Journal of Computational Mathematics, Vol.16, No.1, 1998, 40–50.

A PENALTY TECHNIQUE FOR NONLINEAR COMPLEMENTARITY PROBLEMS*[1]

Dong-hui Li Jin-ping Zeng
(*Department of Applied Mathematics, Hunan University, Changsha, China*)

Abstract

In this paper, we first give a new equivalent optimization form to nonlinear complementarity problems and then establish a damped Newton method in which penalty technique is used. The subproblems of the method are lower-dimensional linear complementarity problems. We prove that the algorithm converges globally for strongly monotone complementarity problems. Under certain conditions, the method possesses quadratic convergence. Few numerical results are also reported.

Key words: Optimization, nonlinear complementarity.

1. Introduction

Consider the following nonlinear complementarity problems NCP(F) of finding an $x \in R^n$, such that

$$x \geq 0,\ F(x) \geq 0 \text{ and } x^T F(x) = 0 \tag{1.1}$$

where F is a mapping from R^n into itself. It is an important form of the following variational inequality VI (F, X) of finding an $x \in X$, such that

$$(y - x)^T F(x) \geq 0, \quad \forall y \in X \tag{1.2}$$

where $X \subset R^n$ is a closed convex set. When $X = R^n_+$, (1.1) is equivalent to (1.2). NCP(F) and VI (F, X) can be transformed into optimization problem to be solved. So, many good techniques for solving optimization problems can be used. The first one may due to Marcotte and Dussault[6] who introduced a line search technique in the traditional linearized Newton method. A gap function was used as the merit function. When F is strongly monotone the algorithm converges globally and local quadratically. However, there is a disadvantage, i.e. the difficulty of the calculation of the merit function. In 1993, a new damped Newton method was established by Taji, Fukushima and Ibaraki[8] based on an equivalent differentiable optimization problem given by Fukushima[2]. The method still possesses global and local quadratic convergence if F is strongly monotone. The drawback of the method is that the merit function relies on a projective operator. Moreover, one has to estimate a positive definite matrix in practice.

* Received March 20, 1995.
1) This work was supported by the National Nature Science Foundation of China.

In both of the methods, the subproblems are linear complementarity problems of dimension n. In this paper, we will give another equivalent optimization of NCP(F) by using penalty technique. We also present a damped Newton method with the subproblems being lower-dimensional linear complementarity. Global convergence is obtained. For some special problems, local quadratic convergence is also established.

The paper is organized as follows: in the next section, we first deduce a new equivalent optimization problem of NCP(F) and then describe the algorithm. In section 3, we prove the global convergence and local quadratic convergence of the algorithm. At last, in section 4, we give some numerical results.

2. The Equivalent Form and the Algorithm

It is easy to see that NCP(F) is equivalent to the following optimization problem (e.g. see [4]):

$$\min f(x) = x^T F(x) \tag{2.1}$$

$$\text{s.t. } x \geq 0, \ F(x) \geq 0 \tag{2.2}$$

with the optimal $f(x^*) = 0$. Generally, the feasible domain $D = \{x \in R^n | x \geq 0, F(x) \geq 0\}$ is not convex. In 1992, Fukushima considered the merit function below

$$f(x) = -F(x)^T(H(x) - x) - \frac{1}{2}(H(x) - x)^T G(H(x) - x). \tag{2.3}$$

and cast NCP(F) as the following optimization problem

$$\min_{x \geq 0} f(x), \tag{2.4}$$

where $H(x) = \text{Proj}_G(x - G^{-1}F(x)))$,and $\text{Proj}_G(x)$ denotes the unique solution of the following mathematical programming:

$$\min_{y \geq 0} \|y - x\|_G = \{(y - x)^T G(y - x)\}^{1/2} \ .$$

Of course, the feasible domain (2.4) is convex. However, the calculation of $f(x)$ relies on the projective operator $H(x)$. To overcome these disadvantages, we give a new equivalent optimization problem of NCP(F).

Our approach follows the way of Fukushima$'$s. We consider the following mathematical programming problem

$$\min \phi_r(x) = x^T \max\{F(x), 0\} + \frac{1}{2} r \| \min\{F(x), 0\}\|^2 \tag{2.5}$$

$$\text{s.t. } x \geq 0 \tag{2.6}$$

Obviously, $\phi_r(x) = 0$ if and only if x solves NCP(F).

The function ϕ_r in (2.5) is not differentiable but directional differentiable. The derivative of ϕ_r at x along direction p is given by

$$\phi_r'(x,p) = \lim_{\alpha \to 0^+} \frac{1}{\alpha}[\phi_r(x + \alpha p) - \phi_r(x)] = p^T \max\{F(x), 0\}$$

$$+\sum_{F_i=0} x_i \max\{\nabla F_i^T p, 0\} + \sum_{F_i>0} x_i \nabla F_i^T p + r \sum_{F_i<0} F_i \nabla F_i^T p. \tag{2.7}$$

We wish to find a direction p to be a descent direction of ϕ_r. In the paper we choose p as the solution of the following lower-dimensional linear complementarity problem:

$$\begin{cases} x_i + p_i = 0, & \text{if } F_i(x) > 0, \quad (2.8)\\ x_i + p_i \geq 0,\ F_i(x) + \nabla F_i^T p \geq 0 & \\ \quad \text{and } (x_i + p_i)(F_i(x) + \nabla F_i^T p) = 0, & \text{if } F_i(x) \leq 0, \quad (2.9)\end{cases}$$

where x_i, p_i and $F_i(x)$ denote the i-th elements of x, p and $F(x)$ respectively. ∇F_i be the gradient of F_i at x. It is clear that if $p = 0$ is a solution of (2.8) and (2.9), then x is a solution of NCP(F).

In the rest of the paper, we assume that

Assumption (A). $F : R^n \to R^n$ is continuously differentiable and is strongly monotone, i.e. there is a constant $\mu > 0$ such that

$$[F(x) - F(y)]^T (x - y) \geq \mu \|x - y\|^2, \quad \forall x, y. \tag{2.10}$$

If we denote F' the Jacobian of F at x, then (2.10) is equivalent to

$$v^T F'(x) v \geq \mu \|v\|^2, \quad \forall x,\ v \in R^n. \tag{2.11}$$

To describe the algorithm, we first justify the descent property of ϕ_r.

Proposition 2.1. *Let assumption (A) hold. $x \geq 0$, p is determined by* (2.8) *and* (2.9). *If $r > 1/(2\mu)$, then*

$$\phi_r'(x, p) \leq -\frac{1}{2} p^T F'(x) p \leq -\frac{\mu}{2} \|p\|^2. \tag{2.12}$$

Proof. Notice that p satisfies (2.8), we have $\nabla F_i^T p \geq 0$ when $F_i(x) = 0$. Thus from (2.7) we deduce that

$$\begin{aligned}
\phi_r'(x,p) &= \sum_{F_i>0} p_i F_i + \sum_{F_i>0} x_i \nabla F_i^T p + \sum_{F_i=0} x_i \nabla F_i^T p + r \sum_{F_i<0} F_i \nabla F_i^T p \\
&= -\sum_{F_i>0} x_i F_i - \sum_{F_i>0} p_i \nabla F_i^T p + \sum_{F_i=0} x_i (F_i + \nabla F_i^T p) + r \sum_{F_i<0} (F_i + \nabla F_i^T p - F_i) F_i \\
&\leq -\sum_{F_i\geq 0} p_i \nabla F_i^T p - r \sum_{F_i<0} F_i^2 = -p^T F'(x) p + \sum_{F_i<0} p_i \nabla F_i^T p - r \|\min\{F(x), 0\}\|^2 \\
&= -p^T F'(x) p + \sum_{F_i<0} p_i (F_i + \nabla F_i^T p) - p^T \min\{F(x), 0\} - r \|\min\{F(x), 0\}\|^2 \\
&\leq -\frac{1}{2} p^T F'(x) p - \Big\{ \frac{1}{2} \frac{p^T F'(x) p}{\|p\|^2} \|p\|^2 + p^T \min\{F(x), 0\} + r \|\min\{F(x), 0\}\|^2 \Big\} \\
&= -\frac{1}{2} p^T F'(x) p - \frac{1}{2} \frac{p^T F'(x) p}{\|p\|^2} \Big\| p + \frac{\|p\|^2}{p^T F'(x) p} \min\{F(x), 0\} \Big\|^2 \\
&\quad - \Big(r - \frac{\|p\|^2}{2 p^T F'(x) p} \Big) \|\min\{F(x), 0\}\|^2
\end{aligned}$$

$$\le -\frac{1}{2}p^T F'(x)p - \Big(r - \frac{\|p\|^2}{2p^T F'(x)p}\Big)\|\min\{F(x),0\}\|^2$$
$$\le -\frac{1}{2}p^T F'(x)p - \Big(r - \frac{1}{2\mu}\Big)\|\min\{F(x),0\}\|^2.$$

If $r > \dfrac{1}{2\mu}$, then (2.12) holds true. Q.E.D.

Now, we state the damped Newton method.

Algorithm 1. Initial. Given constants $\rho, \sigma \in (0,1)$. Take $x^0 \in R^n_+$. $k \Longleftarrow 0$.

Step 1. Solve (2.8) and (2.9) for $x = x^k$ to get p^k.

Step 2. Select $\lambda_k = \rho^{m_k}$, where m_k is the smallest nonnegative integer satisfying that:

$$\phi_r(x^k + \rho^m p^k) - \phi_r(x^k) \le -\frac{1}{2}\sigma\lambda_k (p^k)^T F'(x^k)p^k. \tag{2.13}$$

Step 3. $x^{k+1} = x^k + \lambda_k p^k$, $k+1 \Rightarrow k$, go to Step 1.

3. Global and Local Convergence

In this section, we will prove the global and locally quadratic convergence of algorithm 1. First, we see that the sequence $\{x^k\}$ generated by algorithm 1 are in R^n_+ if $x^0 \in R^n_+$. The following lemma shows that $\{x^k\}$ is bounded.

Lemma 3.1. *For any $x^0 \in R^n_+$, every $r > 0$, if F is strongly monotone, then the level set*

$$\Omega_r = \{x | x \ge 0,\ \phi_r(x) \le \phi_r(x^0)\} \tag{3.1}$$

is bounded.

Proof. Assume that $\{u^k\}$ is an unbounded nonnegative sequence. Then for every $r > 0$, by a simple inequality that

$$\max(a,0) + \min(b,0) \le \max(a+b,0) \le \max(a,0) + \max(b,0) \tag{3.2}$$

we get

$$\begin{aligned}\phi_r(u^k) &\ge (u^k)^T \max\{F(u^k),0\} = (u^k)^T \max\{[F(u^k) - F(0)] + F(0), 0\}\\ &\ge \max\{(u^k)^T[F(u^k) - F(0)], 0\} + \min\{(u^k)^T F(0), 0\}\\ &\ge \mu\|u^k\|^2 - \|u^k\|\|F(0)\|,\end{aligned}$$

this implies that $\phi_r(u^k)$ tends to ∞. Thus Ω_r is bounded since $\phi_r(x) \le \phi_r(x^0)$ for all $x \in \Omega_r$. Q.E.D.

From lemma 3.1, it is easy to see that $\{x^k\}$ generated by algorithm 1 is bounded. So there is a subsequence $\{x^k\}$ with a limit $x^* \in \Omega_r$.

Let $I, J \subset Z_n = \{1, 2, \cdots, n\}$. We say I, J a partition of Z_n if $I \cap J = \emptyset$ and $I \cup J = Z_n$.

Lemma 3.2. *Let assumption (A) hold. Then the sequence $\{p^k\}$ is bounded.*

Proof. Since F is strongly monotone, for every partition I, J of Z_n, we claim that the matrix $G(I,J,x) \equiv ((e)_{i\in I},\ (\nabla F_j(x))_{j\in J})^T$ is uniformly nonsingular, that is there is a constant $\mu_1 > 0$ independent of I, J such that

$$\|G(I,J,x)v\| \ge \mu_1\|v\|,\ \forall v \in R^n,\ \forall x \in R^n_+$$

Recall that for every $i \in Z_n$, we have either

$$x_i^k + p_i^k = 0$$

or

$$F_i(x^k) + \nabla F_i(x^k)^T p^k = 0$$

That is to say, there is a partition I_k, J_k of Z_n such that

$$G(I_k, J_k, x^k)p^k = -H(x^k)$$

where $H(x^k) = (h_1(x^k), h_2(x^k), \cdots, h_n(x^k))$, $h_i(x^k) = x_i^k$ or $F_i(x^k)$. But $\{x^k\} \subset \Omega_r$ is bounded and $F(x)$ is continuous, we conclude the proof by the uniform nonsingularity of $G(I, J, x^k)$. Q.E.D.

The following lemma is useful for the proof of the global convergence of algorithm 1.

Theorem 3.3. *Let $F : R^n \to R^n$ be continuously differentiable. $\{x^k\}$ and $\{p^k\}$ are generated by algorithm 1. If there are subsequences $\{x^k\}_{k\in K}$ and $\{p^k\}_{k\in K}$ taking limits $\bar{x}$ and $\bar{p}$ respectively with $\bar{p} = 0$, then $\bar{x}$ is a solution of NCP(F).*

Proof. We verify the conclusion by partition the index set Z_n into three subsets. Define

$$\bar{\alpha} = \{i | F_i(\bar{x}) > 0\}, \ \bar{\beta} = \{i | F_i(\bar{x}) = 0\}, \ \bar{\gamma} = \{i | F_i(\bar{x}) < 0\}$$

We will show that $\bar{x}_i = 0$, $\forall i \in \bar{\alpha}$, $\bar{x}_i \geq 0$, $\forall i \in \beta$ and $\bar{\gamma}$ is empty. This means that $\bar{x}$ is a solution of NCP(F).

For $i \in \bar{\alpha}$, when $k \in K$ sufficiently large, $F_i(x^k) > 0$. Which implies from (2.8) that $x_i^k + p_i^k = 0$, $\forall i \in \bar{\alpha}$ and $k \in K$ sufficiently large. So we get that $\bar{x}_i = 0$, $\forall i \in \bar{\alpha}$.

For $i \in \bar{\gamma}$, it is clear that when $k \in K$ sufficiently large, (2.9) is always true for $x = x^k$ and $p = p^k$. Thus $F_i(x^k) + \nabla F_i^T(x^k)p^k \geq 0$ for all $i \in \bar{\gamma}$ and $k \in K$ sufficiently. Taking limits in the inequality, we get that $F_i(\bar{x}) \geq 0$, $\forall i \in \bar{\gamma}$. This contradiction means that $\bar{\gamma} = \emptyset$.

For the case $i \in \bar{\beta}$, we have from (2.8) and (2.9) that $x_i^k + p_i^k \geq 0$ for all $i \in \bar{\beta}$ and k. This, of course, implies that $\bar{x}_i \geq 0$, $\forall i \in \bar{\beta}$. Q.E.D.

Now we prove the global convergence of the algorithm 1.

Theorem 3.4. *Let assumption (A) hold. Let also $r > \dfrac{1}{2\mu}$. Then for any $x^0 \in R^n_+$, the sequence $\{x^k\}$ generated by the algorithm converges to the unique solution of NCP(F).*

Proof. Since the level set Ω_r is bounded, there exists a convergent subsequence $\{x^k\}_{k\in K}$. By lemma 3.3, we only need to find a subsequence $\{p^k\}_{k\in K'} \subset \{p^k\}_{k\in K}$ taking the limit $\bar{p} = 0$.

From the line search condition (2.13) and the monotonity of $\{\phi_r(x^k)\}$ we get

$$\lim_{k\to\infty} \lambda_k \|p^k\|^2 = 0. \tag{3.3}$$

Denote $\lambda^* = \inf\{\lambda_k | k \geq 0\}$. If $\lambda^* > 0$, then $p^k \to \bar{p} = 0$.

Now, we consider the case that $\lambda^* = 0$. Without loss of generality, we assume that $\{p^k\}_{k\in K} \to \bar{p}$. By the line search rule, it is clear that when $k \in K$ sufficiently large, $\lambda'_k \equiv \lambda_k/\rho$ does not satisfy (2.13). That is to say that when $k \in K$ sufficiently large,

$$\phi_r(x^k + \lambda'_k p^k) - \phi_r(x^k) > -\frac{1}{2}\sigma\lambda'_k (p^k)^T F'(x^k) p^k. \tag{3.4}$$

For convenience, in the following proof we omit the index k. Consider the left side of (3.4). Denote

$$\begin{aligned}\eta_i =&(x_i + \lambda' p_i)\max\{F_i(x+\lambda' p), 0\} - x_i \max\{F_i(x), 0\}\\ &+ \frac{1}{2}r[\min^2\{F_i(x+\lambda' p), 0\} - \min^2\{F_i(x), 0\}].\end{aligned}$$

Then

$$\phi_r(x^k + \lambda'_k p^k) - \phi_r(x^k) = \sum_{i=1}^n \eta_i. \tag{3.5}$$

Let $\bar{\alpha}, \bar{\beta}$ and $\bar{\gamma}$ be defined by lemma 3.3. If $i \in \bar{\alpha}$, then when $k \in K$ sufficiently large $F_i(x) > 0$, and $F_i(x + \lambda' p) > 0$. Thus

$$\eta_i = x_i[F_i(x+\lambda' p) - F_i(x)] + \lambda' p_i F_i(x+\lambda' p) = \lambda'[x_i \nabla F_i^T p + p_i F_i(x)] + o(\lambda')\ .$$

If $i \in \bar{\gamma}$, then when $k \in K$ sufficiently large $F_i(x) < 0$, and $F_i(x + \lambda' p) < 0$. Thus

$$\eta_i = \frac{1}{2}r[F_i(x+\lambda' p) + F_i(x)][F_i(x+\lambda' p) - F_i(x)] = \lambda' r F_i \nabla F_i^T p + o(\lambda')\ .$$

If $i \in \bar{\beta}$, from the inequality (3.2) and that

$$\min(a+b, 0) \geq \min(a, 0) + \min(b, 0),$$

we deduce that

$$\begin{aligned}\eta_i =&x_i[\max\{F_i(x+\lambda' p), 0\} - \max\{F_i(x), 0\}] + \lambda' p_i \max\{F_i(x+\lambda' p), 0\}\\ &+ \frac{1}{2}r[\min\{F_i(x+\lambda' p), 0\} + \min\{F(x), 0\}][\min\{F_i(x+\lambda' p), 0\} - \min\{F(x), 0\}]\\ \leq& x_i \max\{F_i(x+\lambda' p), 0\} - F_i(x), 0\} + \lambda' p_i \max\{F_i(x+\lambda' p), 0\}\\ &+ \frac{1}{2}r[\min\{F_i(x+\lambda' p), 0\} + \min\{F(x), 0\}] \min\{F_i(x+\lambda' p) - F(x), 0\}\\ =&\lambda'[x_i \max\{\nabla F_i^T p, 0\} + p_i \max\{F_i(x), 0\} + r\lambda' F_i(x) \min\{\nabla F_i^T(x) p, 0\} + o(\lambda')\ .\end{aligned}$$

Substitute all the above estimation to (3.5), we get that

$$\begin{aligned}\phi_r(x^k + \lambda'_k p^k) - \phi_r(x^k) \leq& \lambda'\Big\{\sum_{F_i(\bar{x})>0}[x_i \nabla F_i^T p + p_i F_i(x)] + r\sum_{F_i(\bar{x})<0} F_i(x) \nabla F_i^T p\\ &+ \sum_{F_i(\bar{x})=0}[x_i \max\{\nabla F_i^T p, 0\} + p_i \max\{F_i(x), 0\}\\ &+ r F_i(x) \min\{\nabla F_i^T p, 0\}\Big\} + o(\lambda').\end{aligned}$$

Using this inequality to (3.4) and dividing by λ'_k then taking limit as $k \in K$ and k tends to infinity, we deduce that

$$\begin{aligned}\sum_{F_i(\bar{x})>0} \Big[\bar{x}_i \nabla F_i^T(\bar{x})\bar{p} + \bar{p}_i F_i(\bar{x})\Big] + r \sum_{F_i(\bar{x})<0} F_i(\bar{x}) \nabla F_i^T(\bar{x})\bar{p} \\ + \sum_{F_i(\bar{x})=0} [\bar{x}_i \max\{\nabla F_i(\bar{x})^T \bar{p}, 0\} + \bar{p}_i \max\{F_i(\bar{x}), 0\} \\ + r F_i(\bar{x}) \min\{\nabla F_i^T(\bar{x})\bar{p}, 0\}\Big] \\ \geq -\frac{1}{2}\sigma \bar{p}^T F'(\bar{x})\bar{p}.\end{aligned}$$

But the left side of the above inequality is just $\phi'_r(\bar{x}, \bar{p})$. So by means of (2.12), this implies that

$$-\frac{1}{2}\bar{p}^T F'(\bar{x})\bar{p} \geq -\frac{1}{2}\sigma \bar{p}^T F'(\bar{x})\bar{p}.$$

From this we claim that $\bar{p} = 0$ since $\sigma \in (0, 1)$. Thus $\bar{x}$ solves NCP(F).

The above discussion has shown that there is an accumulation point of $\{x^k\}$ which solves NCP(F). Again by the monotonity of $\{\phi_r(x^k)\}$, every accumulation point of $\{x^k\}$ takes the same value $\phi_r(x^*) = 0$, i.e. a solution of NCP(F). However the strong monotonity of F guarantees the uniqueness of the solution. The proof is completed.

We now analyze the convergent rate of algorithm 1. From the proof of lemma 3.2 and theorem 3.4, we get $\{p^k\} \to 0$.

Theorem 3.5. *Let the conditions of theorem 3.3 hold. Let also that $F'(x)$ is Lipschitz continuous on Ω_r, i.e. there is a constant $L > 0$ such that*

$$\|F'(x) - F'(y)\| \leq L\|x - y\|, \quad \forall x, y \in \Omega_r. \tag{3.6}$$

Then there exists a constant $C > 0$ such that when k is sufficiently large

$$\|x^k + p^k - x^*\| \leq C\|x^k - x^*\|^2, \tag{3.7}$$

where x^* is the unique solution of NCP(F).

Proof. Set

$$\alpha^* = \{i | F_i(x^*) > 0\},\ \beta^* = \{i | F_i(x^*) = x_i^* = 0\},\ \gamma^* = \{i | x_i^* > 0\}. \tag{3.8}$$

Then $\alpha^* \cup \beta^* \cup \gamma^* = Z_n$. For all $i \in \alpha^*$, it follows that $x_i^* = 0$ and that $F_i(x^k) > 0$ when k is sufficiently large. Thus from (2.8), we get

$$x_i^k + p_i^k - x_i^* = 0, \quad \forall i \in \alpha^* \tag{3.10}$$

For every $i \in \beta^*$, we have either (3.10) or

$$F_i(x^k) - F_i(x^*) + \nabla F_i^T(x^k) p^k = 0. \tag{3.11}$$

For $i \in \gamma^*$, we get that $x_i^k + p_i^k > 0$ when k is sufficiently large, and so (3.11) holds.

On the other hand (3.11) can be rewritten as

$$\nabla F_i^T(x^k)(x^k + p^k - x^*) = -(F_i(x^k) - F_i(x^*) - \nabla F_i^T(x^k)(x^k - x^*)) \tag{3.12}$$

The above discussion shows that when k is sufficiently large

$$G_k(x^k + p^k - x^*) = -H_k, \tag{3.13}$$

where $G_k = (g_1^T(x^k), g_2^T(x^k), \cdots, g_n^T(x^k))$ with $g_i(x^k) = e_i$ or $\nabla F_i(x^k)$ and $H_k = (h_1^k, h_2^k, \cdots, h_n^k)$ with $h_i^k = 0$ or the right side of (3.12). Since F is strongly monotone, G_k is uniformly nonsingular. Moreover, (3.6) implies that there is a constant $C_1 > 0$ such that $\|H_k\| \le C_1 \|x^k - x^*\|^2$. Therefore we get (3.7) from (3.13). Q.E.D.

We now want to get the quadratic convergence.

Theorem 3.6. *Let the conditions of theorem* 3.4 *hold. Let also that there is a neighbourhood of* x^**, say* $N(x^*)$*, such that*

$$F_i(y) - F_i(x) \le \nabla F_i^T(x)(y - x), \ \forall i \in \gamma^* \text{ and } x, y \in N(x^*) \cap R_+^n. \tag{3.14}$$

If we take $\sigma < 1/2$ *in algorithm* 1*, then algorithm* 1 *has locally quadratically convergent property.*

Proof. From theorem 3.5, it suffices to verify that when k is sufficiently large, $\lambda_k \equiv 1$. In other words, we only need to verify that

$$\phi_r(x^k + p^k) - \phi_r(x^k) \le \frac{1}{2}\phi_r'(x^k, p^k) + o(\|p^k\|^2). \tag{3.15}$$

Then proposition 2.1 guarantees $\lambda_k = 1$. To prove (3.15), first we note that when k is sufficiently large, both x^k and $x^k + p^k$ are in $N(x^*) \cap R_+^n$. We rewrite (3.15) in another form by means of (2.7)

$$T_1 + T_2 \le o(\|p^k\|^2), \tag{3.16}$$

where

$$\begin{aligned} T_1 \equiv& \sum_{i=0}^n \varepsilon^k \equiv \sum_{i=0}^n [(x_i^k + p_i^k)\max\{F_i(x^k + p^k), 0\} - x_i^k \max\{F_i(x^k), 0\}] \\ &- \frac{1}{2} \sum_{F_i(x^k) \ge 0} [p_i^k F_i(x^k) + x_i^k \nabla F_i^T(x^k) p^k], \\ T_2 \equiv& \frac{1}{2} r \sum_{i=1}^n \bar{\varepsilon}_i^k \equiv \frac{1}{2} r \Big[\sum_{i=0}^n \min{}^2\{F_i(x^k + p^k), 0\} - \sum_{F_i(x^k) < 0} (F_i^2(x^k) + F_i(x^k)\nabla F_i^T(x^k) p^k)\Big], \end{aligned}$$

$$\varepsilon_i^k = \begin{cases} (x_i^k + p_i^k)\max\{F_i(x^k + p^k), 0\} - \frac{1}{2}(x_i^k + p_i^k) F_i(x^k) & \\ \qquad -\frac{1}{2} x_i^k (F_i(x^k) + \nabla F_i^T(x^k) p^k), & \text{if } F_i(x^k) \ge 0, \\ (x_i^k + p_i^k)\max\{F_i(x^k + p_i^k), 0\}, & \text{if } F_i(x^k) < 0 \end{cases}$$

$$\bar{\varepsilon}_i^k = \begin{cases} \min{}^2\{F_i(x^k + p^k), 0\}, & \text{if } F_i(x^k) \ge 0, \\ \min{}^2\{F_i(x^k + p^k), 0\} - F_i(x^k)(F_i(x^k) + \nabla F_i^T(x^k) p^k), & \text{if } F_i(x^k) < 0 \end{cases}$$

We now estimate T_1 and T_2. For convenience, in the later of the proof, we still omit the index k. For every $i \in \alpha^*$, it follows that $F_i(x) > 0$ when k is sufficiently large and thus $x_i + p_i = 0$. Therefore

$$\varepsilon_i = -\frac{1}{2}x_i(F_i(x) + \nabla F_i^T(x)p) \leq 0.$$

For $i \in \beta^*$, $x_i^* = F_i(x^*) = 0$. So

$$\varepsilon_i \leq (x_i + p_i - x^*)\max\{F_i(x+p), 0\} = o(\|p\|^2).$$

For $i \in \gamma^*$, $x_i^* > 0$. This implies that $x_i + p_i > 0$ when k is sufficiently large. So $F_i(x) \leq 0$ and $F_i(x) + \nabla F_i^T(x)p = 0$. In this case, by the assumption of the theorem we have

$$\varepsilon_i = (x_i + p_i)\max\{F_i(x+p) - F_i(x) - \nabla F_i^T(x)p, 0\} = 0.$$

We have now proved that when k is sufficiently large

$$T_1 \leq o(\|p\|^2). \tag{3.17}$$

We turn to estimate T_2. For $i \in \alpha^*$, when k is sufficiently large, $\bar{\varepsilon}_i = 0$. For $i \in \beta^* \cup \gamma^*$, $F_i(x^*) = 0$. From this we deduce that

$$F_i(x+p) = F_i(x+p) - F_i(x^*) = \nabla F_i^T(\tilde{x})(x+p-x^*) = o(\|p\|^2),$$

where $\tilde{x}$ is a point between $x+p$ and x^*. This follows that

$$\min{}^2\{F_i(x+p), 0\} = o(\|p\|^2).$$

So we have the estimation that

$$\bar{\varepsilon}_i = \begin{cases} o(\|p\|^2), & \text{if } F_i(x) \geq 0, \\ o(\|p\|^2) - (F_i(x) - F_i(x^*))(F_i(x) - F_i(x^*) + \nabla F_i^T(x)p), & \text{if } F_i(x) < 0. \end{cases}$$
$$= o(\|p\|^2).$$

Anyway, we have

$$T_2 = o(\|p\|^2). \tag{3.18}$$

The estimation (3.17) and (3.18) imply (3.16). The proof is completed. Q.E.D.

Remark 1. The condition (3.14) means that $\forall i \in \gamma^*$, F_i is concave in some neighbourhood of x^*. For some practical problems such as the piecewise linear elastic-plastic problem etc, these conditions are often satisfied.

Remark 2. For algorithm 1, we see that the descent property and global convergence of algorithm 1 rely on the requirement that $r \geq \dfrac{1}{2\mu}$. This is not convenient in practice since it is difficult to estimate μ in advance. To overcome such disadvantage, we can change r_k successively by obeying the following rule which very like the way used in [7] for constrained optimization problem. If

$$\phi_r'(x_k, p_k) \leq -\frac{1}{2}p_k^T F'(x_k)p_k, \tag{3.19}$$

then we need not change r_k, i.e. $r_{k+1} = r_k$. Otherwise we choose r_{k+1} by

$$r_{k+1} = \max\Big\{2r_k, \frac{\|p_k\|^2}{2p_k^T F'(x_k)p_k}\Big\}$$

The following theorem shows that the global and locally quadratic convergence results for the revised algorithm remain true.

Theorem 3.7. *Let assumption* (*A*) *hold. Then there exists an integer* k_0 *such that when* $k \geq r_0$, $r_k = r_{k_0}$.

Proof. By the rule for the determination of the penalty factor, it suffices to verify that there exists a constant $\eta > 0$ such that (3.19) holds for all $r \geq \eta$. This is just the conclusion of proposition 2.1. Q.E.D.

4. Numerical Results

In this section, we present our numerical results. Two test functions are considered.

Problem 1.

$$F(x) = C(x) + Ax + b,$$

where

$$C(x) = (c_1(x_1), c_2(x_2), \cdots c_n(x_n))^T, \quad c_i(x_i) = \text{arctan } (x_i), \quad i = 1, 2, \cdots, n$$

and

$$A = \begin{pmatrix} 2 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & \cdots & \cdots & & & \\ & & & \cdots & & & \\ & & & & \cdots & & -1 \\ & & & & & -2 & 2 \end{pmatrix},$$

$$b = (-n/2, -n/2+1, -n/2+2, \cdots, -n2+(n-1), -n/2+n)^T.$$

Problem 2. (Simplified as P2)

$$\begin{aligned} F_1(x) &= 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6, \\ F_2(x) &= 2x_1^2 + x_1 + x_2^2 + 3x_3 + 2x_4 - 2, \\ F_3(x) &= 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 3x_4 - 1, \\ F_4(x) &= x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3. \end{aligned}$$

By using $\|p^k\|_\infty \leq 10^{-6}$ as the stopping criterion, the iterative numbers are given in table 1 and table 2. Table 1 shows the iterative number of problem 1. In the table, Taji stands for the method proposed by Taji et al. (1993) with $G = I$ and LZ stands for the algorithm 2 in our paper. We choose the initial penalty factor $r_0 = 1$ in the revised algorithm.

Remark. In the two problems, F is strongly monotone for problem 1 but not for problem 2.

Table 1.

x^0		$(1, \cdots, 1)$	$(0, \cdots, 0)$	$(1, \cdots, n)$	$(n, \cdots, 1)$	$(10^4, \cdots, 10^4)$
$n = 5$	Taji	6	8	7	8	15
	LZ	4	4	4	5	5
$n = 10$	Taji	16	16	16	16	25
	LZ	9	10	9	11	11
$n = 20$	Taji	23	23	23	20	31
	LZ	16	15	15	17	17

Table 2.

x^0	$(1,1,1,1)$	$(10,20,30,40)$	$(1,0,0,0)$	$(1,0,1,0)$	$(10,10,10,10)$	$(10^4, \cdots, 10^4)$
P2	8	7	4	4	7	7

Acknowledgement. The authors thank Professor S.Z. Zhou and the referees for their helpful comments on the paper.

References

[1] M. Fukushima, A successive quadratic programming algorithm with global and superlinear convergence properties, *Mathematical Programming,* **35** (1986), 253–264.

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

[3] S.P. Han, A globally convergent method for nonlinear programming, *Mathematical Programming,* **11** (1976), 87–113.

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

[5] P. Marcotte, J.P. Dussault, A note on global convergent Newton method for solving variational inequalities, *Operations Research Letters,* **6** (1987), 35–42.

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

[7] M.J.P. Powell, Y. Yuan, A recursive quadratic programming algorithm that uses differentiable exact penalty functions, *Mathematical Programming,* **35** (1986), 265–278.

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

[9] K. Taji, M. Fukushima, A globally convergent Newton method for solving variational inequality problems with inequality constraints, In: Recent Advances in Nonsmooth Optimization (eds. D.-Z. Du, L. Qi and R. Womersley), World Scientific Publishing, Singapore-New Jersey-London-Hong Kong, 1995, 405–417.