# A Globally Convergent Newton Method for Solving Variational Inequality Problems with Inequality Constraints

Kouichi TAJI Masao FUKUSHIMA

September, 1994

**Abstract**

Newton method for variational inequality problems can be made globally convergent by incorporating a line search for a merit function. But the existing methods may not be easy to apply to problems with general nonlinear constraints, because the evaluation of a merit function requires the solution of an optimization problem with such nonlinear constraints. In this paper, we propose a new globally convergent Newton method for solving variational inequality problems with general inequality constraints. The method solves at each iteration an affine variational inequality problem, in which not only the mapping of the problem but also the constraints are linearized. In a line search, it makes use of the merit function recently proposed by the authors, which can be evaluated by solving a convex quadratic minimization problem with linear constraints. Thus each step of the algorithm can be carried out finitely even if the constraints are nonlinear. We show that, when the mapping is strongly monotone, the method is globally convergent to the solution, and that, under some additional assumptions, the rate of convergence is superlinear.

# 1 Introduction

We consider the variational inequality problem of finding $x^* \in S$ such that

$$\langle F(x^*), x - x^* \rangle \geq 0 \quad \text{for all} \quad x \in S, \tag{1}$$

where $S$ is a nonempty closed convex subset of $R^n$, $F$ is a continuously differentiable mapping from $R^n$ into $R^n$ and $\langle \cdot, \cdot \rangle$ denotes the inner product in $R^n$. In this paper, we suppose that the set $S$ is specified by

$$S = \{x \in R^n \mid c_i(x) \leq 0, \quad i = 1, \ldots, m\}, \tag{2}$$

where $c_i : R^n \to R$ are twice continuously differentiable convex functions.

Many iterative methods, such as Newton method, projection methods, the linearized Jacobi method and the successive over-relaxation methods, have been proposed to solve the variational inequality problem (1) (see [7] and the references cited therein). Among them, the Newton method generates a sequence $\{x^k\}$, where $x^{k+1}$ is a solution to the linearized variational inequality problem

$$\langle F(x^k) + \nabla F(x^k)^T (x^{k+1} - x^k), x - x^{k+1} \rangle \geq 0 \quad \text{for all} \quad x \in S. \tag{3}$$

It can be shown that, under suitable assumptions, the Newton method converges quadratically to a solution $x^*$, provided that an initial point $x^0$ is chosen sufficiently close to $x^*$ [7, Theorem 4.1].

For the variational inequality problem (1), various merit functions have been proposed and their properties have been studied [1, 2, 4, 8]. Those functions can be used to globalize the Newton method. Marcotte and Dussault [10] obtained a globally convergent Newton method by incorporating an exact line search strategy for the gap function

$$g(x) = \max_y \{\langle F(x), x - y \rangle \mid y \in S\},$$

introduced by Auslender [2]. Another modification is the one recently proposed by Taji, Fukushima and Ibaraki [12], which makes use of Armijo line search for the regularized gap function

$$\hat{f}(x) = \max_y \left\{ -\langle F(x), y - x \rangle - \frac{1}{2} \langle y - x, G(y - x) \rangle \;\middle|\; y \in S \right\},$$

introduced by Fukushima [4], where $G$ is an $n \times n$ symmetric positive definite matrix. Under suitable assumptions, both modifications are shown to be globally convergent to a solution with quadratic rate of convergence [10, 12].

Note that the above mentioned methods tacitly assume that the constraint set $S$ has a relatively simple structure. For example, when $S$ is a polyhedral convex set, that is, the functions $c_i$ are all affine, the variational inequality subproblem (3) of Newton method becomes an affine variational inequality problem and the gap functions $g$ and $\hat{f}$ can be evaluated by solving linear and quadratic programming problems, respectively. However, when $S$ is a general convex set defined by nonlinear convex functions, solving the linearized subproblem (3) and evaluating $g(x)$ and $\hat{f}(x)$ should be considered difficult tasks.

In this paper we propose a new globally convergent Newton method for solving variational inequality problems with general inequality constraints. The method solves at each iteration an affine variational inequality subproblem, in which not only the mapping $F$ but also the constraint functions $c_i$ are linearized. Moreover it makes use of the merit function recently proposed by Taji and Fukushima [13] to obtain global convergence. The method has a clear advantage over the methods that solve subproblems (3) and use the merit function $g$ or $\hat{f}$, in that each step of the algorithm can be carried out finitely even if the set $S$ is specified by nonlinear inequalities. It is shown that, when the mapping is strongly monotone, the method converges globally to the solution, and that, under some additional assumptions, the rate of convergence is superlinear. The method is closely related to a successive quadratic programming method for solving nonlinear programming problems.

# 2 Preliminaries

In this section, we summarize some preliminary facts which will be useful subsequently.

The mapping $F : R^n \to R^n$ is said to be *monotone* if

$$\langle F(x) - F(y), x - y \rangle \geq 0 \quad \text{for all } x, y \in R^n,$$

*strictly monotone* if the above inequality holds strictly whenever $x \neq y$, and *strongly monotone* with modulus $\mu > 0$ if

$$\langle F(x) - F(y), x - y \rangle \geq \mu \, \| x - y \|^2 \quad \text{for all } x, y \in R^n.$$

It is well known [11, Theorem 5.4.3] that, when $F$ is continuously differentiable, $F$ is strongly monotone with modulus $\mu$ if and only if the Jacobian $\nabla F(x)$ satisfies

$$\langle d, \nabla F(x)d\rangle \geq \mu \parallel d \parallel^2 \quad \text{for all } x, d \in R^n. \tag{4}$$

It is also well known [7, Corollary 3.2] that, when the mapping $F$ is strongly monotone, the variational inequality problem (1) admits a unique solution.

A function $\phi : R^n \to [-\infty, +\infty]$ is said to be *directionally differentiable* at $x$ in the direction $d$ if $\phi(x)$ is finite and the limit

$$\lim_{t\to 0, t>0} \frac{\phi(x+td) - \phi(x)}{t}$$

exists. We call the limit the *directional derivative* and denote it by $\phi'(x; d)$.

In the remainder of the paper, we suppose that the Slater's constraint qualification holds for (2), i.e., there exists an $\hat{x} \in R^n$ such that

$$c_i(\hat{x}) < 0 \quad \text{for all} \quad i = 1, \ldots, m. \tag{5}$$

Under this assumption, $x^*$ is a solution to (1) if and only if there exists a Lagrange multiplier vector $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$ such that $(x^*, \lambda^*)$ is a solution to the following mixed nonlinear complementarity problem [7, Proposition 2.2] :

$$\begin{aligned} &F(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) = 0, \\ &c_i(x^*) \leq 0,\ \lambda_i^* \geq 0,\ \lambda_i^* c_i(x^*) = 0,\ i = 1, \ldots, m. \end{aligned} \tag{6}$$

# 3 A merit function

In this section, we review the merit function recently proposed by the authors [13] for the inequality constrained variational inequality problem. The reader may refer to [13] for details.

Choose an $n \times n$ positive definite matrix $G$ and define the function $f : R^n \to R$ by

$$f(x) = \max_y \left\{ -\langle F(x), y - x\rangle - \frac{1}{2}\langle y - x, G(y - x)\rangle \;\middle|\; y \in T(x) \right\}, \tag{7}$$

where $T(x)$ is the polyhedral convex set defined by

$$T(x) = \{y \in R^n \mid c_i(x) + \langle \nabla c_i(x), y - x\rangle \leq 0, \quad i = 1, \ldots, m\}.$$

By the convexity of $c_i$, it is easy to verify that, for all $x \in R^n$, $T(x)$ is a closed convex set containing $S$. Thus $f$ provides an over-estimate of the function $\hat{f}$ introduced by Fukushima [4], i.e., $f(x) \geq \hat{f}(x)$ for all $x \in R^n$. In particular, when $c_i$ are all linear, $f$ coincides with $\hat{f}$. Note that the positive definiteness of $G$ guarantees that the maximum in (7) is always attained by $y = H(x)$ uniquely, where $H(x)$ is the unique solution $y$ to the convex quadratic programming problem

$$\begin{array}{lll} QP(x): & \text{minimize}_y & \frac{1}{2}\langle y-x, G(y-x)\rangle + \langle F(x), y-x\rangle \\ & \text{subject to} & c_i(x) + \langle \nabla c_i(x), y-x\rangle \leq 0, \quad i = 1, \ldots, m. \end{array} \tag{8}$$

Therefore, the function $f$ can be written as

$$f(x) = -\langle F(x), H(x) - x\rangle - \frac{1}{2}\langle H(x) - x, G(H(x) - x)\rangle. \tag{9}$$

Let $\Lambda(x)$ denote the set of optimal Lagrange multiplier vectors for $QP(x)$, that is,

$$\begin{aligned} \Lambda(x) \;=\; \{\lambda \in R^m \mid F(x) + G(H(x) - x) + \sum_{i=1}^{m} \lambda_i \nabla c_i(x) = 0, \lambda_i \geq 0, \\ \lambda_i[c_i(x) + \langle \nabla c_i(x), H(x) - x\rangle] = 0, i = 1, \ldots, m\}. \end{aligned} \tag{10}$$

Since $H(x^*) = x^*$ holds for the solution $x^*$ of (1) [13, Lemma 2.1], $\Lambda(x^*)$ coincides with the set of vectors $\lambda^*$ satisfying (6).

Using the function $f$, we can formulate the optimization problem

$$\text{minimize} \;\; f(x) \;\; \text{subject to} \;\; x \in S. \tag{11}$$

We can prove that this problem is equivalent to the variational inequality problem (1).

**Proposition 3.1** *[13] Let the function $f: R^n \to R$ be defined by (7). Then $f(x) \geq 0$ for all $x \in S$. Moreover, $x \in S$ and $f(x) = 0$ if and only if $x$ solves the variational inequality problem (1). Hence $x$ solves (1) if and only if it solves the optimization problem (11).*

For given $x \in R^n$ and $\lambda \in R^m$, we define the matrix $M(x, \lambda)$ by

$$M(x, \lambda) = \nabla F(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 c_i(x). \tag{12}$$

The next proposition demonstrates the directional differentiability of $f$.

**Proposition 3.2** *[13] Suppose that the mapping $F$ is continuous and the convex functions $c_i$, $i = 1, \ldots, m$, are continuously differentiable. Suppose also that the Slater's constraint qualification (5) holds. Then the function $f$ defined by (7) is continuous on $R^n$. Moreover, if $F$ is continuously differentiable and $c_i$, $i = 1, \ldots, m$, are twice continuously differentiable, then $f$ is directionally differentiable in any direction $d \in R^n$ and its directional derivative $f'(x; d)$ is given by*

$$f'(x; d) = \min_{\lambda \in \Lambda(x)} \langle F(x) - [M(x, \lambda) - G](H(x) - x),\ d \rangle . \tag{13}$$

**Remark 3.1** If the set $\Lambda(x)$ is a singleton $\{\lambda\}$, then $f$ is differentiable at $x$ and the gradient is given by

$$\nabla f(x) = F(x) - [M(x, \lambda) - G](H(x) - x).$$

A sufficient condition [3, Theorem 6] for $\Lambda(x)$ to be a singleton is that the gradient vectors $\nabla c_i(x), i \in I(x)$, are linearly independent, where $I(x) = \{i \mid c_i(x) + \langle \nabla c_i(x), H(x) - x \rangle = 0\}$, and the strict complementarity condition is satisfied, i.e., $\lambda_i = 0$ implies $c_i(x) + \langle \nabla c_i(x), H(x) - x \rangle < 0$.

By Proposition 3.1, $x$ solves the variational inequality problem (1) if and only if $x$ is a global optimal solution of (11). The next proposition gives a condition under which any point satisfying the first order necessary optimality condition for (11) is actually a global optimal solution of (11).

**Proposition 3.3** *[13] Suppose that the mapping $F$ is continuously differentiable, $\nabla F(x)$ is positive definite for all $x$, the convex functions $c_i$, $i = 1, \ldots, m$, are twice continuously differentiable and the Slater's constraint qualification (5) is satisfied. If $x \in S$ and*

$$f'(x; y - x) \geq 0 \quad \textit{for all } y \in S,$$

*then $x$ is a global optimal solution of (11), and hence $x$ is a solution to (1).*

The following lemma will be useful in the next section.

**Lemma 3.1** *For any $x$, we have*

$$\begin{aligned} f(x) &= \frac{1}{2}\langle H(x) - x, G(H(x) - x)\rangle - \sum_{i=1}^{m} \lambda_i c_i(x) \\ &\geq -\sum_{i=1}^{m} \lambda_i c_i(x) \end{aligned}$$

*for any $\lambda \in \Lambda(x)$. In particular, if $x \in S$, then*

$$f(x) \geq \frac{1}{2}\langle H(x) - x, G(H(x) - x)\rangle.$$

**Proof.** Since $H(x)$ solves (8), it follows from the definition (10) of $\Lambda(x)$ that each vector $\lambda \in \Lambda(x)$ satisfies

$$\begin{aligned} F(x) + G(H(x) - x) + \sum_{i=1}^{m} \lambda_i \nabla c_i(x) &= 0, \\ c_i(x) + \langle \nabla c_i(x), H(x) - x\rangle \leq 0, \qquad \lambda_i &\geq 0, \\ \lambda_i[c_i(x) + \langle \nabla c_i(x), H(x) - x\rangle] &= 0, \quad i = 1, \ldots, m. \end{aligned}$$

Hence, we have from (9)

$$\begin{aligned} f(x) &= -\langle F(x), H(x) - x\rangle - \frac{1}{2}\langle H(x) - x, G(H(x) - x)\rangle \\ &= \langle G(H(x) - x), H(x) - x\rangle + \left\langle \sum_{i=1}^{m} \lambda_i \nabla c_i(x), H(x) - x \right\rangle \\ &\quad -\frac{1}{2}\langle H(x) - x, G(H(x) - x)\rangle \\ &= \frac{1}{2}\langle H(x) - x, G(H(x) - x)\rangle + \left\langle \sum_{i=1}^{m} \lambda_i \nabla c_i(x), H(x) - x \right\rangle \\ &= \frac{1}{2}\langle H(x) - x, G(H(x) - x)\rangle - \sum_{i=1}^{m} \lambda_i c_i(x) \\ &\geq -\sum_{i=1}^{m} \lambda_i c_i(x), \end{aligned}$$

where the last inequality follows from the positive definiteness of $G$. Since $\lambda_i \geq 0$ and $c_i(x) \leq 0,\ i = 1, \ldots, m$, for all $x \in S$, the last part of the lemma follows immediately. □

# 4 Globally convergent Newton method

In this section, we present a globally convergent Newton method for the variational inequality problem (1), which incorporates an Armijo-type line search procedure for the penalty function $\theta_r : R^n \to R$ defined by

$$\theta_r(x) = f(x) + r \sum_{i=1}^{m} \max(0, c_i(x)),$$

where $r$ is a sufficiently large positive parameter. By Proposition 3.2 and [6, Lemma 3.1], $\theta_r$ is directionally differentiable and the directional derivative is given by

$$\theta_r'(x;d) = f'(x;d) + r\sum_{i\in I_+}\langle \nabla c_i(x), d\rangle + r\sum_{i\in I_0}\max(0, \langle \nabla c_i(x), d\rangle), \tag{14}$$

where $I_+ = \{i \mid c_i(x) > 0\}$ and $I_0 = \{i \mid c_i(x) = 0\}$. Throughout this section, we assume that the mapping $F$ is continuously differentiable and strongly monotone with modulus $\mu$, so that $\nabla F$ satisfies (4). Note that, since the convexity of $c_i$ guarantees that $\nabla^2 c_i(x)$ is positive semi-definite, (4) implies that the matrix $M(x,\lambda)$ defined by (12) satisfies

$$\langle d, M(x,\lambda)d\rangle \geq \mu \parallel d \parallel^2 \quad \text{for all } x, d \in R^n, \tag{15}$$

whenever $\lambda \geq 0$.

Now we state the algorithm.

## Algorithm

**Step 0** Choose $x^0 \in R^n$, $r > 0$, $0 < \beta < 1$, $0 < \sigma < 1$, and a symmetric positive definite matrix $G$. Let $k := 0$

**Step 1** Find the unique solution $\bar{x}^k \in T(x^k)$ of the linearized variational inequality problem

$$\langle F(x^k) + M(x^k, \bar{\lambda}^k)^T(\bar{x}^k - x^k), x - \bar{x}^k\rangle \geq 0 \quad \text{for all} \quad x \in T(x^k), \tag{16}$$

where $\bar{\lambda}^k$ is an arbitrary vector in $\Lambda(x^k)$. Let $d^k := \bar{x}^k - x^k$.

**Step 2** Set $x^{k+1} := x^k + \beta^{m_k} d^k$, where $m_k$ is the smallest nonnegative integer $m$ such that

$$\theta_r(x^k) - \theta_r(x^k + \beta^m d^k) \geq -\sigma\beta^m \theta_r'(x^k; d^k). \tag{17}$$

Let $k := k + 1$. Go to Step 1.

Note that in Step 1 we need an optimal Lagrange multiplier vector $\bar{\lambda}^k$ for the quadratic programming problem QP($x^k$) (cf. (8)). This has already been obtained in the previous iteration as a by-product of evaluating the function value $f(x^k)$. Note also that, by the positive definiteness of $M$, the linearized problem (16) always has a unique solution. Moreover problem (16) can be rewritten as a linear complementary problem, which can be solved in a finite number of steps using Lemke's complementarity pivoting algorithm [9]. The following theorem shows that the vector $d^k$ generated by the algorithm is a descent direction of $\theta_r$ at $x^k$.

**Theorem 4.1** *Let the mapping $F$ be continuously differentiable and strongly monotone with modulus $\mu$, and let the convex functions $c_i$, $i = 1, \ldots, m$, be twice continuously differentiable. If*

$$\| \lambda \|_\infty \leq r \quad \textit{for all } \lambda \in \Lambda(x^k),$$

*then the vector $d^k = \bar{x}^k - x^k$ satisfies the inequality*

$$\theta_r(x^k; d^k) \leq -\left(\mu - \frac{1}{2} \| G \|\right) \| d^k \|^2 . \tag{18}$$

*In particular, if the matrix $G$ is chosen sufficiently small to satisfy $\| G \| < 2\mu$, then $d^k$ is a descent direction of $\theta_r$ at $x^k$.*

**Proof.** For simplicity of notation, we omit the superscript $k$ in $x^k$ and $d^k$. Let $I_+ = \{i \,|\, c_i(x) > 0\}$ and $I_0 = \{i \,|\, c_i(x) = 0\}$. Note that $d = \bar{x} - x$ together with some Lagrange multiplier vector $\hat{\lambda} \geq 0$ satisfies

$$F(x) + M(x, \bar{\lambda})^T d + \textstyle\sum_{i=1}^m \hat{\lambda}_i \nabla c_i(x) = 0, \tag{19a}$$

$$c_i(x) + \langle \nabla c_i(x), d \rangle \leq 0, \tag{19b}$$

$$\hat{\lambda}_i [c_i(x) + \langle \nabla c_i(x), d \rangle] = 0, \qquad i = 1, \ldots, m. \tag{19c}$$

Then (19b) yields

$$\sum_{i \in I_0} \max(0, \langle \nabla c_i(x), d \rangle) = 0. \tag{20}$$

Since $d = \bar{x} - x$, and since $M(x, \lambda) = \nabla F(x) + \sum_{i=1}^m \lambda_i \nabla^2 c_i(x)$, it follows from (13) that

$$\begin{aligned}
f'(x; d) &= \min_{\lambda \in \Lambda(x)} \langle F(x) - [M(x, \lambda) - G](H(x) - x), \bar{x} - x \rangle \\
&\leq \left\langle F(x) - [M(x, \bar{\lambda}) - G](H(x) - x), \bar{x} - x \right\rangle \\
&= \langle F(x), \bar{x} - x \rangle - \left\langle H(x) - x, M(x, \bar{\lambda})^T (\bar{x} - x) \right\rangle + \langle G(H(x) - x), \bar{x} - x \rangle \\
&= \left\langle F(x) + M(x, \bar{\lambda})^T (\bar{x} - x), \bar{x} - x \right\rangle - \left\langle M(x, \bar{\lambda})^T (\bar{x} - x), \bar{x} - x \right\rangle \\
&\quad - \left\langle F(x) + M(x, \bar{\lambda})^T (\bar{x} - x), H(x) - x \right\rangle + \langle F(x), H(x) - x \rangle \\
&\quad + \langle G(H(x) - x), \bar{x} - x \rangle \\
&= -\left\langle F(x) + M(x, \bar{\lambda})^T (\bar{x} - x), H(x) - \bar{x} \right\rangle
\end{aligned}$$

$$+\left\{\langle F(x), H(x)-x\rangle + \frac{1}{2}\langle H(x)-x, G(H(x)-x)\rangle\right\}$$
$$-\left\langle d, M(x,\bar{\lambda})d\right\rangle + \frac{1}{2}\langle d, Gd\rangle$$
$$-\frac{1}{2}\langle \bar{x}-H(x), G(\bar{x}-H(x))\rangle, \tag{21}$$

where the last equality follows from the equality

$$2\langle \bar{x}-x, G(H(x)-x)\rangle$$
$$= \langle H(x)-x, G(H(x)-x)\rangle + \langle \bar{x}-x, G(\bar{x}-x)\rangle - \langle \bar{x}-H(x), G(\bar{x}-H(x))\rangle.$$

Since $\bar{x}$ is a solution to (16), the first term of (21) is nonpositive. From (9), the second term of (21) equals $-f(x)$. The last term is nonpositive by the positive definiteness of $G$. Hence, we have

$$f'(x;d) \le -f(x) - \left\langle d, M(x,\bar{\lambda})d\right\rangle + \frac{1}{2}\langle d, Gd\rangle.$$

Moreover, since $\bar{\lambda} \in \Lambda(x)$, it follows from Lemma 3.1 that

$$f'(x;d) \le -\langle d, M(x,\bar{\lambda})d\rangle + \frac{1}{2}\langle d, Gd\rangle + \sum_{i=1}^{m}\bar{\lambda}_i c_i(x). \tag{22}$$

Hence, we have

$$\begin{aligned}
\theta_r'(x;d) &\le -\langle d, M(x,\bar{\lambda})^T d\rangle + \frac{1}{2}\langle d, Gd\rangle + \sum_{i=1}^{m}\bar{\lambda}_i c_i(x) + r\sum_{i\in I_+}\langle \nabla c_i(x), d\rangle \\
&\le -\langle d, M(x,\bar{\lambda})^T d\rangle + \frac{1}{2}\langle d, Gd\rangle + \sum_{i\in I_+}(\bar{\lambda}_i - r)c_i(x) \\
&\le -\left(\mu - \frac{1}{2}\,\|\,G\,\|\right)\,\|\,d^k\,\|^2,
\end{aligned}$$

where the first inequality follows from (14), (20) and (22), the second inequality follows from (19b) together with the fact that $\bar{\lambda}_i \ge 0$ for all $i$ and $c_i(x) \le 0$ for $i \notin I_+$, and the third inequality follows from (15) and $\|\,\lambda\,\|_\infty \le r$ for all $\lambda \in \Lambda(x)$. This proves (18). The last part of the theorem follows immediately. □

Next we show the global convergence of the algorithm.

**Theorem 4.2** *Suppose that the mapping $F$ is strongly monotone with modulus $\mu$. Suppose also that the parameter $r$ is chosen sufficiently large. If the matrix $G$ is chosen to satisfy $\|\,G\,\| < 2\mu$, and if the sequence $\{x^k\}$ generated by the algorithm is bounded, then $\{x^k\}$ converges to the unique solution to the variational inequality problem (1).*

**Proof.** Since the sequence $\{x^k\}$ is bounded, it follows from [6, Lemma 3.3] that there exists a positive number $\bar{r} > 0$ such that $\| \lambda^k \|_\infty \leq \bar{r}$ for all $k$, where $\lambda^k$ is any vector in $\Lambda(x^k)$. Assuming that $r \geq \bar{r}$, we have from Theorem 4.1 that $d^k$ satisfies the descent condition (18), whenever $x^k$ is not a solution to (1). Hence, by the line search rule (17), the sequence $\{\theta_r(x^k)\}$ is decreasing. This together with the boundedness of $\{x^k\}$ implies that there is at least an accumulation point. In a way similar to the proof of [12, Theorem 4.1], it can be shown that any accumulation point is a solution to (1). Moreover, under the strong monotonicity assumption, problem (1) must have a unique solution. Therefore we conclude that the entire sequence $\{x^k\}$ converges to the unique solution to (1). $\square$

Next we examine the asymptotic rate of convergence of the algorithm. To this end, we consider the iterates $(x^k, \lambda^k)$ generated by the Newton method directly applied to the mixed nonlinear complementarity problem (6), namely

$$\begin{aligned} &F(x^k) + M(x^k, \lambda^k)^T(x^{k+1} - x^k) + \textstyle\sum_{i=1}^m \lambda_i^{k+1} \nabla c_i(x^k) = 0, \\ &c_i(x^k) + \langle \nabla c_i(x^k), x^{k+1} - x^k \rangle \leq 0, \qquad \lambda_i^{k+1} \geq 0, \\ &\lambda_i^{k+1}[c_i(x^k) + \langle \nabla c_i(x^k), x^{k+1} - x^k \rangle] = 0, \quad i = 1, \ldots, m. \end{aligned} \tag{23}$$

It can be shown [5] that, if $\nabla F(x^*)$ is positive definite, and if the strict complementarity and the linear independence of the active constraints hold, then the sequence generated by the Newton method (23) is quadratically convergent, provided that the starting point is chosen sufficiently close to the solution. (Note that [5] deals with nonlinear programming problems, which correspond to the special case of problem (1) where $F$ is a gradient mapping of some scalar function, so that $F$ is symmetric. But the symmetry assumption is not used in the proof of the theorem in [5].)

Note that a solution $x^{k+1}$ to (23) is a solution of the variational inequality problem

$$\langle F(x^k) + M(x^k, \lambda^k)^T(x^{k+1} - x^k), x - x^{k+1} \rangle \geq 0 \quad \text{for all} \quad x \in T(x^k), \tag{24}$$

which is the same problem as (16) solved in Step 1 of the algorithm, except for the choice of $\lambda^k$. Therefore, if $\| M(x^k, \bar{\lambda}^k) - M(x^k, \lambda^k) \|$ tends to zero as $x^k \to x^*$, then the sequence $\{x^k\}$ generated by solving the linearized variational inequality problem

$$\langle F(x^k) + M(x^k, \bar{\lambda}^k)^T(x^{k+1} - x^k), x - x^{k+1} \rangle \geq 0 \quad \text{for all} \quad x \in T(x^k)$$

with an arbitrary $\bar{\lambda}^k \in \Lambda(x^k)$, is locally superlinearly convergent.

Since the vector $\bar{\lambda}^k$ belongs to $\Lambda(x^k)$ defined by (10), and $\lambda^k$ in (24) is determined at the previous Newton iteration (23), both $\bar{\lambda}^k$ and $\lambda^k$ approach the set $\Lambda(x^*)$ whenever $x^k$ converges to $x^*$. In particular, if $\Lambda(x^*)$ consists of the unique vector $\lambda^*$, then both $\bar{\lambda}^k$ and $\lambda^k$ converge to $\lambda^*$, and hence we have $\| M(x^k, \bar{\lambda}^k) - M(x^k, \lambda^k) \| \to 0$. Note that the uniqueness of the Lagrange multiplier vector $\lambda^*$ is ensured by the linear independence of the active constraints.

These observations are summarized in the following theorem.

**Theorem 4.3** *Let the assumptions of Theorem 4.2 be satisfied. In addition, suppose that the strict complementarity and the linear independence of the active constraints hold at the solution $x^*$. If there is an integer $\bar{k} > 0$ such that the unit step size is accepted in Step 2 of the algorithm for all $k \geq \bar{k}$, then the sequence $\{x^k\}$ generated by the algorithm converges superlinearly to the solution $x^*$.*

# 5 Concluding remarks

We have proposed a Newton method for solving variational inequality problems and shown that, under the strong monotonicity assumption, the method is globally convergent and that, under some additional assumptions, the rate of convergence is superlinear.

When $F$ is a gradient mapping of some differentiable convex function $\varphi$, problem (1) corresponds to a necessary and sufficient optimality condition for the convex programming problem

$$\begin{array}{ll} \text{minimize} & \varphi(x) \\ \text{subject to} & c_i(x) \leq 0, \quad i = 1, \ldots, m. \end{array} \tag{25}$$

Therefore we may apply our method to (25) with the identification $F = \nabla\varphi$. In this case, the matrix $M$ defined by (12) is rewritten as

$$M(x, \lambda) = \nabla^2 \varphi(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 c_i(x),$$

which is the Hessian of the Lagrangean of problem (25). Moreover, since $M$ is symmetric, the subproblem (16) solved in Step 1 can be rewritten as

$$\begin{array}{ll} \text{minimize}_d & \dfrac{1}{2}\langle d, M(x^k, \bar{\lambda}^k) d \rangle + \langle F(x^k), d \rangle \\ \text{subject to} & c_i(x^k) + \langle \nabla c_i(x^k), d \rangle \leq 0 \quad i = 1, \ldots, m. \end{array}$$

Thus our algorithm reduces to a successive quadratic programming (SQP) method. A major difference from other SQP methods is that our algorithm makes use of the function $f$ as a merit function to globalize the convergence, instead of using a penalty function associated with problem (25).

# References


[1] G. Auchmuty, Variational Principles for Variational Inequalities, *Numerical Functional Analysis and Optimization* 10 (1989) 863-874.

[2] A. Auslender, *Optimisation: Méthodes Numériques*, Masson, Paris, 1976.

[3] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, SIAM, Philadelphia, 1990.

[4] M. Fukushima, Equivalent Differentiable Optimization Problems and Descent Methods for Asymmetric Variational Inequality Problems, *Mathematical Programming* 53 (1992) 99-110.

[5] U. M. Garcia Palomares and O. L. Mangasarian, Superlinearly Convergent Quasi-Newton Algorithms for Nonlinearly Constrained Optimization Problems, *Mathematical Programming* 11 (1976) 1-13.

[6] S. P. Han, A Globally Convergent Method for Nonlinear Programming, *Journal of Optimization Theory and Applications* 22 (1977) 297-309.

[7] 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.

[8] T. Larsson and M. Patriksson, A Class of Gap Functions for Variational Inequalities, *Mathematical Programming* 64 (1994) 53-79.

[9] C. E. Lemke, Bimatrix Equilibrium Points and Mathematical Programming, *Management Science* 11 (1965) 681-689.

[10] P. Marcotte and J. P. Dussault, A Note on a Globally Convergent Newton Method for Solving Variational Inequalities, *Operations Research Letters* 6 (1987) 35-42.

[11] J. M. Ortega and W. C. Rheinbolt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[12] K. Taji, M. Fukushima and T. Ibaraki, A Globally Convergent Newton Method for Solving Strongly Monotone Variational Inequalities, *Mathematical Programming* 58 (1993) 369-383.

[13] K. Taji and M. Fukushima, A New Merit Function and a Successive Quadratic Programming Algorithm for Variational Inequality Problems, Technical Report NAIST-IS-TR94013, Graduate School of Information Science, Nara Institute of Science and Technology, Nara, Japan, 1994.

Kouichi TAJI and Masao FUKUSHIMA
Graduate School of Information Science,
Nara Institute of Science and Technology,
Takayama, Ikoma, Nara 630-01, JAPAN