© 1986 The Operations Research Society of Japan

Journal of the Operations Research Society of Japan Vol. 29, No. 4, December 1986

EXTENSION OF NEWTON AND QUASI-NEWTON METHODS TO SYSTEMS OF PC¹ EQUATIONS

Masakazu Kojima Tokyo Institute of Technology Susumu Shindo Sanno College

(Received March 24, 1986: Revised September 25, 1986)

Abstract This paper extends Newton and quasi-Newton methods to systems of PC^1 equations and establishes the quadratic convergence property of the extended Newton method and the Q-superlinear convergence property of the extended quasi-Newton method.

1. Introduction

The Newton method is one of the most popular and practical methods for solving systems of nonlinear equations. Let f be a C^1 (continuously differentiable) mapping from the n-dimensional Euclidean space R^n into itself. Let Df(x) denote the n X n Jacobian matrix, and ||x|| the Euclidean norm of a vector $x \in R^n$.

Algorithm N.

(The Newton method for a system of C^1 equations f(x) = 0). Step 0. Choose an initial point $x^0 \in \mathbb{R}^n$ and $p \leftarrow 0$. Step 1. Solve the system of linear equations, which we will call the Newton equation, in the variable vector $s \in \mathbb{R}^n$:

$$Df(x^p)s = -f(x^p).$$

Step 2. Set $x^{p+1} \leftarrow x^p + s$, $p \leftarrow p + 1$ and go to Step 1.

Let $z \in \mathbb{R}^n$ be a solution of f(x) = 0. It is well-known that if the initial point x^0 is sufficiently close to the solution z and Condition 1 below is satisfied, then the generated sequence { x^p } converges z quadratically, i.e., $\parallel x^{p+1} - z \parallel \leq c \parallel x^p - z \parallel^2$ for sufficiently large p, where c is a positive constant. (See, for example, Theorem 10.2.2 of Ortega and Rheinboldt [16].)

Condition 1.

(a) Df(z) is nonsingular.

(b) Df is Lipschitz continuous in an open convex neighborhood D of z, i.e., there is a Lipschitz constant K such that

 $\| Df(y) - Df(x) \| \leq K \| y - x \| \text{ for every } x, y \in D.$

Here the matrix norm of the left side of the inequality above is the operator norm defined by

 $|| A || = \max \{ || Ax || : x \in \mathbb{R}^{n}, || x || = 1 \}$ for every n X n matrix A.

Although the Newton method has the very nice local convergence property, it has one drawback. That is, we need to evaluate the n X n Jacobian matrix $Df(x^p)$ and solve the Newton equation, the system of linear equations with the coefficient matrix $Df(x^p)$ at each iteration. Solving the Newton equation generally requires $O(n^3)$ arithmetic operations. To avoid this disadvantage, the quasi-Newton methods have been proposed and studied extensively. See, for example, the survey paper Dennis and Moré [5]. In the present paper we shall be concerned with Broyden's quasi-Newton method [1] among the various quasi-Newton methods developed so far. Under certain assumptions, the sequence $\{x^p\}$ generated by the quasi-Newton method converges z Q-superlinearly, i.e., there is a sequence $\{\alpha_p\}$ which converges zero such that

 $\parallel x^{p+1} - z \parallel \leq \alpha_p \parallel x^p - z \parallel$

for every sufficiently large p. See, Dennis and More [5].

The Newton and the quasi-Newton methods are very useful and powerful tools for solving systems of nonlinear equations. However, their applications are restricted to the case where the mappings appeared in the systems are C^1 . On the other hand, many problems arising from the field of operations research and mathematical programming can be formulated as a system of nonlinear equations with a PC^1 (piecewisely continuously

Copyright © *by ORSJ. Unauthorized reproduction of this article is prohibited.*

differentiable) mapping on R^n . The nonlinear complementarity problem (see Example 1 in Section 2) and the Karush-Kuhn-Tucker (abbreviated by KKT below) stationary condition (see, for example, Mangasarian [14]) for inequality constrained nonlinear programs are such examples. We can not apply the Newton or the quasi-Newton methods directly to such problems. Here a PC^1 mapping on R^n is a continuous mapping whose restriction to each piece of a subdivision of R^n is continuously differentiable. More precise definition and some examples will be given in Section 2.

The purpose of this paper is to extend the Newton and the quasi-Newton methods to systems of PC¹ equations. There have been developed some extensions of the Newton and the quasi-Newton methods; Josephy [8,9] for strongly regular generalized equations (Robinson [19]), Pang and Chan [18] for variational inequalities including complementarity problems, In most of those extensions, an original system to be etc.. solved is approximated at an approximate solution by a locally linear but globally piecewise linear subproblem, and then the subproblem is solved to obtain a new approximate solution with a high accuracy. The sequence generated by repeating this process converges locally to a solution of the original system under certain assumptions. When we use those extensions, however, we may find it difficult to solve the piecewise linear subproblem generated. Even if it can be transformed into a linear complementarity problem on which many studies have been done, no unified computational method that can solve all linear complementarity problems efficiently has been developed. Furthermore solving a piecewise linear system usually requires more cost than solving a system of linear equations. We note that Murty [15] has given an example of a linear complementarity problem which requires well-known Lemke's method (Lemke [12]) to consume an exponential order of arithmetic operations.

There have been also developed several extensions of the quasi-Newton methods to approximate a KKT stationary solution (Han [7], Palomares and Mangasarian [17], etc.). In those papers the strict complementarity is assumed at the point to which the generated sequence converges. Under this assumption, the system

354

of PC¹ equations (see Kojima [10]) induced from the KKT condition is locally C¹ in a neighborhood of the solution corresponding to the KKT point. Hence we can apply the Newton and the quasi-Newton methods locally in the neighborhood of the From the theoretical point of view, the strict solution. complementarity assumption is moderate. In fact, almost all optimization problems satisfy the assumption (Fujiwara [6]). More generally, the assumption that the mapping F is C^1 in a neighborhood of a solution of a given system of PC^{1} equations F(x) = 0 may be mathematically moderate and legitimate. From the computational or numerical point of view, however, this fact is not enough to justify their application to PC¹ systems because a solution to be computed is likely to be very close to a common boundary of some different pieces and the generated sequence happens to oscillate between them.

We show in Section 3 that the Newton method with a slight extension works effectively on a system of PC^1 equations even if the solution we want to compute lies on a common boundary of different pieces. Section 4 is devoted to an extension of Broyden's quasi-Newton method to systems of PC^1 equations. Under a certain nonsingularity assumption we establish the quadratic convergence property on the extended Newton method (Theorem 1) and the Q-superlinear convergence property on the extended quasi-Newton method (Theorem 3), respectively. In Section 5, we present some numerical examples on the extended methods.

2. PC¹ Mappings

In this section we introduce a class of PC^1 mappings which we deal with in the remainder of the paper, and show two examples of systems of PC^1 equations. For each subset U of R^n , we employ the symbols cl U and int U to denote the closure of U and the interior of U, respectively.

Definition. Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a continuous mapping. F is a PC^1 mapping if there exists a countable family $\{U_i : i \in \Lambda\}$ of closed subsets of \mathbb{R}^n such that (a) cl (int U_i) = U_i for every $i \in \Lambda$,

- (b) (int U_i) \cap (int U_j) = ϕ whenever $i, j \in \Lambda$ and $i \neq j$, (e) $\bigcup_{i \in \Lambda} U_i^{-} = R^n$,
- (d) { U_i : $i \in \Lambda$ } has a locally finite property, i.e., for any $x \in \mathbb{R}^n$, there exists an open neighborhood V of x such that {i; $V \cap U_i \neq \phi$ } is finite,
- (e) for each $i \in \Lambda$ the restriction $F \mid U_i$ of the mapping to each U_i is a C^1 mapping. More precisely, there exists C^1 mapping f_i from an open neighborhood of U_i into R^n such that $F(x) = f_i(x)$ for any $x \in U_i$.

We call the family $\{U_i : i \in \Lambda\}$ a subdivision of \mathbb{R}^n , and each U_i a piece. So we say that F is PC^1 on a subdivision $\{U_i : i \in \Lambda\}$ of \mathbb{R}^n . For simplicity of discussions, we shall assume throughout the paper that f, is defined on the whole space R^n for each $i \in \Lambda$.

Example 1. (A Nonlinear Complementarity Problem). Let $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}^n$ be a \mathbb{C}^1 mapping. A nonlinear complementarity problem (NCP) is a problem of finding an $x \in \mathbb{R}^n$ such that $x \ge 0$, $f(x) \ge 0$ and $x^T f(x) = 0$.

In order to convert NCP into a system of PC¹ equations, we need the following symbols:

$$a^{+} = \max \{ a, 0 \} \text{ for each } a \in R,$$

$$a^{-} = \min \{ a, 0 \} \text{ for each } a \in R,$$

$$x^{+} = (x_{1}^{+}, \dots, x_{n}^{+}) \text{ for each } x \in R^{n},$$

$$x^{-} = (x_{1}^{-}, \dots, x_{n}^{-}) \text{ for each } x \in R^{n}.$$

We define the mapping F from R^n into itself by

$$F(y) = f(y^{\dagger}) + y^{-}.$$

Then F is a PC¹ mapping on the orthant subdivision. It is easily verified that there is one-to-one correspondence between a solution x of NCP and a solution y of the system of PC¹ equations F(y) = 0 through the transformation $y \rightarrow x = y^+$.

Example 2. For each $x \in R$, let

$$f_{1}(x) = x^{2} - 2x,$$

$$f_{2}(x) = x^{2} + 2x,$$

$$F(x) = \max \{f_{1}(x), f_{2}(x)\}.$$

Then the mapping $F: R \rightarrow R$ is PC^1 on the subdivision consisting of the two pieces

 $U_1 = (-\infty, 0]$ and $U_2 = [0, +\infty)$.

3. Extension of the Newton method to Systems of PC^1 Equations

Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a \mathbb{PC}^1 mapping on a subdivision $\{U_i ; i \in \Lambda\}$. When we apply the Newton method to the system F(x) = 0, we evaluate the values of the mapping F and the Jacobian matrix at the p-th iteration x^p . From the definition of \mathbb{PC}^1 mappings, we see that the Jacobian matrix of F depends not only on the point x^p but also on the piece U_i on which the point x^p lies. Hence before constructing the Newton equation, we need to choose a piece U_i on which x^p lies. This observation leads to:

Algorithm EN.

(The extended Newton method for a system of PC^1 equations F(x) = 0).

Step 0. Choose an $x^0 \in \mathbb{R}^n$ and set $p \leftarrow 0$.

Step 1. Choose a piece U_i that contains x^p . Solve the Newton equation in the variable vector $s \in \mathbb{R}^n$:

 $Df_{i}(x^{p})s = -f_{i}(x^{p}).$

Step 2. Set $x^{p+1} \leftarrow x^p + s$, $p \leftarrow p+1$, and go to Step 1.

Obviously Algorithm EN coincides with Algorithm N if F is C^1 on the whole space R^n . When x^p is contained in more than one piece, we may choose any one of those pieces for U_i . Let

 $I(z) = \{ i \in \Lambda : z \in U_i \} \text{ for each } z \in \mathbb{R}^n, \\ B_{\gamma}(z) = \{ x \in \mathbb{R}^n : \| x - z \| \leq \gamma \}$

for each positive number γ . To present the local convergence property we need the condition and the lemma below.

Condition 2.

- (a) For any $i \in I(z)$, $Df_i(z)$ is nonsingular.
- (b) For any $i \in I(z)$, Df_i is Lipschitz continuous in some open convex neighborhood D_i of z, i.e., there is a Lipschitz

constant K_i such that $\| Df_i(y) - Df_i(x) \| \leq K_i \| y - x \|$ for any $x, y \in D_i$.

Lemma 1. Let z be a solution of the system of PC^1 equations F(x) = 0. Suppose that Condition 2 holds. Let $K = \max \{K_i : i \in I(z)\}$. Then there exists a positive number γ such that (A) $B_{\gamma}(z) \subset int \bigcup \{ \bigcup_i \cap D_i : i \in I(z) \}$, (B) z is a unique solution of F(x) = 0 in $B_{\gamma}(z)$, (C) $Df_i(x)$ is nonsingular for any $x \in B_{\gamma}(z)$ and any $i \in I(z)$, (D) $\parallel Df_i(y) - Df_i(x) \parallel \leq K \parallel y - x \parallel$ for any x, $y \in B_{\gamma}(z)$ and any $i \in I(z)$, (E) $\parallel f_i(y) - f_i(x) - Df_i(z)(y-x) \parallel$ $\leq K \max \{ \parallel y - z \parallel, \parallel x - z \parallel \} \parallel y - x \parallel$ for any x, $y \in B_{\gamma}(z)$ and any $i \in I(z)$, (F) $\parallel f_i(y) - f_i(x) - Df_i(x)(y-x) \parallel \leq (K/2) \parallel y - x \parallel^2$ for any x, $y \in B_{\gamma}(z)$ and any $i \in I(z)$.

Proof: The existence of a positive number γ for which (A) through (D) hold can be easily verified. (E) follows from (D) and Corollary 3.2.5 of Ortega and Rheinboldt [16]. (F) follows from (D) and Theorem 3.2.12 of [16]. The details are omitted here.

Remark. The assumption (a) of Condition 2 does not guarantee that F is locally one-to-one in a neighborhood of z. See Example 2 in Section 2.

Theorem 1. Let z be a solution of the system of PC^1 equations F(x) = 0. Suppose that Condition 2 holds. Then there exists a positive number γ such that if $||x^0 - z|| \leq \gamma$, then the sequence $\{x^p\}$ generated by Algorithm EN converges z quadratically.

<u>Proof</u>: Let γ be the positive number whose existence is ensured by Lemma 1. We see from (C) of Lemma 1 and the compactness of $B_r(z)$ that L = sup { $\| Df_i^{-1}(x) \| : x \in B_{\gamma}(z)$ and $i \in I(z)$ } is positive and finite. Choosing γ sufficiently small if necessary, we may assume (KL/2) $\gamma < 1$. We shall show that if $x^p \in B_{\gamma}(z)$ then

and

 $\| x^{p+1} - z \| \leq (KL/2) \| x^p - z \|^2$.

hold. Then the conclusion of the theorem follows immediately. Let $U_i \ni x^p$ be the piece that has been chosen at Step 1. Since x^p lies in $B_{\gamma}(z)$, the Newton equation can be solved consistently and we have

$$x^{p+1} = x^p - Df_i^{-1}(x^p) f_i(x^p).$$

Hence

$$\| x^{P+1} - z \|$$

$$\leq \| x^{P} - z - Df_{i}^{-1}(x^{P})f_{i}(x^{P}) \|$$

$$\leq \| Df_{i}^{-1}(x^{P}) \| \| f_{i}(x^{P}) - Df_{i}(x^{P})(x^{P} - z) \|$$

$$= \| Df_{i}^{-1}(x^{P}) \| \| f_{i}(x^{P}) - f_{i}(z) - Df_{i}(x^{P})(x^{P} - z) \|$$

$$(since f_{i}(z) = 0)$$

$$\leq (LK/2) \| x^{P} - z \|^{2}$$

$$(by (F) \text{ of Lemma 1})$$

$$\langle \gamma . (since (KL/2) \gamma \langle 1 \text{ and } x^{P} \in B_{\tau}(z)) (Q. E. D)$$

4. Extended Quasi-Newton Method

 $x^{p+1} \in B_r(z)$

In this section, we first present Broyden's quasi-Newton method [1] for solving a system of C^1 equations and its Q-superlinear convergence property. Then we extend the method and the property to the PC^1 case.

Algorithm QN.

(The Broyden method for a system of C^1 equations f(x) = 0). Step 0. Choose an $x^0 \in \mathbb{R}^n$ and an n X n nonsingular matrix A_0 . Set $p \leftarrow 0$.

Step 1. Solve the system of linear equations in the variable vector $s_n \in \mathbb{R}^n$:

$$A_p s_p = -f(x^p)$$

Set

$$\begin{aligned} x^{p+1} \leftarrow x^{p} + s_{p}, \\ u_{p} \leftarrow f(x^{p+1}) - f(x^{p}), \\ A_{p+1} \leftarrow A_{p} + (u_{p} - A_{p}s_{p})s_{p}^{T} \neq || s_{p} ||^{2}. \end{aligned}$$

Step 2. Set $p \leftarrow p+1$ and go to Step 1.

The theorem below states that if an x^0 and an A_0 are good approximations of a solution z of f(x) = 0 and the Jacobian matrix $Df(x^0)$, respectively, then the sequence $\{x^p\}$ generated by Algorithm QN converges z Q-superlinearly.

Theorem 2. (Corollary 5 of Dennis [3], and Theorem 5.2 of Dennis and Moré [5]) Let z be a solution of f(x) = 0. Assume that $f : \mathbb{R}^n \to \mathbb{R}^n$ satisfies Condition 1. There exist positive numbers γ and δ such that if $|| x^0 - z || \leq \gamma$ and $|| A_0 - Df(x^0) || \leq \delta$, then the sequence $\{x^p\}$ generated by Algorithm QN converges z Q-superlinearly.

Now we are in a position to present an extension of Algorithm QN to the system of PC^1 equations. Let $F: \mathbb{R}^n \to \mathbb{R}^n$ be a PC^1 mapping on a subdivision $\{U_i; i \in \Lambda\}$. We will denote a sequence generated by the extended algorithm by $\{y^q: q=0, 1, 2...\}$. It generally traverses more than one piece. When it visits a piece U_i at a point y^q for the first time, we need to give an initial approximation of the Jacobian matrix $Df_i(y^q)$. For this purpose, we introduce an index subset J of Λ and a rule

 $(i, y) \in \Lambda \times U_i \rightarrow C(i, y),$ where C(i, y) is an n×n matrix. We initialize $J = \phi$. When a piece U_i is visited for the first time we update $J \leftarrow J \cup \{i\}$; the set J consists of the indices of the pieces which have been visited. If $y^q \in U_i$ and $i \notin J$ at the beginning of the q-th iteration, we see that the piece U_i has never visited before. In this case we assign the matrix $C(i, y^q)$ to an initial approximation of the Jacobian matrix $Df_i(y^q)$. Letting $x^0 = y^q, A_0 = C(i, y^q)$ and $f = f_i$, we then call Algorithm QN as a subroutine and apply its iteration to the system of C^1 equations $f_i(x) = 0$ as long as the generated point stays inside the

360

piece U_i. When the generated sequence leaves the piece U_i, Algorithm QN outputs a point $x^{r} \notin U_{i}$ and a matrix $A_{r'}$ which can be regarded as an approximation of the Jacobian matrix $Df_{i}(x^{r})$. We set

$$y^{q+p} = x^p$$
 (p=1, 2, ... r),
 $q = q + r$,

and store the matrix A_r into M(i) so that we can utilize this matrix when the sequence $\{y^q\}$ revisits the piece U_i . This process is summarized as follows:

Algorithm EQN.

(The extended Broyden method for a system of PC^1 equations F(y) = 0).

Step 0. Choose a $y^0 \in \mathbb{R}^n$. Set $J \leftarrow \phi$, and $q \leftarrow 0$.

Step 1. Find a piece U_i such that $y^q \in U_i$. If $i \in J$ then let $B_q \leftarrow M(i)$. Otherwise let $B_q \leftarrow C(i, y^q)$,

$$J \leftarrow J \cup \{i\}$$
 and $M(i) \leftarrow B_q$. Set

$$f \leftarrow f_i, x^0 \leftarrow y^q, A_0 \leftarrow B_a, p \leftarrow 0.$$

Step 2. Execute Algorithm QN and set

 $y^{q+p} \leftarrow x^p$,

$$B_{a+p} \leftarrow A_p \text{ if } x^p \in U_i$$

until $x^{p} \notin U_{i}$. If the execution terminates at the *r*-th iteration after generating $x^{r} \notin U_{i}$ then go to Step 3. Step 3. Let $M(i) = A_{r}$. Set $q \leftarrow q + r$ and go to Step 1.

See Figure 1 for an execution of Algorithm EQN.

We need a series of lemmas to establish the generalization of Theorem 2.

Lemma 2. (Broyden [2]). Let *I* and *B* be the $n \times n$ identity matrix and any $n \times n$ matrix, respectively, and $s \in \mathbb{R}^{n}$ such that || s || = 1. Then

 $|| B(I - ss^{\mathrm{T}}) || \leq || B || .$

Let $|| A ||_F$ denote the Frobenius norm of an n×n matrix A, i.e., $|| A ||_F = tr(A^T A)$, where trB is the summation of the diagonal elements of a matrix B.

Figure 1. An execution of Algorithm EQN

 $\{ y^{q} \} : \text{the sequence generated by Algorithm EQN}$ $(y^{q}): q \in Q_{i}$ $A_{q+1} \text{ is stored in } M(i) \text{ since } y^{p} \in U_{i}$ $for \ p = 0, 1, \dots, q \text{ and } y^{q+1} \notin U_{i}.$ $M(i) \text{ is used as an initial matrix since } y^{r}$ $enters \ U_{i} \text{ again.}$

Lemma 3.

(A)
$$||AB||_{F} \leq \min \{ ||A|| ||B||_{F}, ||A||_{F} ||B|| \}$$

for any n×n matrices A and B,
(B) $||E||_{F} \leq n^{1/2} ||E||$ for any n×n matrix E,
(C) $||E[I - ss^{T} / ||s||^{2}] ||_{F}^{2} = ||E||_{F}^{2} - ||Es||^{2} / ||s||^{2}$
for any n×n matrix E and any nonzero vector $s \in R^{n}$,
(D) $||E[I - ss^{T} / ||s||^{2}] ||_{F}$
 $\leq ||E||_{F} - (2 ||E||_{F})^{-1} (||Es|| / ||s||)^{2}$
for any n×n nonzero matrix E and any nonzero vector $s \in R^{n}$.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

Proof:

(A) Let b_j be the *j*-th column vector of B (j = 1, 2, ..., n). Then $|| AB ||_{F}^2 = tr(B^T A^T AB)$

$$= \sum_{j=1}^{n} b_{j}^{\mathrm{T}} A^{\mathrm{T}} A b_{j}$$

$$\leq \|A\|^{2} \sum_{j=1}^{n} b_{j}^{\mathrm{T}} b_{j}^{\mathrm{T}}$$

$$= \|A\|^{2} \|B\|_{\mathrm{F}}^{2}.$$

Thus $||AB||_{F} \leq ||A|| ||B||_{F}$. By interchanging the roles of A and B, we also have

 $||AB||_{F} \leq ||A||_{F} ||B||.$

(B) It is easy to see that $|| I ||_F = n^{1/2}$. Hence (B) follows from (A) by setting A = I and B = E. For (C) and (D), see the proof of Theorem 5.2 of Dennis and Moré [5]. (Q.E.D)

Lemma 4. (Banach Lemma. Theorem 2.3.2 of Ortega and Rheinboldt [16]). Let A and C be n X n matrices and assume that A^{-1} exists and $||A^{-1}|| \leq \alpha$. If $||A - C|| \leq \beta$ and $\alpha \beta < 1$, then C^{-1} exists and $||C^{-1}|| \leq \alpha / (1 - \alpha \beta)$.

Theorem 3. Let z be a solution of F(y) = 0. Suppose that Condition 2 holds. Then there exist positive numbers γ and δ such that the sequence $\{y^q\}$ generated by Algorithm EQN converges z Q-superlinearly whenever

 $(4.1) \parallel y^0 - z \parallel \leq \gamma$ and

(4.2) $\| C(i, y) - Df_{i}(y) \| \leq \delta$

for any $y \in B_{\gamma}(z) \cap U_{i}$ and $i \in I(z)$. **Proof**: Choose a positive number γ satisfying (A)-(F) of Lemma 1. By (C) of Lemma 1 and the compactness of $B_{\gamma}(z)$, we see that

L = sup { $\| Df_i^{-1}(y) \|$; $y \in B_{\gamma}(z)$ and $i \in I(z)$ } is positive and finite. Let $\delta \leq 1/(7L)$. Taking γ sufficiently small if necessary, we may assume that $\gamma \leq (\delta/5K)$, where K = max { K_i ; $i \in I(z)$ }. Suppose that (4.1) and (4.2) hold for these γ and δ . (i) We first prove by induction that (4.3) $\| y^{q} - z \| \leq (1/2) \| y^{q-1} - z \|$ if $q \geq 1$, (4.4) $\| y^{q} - z \| \leq (1/2)^{q} \| y^{0} - z \|$,

and

(4.5)
$$\| B_q - Df_k(y^q) \| \leq (2 - 2^{-q}) \delta$$

if $U_k \ni y^q$ is the piece that has been chosen at Step 1, hold for q = 0, 1, The inequality (4.4) implies that the sequence $\{y^q\}$ converges z. By the assumption, these inequalities hold for q = 0. We assume that the inequalities hold for q = 0, 1, ..., r, and deal with the case q = r+1. By the inequality

$$||y^{r} - z|| \leq (1/2)^{r} ||y^{0} - z|| \leq \gamma$$

and (A) of Lemma 1, $U_i \ni y^r$ is chosen at Step 1 for some $i \in I(z)$. Since the inequalities

 $\| Df_{j}^{-1}(y^{r}) \| \leq L \quad (by \text{ the definition of } L),$ $(2 - 2^{-r}) \delta L \leq 2 \delta L \langle 1 \quad (by \text{ the definition of } \delta),$

and the inequality (4.5) for q = r hold, Lemma 4 ensures that B_r is a nonsingular matrix whose inverse satisfies

(4.6)
$$|| B_r^{-1} || \leq L/(1 - 2 \delta L).$$

Using this inequality we have

$$\| y^{r+1} - z \|$$

$$= \| y^{r} - z - B_{r}^{-1} f_{i}(y^{r}) \| \text{ (since } y^{r+1} = y^{r} - B_{r}^{-1} f_{i}(y^{r}))$$

$$\leq \| B_{r}^{-1} \| \| f_{i}(y^{r}) - f_{i}(z) - B_{r}(y^{r} - z) \|$$

$$\text{ (since } f_{i}(z) = 0)$$

$$\leq L / (1 - 2 \delta L) \| f_{i}(y^{r}) - f_{i}(z) - B_{r}(y^{r} - z) \| \text{ (by (6))}$$

$$\leq L / (1 - 2 \delta L) \{ \| f_{i}(y^{r}) - f_{i}(z) - Df_{i}(y^{r})(y_{r} - z) \|$$

$$+ \| Df_{i}(y^{r}) - B_{r} \| \| y^{r} - z \| \}$$

$$\leq L / (1 - 2 \delta L) \{ (K/2) \| y^{r} - z \| ^{2}$$

$$+ (2 - 2^{-r}) \delta \| y^{r} - z \| \}$$

(by (F) of Lemma 1 and the induction hypothesis)

Copyright © *by ORSJ. Unauthorized reproduction of this article is prohibited.*

364

$$\leq L/(1 - 2\delta L) \{ (K\gamma/2) + 2\delta \} \parallel y^{r} - z \parallel$$
(since $\parallel y^{r} - z \parallel \leq \gamma$)
$$\leq L/(1 - 2\delta L) \{ \delta/10 + 2\delta \} \parallel y^{r} - z \parallel$$
(since $\gamma \leq \delta/(5K)$)
$$\leq (21/50) \parallel y^{r} - z \parallel$$
(since $\delta \leq 1/(7L)$)
$$< (1/2) \parallel y^{r} - z \parallel .$$

Thus we have shown that the inequality (4.3) holds for q = r+1. It follows immediately from the induction hypothesis that (4.4) holds for q = r+1.

Now we show that the inequality (4.5) holds for q = r+1. We have to consider the three cases:

Case (a) $y^{r+1} \in U_i$. Case (b) $y^{r+1} \in U_j \neq U_i$ and $j \notin J$. Case (c) $y^{r+1} \in U_j \neq U_i$ and $j \in J$.

Suppose that Case (a) occurs. In this case we have

$$B_{r+1} = B_r + (u_r - B_r s_r) s_r^T / || s_r ||^2,$$

where $s_r = y^{r+1} - y^r$ and $u_r = f_i (y^{r+1}) - f_i (y^r).$

Hence

$$\| B_{r+1} - Df_{i}(y^{r+1}) \|$$

$$= \| B_{r} - Df_{i}(y^{r+1}) + (u_{r} - B_{r}s_{r})s_{r}^{T} / \| s_{r} \|^{2} \|$$

$$\leq \| \{ B_{r} - Df_{i}(y^{r}) \} (I - s_{r}s_{r}^{T} / \| s_{r} \|^{2}) \|$$

$$+ \| u_{r} - Df_{i}(y^{r})s_{r} \| / \| s_{r} \|$$

$$+ \| Df_{i}(y^{r+1}) - Df_{i}(y^{r}) \|$$

$$\leq \| B_{r} - Df_{i}(y^{r}) \| + (K/2) \| y^{r+1} - y^{r} \|$$

$$+ K \| y^{r+1} - y^{r} \|$$

$$(by \text{ Lemma 2, (D) and (F) of Lemma 1)} \le (2 - 2^{-r}) \delta + (3K/2) \{ \parallel y^{r+1} - z \parallel + \parallel y^r - z \parallel \}$$

$$(by \text{ the induction hypothesis}) \le (2 - 2^{-r}) \delta + (9K/4) \parallel y^r - z \parallel (by (4.3))$$

$$\le (2 - 2^{-r}) \delta + (9K/4) (1/2)^r \gamma (by (4.4))$$

$$\leq (2 - 2^{-r}) \, \delta + 2^{-(r+1)} \, \delta \quad (\text{since } K \, r \leq \delta \, / 5)$$
$$= (2 - 2^{-(r+1)}) \, \delta \, .$$

If Case (b) occurs then the sequence $\{y^q\}$ has visited the piece U_j for the first time. In this case we assign the matrix $C(j, y^{r+1})$ to B_{r+1} . Since $y^{r+1} \in B_{\gamma}(z)$, the desired result follows from the assumption of the theorem.

Now suppose that Case (c) occurs. Then we assign the matrix which has been stored in M(j) at the q-th iterate (q < r) to B_{r+1} . Let \tilde{B}_q denote the matrix assigned. Since the matrix \tilde{B}_q has been generated by

$$\tilde{B}_{q} = B_{q-1} + (u_{q-1} - B_{q-1}s_{q-1})s_{q-1}^{T} / || s_{q-1} ||^{2},$$

by the similar argument as in the case (a), we see (4.7) $\|\tilde{B}_q - Df_j(y^q)\| \leq (2 - 2^{-q}) \delta$. Hence we have

$$\| B_{r+1} - Df_{j}(y^{r+1}) \|$$

$$= \| \tilde{B}_{q} - Df_{j}(y^{r+1}) \|$$

$$\leq \| \tilde{B}_{q} - Df_{j}(y^{q}) \| + \| Df_{j}(y^{q}) - Df_{j}(y^{r+1}) \|$$

$$\leq (2 - 2^{-q}) \delta + K \| y^{q} - y^{r+1} \|$$

$$(by (D) \text{ of Lemma 1 and } (4.7))$$

$$\leq (2 - 2^{-q}) \delta + K \{ \| y^{q} - z \| + \| y^{r+1} - z \| \}$$

$$\leq (2 - 2^{-q}) \delta + K \{ 2^{-q} \gamma + 2^{-r-1} \gamma \}$$

$$(since the inequality (4.4) holds for q \leq$$

$$\leq (2 - 2^{-q}) \delta + (1/5) \delta \{ 2 \times 2^{-q} \}$$

$$(since \gamma \leq \delta / (5K) \text{ and } q < r)$$

$$\leq (2 - 2^{-(r+1)}) \delta . \quad (since q < r)$$

r+1)

Thus we have shown that (4.5) holds for q = r+1.

(ii) Let $I^*(z)$ be the set of the indices $i \in I(z)$ such that the sequence $\{y^q\}$ visits U_i infinitely many times. Let $i \in I^*(z)$ be fixed, and Q_i be the subsequence of the numbers q such that $y^q \in U_i$. We prove that

(4.8)
$$\lim_{q \to \infty} \|y^{q+1} - z\| \neq \|y^{q} - z\| = 0.$$

Define

$$\begin{split} & \widetilde{B}_q = B_q \quad \text{if } q \in Q_i, \\ & \widetilde{B}_q = B_{q-1} + (u_{q-1} - B_{q-1}s_{q-1}) s_{q-1}^T / \parallel s_{q-1} \parallel^2, \\ & \quad \text{if } q-1 \in Q_i \text{ and } q \notin Q_i. \end{split}$$

Note that in the latter case \tilde{B}_q is the matrix which is stored in M(i); hence it coincides with the matrix \tilde{B}_r where r is the successor of q along the subsequence Q_i . (See Figure 1.) Then we have for every $q \in Q_i$,

(4.9)
$$\widetilde{B}_{q}(y^{q+1} - y^{q}) = -f_{i}(y^{q})$$

(4.10) $\widetilde{B}_{q+1} = \widetilde{B}_{q} + (u_{q} - \widetilde{B}_{q}s_{q})s_{q}^{T} || s_{q} ||^{2}$,
For each $q \in Q_{i}$, let
 $\widehat{Q}_{q} = -\frac{||}{2} = -\frac{2}{2} \int f_{q}(q) ||_{q}$

$$\begin{aligned} \theta_{q} &= \| \theta_{q} - Df_{i}(z) \|_{F'} \\ \phi_{q} &= \| (\tilde{B}_{q} - Df_{i}(z)) (y^{q+1} - y^{q}) \| \times \| y^{q+1} - y^{q} \| . \end{aligned}$$

Then there exists a positive number θ such that

$$\theta_q \leq \theta$$
 for every $q \in Q_i$.

In fact we see

$$\theta_q \leq n^{1/2} \| \tilde{B}_q - Df_i(z) \| \quad (by (B) \text{ of Lemma 3})$$

$$\leq n^{1/2} \| B_q - Df_i(y^q) \| + n^{1/2} \| Df_i(y^q) - Df_i(z) \|$$

$$\leq 2n^{1/2} \delta + n^{1/2} K \gamma \qquad (by (D) \text{ of Lemma 1})$$

$$\leq 11n^{1/2} \delta / 5. \qquad (\text{since } K \gamma \leq \delta / 5)$$

On the other hand we have

Thus we obtain

$$(4.11) \qquad (2\theta)^{-1} \sum_{q \in Q_i} \varphi_q^2$$

$$\leq \sum_{q \in Q_i} \varphi_q^{(\theta)} \varphi_q^{-\theta} \varphi_{q+1}^{-1}$$

$$+ K \sum_{q \in Q_i} \|y^q - z\|.$$

The first term of the right side is bounded because if q_1 is the first element of Q_i and $r \in Q_i$ is a finite number then

$$\sum_{q \in Q_{i}} q \in q_{i}, \quad q \leq r \begin{pmatrix} \theta & q^{-1} & \theta \\ q & q^{-1} \end{pmatrix}$$
$$= \theta_{q_{1}} - \theta_{r+1}$$

It follows from (4.4) that the second term on the right side of the inequality (4.11) converges. Hence the left side of the inequality (4.11) also converges. This implies that

(4.12)
$$\lim_{q \to \infty} \phi_{q} = 0.$$

Furthermore we see

$$\begin{array}{l} \not \phi \\ q \\ = & \| & \{ \ \tilde{B}_{q} \ - \ Df_{i}(z) \ \} \ (y^{q+1} \ - \ y^{q}) \ \| \ \not \mid \ y^{q+1} \ - \ y^{q} \ \| \\ = & \| \ -f_{i}(y^{q}) \ - \ Df_{i}(z) \ (y^{q+1} \ - \ y^{q}) \ \| \ \not \mid \ y^{q+1} \ - \ y^{q} \ \| \ (by \ (9)) \\ \geq & \| \ f_{i}(y^{q+1}) \ \| \ \not \mid \ y^{q+1} \ - \ y^{q} \ \| \\ & - & \| \ f_{i}(y^{q+1}) \ -f_{i}(y^{q}) \ - Df_{i}(z) \ (y^{q+1} \ - \ y^{q}) \ \| \ \not \mid \ y^{q+1} \ - \ y^{q} \ \| .$$

Since the second term above converges zero as $q \to \infty$ along the subsequence Q_i , (4.12) ensures that the first term

(4.13)
$$\| f_i(y^{q+1}) \| \neq \| y^{q+1} - y^q \|$$

converges zero as $q \to \infty$ along the subsequence Q_i . Since $f_i(z) = 0$ and $Df_i(z)$ is nonsingular, there exists a positive number β such that

$$\| f_{i}(y^{q+1}) \| \\ = \| f_{i}(y^{q+1}) - f_{i}(z) \| \ge \beta \| y^{q+1} - z \|$$

Hence (4.13) can be evaluated as

$$\| f_{i}(y^{q+1}) \| \neq \| y^{q+1} - y^{q} \|$$

$$\ge \beta \| y^{q+1} - z \| \neq \| y^{q+1} - y^{q} \|$$

$$\ge \beta \| y^{q+1} - z \| \neq \| y^{q+1} - z \| + \| y^{q} - z \| \}$$

$$= \beta (\| y^{q+1} - z \| \neq \| y^{q} - z \|) / \{1 + (\| y^{q+1} - z \| \neq \| y^{q} - z \|)\}$$

368

Thus (4.8) follows.

(iii) Finally we prove that the whole sequence $\{y^q\}$ converges z Q-superlinearly. Let ε be a positive number. It suffices to show that for some number N^* the inequality

$$(4.14) || y^{q+1} - z || / || y^{q} - z || \leq \varepsilon$$

holds whenever $q \ge N^*$. We have shown in (ii) above that (4.8) holds for every $i \in I^*(z)$. Hence, for each $i \in I^*(z)$, we can take a number N_i such that the inequality (4.14) holds if $q \in Q_i$ and $q \ge N_i$. On the other hand, by the definition of $I^*(z)$, there is a number \bar{N} such that if $q \ge \bar{N}$ then $y^q \in U_i$ for some $i \in I^*(z)$. Define

$$N^* = \max \{\overline{N}, N; (i \in I^*(z))\}$$

If $q \ge N^*$ then we can find a number $i \in I^*(z)$ such that $y^q \in U_i$. Thus the inequality (4.14) holds for every $q \ge N^*$. This completes the proof.

5. Numerical Examples

In this section, we solve the following three problems to test the computational behavior of Algorithms EN and EQN. Each of them has a degenerate solution which lies on a common boundary of some different pieces, so the original Newton and quasi-Newton methods might have some difficulty in solving these problems.

Problem 1. (Nonlinear Complementarity Problem)

Consider the following nonlinear complementarity problem : Find a solution $x \in \mathbb{R}^4$ such that $x \ge 0$, $f(x) \ge 0$ and $x^T f(x) = 0$, where $f : \mathbb{R}^4 \to \mathbb{R}^4$ is a C¹-mapping given by $f(x) = 3x^2 + 2x + 2x^2 + x + 3x = 6$

$$f_{1}(x) = 3x_{1} + 2x_{1}x_{2} + 2x_{2} + x_{3} + 3x_{4} - 6,$$

$$f_{2}(x) = 2x_{1}^{2} + x_{1} + x_{2}^{2} + 10x_{3} + 2x_{4} - 2,$$

$$f_{3}(x) = 3x_{1}^{2} + x_{1}x_{2} + 2x_{2}^{2} + 2x_{3} + 9x_{4} - 9,$$

$$f_{4}(x) = x_{1}^{2} + 3x_{2}^{2} + 2x_{3} + 3x_{4} - 3.$$

This problem has a solution $x^* = (\sqrt{6}/2, 0, 0, 1/2)$ such that

 $f(x^*) = (0, 2 + \sqrt{6}/2, 0, 0)$. This solution is degenerate since $x_3^* = 0$ and $f_3(x^*) = 0$.

Problem 2. (2-dimensional System of PC¹ Equations)

Consider the system of PC^1 equations F(x) =

 $(f_1(x), f_2(x)) = 0$, where

$$f_{1}(x) = (x_{2} - x_{1}) \ln \{ (x_{2} - x_{1})^{2} + 1 \} + x_{2} - x_{1},$$

$$f_{2}(x) = \begin{cases} -\exp(-x_{1} - x_{2}) + 1 & (x_{2} \ge 0), \\ (1 - \exp(-x_{1}))/(1 - x_{2}) & (x_{2} \le 0). \end{cases}$$

This equation has a unique solution x = (0,0) which lies on the boundary of two pieces, $\{(x_1, x_2) ; x_2 \ge 0\}$ and $\{(x_1, x_2) ; x_2 \le 0\}$.

Problem 3. (3-person Noncooperative Game)

Let us consider a 3-person noncooperative game as illustrated in Figure 2. Each player i (i = 1, 2, 3) has two pure strategies i.1, i.2. In Figure 2, (a) shows the payoffs when player 3 takes the strategy 3.1, while (b) shows the payoffs when he takes the strategy 3.2. For example, if the players 1, 2, 3 take the strategies 1.1, 2.2, 3.1, respectively, they get the payoffs 4, 2 and 9, respectively. A mixed strategy of player i is denoted by $(p_i, 1 - p_i)$, where p_i is the probability which he assigns to i.1. This game has two equilibrium points

$$w^{*} = (p_{1}^{*}, p_{2}^{*}, p_{3}^{*}) = (3/4, 2/3, 0),$$

$$\tilde{w} = (\tilde{p}_{1}, \tilde{p}_{2}, \tilde{p}_{3}) = (1/2, 1/2, 1).$$

The problem of finding equilibrium points is represented as a system of PC^1 equations. (See Kojima, Okada and Shindoh [11] for details.) The equilibrium point \tilde{w} corresponds to a degenerate solution of the system.

We applied Algorithms EN and EQN to these problems with various initial points. Tables below show the number of iterations and the number of pieces which the generated sequence visited. The computer program was coded in Turbo Pascal and the runs were made on an NEC PC-9801 VM2 personal computer. Initial matrices in Algorithm EQN are generated by the



Figure 2. A 3-person noncooperative game

difference approximation method. From these computational results, we see that Algorithms EN and EQN work effectively even if solutions to be computed are degenerate.

6. Concluding Remark

The following update formula is also well-known:

 $\begin{aligned} x^{p+1} &= x^p - H_p f(x^p), \\ H_{p+1} &= H_p + (s_p - H_p u_p) s_p^T H_p \swarrow s_p^T H_p u_p, \\ \text{where } s_p &= x^{p+1} - x^p, \text{ and } u_p = f(x^{p+1}) - f(x^p) \text{ and } H_p \text{ is an} \\ \text{approximate matrix of the inverse of the Jacobian matrix,} \\ Df^{-1}(x^p). (See, for example, Dennis and Moré [5], Luenberger \\ [13].) Let H_p &= A_p^{-1}. \text{ Then these formulae follow from} \\ \text{Algorithm QN and the Sherman-Morrison formula (Dennis and Moré \\ [5] or Ortega and Rheinboldt [16]). \end{aligned}$

We can extend the method above to the system of PC¹ equations in the similar way. The extension generates the same sequence as Algorithm EQN. Therefore the similar result as Theorem 3 holds.

Init. Point	EN		EQN	
	Iter.	#Pieces	Iter.	#Piece
(2, 2, 2, 2)	12	8	15	
(1, -1, -1, 1)	3	2	12	2
(-1, 1, 1, -1)	9	5	14	5

6

17

Table 1. Computational results on Problem 1

Table 2. Computational results on Problem 2

10

	EN		EQN	
Init. Point	Iter.	#Pieces	Iter.	#Pieces
(-1, -1)	4	2	27	2
(-1, 1)	4	1	14	2

Table 3. Computational results on Problem 3

EN			EQN	
Init. Point	Iter.	#Pieces	Iter.	#Pieces
(0.45, 0.45, 0.9)	3	2	3	2
(0.27, 0.3, 1.02)	3	2	4	2

References

- Broyden, C. G.: A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation*, Vol. 19, No. 92 (1965), 577-593.
- Broyden, C. G.: The convergence of single-rank quasi-Newton methods. Mathematics of Computation, Vol.24, NO.110 (1970), 365-382.
- [3] Dennis, Jr. J. E.: On the convergence of Broyden's method of nonlinear systems of equations. *Mathematics of Computation*, Vol.25, No.115 (1971), 559-567.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

-2 -2

-2)

- [4] Dennis, Jr. J. E. and Moré, J. J.: A characterization of superlinear convergence and its application to Quasi-Newton methods. *Mathematics of Computation*, Vol. 28, No. 126 (1974), 549-560.
- [5] Dennis, Jr. J. E. and Moré, J. J.: Quasi-Newton methods, motivation and theory. SIAM Review, Vol. 19, No. 1 (1977), 46-89.
- [6] Fujiwara, O.: Morse programs: a topological approach to smooth constrained optimization. I. Mathematics of Operations Research. Vol. 7. No. 4 (1982). 602-616.
- [7] Han, S. P.: Superlinearly convergent variable metric algorithm for general nonlinear programming problems. *Mathematical Programming*, Vol. 11, No. 3 (1976), 263-282.
- [8] Josephy, N. H.: Newton's method for generalized equations. Technical Summary Report #1965, Mathematics Research Center, University of Wisconsin-Madison (1979)
- [9] Josephy, N. H.: Quasi-Newton methods for generalized equations. Technical Summary Report #1966, Mathematics Research Center, University of Wisconsin-Madison (1979)
- [10] Kojima, M.: Strongly stable stationary solutions in nonlinear programs. Analysis and Computation of Fixed Points, 93-138, Academic Press, 1980.
- [11] Kojima, M., Okada, A. and Shindoh, S.: Strongly stable equilibrium points of N-person noncooperative games. Mathematics of Operations Research, Vol.10, No4 (1985), 650-663.
- [12] Lemke, C. E.: Bimatrix equilibrium points and mathematical programming. Management Science, Vol.11, No.7 (1965), 681-689.
- [13] Luenberger, D. G.: Linear and nonlinear programming. (second edition) Addison-Wesley, 1984.
- [14] Mangasarian, O. L.: Nonlinear programming. McGraw-Hill, New York, 1969.
- [15] Murty, K. G.: Computational complexity of complementary pivot methods. Mathematical Programming Study, Vol.7 (1978), 61-73.

- [16] Ortega, J. M. and Rheinboldt, W. C.: Iterative solution of nonlinear equations in several variables. Academic Press, New York, 1970.
- [17] Palomares, U. M. G. and Mangasarian, O. L.: Superlinearly convergent quasi-Newton algorithms for nonlinearly constrained optimization problems. *Mathematical Programming*, Vol.11, No.1 (1976), 1-13.
- [18] Pang, J. S. and Chan, D.: Iterative method for variational and complementarity problems. *Mathematical Programming*, Vol.24, No.3 (1982), 284-313.
- [19] Robinson, S. M.: Strongly regular generalized equations. Mathematics of Operations Research, Vol.5, No.1 (1980), 43-62.

Masakazu KOJIMA: Department of Information Sciences, Tokyo Institute of Technology Oh-okayama, Megro-ku, Tokyo 152, Japan

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.