Nonlinear Optimization and Applications 2, pp. 1-000
G. Di Pillo and F. Giannessi, Editors

A new merit function and an SQP method for non-strictly monotone variational inequalities

Michael Patriksson (mipat@math.chalmers.se)
Department of Mathematics
Chalmers University of Technology
S-412 96 Gothenburg, Sweden

Abstract

Merit functions utilized to monitor the convergence of sequential quadratic programming (SQP) methods for nonlinear programs and variational inequality problems have in common that they include a penalty function for the explicit constraints, the value of the penalty parameter for which is subject to the requirement of being large enough compared to estimates of the optimal Lagrange multipliers. In existing applications of the SQP algorithm to variational inequality problems, the possible choices of merit functions used in descent algorithms are also limited by conditions which require the knowledge of problem parameters.

This paper introduces a merit function for use in SQP methods for possibly non-strictly monotone variational inequality problems which does not include an explicit penalty function, and whose application in descent algorithms does not require the knowledge of any problem parameters. This function is an extension of the merit function of Marcotte and Zhu (1995) to primal–dual variational inequality problems. Among its other properties, it is directionally differentiable, and exact for strongly monotone problems. Based on the new merit function, we establish the convergence of a general algorithm for primal–dual variational inequalities which includes an adaptive SQP algorithm as a special case.

Keywords: Variational inequalities, gap function, descent algorithm, sequential quadratic programming.

1 Introduction

Consider the following instance of the variational inequality problem:

$$F(x^*) + N_{X \cap G}(x^*) \ni 0^n, \tag{1}$$

where $X \subset \Re^n$ is a nonempty, closed, convex and bounded set, $G := \{ x \in \Re^n \mid g_i(x) \le 0, \quad i = 1, \dots, m \}$, where each function $g_i : \Re^n \mapsto \Re$ is convex and in C^2. We assume that $X \cap G \ne \emptyset$ and that a Slater constraint qualification holds, that is, that there exists a point $x \in X$ with $g_i(x) < 0$, $i = 1, \dots, m$. Further, we assume that $F : X \mapsto \Re^n$ is monotone, Lipschitz continuous and in C^1 on X. $N_S(x)$ denotes the normal cone to the closed and convex set S at x, that is,

$$N_S(x) = \begin{cases} \{ z \in \Re^n \mid z^{\mathrm{T}}(y - x) \le 0, \quad \forall y \in S \}, & x \in S, \\ \emptyset, & x \notin S. \end{cases} \tag{2}$$

With these assumptions, it is clear that the solution set, $\mathrm{SOL}(F, X \cap G)$, of (1), is nonempty, closed, convex and bounded.

We consider applications of (1) in which the ground set X is simple, for example a polytope, while G is defined by complicating constraints, such as when some of the functions g_i are nonlinear. In the development of a computational method for (1), we consider a relaxation approach wherein the functions g_i are iteratively linearized, thus iteratively replacing G by polyhedral approximations. Examples of such a method are sequential linear programming (SLP) and sequential quadratic programming (SQP). (We may also consider the case where the set X is not included in the original problem formulation, but is an a priori given trust region for the algorithm, although we have chosen not to describe the methodology in the framework of trust region methods.)

We introduce the following equivalent primal–dual formulation of (1). Introducing multipliers $\lambda_i \in \Re_+$, $i = 1, \dots, m$, for the constraints defining G results in

$$\begin{pmatrix} F(x^*) + \nabla g(x^*)^{\mathrm{T}} \lambda^* \\ -g(x^*) \end{pmatrix} + N_{X \times \Re^m_+}(x^*, \lambda^*) \ni 0^{n+m}. \tag{3}$$

This variational inequality in the variables (x, λ) describes the Karush–Kuhn–Tucker (KKT) conditions for (1); under the given Slater constraint qualification, the solutions in x coincide ([Tob86, HaP90]).

In the SQP algorithm for the solution of (3) the search-direction finding subproblem defined at a given iterate, $(x^t, \lambda^t) \in X \times \Re^m_+$, is the linearization of (3) at (x^t, λ^t): find $y^t \in X \cap T(x^t)$ such that

$$[F(x^t) + H(x^t, \lambda^t)(y^t - x^t)]^{\mathrm{T}}(z - y^t) \ge 0, \qquad \forall z \in X \cap T(x^t), \tag{4}$$

where

$$H(x, \lambda) := \nabla_x [F(x) + \nabla g(x)^{\mathrm{T}} \lambda] \tag{5}$$

is the Jacobian of the Lagrangean of (1) with respect to x, and

$$T(x) := \{\, y \in \Re^n \mid g(x) + \nabla g(x)(y - x) \le 0^m \,\}, \qquad x \in X \tag{6}$$

is the linearization of the set G defined at x.

A new iterate is usually defined by means of an inexact line search in the primal space, along the direction of $d^t := y^t - x^t$ in a penalty merit function for (1) [equivalently for (3)]. The penalty merit function utilized is the sum of a merit function for (1) and a penalty function for the set G. So, for nonlinear programming, which is obtained from (1) when $F \equiv \nabla f$ for some real-valued (objective) function $f : X \mapsto \Re$, the function most often considered is (e.g., [Han77, Fle87, Psh94, Ber95])

$$f(x) + r \sum_{i=1}^{m} \max\{0, g_i(x)\}, \qquad x \in X,$$

where r is a positive penalty parameter, while in applications to variational inequality problems, the objective function f must be replaced by a merit (or, gap) function for (1). Taji and Fukushima [TaF95] use the merit function

$$\psi_{\mathrm{TF}}(x) := \underset{y \in X}{\text{maximum}} \quad F(x)^{\mathrm{T}}(x - y) - (1/2)(y - x)^{\mathrm{T}} Q (y - x), \tag{7a}$$

$$\text{subject to} \quad g(x) + \nabla g(x)(y - x) \le 0^m, \tag{7b}$$

defined on X, where $Q \in \Re^{n \times n}$ is a symmetric and positive definite matrix.

Common to the existing analyses in applications of SQP algorithms in nonlinear programming and to variational inequality problems is that the value of the penalty parameter r needs to be sufficiently large in order for the penalty merit function to be exact, and (conversely) for the SQP subproblem to yield descent at every non-stationary point. The value of r is related to predictions on some norm of a vector λ^* of Lagrange multipliers for the explicit constraints; for example, the following condition is present in [Han77, Psh94, TaF95]:

$$r \ge \sup_t \{\, \|\lambda\|_\infty \mid \lambda \in \Lambda(x^t) \,\}, \tag{8}$$

where $\Lambda(x)$ denotes the set of Lagrange multipliers for the linearized constraints (6) in the solution to (4) defined at x, and $\|\cdot\|_\infty$ denotes the maximum norm.

In the case of variational inequality problems, the analysis in [TaF95] introduces the further requirement on the problem that F is strongly monotone (recall that F is strongly monotone on X if, for some $m_F > 0$, $[F(x) - F(y)]^{\mathrm{T}}(x - y) \ge m_F \|x - y\|^2$ holds for all $x, y \in X$), and the possible choices of matrices Q in the merit function (7) for (1) are restricted to those for which $\|Q\| < 2m_F$ holds.

Since an estimate of neither the size of a multiplier vector λ^* nor the modulus m_F of strong monotonicity is available a priori, estimates must be constructed either through experiments or adaptively in the algorithm.

Our approach to solving the problem (3) introduces some novelties in SQP methods. First, we introduce a merit (or, gap) function for (1) which is based on the same linearization of the constraints defining the set G as the SQP subproblem. This merit function extends the primal gap function introduced by Marcotte and Zhu [MaZ95] to the primal–dual setting, as well as the merit function (7) utilized in [TaF95, TaF96] in descent algorithms for (3). The analysis of this merit function (which differs substantially from that in [TaF95, TaF96] and from previous analyses of exact penalty functions) allows for the first time the penalty parameter r to take on any non-negative value. The explanation for this possibility is that the merit function introduced has an inherent strive for feasibility (which, for example, the function f in the case of nonlinear programming has not).

Another novelty is the introduction of a scaling parameter in the merit function. (This was earlier proposed in [MaZ95] in primal applications, but it is here developed further and put to an effective use in the primal–dual setting.) This scaling parameter serves the purpose of adaptively adjusting the merit function so that at any point outside the solution set $\mathrm{SOL}(F, X \cap G)$ of (3), the SQP subproblem solution eventually yields descent with respect to the adjusted merit function. In so doing, we remove the condition on the a priori choice of the merit function present in the analysis in [TaF95]. In particular, the scaling enables us to execute the algorithm without the knowledge (or estimation) of any problem parameters. (In the case where F is strongly monotone, the rescaling will only be made a finite number of times, resulting in an exact penalty merit function satisfying the above mentioned requirement on the merit function in [TaF95, TaF96]; cf. Corollary 3.4.) We stress that the scaling parameter does not have an interpretation as a penalty parameter, since (in the strongly monotone case) its final value mentioned above is only related to the problem data in the sense of the modulus m_F of strong monotonicity of F but not to any optimal vector λ^* of Lagrange multipliers for the explicit constraints.

The new merit function and the new analysis allows us to establish convergence of the SQP algorithm for non-strictly monotone operators F. We note however that, as in [TaF95], the SQP subproblem has an unorthodox feature, in utilizing as a multiplier estimate λ in the Jacobian mapping (5) of the SQP subproblem (4) the multiplier solution that is supplied in the evaluation of the merit function, in place of a multiplier estimate obtained from a previous solution of (4).

The rest of the paper is organized as follows. In the next section, we introduce a class of merit functions for (1), and study its most important properties. In Section 3, we introduce the iterative algorithm, which includes an adaptive SQP algorithm as a special case. The algorithm is analyzed in Section 4, and Section 5 compares the analysis to the ones in [TaF95, TaF96] for a projection algorithm and an SQP algorithm for (3). The paper concludes with some observations and remarks.

2 A merit function

The class of merit functions used defines an extension to primal–dual variational inequality problems of the differentiable gap functions for primal variational inequality problems which have a long history (see, e.g., [Auc89, Fuk92, Pat93a, Pat93b, WFM93, ZhM93, LaP94, Pat94, ZhM94, Fuk96, MaZ95, Pat97, Pat98]).

Let $\varphi : X \times X \mapsto \Re^n$ be a C^1 function on $X \times X$ of the form $\varphi(y, x)$, strictly convex in y for each fixed $x \in X$, and with the additional property that $\varphi(x, x) = 0$ and $\nabla_y \varphi(x, x) = 0^n$ for all $x \in X$. Let $\alpha > 0$, and consider the function

$$\psi_\alpha(x) := \underset{y \in X}{\text{maximum}} \quad L_\alpha(y, x) := F(x)^{\mathrm{T}}(x - y) - (1/\alpha)\varphi(y, x), \tag{9a}$$

$$\text{subject to} \quad g(x) + \nabla g(x)(y - x) \leq 0^m, \tag{9b}$$

defined on X.

We first establish a characterization of $\mathrm{SOL}(F, X \cap G)$ in terms of the fixed points of the problem (9).

Proposition 2.1 (A fixed point characterization of $\mathrm{SOL}(F, X \cap G)$). *$x \in \mathrm{SOL}(F, X \cap G)$ if and only if x solves* (9).

Proof. Let $x \in \mathrm{SOL}(F, X \cap G)$. Then, it is easily seen that $x \in \mathrm{SOL}(F, X \cap T(x))$. Let y solve (9). Then, $[F(x) + (1/\alpha)\nabla_y \varphi(y, x)]^{\mathrm{T}}(z - y) \geq 0$ holds for all $z \in X \cap T(x)$. Since $y \in X \cap T(x)$, $F(x)^{\mathrm{T}}(y - x) \geq 0$ holds, which implies that $(1/\alpha)\nabla_y \varphi(y, x)^{\mathrm{T}}(x - y) \geq 0$. But the strict convexity of $\varphi(\cdot, x)$ then implies that $y = x$ must hold, since otherwise we reach a contradiction from the strict monotonicity of $\nabla_y \varphi(\cdot, x)$ and the assumptions on φ.

Conversely, assume that x solves (9). Then, $x \in X \cap G$ must hold, and by its variational inequality characterization,

$$F(x)^{\mathrm{T}}(z - x) \geq 0, \qquad \forall z \in X \cap T(x) \supseteq X \cap G, \tag{10}$$

and hence, $x \in \mathrm{SOL}(F, X \cap G)$. □

Proposition 2.2 (ψ_α is a merit function for (1)). *$\psi_\alpha(x) \geq 0$ holds for all $x \in X \cap G$. Further, $x \in X \cap G$ and $\psi_\alpha(x) = 0$ hold if and only if $x \in \mathrm{SOL}(F, X \cap G)$.*

Proof. Let $x \in X \cap G$. It is clear from the choice $y = x$ in (9) that $\psi_\alpha(x) \geq 0$ holds.

[$\Longrightarrow$] Let $x \in X \cap G$ and $\psi_\alpha(x) = 0$ hold. That $x \in X \cap T(x)$ implies that $[F(x) + (1/\alpha)\nabla_y \varphi(y, x)]^{\mathrm{T}}(x - y) \geq 0$. But

$$\begin{aligned} 0 = \psi_\alpha(x) &= F(x)^{\mathrm{T}}(x - y) - (1/\alpha)\varphi(y, x) \\ &\geq -(1/\alpha)[\nabla_y \varphi(y, x)^{\mathrm{T}}(x - y) + \varphi(y, x)] \geq (1/\alpha)\varphi(x, x) = 0, \end{aligned}$$

where the last inequality is strict by the strict convexity of $\varphi(\cdot, x)$, unless $y = x$ holds. Proposition 2.1 then implies that $x \in \mathrm{SOL}(F, X \cap G)$.

[$\Longleftarrow$] Let x solve (9). Then, $\psi_\alpha(x) = 0$ immediately follows. Further, since $y = x$ holds, (9b) implies that $x \in X \cap G$. $\square$

We next investigate the continuity and differentiability of the merit function ψ_α.

PROPOSITION 2.3 (Continuity of ψ_α). *The merit function ψ_α is continuous on X.*

PROOF. The mapping $x \mapsto T(x)$ defined in (6) is continuous on X ([TaF96, Prop. 2.5]), and ψ_α is finite on X, whence the result follows from the continuity and boundedness of L_α on $X \times X$. $\square$

The analysis of the differentiability properties of ψ_α requires the explicit use of the KKT conditions for (9), which, given $x \in X$, are:

$$F(x) + (1/\alpha)\nabla_y \varphi(y, x) + \nabla g(x)^{\mathrm{T}} p + N_X(y) \ni 0^n, \tag{11a}$$

$$p \geq 0^m, \tag{11b}$$

$$p^{\mathrm{T}}[g(x) + \nabla g(x)(y - x)] = 0, \tag{11c}$$

$$g(x) + \nabla g(x)(y - x) \leq 0^m. \tag{11d}$$

We let $P(x)$ denote the set of solutions in p to (11), and further introduce the notation

$$H(x, p) := \nabla_x [F(x) + \nabla g(x)^{\mathrm{T}} p], \qquad p \in P(x). \tag{12}$$

PROPOSITION 2.4 (Directional differentiability of ψ_α). *The merit function ψ_α is directionally differentiable on X in any direction $d \in \Re^n$, with*

$$\psi'_\alpha(x; d) = \min_{p \in P(x)} \{[F(x) - H(x, p)(y - x) - (1/\alpha)\nabla_x \varphi(y, x)]^{\mathrm{T}} d\},$$

where y is the unique solution to (9).

PROOF. The result follows from [Hog73, Thm. 2]. $\square$

Although the merit function ψ_α has an inherent strive for feasibility with respect to G, we consider augmenting ψ_α with a penalty function, as follows:

$$\psi_{\alpha,r}(x) := \psi_\alpha(x) + r \sum_{i=1}^{m} \max\{0, g_i(x)\}, \qquad r \geq 0, \qquad x \in X.$$

Clearly, $\psi_{\alpha,r}$ is a continuous merit function for (1). The following result establishes that the directional derivative of $\psi_{\alpha,r}$ in the direction of $y - x$ is strictly monotonically decreasing in $r \geq 0$. While $r = 0$ will be shown to be sufficient for establishing the validity of the algorithm, a positive value of r may still be beneficial, through the addition of another force directed towards feasibility.

Proposition 2.5 ($\psi'_{\alpha,r}(x; y-x)$ is strictly monotonically decreasing in $r \geq 0$). *It holds that $\psi'_{\alpha,r}(x; y-x)$ is strictly monotonically decreasing in r. In particular, then,*

$$\psi'_{\alpha,r}(x; y-x) \leq \psi'_{\alpha}(x; y-x), \qquad r \geq 0. \tag{13}$$

Proof. The additional terms in the expression of $\psi'_{\alpha,r}(x; d)$ are

$$r \sum_{i \in \mathcal{I}_+(x)} \nabla g_i(x)^{\mathrm{T}} d + r \sum_{i \in \mathcal{I}_0(x)} \max\{0, \nabla g_i(x)^{\mathrm{T}} d\}, \tag{14}$$

where

$$\mathcal{I}_+(x) := \{\, i = 1, \ldots, m \mid g_i(x) > 0 \,\}; \qquad \mathcal{I}_0(x) := \{\, i = 1, \ldots, m \mid g_i(x) = 0 \,\}.$$

By (11d), $\nabla g_i(x)^{\mathrm{T}} d \leq -g_i(x) = 0$, $i \in \mathcal{I}_0(x)$, so the second term in (14) is zero. Further, by the same relation, $\nabla g_i(x)^{\mathrm{T}} d \leq -g_i(x) < 0$, $i \in \mathcal{I}_+(x)$, so the first term in (14) is non-positive, and strictly monotonically decreasing in r. The result follows. □

3 A modified descent algorithm for variational inequalities

We introduce a class of search direction subproblems in the class of cost approximation methods ([Pat93b]) for the solution (3) which extend the search direction finding subproblems of the algorithms of [Pat93a, Pat93b, WFM93, ZhM93, Pat94, ZhM94, MaZ95] to the explicit treatment of constraints and to the present primal–dual setting.

For any $\alpha > 0$, let $\Phi_\alpha : X \times X \times \Re^m_+ \mapsto \Re^n$ be a continuous mapping on $X \times X \times \Re^m_+$ of the form $\Phi_\alpha(y, x, p)$, monotone and in C^1 on X with respect to y. Given $(x, p) \in X \times \Re^m_+$, we consider replacing F by $\Phi_\alpha(\cdot, x, p)$; the error made is subtracted off through a fixed error term. The search-direction finding subproblem thus is to find $y \in X \cap T(x)$ such that

$$[\Phi_\alpha(y, x, p) + F(x) - \Phi_\alpha(x, x, p)]^{\mathrm{T}}(z - y) \geq 0, \qquad \forall z \in X \cap T(x). \tag{15}$$

If

$$\Phi_\alpha(\cdot, x, p) = (1/\alpha)\nabla_y \varphi(\cdot, x), \qquad x \in X, \tag{16}$$

then the problems (9) and (15) coincide, whence the search direction is defined by the point y that defines the current value of the merit function; the possible choice of Φ_α is however more general. In our convergence analysis, we will assume that either (16) holds, or the choice of Φ_α corresponds to an (approximate) Newton subproblem; the latter includes the SQP subproblem as a special case.

We next establish the descent properties of the search directions obtained from the solution of (15). We then need the following definition, which extends the one in [MaZ95] to the primal–dual setting.

Definition 3.1 (γ-strong descent). *The vector $d \in \Re^n$ defines a γ-strong descent direction for the merit function $\psi_{\alpha,r}$ at x if*

$$\psi'_{\alpha,r}(x;d) \le -\gamma\psi_\alpha(x), \qquad \gamma \in (0,1). \tag{17}$$

The next result extends one from [MaZ95] to the primal–dual setting.

Proposition 3.2 (Descent properties). *Let $\alpha > 0$. Let either Φ_α be chosen such that*

$$\|\nabla_y \Phi_\alpha(y,x,p) - H(x,p)\| \le \delta(\alpha), \qquad x, y \in Y, \quad p \in P(x), \tag{18}$$

for some function δ depending on problem data only, or (16) hold. Let $x \in X \setminus \mathrm{SOL}(F, X \cap G)$, y solve (15), and $d := y - x$. Then, either d defines a γ-strong descent direction for $\psi_{\alpha,r}$ at x, for any $r \ge 0$, or

$$\psi_\alpha(x) \le 1/(1-\gamma)[\delta(\alpha) + 1/\alpha]b, \tag{19}$$

for some $b > 0$.

Proof. We consider the first case for Φ_α. We first note that the direction $d := y - x$ also satisfies the conclusion (13) of Proposition 2.5, arguments for which are very similar, based on the KKT conditions for (15). Let w denote the solution to (9). Then, by Proposition 2.4,

$$\begin{aligned}
\psi'_{\alpha,r}(x;d) &\le [F(x) - H(x,p)(w-x) - (1/\alpha)\nabla_x\varphi(w,x)]^{\mathrm{T}} d \\
&\le -\psi_\alpha(x) + F(x)^{\mathrm{T}}(x-w) - (1/\alpha)\varphi(w,x) \\
&\quad + [F(x) - H(x,p)(w-x) - (1/\alpha)\nabla_x\varphi(w,x)]^{\mathrm{T}} d \\
&\quad + [F(x) + \Phi_\alpha(y,x,p) - \Phi_\alpha(x,x,p)]^{\mathrm{T}}(w-y) \\
&= -\psi_\alpha(x) - d^{\mathrm{T}} H(x,p) d \\
&\quad + [\Phi_\alpha(y,x,p) - \Phi_\alpha(x,x,p) - H(x,p)d]^{\mathrm{T}}(w-y) \\
&\quad - (1/\alpha)[\varphi(w,x) + \nabla_x\varphi(w,x)^{\mathrm{T}} d].
\end{aligned}$$

The second row in the last equality is written as

$$[\Phi_\alpha(y,x,p) - \Phi_\alpha(x,x,p) - H(x,p)d]^{\mathrm{T}}([x-y] + [w-x]);$$

the first term is less than $d^{\mathrm{T}}H(x,p)d$, by the monotonicity of $\Phi_\alpha(\cdot,x,p)$, while the second term can be shown to be less than $(1/2)\delta(\alpha)[\|d\|^2 + \|w-x\|^2]$ by the assumption (18). We conclude that, with

$$\theta(x,\alpha) := (1/2)\delta(\alpha)[\|d\|^2 + \|w-x\|^2] - (1/\alpha)[\varphi(w,x) + \nabla_x\varphi(w,x)^{\mathrm{T}} d], \tag{20}$$

for some $b > 0$,

$$\psi'_{\alpha,r}(x;d) \le -\psi_\alpha(x) + \theta(x,\alpha) \le -\psi_\alpha(x) + [\delta(\alpha) + 1/\alpha]b$$

holds, since X is bounded and φ and $\nabla_x\varphi$ are continuous. If $x \notin \mathrm{SOL}(F, X \cap G)$, then it follows that either d defines a γ-strong descent direction for $\psi_{\alpha,r}$ at x or that (19) holds.

In the second case for Φ_α, $w = y$ holds; the second row thus cancels out, and the conclusion holds with $\delta(\alpha) = 0$. □

INSTANCE 3.3 (An SQP algorithm). Let $\Phi_\alpha(y, x, p) := H(x, p)y$, $p \in P(x)$. Then, the CA subproblem is to find a $y \in X \cap T(x)$ such that

$$[F(x) + H(x, p)(y - x)]^{\mathrm{T}} (z - y) \geq 0, \qquad \forall z \in X \cap T(x). \tag{21}$$

This is the subproblem of an SQP algorithm in which a subset of the constraints are linearized, and where the multiplier estimate p is taken from the calculation of $\psi_\alpha(x)$.

We note that in an SQP algorithm that uses $\psi_{\alpha,r}$ as a merit function, the above result is applicable to monotone mappings F and for arbitrary choices of the penalty parameter $r \geq 0$, because, by construction, (18) holds with $\delta(\alpha) = 0$. □

COROLLARY 3.4 (Descent under strong monotonicity). *In addition to the assumptions of Proposition* 3.2*, assume that F is strongly monotone on X. Let either the search direction d be given by the SQP subproblem described above or let* (16) *hold.*

(a) *For every $x \in X \setminus \mathrm{SOL}(F, X \cap G)$, there exists an $\overline{\alpha} > 0$ such that*

$$\alpha \geq \overline{\alpha} \qquad \Longrightarrow \qquad \psi'_{\alpha,r}(x; d) \leq -\psi_\alpha(x).$$

(b) *Let further $\varphi(y, x) := (1/2)(y - x)^{\mathrm{T}} Q(y - x)$, for some symmetric and positive definite matrix Q. Then, for every $x \in X \setminus \mathrm{SOL}(F, X \cap G)$,*

$$\alpha \geq \|Q\|/(2m_F) \qquad \Longrightarrow \qquad \psi'_{\alpha,r}(x; d) \leq -\psi_\alpha(x).$$

PROOF.

(a) The proof of Proposition 3.2 shows that

$$\psi'_{\alpha,r}(x; d) \leq -\psi_\alpha(x) - d^{\mathrm{T}} \nabla F(x) d - (1/\alpha)[\varphi(y, x) + \nabla_x \varphi(y, x)^{\mathrm{T}} d],$$

whence the result follows by the boundedness of X, and the continuity of φ and $\nabla_x\varphi$.

(b) The above estimate reduces to

$$\psi'_{\alpha,r}(x; d) \leq -\psi_\alpha(x) - d^{\mathrm{T}}[\nabla F(x) - 1/(2\alpha)Q]d,$$

and the result follows. □

The result establishes that the merit function constructed is exact for large enough values of α whenever F is strongly monotone, in the sense that

$$\psi'_{\alpha,r}(x;d) \geq 0 \text{ and } x \in X \cap G \qquad \Longleftrightarrow \qquad x \in \mathrm{SOL}(F, X \cap G).$$

The requirement in (b) on the value of α (albeit with strict inequality) is present in the convergence conditions for the SQP algorithm of Taji and Fukushima [TaF95] based on the same merit function, but their convergence analysis is different, and requires in addition that the value of r fulfills a condition of the form (8).

We are now ready to state the algorithm; it is given in Table 1.

TABLE 1: A modified descent algorithm

---

0. (*Initialization*): Choose an initial point $x^0 \in X$, let $\alpha_0 > 0$, $\Delta\alpha > 0$, $\gamma \in (0,1)$, $\beta \in (0,1)$, $\eta \in (0,1)$, $s \in (0,1]$, and $r \geq 0$. Set $t := 0$.
1. (*Search direction generation*): Let

$$\psi_{\alpha_t}(x^t) := \underset{y \in X \cap T(x^t)}{\text{maximum}} \quad L_{\alpha_t}(y, x^t) := F(x^t)^{\mathrm{T}}(x^t - y) - (1/\alpha_t)\varphi(y, x^t). \quad (22)$$

   If the second alternative for Φ_α is chosen, then let y^t be the solution to (22), otherwise, let $y^t \in X \cap T(x^t)$ be a vector such that

$$[\Phi_{\alpha_t}(y^t, x^t, p^t) + F(x^t) - \Phi_{\alpha_t}(x^t, x^t, p^t)]^{\mathrm{T}}(z - y^t) \geq 0, \quad \forall z \in X \cap T(x^t), \quad (23)$$

   where $p^t \in P(x^t)$ is arbitrary.
   The resulting search direction is $d^t := y^t - x^t$.
2. (*Termination criterion*): If x^t solves (23) or (22) $\rightarrow$ Stop [(x^t, p^t) solves (3)]. Otherwise, continue.
3. (*Armijo line search or null step*): If $\psi_{\alpha_t}(x^t) \leq \theta(x^t, \alpha_t)/(1-\gamma)$, where $\theta(x^t, \alpha_t)$ is defined by (20), then let $\alpha_{t+1} := \alpha_t + \Delta\alpha$ and $x^{t+1} := x^t$; otherwise, let $\alpha_{t+1} := \alpha_t$ and $x^{t+1} := x^t + \ell_t d^t$, where $\ell_t := s\beta^i$ for the smallest non-negative integer i such that

$$\psi_{\alpha_t,r}(x^t + s\beta^i d^t) - \psi_{\alpha_t,r}(x^t) \leq \eta s \beta^i \psi'_{\alpha_t,r}(x^t; d^t). \quad (24)$$

4. (*Termination criterion*): If x^t is acceptable $\rightarrow$ Stop. Otherwise, go to Step 1 with $t := t+1$.

REMARK 3.5 (Observations on the algorithm). (a) The validity of the termination criterion in Step 2 is clear from the relation (10) [the case (23)] and Proposition 2.1 [the case (22)].

(b) The rationale for the updating of α_t is clear from the proof of Proposition 3.2; if either (18) holds for a function δ with the property that $\{\delta(\alpha_t)\} \to 0$ when $\{\alpha_t\} \to \infty$ or if (16) holds, then the last equality in that proof shows that whenever x^t is not a solution to (1), by increasing the value of α_t, the directional derivative $\psi'_{\alpha_t,r}(x^t; d^t)$ will eventually be negative.

(c) The updating rule for α_t can also be motivated from the point of view of obtaining stationary points to the correct merit function: if a given iterate x^t is a stationary point for the current merit function used but x^t does not solve VIP$(F, X \cap G)$, then the updating rule is so designed that it adjusts the merit function so that eventually x^t is no longer stationary.

(4) If iteration $t-1$ included an Armijo line search, then, at iteration t, the solution to (22), the value of p^t used in (23), and the value of $\psi_{\alpha_t,r}(x^t)$ all are supplied automatically from the use of the Armijo rule at iteration $t-1$. □

4 Convergence analysis

Before we turn to the convergence of the proposed algorithm, we need the following technical lemmas.

Lemma 4.1 (Boundedness of $P(x)$). *For each $x \in X$, the set $P(x)$ is nonempty, closed and bounded.*

Proof. The first two properties follow from the Slater CQ and (11). Let y be the solution to (9b), $p \in P(x)$, and $\nu \in N_X(y)$ satisfy (11a). Let $\overline{x} \in X$ be such that $g_i(\overline{x}) < 0$ for all $i = 1, \dots, m$. Then,

$$\begin{aligned}
0 &\le [F(x) + (1/\alpha)\nabla_y \varphi(y,x) + \nabla g(x)^{\mathrm{T}} p + \nu]^{\mathrm{T}}(\overline{x} - y) \\
&\le [F(x) + (1/\alpha)\nabla_y \varphi(y,x)]^{\mathrm{T}}(\overline{x} - y) \\
&\quad + \sum_{i=1}^{m} p_i [\nabla g_i(x)^{\mathrm{T}}(x - y) + g_i(\overline{x}) - g_i(x)] \\
&= [F(x) + (1/\alpha)\nabla_y \varphi(y,x)]^{\mathrm{T}}(\overline{x} - y) + \sum_{i=1}^{m} p_i g_i(\overline{x}),
\end{aligned}$$

where we have used (11b) and (11c), the definition (2) of the normal cone to X at y, and the convexity of g_i, $i = 1, \dots, m$. It therefore follows that

$$0 \le \sum_{i=1}^{m} p_i \le \frac{[F(x) + (1/\alpha)\nabla_y \varphi(y,x)]^{\mathrm{T}}(\overline{x} - y)}{\min_{i \in \{1,\dots,m\}} \{-g_i(\overline{x})\}},$$

and the result follows. □

Lemma 4.2 (Continuity of $\psi_\cdot(\cdot)$). *Assume that $X \supset \{x^t\} \to x \in X$ and $\Re_{++} \supset \{\alpha_t\} \to +\infty$. Then,*

$$\{\psi_{\alpha_t}(x^t)\} \to \psi_\infty(x) := \underset{y \in X \cap T(x)}{\text{maximum}} \; F(x)^{\mathrm{T}}(x-y).$$

Proof. Let $\varepsilon > 0$ be arbitrary. First, we note that ψ_∞ is continuous on X (see the proof of Proposition 2.3), and thus

$$|\psi_\infty(x^t) - \psi_\infty(x)| \leq \varepsilon/2, \qquad t \geq \tau_1.$$

Second, $\psi_\cdot(x)$ is continuous as a function of α for any fixed value of $x \in X$ (by arguments similar to those for $\psi_\alpha(\cdot)$ in Proposition 2.3, and hence,

$$|\psi_{\alpha_t}(x) - \psi_\infty(x)| \leq \varepsilon/2, \qquad t \geq \tau_2.$$

By the above two inequalities,

$$|\psi_{\alpha_t}(x^t) - \psi_\infty(x)| \leq \varepsilon, \qquad t \geq \max\{\tau_1, \tau_2\},$$

and the result follows. □

Theorem 4.3 (Global convergence). *In addition to the assumptions of Proposition 3.2, let the function δ be such that $\{\delta(\alpha)\} \to 0$ when $\alpha \to \infty$. Let $\{(x^t, p^t)\}$ be a sequence of primal and dual iterates generated by the algorithm of Table 1. Then, either the sequence $\{(x^t, p^t)\}$ is finite, whence the last iterate (x^t, p^t) is a solution to (3), or the sequence $\{(x^t, p^t)\}$ is infinite, whence it is bounded, any accumulation point is a solution to (3), and*

$$\left\{ \underset{(x,\lambda) \text{ solves } (3)}{\text{minimum}} \; \|(x^t, p^t) - (x, \lambda)\| \right\} \to 0. \tag{25}$$

Proof. The conclusion in the case of finite termination is clear from the construction of (23) and (22).

Clearly, the sequence $\{x^t\}$ is bounded, since X is bounded. The same is true for the sequence $\{p^t\}$, by the boundedness of $\{x^t\}$ and by Lemma 4.1.

We next distinguish two cases.

($\{\alpha_t\} \to \infty$) Let x^∞ be any accumulation point of $\{x^t\}$, corresponding to some subsequence $\mathcal{T}$. This assumption, together with the proof of Proposition 3.2, establish that then, $\psi_{\alpha_t}(x^t) \leq [\delta(\alpha_t) + 1/\alpha_t]b/(1-\gamma)$ must hold for infinitely many $t \in \mathcal{T}$. Since $\{\delta(\alpha_t)\} \to 0$ holds by Step 3, $\{\psi_{\alpha_t}(x^t)\}_{t \in \mathcal{T}} \to 0$. Lemma 4.2 establishes that $\psi_\infty(x^\infty) = 0$. In a manner similar to that in the proof of Proposition 2.2, we conclude that x^∞ solves (9) at x^∞ for $\alpha = +\infty$, and that $x^\infty \in G$ therefore holds. Hence,

$x^\infty \in \mathrm{SOL}(F, X \cap G)$ follows. Turning to the proof for the sequence $\{p^t\}$, let p^∞ be any of its accumulation points. The fact that x^∞ solves (9) at x^∞ for $\alpha = +\infty$ then implies that (11) is satisfied with $x = y = x^\infty$, $p = p^\infty$ and $\alpha = +\infty$, which implies that (x^∞, p^∞) is a solution to (3).

($\alpha_t = \overline{\alpha}$, $t \geq \tau$) By the continuity of $\psi_{\overline{\alpha},r}$, the Armijo rule (24) is well-defined. We will show that $\{d^t\} \to 0^n$. The proof proceeds by contradiction. If $\{d^t\} \not\to 0^n$, then as $\{\psi_{\overline{\alpha},r}(x^{t+1}) - \psi_{\overline{\alpha},r}(x^t)\} \to 0$ must hold, there must be a subsequence $\mathcal{T}$ such that $\{\ell_t\}_{t \in \mathcal{T}} \to 0$ holds. There must then be a $\overline{\tau} \geq \tau$ such that

$$\psi_{\overline{\alpha},r}(x^t + [\ell_t/\beta]d^t) - \psi_{\overline{\alpha},r}(x^t) > -\eta[\ell_t/\beta]\psi'_{\overline{\alpha},r}(x^t; d^t), \qquad t \geq \overline{\tau}, \quad t \in \mathcal{T}. \tag{26}$$

Let d^∞ be an accumulation point of $\{d^t\}_{t \in \mathcal{T}}$; the existence of an accumulation point follows since $\{d^t\}$ is bounded. Dividing (26) by ℓ_t/β and taking the limit of $\mathcal{T}$, we obtain by the directional differentiability of $\psi_{\overline{\alpha},r}$ that $\psi'_{\overline{\alpha},r}(x^\infty; d^\infty) \geq 0$. In the limit of $\mathcal{T}$ of the inequality (17), we obtain that $\psi'_{\overline{\alpha},r}(x^\infty; d^\infty) \leq -\gamma\psi_{\overline{\alpha},r}(x^\infty)$. If the right-hand side of this inequality is zero, then it follows that x^∞ solves (9) at x^∞, and therefore $x^\infty \in \mathrm{SOL}(F, X \cap G)$ holds. If $\psi_{\overline{\alpha},r}(x^\infty) > 0$ holds, then we have reached a contradiction. In this case, we conclude that $\{d^t\} \to 0^n$, and we again conclude that $x^\infty \in \mathrm{SOL}(F, X \cap G)$ holds, this time by Proposition 2.1. Turning to the proof for the sequence $\{p^t\}$, we let p^∞ be any accumulation point thereof, and note that $\{d^t\} \to 0^n$ implies from (11), together with the fact that $\nabla_y \varphi(x^\infty, x^\infty) = 0^n$, that (x^∞, p^∞) is a solution to (3). □

Remark 4.4 (On step length selection rules). We observe that the proof establishes that any step length rule that provides a decrease in $\psi_{\alpha_t,r}$ that is at least as large as that of the Armijo rule (24) is a valid choice. This facilitates a plentiful of alternatives for more efficient choices of step length rules. For example, the line search can be replaced by a more general multi-dimensional search along the lines of simplicial decomposition methods; this approach is analyzed further in [Pat98, Chapter 9]. □

The following result shows that the modified descent algorithm finitely turns into a descent approach when F is strongly monotone. [The SQP method under case (b) may be viewed as an adaptive version of the SQP method of [TaF95], albeit under milder convergence conditions.]

Corollary 4.5 (Finite reduction to a descent algorithm). *In addition to the assumptions of Theorem* 4.3, *assume that F is strongly monotone on X. Let either* (16) *hold or* (18) *hold with $\delta(\alpha) = 0$.*

(a) *There exists a finite integer τ such that $\alpha_t = \overline{\alpha}$, for all $t \geq \tau$.*

(b) *Assume further that $\varphi(y, x) := (1/2)(y - x)^{\mathrm{T}}Q(y - x)$, for some symmetric and positive definite matrix Q. Then, $\overline{\alpha} \leq \max\{\alpha_0, \|Q\|/(2m_F) + \Delta\alpha\}$.*

PROOF. The result follows from Theorem 4.3 and Corollary 3.4. ◻

The following result validates a more active involvement of the penalty parameter r in the algorithm.

COROLLARY 4.6 (Adaptive penalty parameter). *Let the assumptions of Theorem* 4.3 *hold. In the algorithm of Table* 1, *consider allowing the parameter r to be updated arbitrarily within a bounded interval* $0 \leq r \leq \overline{r}$, *whenever a null step is taken. Then, the conclusion of Theorem* 4.3 *holds for this modified algorithm.*

PROOF. In the first case ($\{\alpha_t\} \to \infty$), the value of r is immaterial to the proof. In the second case ($\alpha_t = \overline{\alpha}$, $t \geq \tau$), the value of r is not updated after iteration τ, whence the proof reduces to the original one. ◻

REMARK 4.7 (On the convergence rate of the SQP method). The merit function ψ_α is differentiable at some point $x \in X$ whenever the set $P(x)$ of multipliers for (9b) is a singleton. (This condition holds if the normals $\nabla g_i(x)$ of the constraints active at y are linearly independent and strict complementarity holds there; see [FiM68, Thm. 6].) If this condition holds uniformly over $x \in X$, then it is possible to establish a linear convergence rate in the primal space of the algorithm whenever F is strongly monotone, along the lines of existing analyses of primal SQP methods (e.g., [Pat98]).

Under these assumptions, the SQP method based on the subproblem (4) is locally superlinearly convergent, and quadratically convergent under additional assumptions; see [GaM76]. It is not yet clear if such results are precluded for the algorithm of this paper by the fact that the multiplier estimate p^t is instead chosen in the set $P(x^t)$ associated with the merit function. As remarked in [TaF95], however, under the above-mentioned additional assumptions, the multiplier vector λ^* is unique, in which case the two sequences converge to the same point; if we can guarantee that the unit step is accepted by the Armijo rule close to the solution, then at least a superlinear convergence rate is achieved. ◻

5 Two applications

We discuss two applications of Theorem 4.3, to a projection algorithm and to an SQP algorithm.

5.1 A modified projection algorithm

Let $\varphi(y, x) := (1/2)(y-x)^{\mathrm{T}}Q(y-x)$, for some symmetric and positive definite matrix Q, and let (16) hold. Then, for any $r \geq 0$, Theorem 4.3 applies. The algorithm thus constructed is a modified projection algorithm, wherein the strictly convex, quadratic subproblems are solved over the set $X \cap T(x^t)$, and the matrix Q is adaptively scaled by the parameter α_t whenever the direction obtained from (22) does not yield descent.

A related projection algorithm is given by Taji and Fukushima [TaF96]. Their problem is obtained from (1) by letting $X := \Re^n$, and we identify their algorithm in this case by setting $\alpha_t := 1$ [cf. (7)]. Their convergence result relies on the monotone decrease in their merit function; in order to guarantee this descent property, the following assumptions must be introduced in their analysis:

1. $\nabla F(x)$ is positive definite on $\Re^n$; and
2. r is chosen so that $r \geq \sup_t \{ \|p\|_\infty \mid p \in P(x^t) \}$, where $P(x^t)$ is the set of multipliers for the linearized constraints (7b).

They also include the assumption that the generated sequence $\{x^t\}$ is bounded. (In the present paper, we have instead assumed that the ground set X is bounded.)

5.2 An SQP algorithm

Our second, and most interesting application, is to SQP methods for (1). Given $x^t \in X$, the SQP method proceeds as follows. We construct a prediction of the multiplier vector λ^* of (3) by solving (22) and extracting an arbitrary multiplier vector $p^t \in P(x^t)$. The search-direction finding subproblem of the SQP scheme then is to find $y^t \in X \cap T(x^t)$ such that (4) holds, where λ^t is replaced by p^t.

We then note that (4) is obtained from (23) by choosing $\Phi_\alpha(y, x, p) := H(x, p)(y - x)$, $p \in P(x)$. Further, (18) holds with $\delta(\alpha) = 0$, and therefore Theorem 4.3 applies. The analysis provided in Theorem 4.3 shows that this SQP method for (1) is globally convergent for non-strictly monotone problem mappings F, and further that the penalty parameter r can be chosen arbitrarily in $\Re_+$.

We also note that in the case where $F \equiv \nabla f$ for some real-valued function $f : X \mapsto \Re$, the above adaptive SQP method is applicable to the nonlinear programming problem whenever f is convex.

This analysis improves that of the SQP method in [TaF95]. The problem is obtained by letting $X := \Re^n$, and their algorithm by choosing $\alpha_t := 1$ and moreover the merit function as the projection-based measure (7). The progress of the algorithm relies on the descent properties of the merit function, which makes it necessary to introduce the following assumptions in the analysis:

1. F is strongly monotone on $\Re^n$;
2. r is chosen so that $r \geq \sup_t \{ \|p\|_\infty \mid p \in P(x^t) \}$; and
3. $\|Q\| < 2m_F$.

Convergence is established under the additional presumption that the sequence $\{x^t\}$ is bounded.

6 Observations and final remarks

The convergence analyses of [TaF95, TaF96] for the above two algorithms do not cover the convergence of the sequence $\{p^t\}$ of multiplier estimates.

The conditions 2. and 3. above are difficult to observe a priori. We note also again in passing that condition 2. is present in some analyses of SQP algorithms applied to nonlinear optimization problems (e.g., [Han77, Psh94]). Further, as is illustrated in Corollary 4.5, the requirement 3. present in the analysis of [TaF95] is not present here, since the function φ is automatically scaled in order to yield descent. By doing so, we gain the additional advantage of not requiring the knowledge or estimation of any problem parameters, instead fulfilling the requirement 3. eventually in the case of strongly monotone problems (cf. Corollary 3.4).

The present SQP algorithm and its convergence analysis extends and improves, respectively, that of primal Newton methods in [MaD87, TFI93], and primal–dual Newton methods in [MaD89].

Recent articles on iterative methods for variational inequality problems, for example [ZhM93, MaZ95, Kon97], and in particular on modified projection methods ([Ius94, SoT96, Sun96, He97]), show that one may reach convergence results under milder conditions when one introduces adaptivity. The above analysis serves as a further illustration that this route of development is worthy of further study also for other algorithms.

We finally note that in the context of nonlinear optimization, recent experience with modifications of SQP algorithms that do not utilize a penalty function, but rather a filter using a dominance concept adapted from multiobjective optimization, shows promise, albeit no convergence analysis for these methods is available at present ([Zop95, FlL97]). A convergence result is however available for the more basic approach of SLP in combination with trust region approaches ([FLT98]), which would suggest the possibility of applying more general non-penalty methods.

Acknowledgement

The presentation of this article benefitted from discussions with Professor Torbjörn Larsson.

References

[Auc89] G. Auchmuty (1989), "Variational principles for variational inequalities," *Numerical Functional Analysis and Optimization*, vol. 10, pp. 863–874.

[Ber95] D. P. Bertsekas. *Nonlinear Programming*, Athena Scientific, Belmont, MA, 1995.

[FiM68] A. V. Fiacco and G. McCormick (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley & Sons,, New York, NY. Also published as volume 4 in the Classics in Applied Mathematics Series, SIAM, Philadelphia, PA, 1990.

[Fle87] R. Fletcher (1987). *Practical Methods of Optimization*, John Wiley & Sons, Chichester, U.K., second ed.

[FlL97] R. Fletcher and S. Leyffer (1997), "Nonlinear programming without a penalty function," Numerical Analysis Report NA/171, Department of Mathematics, University of Dundee, Dundee, Scotland, U.K.

[FLT98] R. Fletcher, S. Leyffer and Ph. L. Toint (1998), "On the global convergence of an SLP-filter algorithm," Numerical Analysis Report NA/183, Department of Mathematics, University of Dundee, Dundee, Scotland, U.K.

[Fuk92] M. Fukushima (1992), "Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems," *Mathematical Programming*, vol. 53, pp. 99–110.

[Fuk96] M. Fukushima (1996), "Merit functions for variational inequality and complementarity problems," In: *Nonlinear Optimization and Applications, Proceedings of the International School of Mathematics "G. Stampacchia" 21st Workshop on Nonlinear Optimization and Applications*, held June 13–21, 1995, in Erice, Italy, G. Di Pillo and F. Giannessi (eds.), Plenum Press, New York, NY, pp. 155–170.

[GaM76] U. M. Garcia Palomares and O. L. Mangasarian (1976), "Superlinearly convergent quasi-Newton algorithms for nonlinearly constrained optimization problems," *Mathematical Programming*, vol. 11, pp. 1–13.

[Han77] S. P. Han (1977), "A globally convergent method for nonlinear programming," *Journal of Optimization Theory and Applications*, vol. 22, pp. 297–309.

[HaP90] P. T. Harker and J.-S. Pang (1990), "Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications," *Mathematical Programming*, vol. 48, pp. 161–220.

[He97] B. He (1997), "A class of projection and contraction methods for monotone variational inequalities," *Applied Mathematics & Optimization*, vol. 35, pp. 69–76.

[Hog73] W. W. Hogan (1973), "Directional derivatives for extremal-value functions with applications to the completely convex case," *Operations Research*, vol. 21, pp. 188–209.

[Ius94] A. I. Iusem (1994), "An iterative algorithm for the variational inequality problem," *Computational and Applied Mathematics*, vol. 13, pp. 103–114.

[Kon97] I. V. Konnov (1997), "A class of combined iterative methods for solving variational inequalities," *Journal of Optimization Theory and Applications*, vol. 94, pp. 677–693.

[LaP94] T. Larsson and M. Patriksson (1994), "A class of gap functions for variational inequalities," *Mathematical Programming*, vol. 64, pp. 53–79.

[MaD87] P. Marcotte and J.-P. Dussault (1987), "A note on a globally convergent Ne"wton method for solving monotone variational inequalities, *Operations Research Letters*, vol. 6, pp. 35–42.

[MaD89] P. Marcotte and J.-P. Dussault (1989), "A sequential linear programming algorithm for solving monotone variational inequalities," *SIAM Journal on Control and Optimization*, vol. 27, pp. 1260–1278.

[MaZ95] P. Marcotte and D. Zhu (1995), "Global convergence of descent processes for solving non strictly monotone variational inequalities," *Computational Optimization and Applications*, vol. 4, pp. 127–138.

[Pat93a] M. Patriksson (1993), "A unified description of iterative algorithms for traffic equilibria," *European Journal of Operational Research*, vol. 71, pp. 154–176.

[Pat93b] M. Patriksson (1993), *A unified framework of descent algorithms for nonlinear programs and variational inequalities*, PhD thesis, Department of Mathematics, Linköping Institute of Technology, Linköping, Sweden.

[Pat94] M. Patriksson (1994), "On the convergence of descent methods for monotone variational inequalities," *Operations Research Letters*, vol. 16, pp. 265–269.

[Pat97] M. Patriksson (1997), "Merit functions and descent algorithms for a class of variational inequality problems," *Optimization*, vol. 41, pp. 37–55.

[Pat98] M. Patriksson (1998). *Nonlinear Programming and Variational Inequality Problems: A Unified Approach*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

[Psh94] B. N. Pshenichnyj (1994). *The Linearization Method for Constrained Optimization*, vol. 22 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin. Translated from the Russian by S. S. Wilson.

[SoT96] M. V. Solodov and P. Tseng (1996), "Modified projection-type method for monotone variational inequalities," *SIAM Journal on Control and Optimization*, vol. 34, pp. 1814–1830.

[Sun96] D. Sun (1996), "A class of iterative methods for solving nonlinear projection equations," *Journal of Optimization Theory and Applications*, vol. 91, pp. 123–140.

[TaF95] K. Taji and M. Fukushima (1995), "A globally convergent Newton method for solving variational inequality problems with inequality constraints," in *Recent Advances in Nonsmooth Optimization*, D.-Z. Du, L. Qi, and R. S. Womersley, eds., World Scientific Publishing Co, Singapore, pp. 405–417.

[TaF96] K. Taji and M. Fukushima (1996), "A new merit function and a successive quadratic programming algorithm for variational inequality problems," *SIAM Journal on Optimization*, vol. 6, pp. 704–713.

[TFI93] K. Taji, M. Fukushima, and T. Ibaraki (1993), "A globally convergent Newton method for solving strongly monotone variational inequalities," *Mathematical Programming*, vol. 58, pp. 369–383.

[Tob86] R. L. Tobin (1986), "Sensitivity analysis for variational inequalities," *Journal of Optimization Theory and Applications*, vol. 48, pp. 191–204.

[WFM93] J. H. Wu, M. Florian, and P. Marcotte (1993), "A general descent framework for the monotone variational inequality problem," *Mathematical Programming*, vol. 61, pp. 281–300.

[ZhM93] D. L. Zhu and P. Marcotte (1993), "Modified descent methods for solving the monotone variational inequality problem," *Operations Research Letters*, vol. 14, pp. 111–120.

[ZhM94] D. L. Zhu and P. Marcotte (1994), "An extended descent framework for variational inequalities," *Journal of Optimization Theory and Applications*, vol. 80, pp. 349–366.

[Zop95] C. Zoppke-Donaldson (1995), *A tolerance-tube approach to sequential quadratic programming with applications*, PhD thesis, Department of Mathematics and Computer Science, University of Dundee, Dundee, Scotland, U.K.