A Class of Gap Functions for Variational Inequalities

by

Torbjörn Larsson
Michael Patriksson

Department of Mathematics
Linköping Institute of Technology
S-581 83 Linköping
Sweden

Abstract

Recently, Auchmuty (1989) has introduced a new class of merit functions, or optimization formulations, for variational inequalities in finite-dimensional space. We develop and generalize Auchmuty's results, and relate his class of merit functions to other works done in this field. Especially, we investigate differentiability and convexity properties, and present characterizations of the set of solutions to variational inequalities. We then present new descent algorithms for variational inequalities within this framework, including approximate solutions of the direction finding and line search problems. The new class of merit functions include the primal and dual gap functions, introduced by Zuhovickiĭ *et al.* (1969a, 1969b), and the differentiable merit function recently presented by Fukushima (1992); also, the descent algorithm proposed by Fukushima is a special case from the class of descent methods developed in this paper. Through a generalization of Auchmuty's class of merit functions we extend those inherent in the works of Dafermos (1983), Cohen (1988) and Wu *et al.* (1991); new algorithmic equivalence results, relating these algorithm classes to each other and to Auchmuty's framework, are also given.

Key Words: Finite-Dimensional Variational Inequalities, Merit Functions, Variational Principles, Successive Approximation Algorithms, Descent Algorithms

Abbreviated title: Gap Functions for Variational Inequalities

1 Introduction

Let X be a nonempty, closed and convex subset of $\Re^n$, and $F : X \mapsto \Re^n$ be a continuous mapping. The *variational inequality problem* is to find an $x^* \in X$ such that

$$\text{[VIP]} \qquad F(x^*)^{\mathrm{T}}(x - x^*) \geq 0, \;\; \forall\, x \in X.$$

Here, T denotes transpose. We will throughout the paper assume that [VIP] admits at least one solution, and denote the solution set of [VIP] by Ω. The assumptions made on X and F are assumed to hold throughout the paper.

Variational inequality problems occur in a number of applications, and were first studied in infinite-dimensional space in connection with nonlinear partial differential equations. More recently, finite-dimensional applications of variational inequality theory, such as non-cooperative games and other equilibrium problems, have arisen.

The solution set is nonempty if, for instance, the feasible set, X, is bounded or if F is *coercive* on X, i.e., if there is a vector $\overline{x} \in X$ such that

$$\lim_{\substack{x \in X \\ \|x\| \to +\infty}} \left\{ F(x)^{\mathrm{T}}(x - \overline{x}) \,/\, \|x\| \right\} = +\infty,$$

where $\|\cdot\|$ is some appropriate vector norm, see, e.g., Kinderlehrer and Stampacchia (1980). Moreover, the set Ω is a singleton if it is nonempty and if F is *strictly monotone* on X, i.e., if

$$[F(x) - F(y)]^{\mathrm{T}}(x - y) > 0, \quad \forall\, x, y \in X \text{ and } x \neq y.$$

The mapping F is said to be *monotone* on X if the strict inequality above is replaced by an inequality, and *strongly monotone* (with modulus m_F) on X if, in addition, the right hand side of the inequality is replaced by $m_F \|x - y\|^2$. (Note that strong monotonicity implies both coercivity and strict monotonicity, so that strong monotonicity implies the existence of a unique solution to [VIP].) If F is the gradient of a function $\phi : \Re^n \mapsto \Re$, then the above properties are equivalent to strict convexity, convexity and strong convexity of ϕ on X, respectively, and, if F is continuously differentiable on an open neighbourhood of X, then the monotonicity properties of F on X may be characterized in terms of the positive definiteness of its Jacobian, ∇F, see e.g. Ortega and Rheinboldt (1970, Section 5.4). For an overview of theory, algorithms and applications of [VIP], we refer to the recent survey by Harker and Pang (1990).

It is well known that the problem [VIP] is equivalent to the first order optimality conditions for the convex problem

$$\text{[P]} \qquad \min_{x \in X} \; \phi(x) \stackrel{\text{def}}{=} \int_0^x F(s)^{\mathrm{T}} ds,$$

if the Jacobian, ∇F, of the mapping F is symmetric and positive semidefinite on an open neighbourhood of X. If the mapping F has a Jacobian which is not symmetric, the integral in [P] is not unambiguously defined, and [VIP] can in this case not be converted into an equivalent minimization problem of the form [P]. (In this case, we say that [VIP] is *asymmetric*.) It is the purpose of this paper to review and develop reformulations of asymmetric variational inequalities as equivalent minimization problems, investigate their properties, and derive descent methods for the solution of [VIP], utilizing a class of such optimization formulations. The basis for our development is the article by Auchmuty (1989).

We shall make use of the following conventions. All vectors are assumed to be column vectors, $\|\cdot\|$ denotes the Euclidean vector norm in $\Re^n$, and its induced matrix and operator norms. For a given function $h : X \mapsto \Re^l$ the Lipschitz constant is denoted by M_h. The constant in the definition of strong monotonicity will be denoted by m_h, and the same notation will be used for the constant in the definition of strong convexity of h (when $l = 1$). We shall make use of the abbreviations u.s.c. and l.s.c. for upper and lower semicontinuity, respectively, and write $h \in C^p$ on X meaning that h is p times continuously differentiable on an open neighbourhood of X. (We write $h \in C^p$ if this holds on $\Re^n$.)

The remainder of the paper is organized as follows. In Section 2 we present a class of merit functions for asymmetric variational inequalities, originally proposed by Auchmuty (1989). We summarize some of the most important results of that paper, and provide an interpretation and a motivation of the use of these merit functions. In Section 3 we review and relate to Auchmuty's class of merit functions some earlier developed merit functions, with the common feature that they can be described as *gap functions*. In particular, we show that the primal and dual gap functions, which were introduced by Zuhovickiĭ *et al.* (1969a, 1969b), and the recently presented differentiable merit function of Fukushima (1992), belong to this class. In Section 4 we further investigate properties of Auchmuty's merit functions, provide equivalent characterizations of the solution set, and establish descent algorithms, constructed in a manner similar to Fukushima's. In Section 5 we present a generalization of his merit functions, and give some algorithmic equivalence results, relating our work to that of Dafermos (1983), Cohen (1988) and Wu *et al.* (1991). Finally, in Section 6 we make some concluding remarks and present research opportunities.

2 Background and motivation

The merit functions discussed in this paper are all of the type given by the following definition.

Definition 2.1 (Gap function) *A function $\gamma : X \mapsto \Re \cup \{-\infty, +\infty\}$ is a gap function for* [VIP] *if*

1. *γ is restricted in sign on X, and*
2. *$\gamma(x) = 0 \Leftrightarrow x \in \Omega$.*

A gap function γ provides a measure of the violation of [VIP] at any point $x \in X$, and by minimizing γ over X (assuming $\gamma(x) \geq 0$ on X) a point in Ω is obtained. Therefore, gap functions may be used as merit functions for variational inequalities.

We now summarize the results of Auchmuty (1989). Auchmuty develops merit functions for [VIP], based on a function $\tilde{L} : X \times X \mapsto \Re$, with

$$\tilde{L}(x, y) = f(x) - f(y) + [F(x) - \nabla f(x)]^{\mathrm{T}} (x - y), \tag{2.1}$$

where $f : \Re^n \mapsto \Re \cup \{+\infty\}$ is convex and l.s.c., and $f \in C^1$ on X.

The following theorem relates saddle points to $\tilde{L}$ to the set of solutions to [VIP]. [For proofs of the theorems stated in this section, we refer to Auchmuty (1989).]

Theorem 2.1 (Saddle points are in Ω) *If $(x^*, y^*) \in X \times X$ satisfies*

$$\tilde{L}(x^*, y) \leq \tilde{L}(x^*, y^*) \leq \tilde{L}(x, y^*)$$

for all $(x, y) \in X \times X$, then x^* solves [VIP].

Remark 2.1 The point y^* does not, in general, solve [VIP] and, since $\tilde{L}(x,y)$ is not convex in the x argument, it is, in general, not a true saddle function.

The saddle point characterization of a solution to [VIP] defines two possible optimization formulations for the variational inequality problem by the introduction of two different gap functions. The first, which we will refer to as the *primal generalized gap function*, for reasons to be evident later, is defined as

$$\text{[PGG]} \qquad \tilde{G}(x) \stackrel{\text{def}}{=} \sup_{y \in X} \tilde{L}(x,y), \quad x \in X,$$

and the corresponding optimization formulation is

$$\text{[OPGG]} \qquad \inf_{x \in X} \tilde{G}(x).$$

Note that [PGG] is a convex program, since $\tilde{L}$ is concave in y, and that $\tilde{L}$ is invariant under the addition of an affine function to f. For the primal generalized gap function $\tilde{G} : X \mapsto \Re \cup \{+\infty\}$, Auchmuty's results are summarized in the following theorem.

Theorem 2.2 (Properties of $\tilde{G}$)

1. *$\tilde{G}(x)$ is l.s.c. and nonnegative for all $x \in X$, and*
2. *$\tilde{G}(x) = 0 \Leftrightarrow x \in \Omega$.*

In particular, when X is bounded, $\tilde{G}$ is finite on X.

Remark 2.2 Finiteness of $\tilde{G}$ may be accomplished also for unbounded sets for instance by imposing strong convexity of f. Since a minimum of $\tilde{G}$ over X always exists whenever $\tilde{G}$ is finite, [OPGG] can be used also to detect the non-existence of a solution to [VIP], since, in this case, $\min_{x \in X} \tilde{G}(x) > 0$; the properties of such minimizers are subjects for further research.

Hence, solving [VIP] is equivalent to minimizing $\tilde{G}$ over X, so that $\tilde{G}$ is a merit function for variational inequalities.

Turning to the dual formulation, we introduce the function $\tilde{g} : X \mapsto \Re \cup \{-\infty\}$ through

[DGG] $$\tilde{g}(y) \stackrel{\text{def}}{=} \inf_{x \in X} \tilde{L}(x, y), \quad y \in X,$$

and the resulting optimization formulation is

[ODGG] $$\sup_{y \in X} \tilde{g}(y).$$

For the function $\tilde{g}$ we have the following results, corresponding to Theorem 2.2.

Theorem 2.3 (Properties of $\tilde{g}$)

1. *$\tilde{g}(y)$ is nonpositive for all $y \in X$, and*
2. *$\tilde{g}(y) = 0 \Rightarrow y \in \Omega$.*

In general, the converse of 2. does not hold, as the example below shows. A consequence of this is that $\tilde{g}$ is *not* a gap function. (However, for pseudo-monotone maps F and with $f \equiv 0$, the converse of 2. does hold, see Section 3.)

Example 2.1 Let $F(x) = x$ and $X = \{x \mid 0 \leq x \leq 1\}$. The problem is then equivalent to minimizing $\frac{1}{2}x^2$ over X, with the unique solution $x^* = 0$ [$\Omega = \{0\}$]. Choose $f(x) = x^4$, which is strictly convex. Then $\tilde{g}(0) = -2 \neq 0$. □

We end this section by motivating the use of $\tilde{G}$ as a merit function, by providing an interpretation of it in terms of cost approximations of [VIP].

Since [PGG] is a convex program, evaluating $\tilde{G}(x)$ is equivalent to finding a solution $y(x)$ to the following variational inequality, defining the first order optimality conditions of [PGG] (e.g., Bazaraa and Shetty, 1979, Theorem 3.4.3).

$$\nabla_y \tilde{L}\left(x, y(x)\right)^{\mathrm{T}} \left(y - y(x)\right) \leq 0, \ \forall \ y \in X \tag{2.2}$$

Here, ∇_y denotes differentiation with respect to y.

Now, assume that, in [VIP], the mapping F is approximated by the gradient of a convex function f. Obviously, the error made when replacing the original map is $F - \nabla f$. If, for some fixed $x \in X$, this error is taken into account by adding to ∇f the error term $F(x) - \nabla f(x)$, then the approximate *symmetric* variational inequality is to find a $y(x) \in X$ such that

[AVIP] $$\left[\nabla f\left(y(x)\right) + F(x) - \nabla f(x)\right]^{\mathrm{T}} \left(y - y(x)\right) \geq 0, \ \forall \ y \in X.$$

But from (2.1) it then follows that (2.2) and [AVIP] are equivalent. We may thus conclude that the calculation of $\tilde{G}(x)$ amounts to solving an approximate, symmetric, variational inequality obtained through a cost approximation. [This interpretation was first made by Patriksson (1991b).] Obviously, the better the gradient ∇f approximates the, possibly asymmetric, mapping F, the closer $y(x)$ will be to Ω. This is further illustrated below.

Assume now that ∇F is symmetric on X. We may then choose the function f such that $\nabla f = F$, and [AVIP] reduces to [VIP]. [Also, $\tilde{G}(x) = f(x) + c$, where $c = -\inf_{x \in X} f(x)$, so that minimizing $\tilde{G}$ is equivalent to minimizing f over X, see Auchmuty (1989).] This possibility of choosing ∇f so as to approximate F exactly for symmetric variational inequalities suggests a symmetrization strategy for the asymmetric case. The non-existence of the optimization problem [P] is due to the fact that the value of the line integral in [P] is dependent of the curve, Γ, taken from 0 to $x \in X$. To define an unambiguous integral, we may choose the curve Γ to be the straight line from 0 to x, and hence introduce a parametrization $s = s(t) = tx$, with $t \in [0, 1]$. We then define

$$f(x) \stackrel{\text{def}}{=} \int_{\Gamma} F(s)^{\mathrm{T}} ds = \int_0^1 \Psi(t) dt, \tag{2.3}$$

where $\Psi : \Re \mapsto \Re$ is given by $\Psi(t) = F(tx)^{\mathrm{T}} x$. This approximation is exact ($f \equiv \phi$ in [P]) if F is the gradient of f, and should be a good approximation for modest asymmetries of ∇F. The following example shows that a well known symmetrization of affine variational inequalities is obtained from the application of (2.3).

Example 2.2 Consider the asymmetric and affine variational inequality, i.e., $F(x) = Ax - b$, where A is an asymmetric matrix in $\Re^{n \times n}$. By the choice of f as in (2.3), we obtain that

$$f(x) = \int_0^1 (tAx - b)^{\mathrm{T}} x dt = \frac{1}{2} x^{\mathrm{T}} Ax - b^{\mathrm{T}} x,$$

which indeed defines a symmetrization of ∇F, since $\nabla^2 f = \frac{1}{2}\left(A + A^{\mathrm{T}}\right)$ is the symmetric part of $\nabla F = A$. □

In the next section we relate the results obtained by Auchmuty to some earlier results on merit functions for [VIP].

3 Gap functions for variational inequalities

3.1 The primal and dual gap functions

We begin by studying the important special case of Auchmuty's merit functions, resulting from letting $f \equiv 0$ in (2.1). (Auchmuty treats this case separately, but studies into the resulting merit functions in fact traces back to the late 1960's.) The corresponding saddle function to (2.1) then is

$$L(x, y) \stackrel{\text{def}}{=} F(x)^{\mathrm{T}}(x - y). \tag{3.1}$$

In the context of non-cooperative game theory this function was studied already by Zuhovickiĭ *et al.* (1969a, 1969b, 1970, 1973); they showed that, under a monotonicity assumption on F, a solution to the equilibrium problem is found by searching for a saddle point to L, i.e., a point (x^*, y^*) such that

[SP] $$\sup_{y \in X} \inf_{x \in X} F(x)^{\mathrm{T}}(x - y) = F(x^*)^{\mathrm{T}}(x^* - y^*) = \inf_{x \in X} \sup_{y \in X} F(x)^{\mathrm{T}}(x - y).$$

Exploring the rightmost equality of [SP] we obtain the *primal gap function* $G : X \mapsto \Re \cup \{+\infty\}$ defined by

[PG] $$G(x) \stackrel{\text{def}}{=} \sup_{y \in X} F(x)^{\mathrm{T}}(x - y),$$

which has been extensively studied in various contexts. Below we summarize its most important properties.

Theorem 3.1 (Properties of G) *For any $x \in X$, let $Y(x)$ denote the (possibly empty) set of optimal solutions to the problem* [PG].

1. *G is a gap function.*
2. *G is l.s.c. on X.*
3. *If X is bounded and $F \in C^1$ on X, then G is Lipschitz continuous on X, i.e., there exists a constant $M_G \geq 0$ such that, for all $x^1, x^2 \in X$,*
$$\left\| G\left(x^1\right) - G\left(x^2\right) \right\| \leq M_G \left\| x^1 - x^2 \right\|.$$
4. *If $F \in C^1$ on X, then G is differentiable at $x \in X$ if $Y(x) = \{y(x)\}$, with $\nabla G(x) = F(x) + \nabla F(x)^{\mathrm{T}}(x - y(x))$.*

5. *If $F \in C^1$ on X, and monotone on X, then if $x \notin \Omega$ and $Y(x) = \{y(x)\}$, the direction $d = y(x) - x$ is a feasible direction of descent with respect to G at x, and the directional derivative $G'(x; d) = \nabla G(x)^{\mathrm{T}} d \leq -G(x)$.*

6. *G is convex on X if $F(x)^{\mathrm{T}} x$ is convex on X and each component of F is concave on X.*

7. (A fixed point characterization of Ω) $\quad x \in \Omega \Leftrightarrow x \in Y(x)$.

8. (A stationary point characterization of Ω) *Under the conditions on F in 5., $x \in \Omega \Leftrightarrow G'(x; y - x) \geq 0, \forall\, y \in X$.*

Proof

1. See Zuhovickiĭ *et al.* (1969a, 1969b, 1970, 1973). See also Demyanov and Rubinov (1970, Section 3.1) and Hearn (1978, 1982).

2. See Dunn (1980).

3. See Marcotte (1985).

4. See Demyanov and Pevnyi (1972). See also Zuhovickiĭ *et al.* (1973), Auslender (1976, Section VII.3), Hearn (1978, 1982) and Hearn *et al.* (1984).

5. See Demyanov and Pevnyi (1972). See also Hearn (1982) for the first result, and Marcotte (1985).

6. See Hearn (1978, 1982) and Hearn *et al.* (1984).

7. See Marcotte (1985). See also Evtushenko (1985, p. 34).

8. See Marcotte (1985). □

Thus, if [VIP] has a solution, the set of solutions to

[OPG] $$\inf_{x \in X} G(x)$$

equals the set Ω. Note that when X is polyhedral, [PG] reduces to a linear program which, in the symmetric case, is equivalent to that solved in the Frank-Wolfe algorithm, see e.g. Hearn (1978, 1982). The function G has been used as a merit function in several algorithms for the solution of [VIP], see e.g. Zuhovickiĭ *et al.* (1970, 1973), Demyanov and Pevnyi (1972), Primak (1975), Lawphongpanich and Hearn (1984), Marcotte (1985, 1986), Marcotte and Dussault (1985, 1987, 1989), and Marcotte and Guélat (1988).

In the next section we will extend the results given in Theorem 3.1 to the function $\tilde{G}$.

Turning to the leftmost equality of [SP] we obtain the *dual gap function* $g : X \mapsto \Re \cup \{-\infty\}$ defined by

[DG] $$g(y) \stackrel{\text{def}}{=} \inf_{x \in X} F(x)^{\mathrm{T}}(x-y).$$

As established earlier, in the general case ($f \not\equiv 0$) the dual formulation does *not* constitute a proper gap function. For the case $f \equiv 0$, however, it is possible to show that g is a gap function, provided that the mapping F is *pseudo-monotone* on X, i.e.,

$$F(y)^{\mathrm{T}}(x-y) \geq 0 \Rightarrow F(x)^{\mathrm{T}}(x-y) \geq 0, \forall\, x, y \in X. \tag{3.2}$$

(This property is, for instance, implied by monotonicity.) Under this condition [VIP] is equivalent to finding an $x^* \in X$ such that

[VIP'] $$F(x)^{\mathrm{T}}(x-x^*) \geq 0,\ \forall\, x \in X,$$

see e.g. Karamardian (1976) and Auslender (1976, Section VII.1). We summarize some properties of g below.

Theorem 3.2 (Properties of g) *Let $F : X \mapsto \Re^n$ be pseudo-monotone on X. Also, for any $y \in X$, let $X(y)$ denote the (possibly empty) set of optimal solutions to the problem* [DG].

1. *g is a gap function.*
2. *g is concave on X.*
3. *If $F \in C^1$ on X, then g is differentiable at $y \in X$ if $X(y) = \{x(y)\}$, with $\nabla g(y) = -\nabla F(x(y))$.*

Proof

1. See Zuhovickiĭ *et al.* (1969a, 1969b). See also Demyanov and Pevnyi (1972) and Auslender (1976, Section VII.5).
2. See Hearn and Nguyen (1982), Hearn *et al.* (1984) and Nguyen and Dupuis (1984).
3. See Hearn and Nguyen (1982) and Hearn *et al.* (1984). □

As a consequence of the gap function properties of g, the set of solutions to

[ODG] $$\sup_{y \in X} g(y)$$

equals the set Ω. Methods developed for [ODG] mostly apply constraint generation techniques to the semi-infinite linear program

[SIP]
$$\begin{aligned} \max \quad & z \\ \text{s.t.} \quad & z \leq F(x)^{\mathrm{T}}(x-y), \quad \forall\, x \in X \\ & y \in X. \end{aligned}$$

A line of development of such methods can be traced in Zuhovickiĭ *et al.* (1969a, 1969b, 1973), Demyanov and Pevnyi (1972), Auslender (1976, Section VII.5), Nguyen and Dupuis (1984) and Hearn and Lawphongpanich (1989).

Summarizing the properties of the saddle function L, although it is not convex in the x argument, we have, for pseudo-monotone F,

$$\begin{array}{rcccll} g(y) & \leq & 0 & \leq & G(x) & \forall\,(x,y)\in X\times X, \quad \text{and} \\ \sup_{y\in X} g(y) & = & 0 & = & \inf_{x\in X} G(x), & \end{array}$$

and (x^*, y^*) is a saddle point to L if and only if $x^*, y^* \in \Omega$.

One may note the nice symmetry in the properties of G and g. While the calculation of $G(x)$ is a convex problem, the minimization of G is in general both a nonconvex and nondifferentiable problem; on the other hand, while the evaluation of $g(y)$ is, in general, a nonconvex problem, the maximization of g is always a convex problem. Also, convexity of G holds for affine and monotone maps F, and under the same conditions the calculation of $g(y)$ is a convex problem.

Finally, it can be mentioned that subgradient optimization methods have been devised for the reformulations [OPG] and [ODG], see e.g. Hearn (1978, 1982) and Hammond and Magnanti (1987).

3.2 Smith's class of gap functions

In the context of traffic equilibrium, Smith (1983a, 1983b) has developed a family of gap-type merit functions for variational inequalities on polyhedral sets. If we define

$$[x]_+ = \max\{0, x\},$$

the family of objectives can be written as:

$$G^p(x) = \sum_{i\in\mathcal{E}} \left(\left[F(x)^{\mathrm{T}}\left(x - \hat{y}^i\right)\right]_+\right)^p, \; x \in X,$$

where $p > 0$ and $\{\hat{y}^i : i \in \mathcal{E}\}$ is the set of extreme points of the bounded polyhedron X. The functions have the following properties, similar to the ones for $\tilde{G}$, G and g: for all $p \geq 1$, $G^p(x) > 0$, $\forall\, x \notin \Omega$ and $G^p(x^*) = 0 \Leftrightarrow x^* \in \Omega$. Further studies of the properties of $G^p(x)$ are made by Hearn *et al.* (1984). They show that the primal gap function G is a limiting case of this class of functions, since, for all $x \in X$,

$$\begin{array}{rcl} \lim_{p\to+\infty} \left(G^p(x)\right)^{1/p} & = & \max_{i\in\mathcal{E}} F(x)^{\mathrm{T}}\left(x - \hat{y}^i\right) \\ & = & G(x). \end{array}$$

They proceed to show that $G^p(x)$ is differentiable for $2 \leq p < +\infty$, and convex for $p \geq 2$ and $p = +\infty$ e.g. if F is affine and monotone. Note that the extreme points must be known explicitly in order to evaluate $G^p(x)$, for $p < +\infty$. The practical use of these gap functions is thus limited to polyhedral sets, in combination with simplicial decomposition/column generation techniques, see Smith (1983a, 1983b).

3.3 Fukushima's special case of gap function

Fukushima (1992) introduces a gap function, which may be identified as a special case from Auchmuty's class, by using $f(x) = \frac{1}{2}\|x\|_Q^2 \stackrel{\text{def}}{=} \frac{1}{2}x^{\mathrm{T}}Qx$, where Q is a given symmetric and positive definite matrix in $\Re^{n\times n}$. The problem [PGG] is shown to reduce to a projection problem, i.e., to find the unique point $y(x)$ minimizing $\|y-(x-Q^{-1}F(x))\|_Q$ over $y \in X$. [This problem is equivalent to that solved iteratively in the *projection algorithm*, see e.g. Dafermos (1983).] Introducing the notation $\tilde{G}_{\mathrm{Fu}}$ for Fukushima's special case of gap function, we have the following results. [For proofs of the results in this section, we refer to Fukushima (1992).]

Theorem 3.3 (Properties of $\tilde{G}_{\mathrm{Fu}}$)

1. *$\tilde{G}_{\mathrm{Fu}}$ is a gap function.*
2. *(A fixed point characterization of Ω)* $x \in \Omega \Leftrightarrow x = y(x)$.
3. *If $F : \Re^n \mapsto \Re^n$ is continuous, then $\tilde{G}_{\mathrm{Fu}}$ is also continuous.*
4. *If $F : \Re^n \mapsto \Re^n$ is in C^1 then $\tilde{G}_{\mathrm{Fu}}$ is also in C^1.*
5. *(Stationary points are in Ω) If $F \in C^1$, with a positive definite Jacobian ∇F on X, we have that*
$$\nabla\tilde{G}_{\mathrm{Fu}}(x)^{\mathrm{T}}(y-x) \geq 0, \forall\ y \in X \Rightarrow x \in \Omega. \tag{3.3}$$

Fukushima also devises descent methods for asymmetric variational inequalities with positive definite Jacobian. Global convergence is guaranteed from the stationary point property (3.3), and the fact that the direction $y(x) - x$ is a feasible direction of descent with respect to $\tilde{G}_{\mathrm{Fu}}$ whenever $x \notin \Omega$. We summarize these results below.

Theorem 3.4 (Descent properties) *Let $F : \Re^n \mapsto \Re^n$ be in C^1, with positive definite Jacobian on X.*

1. *If $x \notin \Omega$, then*
$$\nabla\tilde{G}_{\mathrm{Fu}}(x)^{\mathrm{T}}(y(x)-x) < 0.$$
2. *If F also is strongly monotone on X, then for all $x \in X$,*
$$\nabla\tilde{G}_{\mathrm{Fu}}(x)^{\mathrm{T}}(y(x)-x) \leq -m_F\,\|y(x)-x\|^2.$$

Theorem 3.5 (Global convergence) *Let X be bounded, and let $F : \Re^n \mapsto \Re^n$ be in C^1 with positive definite Jacobian on X.*

1. *Assume that* x^0 *is chosen arbitrary in* X*, and that* x^{k+1} *is determined from* x^k *through the exact line search rule*

$$\min_{0 \le t \le 1} \tilde{G}_{\mathrm{Fu}}\left(x^k + t\left(y\left(x^k\right) - x^k\right)\right). \tag{3.4}$$

   Then, $\lim_{k \to +\infty} x^k = x^*$*, i.e., the unique solution to* [VIP].

2. *Assume, in addition, that both* F *and its Jacobian are Lipschitz continuous on* X*, and that* F *is strongly monotone on* X*.*

   Replace the exact line search rule (3.4) by the Armijo (1964) step length rule, where the step length t_k *in iteration* k *is chosen as* $t_k = \beta^{l_k}$*, where* l_k *is the smallest nonnegative integer* l *such that*

$$\tilde{G}_{\mathrm{Fu}}\left(x^k + \beta^l\left(y\left(x^k\right) - x^k\right)\right) \le \tilde{G}_{\mathrm{Fu}}\left(x^k\right) - \alpha\beta^l\left\|y\left(x^k\right) - x^k\right\|^2, \tag{3.5}$$

   with $\alpha > 0$ *and* $0 < \beta < 1$*.*

   Then the same conclusions as above hold, provided that $\alpha < m_F$*.*

Fukushima's special case of gap function has also been used in a descent algorithm based on Newton's method, see Taji *et al.* (1991). This algorithm extends the globally convergent Newton methods by Marcotte and Dussault (1985, 1987, 1989), based on the primal gap function, G.

Remark 3.1 The authors became aware of Fukushima's work after the original manuscript had been submitted. Some extensions (Theorems 4.4 and 4.9) of his results have been added in this revised version. We would like to point out that, from the interpretation and motivation made in Section 2, a gap function based on a symmetric approximation ∇f of F seems to be a more natural choice than a gap function based on a fixed matrix Q.

4 The primal generalized gap function

Based on the discussions made in the previous sections, we will here further investigate the characteristics of the primal generalized gap function and the mathematical program [OPGG]. We will in particular provide characterizations of Ω in terms of fixed points and stationary points, give sufficient conditions under which $\tilde{G}$ is continuously differentiable and convex, and establish globally convergent descent algorithms.

4.1 The properties of the generalized gap function

The following theorem provides a new fixed point characterization of solutions to variational inequalities, and extends the results of Theorems 3.1.7 and 3.3.2.

Theorem 4.1 (A fixed point characterization of Ω) *For any $x \in X$, let $Y(x)$ denote the set of optimal solutions to* [PGG]*. Then,*

$$x \in \Omega \Leftrightarrow x \in Y(x).$$

Proof

$\Rightarrow$) If $x \in \Omega$, then $\tilde{G}(x) = 0$. From $\tilde{L}(x, x) = 0$, the result then follows immediately.

$\Leftarrow$) If $x \in Y(x)$, then from [AVIP] it follows that

$$[-\nabla f(x) - F(x) + \nabla f(x)]^{\mathrm{T}} (y - x) = -F(x)^{\mathrm{T}}(y - x) \leq 0, \forall\, y \in X,$$

which implies that $x \in \Omega$. $\square$

To simplify the subsequent discussion we shall henceforth let f be continuous on X. An immediate consequence of this property is that $\tilde{G}$ then also is continuous whenever it is everywhere finite; continuity of $\tilde{G}$ is important when developing descent algorithms, since the algorithmic map associated with the line search with respect to $\tilde{G}$ then is closed in the sense of Zangwill (1969).

Theorem 4.2 (Continuity properties of $\tilde{G}$ and Y)

1. *For any $x \in X$, the set $Y(x)$ is closed and convex, the mapping $x \mapsto Y(x)$ is closed and $\tilde{G}$ is l.s.c.*
2. *If X is bounded or if f is strongly convex on X, then for any $x \in X$, $Y(x)$ is nonempty and bounded, and $\tilde{G}$ is continuous on X.*
3. *If X is bounded and $F \in C^1$ on X, and $f \in C^2$ on X, then $\tilde{G}$ is Lipschitz continuous on X.*
4. *If X is bounded and f is strictly convex on X, or if f is strongly convex on X, then, for any $x \in X$, $Y(x)$ is a singleton and $x \mapsto Y(x)$ is continuous.*

Proof

1. The closedness of $Y(x)$ follows from the continuity of $\tilde{L}$ (Hogan, 1973), and its convexity from the convexity of the problem [PGG] (e.g. Minty, 1962). The l.s.c. property of $\tilde{G}$ is given in Theorem 2.2.
2. The first two results are well known. Continuity of $\tilde{G}$ follows from the continuity and boundedness of $\tilde{L}$.

3. The assumptions imply the continuity and boundedness of $\nabla_x \tilde{L}(x,y)$, which in turn implies the Lipschitz continuity of $\tilde{L}$ (Ortega and Rheinboldt, 1970, Corollary 3.2.4). By Theorem 2.1 of Clarke (1975) $\tilde{G}$ is then also Lipschitz continuous.

4. The first result is obvious. For the second result, see e.g. Hogan (1973). □

Under appropriate assumptions on F and f, it is also possible to show that the mapping $x \mapsto Y(x)$ is Lipschitz continuous, and this is therefore also true for $\nabla \tilde{G}$. This property is crucial when establishing convergence using inexact line searches, see Lemma 4.1. The next theorem gives sufficient conditions for $\tilde{G}$ to be continuously differentiable on X.

Theorem 4.3 (Differentiability of $\tilde{G}$) *Let $F \in C^1$ on X and $f \in C^2$ on X. If X is bounded and f is strictly convex on X, or if f is strongly convex on X, then $\tilde{G} \in C^1$ on X, with*

$$\nabla \tilde{G}(x) = F(x) + \left[\nabla F(x)^{\mathrm{T}} - \nabla^2 f(x)\right](x - y(x)), \tag{4.1}$$

where $y(x)$ is the unique solution to [PGG].

Proof $\tilde{G}$ is differentiable since $Y(x)$ is a singleton from Theorem 4.2.4 [see also Rockafellar (1981, Proposition 3B)]. The continuity of $\nabla \tilde{G}$ then follows from the continuity of $\nabla_x \tilde{L}$, and $\nabla \tilde{G}(x)$ is given by $\nabla_x \tilde{L}(x, y(x))$ [cf. Rockafellar (1981, Proposition 3H)]. □

It is therefore possible to equivalently reformulate asymmetric variational inequalities, with a differentiable mapping F, as a continuously differentiable minimization problem, see Harker and Pang (1990, p. 165).

One may note the major difference between the differentiability properties of $\tilde{G}$ compared to the ones for G. (Differentiability of G holds in general only under the restrictive assumption that the set X is *strongly convex* in the sense of Zuhovickiĭ *et al.* (1969a), which excludes applications where X is polyhedral.)

The next theorem characterizes solutions to [VIP] as stationary points of the problem [OPGG], for different monotonicity properties of F, and, correspondingly, different convexity properties of f.

Theorem 4.4 (Stationary point characterization of Ω) *Let $F \in C^1$ on X, and $f \in C^2$ on X. Under either one of the following three additional assumptions, x^* solves* [VIP] *if and only if x^* is a stationary point of the problem* [OPGG], *i.e., if and only if*

$$\tilde{G}'(x^*; x - x^*) \geq 0, \forall\, x \in X. \tag{4.2}$$

1. *X is bounded, and the matrix $\nabla F(x^*)^{\mathrm{T}} - \nabla^2 f(x^*)$ is positive definite.*

2. *X is bounded, f is strictly convex on X, and the matrix $\nabla F(x^*)^{\mathrm{T}} - \nabla^2 f(x^*)$ is positive semidefinite.*

3. *F is strongly monotone on X, f is strongly convex on X, ∇f is Lipschitz continuous on X, and*
$$M_{\nabla f} - m_f < m_F. \tag{4.3}$$

Proof Assume first that (4.2) does not hold for some $\overline{x} \in X$. For a sufficiently small $\varepsilon > 0$ it must then hold that $\tilde{G}(x^* + \varepsilon(\overline{x} - x^*)) < \tilde{G}(x^*) = 0$, which contradicts the optimality of x^*. We have thus shown that a solution x^* to [VIP] must satisfy (4.2).

We next turn to prove the reverse, for the three sets of assumptions, respectively. We first note that the directional derivative $\tilde{G}'(x^*; x - x^*)$ is given by [e.g., Rockafellar (1981, Proposition 3B)]

$$\tilde{G}'(x^*; x - x^*) = \sup_{y \in Y(x^*)} \left(F(x^*) + \left[\nabla F(x^*)^{\mathrm{T}} - \nabla^2 f(x^*)\right](x^* - y)\right)^{\mathrm{T}} (x - x^*) \geq 0, \forall\, x \in X. \tag{4.4}$$

The set $Y(x^*)$ is convex and compact, cf. Theorem 4.2.1 and Theorem 4.2.2. The restriction of (4.4) to $x \in Y(x^*)$ defines a point-to-set map from $Y(x^*)$ into itself, i.e., a map of the form $x \mapsto S(x)$, where $S(x)$ is the set of optimal solutions to the problem defined in (4.4). The set $S(x)$ inherits the properties of $Y(x^*)$, and furthermore, the map $x \mapsto S(x)$ is u.s.c., since the inner problem is linear in x. Using Kakutani's (1941) fixed point theorem, we conclude that there is a $\overline{y} \in Y(x^*)$ such that $\overline{y} \in S(\overline{y})$. From (4.4) we then have that

$$\left(F(x^*) + \left[\nabla F(x^*)^{\mathrm{T}} - \nabla^2 f(x^*)\right](x^* - \overline{y})\right)^{\mathrm{T}} (\overline{y} - x^*) \geq 0. \tag{4.5}$$

(Note that for a strictly convex f this holds directly from $\nabla \tilde{G}(x^*)^{\mathrm{T}} (y(x^*) - x^*) \geq 0$, by replacing $\overline{y}$ with $y(x^*)$, since, in this case, $Y(x^*)$ is a singleton.)

Since $\overline{y} \in Y(x^*)$, we also have from (2.2) that

$$[-\nabla f(\overline{y}) - F(x^*) + \nabla f(x^*)]^{\mathrm{T}} (x - \overline{y}) \leq 0, \forall\, x \in X. \tag{4.6}$$

Combining (4.5) and (4.6), with $x = x^*$, we then have that

$$(\overline{y} - x^*)^{\mathrm{T}} \left[\nabla F(x^*)^{\mathrm{T}} - \nabla^2 f(x^*)\right] (\overline{y} - x^*) \leq [\nabla f(\overline{y}) - \nabla f(x^*)]^{\mathrm{T}} (x^* - \overline{y}). \tag{4.7}$$

We now turn to the three sets of assumptions.

1. From the positive definiteness of $\nabla F(x^*)^{\mathrm{T}} - \nabla^2 f(x^*)$ and the convexity of f, the inequality (4.7) can only hold if $\overline{y} = x^*$. Replacing $\overline{y}$ by x^* in (4.6), we obtain that x^* solves [VIP].

2. Follows from the same line of arguments as in 1.

3. Using the monotonicity and Lipschitz continuity properties of the maps F and ∇f in (4.7), it follows that

$$(m_F - M_{\nabla f} + m_f)\,\|\overline{y} - x^*\|^2 \leq 0,$$

which, together with (4.3), yields $\overline{y} = x^*$, and as in 1., we conclude that x^* solves [VIP]. □

It is a direct consequence of this theorem, that if a feasible point x is not a solution to [VIP], then there is a $y \in Y(x)$ such that the direction $y - x$ is a feasible direction of descent with respect to $\tilde{G}$. Based on this observation we shall, later on, develop descent algorithms.

Since the descent methods to be presented here are all asymptotically convergent, an optimal solution is, in general, not reached finitely, and it is of great practical interest to obtain a measure of the distance from an approximate solution point to the set Ω. The following theorem establishes an a posteriori error bound for strongly monotone variational inequalities, measured in terms of the value of the function $\tilde{G}$.

Theorem 4.5 (An a posteriori error bound) *Let F be strongly monotone on X, and $f \in C^2$ on X, with a Lipschitz continuous gradient on X. If f is chosen such that $M_{\nabla f} < 2m_F$, then for any $x \in X$,*

$$\|x - x^*\|^2 \leq \frac{2\tilde{G}(x)}{2m_F - M_{\nabla f}}. \tag{4.8}$$

Proof Combining the inequalities obtained from the strong monotonicity of F, and the fact that x^* is the solution of [VIP], we obtain for any $x \in X$ that

$$F(x)^{\mathrm{T}}(x - x^*) \geq m_F \|x - x^*\|^2. \tag{4.9}$$

Now,

$$\begin{aligned}
\tilde{G}(x) &= f(x) - f(y(x)) + [F(x) - \nabla f(x)]^{\mathrm{T}}(x - y(x)) \\
&\geq f(x) - f(x^*) + [F(x) - \nabla f(x)]^{\mathrm{T}}(x - x^*) && (y(x) \text{ solves [PGG]}) \\
&\geq f(x) - f(x^*) - \nabla f(x)^{\mathrm{T}}(x - x^*) + m_F\|x - x^*\|^2 && (\text{Equation (4.9)}) \\
&\geq \left(m_F - \frac{M_{\nabla f}}{2}\right)\|x - x^*\|^2, && (\nabla f \text{ Lipschitz})
\end{aligned}$$

and the theorem is proved. □

The error bound (4.8) generalizes those obtained by Pang (1987) for the choices of $f(x) = \frac{1}{2}\|x\|^2$ (projection measure, Theorem 3.1) and $f \equiv 0$ (linear programming measure, Theorem 4.1), and the result obtained by Taji *et al.* (1991) for the choice of $f(x) = \frac{1}{2}\|x\|_Q^2$. Notice that neither Lipschitz continuity [cf. Pang (1987)], nor differentiability, of F is necessary, and that the result holds regardless of any convexity properties of the function f.

The following corollary gives some interesting properties of $\tilde{G}$.

Corollary 4.1 (Coercivity of $\tilde{G}$) *Under the assumptions of the theorem, $\tilde{G}$ has bounded level sets and is weakly coercive on X, i.e.,*

$$\lim_{\substack{x \in X \\ \|x\| \to +\infty}} \tilde{G}(x) = +\infty. \tag{4.10}$$

Proof Follows directly from (4.8). □

Remark 4.1 Although $\tilde{G}$ is, in general, not convex, the corollary establishes properties of $\tilde{G}$ shared with convex functions. (Weak coercivity and boundedness of level sets are well known properties of *strongly* convex functions.)

Below we show that, for affine variational inequalities, $\tilde{G}$ can be constructed strictly convex.

Theorem 4.6 (Convexity of $\tilde{G}$) *Let F be affine, i.e., $F(x) = Ax - b$. Let f be given by $f(x) = \frac{1}{2}x^{\mathrm{T}}Qx$, where Q is symmetric and positive definite. Then $\tilde{G}$ is convex on X if the matrix $A + A^{\mathrm{T}} - Q$ is positive semidefinite, and strictly convex if this matrix is positive definite.*

Proof If, for any fixed $y \in X$, the function $\tilde{L}$ is (strictly) convex in its x argument on the set X, then $\tilde{G}$, which is defined as the pointwise supremum of $\tilde{L}$, is also (strictly) convex. But

$$\nabla_x \tilde{L}(x, y) = \left[A + A^{\mathrm{T}} - Q\right]^{\mathrm{T}} (x - y),$$

so that

$$\nabla^2_{xx} \tilde{L}(x, y) = A + A^{\mathrm{T}} - Q,$$

and the result follows directly. □

Note that the choice of f as the symmetrization made in Example 2.2 leads to a strictly convex gap function whenever A is positive definite. Moreover, this condition also leads to the stationary point property, see Theorem 4.4.2. Whether $\tilde{G}$ can be constructed convex for non-affine mappings F is an open question.

4.2 A descent algorithm for asymmetric variational inequalities

We conclude this section by developing descent algorithms, based on the stationary point characterization of Theorem 4.4. We also investigate the possibility of obtaining descent directions from inexact solutions of [PGG], and using inexact line searches; both these extensions of the basic scheme are of interest in implementations.

Theorem 4.7 (Descent properties) *Let $F \in C^1$ on X and $f \in C^2$ on X. Also, for any $x \in X$, let $Y(x)$ denote the set of optimal solutions to* [PGG].

1. *If X is bounded, $x \notin \Omega$, and the matrix $\nabla F(x)^{\mathrm{T}} - \nabla^2 f(x)$ is positive definite, then there exists a $y \in Y(x)$ such that*

$$\tilde{G}'(x; y - x) < 0.$$

2. *If X is bounded, f is strictly convex on X, $x \notin \Omega$ and the matrix $\nabla F(x)^{\mathrm{T}} - \nabla^2 f(x)$ is positive semidefinite, then*

$$\nabla \tilde{G}(x)^{\mathrm{T}}(y(x) - x) < 0.$$

3. *Assume that F is strongly monotone on X, and that f is strongly convex with a Lipschitz continuous gradient on X. Then, for any $x \in X$,*

$$\nabla \tilde{G}(x)^{\mathrm{T}}(y(x) - x) \leq -\left(m_F + m_f - M_{\nabla f}\right) \|y(x) - x\|^2. \tag{4.11}$$

Proof

1. Follows from expressing $\tilde{G}'(x; y - x)$ as in (4.4), and using the same technique as in the proof of Theorem 4.4, together with (2.2), to establish the existence of a $\overline{y} \in Y(x)$ such that $\tilde{G}'(x; \overline{y} - x) < 0$.

2. Follows from the expression

$$\tilde{G}'(x; y(x) - x) = \nabla \tilde{G}(x)^{\mathrm{T}}(y(x) - x),$$

combined with (2.2).

3. Proven as in 2. □

Note that for affine F and f quadratic, the conditions 2. reduce to the positive semidefiniteness of the matrix $A^{\mathrm{T}} - Q$. The symmetrization given in Example 2.2 has the descent property whenever A is positive definite. Note also that, because of the generality of f, the conditions of Theorem 4.7 do not reduce to those of Theorem 3.4 when choosing f quadratic.

Using the stationary point property we have thus shown that the solution to [PGG], obtained when calculating $\tilde{G}(x)$, defines a feasible descent direction with respect to $\tilde{G}$, whenever $x \notin \Omega$. We next show that, using an exact line search rule, the sequence of iteration points converges towards the unique solution x^*, from any feasible starting point.

Theorem 4.8 (Global convergence) *Let $F \in C^1$ on X and $f \in C^2$ on X. Let x^0 be arbitrary in X, and let x^{k+1} be determined from x^k through the exact line search rule*

$$\min_{0 \leq t \leq 1} \tilde{G}\left(x^k + td^k\right),$$

where $d^k = y\left(x^k\right) - x^k$ with $y\left(x^k\right) \in Y\left(x^k\right)$. Consider the following sets of assumptions.

1. *X is bounded, the matrix $\nabla F(x)^{\mathrm{T}} - \nabla^2 f(x)$ is positive definite on X, and the point $y\left(x^k\right)$ is chosen as in Theorem 4.7.1.*
2. *X is bounded, f is strictly convex on X, and the matrix $\nabla F(x)^{\mathrm{T}} - \nabla^2 f(x)$ is positive semidefinite on X.*
3. *F is strongly monotone on X, f is strongly convex with a Lipschitz continuous gradient on X and chosen such that (4.3) holds. Further, X is bounded or $M_{\nabla f} < 2m_F$ holds.*

Under either one of the above sets of assumptions, [VIP] *has a unique solution, and the sequence $\left\{x^k\right\}$ converges to it.*

Proof We begin by showing that the assumptions of Theorem A of Zangwill (1969) are fulfilled in all the three cases, i.e., $i)$ that $\left\{x^k\right\}$ stay in a compact subset of X, $ii)$ that $\tilde{G}\left(x^{k+1}\right) < \tilde{G}\left(x^k\right)$ whenever $x^k \notin \Omega$, and $x^k \in \Omega$ if $\tilde{G}\left(x^{k+1}\right) \geq \tilde{G}\left(x^k\right)$, and $iii)$ that the algorithmic map is closed.

$i)$ For cases 1. and 2., this is obvious since X is compact. For case 3., either it holds from the compactness of X, or from the compactness of the level sets (cf. Corollary 4.1).

$ii)$ Follows directly from Theorems 4.4 and 4.7, and the choice of step length.

$iii)$ For case 1., it follows from Theorem 4.2.1 and the closedness of $S(x)$ in the proof of Theorem 4.4, that the point-to-set map assigning the search direction d^k to the point x^k is closed. For cases 2. and 3., the continuity of this map follows from Theorem 4.2.4. Since the exact line search is well known to be closed for a continuous objective, the composite algorithmic map is closed for all three cases.

Thus, by Zangwill's Theorem A, every accumulation point [from i), at least one such point exists] is optimal in [VIP]. For all three cases, it is straightforward to show that Ω is a singleton, and we conclude that $\left\{x^k\right\}$ converges to the unique solution of [VIP]. $\square$

Remark 4.2 Boundedness of X, which was assumed by Fukushima (1992), is not required for the global convergence of the algorithm. If x^0 is infeasible, x^0 can be replaced by $y\left(x^0\right)$.

Global convergence with an Armijo step length rule can also be established, extending the result of Fukushima for the gap function $\tilde{G}_{\mathrm{Fu}}$. It is well known that the use of the Armijo rule requires the gradient of the objective to be Lipschitz continuous; in this case, this property is a consequence of the Lipschitz continuity of the map $x \mapsto Y(x)$, which is established in the below lemma.

Lemma 4.1 (Lipschitz continuity of Y) *Let F be Lipschitz continuous on X, and f strongly convex with a Lipschitz continuous gradient on X. Then the mapping $x \mapsto Y(x)$ is Lipschitz continuous on X with modulus $\frac{M_F+M_{\nabla f}}{m_f}$.*

Proof Given arbitrary $x^1, x^2 \in X$, combining the two inequalities (2.2), where we choose $y = y\left(x^2\right)$ and $y = y\left(x^1\right)$, respectively, and rearranging terms, we obtain

$$\begin{aligned}
&\left[F\left(x^2\right)-F\left(x^1\right)\right]^{\mathrm{T}}\left(y\left(x^2\right)-y\left(x^1\right)\right)+ \\
&\quad\left[\nabla f\left(x^1\right)-\nabla f\left(x^2\right)\right]^{\mathrm{T}}\left(y\left(x^2\right)-y\left(x^1\right)\right) \leq \left[\nabla f\left(y\left(x^1\right)\right)-\nabla f\left(y\left(x^2\right)\right)\right]^{\mathrm{T}}\left(y\left(x^2\right)-y\left(x^1\right)\right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \leq -m_f\left\|y\left(x^2\right)-y\left(x^1\right)\right\|^2,
\end{aligned}$$

which implies that

$$\begin{aligned}
m_f\left\|y\left(x^2\right)-y\left(x^1\right)\right\|^2 &\leq \left(\left\|F\left(x^2\right)-F\left(x^1\right)\right\|+\left\|\nabla f\left(x^1\right)-\nabla f\left(x^2\right)\right\|\right)\left\|y\left(x^2\right)-y\left(x^1\right)\right\| \\
&\leq \left(M_F+M_{\nabla f}\right)\left\|x^2-x^1\right\|\left\|y\left(x^2\right)-y\left(x^1\right)\right\|,
\end{aligned}$$

and the result follows. $\square$

Global convergence under an Armijo step length rule is shown below for the case of strongly monotone variational inequalities.

Theorem 4.9 (Global convergence under Armijo rule) *Let the assumptions of Theorem 4.8.3 and Lemma 4.1 hold. Furthermore, let ∇F and $\nabla^2 f$ be Lipschitz continuous on X, and replace the exact line search with the Armijo step length rule (3.5), where $\tilde{G}$ replaces $\tilde{G}_{\mathrm{Fu}}$. Letting $\mu \stackrel{def}{=} m_F + m_f - M_{\nabla f} > 0$, the conclusions of Theorem 4.8.3 hold if α is chosen such that $\alpha < \mu$.*

Proof From Lemma 4.1 and the Lipschitz continuity of ∇F and $\nabla^2 f$, it follows that $\nabla \tilde{G}$ is also Lipschitz continuous on X. Using this fact together with (4.7), we obtain, for all $0 \leq t \leq 1$, that

$$\tilde{G}\left(x^k + td^k\right) - \tilde{G}\left(x^k\right) \leq -t\left(\mu - \frac{t}{2} M_{\nabla \tilde{G}}\right) \left\|d^k\right\|^2,$$

where $d^k = y\left(x^k\right) - x^k$. Tracing the proof of Theorem 4.2 of Fukushima (1992) we may conclude that the step length $t_k = \beta^{l_k}$ satisfies

$$t_k > \min\left\{\beta, 2\beta(\mu - \alpha)/M_{\nabla \tilde{G}}\right\}, \tag{4.12}$$

and

$$\tilde{G}\left(x^{k+1}\right) \leq \tilde{G}\left(x^k\right) - \alpha t_k \left\|d^k\right\|^2. \tag{4.13}$$

Since, from (4.13), the sequence $\left\{\tilde{G}\left(x^k\right)\right\}$ is monotonically decreasing, it follows from (4.12) that $\lim_{k \to +\infty} \left\|d^k\right\| = 0$.

The sequence $\left\{x^k\right\}$ is contained in a compact subset of X, and hence has at least one accumulation point, say $\overline{x}$. From the continuity property of $y(x)$ (cf. Theorem 4.2), $\overline{x} = y(\overline{x})$, so that, by Theorem 4.1, $\overline{x}$ solves [VIP]. The strong monotonicity of F then implies that $\overline{x}$ must be the unique accumulation point, and we conclude that the sequence $\left\{x^k\right\}$ converges to the unique solution to [VIP]. □

Remark 4.3 The subproblem solution $y\left(x^{k+1}\right)$ is obtained in the Armijo line search for t_k, and is therefore provided with no extra computation.

Assume now that the algorithm for solving [PGG] is truncated prematurely, in the sense that it is terminated when, for some a priori determined $\varepsilon > 0$, a point $y_\varepsilon(x)$ is found, satisfying

$$\nabla_y \tilde{L}\left(x, y_\varepsilon(x)\right)^{\mathrm{T}} \left(y - y_\varepsilon(x)\right) \leq \varepsilon, \forall\, y \in X. \tag{4.14}$$

This implies, by the concavity of $\tilde{L}$ with respect to y, that $\tilde{G}(x) \leq \tilde{L}\left(x, y_\varepsilon(x)\right) + \varepsilon$, i.e., that $y_\varepsilon(x)$ is an ε-optimal solution to [PGG]. [Note that when X is polyhedral the termination criterion (4.14) is easily checked by solving the linear subproblem of the Frank-Wolfe algorithm, see, e.g., Migdalas (1990).] An interesting question then is if a point $y_\varepsilon(x)$ satisfying (4.14) also satisfies $\nabla \tilde{G}(x)^{\mathrm{T}} \left(y_\varepsilon(x) - x\right) < 0$, for all sufficiently small ε. If the assumptions of Theorem 4.7.3 are satisfied, the theorem below establishes an affirmative answer.

Theorem 4.10 (Descent obtained from truncated subproblems) *Let the assumptions of Theorem 4.7.3 be satisfied. Then, for any $x \in X$ and $\varepsilon \geq 0$,*

$$\|y(x) - y_\varepsilon(x)\|^2 \leq \frac{\varepsilon}{m_f}. \tag{4.15}$$

Furthermore, if $x \notin \Omega$, (4.3) holds, and $\varepsilon > 0$ is chosen small enough (dependent of x), then

$$\nabla\tilde{G}(x)^{\mathrm{T}}\left(y_\varepsilon(x) - x\right) < 0.$$

Proof From (2.2) and (4.14) we have

1. $\left[-\nabla f(y(x)) - F(x) + \nabla f(x)\right]^{\mathrm{T}}(y - y(x)) \leq 0, \forall\, y \in X$, and
2. $\left[-\nabla f\left(y_\varepsilon(x)\right) - F(x) + \nabla f(x)\right]^{\mathrm{T}}(y - y_\varepsilon(x)) \leq \varepsilon, \forall\, y \in X$.

Combining these inequalities, with $y = y_\varepsilon(x)$ and $y = y(x)$, respectively, we obtain, also using the strong convexity of f, that

$$m_f \left\|y(x) - y_\varepsilon(x)\right\|^2 \leq \left[\nabla f(y(x)) - \nabla f\left(y_\varepsilon(x)\right)\right]^{\mathrm{T}}\left(y(x) - y_\varepsilon(x)\right) \leq \varepsilon,$$

which yields (4.15).

Now, from Theorems 4.3 and 4.7.3,

$$\begin{aligned}
\nabla\tilde{G}(x)^{\mathrm{T}}(y_\varepsilon(x) - x) &= \nabla\tilde{G}(x)^{\mathrm{T}}(y(x) - x) + \nabla\tilde{G}(x)^{\mathrm{T}}(y_\varepsilon(x) - y(x)) \\
&\leq -\left(m_F + m_f - M_{\nabla f}\right)\|y(x) - x\|^2 + F(x)^{\mathrm{T}}(y_\varepsilon(x) - y(x)) \\
&\quad +(x - y(x))^{\mathrm{T}}\left[\nabla F(x)^{\mathrm{T}} - \nabla^2 f(x)\right](y_\varepsilon(x) - y(x)) \\
&\leq -\left(m_F + m_f - M_{\nabla f}\right)\|y(x) - x\|^2 + \\
&\quad \left(\|F(x)\| + \left\|\nabla F(x)^{\mathrm{T}} - \nabla^2 f(x)\right\| \cdot \|x - y(x)\|\right)\|y_\varepsilon(x) - y(x)\| \\
&< 0,
\end{aligned}$$

if

$$\|y_\varepsilon(x) - y(x)\| < -\frac{\left(m_F + m_f - M_{\nabla f}\right)\|y(x) - x\|^2}{\|F(x)\| + \|\nabla F(x)^{\mathrm{T}} - \nabla^2 f(x)\| \cdot \|x - y(x)\|}. \tag{4.16}$$

Since the right hand side of this strict inequality is strictly positive whenever $x \notin \Omega$, it follows from (4.15) that (4.16) holds by choosing ε small enough. □

Remark 4.4 By further imposing Lipschitz continuity of f on X, the term $\|y(x) - x\|$ may be estimated both from above and below by bounded positive measures in x, ε and $y_\varepsilon(x)$ only, thus enabling the right hand side of (4.16) to be bounded from below. The resulting bound may then be used in a termination criterion for the solution of [PGG]. Also, this bound tends to zero when x approaches Ω, so that, in effect, the condition in (4.16) amounts to solving [PGG] accurately in the limit. Introducing a sequence $\{\varepsilon_k\}$ with $\varepsilon_k \to 0$, the global convergence of the overall scheme is also easily established from Theorem 4.8.

In the next section we shall extend Auchmuty's class of gap functions, by introducing a sequence of functions f_k, with the same properties as the function f in Auchmuty's framework. By associating with each step k in an iterative algorithm an appropriate function f_k, the problem [PGG] may thus even better approximate [VIP]. The introduction of this more general class of gap functions also results in interesting relationships with successive approximation-based algorithms of Cohen (1988), Noor (1992) and Dafermos (1983), and with the descent algorithms of Wu *et al.* (1991).

5 Generalizations of Auchmuty's class of gap functions

Cohen (1988) develops an iterative algorithm for [VIP] using a successive approximation approach. Below we show that the subproblem solved in this algorithm is *equivalent* to [PGG], so that, immediately, a successive approximation algorithm based on the calculation of $\tilde{G}$ is established.

In our notation the subproblem, given $x^k \in X$, is to find $x^{k+1} \in X$ solving

$$\text{[C]} \qquad \min_{x \in X} \left\{ f(x) + \left[\varepsilon F\left(x^k\right) - \nabla f\left(x^k\right) \right]^{\mathrm{T}} \left(x - x^k\right) \right\},$$

where ε is a positive constant. (Cohen refers to this problem as the *auxiliary problem*, and to the algorithm as the *auxiliary problem principle.*) If we, in Auchmuty's framework, replace f by $\frac{1}{\varepsilon} f$, then it is not difficult to see that [PGG] is equivalent to [C]. The successive approximation algorithm, which thus is defined by letting $x^{k+1} = y\left(x^k\right)$ in [PGG], is globally convergent through Cohen's analysis. (A successive approximation algorithm can then be seen as an alternative to the descent approach, if the line search with respect to $\tilde{G}$ is considered computationally too expensive.)

The global convergence of the successive approximation variant of the descent methods of Section 4.2 is given below. The proof follows from Theorem 2.2 of Cohen (1988).

Theorem 5.1 (A successive approximation algorithm) *Let F be strongly monotone and Lipschitz continuous on X, and let f be strongly convex on X. Let x^0 be arbitrary in X, and let x^{k+1} be the (unique) solution to* [PGG]*, defined at x^k. Then the sequence $\left\{x^k\right\}$ converges towards the unique solution to* [VIP]*, provided that*

$$2m_f > \frac{M_F^2}{m_F}. \tag{5.1}$$

Remark 5.1 The algorithm defined in the theorem includes well known algorithms such as the projection algorithm [e.g. Dafermos (1983)]. It is interesting to note that the convergence criterion (5.1) amounts to choosing the function f so that it, in a sense, dominates the original mapping F, while in the descent approach the convergence criteria [e.g. (4.3)] amounts to precisely the opposite.

Remark 5.2 Noor (1992) recently extended Cohen's algorithm to allow for variational inequality subproblems [cf. Remark 2.4 in Cohen (1988)]. This means that, in [AVIP], ∇f is replaced by a mapping $a : \Re^n \mapsto \Re^n$ that is *not* necessarily a gradient; thus [AVIP] does not reduce to the problem [PGG], i.e., to the calculation of a gap function.

Cohen develops his approach to also allow for F being a point-to-set map, and a varying parameter, ε_k, in [C]. Global convergence is now ensured by also letting $\{\varepsilon_k\}$ fulfill the conditions

$$\varepsilon_k > 0, \quad \sum_{k=0}^{+\infty} \varepsilon_k = +\infty \quad \text{and} \quad \sum_{k=0}^{+\infty} \varepsilon_k^2 < +\infty.$$

This result of course carries over directly to Auchmuty's framework, provided that the function f is allowed to change with iteration (which would correspond to letting $f_k \stackrel{\text{def}}{=} \frac{1}{\varepsilon_k} \cdot f$).

From the above discussion we are led to introduce a sequence $\{f_k\}$ of functions, where each function f_k inherits the properties of f in Auchmuty's framework. The rationale behind this generalization is that the gradients ∇f_k can be chosen so as to better approximate $F\left(x^k\right)$ than the fixed gradient ∇f (cf. Section 2).

Introducing the sequence $\{f_k\}$ thus introduces a sequence of gap functions $\left\{\tilde{G}_k\right\}$, where each function $\tilde{G}_k$ inherits the properties of $\tilde{G}$, given in Theorems 2.1, 2.2 and 4.1. Without imposing a continuity assumption [such as *equicontinuity* in the sense of Rockafellar (1970)] on the sequence of functions f_k, the gap decreasing algorithms of Section 4.2 can not be easily extended to the more general case. Global convergence under equicontinuity assumptions is an open problem.

A natural way of ensuring continuity is to restrict the functions f_k to vary continuously, in the following sense. Introduce $f : \Re^n \times X \mapsto \Re$ as a continuous function with respect to its second argument, and convex and continuously differentiable with respect to its first. In the point x^k, the function $f_k = f\left(\cdot, x^k\right)$ then replaces f in [PGG]. Since this function defines a special case of the sequence of functions f_k, the Theorems 2.1, 2.2 and 4.1 apply directly. [A proof of this fact is given in Patriksson (1991b).] Furthermore, the gap function $\tilde{G}$, defined through this special case of $\left\{\tilde{G}_k\right\}$, is continuous under the same conditions as in Theorem 4.2.2.

In the rest of this section we shall assume f to be a continuous function on $\Re^n \times X$ as defined above (and let $\tilde{G}$, [PGG] and [AVIP] be defined accordingly), and consider this generalization of Auchmuty's framework. We will show that the descent algorithms derived in Section 4.2 may be extended to the more general case of function f, and relate this class of algorithms to those of Dafermos (1983) and Wu *et al.* (1991). We first consider the works of Dafermos (1983).

Dafermos presents a general framework of successive approximation algorithms for [VIP]. Dafermos defines a smooth function $h : X \times X \mapsto \Re^n$, with the properties *i)* $h(x, x) = F(x), \forall\, x \in X$, and *ii)* $\nabla_x h(x, y)$ is symmetric and positive definite for any $x, y \in X$. A successive approximation algorithm is then defined through the subproblem

$$h\left(x^{k+1}, x^k\right)^{\mathrm{T}}\left(x - x^{k+1}\right) \geq 0, \forall\, x \in X. \tag{5.2}$$

(Due to assumption *ii*) above, (5.2) reduces to a strictly convex minimization problem.)

The sequence $\left\{x^k\right\}$ is convergent towards the unique solution of [VIP] if

$$\left\| \nabla_x h^{-1/2}\left(x^1, y^1\right) \cdot \nabla_y h\left(x^2, y^2\right) \cdot \nabla_x h^{-1/2}\left(x^3, y^3\right) \right\| < 1, \forall\, x^i, y^i \in X, \tag{5.3}$$

for $i = 1, 2, 3$. (Dafermos shows that Condition (5.3) implies strict monotonicity of F.) Several variational inequality algorithms have been analyzed in Dafermos' framework, especially for economic equilibrium problems, see, e.g., Zhao and Dafermos (1991), and the references cited therein.

Patriksson (1991b) shows that the subproblem (5.2) and [PGG] are equivalent, in the sense that for any choice of function f, a function h may be chosen so that the subproblem (5.2) equals [PGG], and vice versa. (Dafermos' class of methods may therefore be viewed as successive approximation variants of the descent algorithms presented in this paper, for the choice of functions $f : \Re^n \times X \mapsto \Re$.) The class of algorithms defined by Cohen (1988) is then clearly a special case of Dafermos' class of algorithms. Since her convergence results actually do not require the mapping h to be symmetric [cf. Remark 2.2 in Dafermos (1983)], the same conclusion follows for Noor's class of methods.

An interesting question provoked by this algorithmic equivalence result is whether the gap decreasing algorithms of Section 4.2 can be extended to include functions $f : \Re^n \times X \mapsto \Re$, and thereby introduce an accelerating line search in Dafermos' class of algorithms. As is shown below (through another interesting algorithmic equivalence result), the answer is in the affirmative, at least for some choices of such functions f.

Let f be such that $\nabla_y f(x, x) \stackrel{\text{def}}{=} \nabla_y f(y, x)|_{y=x} = 0$, for all $x \in X$. Then [PGG] may be written as

$$\min_{y \in X} \left\{ f(y, x) + F(x)^{\mathrm{T}}(y - x) \right\}, \tag{5.4}$$

which is the subproblem of Wu *et al.* (1991). [Actually, Patriksson (1991b) shows that [PGG] and (5.4) are *equivalent*, in the same sense as for [PGG] and (5.2).] Wu *et al.* develop descent algorithms for [VIP] through the subproblem (5.4), and exact line searches with respect to the corresponding gap function. In the convergence analysis they consider the special case of choosing $f(y, x) = \frac{1}{2}(y - x)^{\mathrm{T}} Q(x)(y - x)$, where $Q(x)$ is a symmetric and positive definite matrix on X. (This implies a direct generalization of Fukushima's (1992) algorithm, where the matrix Q is fixed.) Their basic convergence result (Proposition 3.11) requires the matrix $\nabla Q(x)(y(x) - x)$ to be positive definite for all $x \in X$. This condition is, however, difficult to verify in practice.

Remark 5.3 In the case of nonlinear programming, the algorithmic classes of Cohen (1988) and Wu *et al.* (1991) reduce to the ones of Cohen (1980) and Patriksson (1991a), and to the ones of Migdalas (1990) and Tseng (1991), respectively. A partial list of corresponding algorithmic equivalence results can be found in Patriksson (1991b); further studies of algorithmic equivalence results are made in Patriksson (1993).

Remark 5.4 It seems, from the discussions made above, that it is an inherent property of the algorithms that their convergence conditions are *severe* and very difficult to verify in cases when f is allowed to be iteration dependent [cf. Condition (5.3) in Dafermos' successive approximation method and the conditions in Wu's *et al.* descent method], while the conditions in the case of iteration independent functions f may be stated in terms of the original data (including, typically, Lipschitz and strongly monotone constants) [cf. Condition (4.3) in Auchmuty's framework, Condition (5.1) in Cohen's framework and the conditions of Theorem 3.1 of Noor (1992)].

6 Concluding remarks

In this paper we have developed and unified mathematical programming reformulations of variational inequalities, as well as proposed new descent algorithms derived from these reformulations. We have also provided algorithmic equivalence results, relating Auchmuty's framework to the algorithm classes of Dafermos (1983), Cohen (1988), Wu *et al.* (1991) and Noor (1992).

The framework of Auchmuty extends hitherto proposed optimization formulations, in the sense that the function f defining [OPG] may be chosen rather arbitrarily, and may therefore be used to more closely approximate the original mapping F (cf. Section 2). Moreover, it is possible to take problem structure into account when choosing the function f. As an example [due to Cohen (1988)], consider the case where the feasible set X is a Cartesian product, i.e., $X = \prod_{i \in \mathcal{C}} X_i$, where X_i is a closed and convex set in $\Re^{n_i}$, and $\sum_{i \in \mathcal{C}} n_i = n$. [A variety of equilibrium problems can be cast in terms of such variational inequalities; see Pang (1985), Bertsekas and Tsitsiklis (1989, Section 3.5), and Harker and Pang (1990).] It may then be appropriate to choose

$$f(x) = \sum_{i \in \mathcal{C}} f_i\left(x_i\right),$$

where each function $f_i : X_i \mapsto \Re$ is convex and continuously differentiable on X_i. The evaluation of $\tilde{G}(x)$, i.e., the solution of the convex problem [PGG], then decomposes into $|\mathcal{C}|$ problems of smaller dimensions. Since a decomposition takes place in the line search phase also, the computations involved in the descent algorithms of Section 4.2 could in this example effectively utilize parallel computing environments.

Patriksson (1993) combines the gap decreasing algorithm outlined above with cyclic decomposition with respect to the individual sets X_i. In this combined algorithm, search directions for $\tilde{G}$ are defined through the solution of [PGG] where, in some iteration k, all variables but x_{i_k}, $i_k \in \mathcal{C}$, are fixed to their current values. The algorithm, which generalizes the standard cyclic (Gauss-Seidel) decomposition methods (see, e.g., Dafermos, 1983) to a descent algorithm, can be established convergent under conditions similar to those of Theorem 4.9, and if further each index $i \in \mathcal{C}$ is chosen at least once every $B \geq |\mathcal{C}|$ successive iterations.

If the feasible set X is a closed and convex cone, then [VIP] reduces to the *nonlinear complementarity problem*, i.e., to find $x^* \in X$ such that

[NCP] $$F(x^*) \in X^* \text{ and } F(x^*)^{\mathrm{T}} x^* = 0,$$

where X^* is the polar set of X. In the special case of $X = \Re^n_+$, Cottle (1966) (see also Karamardian, 1971) shows that x^* solves [NCP] if and only if x^* minimizes $F(x)^{\mathrm{T}}x$ on the set $\{x \geq 0 \mid F(x) \geq 0\}$, and $F(x^*)^{\mathrm{T}} x^* = 0$. Auchmuty shows that [OPG] extends this optimization reformulation to a general nonlinear complementarity problem, by establishing that [OPG] in fact is equivalent to minimizing $F(x)^{\mathrm{T}}x$ on the set $\{x \in X \mid F(x) \in X^*\}$, and that x^* solves [NCP] if and only if $F(x^*)^{\mathrm{T}} x^* = 0$ (i.e., if and only if $G(x^*) = 0$). Noting that, in this case, $f \equiv 0$, we may therefore conclude that the generalized gap function $\tilde{G}$ provides a means for making optimization reformulations of nonlinear complementarity problems; the one of Cottle is indeed a very special case of such an optimization reformulation. (New descent algorithms for [NCP] may thus be constructed, by specializing those algorithms developed in this paper that are valid also for variational inequalities over unbounded feasible sets.)

An interesting question for further studies regards the convexity properties of the gap function $\tilde{G}$. There are, seemingly (see Section 4.1), some connections between the descent, convexity and coercivity properties of $\tilde{G}$ that is not covered by the existing theory. Is it possible to characterize these connections, and to extend them to non-affine variational inequalities?

Some extensions of the developed theory and algorithms would be of interest to investigate. Such extensions include infinite-dimensional settings, as considered by Cohen (1988) and Noor (1992), an introduction of point-to-set maps F, as in Cohen (1988), equicontinuity of f_k, and the possibility to combine approximate line searches and truncated subproblems. Newton methods with line searches on gap functions have been shown to be globally convergent for the two special choices of $f \equiv 0$ and $f(x) = \frac{1}{2}x^{\mathrm{T}}Qx$ (see, e.g., Marcotte and Dussault, 1985, and Taji *et al.*, 1991). The possibility of extending these algorithms to more general functions f is currently under investigation. Finally, the practical as well as the theoretical convergence rates obtained for various choices of functions f, and line search and subproblem truncation strategies, must be studied further.

Acknowledgements

The authors would like to thank Dr. Athanasios Migdalas for valuable discussions, and the referees for several remarks leading to improvements of the paper.

References

[] L. Armijo, "Minimization of functions having Lipschitz continuous partial derivatives," *Pacific Journal of Mathematics* 16 (1964) 1–3.

[] G. Auchmuty, "Variational principles for variational inequalities," *Numerical Functional Analysis and Optimization* 10 (1989) 863–874.

A. Auslender, *Optimisation: Méthodes numériques* (Masson, Paris, France, 1976).

M.S. Bazaraa and C.M. Shetty, *Nonlinear programming: Theory and algorithms* (John Wiley & Sons, New York, NY, 1979).

D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and distributed computation: Numerical methods* (Prentice-Hall, London, 1989).

F.H. Clarke, "Generalized gradients and applications," *Transactions of the American Mathematical Society* 205 (1975) 247–262.

G. Cohen, "Auxiliary problem principle and decomposition of optimization problems," *Journal of Optimization Theory and Applications* 32 (1980) 277–305.

G. Cohen, "Auxiliary problem principle extended to variational inequalities," *Journal of Optimization Theory and Applications* 59 (1988) 325–333.

R.W. Cottle, "Nonlinear programs with positively bounded Jacobians," *SIAM Journal on Applied Mathematics* 14 (1966) 147–158.

S. Dafermos, "An iterative scheme for variational inequalities," *Mathematical Programming* 26 (1983) 40–47.

V.F. Demyanov and A.B. Pevnyi, "Numerical methods for finding saddle points," *USSR Computational Mathematics and Mathematical Physics* 12 (1972) 11–52.

V.F. Demyanov and A.M. Rubinov, *Approximate methods in optimization problems* (Elsevier, New York, NY, 1970).

J.C. Dunn, "Convergence rates for conditional gradient sequences generated by implicit step length rules," *SIAM Journal of Control and Optimization* 18 (1980) 473–487.

Yu.G. Evtushenko, *Numerical optimization techniques* (Optimization Software, Inc., New York, NY, 1985).

M. Fukushima, "Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems," *Mathematical Programming* 53 (1992) 99–110.

J.C. Hammond and T.L. Magnanti, "A contracting ellipsoid method for variational inequality problems," Working Paper OR 160-87, Operations Research Center, Massachusetts Institute of Technology (Cambridge, MA, 1987).

P.T. Harker and J.-S. Pang, "Finite dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications," *Mathematical Programming* 48 (1990) 161–220.

D.W. Hearn, "Network aggregation in transportation planning models, Part I," Mathtec Final Report DOT-TSC-RSPD-78-8, Mathtec, Inc. (Princeton, NJ, 1978).

D.W. Hearn, "The gap function of a convex program," *Operation Research Letters* 1 (1982) 67–71.

D.W. Hearn and S. Lawphongpanich, "A dual ascent algorithm for traffic assignment problems," Conference paper, Joint Seminar on Urban Traffic Networks, Capri, Italy, 1989.

D.W. Hearn, S. Lawphongpanich, and S. Nguyen, "Convex programming formulations of the asymmetric traffic assignment problem," *Transportation Research* 18B (1984) 357–365.

D.W. Hearn and S. Nguyen, "Dual and saddle functions related to the gap function," Research Report 82-4, Department of Industrial and Systems Engineering, University of Florida (Gainesville, FL, 1982).

W.W. Hogan, "Point-to-set maps in mathematical programming," *SIAM Review* 15 (1973) 591–603.

S. Kakutani, "A generalization of Brouwer's fixed point theorem," *Duke Mathematics Journal* 8 (1941) 457–459.

S. Karamardian, "Generalized complementarity problem," *Journal of Optimization Theory and Applications* 8 (1971) 161–168.

S. Karamardian, "An existence theorem for the complementarity problem," *Journal of Optimization Theory and Applications* 18 (1976) 445–454.

D. Kinderlehrer and G. Stampacchia, *An introduction to variational inequalities and their applications* (Academic Press, New York, NY, 1980).

S. Lawphongpanich and D.W. Hearn, "Simplicial decomposition of the asymmetric traffic assignment problem," *Transportation Research* 18B (1984) 123–133.

P. Marcotte, "A new algorithm for solving variational inequalities with application to the traffic assignment problem," *Mathematical Programming Study* 33 (1985) 339–351.

P. Marcotte, "Gap-decreasing algorithms for monotone variational inequalities," Conference paper, ORSA/TIMS Joint National Meeting, Miami Beach, FL, 1986.

P. Marcotte and J.-P. Dussault, "A modified Newton method for solving variational inequalities," in: *Proceedings of the 24th IEEE Conference on Decision and Control* (Fort Lauderdale, FL, 1985) pp. 1433–1436.

P. Marcotte and J.-P. Dussault, "A note on a globally convergent Newton method for solving monotone variational inequalities," *Operations Research Letters* 6 (1987) 35–42.

P. Marcotte and J.-P. Dussault, “A sequential linear programming algorithm for solving monotone variational inequalities,” *SIAM Journal of Control and Optimization* 27 (1989) 1260–1278.

P. Marcotte and J. Guélat, “Adaptation of a modified Newton method for solving the asymmetric traffic equilibrium problem,” *Transportation Science* 22 (1988) 112–124.

A. Migdalas, “A regularization of the Frank-Wolfe algorithm,” Report LiTH-MAT-R-90-10, Linköping Institute of Technology (Linköping, Sweden, 1990).

G.J. Minty, “Monotone (non-linear) operators in Hilbert space,” *Duke Mathematics Journal* 29 (1962) 341–346.

S. Nguyen and C. Dupuis, “An efficient method for computing traffic equilibria in networks with asymmetric transportation costs,” *Transportation Science* 18 (1984) 185–202.

M.A. Noor, “General algorithm for variational inequalities,” *Journal of Optimization Theory and Applications* 73 (1992) 409–413.

J.M. Ortega and W.C. Rheinboldt, *Iterative solution of nonlinear equations in several variables* (Academic Press, New York, NY, 1970).

J.-S. Pang, “Asymmetric variational inequality problems over Cartesian product sets: applications and iterative methods,” *Mathematical Programming* 31 (1985) 206–219.

J.-S. Pang, “A posteriori error bounds for the linearly-constrained variational inequality problem,” *Mathematics of Operations Research* 12 (1987) 474–484.

M. Patriksson, “Partial linearization methods in nonlinear programming,” Report LiTH-MAT-R-91-11, Linköping Institute of Technology (Linköping, Sweden, 1991). Forthcoming in *Journal of Optimization Theory and Applications* 78 (1993).

M. Patriksson, “A unified description of iterative algorithms for traffic equilibria,” Report LiTH-MAT-R-91-34, Linköping Institute of Technology (Linköping, Sweden, 1991). Forthcoming in *European Journal of Operational Research.*

M. Patriksson, “A unified framework of descent algorithms for nonlinear programming and variational inequalities,” Ph.D. thesis, Linköping University (Linköping, Sweden, 1993).

M.E. Primak, “A computational process of search for equilibrium points,” *Cybernetics* 9 (1975) 106–113.

R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, NJ, 1970).

R.T. Rockafellar, *The theory of subgradients and its applications to problems of optimization: Convex and nonconvex functions* (Heldermann Verlag, Berlin, 1981).

M.J. Smith, "The existence and calculation of traffic equilibria," *Transportation Research* 17B (1983) 291–303.

M.J. Smith, "An algorithm for solving asymmetric equilibrium problems with a continuous cost-flow function," *Transportation Research* 17B (1983) 365–371.

K. Taji, M. Fukushima, and T. Ibaraki, "A globally convergent Newton method for solving strongly monotone variational inequalities," Technical Report 91006, Kyoto University (Kyoto, Japan, 1991). Forthcoming in *Mathematical Programming.*

P. Tseng, "Decomposition algorithm for convex differentiable minimization," *Journal of Optimization Theory and Applications* 70 (1991) 109–135.

J.H. Wu, M. Florian, and P. Marcotte, "A general descent framework for the monotone variational inequality problem," Publication 723, Centre de recherche sur les transports, Université de Montréal (Montréal, Canada, 1991).

W.I. Zangwill, *Nonlinear programming: A unified approach* (Prentice-Hall, Englewood Cliffs, NJ, 1969).

L. Zhao and S. Dafermos, "General economic equilibrium and variational inequalities," *Operations Research Letters* 10 (1991) 369–376.

S.I. Zuhovickiĭ, R.A. Poljak, and M.E. Primak, "Two methods of search for equilibrium points of n-person concave games," *Soviet Mathematics Doklady* 10 (1969) 279–282.

S.I. Zuhovickiĭ, R.A. Poljak, and M.E. Primak, "Numerical methods of finding equilibrium points of n-person games," in: *Proceedings of the First Winter School of Mathematical Programming at Drogobych* (1969) no. 1, pp. 93–130.

S.I. Zuhovickiĭ, R.A. Poljak, and M.E. Primak, "On an n-person concave game and a production model," *Soviet Mathematics Doklady* 11 (1970) 522–526.

S.I. Zuhovickiĭ, R.A. Poljak, and M.E. Primak, "Concave multiperson games. Numerical methods," *Matekon* 9 (1973) 10–30.