Optimization Based Globally Convergent Methods for the Nonlinear Complementarity Problem

Kouichi TAJI Masao FUKUSHIMA

Abstract

The nonlinear complementarity problem has been used to study and formulate various equilibrium problems including the traffic equilibrium problem, the spatial equilibrium problem and the Nash equilibrium problem. To solve the nonlinear complementarity problem, various iterative methods such as projection methods, linearized methods and Newton method have been proposed and their convergence results have been established. In this paper, we propose globally convergent methods based on differentiable optimization formulation of the problem. The methods are applications of a recently proposed method for solving variational inequality problems, but they take full advantage of the special structure of nonlinear complementarity problem. We establish global convergence of the proposed methods, which is a refinement of the results obtained for variational inequality counterparts. Some computational experience indicates that the proposed methods are practically efficient.

Key Words. Nonlinear complementarity problem, optimization formulation, descent method, Newton method, global convergence.

1 Introduction

We consider the nonlinear complementarity problem, which is to find a vector $x \in R^n$ such that

$$x \geq 0, F(x) \geq 0 \text{ and } x^T F(x) = 0, \tag{1}$$

where $F(x) = (F_1(x), F_2(x), \ldots, F_n(x))^T$ is a given continuously differentiable mapping from R^n into itself and ^{T} denotes transposition. This problem has been used to formulate and study various equilibrium problems including the traffic equilibrium problem, the spatial economic equilibrium problem and Nash equilibrium problem [4, 10, 12, 15].

To solve the nonlinear complementarity problem (1), various iterative algorithms, such as fixed point algorithms, projection methods, nonlinear Jacobi method, successive over-relaxation methods and Newton method, have been proposed [6, 7, 13]. Many of these methods are generalizations of classical methods for systems of nonlinear equations and their convergence results have been studied extensively [7, 13].

Assuming the monotonicity of mapping F, Fukushima [5] has recently proposed a differentiable optimization formulation for variational inequality problem and proposed a decent algorithm to solve variational inequality problem. Based on this optimization formulation, Taji et al. [14] proposed a modification of Newton method for solving the variational inequality problem, and proved that, under the strong monotonicity assumption, the method is globally and quadratically convergent.

In this paper we apply the methods of Fukushima [5] and Taji et al. [14] to the nonlinear complementarity problem. We show that those methods can take full advantage of the special structure of problem (1), thereby yielding new algorithms for solving monotone complementarity problems. Moreover, we establish global convergence of the proposed methods, which is a refinement of the results obtained for the variational inequality counterparts in several respects. In particular, we remark that the constraint set of nonlinear complementarity problem is R^n_+ which is clearly not compact, while the descent method by Fukushima [5] assumes that the constraint set is compact. In this paper we show that the compactness assumption can be removed for the nonlinear complementarity problem. We also present some computational results to demonstrate that the proposed methods are practically efficient.

2 Equivalent optimization problem

In this section, we introduce a merit function for the nonlinear complementarity problem (1) and present some of its properties.

Choose positive parameters $\delta_i > 0, i = 1, 2, \ldots, n$ and define the function $f : R^n \to R^n$ by

$$f(x) = \sum_{i=1}^{n} \frac{1}{2\delta_i} \left\{ F_i(x)^2 - (\max(0, F_i(x) - \delta_i x_i))^2 \right\}. \tag{2}$$

This function is a special case of the function originally introduced by Fukushima [5] for variational inequality problem. Though, some of its properties can be derived from the results of [5], we give here simple and direct proofs for these properties, which utilize the special structure of problem (1).

For convenience, we define

$$f_i(x) = \frac{1}{2\delta_i} \{ F_i(x)^2 - (\max(0, F_i(x) - \delta_i x_i))^2 \}, \tag{3}$$

hence f is written as $f(x) = \sum_{i=1}^{n} f_i(x)$. We denote D as a diagonal matrix such that $D = diag(\delta_1, \delta_2, \ldots, \delta_n)$, and e_i as the i-th unit vector such that

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} < i.$$

We also denote

$$M(x) = \max(0, x - D^{-1} F(x)), \tag{4}$$

where max operator is taken component-wise, i.e.,

$$M_i(x) = \max(0, x_i - \frac{1}{\delta_i} F_i(x)).$$

Using this notation, we have the following result.

Lemma 2.1 *Let the mapping $M : R^n \to R^n_+$ be defined by (4). Then x^* solves (1) if and only if $x^* = M(x^*)$.*

Proof. Suppose that x^* is a solution to (1). Then either

$$\begin{cases} x_i^* = 0 \\ F_i(x^*) \geq 0 \end{cases} \quad \text{or} \quad \begin{cases} x_i^* \geq 0 \\ F_i(x^*) = 0 \end{cases}$$

holds for all $i = 1, 2, \ldots, n$. Since $\delta_i > 0$, we have

$$x_i^* = 0 \Longrightarrow M_i(x^*) = \max(0, -\frac{1}{\delta_i} F_i(x)) = 0,$$

and

$$F_i(x^*) = 0 \Longrightarrow M_i(x^*) = \max(0, x_i^*) = x_i^*.$$

Thus $M_i(x^*) = x_i^*$ for all i.

On the other hand, suppose that x^* satisfies $x^* = M(x^*)$. Then, for all i, either

$$\begin{cases} x_i^* = 0 \\ x_i^* - \delta_i^{-1} F_i(x^*) \leq 0 \end{cases} \quad \text{or} \quad \begin{cases} x_i^* = x_i^* - \delta_i^{-1} F_i(x^*) \\ x_i^* - \delta_i^{-1} F_i(x^*) \geq 0 \end{cases}$$

holds. Hence, it follows from $\delta_i > 0$ that either $x_i^* = 0$ and $F_i(x^*) \geq 0$, or $x_i^* \geq 0$ and $F_i(x^*) = 0$ holds for all i. Thus x^* solves (1). □

Using the function (2), an equivalent optimization problem is obtained for nonlinear complementarity problem (1).

Proposition 2.1 *Let the function $f : R^n \to R$ be defined by (2). Then $f(x) \geq 0$ for all $x \geq 0$, and $f(x) = 0$ if and only if x solves (1). Hence, x solves (1) if and only if it is a solution to the following optimization problem:*

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad x \geq 0. \tag{5}$$

Proof. We first show that $f_i(x) \geq 0$ for all $x \geq 0$, so that $f(x) \geq 0$ for all $x \geq 0$. If $F_i(x) - \delta_i x_i \leq 0$, then $f_i(x) = (2\delta_i)^{-1} F_i(x)^2 \geq 0$. So we consider the case $F_i(x) - \delta_i x_i > 0$. Since $\delta_i > 0$, $x_i \geq 0$ and $F_i(x) > \delta_i x_i$ hold, we see, from (3),

$$\begin{aligned} f_i(x) &= \frac{1}{2\delta_i}\{F_i(x)^2 - (F_i(x) - \delta_i x_i)^2\} \\ &= x_i F_i(x) - \frac{\delta_i}{2} x_i^2 \\ &\geq \frac{\delta_i}{2} x_i^2 \\ &\geq 0. \end{aligned}$$

Therefore, $f(x) \geq 0$ for all $x \geq 0$.

Next, suppose $f(x) = 0$. Then $f_i(x) = 0$ must hold for all i. Hence, as shown in the above, either $F_i(x) = 0$ and $F_i(x) - \delta_i x_i \leq 0$, or $x_i = 0$ and $F_i(x) - \delta_i x_i > 0$ holds for all i. Therefore, x is a solution of (1).

On the other hand, suppose that x solves (1). Then either $F(x_i) = 0$ or $x_i = 0$ holds for all i. If $F(x_i) = 0$, then from (3) we have

$$f_i(x) = -\frac{1}{2\delta_i}(\max(0, -\delta_i x_i))^2 = 0.$$

Also if $x_i = 0$, then we have

$$\begin{aligned} f_i(x) &= \frac{1}{2\delta_i}\{F_i(x)^2 - (\max(0, F_i(x)))^2\} \\ &= \frac{1}{2\delta_i}\{F_i(x)^2 - F_i(x)^2\} \\ &= 0. \end{aligned}$$

Therefore, $f(x) = 0$. □

It can be shown that the function f is continuously differentiable whenever so is the mapping F.

Proposition 2.2 *If the mapping F is continuously differentiable, then so is the function f defined by (2), and the gradient of f is given by*

$$\nabla f(x) = F(x) - (\nabla F(x) - D)(M(x) - x). \tag{6}$$

Proof. We first note that, if a function $\theta : R^n \to R$ is continuously differentiable and $\Theta(x) = (\max(0, \theta(x)))^2$, then Θ is continuously differentiable and the gradient of Θ is given by

$$\nabla\Theta(x) = \max(0, \theta(x))\nabla\theta(x).$$

Hence, from (3) we have

$$\nabla f_i(x) = \frac{1}{\delta_i}(F_i(x) - \max(0, F_i(x) - \delta_i x_i))\nabla F_i(x) + \max(0, F_i(x) - \delta_i x_i)e_i. \tag{7}$$

Since

$$\begin{aligned} \max(0, F_i(x) - \delta_i x_i) &= F_i(x) - \delta_i x_i + \delta_i \max(0, x_i - \frac{1}{\delta_i}F_i(x)) \\ &= F_i(x) - \delta_i x_i + \delta_i M_i(x) \end{aligned}$$

holds, we have from (7)

$$\begin{aligned}\nabla f_i(x) &= \frac{1}{\delta_i}(F_i(x) - \max(0, F_i(x) - \delta_i x_i))\nabla F_i(x) + \max(0, F_i(x) - \delta_i x_i)e_i \\ &= (x_i - M_i(x))\nabla F_i(x) + (F_i(x) - \delta_i x_i + \delta_i M_i(x))e_i. \qquad (8)\end{aligned}$$

Therefore, we have from (8)

$$\begin{aligned}\nabla f(x) &= \sum_{i=1}^{n} \nabla f_i(x) \\ &= \sum_{i=1}^{n} \{(x_i - M_i(x))\nabla F_i(x) + (F_i(x) - \delta_i x_i + \delta_i M_i(x))e_i\} \\ &= F(x) - (\nabla F(x) - D)(M(x) - x). \quad \square\end{aligned}$$

In addition, when $\nabla F(x)$ is positive definite for all $x \geq 0$, it can be shown that, if $x \geq 0$ is a stationary point of problem (5), then x is a global optimal solution of problem (1).

Proposition 2.3 *Suppose that $\nabla F(x)$ is positive definite for all $x \geq 0$. If $x \geq 0$ is a stationary point of problem (5), i.e.,*

$$(y - x)^T \nabla f(x) \geq 0 \quad \text{for all} \quad y \geq 0, \qquad (9)$$

then x is a global optimal solution of problem (5), and hence it solves the nonlinear complementarity problem (1).

Proof. Suppose that x satisfies (9). Then from (6) we have

$$\begin{aligned}&(M(x) - x)^T \nabla f(x) \\ &\quad = (M(x) - x)^T \{(F(x) - Dx + DM(x)) - \nabla F(x)(M(x) - x)\} \\ &\quad = (M(x) - x)^T (F(x) - Dx + DM(x)) - (M(x) - x)^T \nabla F(x)(M(x) - x). \qquad (10)\end{aligned}$$

It is easy to see that

$$\begin{aligned}&(M(x) - x)^T (F(x) - Dx + DM(x)) \\ &\quad = \sum_{i=1}^{n} \{\delta_i (\max(0, x_i - \frac{1}{\delta_i} F_i(x)) - x_i)^2 + (\max(0, x_i - \frac{1}{\delta_i} F_i(x)) - x_i) F_i(x)\} \\ &\quad \leq 0. \qquad (11)\end{aligned}$$

Since x satisfies (9), it follows from (10) and (11) that

$$(M(x) - x)^T \nabla F(x)(M(x) - x) \leq 0.$$

However, since $\nabla F(x)$ is positive definite, we have $x = M(x)$. Therefore, it follows from Lemma 2.1 that x is a solution to (1). $\Box$

We say that the mapping F is strongly monotone on R^n_+ with modulus $\mu > 0$, if

$$(x - y)^T (F(x) - F(y)) \geq \mu \parallel x - y \parallel^2 \quad \text{for all} \quad x, y \geq 0. \tag{12}$$

The following result establishes an asymptotic behavior of the function f. Note that similar results have not been obtained for variational inequality problems.

Proposition 2.4 *If F is strongly monotone on R^n_+, then*

$$\lim_{x \geq 0,\ \|x\| \to \infty} f(x) = +\infty.$$

Proof. Let $\{x^k\}$ be a sequence such that $x^k \geq 0$ and $\| x^k \| \to \infty$. Taking a subsequence if necessary, we may suppose that there exists a set $I \subset \{1, 2, \ldots, n\}$ such that $x^k_i \to +\infty$ for $i \in I$ and $\{x^k_i\}$ is bounded for $i \notin I$. From $\{x^k\}$, we define another sequence $\{y^k\}$ such that $y^k_i = 0$ if $i \in I$ and $y^k_i = x^k_i$ if $i \notin I$. From (12) and the definition of y^k, we have

$$\sum_{i \in I} (F_i(x^k) - F_i(y^k)) x^k_i \geq \mu \sum_{i \in I} (x^k_i)^2. \tag{13}$$

By Cauchy's inequality we have

$$(\sum_{i \in I} (F_i(x^k) - F_i(y^k))^2)^{\frac{1}{2}} (\sum_{i \in I} (x^k_i)^2)^{\frac{1}{2}} \geq \sum_{i \in I} (F_i(x^k) - F_i(y^k)) x^k_i. \tag{14}$$

It then follows from (13) and (14) that

$$\sum_{i \in I} (F_i(x^k) - F_i(y^k))^2 \geq \mu^2 \sum_{i \in I} (x^k_i)^2. \tag{15}$$

Since $\{y^k\}$ is bounded by definition, $\{F_i(y^k)\}$ is also bounded. Therefore, since $x^k_i \to +\infty$ for all $i \in I$, (15) implies

$$\sum_{i \in I} F_i(x^k)^2 \to \infty.$$

As shown at the beginning of the proof of Proposition 2.1, $f_i(x^k) = \dfrac{1}{2\delta_i} F_i(x^k)^2 \geq 0$ if $F_i(x^k) - \delta_i x_i^k \leq 0$, and $f_i(x^k) \geq \dfrac{\delta_i}{2}(x_i^k)^2$ if $F_i(x^k) - \delta_i x_i^k > 0$. Therefore we have

$$\begin{aligned} f(x^k) &= \sum_{i=1}^{n} f_i(x^k) \\ &\geq \sum_{i \in I} f_i(x^k) \\ &\geq \sum_{i \in I} \frac{1}{2\delta_i} \min(F_i(x^k)^2, (\delta_i x_i^k)^2). \end{aligned}$$

Since $x_i^k \to +\infty$ for all $i \in I$ and $\sum_{i \in I} F_i(x^k)^2 \to \infty$, it follows that $f(x^k) \to +\infty$. □

3 Algorithms

In this section, we propose two globally convergent methods for solving (5). One is based on the methods of Fukushima [5] and the other the method of Taji et al. [14], which were originally proposed for variational inequality problems. Throughout this section, we suppose that the mapping F is strongly monotone on R_+^n with modulus $\mu > 0$.

3.1 Descent method

The first method uses the vector

$$\begin{aligned} d &= M(x) - x \\ &= \max(0, x - D^{-1}F(x)) - x \end{aligned} \tag{16}$$

as a search direction at x. When the mapping F is strongly monotone, it can be shown that the vector d given by (16) is a descent direction.

Lemma 3.1 *If F is strongly monotone with modulus μ, then the vector d given by* (16) *satisfies the descent condition*

$$\nabla f(x)^T d \leq -\mu \, \| \, d \|^2 \, .$$

Proof. From (10) and (11), we have

$$(M(x) - x)^T \nabla f(x) \leq -(M(x) - x)^T \nabla F(x)(M(x) - x). \tag{17}$$

It is known that, when F is differentiable and strongly monotone on R^n_+, ∇F satisfies

$$(y-x)^T \nabla F(x)(y-x) \geq \mu \parallel y-x \parallel^2 \text{ for all } x, y \geq 0.$$

Therefore, from (17) we have

$$\nabla f(x)^T d \leq -\mu \parallel d \parallel^2 . \ \square$$

Thus the direction d can be used to determine the next iterate by using the following Armijo-type line search rule: Let $\alpha := \beta^{\hat{m}}$, where $\hat{m}$ is the smallest nonnegative integer m such that

$$f(x) - f(x+\beta^m d) \geq \sigma \beta^m \parallel d \parallel^2,$$

where $0 < \beta < 1$ and $\sigma > 0$. Note that, in the descent method originally proposed by Fukushima [5] for variational inequality problems, the line search only examines step sizes shorter than unity. Here, we propose the algorithm that allows longer step sizes at each iteration.

Algorithm 1:

Choose $x^0 \geq 0$, $\alpha > 1$, $\beta \in (0,1)$, $\sigma > 0$ and a positive diagonal matrix D;
$k := 0$
while convergence criterion is not satisfied **do**
  $d^k := \max(0, x^k - D^{-1}F(x^k)) - x^k$;
  $\hat{t} := \max\{t \mid x^k + td^k \geq 0,\ t \geq 0\}$;
  $m := 0$
  if $f(x^k) - f(x^k + d^k) \geq \sigma \parallel d^k \parallel^2$ **then**
    while $\alpha^m \leq \hat{t}$ and $f(x^k) - f(x^k + \alpha^m d^k) \geq \sigma \alpha^m \parallel d^k \parallel^2$
      and $f(x^k + \alpha^{m+1} d^k) \leq f(x^k + \alpha^m d^k)$ **do**
      $m := m + 1$
    endwhile
    $x^{k+1} := x^k + \alpha^m d^k$
  else
    while $f(x^k) - f(x^k + \beta^m d^k) < \sigma \beta^m \parallel d^k \parallel^2$ **do**
      $m := m + 1$

endwhile

$x^{k+1} := x^k + \beta^m d^k$

endif

$k := k + 1$

endwhile

Note that the vector $M(x^{k+1}) = \max(0, x^{k+1} - D^{-1}F(x^{k+1}))$ has already been found at the previous iteration as a by-product of evaluating f. Therefore one need not compute again the search direction d^k at the beginning of each iteration.

Theorem 3.1 *Suppose that the mapping F is continuously differentiable and strongly monotone with modulus μ on R^n_+. Suppose also that ∇F is Lipschitz continuous on any bounded subset of R^n_+. Then, for any starting point $x^0 \geq 0$, the sequence $\{x^k\}$ generated by Algorithm 1 converges to the unique solution of problem (1) if the positive constant σ is chosen to be sufficiently small such that $\sigma < \mu$.*

Proof. By Proposition 2.4, the level set $S = \{x \mid f(x) \leq f(x^0)\}$ is bounded. Hence ∇F is Lipschitz continuous on S. Since F is continuously differentiable, it is easy to show that F is also Lipschitz continuous on S. Under these conditions, it is not difficult to show that ∇f is Lipschitz continuous on S, i.e., there exists a constant $L > 0$ such that

$$\| \nabla f(x) - \nabla f(y) \| \leq L \| x - y \| \quad \text{for all} \quad x, y \in S.$$

Therefore, as shown in the proof of [5, Theorem 4.2], any accumulation point of $\{x^k\}$ satisfies $x = M(x)$, and hence solves (1) by Lemma 2.1. Since strong monotonicity of F ensures that problem (1) has a unique solution, we can conclude that the entire sequence converges to the unique solution of (1). □

Remark 3.1 In Fukushima [5], global convergence theorem assumes not only the strong monotonicity of mapping F but the *compactness* of the constraint set, which is not the case for nonlinear complementarity problems. Theorem 3.1 above establishes global convergence under strong monotonicity of the mapping F only.

3.2 Modification of Newton method

The second method is a modification of Newton method, which incorporates a line search strategy. The original Newton method for solving the nonlinear complementarity problem (1) generates a sequence $\{x^k\}$ such that $x^0 \geq 0$ and x^{k+1} is determined as $x^{k+1} := \bar{x}$, where $\bar{x}$ is a solution to the following linearized complementarity problem:

$$x \geq 0, F(x^k) + \nabla F(x^k)^T(x - x^k) \geq 0 \text{ and } x^T(F(x^k) + \nabla F(x^k)^T(x - x^k)) = 0. \quad (18)$$

It is shown [13] that, under suitable assumptions, the sequence generated by (18) quadratically converges to a solution x^* of the problem (1), provided that the starting point x^0 is chosen sufficiently close to x^*.

Lemma 3.2 *When the mapping F is strongly monotone with modulus μ, the vector $d^k := \bar{x} - x^k$ obtained by solving the linearized complementarity problem* (18) *satisfies the inequality*

$$\nabla f(x^k)^T d^k < -(\mu - \frac{1}{4} \| D \|) \| d^k \|^2 .$$

Therefore, d^k is actually a feasible descent direction of f at x^k, if the matrix D is chosen to satisfy $\| D \| = \max_i(\delta_i) < 4\mu$.

Proof. For simplicity of presentation, we omit the superscript k in x^k and d^k. Since $d := \bar{x} - x$, it follows from (6) that

$$\begin{aligned}
\nabla f(x)^T d &= (\bar{x} - x)^T F(x) + (\bar{x} - x)^T(\nabla F(x) - D)(x - M(x)) \\
&= (\bar{x} - x)^T(F(x) + \nabla F(x)^T(\bar{x} - x)) - (\bar{x} - x)^T \nabla F(x)^T(\bar{x} - x)) \\
&\quad +(x - M(x))^T(F(x) + \nabla F(x)^T(\bar{x} - x)) \\
&\quad -(x - M(x))^T F(x) - (\bar{x} - x)^T D(x - M(x)) \\
&= -(M(x) - \bar{x})^T(F(x) + \nabla F(x)^T(\bar{x} - x)) \\
&\quad +(M(x) - x)^T F(x) + (\bar{x} - x)^T D(M(x) - x) \\
&\quad -(\bar{x} - x)^T \nabla F(x)^T(\bar{x} - x). \qquad (19)
\end{aligned}$$

Since $\bar{x}$ is a solution to (18) and $M(x) \geq 0$, the first term of (19) is nonnegative. From (11), we have

$$(M(x)-x)^T F(x) \le -(M(x)-x)^T D(M(x)-x).$$

Then it follows from the second term of (19) that

$$\begin{aligned}
&(M(x)-x)^T F(x) + (\bar{x}-x)^T D(M(x)-x) \\
&\quad \le (\bar{x}-x)^T D(M(x)-x) - (M(x)-x)^T D(M(x)-x) \\
&\quad = \sum_{i=1}^{n} \delta_i \{(\bar{x}_i - x_i)(M_i(x)-x_i) - (M_i(x)-x_i)^2\} \\
&\quad = \sum_{i=1}^{n} \frac{\delta_i}{2} \{(\bar{x}_i - x_i)^2 - (M_i(x)-x_i)^2 - (\bar{x}_i - M_i(x))^2\} \\
&\quad \le \sum_{i=1}^{n} \frac{\delta_i}{2} \{(\bar{x}_i - x_i)^2 - \frac{1}{2}(\bar{x}_i - x_i)^2\} \\
&\quad = \sum_{i=1}^{n} \frac{\delta_i}{4} (\bar{x}_i - x_i)^2 \\
&\quad = \frac{1}{4}(\bar{x}-x)^T D(\bar{x}-x). \qquad (20)
\end{aligned}$$

Hence, we have from (19) and (20)

$$\nabla f(x)^T d \le -d^T \nabla F(x)^T d + \frac{1}{4} d^T D d.$$

But since strong monotonicity of F implies $d^T \nabla F(x) d \ge \mu \| d \|^2$ and since $d^T D d \le \| D \| \| d \|^2$, we have

$$\nabla f(x)^T d < -(\mu - \frac{1}{4} \| D \|) \| d \|^2 .$$

The last half of the proposition then follows immediately. □

Using this result, we can construct a modified Newton method for solving the nonlinear complementarity problem (1).

Algorithm 2:

Choose $x^0 \ge 0$, $\beta \in (0,1)$, $\sigma \in (0, \frac{1}{2})$, and a positive diagonal matrix D;

$k := 0$

while convergence criterion is not satisfied **do**

find $\bar{x}$ such that

$x \ge 0, F(x^k) + \nabla F(x^k)^T (x - x^k) \ge 0$ and $x^T (F(x^k) + \nabla F(x^k)^T (x - x^k)) = 0$;

$d^k := \bar{x} - x^k$;

$m := 0$

while $f(x^k) - f(x^k + \beta^m d^k) < -\sigma\beta^m \nabla f(x^k)^T d^k$ **do**

$m := m + 1$

endwhile

$x^{k+1} := x^k + \beta^m d^k$;

$k := k + 1$

endwhile

When the mapping F is strongly monotone, we can establish the global convergence of Algorithm 2.

Theorem 3.2 *Suppose that the mapping F is continuously differentiable and strongly monotone with modulus μ. If the matrix D is chosen such that $\| D \| = \max_i(\delta_i) < 4\mu$, then, for any starting point $x^0 \geq 0$, the sequence $\{x^k\}$ generated by Algorithm 2 converges to the solution of* (1).

Proof. By Lemma 3.2 and the Armijo line search rule, the sequence $\{f(x^k)\}$ is nonincreasing. It then follows from Proposition 2.4 that the sequence $\{x^k\}$ is bounded, and hence it contains at least one accumulation point. As shown in the proof of [14, Theorem 4.1], any accumulation point of $\{x^k\}$ is a solution of (1). Since strong monotonicity of F ensures that problem (1) has a unique solution, we can conclude that the entire sequence converges to the unique solution of (1). □

We can also show that the rate of convergence of Algorithm 2 is quadratic if $F \in C^2$ and the strict complementarity condition holds at the unique solution x^* of (1).

Theorem 3.3 *Suppose that the sequence $\{x^k\}$ generated by Algorithm 2 converges to the solution x^* to problem* (1). *Suppose also that the mapping F belongs to class C^2 and is strongly monotone on R^n_+. If the strict complementarity condition holds at x^*, i.e., $x^*_i = 0$ implies $F_i(x^*) > 0$ for all $i = 1, 2, \ldots, n$, then there exists an integer $\bar{k}$ such that the unit step size is accepted for all $k \geq \bar{k}$. Therefore, the sequence $\{x^k\}$ converges quadratically to the solution x^*.*

Before proving Theorem 3.3, we show the following lemma.

Lemma 3.3 *Let x^* be a solution to problem (1). If $F \in C^2$ and the strict complementarity condition holds at x^*, then $f_i \in C^2$ on a neighborhood of x^*, and the gradient and the Hessian of f_i are given by*

$$\nabla f_i(x) = \begin{cases} (x_i \nabla F_i(x) + F_i(x) e_i) - \delta_i x_i e_i & \text{if } i \in I^* \\ \frac{1}{\delta_i} F_i(x) \nabla F_i(x) & \text{if } i \in \bar{I}^* \end{cases} \tag{21}$$

and

$$\nabla^2 f_i(x) = \begin{cases} x_i \nabla^2 F_i(x) + 2 \nabla F_i(x) e_i^T - \delta_i e_i e_i^T & \text{if } i \in I^* \\ \frac{1}{\delta_i}(F_i(x) \nabla^2 F_i(x) + \nabla F_i(x) \nabla F_i(x)^T) & \text{if } i \in \bar{I}^*, \end{cases} \tag{22}$$

respectively, where $I^* = \{i \,|\, x_i^* = 0\}$ and $\bar{I}^* = \{i \,|\, x_i^* > 0\}$.

Proof. The strict complementarity and the continuity of F ensure that there is a neighborhood B of x^* such that

$$\begin{cases} F_i(x) - \delta_i x_i > 0 & \text{if } i \in I^* \\ F_i(x) - \delta_i x_i < 0 & \text{if } i \in \bar{I}^* \end{cases} \tag{23}$$

holds for all $x \in B$. Hence, we have from (3)

$$f_i(x) = \begin{cases} x_i F_i(x) - \frac{\delta_i}{2} x_i^2 & \text{if } i \in I^* \\ \frac{1}{\delta_i} F_i(x)^2 & \text{if } i \in \bar{I}^*. \end{cases} \tag{24}$$

Therefore, by differentiating (24) directly, we have (21) and (22). □

Proof of Theorem 3.3. Without loss of generality, we assume $\bar{I}^* = \{l, l+1, \ldots, n\}$, where $1 \le l \le n$, and denote

$$x = \begin{pmatrix} x_{I^*} \\ x_{\bar{I}^*} \end{pmatrix}, \quad F(x) = \begin{pmatrix} F_{I^*}(x) \\ F_{\bar{I}^*}(x) \end{pmatrix}, \quad \nabla F(x) = \begin{pmatrix} \nabla_{I^*} F_{I^*}(x) & \nabla_{I^*} F_{\bar{I}^*}(x) \\ \nabla_{\bar{I}^*} F_{I^*}(x) & \nabla_{\bar{I}^*} F_{\bar{I}^*}(x) \end{pmatrix}.$$

Since the strict complementarity holds at x^*, there is an integer K_1 such that x^k satisfies

$$\begin{cases} F_i(x^k) - \delta_i x_i^k > 0 & \text{if } i \in I^* \\ F_i(x^k) - \delta_i x_i^k < 0 & \text{if } i \in \bar{I}^* \end{cases} \tag{25}$$

for all $k \ge K_1$. Under the given assumptions, Newton method (18) is locally quadratically convergent to the solution x^* [13]. Hence, it follows from the strict complementarity and the continuity of F and ∇F that there is an integer K_2 such that

$$\begin{cases} \bar{x}_i = 0 \quad \text{and} \quad F_i(x^k) + \nabla F_i(x^k)^T(\bar{x} - x^k) > 0 \quad \text{if} \quad i \in I^* \\ \bar{x}_i > 0 \quad \text{and} \quad F_i(x^k) + \nabla F_i(x^k)^T(\bar{x} - x^k) = 0 \quad \text{if} \quad i \in \bar{I}^* \end{cases} \tag{26}$$

for all $k \geq K_2$. Moreover, from strict complementarity, it is easy to show that there is an integer K_3 such that if $i \in I^*$ then $x_i^k = 0$ holds for all $k \geq K_3$. Let $K = \max(K_1, K_2, K_3)$. Then we have $x_{I^*}^k = 0$ for all $k \geq K$. It follows from (26) that the search direction d^k is given by $d^k = \begin{pmatrix} 0 \\ d_{\bar{I}^*}^k \end{pmatrix}$, where $d_{\bar{I}^*}^k$ is a solution of the linear equation

$$F_{\bar{I}^*}(x^k) + \nabla F_{\bar{I}^*}(x^k)^T d_{\bar{I}^*}^k = 0. \tag{27}$$

It also follows from (21), (24) and (26) that, for each $i \in I^*$

$$\begin{aligned} & f_i(x^k) - f_i(\bar{x}) + \sigma \nabla f_i(x^k)^T(\bar{x} - x) \\ &= x_i^k F_i(x^k) - \frac{\delta_i}{2}(x_i^k)^2 - \bar{x}_i F_i(\bar{x}) - \frac{\delta_i}{2}\bar{x}_i^2 + \sigma(x_i^k \nabla F_i(x^k) + F_i(x^k)e_i - \delta_i x_i^k e_i)^T(\bar{x} - x^k) \\ &= \sigma(F_i(x^k)(\bar{x}_i - x_i^k) - \delta_i x_i^k(\bar{x}_i - x_i^k)) \\ &= 0, \end{aligned} \tag{28}$$

provided that $k \geq K$.

Therefore, Algorithm 2 is reduced to Newton method for solving the $n - l + 1$-dimensional equation

$$F_{\bar{I}^*} \begin{pmatrix} 0 \\ x_{\bar{I}^*} \end{pmatrix} = 0$$

with Armijo line search procedure for the norm function

$$\bar{f}(x) = \sum_{i \in \bar{I}^*} \frac{1}{2\delta_i} F_i^2 \begin{pmatrix} 0 \\ x_{\bar{I}^*} \end{pmatrix}.$$

Since the strong monotonicity of F ensures the positive definiteness of the matrix ∇F, the submatrix $\nabla_{\bar{I}^*} F_{\bar{I}^*}$ is also positive definite. Hence, it follows from [3] that the unit step is always accepted for sufficiently large k. □

Remark 3.2 Taji et. al. [14] have obtained a globally convergent Newton method for variational inequality problems. In their method, to obtain quadratic convergence, the following line search procedure is used:

Let $0 < \beta < 1$, $0 < \gamma < 1$ and $\sigma \in (0, 1)$;

if $f(x^k + d^k) \leq \gamma f(x^k)$ **then**

$\alpha_k := 1$

else

$m := 0$

while $f(x^k) - f(x^k + \beta^m d^k) < -\sigma \beta^m {d^k}^T \nabla f(x^k)$ **do**

$m := m + 1$

endwhile

$\alpha_k := \beta^m$

endif

$x^{k+1} := x^k + \alpha_k d^k$

Note that this line search procedure, which is similar to the one used by Marcotte and Dussault [11], first checks if the unit step size is acceptable. On the other hand, Algorithm 2 employs Armijo rule in a more direct manner.

4 Computational results

In this section, we report some numerical results for Algorithms 1 and 2 discussed in the previous section. All computer programs were coded in FORTRAN and the runs were made in double precision on a personal computer called Fujitsu FMR-70.

Throughout the computational experiments, the parameters used in the algorithms were set as $\alpha = 2$, $\beta = 0.5$, $\gamma = 0.5$ and $\sigma = 0.0001$. The positive diagonal matrix D was chosen to be the identity matrix multiplied by a positive parameter $\delta > 0$. Therefore the merit function (2) can be written simply as

$$f(x) = \frac{1}{2\delta} \sum_{i=1}^{n} \left\{ F_i(x)^2 - (\max(0, F_i(x) - \delta x_i))^2 \right\}. \tag{29}$$

The search direction of Algorithms 1 can also be written as

$$d^k := \max\left(0, x^k - \frac{1}{\delta} F(x^k)\right) - x^k.$$

The convergence criterion was

$$|\min(x_i, F_i(x))| \leq 10^{-5} \quad \text{for all } i = 1, 2, \ldots, n. \tag{30}$$

For comparison purposes, we also tested two popular methods for solving the nonlinear complementarity problem, the projection method [2] and Newton method without line search [9]. The projection method generates a sequence $\{x^k\}$ such that $x^0 \geq 0$ and x^{k+1} is determined from x^k by

$$x^{k+1} := \max\left(0, x^k - \frac{1}{\delta}F(x^k)\right), \tag{31}$$

for all k. Note that this method may be considered a fixed step-size variant of Algorithm 1. When the mapping F is strongly monotone and Lipschitz continuous with constants μ and L, respectively, this method is globally convergent if δ is chosen large enough to satisfy $\delta > L^2/2\mu$ (see [13, Corollary 2.11.]).

The mappings used in our numerical experiments are of the form

$$F(x) = Ix + \rho(N - N^T)x + \phi(x) + c, \tag{32}$$

where I is the $n \times n$ identity matrix, N is an $n \times n$ matrix such that each row contains only one nonzero element, and $\phi(x)$ is a nonlinear monotone mapping with components $\phi_i(x_i) = p_i x_i^4$, where p_i are positive constants. Elements of matrix N and vector c as well as coefficients p_i are randomly generated such that $-5 \leq N_{ij} \leq 5$, $-25 \leq c_i \leq 25$ and $0.001 \leq p_i \leq 0.006$. The results are shown in Tables $1 \sim 4$. All starting points were chosen to be $(0, 0, \ldots, 0)$. In the tables, $\#f$ is the total number of evaluating the merit function f and all CPU times are in seconds and exclude input/output times. The parameter ρ is used to change the degree of asymmetry of F, namely F deviates from symmetry as ρ becomes large. Since the matrix $I + \rho(N - N^T)$ is positive definite for any ρ and $\phi_i(x_i)$ are monotonically increasing for $x_i \geq 0$, the mapping F defined by (32) is strongly monotone on R^n_+.

4.1 Comparison of Algorithm 1 and the projection method

First we have compared Algorithms 1 and the projection method (31) by using a 10-dimensional example, in which mapping F is given by

$$F(x) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -2 & 0 & -3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & -5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 & 1 & 0 & -5 & 0 & 0 \\ 0 & 0 & -3 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ -5 & 0 & 0 & 0 & 0 & 5 & 0 & 1 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5 & 4 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \end{pmatrix} + \begin{pmatrix} 0.004x_1^4 \\ 0.004x_2^4 \\ 0.003x_3^4 \\ 0.003x_4^4 \\ 0.006x_5^4 \\ 0.006x_6^4 \\ 0.004x_7^4 \\ 0.004x_8^4 \\ 0.004x_9^4 \\ 0.002x_{10}^4 \end{pmatrix} + \begin{pmatrix} 2 \\ 10 \\ 2 \\ 9 \\ -15 \\ 12 \\ -9 \\ 5 \\ 7 \\ -17 \end{pmatrix} . \quad (33)$$

The results for this problem are shown in Table 1.

In general, the projection method is guaranteed to converge, only if the parameter δ is chosen sufficiently large. In fact, Table 1 shows that when δ is large, the projection method is always convergent, but as δ becomes small, the behavior of the method becomes unstable and eventually it fails to converge.

Table 1 also shows that Algorithm 1 is always convergent even if δ is chosen small, since the line search determines an adequate step size at each iteration. In Algorithm 1, the number of iterations is almost constant. This is because we may choose a larger step size when the magnitude of vector d^k is small, i.e. δ is large.

Algorithm 1 spends more CPU times per iteration than the projection method, because the former algorithm requires overheads of evaluating the merit function f. But, when δ becomes large, Algorithm 1 tends to spend less CPU time than the projection method, because the number of iterations of Algorithm 1 increases mildly.

4.2 Comparison of Algorithm 2 and Newton method

Next we have compared Algorithm 2 and the pure Newton method (18) without line search. For each of the problem sizes $n = 30, 50$ and 90, we randomly generated five test problems. The parameters ρ and δ were set as $\rho = 1$ and $\delta = 1$. The starting point was chosen to be $x = 0$. In solving the linearized subproblem at each iteration of Algorithm 2 and Newton method, we used Lemke's complementarity pivoting method [10] coded by Fukushima [8]. The results are given in Table 2. All numbers shown in Table 2 are averages of the results for five problems for each case and #Lemke is the total number of pivotings in Lemke's method.

Table 1: Comparison of Algorithm 1 and the projection method ($n = 10$, $\rho = 1$).

	Algorithm 1			projection method	
δ	#Iterations	$\#f$	CPU	#Iterations	CPU
0.1	1380	9683	18.45	oscillation	
0.5	307	1527	3.04	oscillation	
1	328	1305	2.63	oscillation	
2	338	1009	2.13	oscillation	
3	353	951	2.04	oscillation	
4	342	685	1.55	oscillation	
5	256	511	1.16	oscillation	
6	351	701	1.72	oscillation	
6.2	337	674	1.66	oscillation	
6.3	377	754	1.91	9118	13.94
6.5	376	752	1.93	1594	2.43
7	385	770	2.07	610	0.93
8	337	676	1.98	338	0.51
9	270	542	1.59	271	0.43
10	242	487	1.41	244	0.36
12	232	468	1.36	229	0.34
15	239	488	1.41	239	0.35
20	254	636	1.72	272	0.41
50	229	920	2.17	549	0.79
100	229	1149	2.63	1036	1.47
200	239	1385	3.05	2008	2.88
500	239	1674	3.59	5007	7.08
1000	372	2605	5.55	9998	14.18

Table 2: Comparison of Algorithm 2 and Newton method ($\rho = 1$).

	Algorithm 2				Newton method		
n	#Iterations	$\#f$	#Lemke	CPU	#Iterations	#Lemke	CPU
30	5.6	7.6	80.6	4.294	8	115.2	5.840
50	5.6	7.6	156.0	19.880	8	216.4	26.142
90	6.0	8.0	275.2	105.400	8	358.8	135.690

Table 2 shows that the number of iterations of Newton method is consistently larger than that of Algorithm 2 as far as the test problems used in the experiments are concerned. Therefore, since it is time consuming to solve a linear subproblem at each iteration, Algorithm 2 required less CPU time than the pure Newton method in spite of the overheads in line search. Finally we note that, the pure Newton method (18) is not guaranteed to be globally convergent, although it actually converged for all test problems reported in Table 2.

4.3 Comparison of Algorithms 1 and 2

Finally we have compared Algorithms 1 and 2. For each of the problem sizes $n = 30, 50$ and 90, we randomly generated five test problems. To see how these algorithms behave for various degrees of asymmetry of the mapping F, we have tested several values of ρ between 0.1 and 2.0. The starting point was chosen to be $x = 0$. The results are given in Table 3. All numbers shown in Table 3 are averages of the results for five test problems for each case.

Table 3 shows that when the mapping F is close to symmetry, Algorithm 1 converges very fast, and when the mapping becomes asymmetric, the number of iterations and CPU time of Algorithm 1 increase rapidly. On the other hand, in Algorithm 2, while the total number of pivotings of Lemke's method increases in proportion to problem size n, the number of iterations stays constant even when the problem size and the degree of asymmetry of F are varied. Hence, when the degree of asymmetry of F is relatively small, that is, when ρ is smaller than 1.0 in our test problems, Algorithm 1 requires less CPU time than Algorithm 2.

Table 3: Comparison of Algorithms 1 and 2.

		Algorithm 1			Algorithm 2			
ρ	n	#Iterations	$\#f$	CPU	#Iterations	$\#f$	#Lemke	CPU
	30	31.6	96.0	0.672	6.0	8.4	87.4	4.956
0.1	50	28.4	84.6	0.966	6.0	9.2	148.0	19.050
	90	38.2	114.0	2.322	6.2	8.2	259.0	99.156
	30	40.0	119.6	0.838	6.2	8.2	92.6	4.872
0.2	50	37.6	110.0	1.264	6.4	8.8	162.6	20.888
	90	40.4	120.4	2.448	6.6	8.6	277.2	106.672
	30	33.8	99.2	0.690	6.4	8.4	90.8	4.812
0.3	50	39.2	112.2	1.292	6.2	8.6	157.8	20.270
	90	41.4	119.6	2.408	6.6	8.6	275.2	105.890
	30	45.8	127.0	0.892	6.0	8.2	88.2	4.662
0.5	50	58.6	161.8	1.848	6.0	8.0	157.8	20.242
	90	110.8	273.8	5.780	6.6	8.6	281.4	108.260
	30	152.8	322.0	2.436	5.8	7.8	84.4	4.474
0.8	50	290.4	586.4	7.450	6.0	8.0	160.4	20.476
	90	780.2	1557.4	33.960	6.0	8.0	269.4	103.266
	30	394.2	792.4	5.270	5.6	7.6	80.6	4.294
1.0	50	519.6	1077.6	11.068	5.6	7.6	156.0	19.880
	90	866.0	2129.6	35.880	6.0	8.0	275.2	105.400
	30	1197.0	3793.2	21.518	5.4	7.4	80.0	4.266
1.5	50	1604.0	4927.4	45.280	5.2	7.2	145.8	18.648
	90	2928.0	9777.8	158.162	6.0	8.0	275.6	105.492
	30	3195.2	12694.8	69.636	5.2	7.2	79.0	4.168
2.0	50	3842.6	15929.0	141.944	5.2	7.2	145.2	18.494
	90	4957.6	20905.0	332.510	5.8	7.8	279.0	106.514

Table 4: Results for Algorithm 2 (Gauss-Seidel version).

		Algorithm 2			
ρ	n	#Iterations	$\#f$	#Gauss	CPU
	30	6.0	8.4	39.2	0.274
0.1	50	6.0	9.2	39.6	0.466
	90	6.2	8.2	48.0	0.916
	30	3 of 5 failed			
0.2	50	3 of 5 failed			
	90	failed			
	30				
$0.3 \sim 2.0$	50	failed			
	90				

Note that, since the mapping F used in our computational experience is sparse, complexity of each iteration in Algorithm 1 is small. On the other hand, the code [8] of Lemke's method used in Algorithm 2 to solve a linear subproblem does not make use of sparsity, so that it requires a significant amount of CPU time at each iteration for large problems. If a method that can make use of sparsity is available to solve a linear subproblem, CPU time of Algorithm 2 may decrease. The projected Gauss-Seidel method [1, pp. 397] for solving linear complementarity problem is one of such methods. In Table 4, results of Algorithm 2 using the projected Gauss-Seidel method in place of Lemke's method are given. Table 4 shows that, if the mapping F is almost symmetric, Algorithm 2 converges very fast. But Algorithm 2 fails to converge when the degree of asymmetry increased, because the projected Gauss-Seidel method failed to solve linear subproblems.

Figure 1 illustrates how Algorithms 1 and 2 converged for two typical test problems with $n = 30$ and 50. In the figure, the vertical axis represents the accuracy of a generated point to the solution, which is evaluated by

$$\mathrm{ACC} = \max_i \{ |\min(x_i, F_i(x))| \mid i = 1, 2, \ldots, n \}. \tag{34}$$

Figure 1 indicates that Algorithm 2 is quadratically convergent near the solution. Figure 1 also indicates that Algorithm 1 is linearly convergent though it has not been proved

theoretically.

Acknowledgement The authors are grateful to Professor T. Ibaraki of Kyoto University for his helpful comments and support.

References

[1] Cottle, R. W., Pang, J. S. and Stone, R. E.: *The Linear Complementarity Problem,* Academic Press, San Diego, 1992.

[2] Dafermos, S.: Traffic Equilibrium and Variational Inequalities. *Transportation Science,* vol. 14 (1980), 42-54.

[3] Dennis, J. E. and Schnabel, R. B.: *Numerical Methods for Unconstrained Optimization and Nonlinear Equations,* Prentice-Hall, Eaglewood Cliffs NJ, 1983.

[4] Florian, M.: Mathematical Programming Applications in National, Regional and Urban Planning. *Mathematical Programming: Recent Developments and Applications,* Iri, M. and Tanabe, K. (eds.), KTK Scientific Publishers, Tokyo, 1989.

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

[6] Garcia, C. B. and Zangwill, W. I.: *Pathways to Solutions, Fixed Points, and Equilibria* Prentice-Hall, Eaglewood Cliffs NJ, 1981.

[7] Harker, P. T. and Pang, J. S.: Finite-Dimensional Variational Inequality and Nonlinear Complementarity Problems: A Survey of Theory, Algorithms and Applications. *Mathematical Programming,* vol. 48 (1990), 161-220.

[8] Ibaraki, T. and Fukushima, M.: *FORTRAN 77 Optimization Programming* (in Japanese), Iwanami-Shoten, Tokyo, 1991.

[9] Josephy, N. H.: Newton's Method for Generalized Equations. Technical Report No. 1965, Mathematics Research Center, University of Wisconsin, Madison, WI, 1979.

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

[11] Marcotte, P. and Dussault, J. P.: A Sequential Linear Programming Algorithm for Solving Monotone Variational Inequalities. *SIAM Journal on Control and Optimization,* vol. 27 (1989), 1260-1278.

[12] Mathiesen, L.: An Algorithm based on a Sequence of Linear Complementarity Problems Applied to a Walrasian Equilibrium Model : An Example. *Mathematical Programming,* vol. 37 (1987), 1-18.

[13] Pang, J. S. and Chan, D.: Iterative Methods for Variational and Complementarity Problems. *Mathematical Programming,* vol. 24 (1982), 284-313.

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

[15] Tobin, R. L.: A Variable Dimension Solution Approach for the General Spatial Price Equilibrium Problem. *Mathematical Programming*, vol. 40 (1988), 33-51.

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