of the D-gap function associated with a regularized VI. Preliminary numerical experience with the two methods is reported in the last section.

The assumption of X being bounded seems reasonable since in practice, upper bounds are frequently placed on variables based on physical considerations. If X is unbounded but it is known that the solution set S makes a nonempty intersection with the interior of a closed bounded convex set B, then we can replace the feasible set X by $X \cap B$ and the new VI would have $S \cap B$ as its solution set.[1] Thus, if we have an *a priori* error bound of the form:

$$\|x^0 - x^*\| < \tau g_\alpha(x^0)\,,$$

with $x^0 \in X$ and $\alpha > 0$, $\tau > 0$ known and x^* an (unknown) solution, then we can take B to be the closed Euclidean ball of radius $\tau g_\alpha(x^0)$ and centered at x^0. The above error bound holds in the case where F is strongly monotone [26] or where X is a box and F is a uniformly P-function (see Theorem 1). Of course, an error bound based on any other merit function can be used (although other bounds typically require stronger assumptions than the one using the regularized-gap function). For the case of monotone LCP, even more choices are available (see the survey paper [19] and references therein).

In our notation, all vectors are column vectors, $\Re^n$ denotes the space of n–dimensional real column vectors, ^{T} denotes transpose, $\Re^n_+$ denotes the nonnegative orthant in $\Re^n$. For any vector $x \in \Re^n$, we denote by x_i the ith component of x and by $\|x\|$ the Euclidean norm of x, so $\|x\| = \sqrt{x^T x}$. We denote the Jacobian of F by $\nabla F = [\nabla F_1 \ \cdots \ \nabla F_n]$, where ∇F_i denotes the gradient of F_i. Also, we denote $\mathrm{diam}(X) = \max_{x,y \in X} \|x - y\|$.

2. Regularized-gap function

By definition of $y_\alpha(x)$ and properties of projection, we have

$$\left(F(x) + \frac{1}{\alpha}(y_\alpha(x) - x)\right)^T (x - y_\alpha(x)) \geq 0 \quad \forall x \in X,$$

so that (2) yields

$$g_\alpha(x) \geq \frac{1}{2\alpha}\|x - y_\alpha(x)\|^2 \quad \forall x \in X. \tag{5}$$

Thus, for any $\alpha > 0$, $x \in X$ and $g_\alpha(x) = 0$ imply x solves VI.

It is well known that g_α is continuously differentiable on $\Re^n$ and that

$$\nabla g_\alpha(x) = \nabla F(x)(x - y_\alpha(x)) + F(x) + \frac{1}{\alpha}(y_\alpha(x) - x) \quad \forall x \in \Re^n. \tag{6}$$

Then (6) and the positive semidefinite property of $\nabla F(x)$ yield

$$\begin{aligned}
&\nabla g_\alpha(x)^T(x - y_\alpha(x)) \\
&\quad = (x - y_\alpha(x))^T \nabla F(x)(x - y_\alpha(x)) + F(x)^T(x - y_\alpha(x)) - \frac{1}{\alpha}\|x - y_\alpha(x)\|^2 \\
&\quad \geq F(x)^T(x - y_\alpha(x)) - \frac{1}{\alpha}\|x - y_\alpha(x)\|^2 \\
&\quad = g_\alpha(x) - \frac{1}{2\alpha}\|x - y_\alpha(x)\|^2,
\end{aligned} \tag{7}$$

where the last equality uses (2). Thus, for any fixed $\gamma \in (0, 1)$,

either (i) $\nabla g_\alpha(x)^T(y_\alpha(x) - x) \leq -\gamma g_\alpha(x)$,
or else (ii) $(1 - \gamma)g_\alpha(x) \leq \frac{1}{2\alpha}\|x - y_\alpha(x)\|^2$.

This fact was shown by Zhu and Marcotte [33, Cor. 1], although our proof seems simpler. Assuming X is bounded, Zhu and Marcotte then used this fact to develop a modification of a feasible descent method of Fukushima [8] that, given a current iterate $x \in X$ and $\alpha > 0$, either takes a feasible descent step with direction $y_\alpha(x) - x$ (and using g_α as the merit function) if case (i) occurs or else increases α with x fixed until case (i) occurs.[2]

We conclude this section with an error bound result for the case where X is a box and F is a uniformly P-function on the set X, i.e., there exists some $\mu > 0$ such that

$$\forall\, x, y \in X \quad \exists j \in \{1, \ldots, n\} \text{ such that } (F_j(x) - F_j(y))(x_j - y_j) \geq \mu\|x - y\|^2.$$

This result seems to be the first error bound for the case of box-constrained VI which does not require F to be Lipschitz continuous. For the case of NCP, such bounds (based on other merit functions) have been obtained in [13, 29]. Our proof is in the spirit of one used in [26, Prop. 3.4] for the case where F is strongly monotone.

Theorem 1. *Assume $X = \Pi_{i=1}^n [l_i, u_i]$ for some $-\infty \le l_i < u_i \le \infty$, and F is a uniformly P-function on X with modulus $\mu > 0$. Then for any $\alpha > 1/(2\mu)$ it holds that*

$$\forall x \in X, \quad g_\alpha(x) \ge \left(\mu - \frac{1}{2\alpha}\right)\|x - x^*\|^2 ,$$

where x^* is the solution of VI (known to exist and be unique).

Proof: Since x^* is the solution of VI and X is a box, it is known and easily seen that

$$\forall x \in X , \quad F_i(x^*)(x_i - x_i^*) \ge 0, \quad i = 1, \ldots, n .$$

Since F is a uniformly P-function on X, for any fixed $x \in X$ we have that there exists some index j such that

$$(F_j(x) - F_j(x^*))(x_j - x_j^*) \ge \mu\|x - x^*\|^2 .$$

Adding the two inequalities, we obtain that

$$F_j(x)(x_j - x_j^*) \ge \mu\|x - x^*\|^2 . \tag{8}$$

Observe further that

$$\begin{aligned}
g_\alpha(x) &= \max_{y\in X} F(x)^T(x - y) - \frac{1}{2\alpha}\|x - y\|^2 \\
&= \max_{y\in X} \sum_{i=1}^n \left(F_i(x)(x_i - y_i) - \frac{1}{2\alpha}(x_i - y_i)^2\right) \\
&= \sum_{i=1}^n \max_{l_i\le y_i\le u_i} \left(F_i(x)(x_i - y_i) - \frac{1}{2\alpha}(x_i - y_i)^2\right) \\
&= \sum_{i=1}^n g_\alpha^i(x),
\end{aligned}$$

where

$$g_\alpha^i(x) := \max_{l_i\le y_i\le u_i} \left(F_i(x)(x_i - y_i) - \frac{1}{2\alpha}(x_i - y_i)^2\right).$$

Clearly, if $x \in X$ then $g_\alpha^i(x) \ge 0$ for all i (because $y_i = x_i$ is feasible). Hence,

$$g_\alpha(x) \ge g_\alpha^j(x) . \tag{9}$$

Furthermore, since $x^* \in X$ we have that

$$\begin{aligned} g_\alpha^j(x) &= \max_{l_j \le y_j \le u_j} \left(F_j(x)(x_j - y_j) - \frac{1}{2\alpha}(x_j - y_j)^2 \right) \\ &\ge F_j(x)(x_j - x_j^*) - \frac{1}{2\alpha}(x_j - x_j^*)^2 \\ &\ge \mu \|x - x^*\|^2 - \frac{1}{2\alpha}\|x - x^*\|^2 \\ &= \left(\mu - \frac{1}{2\alpha} \right) \|x - x^*\|^2, \end{aligned}$$

where the last inequality follows from (8) and from the monotonicity of the norm. Taking into account (9) completes the proof. □

3. D-gap function

In this section we derive useful properties of the D-gap function $h_{\alpha,\beta}$ which we use in the next section to develop two methods for solving VI when X is bounded. We begin with the following two-sided bound on $h_{\alpha,\beta}$, which is an analog of (5).

Lemma 1. *For any $\alpha > \beta > 0$ and $x \in \Re^n$, we have*

$$h_{\alpha,\beta}(x) \ge \frac{1}{2}\left(\frac{1}{\beta} - \frac{1}{\alpha} \right) \|x - y_\beta(x)\|^2 + \frac{1}{2\alpha}\|y_\alpha(x) - y_\beta(x)\|^2, \tag{10}$$

$$h_{\alpha,\beta}(x) \le \frac{1}{2}\left(\frac{1}{\beta} - \frac{1}{\alpha} \right) \|x - y_\alpha(x)\|^2 - \frac{1}{2\beta}\|y_\alpha(x) - y_\beta(x)\|^2. \tag{11}$$

Proof: First note that by (3) and (2), we have that

$$h_{\alpha,\beta}(x) = F(x)^T (y_\beta(x) - y_\alpha(x)) - \frac{1}{2\alpha}\|x - y_\alpha(x)\|^2 + \frac{1}{2\beta}\|x - y_\beta(x)\|^2.$$

Using this relation, the projection properties of $y_\alpha(x)$ imply

$$\begin{aligned} 0 &\le \left(F(x) + \frac{1}{\alpha}(y_\alpha(x) - x) \right)^T (y_\beta(x) - y_\alpha(x)) \qquad (12) \\ &= h_{\alpha,\beta}(x) - \frac{1}{2}\left(\frac{1}{\beta} - \frac{1}{\alpha} \right) \|x - y_\beta(x)\|^2 - \frac{1}{2\alpha}\|y_\alpha(x) - y_\beta(x)\|^2 . \end{aligned}$$

This proves (10). Similarly, the projection properties of $y_\beta(x)$ imply

$$\begin{aligned} 0 &\le \left(F(x) + \frac{1}{\beta}(y_\beta(x) - x) \right)^T (y_\alpha(x) - y_\beta(x)) \qquad (13) \\ &= -h_{\alpha,\beta}(x) - \frac{1}{2}\left(\frac{1}{\alpha} - \frac{1}{\beta} \right) \|x - y_\alpha(x)\|^2 - \frac{1}{2\beta}\|y_\beta(x) - y_\alpha(x)\|^2 , \end{aligned}$$

which proves (11). □

Thus, for any $\alpha > \beta > 0$, $h_{\alpha,\beta}(x) = 0$ implies x solves VI. The bounds in Lemma 1 are tight. In particular, suppose F is a constant (but nonzero) function and $x = x^* - tF(x^*)$, where x^* is any solution of the VI and t is any positive number. Then $x_\alpha(x) = x_\beta(x) = x^*$ for every $\alpha > \beta > 0$, so (10) and (11) hold with equality. Notice that x is far away from X for large t.

Using (3) and (6), we have that $h_{\alpha,\beta}$ is continuously differentiable on $\Re^n$ and that

$$\nabla h_{\alpha,\beta}(x) = \nabla F(x)(y_\beta(x) - y_\alpha(x)) + \frac{1}{\alpha}(y_\alpha(x) - x) - \frac{1}{\beta}(y_\beta(x) - x) \quad \forall x \in \Re^n. \tag{14}$$

Then (14) and the positive semidefinite property of $\nabla F(x)$ yield

$$\begin{aligned}
&\nabla h_{\alpha,\beta}(x)^T(y_\beta(x) - y_\alpha(x)) \\
&\quad = (y_\beta(x) - y_\alpha(x))^T \nabla F(x)(y_\beta(x) - y_\alpha(x)) \\
&\qquad + \left(\frac{1}{\alpha}(y_\alpha(x) - x) - \frac{1}{\beta}(y_\beta(x) - x)\right)^T (y_\beta(x) - y_\alpha(x)) \\
&\quad \geq \left(\frac{1}{\alpha}(y_\alpha(x) - x) - \frac{1}{\beta}(y_\beta(x) - x)\right)^T (y_\beta(x) - y_\alpha(x)) \\
&\quad =: e_{\alpha,\beta}(x) \geq 0\,,
\end{aligned} \tag{15}$$

where nonnegativity of $e_{\alpha,\beta}(x)$ follows from adding the inequalities (12) and (13) (see also [31, Lemma 3.2]). The inequality (15) is somewhat analogous to (7). In particular, either $e_{\alpha,\beta}(x)$ is above a tolerance $\epsilon > 0$, in which case $y_\alpha(x) - y_\beta(x)$ is a direction of sufficient descent for $h_{\alpha,\beta}$ at x or else, as we show in the lemmas below, x is an approximate solution of the VI with accuracy depending on ϵ, α, β. This result will lead to our methods based on the D-gap function.

Lemma 2. *For any $\alpha > \beta > 0$ and $x \in \Re^n$, we have*

$$\|x - y_\beta(x)\| \leq \sqrt{2h_{\alpha,\beta}(x)/(1/\beta - 1/\alpha)} \qquad \textit{and} \tag{16}$$

$$r_{\alpha,\beta}(x) \leq g_\alpha(x) \leq r_{\alpha,\beta}(x) + e_{\alpha,\beta}(x) + \frac{1}{2\alpha}\|y_\alpha(x) - y_\beta(x)\|^2\,, \tag{17}$$

where we let $r_{\alpha,\beta}(x) := F(x)^T(x - y_\beta(x)) - \frac{1}{2\alpha}\|x - y_\beta(x)\|^2$.

Proof: Inequality (16) follows immediately from (10) of Lemma 1.

The definition of the regularized-gap function (2) implies that

$$g_\alpha(x) = \max_{y \in X} F(x)^T(x - y) - \frac{1}{2\alpha}\|x - y\|^2 \geq r_{\alpha,\beta}(x)\,,$$

proving the first inequality in (17).

Since $e_{\alpha,\beta}(x)$ is the sum of the nonnegative quantity $(F(x) + \frac{1}{\alpha}(y_\alpha(x) - x))^T(y_\beta(x) - y_\alpha(x))$ with another nonnegative quantity (see (12) and (13)), we have

$$e_{\alpha,\beta}(x) \geq \left(F(x) + \frac{1}{\alpha}(y_\alpha(x) - x)\right)^T (y_\beta(x) - y_\alpha(x))$$
$$= g_\alpha(x) - r_{\alpha,\beta}(x) - \frac{1}{2\alpha}\|y_\alpha(x) - y_\beta(x)\|^2,$$

where the equality also uses (2). Thus,

$$g_\alpha(x) \leq r_{\alpha,\beta}(x) + e_{\alpha,\beta}(x) + \frac{1}{2\alpha}\|y_\alpha(x) - y_\beta(x)\|^2,$$

which is the second inequality in (17). □

In the special case where X is a box, we can sharpen Lemma 2 further to obtain explicit bounds on $y_\alpha(x) - x$ and $y_\beta(x) - x$ in terms of $e_{\alpha,\beta}(x)$, α, β.

Lemma 3. *Assume $X = \Pi_{i=1}^n [l_i, u_i]$ for some $-\infty < l_i < u_i < \infty$. For any $\alpha > \beta > 0$ and $x \in \Re^n$, we have for each $i \in \{1, \ldots, n\}$ that*

$$\text{either} \quad |(y_\alpha(x) - x)_i| \leq \sqrt{e_{\alpha,\beta}(x)} + \sqrt{2h_{\alpha,\beta}(x)/(1/\beta - 1/\alpha)}$$
$$\text{or} \quad \frac{1}{\beta}|(y_\beta(x) - x)_i| \leq \sqrt{e_{\alpha,\beta}(x)} + \frac{1}{\alpha}\left(u_i - l_i + \sqrt{2h_{\alpha,\beta}(x)/(1/\beta - 1/\alpha)}\right). \quad (18)$$

Proof: First, it is easily seen that the projection operator in this case is separable, and therefore the right-hand side of (15) is nonnegative componentwise, i.e.,

$$\left(\frac{1}{\alpha}(y_\alpha(x) - x)_i - \frac{1}{\beta}(y_\beta(x) - x)_i\right)(y_\beta(x) - y_\alpha(x))_i \geq 0, \quad i = 1, \ldots, n.$$

Since their sum equals $\epsilon := e_{\alpha,\beta}(x)$, then each term is below ϵ. Thus, for each $i \in \{1, \ldots, n\}$, either (i) $|\frac{1}{\alpha}(y_\alpha(x) - x)_i - \frac{1}{\beta}(y_\beta(x) - x)_i| \leq \sqrt{\epsilon}$ or (ii) $|(y_\beta(x) - y_\alpha(x))_i| \leq \sqrt{\epsilon}$. In case (ii),

$$\sqrt{\epsilon} \geq |(y_\alpha(x) - y_\beta(x))_i| = |(y_\alpha(x) - x)_i + (x - y_\beta(x))_i|$$
$$\geq |(y_\alpha(x) - x)_i| - |(x - y_\beta(x))_i|,$$

so that (10) yields

$$|(y_\alpha(x) - x)_i| \leq \sqrt{\epsilon} + |(x - y_\beta(x))_i| \leq \sqrt{\epsilon} + \sqrt{2h_{\alpha,\beta}(x)/(1/\beta - 1/\alpha)}.$$

In case (i),

$$\sqrt{\epsilon} \geq \left|\frac{1}{\beta}(y_\beta(x) - x)_i - \frac{1}{\alpha}(y_\alpha(x) - x)_i\right| \geq \frac{1}{\beta}|(y_\beta(x) - x)_i| - \frac{1}{\alpha}|(y_\alpha(x) - x)_i|,$$

so that (10) yields

$$\begin{aligned}\frac{1}{\beta}|(y_\beta(x)-x)_i| &\le \sqrt{\epsilon}+\frac{1}{\alpha}|(y_\alpha(x)-x)_i|\\ &\le \sqrt{\epsilon}+\frac{1}{\alpha}(|(y_\alpha(x)-y_\beta(x))_i|+|(y_\beta(x)-x)_i|)\\ &\le \sqrt{\epsilon}+\frac{1}{\alpha}(u_i-l_i+\sqrt{2h_{\alpha,\beta}(x)/(1/\beta-1/\alpha)}).\end{aligned}$$

□

Lemmas 2 and 3 show that the accuracy of solution x depends on α,β. In particular, to improve on the accuracy, we need to increase α and/or decrease β and decrease $h_{\alpha,\beta}(x)/(1/\beta-1/\alpha)$. The next lemma shows that, for x fixed, we can increase α and decrease β without increasing $h_{\alpha,\beta}(x)/(1/\beta-1/\alpha)$. In fact, the latter would go to zero when $x\in X$. This is key to developing a strategy for adjusting α and β to improve the accuracy of solution.

Lemma 4. *Assume X is bounded. For any $\alpha>\beta>0$ and $x\in\Re^n$, we have*

$$\lim_{\beta'\to 0}\sup_{\alpha'\ge\alpha} h_{\alpha',\beta'}(x)/(1/\beta'-1/\alpha') \;\le\; h_{\alpha,\beta}(x)/(1/\beta-1/\alpha),$$

with equality holding only if $y_\alpha(x)=y_\beta(x)=\mathrm{Proj}_X[x]$. If $x\in X$, then the left-hand side equals zero.

Proof: Let $\bar{x}=\mathrm{Proj}_X[x]$. We have from (10) that

$$\begin{aligned}2h_{\alpha,\beta}(x)/(1/\beta-1/\alpha) &\ge \|x-y_\beta(x)\|^2+\frac{1}{\alpha(1/\beta-1/\alpha)}\|y_\alpha(x)-y_\beta(x)\|^2\\ &\ge \|x-\bar{x}\|^2,\end{aligned}$$

with equality holding throughout only if $y_\alpha(x)=y_\beta(x)=\bar{x}$. [Notice that $\|x-y_\beta(x)\|\ge\|x-\bar{x}\|$, with equality holding if and only if $y_\beta(x)=\bar{x}$.] Using (2) and (3), we also have for all $\alpha'\ge\alpha$ and $0<\beta'\le\beta$ that

$$\begin{aligned}&2h_{\alpha',\beta'}(x)/(1/\beta'-1/\alpha')\\ &\quad\le 2h_{\alpha',\beta'}(x)/(1/\beta'-1/\alpha)\\ &\quad=\left(2F(x)^T(y_{\beta'}(x)-y_{\alpha'}(x))-\frac{1}{\alpha'}\|x-y_{\alpha'}(x)\|^2+\frac{1}{\beta'}\|x-y_{\beta'}(x)\|^2\right)\\ &\qquad\times(1/\beta'-1/\alpha)^{-1}\\ &\quad\le\left(2\|F(x)\|\mathrm{diam}(X)+\frac{1}{\beta'}\|x-y_{\beta'}(x)\|^2\right)\Big/(1/\beta'-1/\alpha).\end{aligned}$$

The right-hand side approaches $\|x-\bar{x}\|^2$ as $\beta'\to 0$, yielding an upper bound. A similar argument shows this is also a lower bound. If $x\in X$, then $x=\bar{x}$ and the right-hand side approaches zero as $\beta' \to 0$. [In fact, it can be seen that $\|x - y_{\beta'}(x)\| \le \beta' \|F(x)\|$, so the right-hand side is in the order of β'.] □

Recall that a stationary point x of $h_{\alpha,\beta}$ is a global minimum whenever $\nabla F(x)$ is positive definite [20, 30, 31]. The following lemma extends this result to give an explicit bound on $h_{\alpha,\beta}(x)$ in terms of $\|\nabla h_{\alpha,\beta}(x)\|$ whenever $\nabla F(x)$ is positive definite (x need not be a stationary point). This bound will suggest good choices of the function ϕ used in Algorithm 2 (see Theorem 3(b)).

Lemma 5. *For any $\alpha > \beta > 0$ and $x \in \Re^n$, if $\min_{\|u\|=1} u^T \nabla F(x) u \ge \sigma$ for some $\sigma > 0$, then*

$$\frac{h_{\alpha,\beta}(x)}{1/\beta - 1/\alpha} \le \frac{1}{2}\left(\frac{\|\nabla F(x)\|}{\sigma} + \frac{1}{\beta\sigma} + 1\right)^2 \left(\frac{\|\nabla h_{\alpha,\beta}(x)\|}{1/\beta - 1/\alpha}\right)^2.$$

Proof: First note that by (11),

$$h_{\alpha,\beta}(x) \le \frac{1}{2}\left(\frac{1}{\beta} - \frac{1}{\alpha}\right)\|x - y_\alpha(x)\|^2. \tag{19}$$

We have from (15) that

$$\begin{aligned}\nabla h_{\alpha,\beta}(x)^T (y_\beta(x) - y_\alpha(x)) &= (y_\beta(x) - y_\alpha(x))^T \nabla F(x)(y_\beta(x) - y_\alpha(x)) + e_{\alpha,\beta}(x) \\ &\ge \sigma \|y_\beta(x) - y_\alpha(x)\|^2.\end{aligned}$$

Letting $\epsilon := \|\nabla h_{\alpha,\beta}(x)\|$, this implies

$$\|y_\beta(x) - y_\alpha(x)\| \le \|\nabla h_{\alpha,\beta}(x)\|/\sigma = \epsilon/\sigma,$$

which together with (14) yields

$$\begin{aligned}\epsilon &= \left\|\left(\nabla F(x) - \frac{1}{\beta}I\right)(y_\beta(x) - y_\alpha(x)) + \left(\frac{1}{\alpha} - \frac{1}{\beta}\right)(y_\alpha(x) - x)\right\| \\ &\ge \left(\frac{1}{\beta} - \frac{1}{\alpha}\right)\|y_\alpha(x) - x\| - \left\|\nabla F(x) - \frac{1}{\beta}I\right\| \|y_\beta(x) - y_\alpha(x)\| \\ &\ge \left(\frac{1}{\beta} - \frac{1}{\alpha}\right)\|y_\alpha(x) - x\| - \left(\|\nabla F(x)\| + \frac{1}{\beta}\right)\frac{\epsilon}{\sigma}.\end{aligned}$$

Rearranging terms gives

$$\left(\frac{1}{\beta} - \frac{1}{\alpha}\right)\|y_\alpha(x) - x\| \le \left(\frac{\|\nabla F(x)\|}{\sigma} + \frac{1}{\beta\sigma} + 1\right)\epsilon.$$

Combining the latter relation with (19) yields the desired result. □

4. Two methods based on D-gap function

We formally describe our first method below. We then analyze its convergence using Lemma 2.

Algorithm 1. *Choose any* $x^0 \in \Re^n$, *any* $\alpha^0 > \beta^0 > 0$. *Choose any sequences of numbers* $\epsilon^k > 0, \eta^k \geq 0, \theta^k \in [0, 1)$, $k = 1, 2, \ldots$, *such that*

$$\lim_{k\to\infty} \epsilon^k = \lim_{k\to\infty} \eta^k/(1-\theta^k) = 0, \quad \sum_{k=1}^{\infty}(1-\theta^k) = \infty. \tag{20}$$

For $k = 1, 2, \ldots$, *we iterate as follows*:
Iteration k**.** *Choose any* $\alpha^k \geq \alpha^{k-1}$. *Choose* $\beta^k \leq \beta^{k-1}$ *and* $\hat{x}^k \in \Re^n$ *satisfying*

$$h_{\alpha^k,\beta^k}(\hat{x}^k)/(1/\beta^k - 1/\alpha^k) \leq \eta^k + \theta^k h_{\alpha^{k-1},\beta^{k-1}}(x^{k-1})/(1/\beta^{k-1} - 1/\alpha^{k-1}), \tag{21}$$

Apply a descent method to the unconstrained minimization of the function h_{α^k,β^k}, *with* $\hat{x}^k$ *as the starting point and using* $y_{\alpha^k}(x) - y_{\beta^k}(x)$ *as a safeguard descent direction at* x, *until the method generates an* $x \in \Re^n$ *satisfying* $e_{\alpha^k,\beta^k}(x) \leq \epsilon^k$. *The resulting* x *is denoted* x^k.

Note 1. Since we apply a descent method at iteration k, we have $h_{\alpha^k,\beta^k}(x^k) \leq h_{\alpha^k,\beta^k}(\hat{x}^k)$, so (21) implies

$$h_{\alpha^k,\beta^k}(x^k)/(1/\beta^k - 1/\alpha^k) \leq \eta^k + \theta^k h_{\alpha^{k-1},\beta^{k-1}}(x^{k-1})/(1/\beta^{k-1} - 1/\alpha^{k-1})\,. \tag{22}$$

In fact, this and $e_{\alpha^k,\beta^k}(x^k) \leq \epsilon^k$ are the only conditions on x^k needed for the convergence proof of Theorem 2 so, in particular, the descent method can use nonmonotone linesearch for practical efficiency [11].

Note 2. There are many choices for ϵ^k, such as $\epsilon^k = 1/k$ or $\epsilon^k = (1/2)^k$, and similarly for η^k. We can also choose them adaptively, say, depending on x^{k-1}. For example, we can set $\epsilon^k = \epsilon^{k-1}$ if $\|x^{k-1} - y_{\beta^{k-1}}(x^{k-1})\|/\beta^{k-1}$ is below a desired threshold; otherwise we set $\epsilon^k = \epsilon^{k-1}/2$, say. Alternatively, as will be clear from the subsequent analysis, we can set ϵ^k to be any quantity that is known to go to zero whenever $a^k := h_{\alpha^k,\beta^k}(x^k)/(1/\beta^k - 1/\alpha^k)$ goes to zero. For example, ϵ^k can be any forcing function of a^k or of $\|x^k - y_{\beta^k}(x^k)\|$.

Note 3. There are many choices for α^k. For example, we can choose it adaptively such as

$$\alpha^k = \begin{cases} \alpha^{k-1} & \text{if } h_{\alpha^{k-1},\beta^{k-1}}(x^{k-1}) \leq \nu^{k-1} \\ \mu\alpha^{k-1} & \text{else,} \end{cases} \tag{23}$$

where $\mu > 1$ and $\{\nu^k\}_{k=0,1,\ldots}$ is a sequence of positive numbers tending to zero. Given α^k, we can choose $\hat{x}^k$ to be any element of X and then choose $\beta^k \leq \beta^{k-1}$ sufficiently small so that (21) holds. By Lemma 4, this is possible. For warm start, the choices

$$\hat{x}^k = y_{\beta^{k-1}}(x^{k-1}) \quad \text{or} \quad \hat{x}^k = \mathrm{Proj}_X[x^{k-1}]$$

are possibilities. If $h_{\alpha^k,\beta^k}(x^{k-1})/(1/\beta^k - 1/\alpha^k)$ is strictly less than the right-hand side of (21), we can alternatively choose

$$\hat{x}^k = x^{k-1}$$

and choose $\beta^k \leq \beta^{k-1}$ sufficiently small so that (21) holds. By Lemma 4, this is possible.

Note 4. Algorithm 1 is a *derivative-free* method in the sense that one need not compute the gradient of the D-gap function (and hence, the Jacobian of F) to obtain a direction of descent or to perform a linesearch along this direction. This is a useful feature for the case where the gradient is not available or is costly to compute. In this respect, the method is much in the spirit of the algorithms in [8, 15, 17, 29, 30].

Theorem 2. *Assume X is bounded. Let $\{(x^k, \alpha^k, \beta^k, \epsilon^k, \eta^k, \theta^k)\}_{k=0,1,\ldots}$ be generated by Algorithm* 1. *Then,* $\{x^k\}$ *is bounded and its every cluster point is in X. If we choose $\alpha^k \to \infty$ or α^k is chosen by* (23), *then every cluster point of $\{x^k\}$ is a solution of VI.*

Proof: Denote $a^k := h_{\alpha^k,\beta^k}(x^k)/(1/\beta^k - 1/\alpha^k)$. By (22), we have $a^k \leq \eta^k + \theta^k a^{k-1}$ for $k = 1, 2, \ldots,$ and it follows from (20) that $a^k \to 0$ [23, Lemma 3, p. 45]. For each $k \in \{1, 2, \ldots\}$, we have from Lemma 2 that (16) and (17) hold with $\alpha = \alpha^k, \beta = \beta^k$, $x = x^k$. This together with $e_{\alpha^k,\beta^k}(x^k) \leq \epsilon^k$ and $\|y_{\alpha^k}(x^k) - y_{\beta^k}(x^k)\| \leq \text{diam}(X)$ yields

$$\|x^k - y_{\beta^k}(x^k)\| \leq \sqrt{2a^k}, \quad r^k \leq g_{\alpha^k}(x^k) \leq r^k + \epsilon^k + \frac{\text{diam}(X)^2}{2\alpha^k}, \tag{24}$$

where $r^k := F(x^k)^T(x^k - y_{\beta^k}(x^k)) - \frac{1}{2\alpha^k}\|x^k - y_{\beta^k}(x^k)\|^2$. Since $a^k \to 0$, the first inequality in (24) implies x^k is bounded and its every cluster point is in X. Moreover, this also implies $r^k \to 0$.

Thus, if $\alpha^k \to \infty$, then the last two inequalities in (24) yield $g_{\alpha^k}(x^k) \to 0$. Since for each $y \in X$, we have from (2) that

$$g_{\alpha^k}(x^k) \geq F(x^k)^T(x^k - y) - \frac{1}{2\alpha^k}\|x^k - y\|^2,$$

this yields $0 \geq F(x^\infty)^T(x^\infty - y)$ for each cluster point x^∞ of $\{x^k\}$. Thus, each cluster point x^∞ solves VI.

Alternatively, if α^k is chosen by (23), with $\mu > 1$ and $\nu^k \to 0$, then either (i) $\alpha^k \to \infty$, which can be treated as above, or (ii) α^k is constant for all k exceeding some $\bar{k}$, which implies for all $k \geq \bar{k}$ that $h_{\alpha^{\bar{k}},\beta^k}(x^k) \leq \nu^k \to 0$. Since, by (10),

$$h_{\alpha^{\bar{k}},\beta^k}(x^k) \geq \frac{1}{2\alpha^{\bar{k}}}\|y_{\alpha^{\bar{k}}}(x^k) - y_{\beta^k}(x^k)\|^2 \quad \forall k \geq \bar{k},$$

it follows that in case (ii) $\|y_{\alpha^{\bar{k}}}(x^k) - y_{\beta^k}(x^k)\| \to 0$. Taking further into account the first relation in (24) and $a^k \to 0$, we conclude that $\|x^k - y_{\alpha^{\bar{k}}}(x^k)\| \to 0$ in this case, so each cluster point x^∞ of $\{x^k\}$ satisfies $x^\infty = y_{\alpha^{\bar{k}}}(x^\infty)$ and hence solves the VI. □

If in Algorithm 1 we choose $\eta^k = 0$ and $\theta^k \leq \theta < 1$ for all k, then $a^k \to 0$ linearly and, by (24), $\|x^k - y_{\beta^k}(x^k)\| \to 0$ linearly (in the root sense) as does r^k. Thus if we further choose $\epsilon^k \to 0$ and $1/\alpha^k \to 0$ linearly, then, by (24), $g_{\alpha^k}(x^k) \to 0$ linearly.

Below we consider a second method that increases α and decreases β whenever $\|\nabla h_{\alpha,\beta}(x)\|$ is too small relative to $h_{\alpha,\beta}(x)$ to achieve sufficient descent. Unlike Algorithm 1, this method always uses x^{k-1} to initialize iteration k.

Algorithm 2. *Choose any $x^0 \in \Re^n$, any $\alpha^0 > \beta^0 > 0$, and two sequences of nonnegative numbers ρ^k, η^k, $k = 1, 2, \ldots,$ such that*

$$\rho^k + \eta^k > 0 \; \forall k, \quad \sum_{k=1}^{\infty} \rho^k < \infty, \quad \sum_{k=1}^{\infty} \eta^k < \infty . \tag{25}$$

Choose any continuous function $\phi : \Re_+ \mapsto \Re_+$ with $\phi(t) = 0 \Leftrightarrow t = 0$ and any $\omega \in (0, 1)$. For $k = 1, 2, \ldots,$ we iterate as follows:

Iteration k**.** *Choose any $\alpha^k \geq \alpha^{k-1}$ and then choose $0 < \beta^k \leq \omega\beta^{k-1}$ satisfying*

$$h_{\alpha^k,\beta^k}(x^{k-1})/(1/\beta^k - 1/\alpha^k) \leq (1 + \rho^k) h_{\alpha^{k-1},\beta^{k-1}}(x^{k-1})/(1/\beta^{k-1} - 1/\alpha^{k-1}) + \eta^k. \tag{26}$$

Apply a descent method to the unconstrained minimization of the function h_{α^k,β^k}, with x^{k-1} as the starting point. We assume the descent method has the property that the amount of descent achieved at x per step is bounded away from zero whenever x is bounded and $\|\nabla h_{\alpha^k,\beta^k}(x)\|$ is bounded away from zero. Then, either the method in a finite number of steps generates an x satisfying

$$\|\nabla h_{\alpha^k,\beta^k}(x)\| \leq \phi(h_{\alpha^k,\beta^k}(x)/(1/\beta^k - 1/\alpha^k)) , \tag{27}$$

which we denote by x^k, or else $h_{\alpha^k,\beta^k}(x)$ must decrease towards zero,[3] *in which case any cluster point of x solves VI.*

Note 5. Since $\rho^k + \eta^k > 0$, Lemma 4 shows that, for any $\alpha^k \geq \alpha^{k-1}$, (26) is satisfied by choosing β^k sufficiently small.

Note 6. There are many choices for the descent method. For example, we can use any gradient method whose direction is, in the terminology of [4, Section 1.2], "gradient-related" and whose stepsize is chosen by the Armijo rule or the Goldstein rule. Such a method has the desired properties, as can be seen from the proofs of [4, Propositions 1.2.1 and 1.2.2].

Note 7. If the descent method is memoryless, Algorithm 2 may alternatively be written in a form similar to the Zhu-Marcotte method: Given k, α^k, β^k and x, update these four quantities as follows:

if (27), then (null step) set $x^k = x$, increment k by 1, choose $\alpha^k \geq \alpha^{k-1}$, and then choose $0 < \beta^k \leq \omega\beta^{k-1}$ satisfying (26). [x is unchanged.]

else (descent step) apply one step of the descent method to h_{α^k,β^k} at x to obtain direction $d \in \Re^n$ and stepsize $\lambda > 0$. Replace x by $x + \lambda d$. [k, α^k, β^k are unchanged.]

Analogous to Algorithm 1 (see Note 1), $h_{\alpha^k,\beta^k}(x^k) \le h_{\alpha^k,\beta^k}(x^{k-1})$ and (27) are the only conditions on x^k needed for the convergence proof of Theorem 3 so, in particular, the descent method can use nonmonotone linesearch.

Note 8. We can modify the criterion (27) by adding to its right-hand side a positive scalar ϵ^k tending to zero. This modified criterion does not affect the convergence result below (i.e., Theorem 3(a)) and is always satisfied after a finite number of steps of the descent method. However, this would entail $\beta^k \to 0$ always and would preclude the possibility of finding a solution of the VI at some iteration k, i.e., a finite number of null steps.

Theorem 3. *Assume X is bounded. Let* $\{(x^k, \alpha^k, \beta^k, \rho^k, \eta^k)\}_{k=0,1,\ldots}$ *be generated by Algorithm* 2.
(*a*). *Suppose* x^k *is obtained for all* k. *Then*, $\{x^k\}$ *is bounded*, $\beta^k \to 0$, *and every cluster point of* $\{x^k\}$ *is in* X. *If we choose* $\alpha^k \to \infty$ *or* α^k *is chosen by* (23), *then every cluster point of* $\{x^k\}$ *is a solution of VI.*
(*b*). *Suppose* x^k *is not obtained for some* k. *Then*, *the descent method generates a bounded sequence of* x *with* $h_{\alpha^k,\beta^k}(x) \to 0$ (*so every cluster point of* x *solves the VI*). *In particular*, *this occurs when* F *is strongly monotone and we choose* ϕ *such that* $\lim_{t\to 0}\sup \phi(t)/\sqrt{t} = 0$.

Proof: (a). Since we use a descent method at iteration k to obtain x^k from x^{k-1}, then $h_{\alpha^k,\beta^k}(x^k) \le h_{\alpha^k,\beta^k}(x^{k-1})$, so (26) yields

$$h_{\alpha^k,\beta^k}(x^k)/(1/\beta^k - 1/\alpha^k) \le (1+\rho^k)h_{\alpha^{k-1},\beta^{k-1}}(x^{k-1})/(1/\beta^{k-1} - 1/\alpha^{k-1}) + \eta^k .$$

Denote $a^k := h_{\alpha^k,\beta^k}(x^k)/(1/\beta^k - 1/\alpha^k)$. This can then be written as $a^k \le (1+\rho^k)a^{k-1} + \eta^k$ for $k = 1, 2, \ldots$ Using $a^k \ge 0$ and (25), it follows that the sequence $\{a^k\}$ converges to some $\bar{a} \ge 0$ [23, Lemma 2, p. 44]. Since (16) implies

$$\|x^k - y_{\beta^k}(x^k)\| \le \sqrt{2a^k}, \quad \forall k, \tag{28}$$

the sequence $\{x^k\}$ is bounded.

We claim that $\bar{a} = 0$. Suppose the contrary. Then for all k sufficiently large, it holds that $a^k \ge \bar{a}/2 > 0$. Then, (2) and (3) imply

$$\frac{\bar{a}}{2} \le \left(2F(x^k)^T(y_{\beta^k}(x^k) - y_{\alpha^k}(x^k)) - \frac{1}{\alpha^k}\|x^k - y_{\alpha^k}(x^k)\|^2 + \frac{1}{\beta^k}\|x^k - y_{\beta^k}(x^k)\|^2\right) \times (1/\beta^k - 1/\alpha^k)^{-1} .$$

Since, by the construction of the algorithm, $\beta^k \to 0$ and $\alpha^k \ge \alpha^0$, and we already established that x^k is bounded (as are $y_{\beta^k}(x^k)$ and $y_{\alpha^k}(x^k)$), we obtain in the limit

$$0 < \frac{\bar{a}}{2} \le \liminf_{k\to\infty} \|x^k - y_{\beta^k}(x^k)\|^2 .$$

Then $\lim_{k\to\infty} \|x^k - y_{\beta^k}(x^k)\|/\beta^k = \infty$, so (14) implies

$$\lim_{k\to\infty} \|\nabla h_{\alpha^k,\beta^k}(x^k)\| = \infty.$$

Since x^k satisfies (27) so that $\|\nabla h_{\alpha^k,\beta^k}(x^k)\| \le \phi(a^k)$ for all k, this contradicts convergence of $\{\phi(a^k)\}$ (recall that ϕ is a continuous function). Hence, $\bar{a} = 0$.

For each $k \in \{1, 2, \ldots\}$, we have from Lemma 2 that (17) holds with $\alpha = \alpha^k$, $\beta = \beta^k$, $x = x^k$ and from (15) that $e_{\alpha^k,\beta^k}(x^k) \le \epsilon^k := \nabla h_{\alpha^k,\beta^k}(x^k)^T(y_{\beta^k}(x^k) - y_{\alpha^k}(x^k))$. This together with $\|y_{\alpha^k}(x^k) - y_{\beta^k}(x^k)\| \le \operatorname{diam}(X)$ yields

$$r^k \le g_{\alpha^k}(x^k) \le r^k + \epsilon^k + \frac{\operatorname{diam}(X)^2}{2\alpha^k}, \tag{29}$$

where $r^k := F(x^k)^T(x^k - y_{\beta^k}(x^k)) - \frac{1}{2\alpha^k}\|x^k - y_{\beta^k}(x^k)\|^2$. Since $a^k \to 0$, (28) implies $\{x^k\}$ is bounded and every one of its cluster points is in X. Moreover, (28) implies $r^k \to 0$. Also, we have $\|\nabla h_{\alpha^k,\beta^k}(x^k)\| \le \phi(a^k) \to 0$, so $\epsilon^k \to 0$.

If $\alpha^k \to \infty$, then (29) and $r^k \to 0$, $\epsilon^k \to 0$ yield $g_{\alpha^k}(x^k) \to 0$. Since for each $y \in X$, we have from (2) that

$$g_{\alpha^k}(x^k) \ge F(x^k)^T(x^k - y) - \frac{1}{2\alpha^k}\|x^k - y\|^2,$$

this yields $0 \ge F(x^\infty)^T(x^\infty - y)$ for each cluster point x^∞ of $\{x^k\}$. Thus, each cluster point x^∞ solves VI.

If α^k is chosen by (23), then, by an identical argument as in the last paragraph of the proof of Theorem 2, we obtain that each cluster point of $\{x^k\}$ is a solution of the VI.

(b). It suffices to show that x^k is not obtained for some k under the assumptions on F and ϕ. Suppose the contrary, so x^k is obtained for all k. Then, as in (a), we obtain

$$\|\nabla h_{\alpha^k,\beta^k}(x^k)\| \le \phi(a^k) \ \ \forall k, \quad a^k \to 0\,, \tag{30}$$

and $\beta^k \to 0$, $\{x^k\}$ is bounded. Since $\lim_{t\to 0}\sup \phi(t)/\sqrt{t} = 0$, then $\phi(a^k)/\sqrt{a^k} \to 0$. Also, strong monotonicity of F and Lemma 5 imply that there exists $\sigma > 0$ such that

$$\begin{aligned}\phi(a^k) &= \frac{\phi(a^k)}{\sqrt{a^k}}\sqrt{a^k} \\ &\le \frac{\phi(a^k)}{\sqrt{a^k}}\left(\frac{\|\nabla F(x^k)\|/\sigma + 1/(\beta^k\sigma) + 1}{1/\beta^k - 1/\alpha^k}\right)\|\nabla h_{\alpha^k,\beta^k}(x^k)\| \ \ \forall k.\end{aligned} \tag{31}$$

Since $\{x^k\}$ is bounded, $\beta^k \to 0$, and $\alpha^k \ge \alpha^0$, it follows that the right-hand side of (31) tends to zero faster than $\|\nabla h_{\alpha^k,\beta^k}(x^k)\|$, and hence so does $\phi(a^k)$. On the other hand, by (30) $\phi(a^k)$ tends to zero no faster than $\|\nabla h_{\alpha^k,\beta^k}(x^k)\|$, which gives a contradiction. □

Observe further that if F is strongly monotone, then $h_{\alpha^k,\beta^k}(x) \to 0$ (with k fixed) immediately implies that x converges to the unique solution of VI. In fact, the rate of convergence in this case is linear, assuming that at every step the descent method achieves decrease of the objective function in the order of $\|\nabla h_{\alpha^k,\beta^k}(x)\|^2$ at x (which is true for most gradient-type descent methods using an Armijo-type stepsize rule). This is because Lemma 5 implies that the amount of decrease is in the order of $h_{\alpha^k,\beta^k}(x)$, from which linear convergence of $h_{\alpha^k,\beta^k}(x)$ readily follows. Since F being strongly monotone and X being bounded implies that the distance from x to the solution of VI is in the order of $\sqrt{h_{\alpha^k,\beta^k}(x)}$ [31], it follows that x converges to the solution at a linear rate (in the root sense).

We conclude this section with a finite termination result, assuming the following *nondegeneracy* condition: there exists a solution x^* of VI satisfying

$$-F(x^*) \in \operatorname{int} N_X(x^*)\,, \tag{32}$$

where N_X denotes the normal cone to the (convex) set X, and "int" stands for its interior. Note that this condition and the monotonicity of F imply that the solution is unique. Condition (32) was used in [1] to force finite termination of convergent algorithms by modifying them to periodically solve an auxiliary linearized subproblem. Here, we show that no such modification is needed to obtain finite termination. A weaker nondegeneracy condition is used in the recent paper [18]. However, there it assumes that F is "pseudomonotone$^+$" which need not hold under our assumption of F being monotone. We note that finite termination has been studied fairly extensively in the context of optimization problems (see, e.g., [5] and references therein).

Theorem 4. *Suppose F is monotone and continuous, and there exists a solution x^* of VI satisfying* (32). *Let $\{x^k\}_{k=0,1,\dots}$ be any bounded sequence of points in $\Re^n$ such that every cluster point of $\{x^k\}$ is a solution of VI. Then, for each $\bar\alpha > 0$, there exists an index $\bar k$ such that $y_\alpha(x^k) = x^*$ for all $\alpha \geq \bar\alpha$ and $k \geq \bar k$.*

Proof: Fix any $\bar\alpha > 0$. Since every cluster point of $\{x^k\}$ is a solution of VI, and x^* is the unique solution by (32), it follows that $\{x^k\}$ converges to x^*. Furthermore, (32) implies there exists some $\delta > 0$ such that

$$-F(x^*) + b \in N_X(x^*) \quad \forall b \text{ with } \|b\| \leq \delta\,.$$

This remains true when the left-hand side is scaled by any $\alpha > 0$ and it can be written equivalently as

$$x^* = \operatorname{Proj}_X[x^* - \alpha F(x^*) + \alpha b] \quad \forall \alpha > 0,\ \forall b \text{ with } \|b\| \leq \delta\,. \tag{33}$$

On the other hand,

$$y_\alpha(x^k) = \operatorname{Proj}_X[x^k - \alpha F(x^k)] = \operatorname{Proj}_X[x^* - \alpha F(x^*) + \alpha b^k]\,, \tag{34}$$

where $b^k := (x^k - x^*)/\alpha - F(x^k) + F(x^*)$. For $\alpha \geq \bar\alpha$, we further obtain

$$\|b^k\| \leq \|x^k - x^*\|/\bar\alpha + \|F(x^k) - F(x^*)\| \leq \delta\,,$$

where the last inequality holds for all k exceeding some index $\bar{k}$, due to $\{x^k\}$ converging to x^* and F being continuous. Then (33) and (34) yield that $y_\alpha(x^k) = x^*$ for all $\alpha \geq \bar{\alpha}$ and $k \geq \bar{k}$. □

Theorem 4 establishes finite termination of any convergent algorithm for solving a monotone VI satisfying the nondegeneracy condition (32), provided one monitors the value of the projection residual in the stopping test. For example, for $\{x^k\}$ generated by either of our two algorithms, we can test periodically whether $y_{\alpha^k}(x^k)$ or $y_{\beta^k}(x^k)$ satisfies $y_1(x) = x$ and terminate accordingly.

5. Methods based on D-gap function applied to regularized VI

Another approach to solving monotone VI is to regularize the problem by adding to the mapping F a positive δ multiple of the identity mapping (or any Lipschitz continuous strongly monotone mapping), thus obtaining the strongly monotone mapping:

$$F^\delta(x) = F(x) + \delta x \quad \forall x \in \Re^n,$$

and to use the D-gap function to solve the regularized VI approximately. As the regularization parameter δ and the solution accuracy go to zero, any cluster point of the approximate solution of the regularized problem would approach a solution of the original VI. Billups and Ferris [6] reported good numerical experience using this regularization technique in a QP-based method for solving box-constrained VI. We study this in more depth below.

We define $y_\alpha^\delta(x)$, $g_\alpha^\delta(x)$, $h_{\alpha,\beta}^\delta(x)$ analogously as in (1), (2), (3), but with F replaced by F^δ. This leads to the following method:

Algorithm 3. *Choose any $x^0 \in \Re^n$, any $\alpha^0 > \beta^0 > 0$. Choose any sequence of positive numbers ϵ^k, δ^k, $k = 1, 2, \ldots,$ tending to zero. For $k = 1, 2, \ldots,$ we iterate as follows*:

Iteration k**.** *Choose an $\alpha^k \geq \alpha^{k-1}$, an $\beta^k \leq \beta^{k-1}$ and an $\hat{x}^k \in \Re^n$. Apply a descent method to the unconstrained minimization of the function $h_{\alpha^k,\beta^k}^{\delta^k}$, with $\hat{x}^k$ as the starting point. We assume the descent method has the property that the amount of descent achieved at x per step is bounded away from zero whenever x is bounded and $\|\nabla h_{\alpha^k,\beta^k}^{\delta^k}(x)\|$ is bounded away from zero. Then, the method in a finite number of steps generates an x satisfying*

$$h_{\alpha^k,\beta^k}^{\delta^k}(x) \leq \epsilon^k. \tag{35}$$

The resulting x is denoted x^k.[4]

Note 9. There are many choices for the descent method used in Algorithm 3, as was discussed in Note 6. For example, it can be any descent method using $-\nabla h_{\alpha^k,\beta^k}^{\delta^k}(x)$ as a safeguard direction at x and an Armijo-type stepsize rule, such as Algorithm 7.1 in [11]. The derivative-free descent method proposed in [31] is another possibility. Also, one can replace the left-hand side in the accuracy criterion (35) by either the projection residual $\|x - y_1^{\delta^k}(x)\|$ or $\|\nabla h_{\alpha^k,\beta^k}^{\delta^k}(x)\|$.

Note 10. Instead of using the Tikhonov regularization $F^{\delta^k}(x) = F(x) + \delta^k x$ at iteration k, we can alternatively use the proximal regularization

$$F^{\delta^k}(x) = F(x) + \delta^k(x - x^{k-1}).$$

The convergence result below still applies and, in fact, can be strengthened by invoking convergence theory for (inexact) proximal point algorithms. Furthermore, in that case δ^k need not go to zero and more adaptive accuracy criteria involving x^{k-1} instead of ϵ^k can be used (e.g., [25]).

Note 11. There are many choices for α^k, β^k and $\hat{x}^k$. For warm start, we can choose

$$\hat{x}^k = x^{k-1}.$$

However, unlike Algorithms 1 and 2 (see Lemma 4), we do not have active control on the growth in $h^\delta_{\alpha,\beta}(x)$ as we adjust δ.

Theorem 5. *Let* $\{(x^k, \alpha^k, \beta^k, \epsilon^k, \delta^k)\}_{k=0,1,\ldots}$ *be generated by Algorithm* 3. *If the parameters are chosen so that*

$$\limsup_{k\to\infty} \alpha^k < \infty \text{ or } \lim_{k\to\infty} \epsilon^k/\beta^k = 0\,,$$

then every cluster point of $\{x^k\}$ *is a solution of VI.*

Proof: Since x^k satisfies (35) and $\epsilon^k \to 0$, then $h^{\delta^k}_{\alpha^k,\beta^k}(x^k) \to 0$. Using (10) with $h^\delta_{\alpha,\beta}$ in place of $h_{\alpha,\beta}$, we then obtain

$$\left(\frac{1}{\beta^k} - \frac{1}{\alpha^k}\right)\left\|x^k - y^{\delta^k}_{\beta^k}(x^k)\right\|^2 \to 0, \qquad \frac{1}{\alpha^k}\left\|y^{\delta^k}_{\alpha^k}(x^k) - y^{\delta^k}_{\beta^k}(x^k)\right\|^2 \to 0. \tag{36}$$

In particular, the first inequality implies that $\{x^k\}$ is bounded.

If $\lim_{k\to\infty}\sup \alpha^k < \infty$, then (36) together with $\beta^k \le \beta^0$, $\delta^k \to 0$ and continuity of $y^\delta_\beta(x)$ in (x, β, δ) would imply that any cluster point x of $\{x^k\}$ satisfies $\|x - y_\beta(x)\| = 0$ and $\|x_\alpha(x) - y_\beta(x)\| = 0$ for some $\alpha > 0$ and some $\beta \ge 0$. Hence x solves VI.

If $\lim_{k\to\infty} \epsilon^k/\beta^k = 0$, then using (10) with $h^\delta_{\alpha,\beta}$ in place of $h_{\alpha,\beta}$ gives

$$\frac{1}{\beta^k}\left\|x^k - y^{\delta^k}_{\beta^k}(x^k)\right\| \le \frac{1}{\beta^k}\sqrt{\frac{2h^{\delta^k}_{\alpha^k,\beta^k}(x^k)}{1/\beta^k - 1/\alpha^k}} \;\le\; \sqrt{\frac{2\epsilon^k}{\beta^k - (\beta^k)^2/\alpha^k}}\,,$$

where the second inequality is due to x^k satisfying (35). Since $\epsilon^k/\beta^k \to 0$ and $\alpha^k \ge \alpha^0$, this implies $\|x^k - y^{\delta^k}_{\beta^k}(x^k)\|/\beta^k \to 0$. By definition of $y^{\delta^k}_{\beta^k}(x^k)$ and properties of projection, we have

$$\left(F(x^k) + \delta^k x^k + \left(y^{\delta^k}_{\beta^k}(x^k) - x^k\right)/\beta^k\right)^T\left(y - y^{\delta^k}_{\beta^k}(x^k)\right) \ge 0 \quad \forall y \in X,$$

so this together with $\{x^k\}$ being bounded and $\delta^k \to 0$ yields that every cluster point x of $\{x^k\}$ satisfies $F(x)^T(y-x) \geq 0$ for all $y \in X$. Hence x solves VI. □

6. Preliminary numerical results

To gain some understanding of the numerical behavior/performance of Algorithms 1 and 2, we implemented these methods in Matlab and ran them on a set of test problems. In this section we describe the implementation and report on our preliminary numerical experience with them.

In our Matlab implementation of Algorithm 2, we choose the parameters

$$\rho^k := 1/k^2, \quad \eta^k := 0, \quad \phi(t) := t^2, \quad \omega := .5,$$

To ensure β is not decreased too rapidly, we modify (27) by taking the minimum of the right-hand side with $.01\|x - y_1(x)\|$. It can be seen that this does not affect the convergence properties. We adjust α^k by (23) with $\mu := 2$ and $\nu^k := \|x^0 - y_1(x^0)\|/\ln(k+1)$, and we choose β^k to be the largest element of $\{\omega\beta^{k-1}, \omega^2\beta^{k-1}, \ldots\}$ for which (26) holds. For the descent method, we use a slightly modified version of the Gauss-Newton method of [11], which was shown to have good numerical performance on MCPLIB problems. More precisely, for a fixed $\alpha > \beta > 0$, the method minimizes $h_{\alpha,\beta}$ by successively applying the (nonmonotone) descent step:

$$x^{new} := x + td, \tag{37}$$

where t is the largest element of $\{1, .1, (.1)^2, \ldots\}$ such that

$$h_{\alpha,\beta}(x+td) \leq \mathcal{R} + 10^{-4} t d^T \nabla h_{\alpha,\beta}(x), \tag{38}$$

and

$$d := \begin{cases} \tilde{d} & \text{if } \tilde{d}^T \nabla h_{\alpha,\beta}(x) \leq -10^{-8}\|\tilde{d}\|^{2.1} \\ \hat{d} & \text{else if } \hat{d}^T \nabla h_{\alpha,\beta}(x) \leq -10^{-8}\|\hat{d}\|^{2.1} \\ -\nabla h_{\alpha,\beta}(x) & \text{else} \end{cases}$$

Here $\mathcal{R}$ is updated as described in [11, pp. 81, 83]; $\tilde{V}$ is an element of $\tilde{\mathcal{V}}(x)$ as defined in [11, p. 75];

$$\tilde{V}\tilde{d} = -\nabla h_{\alpha,\beta}(x),$$

if $\text{cond}(\tilde{V}) \leq 10^9$ and otherwise $\tilde{d}$ is obtained by applying preconditioned conjugate gradient (CG) algorithm to the above equation until either the residual is below the threshold given in [11, p. 81] or the number of CG steps exceeds $2n$; and

$$\hat{d} = -(I + 10\tilde{V}/\|\tilde{V}\|)^{-1}\nabla h_{\alpha,\beta}(x).$$

Table 1. Performance of Algorithm 2 on eight test problems, as indicated by the number of iterations to termination (nit), the number of descent steps (nstep), the number of F-evaluations (nf), and the residual $\|x - y_1(x)\|$ upon termination (res).

Name	n	Problem (nit/nstep/nf/res)		
		$x^0 = (.1, \dots, .1)^T$	$x^0 = (1, \dots, 1)^T$	$x^0 = (10, \dots, 10)^T$
Yamashita-Fukushima	1	0/5/6/1 $\times 10^{-4}$	22/15/48/3$\times 10^{-5}$	2/8/13/7$\times 10^{-4}$
Kojima-Shindo	4	1/26/43/2 $\times 10^{-4}$	0/10/16/3$\times 10^{-7}$	1/24/38/2$\times 10^{-4}$
Watson	5	0/59/82/3$\times 10^{-9}$	0/15/15/1$\times 10^{-4}$	0/89/600/1$\times 10^{-4}$
Nash-Cournot	10	0/9/9/3 $\times 10^{-5}$	0/6/6/8$\times 10^{-6}$	0/24/46/5$\times 10^{-5}$
Lemke	30	0/55/102/1 $\times 10^{-16}$	0/1/1/0.00	0/1/1/4$\times 10^{-14}$
Harker-Pang	30	0/8/15/6 $\times 10^{-11}$	0/7/12/4$\times 10^{-10}$	0/10/18/6$\times 10^{-11}$
SkewSymLCP	30	0/23/32/1 $\times 10^{-6}$	0/54/94/2$\times 10^{-6}$	0/44/72/7$\times 10^{-7}$
SymPSDLCP	30	0/7/7/4 $\times 10^{-6}$	0/11/11/5$\times 10^{-4}$	0/22/36/2$\times 10^{-4}$

The regularized Gauss-Newton direction $\hat{d}$, not used in [11], helps to accelerate the convergence on a few problems. For problems where F is undefined or highly nonmonotone outside of $\Re^n_+$, we use a gradient projection technique to maintain x nonnegative. In particular, we replace "$x + td$" in (37) and (38) by "$\max\{0, x + td\}$" and accordingly replace "td^T" in (38) by "$(\max\{0, x + td\} - x)^T$". This technique was beneficial on the Kojima-Shindo problem. We also experimented with an alternative technique described in [11, p. 83], but it was less successful. We remark that, while the above parameter choices are reasonable, they were made without much fine-tuning and can conceivably be improved.

In our tests, we set the starting point x^0 to be $(1, \dots, 1)^T$ multiplied by either .1 or 1 or 10, and choose $\alpha^0 = 1/.9$ and $\beta^0 = 1/1.1$, as suggested by [11]. We terminate a method whenever $\|x - y_1(x)\| \leq 10^{-3}$ is satisfied. For comparison, we also ran the Gauss-Newton method by itself (with fixed $\alpha = 1/.9$, $\beta = 1/1.1$). The performance of the methods on a set of eight NCP test problems is summarized in Table 1. Since the feasible set should be bounded, we put an artificial upper bound of 10^5 on each variable, so for all our problems, $X = [0, 10^5]^n$. For the first problem, F is given by (4). For the remaining seven problems, F is as described in [15, §5].

As can be seen from Tables 1 and 2, Algorithm 2 has better performance than the Gauss-Newton method in three cases and in other cases the two methods have identical performance. In the three cases, the Gauss-Newton method converged to a non-solution that is a stationary point of $h_{\alpha,\beta}$, and at which ∇F is not positive definite but positive semidefinite (or nearly so). In two cases, only β is adjusted in Algorithm 2, but on the Yamashita-Fukushima problem, adjusting α is also needed (and the correct solution $x = 2$ is generated). This shows that dynamically adjusting α and β can perhaps improve the performance of descent methods based on the D-gap function, specifically when these methods are attracted to stationary points which are not solutions of the underlying VI. For most cases, the number of steps (each step requires one evaluation of the Jacobian $\nabla F(x)$ and one/two $n \times n$ linear-equations solve), seems reasonable for the level of accuracy

Table 2. Performance of Gauss-Newton method on eight test problems, as indicated by the the number of descent steps (nstep), the number of F-evaluations (nf), and the residual $\|x - y_1(x)\|$ upon termination (res).

		Problem (nstep/nf/res)		
Name	n	$x^0 = (.1, \ldots, .1)^T$	$x^0 = (1, \ldots, 1)^T$	$x^0 = (10, \ldots, 10)^T$
Yamashita-Fukushima	1	$5/6/1\times10^{-4}$	stationary[a]	$8/13/7\times10^{-4}$
Kojima-Shindo	4	stationary[b]	$10/16/3\times10^{-7}$	stationary[b]
Watson	5	$59/82/3\times10^{-9}$	$15/15/1\times10^{-9}$	$89/600/1\times10^{-4}$
Nash-Cournot	10	$9/9/3\times10^{-5}$	$6/6/8\times10^{-6}$	$24/46/5\times10^{-5}$
Lemke	30	$55/102/1\times10^{-16}$	1/1/0.00	$1/1/4\times10^{-14}$
Harker-Pang	30	$8/15/6\times10^{-11}$	$7/12/4\times10^{-10}$	$10/18/6\times10^{-11}$
SkewSymLCP	30	$23/32/1\times10^{-6}$	$54/94/2\times10^{-6}$	$44/72/7\times10^{-7}$
SymPSDLCP	30	$7/7/4\times10^{-6}$	$11/11/5\times10^{-4}$	$22/36/2\times10^{-4}$

[a]Iterates remain at non-solution $x = 1$.
[b]Iterates converge to non-solution $x = (0, 0, 0.54, 1.64)^T$.

achieved. The slow convergence of the methods on SkewSymLCP and on Lemke (with starting point $(.1, \ldots, .1)^T$) appears to be due to $\tilde{V}$ having a large condition number, in the order of 10^6, during the early iterations. Similar difficulty is encountered on the Watson problem with poor starting point. For Algorithm 2 and Gauss-Newton method, the direction $\hat{d}$ and the safeguard steepest-descent direction $-\nabla h_{\alpha,\beta}(x)$ were invoked on the first three test problems only.

We also experimented with a derivative-free version of Algorithm 1, but it was overall less successful. In our implementation, we chose $\eta^k = 0,\ \theta^k = .9$ for all k. The parameters $\alpha^k,\ \beta^k$ were obtained by increasing α^{k-1} by multiples of 2 and decreasing β^{k-1} by factors of 1.2 until (21) is satisfied. The descent method was based on the derivative-free direction $y_{\alpha^k}(x) - y_{\beta^k}(x)$ and a monotone linesearch rule. The descent method was terminated, and $\alpha^k,\ \beta^k$ updated, when the criterion specified in Algorithm 4 was satisfied with $\epsilon^k = e_{\alpha^k,\beta^k}(\hat{x}^k)/10$. The method was restarted from $\hat{x}^k = \mathrm{Proj}_X[x^{k-1}]$, although this restart was rarely necessary unless the initial x^0 is chosen outside of X. For smaller problems (e.g., our first four test problems), the performance was satisfactory and not much different from Algorithm 2. However, convergence on bigger LCPs was rather slow. We therefore would not recommend using the derivative-free version of Algorithm 1 unless computing the derivatives of F is impossible or cost-prohibitive. When we changed the descent method in this algorithm to the Gauss-Newton method described above, the performance was very similar to that reported for Algorithm 2. However, in that experiment we also changed the safeguard descent direction to the direction of steepest descent. So strictly speaking, this implementation is not within the framework of Algorithm 1.

To summarize, our numerical experience seems to confirm that dynamically adjusting α and β while minimizing the D-gap function $h_{\alpha,\beta}$ helps to ensure convergence to solutions of the underlying VI. This is precisely one the main messages of this paper. Our strategy is of particular relevance in the case when the D-gap function has stationary points which are not solutions of VI (i.e., not global minima of the D-gap function).

Acknowledgments

Research of the first author is supported by CNPq Grant 300734/95-6 and by PRONEX–Optimization. The second author is supported by National Science Foundation Grant CCR-9731273.

Notes

1. Why? Let S' denote the solution set of the new VI. It is easy to see that $S \cap B \subset S'$. To see the converse, consider any $x^* \in S' \cap \text{int}B$. Such x^* exists by our choice of B. Then, for any $y \in X$ we can find $t \in (0, 1]$ such that $y^t = x^* + t(y - x^*) \in B$, yielding $y^t \in X \cap B$ and hence $0 \le F(x^*)^T(y^t - x^*) = tF(x^*)^T(y - x^*)$, implying $x^* \in S \cap B$. Thus, $S' \cap \text{int}B \subset S \cap B$. Since S' is closed convex, it is the closure of $S' \cap \text{int}B$. Since $S \cap B$ is closed, this shows $S' \subset S \cap B$.
2. As α is increased, the right-hand side of the inequality describing case (ii) tends to zero, while the left-hand side is nondecreasing and hence is bounded away from zero (assuming x is not a solution of VI). Thus, for α sufficiently large, case (ii) cannot occur and so case (i) must occur.
3. This is because, by (16) and assuming X is bounded, x is bounded and, by assumption on the descent method, $\|\nabla h_{\alpha^k,\beta^k}(x)\|$ must go to zero.
4. This is because, by (16), $\|x - y_{\beta^k}(x)\|$ is bounded, so assuming X is bounded or F is Lipschitz continuous on $\Re^n$, then x is bounded (see, e.g., [31, Lemma 4.1]). By assumption on the descent method, $\|\nabla h^{\delta^k}_{\alpha^k,\beta^k}(x)\|$ must go to zero and, using the strong monotonicity of F, it can be shown (e.g., [31, Theorem 3.3]) that $h^{\delta^k}_{\alpha^k,\beta^k}(x)$ also goes to zero.

References

1. F. Al-Khayyal and J. Kyparisis, “Finite convergence of algorithms for nonlinear programs and variational inequalities,” Journal of Optimization Theory and Applications, vol. 70, pp. 319–332, 1991.
2. G. Auchmuty, “Variational principles for variational inequalities,” Numerical Functional Analysis and Optimization, vol. 10, pp. 863–874, 1989.
3. A. Auslender, Optimisation: Méthodes Numériques, Masson: Paris, 1976.
4. D.P. Bertsekas, Nonlinear Programming, Athena Scientific: Belmont, 1995.
5. J.V. Burke and M.C. Ferris, “Weak sharp minima in mathematical programming,” SIAM Journal on Control and Optimization, vol. 31, pp. 1340–1359, 1993.
6. S.C. Billups and M.C. Ferris, “QPCOMP: A quadratic program based solver for mixed complementarity problems,” Mathematical Programming, vol. 76, pp. 533–562, 1997.
7. M.C. Ferris and J.-S. Pang (Eds.), Complementarity and Variational Problems: State of the Art, SIAM: Philadelphia, 1997.
8. M. Fukushima, “Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems,” Mathematical Programming, vol. 53, pp. 99–110, 1992.
9. 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.
10. H. Jiang, “Unconstrained minimization approaches to nonlinear complementarity problems,” Journal of Global Optimization, vol. 9, pp. 169–181, 1996.
11. C. Kanzow and M. Fukushima, “Theoretical and numerical investigation of the D-gap function for box constrained variational inequalities,” Mathematical Programming, vol. 83, pp. 55–87, 1998.
12. C. Kanzow and M. Fukushima, “Solving box constrained variational inequalities by using the natural residual with D-gap function globalization,” Operations Research Letters, vol. 23, pp. 45–51, 1998.
13. C. Kanzow, N. Yamashita, and M. Fukushima, “New NCP-functions and their properties,” Journal of Optimization Theory and Applications, vol. 94, pp. 115–135, 1997.

14. T. Larsson and M. Patriksson, “A class of gap functions for variational inequalities,” Mathematical Programming, vol. 64, pp. 53–79, 1994.
15. 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, 1997.
16. O.L. Mangasarian and M.V. Solodov, “Nonlinear complementarity as unconstrained and constrained minimization,” Mathematical Programming, vol. 62, pp. 277–297, 1993.
17. O.L. Mangasarian and M.V. Solodov, “A linearly convergent derivative-free descent method for strongly monotone complementarity problems,” Computational Optimization and Applications, vol. 14, pp. 5–16, 1999.
18. P. Marcotte and D.L. Zhu, “Weak sharp solutions of variational inequalities,” SIAM Journal on Optimization, vol. 9, pp. 179–189, 1999.
19. J.-S. Pang, “Error bounds in mathematical programming,” Mathematical Programming, vol. 79, pp. 299–332, 1997.
20. J.-M. Peng, “Equivalence of variational inequality problems to unconstrained minimization,” Mathematical Programming, vol. 78, pp. 347–355, 1997.
21. J.-M. Peng and M. Fukushima, “A hybrid Newton method for solving the variational inequality problem via the D-gap function,” Mathematical Programming, vol. 86, pp. 367–386, 1999.
22. J.-M. Peng, M. Fukushima, and C. Kanzow, “A hybrid Josephy-Newton method for solving box constrained variational inequality problems via the D-gap function,” Optimization Methods and Software, vol. 10, pp. 687–710, 1999.
23. B.T. Polyak, Introduction to Optimization, Optimization Software, Inc., Publications Division: New York, 1987.
24. M.V. Solodov, “Implicit Lagrangian,” in Encyclopedia of Optimization, C. Floudas and P. Pardalos (Eds.), Kluwer Academic Publishers, 2000.
25. M.V. Solodov and B.F. Svaiter, “A hybrid projection–proximal point algorithm,” Journal of Convex Analysis, vol. 6, pp. 59–70, 1999.
26. K. Taji, M. Fukushima, and T. Ibaraki, “A globally convergent Newton method for solving strongly monotone variational inequalities,” Mathematical Programming, vol. 58, pp. 369–383, 1993.
27. 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, 1997, pp. 452–473.
28. J.H. Wu, M. Florian, and P. Marcotte, “A general descent framework for the monotone variational inequality problem,” Mathematical Programming, vol. 61, pp. 281–300, 1993.
29. K. Yamada, N. Yamashita, and M. Fukushima, “A new derivative-free descent method for the nonlinear complementarity problem,” in Nonlinear Optimization and Applications 2, G. Di Pillo and F. Giannessi (Eds.), Kluwer Academic Publishers, pp. 463–487, 2000.
30. N. Yamashita and M. Fukushima, “On stationary points of the implicit Lagrangian for nonlinear complementarity problems,” Journal of Optimization Theory and Applications, vol. 19, pp. 653–663, 1995.
31. N. Yamashita, K. Taji, and M. Fukushima, “Unconstrained optimization reformulations of variational inequality problems,” Journal of Optimization Theory and Applications, vol. 92, pp. 439–456, 1997.
32. H. Xu, “Fukushima’s regularized gap function and Peng’s D-gap function for nonsmooth variational inequalities,” Research Report, School of Information Technology and Mathematical Sciences, University of Ballarat, Ballarat, Australia, 1998.
33. D.L. Zhu and P. Marcotte, “Modified descent methods for solving the monotone variational inequality problem,” Operations Research Letters, vol. 14, pp. 111–120, 1993.