JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 94, No. 3, pp. 659-676, SEPTEMBER 1997

Newton and Quasi-Newton Methods for Normal Maps with Polyhedral Sets[1]

J. HAN[2] AND D. SUN[3]

Communicated by O. L. Mangasarian

Abstract. We present a generalized Newton method and a quasi-Newton method for solving $H(x) := F(\Pi_C(x)) + x - \Pi_C(x) = 0$, when C is a polyhedral set. For both the Newton and quasi-Newton methods considered here, the subproblem to be solved is a linear system of equations per iteration. The other characteristics of the quasi-Newton method include: (i) a Q-superlinear convergence theorem is established without assuming the existence of H' at a solution x^* of $H(x) = 0$; (ii) only one approximate matrix is needed; (iii) the linear independence condition is not assumed; (iv) Q-superlinear convergence is established on the original variable x; and (v) from the QR-factorization of the kth iterative matrix, we need at most $O((1 + 2|J_k| + 2|L_k|)n^2)$ arithmetic operations to get the QR-factorization of the $(k+1)$th iterative matrix.

Key Words. Normal maps, Newton methods, quasi-Newton methods, Q-superlinear convergence.

1. Introduction

Let C be a nonempty closed convex set in $\Re^n$, and let F be a continuous function from $\Re^n$ to itself. A very common problem arising in optimization and equilibrium analysis is that of finding a point x that is a solution of the following equation:

$$H(x) := F(\Pi_C(x)) + x - \Pi_C(x) = 0, \tag{1}$$

[1]This work was supported by the National Natural Science Foundation of China. The authors would like to thank two referees for careful reading and helpful suggestions.
[2]Professor, Institute of Applied Mathematics, Academia Sinica, Beijing, P.R. China.
[3]PhD, Institute of Applied Mathematics, Academia Sinica, Beijing, P.R. China; presently at the School of Mathematics, University of New South Wales, Sydney, NSW, Australia.

where Π_C is the Euclidean projector on C and H is called a normal map (Ref. 1). For example, the variational inequality problem defined on C is to find $y \in C$ such that

$$(z-y)^T F(y) \geq 0, \qquad \forall z \in C. \tag{2}$$

It is easy to verify that, if $H(x)=0$, then the point $y := \Pi_C(x)$ solves (2); conversely, if y solves (2), then with $x := y - F(y)$, one has $H(x) = 0$. Therefore the equation $H(x) = 0$ is an equivalent way of formulating the variational inequality problem (2). The function H defined in (1) is in general nonsmooth.

Recently, many authors have considered Newton and/or quasi-Newton methods for solving nonsmooth equations and related problems; see, e.g., Refs. 2–10 and references therein. For solving the variational inequality problem, some early methods are Josephy's Newton method (Ref. 4) and quasi-Newton methods (Ref. 5). At each step, the Josephy methods solve a linear variational inequality problem defined on the set C. This is a nonlinear and nonconvex subproblem. Kojima and Shindo (Ref. 6) generalized Newton and quasi-Newton methods to piecewise smooth equations. For quasi-Newton methods, they need a new approximate starting matrix when the iteration sequence moves to a new C^1-piece. This may require storing lots of initial matrices. Ip and Kyparisis (Ref. 3) discussed quasi-Newton methods directly applied to nonsmooth equations. The Q-superlinear convergence of quasi-Newton methods was established by them on the assumption that the underlying mapping is strongly Fréchet differentiable at the solution (Ref. 11). This is somewhat too restrictive for (1). The results of Chen and Qi (Ref. 2) are not far from this. Qi and Jiang (Ref. 9) discussed applications of Newton and quasi-Newton methods for solving semismooth Karush–Kuhn–Tucker (KKT) equations of a nonlinear programming problem and established superlinear convergence of the proposed methods without assuming a strict complementarity condition. Sun and Han (Ref. 10) considered Newton and quasi-Newton methods for a class of nonsmooth equations and related problems, which includes the general nonlinear complementarity problem, the variational inequality problem with simple bound constraints, and the KKT system of a nonlinear programming problem. The quasi-Newton method of Sun and Han needs one approximate initial matrix, and at each step it solves only a linear system of equations. Furthermore, for the quasi-Newton method, they discussed how to update the QR-factorization of the present iterative matrix to the QR-factorization of the next iterative matrix in less than $O(n^3)$ arithmetic operations.

In this paper, we generalize the methods developed in Ref. 10 to solve (1) when C is a general polyhedral set. Since H is piecewise smooth if F is continuously differentiable and C is polyhedral [even if C is nonpolyhedral,

H may still be piecewise smooth under some constraint qualification assumption (Ref. 12)], the Newton-type methods developed in Ref. 6 for solving general piecewise smooth equations can be used directly. Suppose that H is piecewise smooth with a continuous selection of a family of finitely many continuous differentiable functions $\{H^i | i \in E\}$ such that, for any $x \in \mathfrak{R}^n$,

$$H(x) \in \{H^i(x) | i \in E\}.$$

In order to use the methods developed in Ref. 6 for solving (1), one must first find an essentially active piece H^i, $i \in E$, at a point x (Ref. 13); i.e.,

$$x \in \text{cl}(\text{int}\{y \in \mathfrak{R}^n | H(y) = H^i(y)\}),$$

where for any set $S \subseteq \mathfrak{R}^n$, cl(int$\{S\}$) denotes the closure of the interior of S. This is not an easy task when C is a general polyhedral set. Based on the work of Ref. 12, in this paper we first modify slightly the Newton methods developed in Refs. 6 and 8 for solving (1). This modification allows us to use nonessentially active smooth pieces, making the resulting Newton method more computable. Based on the modified Newton method and on Josephy's quasi-Newton methods (Ref. 5), we develop a new quasi-Newton method. The subproblem involved in the new method is a linear system of equations, and only one approximate starting matrix is used.

In this paper, we assume that C has the form

$$C = \{x | Ax \leq a, Bx = b\}, \tag{3}$$

where $A: \mathfrak{R}^n \to \mathfrak{R}^m$, $B: \mathfrak{R}^n \to \mathfrak{R}^p$, $a \in \mathfrak{R}^m$, and $b \in \mathfrak{R}^p$. Also, we assume that rank$(B) = p$, $p \leq n$. We discuss Newton and quasi-Newton methods that use a linear system of equations as the subproblem per iteration. It is noted that a linear constrained least-square problem must be solved for obtaining the projection of a point x over the set C. This computation work cannot be avoided, since the function value of H at x is needed. We treat this work as part of the computation involved in computing $H(x)$, rather than that involved in each subproblem.

The main characteristics of the quasi-Newton method considered in this paper are as follows:

(i) without assuming the existence of H' at a solution x^* of (1), we establish a Q-superlinear convergence theorem;
(ii) only one approximate matrix is needed;
(iii) the linear independence condition is not assumed;
(iv) Q-superlinear convergence is established on the original variable x;

(v) from the QR-factorization of the kth iterative matrix, we need at most $O((1+2|J_k|+2|L_k|)n^2)$ arithmetic operations to get the QR-factorization of the $(k+1)$th iterative matrix; see (38) for the definition of J_k and L_k.

The rest of this paper is organized as follows. In Section 2, we discuss some properties of the normal map in (1). A generalized Newton method and a quasi-Newton method are given in Sections 3 and 4. In Section 5, we discuss the implementation aspects of Newton and quasi-Newton methods.

2. Basic Preliminaries

Let $\|\cdot\|$ denote the l_2-vector norm or its induced matrix norm. For any $x\in\Re^n$, let $\Pi_C(x)$ be the Euclidean projection of x on C. Since C is of the form (3), there exist multipliers $\lambda\in\Re^m_+$, $\mu\in\Re^p$ such that

$$\Pi_C(x)-x+A^T\lambda+B^T\mu=0, \tag{4a}$$

$$\lambda\geq 0, \qquad a-A\Pi_C(x)\geq 0, \qquad \lambda^T[a-A\Pi_C(x)]=0, \tag{4b}$$

$$b-B\Pi_C(x)=0. \tag{4c}$$

Let $\mathcal{M}(x)$ denote the nonempty set of multipliers $(\lambda,\mu)\in\Re^m_+\times\Re^p$ that satisfy the KKT conditions (4). For a nonnegative vector $d\in\Re^m$, let supp(d), the support of d, be the subset of $\{1,\ldots,m\}$ consisting of the indexes i for which $d_i>0$. Denote

$$I(x)=\{i\,|\,A_i\Pi_C(x)=a_i,\ i=1,\ldots,m\}. \tag{5}$$

Define the family $\mathcal{B}(x)$ of indexes of $\{1,\ldots,m\}$ as follows: $K\in\mathcal{B}(x)$ if and only if supp$(\lambda)\subseteq K\subseteq I(x)$, for some $(\lambda,\mu)\in\mathcal{M}(x)$, and the vectors

$$\{A_i^T,\ i\in K\}\cup\{B_j^T,\ j=1,\ldots,p\} \tag{6}$$

are linearly independent. The family $\mathcal{B}(x)$ is nonempty, because $\mathcal{M}(x)$ has an extreme point which yields easily a desired index set K with the stated properties.

Define

$$\mathcal{P}(x)=\left\{P\in\Re^{n\times n}\,|\,P=I-[A_K^T,B^T]\left[\begin{bmatrix}A_K\\B\end{bmatrix}[A_K^T,B^T]\right]^{-1}\begin{bmatrix}A_K\\B\end{bmatrix},\ K\in\mathcal{B}(x)\right\}, \tag{7}$$

where I is the identity matrix of $\Re^{n\times n}$ and A_K is the matrix consisting of the rows of A, indexed by K.

Remark 2.1. The existence of the inverse in (7) comes from the linear independence of the vectors

$$\{A_i^T, i\in K\} \cup \{B_j^T, j=1, \ldots, p\}.$$

Note that, for all $P\in\mathscr{P}(x)$, we have

$$P^T=P, \qquad P^2=P, \qquad \|P\|\leq 1.$$

These simple facts will be used later.

In the following lemma, part (i) is a consequence of Pang and Ralph (Ref. 12). For completeness, we prove it here.

Lemma 2.1.

(i) There exists a neighborhood $N(x)$ of x such that, when $y\in N(x)$, we have

$$\mathscr{B}(y)\subseteq\mathscr{B}(x) \quad \text{and} \quad \mathscr{P}(y)\subseteq\mathscr{P}(x).$$

(ii) When $\mathscr{B}(y)\subseteq\mathscr{B}(x)$, then $\Pi_C(y)=\Pi_C(x)+P(y-x)$, $\forall P\in\mathscr{P}(y)$.

Proof.

(i) According to the definition of $\mathscr{P}(\cdot)$, we only need to prove that there exists a neighborhood $N(x)$ of x such that

$$\mathscr{B}(y)\subseteq\mathscr{B}(x), \qquad \forall y\in N(x). \tag{8}$$

If not, then there exists a sequence $\{y^k\}$ converging to x such that, for all k, there is an index set $K^k\in\mathscr{B}(y^k)\setminus\mathscr{B}(x)$. Since there are only finitely many such index sets, if necessary by taking a subsequence we assume that these index sets K^k are the same for all k. By letting K be the common index set, we have that the vectors

$$\{A_i^T, i\in K\} \cup \{B_j^T, j=1, \ldots, p\}$$

are linearly independent and there exists $(\lambda^k, \mu^k)\in\mathscr{M}(y^k)$ such that $\mathrm{supp}(\lambda^k)\subseteq K\subseteq I(y^k)$, but $K\notin\mathscr{B}(x)$. Clearly, $K\subseteq I(x)$. The only way for $K\notin\mathscr{B}(x)$ is that there exists no $(\lambda, \mu)\in\mathscr{M}(x)$ such that $\mathrm{supp}(\lambda)\subseteq K$. But we have

$$\Pi_C(y^k)-y^k+\sum_{i\in K}\lambda_i^k A_i^T+\sum_{j=1}^{p}\mu_j^k B_j^T=0.$$

Since $y^k\to x$, and since $\{A_i^T, i\in K\} \cup \{B_j^T, j=1, \ldots, p\}$ are linearly independent, it follows that $\{\lambda_i^k, i\in K\}$ and $\{\mu_j^k, j=1, \ldots, p\}$ are bounded; thus, the full sequence $\{\lambda^k\} \cup \{\mu^k\}$ must have an accumulation point which must be an element in $\mathscr{M}(x)$ and whose support is a subset of K. This is a contradiction.

(ii) When $\mathscr{B}(y) \subseteq \mathscr{B}(x)$, from the KKT conditions (4) we know that, for any $P \in \mathscr{P}(y)$, there exists $K \in \mathscr{B}(y) \subseteq \mathscr{B}(x)$ such that

$$P = I - [A_K^T, B^T]\left[\begin{bmatrix} A_K \\ B \end{bmatrix}[A_K^T, B^T]\right]^{-1}\begin{bmatrix} A_K \\ B \end{bmatrix},$$

$$\Pi_C(y) = Py + c_K, \qquad \Pi_C(x) = Px + c_K,$$

where

$$c_K = [A_K^T, B^T]\left[\begin{bmatrix} A_K \\ B \end{bmatrix}[A_K^T, B^T]\right]^{-1}\begin{bmatrix} a_K \\ b \end{bmatrix}$$

and a_K is the vector consisting of the components of a, indexed by K. Thus,

$$\Pi_C(y) = \Pi_C(x) + P(y - x), \qquad P \in \mathscr{P}(y).$$

This completes the proof. □

By the above lemma, we have

$$\partial_B \Pi_C(x) \subseteq \mathscr{P}(x), \tag{9}$$

where for any locally Lipschitz continuous function $G: \Re^n \to \Re^n$, $\partial_B G(x)$ is defined as (Ref. 8)

$$\partial_B G(x) = \left\{ V \in \Re^{n \times n} \,\middle|\, V = \lim_{x^k \to x} G'(x^k),\ G \text{ is differentiable at } x^k \right\}.$$

The equality in (9) does not hold in general. To see this, consider

$$C = \{x \in \Re^2 \,|\, -2x_1 + x_2 \leq 0,\ x_1 - x_2 \leq 0,\ -x_2 \leq 0\}.$$

After simple computations, we have

$$\mathscr{P}(0) = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1/5 & 2/5 \\ 2/5 & 4/5 \end{bmatrix}, \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\},$$

$$\partial_B \Pi_C(0) = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1/5 & 2/5 \\ 2/5 & 4/5 \end{bmatrix}, \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \right\}.$$

So,

$$\partial_B \Pi_C(0) \neq \mathscr{P}(0).$$

Ideally, for a given point $x \in \mathfrak{R}^n$, it is better to use an element of $\partial_B \Pi_C(x)$ instead of $\mathscr{P}(x)$ to design Newton-type methods for solving (1). However, it is not easy to compute an element of $\partial_B \Pi_C(x)$, while it is relatively easier to compute an element of $\mathscr{P}(x)$. In Sections 3 and 4, we will use $\mathscr{P}(x)$ to design Newton-type methods.

3. Generalized Newton Method

In the following sections, suppose that F is continuously differentiable. Denote

$$\mathscr{W}(x) = \{W \in \mathfrak{R}^{n \times n} \mid W = F'(\Pi_C(x))P + I - P,\ P \in \mathscr{P}(x)\}.$$

A generalized Newton method for solving (1) can be described as follows:

Step 0. Given $x^0 \in \mathfrak{R}^n$.
Step 1. For $k = 0, 1, \ldots,$ choose $P_k \in \mathscr{P}(x^k)$ and compute

$$W_k := F'(\Pi_C(x^k))P_k + I - P_k \in \mathscr{W}(x^k).$$

Solve the following equation for s^k:

$$W_k s + H(x^k) = 0. \tag{10}$$

Set

$$x^{k+1} = x^k + s^k. \tag{11}$$

Before giving a convergence analysis of the above Newton method, let us first consider a generalized Newton method for solving piecewise smooth equations. Suppose that a function $G: \mathfrak{R}^n \to \mathfrak{R}^n$ has a continuous selection of a family of finitely many continuously differentiable functions $\{G^i \mid i \in E\}$. Denote the active set of G at $x \in \mathfrak{R}^n$ by

$$E(x) := \{i \in E \mid G^i(x) = G(x)\}.$$

Given $x^0 \in \mathfrak{R}^n$, a generalized Newton method for solving the piecewise smooth equation

$$G(x) = 0 \tag{12}$$

is as follows:

$$x^{k+1}=x^k+d^k, \qquad k=0,1,\ldots, \tag{13}$$

where d^k is the solution of the following equation:

$$G(x^k)+V_k d=0,$$
$$V_k\in\{(G^i)'(x^k)|i\in E(x^k)\}.$$

If all G^i, $i\in E(x)$, are essentially active pieces of G at any $x\in\mathfrak{R}^n$, the convergence analysis of the iteration method (13) was given in Ref. 6. Since not all active pieces G^i, $i\in E(x)$, of G at x are essentially active pieces of G at x (see the example in Section 2), and since an essentially active piece of G at x, which exists by Proposition 4.1.1 of Ref. 14, may not easily be identified, it is better to use any active pieces of G at x in the algorithm. By noting that $\{G^i|i\in E\}$ is a continuous selection of a family of several continuously differentiable functions of G, similarly to the proof of Theorem 1 of Ref. 6, we have the following lemma.

Lemma 3.1. Suppose that x^* is a solution of (12). If all $V_*\in\{(G^i)'(x^*)|i\in E(x^*)\}$ are nonsingular, then there exists a neighborhood N of x^* such that, when the initial vector x^0 is chosen in N, the entire sequence $\{x^k\}$ generated by (13) is well defined and converges Q-superlinearly to x^*. Furthermore, if all $(G^i)'$, $i\in E(x^*)$, are Lipschitz continuous around x^*, then the convergence is quadratic.

We may prove the convergence of the iteration method (11) by showing that (11) is a special case of (13) in a neighborhood of a solution x^* of (1).

Theorem 3.1. Suppose that F: $\mathfrak{R}^n\to\mathfrak{R}^n$ is continuously differentiable, C is of the form (3), and x^* is a solution of (1). If all $W_*\in\mathscr{W}(x^*)$ are nonsingular, then there exists a neighborhood N of x^* such that, when the initial vector x^0 is chosen in N, the entire sequence $\{x^k\}$ generated by (11) is well defined and converges to x^* Q-superlinearly. Furthermore, if F' is Lipschitz continuous around $\Pi_C(x^*)$, then the convergence is quadratic.

Proof. From Lemma 2.1, we know that there exists a neighborhood N of x^* such that

$$\mathscr{B}(x)\subseteq\mathscr{B}(x^*) \quad \text{and} \quad \mathscr{P}(x)\subseteq\mathscr{P}(x^*)$$

hold for all $x\in N$. Then, for all $x\in N$,

$$H(x)\in\{F(c^*+P(x-x^*))+x-(c^*+P(x-x^*))|P\in\mathscr{P}(x^*)\},$$

where $c^* := \Pi_C(x^*)$ and $\mathscr{P}(x^*)$ is the family of finitely many affine functions defined in Eq. (7) at x^*. By the definition of $\mathscr{P}(\cdot)$, it is easy to see that, for any $x^k \in N$ and any $P_k \in \mathscr{P}(x^k)$, $H(x^k)$ and W_k are the value and the Jacobian of the following function at x^k:

$$F(x^* + P_k(x - x^*)) + x - (c^* + P_k(x - x^*)).$$

Thus, for all $x^k \in N$, (11) is a special case of (13). By the assumptions and Lemma 3.1, the results of this theorem hold. □

Since in general $\mathscr{P}(x) \neq \partial_B \Pi_C(x)$, and so $\mathscr{W}(x) \neq \partial_B H(x)$, the Newton method established here is slightly different from those in Refs. 6 and 8. Concerning the assumption of nonsingularity of all $W_* \in \mathscr{W}(x^*)$, we have the following result.

Proposition 3.1. Suppose that $V := F'(\Pi_C(x))$ is strictly positive definite on

$$\mathscr{C}(x; C) = \bigcup_K \{v \mid A_K v = 0, Bv = 0, K \in \mathscr{B}(x)\},$$

i.e.,

$$v^T V v > 0, \qquad \forall v \in \mathscr{C}(x; C) \backslash 0. \tag{14}$$

Then, all $W \in \mathscr{W}(x)$ are nonsingular.

Proof. For any $W \in \mathscr{W}(x)$, there exists an index set $K \in \mathscr{B}(x)$ such that

$$W = VP + I - P,$$

where

$$P = I - [A_K^T, B^T] \left[\begin{bmatrix} A_K \\ B \end{bmatrix} [A_K^T, B^T] \right]^{-1} \begin{bmatrix} A_K \\ B \end{bmatrix}$$

is an element of $\mathscr{P}(x)$. Assume that v is such that

$$Wv = 0,$$

i.e.,

$$VPv + v - Pv = 0. \tag{15}$$

Multiplying by $(Pv)^T$ both sides of (15), and noting that $P^T = P$ and $P^2 = P$, we have

$$\begin{aligned} 0 &= (Pv)^T VPv + (Pv)^T v - (Pv)^T Pv \\ &= (Pv)^T VPv + v^T Pv - v^T P^2 v \\ &= (Pv)^T VPv + v^T Pv - v^T Pv \\ &= (Pv)^T VPv. \end{aligned}$$

Therefore,

$$(Pv)^T VPv = 0. \tag{16}$$

But

$$\begin{aligned} \begin{bmatrix} A_K \\ B \end{bmatrix} Pv &= \begin{bmatrix} A_K \\ B \end{bmatrix} v - \begin{bmatrix} A_K \\ B \end{bmatrix} [A_K^T, B^T] \left[\begin{bmatrix} A_K \\ B \end{bmatrix} [A_K^T, B^T] \right]^{-1} \begin{bmatrix} A_K \\ B \end{bmatrix} v \\ &= \begin{bmatrix} A_K \\ B \end{bmatrix} v - \begin{bmatrix} A_K \\ B \end{bmatrix} v \\ &= 0, \end{aligned}$$

which means that

$$Pv \in \mathscr{C}(x; C).$$

By (14), (16), and the above formula, we have

$$Pv = 0.$$

Substituting this into (15) gives

$$v = 0,$$

which means that W is nonsingular. □

Remark 3.1. In Proposition 3.1, we do not need the condition of linear independence of the vectors

$$\{A_i^T, i \in I(x)\} \cup \{B_j^T, j = 1, \ldots, p\}.$$

If this linear independence condition is satisfied, then Condition (14) is equivalent to the Robinson strong sufficiency condition (Ref. 15), which is implied by the sufficiency condition and the strict complementarity condition; i.e., there exists no $i \in I(x)$ such that $\lambda_i = 0$, where $(\lambda, \mu) \in \mathscr{M}(x)$.

In the linear case, Lemma 2.1, Theorem 3.1, and Proposition 3.1 give the following direct result.

Corollary 3.1. Suppose that F is affine, i.e., $F(z) = Mz + c$, with $M \in \Re^{n \times n}$ and $c \in \Re^n$, and x^* is a solution of (1). If all $W_* \in \mathscr{W}(x^*)$ are nonsingular, in particular if M is positive definite, then there exists a neighborhood N of x^* such that, when x^0 is chosen in N, we have

$$x^1 = x^*;$$

i.e., the iteration (11) will terminate in one step.

4. Quasi-Newton Method

Based on the generalized Newton method established in Section 3, we describe a quasi-Newton method for solving (1) [Broyden's Case (Ref. 16)]:

Step 0. Given $x^0 \in \Re^n$, $D_0 \in \Re^{n \times n}$, an approximation of $F'(\Pi_C(x^0))$.

Step 1. For $k = 0, 1, \ldots$, choose $P_k \in \mathscr{P}(x^k)$ and compute

$$V_k := D_k P_k + I - P_k.$$

Solve

$$V_k s + H(x^k) = 0$$

for s^k. Set

$$x^{k+1} = x^k + s^k,$$

$$\delta^k = \Pi_C(x^{k+1}) - \Pi_C(x^k),$$

$$y^k = F(\Pi_C(x^{k+1})) - F(\Pi_C(x^k)),$$

$$D_{k+1} = D_k + (y^k - D_k \delta^k)\delta^{k^T}/\delta^{k^T}\delta^k.$$

Remark 4.1. The above Broyden update should be credited to Josephy (Ref. 5), who solves a linear variational inequality problem defined on C to get s^k.

Theorem 4.1. Suppose that $F: \Re^n \to \Re^n$ is continuously differentiable, x^* is a solution of (1), F' is Lipschitz continuous in a neighborhood of $\Pi_C(x^*)$, and the Lipschitz constant is γ. Suppose that all $W_* \in \mathscr{W}(x^*)$ are nonsingular. Then, there exist positive constants ϵ, δ such that, when $\|x^0 - x^*\| \le \epsilon$ and $\|D_0 - F'(\Pi_C(x^*))\| \le \delta$, the sequence $\{x^k\}$ generated by the above quasi-Newton method (Broyden's case) is well defined and converges Q-superlinearly to x^*.

Proof. From Lemma 2.1 and the nonsingularity assumption of all $W_* \in \mathscr{W}(x^*)$, we know that there exist a neighborhood $N_0(x^*)$ of x^* and a positive number $\beta > 0$ such that $\mathscr{B}(x) \subseteq \mathscr{B}(x^*)$ and $\|W^{-1}\| \le \beta$ for any $x \in N_0(x^*)$ and any $W \in \mathscr{W}(x)$. Choose ϵ and δ such that

$$\mathscr{B}(x) \subseteq \mathscr{B}(x^*), \tag{17}$$

$$\|F'(\Pi_C(x)) - F'(\Pi_C(x^*))\| \le \gamma \|\Pi_C(x) - \Pi_C(x^*)\|, \tag{18}$$

$$7\beta\delta \le 1, \tag{19}$$

$$3\gamma\epsilon \le 2\delta, \tag{20}$$

$$\|W^{-1}\| \le \beta, \tag{21}$$

$$\|F(\Pi_C(x)) - F(\Pi_C(x^*)) - F'(\Pi_C(x))(\Pi_C(x) - \Pi_C(x^*))\|$$
$$\le (\delta/2)\|\Pi_C(x) - \Pi_C(x^*)\|, \tag{22}$$

for any $x \in N(x^*) := \{x \mid \|x - x^*\| \le \epsilon\}$ and any $W \in \mathscr{W}(x)$. Denote $e^k = x^k - x^*$.

We first prove that $\{x^k\}$ is locally Q-linearly convergent. The local proof consists of showing by induction that

$$\|D_k - F'(\Pi_C(x^*))\| \le (2 - 2^{-k})\delta, \tag{23}$$

$$\|V_k^{-1}\| \le (7/5)\beta, \tag{24}$$

$$\|e^{k+1}\| \le (1/2)\|e^k\|, \tag{25}$$

for $k = 0, 1, \ldots$.

For $k=0$, (23) is trivially true. The proof of (24) and (25) is identical to the proof at the induction step, so we omit it here.

Now, assume that (23)–(25) hold for $k = 0, 1, \ldots, i-1$. For $k = i$, we have from Dennis and Moré (Ref. 17) [also see Lemma 8.2.1 of Dennis and Schnabel (Ref. 18)] and the induction hypothesis that

$$\|D_i - F'(\Pi_C(x^*))\|$$
$$\le \|D_{i-1} - F'(\Pi_C(x^*))\|$$
$$+ (\gamma/2)[\|\Pi_C(x^i) - \Pi_C(x^*)\| + \|\Pi_C(x^{i-1}) - \Pi_C(x^*)\|]$$
$$\le [2 - 2^{-(i-1)}]\delta + (\gamma/2)(\|e^i\| + \|e^{i-1}\|)$$
$$\le [2 - 2^{-(i-1)}]\delta + (3/4)\gamma\|e^{i-1}\|. \tag{26}$$

From (25) and $\|e^0\| \le \epsilon$, we get

$$\|e^{i-1}\| \le 2^{-(i-1)}\|e^0\| \le 2^{-(i-1)}\epsilon.$$

Substituting this into (26), and using (20), gives

$$\begin{aligned}\|D_i - F'(\Pi_C(x^*))\| &\le [2-2^{-(i-1)}]\delta + (3/4)\gamma \cdot 2^{-(i-1)}\epsilon \\ &\le [2-2^{-(i-1)}]\delta + 2^{-i}\delta \\ &= (2-2^{-i})\delta,\end{aligned}$$

which verifies (23).

To verify (24), we must first show that V_i is invertible. From the definition of V_i, there exists $P_i \in \mathscr{P}(x^i)$ such that

$$V_i = D_i P_i + I - P_i.$$

Denote

$$W_i = F'(\Pi_C(x^i))P_i + I - P_i.$$

Then, $W_i \in \mathscr{W}(x^i)$ and

$$\begin{aligned}\|V_i - W_i\| &\le \|D_i - F'(\Pi_C(x^i))\| \, \|P_i\| \\ &\le \|D_i - F'(\Pi_C(x^i))\| \\ &\le \|D_i - F'(\Pi_C(x^*))\| \\ &\quad + \|F'(\Pi_C(x^i)) - F'(\Pi_C(x^*))\|.\end{aligned} \tag{27}$$

Using (23) for $k=i$ and the Lipschitz condition (18) gives

$$\begin{aligned}\|V_i - W_i\| &\le (2-2^{-i})\delta + \gamma\|\Pi_C(x^i) - \Pi_C(x^*)\| \\ &\le (2-2^{-i})\delta + \gamma\|e^i\|.\end{aligned} \tag{28}$$

From (25), $\|e^0\| \le \epsilon$, and (20),

$$\gamma\|e^i\| \le 2^{-i}\epsilon\gamma \le (2/3)2^{-i}\delta,$$

which substituted into (28) gives

$$\|V_i - W_i\| \le (2-2^{-i})\delta + (2/3)2^{-i}\delta \le 2\delta. \tag{29}$$

From (21), (29), and (19), we get

$$\|W_i^{-1}(W_i - V_i)\| \le \beta \cdot 2\delta \le 2/7 < 1.$$

So, from Theorem 2.3.2 of Ortega and Rheinboldt (Ref. 11) we have that V_i is invertible and

$$\|V_i^{-1}\| \le \|W_i^{-1}\| / [1 - \|W_i^{-1}(W_i - V_i)\|] \le \beta/(1-2/7) = (7/5)\beta,$$

which verifies (24).

To complete the induction, we verify (25). From

$$V_i(x^{i+1}-x^i)+H(x^i)=0,$$

we have

$$\begin{aligned} V_i e^{i+1} &= -H(x^i)+V_i e^i \\ &= -[H(x^i)-H(x^*)-V_i e^i]. \end{aligned}$$

From Lemma 2.1 and (17), we know that

$$\Pi_C(x^i)-\Pi_C(x^*)=P_i(x^i-x^*). \tag{30}$$

Therefore,

$$\begin{aligned} \|e^{i+1}\| &\le \|V_i^{-1}\|\,\|H(x^i)-H(x^*)-V_i e^i\| \\ &= \|V_i^{-1}\|\,\|[F(\Pi_C(x^i))-F(\Pi_C(x^*))-D_iP_i(x^i-x^*)] \\ &\qquad +[x^i-x^*-I(x^i-x^*)] \\ &\qquad -[\Pi_C(x^i)-\Pi_C(x^*)-P_i(x^i-x^*)]\| \\ &= \|V_i^{-1}\|\,\|F(\Pi_C(x^i))-F(\Pi_C(x^*))-D_i[\Pi_C(x^i)-\Pi_C(x^*)]\| \\ &\le \|V_i^{-1}\|\{\|F(\Pi_C(x^i))-F(\Pi_C(x^*)) \\ &\qquad -F'(\Pi_C(x^i))[\Pi_C(x^i)-\Pi_C(x^*)]\| \\ &\qquad +\|[F'(\Pi_C(x^i))-D_i]P_i(x^i-x^*)\|\} \\ &= \|V_i^{-1}\|\{\|F(\Pi_C(x^i))-F(\Pi_C(x^*)) \\ &\qquad -F'(\Pi_C(x^i))[\Pi_C(x^i)-\Pi_C(x^*)]\| \\ &\qquad +\|(W_i-V_i)(x^i-x^*)\|\}. \end{aligned} \tag{31}$$

From (31), (24), (22), (29), and (19), we get

$$\begin{aligned} \|e^{i+1}\| &\le (7/5)\beta[(\delta/2)\|\Pi_C(x^i)-\Pi_C(x^*)\|+2\delta\|e_i\|] \\ &\le (7/5)\beta[(\delta/2)\|e^i\|+2\delta\|e^i\|] \\ &\le (1/2)\|e^i\|. \end{aligned}$$

This proves (25) and completes the Q-linear convergence proof.

Next, we prove the Q-superlinear convergence of $\{x^k\}$ under the assumptions. Let

$$E_k=D_k-F'(\Pi_C(x^*)).$$

From Ref. 17 or the last part of the proof of Theorem 8.2.2 of Ref. 18, we get

$$\lim_{k\to\infty} \|E_k \delta^k\|/\|\delta^k\| = 0. \tag{32}$$

So, from (31), (24), (30), (23), (32), and (19), we have

$$\begin{aligned}
\|e^{k+1}\| &\le (7/5)\beta\{\|F(\Pi_C(x^i)) - F(\Pi_C(x^*)) \\
&\qquad - F'(\Pi_C(x^i))[\Pi_C(x^i) - \Pi_C(x^*)]\| \\
&\qquad + \|[F'(\Pi_C(x^k)) - D_k][\Pi_C(x^k) - \Pi_C(x^*)]\|\} \\
&\le o(\|\Pi_C(x^k) - \Pi_C(x^*)\|) \\
&\quad + (7/5)\beta\{\|[D_k - F'(\Pi_C(x^*))][\Pi_C(x^k) - \Pi_C(x^*)]\| \\
&\qquad + \|[F'(\Pi_C(x^k)) - F'(\Pi_C(x^*))] \\
&\qquad \times [\Pi_C(c^k) - \Pi_C(x^*)]\|\} \\
&\le o(\|\Pi_C(x^k) - \Pi_C(x^*)\|) \\
&\quad + (7/5)\beta\{\|[D_k - F(\Pi_C(x^*))][\Pi_C(x^{k+1}) - \Pi_C(x^k)]\| \\
&\qquad + \|[D_k - F'(\Pi_C(x^*))][\Pi_C(x^{k+1}) - \Pi_C(x^*)]\|\} \\
&\quad + o(\|\Pi_C(x^k) - \Pi_C(x^*)\|) \\
&\le o(\|\Pi_C(x^k) - \Pi_C(x^*)\|) + (7/5)\beta\|E_k\delta^k\| \\
&\quad + (14/5)\beta\delta\|\Pi_C(x^{k+1}) - \Pi_C(x^*)\| \\
&\le o(\|e^k\|) + o(\|\delta^k\|) + (2/5)\|e^{k+1}\| \\
&\le o(\|e^k\|) + o(\|e^k\|) + o(\|e^{k+1}\|) + (2/5)\|e^{k+1}\|,
\end{aligned}$$

which means that

$$\lim_{k\to\infty} \|e^{k+1}\|/\|e^k\| = 0.$$

This completes the Q-superlinear convergence of $\{x^k\}$. □

When $C = \Re^n_+$, H defined by (1) is essentially equivalent to the function G defined in Refs. 6 and 3. In Ref. 3, Ip and Kyparisis discussed the convergence properties of quasi-Newton methods applied directly to nonsmooth equations. For the nonlinear complementarity problem (i.e., $C = \Re^n_+$), they described the sufficient conditions to guarantee the convergence of the quasi-Newton method; see Theorem 5.4 of Ref. 3. A restrictive condition in Theorem 5.4 of Ref. 3 is that

$$\nabla F_i(\Pi_C(x^*)) = I_i^T, \qquad \text{for all } i \in \{j \mid x_j^* = 0\},$$

where I_i^T is the ith column of the identity matrix I. This condition restricts the class F to which Theorem 5.4 of Ref. 3 applies. Here, to guarantee the convergence of our new quasi-Newton method, we need a nonsingularity assumption for all $W_* \in \mathscr{W}(x^*)$, instead of the existence and invertibility of $H'(x^*)$ as in Ref. 3. For the nonlinear complementarity problem, the nonsingularity assumption used here is similar to that of Ref. 10.

5. Implementation Aspects

For implementing the generalized Newton method established in this paper, there is not much difference from the smooth case except for choosing the iterative matrices. For implementing the quasi-Newton method, there exist some differences from the smooth case, especially for the factorization of the iterative matrix V_k. The entire QR-factorization of V_k costs $O(n^3)$ arithmetic operations. If we do this per iteration, then the advantages of quasi-Newton methods diminish. In this section, we discuss how to update the QR-factorization of V_k into the QR-factorization of V_{k+1} in much less than $O(n^3)$ operations.

Denote

$$\bar{V}_k = D_{k+1} P_k + I - P_k. \tag{33}$$

Then,

$$\bar{V}_k = V_k + (y_k - D_k \delta^k) \delta^{k^T} P_k / \delta^{k^T} \delta^k, \tag{34}$$

$$V_{k+1} = \bar{V}_k + (D_{k+1} - I)(P_{k+1} - P_k). \tag{35}$$

It is well known that we can update the QR-factorization of V_k into the QR-factorization of $\bar{V}_k$ in $O(n^2)$ operations; see, e.g., Refs. 19 and 20.

According to the definition of P_k and P_{k+1}, there exist $K \in \mathscr{B}(x^k)$ and $\tilde{K} \in \mathscr{B}(x^{k+1})$ such that

$$P_k = I - [A_K^T, B^T] \left[\begin{bmatrix} A_K \\ B \end{bmatrix} [A_K^T, B^T] \right]^{-1} \begin{bmatrix} A_K \\ B \end{bmatrix}, \tag{36}$$

$$P_{k+1} = I - [A_{\tilde{K}}^T, B^T] \left[\begin{bmatrix} A_{\tilde{K}} \\ B \end{bmatrix} [A_{\tilde{K}}^T, B^T] \right]^{-1} \begin{bmatrix} A_{\tilde{K}} \\ B \end{bmatrix}. \tag{37}$$

Denote

$$\bar{K} = K \cap \tilde{K}, \; J_k = K \backslash \bar{K}, \; L_k = \tilde{K} \backslash \bar{K}. \tag{38}$$

Define

$$\bar{P}_k = I - [A_{\bar{K}}^T, B^T]\left[\begin{bmatrix} A_{\bar{K}} \\ B \end{bmatrix}(A_{\bar{K}}^T, B^T)\right]^{-1}\begin{bmatrix} A_{\bar{K}} \\ B \end{bmatrix}, \tag{39}$$

$$\bar{V}_{k+1} = \bar{V}_k + (D_{k+1} - I)(\bar{P}_k - P_k). \tag{40}$$

After simple computations, we know that $(D_{k+1} - I)(\bar{P}_k - P_k)$ is at most a rank-$2|J_k|$ matrix and $(D_{k+1} - I)(P_{k+1} - \bar{P}_k)$ is at most a rank-$2|L_k|$ matrix. But from (35) and (40), we know that

$$V_{k+1} = \bar{V}_{k+1} + (D_{k+1} - I)(P_{k+1} - \bar{P}_k).$$

So, we can update the QR-factorization of $\bar{V}_k$ into the QR-factorization of V_{k+1} in $O(2(|J_k| + |L_k|)n^2)$ operations; see, e.g., Refs. 19 and 20. Therefore, we get the following theorem.

Theorem 5.1. The cost of updating the QR-factorization of V_k into the QR-factorization of V_{k+1} is at most $O((1 + 2|J_k| + 2|L_k|)n^2)$ arithmetic operations.

In Ref. 21, Harker and Pang proposed the following question: How to reduce the work to solve the subproblems involved in Josephy's quasi-Newton methods for solving the variational inequality problem. Theorem 5.1 says that, except for the first step, the subproblem involved in the quasi-Newton proposed here can be solved by using update technique. This may reduce greatly the work to solve subproblems.

References

1. ROBINSON, S. M., *Normal Maps Induced by Linear Transformation*, Mathematics of Operations Research, Vol. 17, pp. 691–714, 1992.
2. CHEN, X., and QI, L., *A Parameterized Newton Method and a Broyden-Like Method for Solving Nonsmooth Equations*, Computational Optimization and Applications, Vol. 3, pp. 157–179, 1994.
3. IP, C. M., and KYPARISIS, T., *Local Convergence of Quasi-Newton Methods for B-Differentiable Equations*, Mathematical Programming, Vol. 56, pp. 71–89, 1992.
4. JOSEPHY, N. H., *Newton's Method for Generalized Equations*, Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin, 1979.
5. JOSEPHY, N. H., *Quasi-Newton Methods for Generalized Equations*, Technical Summary Report No. 1966, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin, 1979.

6. KOJIMA, M., and SHINDO, S., *Extensions of Newton and Quasi-Newton Methods to Systems of PC^1-Equations*, Journal of the Operations Research Society of Japan, Vol. 29, pp. 352–374, 1986.
7. PANG, J. S., and QI, L., *Nonsmooth Equations: Motivation and Algorithms*, SIAM Journal on Optimization, Vol. 3, pp. 443–465, 1993.
8. QI, L., *Convergence Analysis of Some Algorithms for Solving Nonsmooth Equations*, Mathematics of Operations Research, Vol. 18, pp. 227–244, 1993.
9. QI, L., and JIANG, H., *Semismooth Karush–Kuhn–Tucker Equations and Convergence Analysis of Newton and Quasi-Newton Methods for Solving These Equations*, Mathematics of Operations Research (to appear).
10. SUN, D., and HAN, J., *Newton and Quasi-Newton Methods for a Class of Nonsmooth Equations and Related Problems*, SIAM Journal on Optimization, Vol. 7, pp. 463–480, 1997.
11. ORTEGA, J. M., and RHEINBOLDT, W. C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, New York, 1970.
12. PANG, J. S., and RALPH, D., *Piecewise Smoothness, Local Invertibility, and Parametric Analysis of Normal Maps*, Mathematics of Operations Research, Vol. 21, pp. 401–426, 1996.
13. KUNTZ, L., and SCHOLTES, S., *Structural Analysis of Nonsmooth Mappings, Inverse Functions, and Metric Projections*, Journal of Mathematical Analysis and Applications, Vol. 188, pp. 346–386, 1994.
14. SCHOLTES, S., *Introduction to Piecewise Differentiable Equations*, Habilitation Thesis, Institut für Statistik und Mathematische Wirtschaftstheorie, Universität Karlsruhe, Karlsruhe, Germany, 1994.
15. ROBINSON, S. M., *Strongly Regular Generalized Equations*, Mathematics of Operations Research, Vol. 5, pp. 43–62, 1980.
16. BROYDEN, C. G., *A Class of Methods for Solving Nonlinear Simultaneous Equations*, Mathematics of Computation, Vol. 19, pp. 577–593, 1965.
17. DENNIS, J. E., and MORÉ, J. J., *A Characterization of Superlinear Convergence and Its Application to Quasi-Newton Methods*, Mathematics of Computation, Vol. 28, pp. 549–560, 1974.
18. DENNIS, J. E., and SCHNABEL, R. B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, New Jersey, 1983.
19. GILL, P. E., and MURRAY, M., *Quasi-Newton Methods for Unconstrained Optimization*, Journal of the Institute of Mathematics and Its Applications, Vol. 9, pp. 91–108, 1972.
20. GOLUB, G. H., and VAN LOAN, C., *Matrix Computations*, Johns Hopkins University Press, Baltimore, Maryland, 1983.
21. HARKER, P. T., and PANG, J. S., *Finite-Dimensional Variational Inequality and Nonlinear Complementarity Problem: A Survey of Theory, Algorithms, and Applications*, Mathematical Programming, Vol. 48, pp. 161–220, 1990.