Mathematical Programming 56 (1992) 71-89
North-Holland

Local convergence of quasi-Newton methods for B-differentiable equations

Chi-Ming Ip

Department of Industrial Engineering, College of Engineering, University of Miami, Coral Gables, FL 33124, USA

Jerzy Kyparisis

Department of Decision Sciences and Information Systems, College of Business Administration, Florida International University, Miami, FL 33199, USA

Received 21 February 1990
Revised manuscript received 15 November 1990

We study local convergence of quasi-Newton methods for solving systems of nonlinear equations defined by B-differentiable functions. We extend the classical linear and superlinear convergence results for general quasi-Newton methods as well as for Broyden's method. We also show how Broyden's method may be applied to nonlinear complementarity problems and illustrate its computational performance on two small examples.

Key words: B-differentiable functions, quasi-Newton methods, local convergence, nonlinear equations, nonlinear complementarity problems.

1. Introduction

In this paper we study local convergence of quasi-Newton methods for solving systems of nonlinear equations defined by B-differentiable functions. This analysis extends the class of equation functions to which such methods may be applied beyond the familiar Fréchet differentiable functions. B-differentiable functions, introduced by Robinson [16], have properties of Lipschitz continuity and directional differentiability and they allow natural extensions of a number of key analytic results useful in proving convergence results. In two related recent papers, Pang [13, 14] extended the classical Newton method for solving nonlinear equations to B-differentiable functions and applied it to nonlinear complementarity and variational inequality problems. A similar extension of the classical Newton method to B-differentiable nonlinear equations was also proposed by Harker and Xiao [8] who applied it to nonlinear complementarity problems. These Newton-type methods were subsequently further generalized to semismooth nonlinear equations by Qi and Sun [15]. Another extension of Newton's method to a different class of non-smooth equations was studied by Robinson [17].

The objectives of this paper include: (i) extension of the classical linear and superlinear convergence results for general quasi-Newton methods as well as for Broyden's method, and (ii) a novel application of Broyden's method to nonlinear complementarity problems. The first objective is motivated by the fact that quasi-Newton methods combine excellent convergence rates with relatively minimal computational effort and are therefore often competitive with Newton's method [1-6]. Also, Newton's method of Pang [13, 14] requires a solution of nonlinear subproblems in each iteration which may be, in general, computationally inefficient. Thus, solution of nonlinear equations defined by non-Fréchet differentiable functions using quasi-Newton methods, which require only low rank update of the matrix inverse, is of practical interest. The only other paper dealing with quasi-Newton methods for nondifferentiable equations is due to Kojima and Shindo [11] where they considered systems of piecewise continuously differentiable equations.

The second objective is motivated by the simplicity of the specialization of Broyden's method to a nonlinear complementarity problem via equivalent systems of nonlinear equations. Computational methods for nonlinear complementarity problems typically require a solution of a mixed linear complementarity problem in each iteration which is in contrast to the method proposed here where only rank one update of the matrix inverse is done in each iteration. The solution of a mixed linear complementarity problem is required, for example, in the Newton and quasi-Newton methods of Josephy [9, 10] which extend the classical methods for nonlinear equations to variational inequalities and nonlinear complementarity problems via the framework of generalized equations as well as in the recent Newton-type methods of Pang [13, 14] and Harker and Xiao [8]. The computational performance of the proposed method is illustrated on two small examples.

The remainder of the paper is organized as follows. In Section 2 we review the notion and key analytic properties of a B-differentiable function. In Section 3 we study local convergence of quasi-Newton methods for solving B-differentiable systems of nonlinear equations and generalize the classical results to B-differentiable equation functions. In Section 4 we obtain local convergence properties of Broyden's method for nonlinear B-differentiable equations. In Section 5 we propose a new method for solving nonlinear complementarity problems based on the application of Broyden's method to two equivalent systems of nonlinear equations and analyze its local convergence properties. The computational performance of the proposed method is illustrated on two small examples in Section 6.

2. B-differentiable functions

In this section, we review the notion of a B-differentiable function and present some properties of such a function. We refer the reader to the recent papers by Robinson [16] and Pang [13] for a more detailed exposition.

The following definition is due to Robinson [16].

Definition 2.1 [16]. A function $H:\mathbb{R}^m \to \mathbb{R}^n$ is *B-differentiable* at a point $z \in D$ if there exists a positively homogeneous function $BH(z):\mathbb{R}^m \to \mathbb{R}^n$ (i.e., $BH(z)(\lambda v) = \lambda BH(z)v$ for all $\lambda \geq 0$ and $v \in \mathbb{R}^m$), called the *B-derivative* of H at z, such that

$$\lim_{v\to 0}[H(z+v)-H(z)-BH(z)v]/\|v\|=0.$$

If H is B-differentiable at all points $z \in D$, then H is called B-differentiable on D.

The basic properties of a B-differentiable function are summarized in the next proposition.

Proposition 2.1 [13, 16]. *Let $H:\mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitz continuous at a point z.*

(1) *If H is Fréchet differentiable at z, then it is B-differentiable at z and $BH(z) = \nabla H(z)$. Conversely, if H is B-differentiable at z and if the B-derivative $BH(z)v$ is linear in v, then H is Fréchet differentiable at z.*

(2) *If H is B-differentiable at z, the B-derivative is unique. Moreover, $BH(z)$ is Lipschitz continuous with the same modulus as H.*

(3) *If H is B-differentiable at z, then H is directionally differentiable at z in any direction and $H'(z, d) = BH(z)d$.*

(4) *The addition, subtraction, and chain rules hold for the B-derivative.* □

Throughout this section we assume that the function $H:\mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous in the region of interest. The following theorem was proved in Pang [13].

Theorem 2.1 [13]. *Let $H:\mathbb{R}^n \to \mathbb{R}^n$ be B-differentiable on an open, convex set $D \subset \mathbb{R}^n$. Then, for any x, y, $z \in D$,*

$$\|H(x)-H(y)-BH(z)(x-y)\| \leq \sup_{0\leq t\leq 1}\|(BH(y+t(x-y))-BH(z))(x-y)\|.$$

Moreover, the following statements are equivalent:

(1) *$BH(\cdot)$ is continuous at $z \in D$, i.e., for every $\varepsilon > 0$ there exists a neighborhood N of z such that, for all $x \in N$ and all $v \in \mathbb{R}^n$ with $\|v\| = 1$,*

$$\|(BH(x)-BH(z))v\| \leq \varepsilon.$$

(2) *H is Fréchet differentiable at z in the strong sense (see Ortega and Rheinboldt [12]) which, in particular, implies that H is Fréchet differentiable at z.*

(3) *The B-derivative $BH(z)$ satisfies the stronger limit property*

$$\lim_{(x,y)\to(z,z)}[H(x)-H(y)-BH(z)(x-y)]/\|x-y\|=0. \qquad \square$$

The above theorem as well as the next two lemmas will be utilized in subsequent sections.

Lemma 2.1. *Let $H:\mathbb{R}^n\to\mathbb{R}^n$ be B-differentiable on an open, convex set $D\subset\mathbb{R}^n$, and suppose that $BH(\cdot)$ is continuous at $z\in D$, and that $\nabla H(z)$ is nonsingular ($\nabla H(z)$ exists by Theorem 2.1). Then there exists a neighborhood N of z and $\alpha>0$ such that, for all $x, y\in N$,*

$$\|H(x)-H(y)\|\geqslant\alpha\|x-y\|.$$

Proof. By Proposition 2.1 and Theorem 2.1, for any $\varepsilon>0$ there exists a neighborhood N of z such that, for all $x, y\in N$,

$$\begin{aligned}\|H(x)-H(y)\|&\geqslant\|\nabla H(z)(x-y)\|-\|H(x)-H(y)-\nabla H(z)(x-y)\|\\&\geqslant\|x-y\|/\|\nabla H(z)^{-1}\|-\varepsilon\|x-y\|\\&=(1/\|\nabla H(z)^{-1}\|-\varepsilon)\|x-y\|,\end{aligned}$$

where we used

$$\|x-y\|=\|\nabla H(z)^{-1}\nabla H(z)(x-y)\|\leqslant\|\nabla H(z)^{-1}\|\,\|\nabla H(z)(x-y)\|,$$

and where the symbol $\|\cdot\|$ also denotes a matrix norm consistent with the vector norm $\|\cdot\|$ in $\mathbb{R}^n$. Thus, if $\varepsilon<1/\|\nabla H(z)^{-1}\|$, the conclusion holds with $\alpha=1/\|\nabla H(z)^{-1}\|-\varepsilon>0$. □

Lemma 2.2. *Let $H:\mathbb{R}^n\to\mathbb{R}^n$ be B-differentiable on an open, convex set $D\subset\mathbb{R}^n$, and suppose that $BH(\cdot)$ is Lipschitz continuous at $z\in D$, i.e., there exists a neighborhood N of z and some $L>0$ such that, for all $x\in N$ and all $v\in\mathbb{R}^n$ with $\|v\|=1$,*

$$\|(BH(x)-BH(z))v\|\leqslant L\|x-z\|.$$

Then, for all $x, y\in N$,

$$\|H(x)-H(y)-BH(z)(x-y)\|\leqslant L\max\{\|x-z\|,\|y-z\|\}\|x-y\|.$$

Proof. By Theorem 2.1 and Lipschitz continuity of $BH(\cdot)$,

$$\|H(x)-H(y)-BH(z)(x-y)\|\leqslant L\Big(\sup_{0\leqslant t\leqslant 1}\|(y+t(x-y))-z\|\Big)\|x-y\|.$$

The result now follows since

$$\sup_{0\leqslant t\leqslant 1}\|(y+t(x-y))-z\|\leqslant\max\{\|x-z\|,\|y-z\|\}.\qquad\square$$

3. Local convergence of quasi-Newton methods

In this section we study local convergence of quasi-Newton methods for solving B-differentiable systems of nonlinear equations. We generalize classical results in [2-6] by relaxing the standard Fréchet differentiability assumptions on the equation functions. The proofs of all the results in this section parallel those of the corresponding classical results.

Let $F:\mathbb{R}^n \to \mathbb{R}^n$ be an arbitrary function and consider a system of simultaneous nonlinear equations

$$F(x)=0.$$

General *quasi-Newton* methods for solving this system have the form [4, 5, 12],

$$x_{k+1}=x_k-B_k^{-1}F(x_k), \quad B_k \in \mathbb{R}^{n\times n}, \qquad k=0,1,\dots. \tag{3.1}$$

Denote a vector norm by $\|v\|$ for $v\in\mathbb{R}^n$ and the subordinate matrix operator norm by $\|A\|$ for $A\in\mathbb{R}^{n\times n}$. The notation $|||A|||$ for $A\in\mathbb{R}^{n\times n}$ stands for any arbitrary but fixed norm on $\mathbb{R}^{n\times n}$ which may not be subordinate to a vector norm. We shall utilize the fact that all norms on $\mathbb{R}^{n\times n}$ are equivalent. In particular, for $\|\cdot\|$ and $|||\cdot|||$ we assume for some μ, $\eta>0$ and any $A\in\mathbb{R}^{n\times n}$, that

$$\mu|||A|||\leqslant\|A\|\leqslant\eta|||A|||.$$

The following theorem extends Theorem A2.1 in Dennis and Walker [6] to B-differentiable functions.

Theorem 3.1. *Let $F: D\subset\mathbb{R}^n\to\mathbb{R}^n$ be locally Lipschitz continuous and B-differentiable at $x^*\in D$, where D is an open, convex subset of $\mathbb{R}^n$ and $F(x^*)=0$, and let $B_*\in\mathbb{R}^{n\times n}$ be such that B_*^{-1} exists and for all $v\in\mathbb{R}^n$ with $\|v\|=1$,*

$$\|[I-B_*^{-1}BF(x^*)]v\|\leqslant r_*<1. \tag{3.2}$$

Let $U:\mathbb{R}^n\times\mathbb{R}^{n\times n}\to 2^{\mathbb{R}^{n\times n}}$ be defined in a neighborhood $N=N_1\times N_2$ of (x^*, B_*), where $N_1\subset D$, $N_2\subset\{A\in\mathbb{R}^{n\times n}\,|\,A^{-1}$ exists$\}$. Suppose that there exist $\alpha_1\geqslant 0$ and $\alpha_2\geqslant 0$ such that, for each $(x,B)\in N$, and for $x_+=x-B^{-1}F(x)$, every $B_+\in U(x,B)$ satisfies

$$|||B_+-B_*|||\leqslant[1+\alpha_1\sigma(x,x_+)^p]|||B-B_*|||+\alpha_2\sigma(x,x_+)^p \tag{3.3}$$

where $\sigma(x,x_+)=\max\{\|x-x^*\|, \|x_+-x^*\|\}$ and $p\in(0,1]$.

Then, for any $r\in(r_*,1)$, there exist ε_r, $\delta_r>0$, such that if $\|x_0-x^*\|<\varepsilon_r$ and $\|B_0-B_*\|<\delta_r$, then any sequence $\{x_k\}$ defined by

$$x_{k+1}=x_k-B_k^{-1}F(x_k), \quad B_{k+1}\in U(x_k,B_k), \qquad k=0,1,\dots,$$

converges q-linearly to x^* with

$$\|x_{k+1}-x^*\|\leqslant r\|x_k-x^*\|,$$

and has the property that $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ are uniformly bounded.

Proof. Since F is B-differentiable at x^*, for any $\rho>0$, there is $\varepsilon_r>0$ such that

$$\|x-x^*\|<\varepsilon_r \Rightarrow \|F(x)-F(x^*)-BF(x^*)(x-x^*)\|/\|x-x^*\|\leqslant\rho. \tag{3.4}$$

Let $r\in(r_*,1)$ and choose δ and ρ sufficiently small so that for $\beta\geqslant\|B_*^{-1}\|$ and $\psi\geqslant\sup_{\|v\|=1}\|BF(x^*)v\|$, one has $2\beta\eta\delta<1$ and

$$r\geqslant r_*+\frac{\beta}{1-2\beta\eta\delta}(\rho+2\beta\eta\psi\delta).$$

Now, choose ε_r based on (3.4) so that

$$(2\alpha_1\delta+\alpha_2)\frac{\varepsilon_r^p}{1-r^p}\leq\delta.$$

Select δ_r small enough so that $|||B-B_*|||<\delta$ if $\|B-B_*\|<\delta_r$. Restrict ε_r, δ_r so that $(x, B)\in N$ if $\|B-B_*\|<2\eta\delta$ and $\|x-x^*\|<\varepsilon_r$. Let $\|B_0-B_*\|<\delta_r$ and $\|x_0-x^*|<\varepsilon_r$.

From standard arguments,

$$x_1-x^*=[I-B_0^{-1}\,BF(x^*)](x_0-x^*)-B_0^{-1}[F(x_0)-F(x^*)-BF(x^*)(x_0-x^*)].$$

Thus,

$$\begin{aligned}\|x_1-x^*\| &\leq \|B_0^{-1}\|\,\|F(x_0)-F(x^*)-BF(x^*)(x_0-x^*)\|\\ &\quad+\|[I-B_0^{-1}\,BF(x^*)](x_0-x^*)\|\\ &\leq \|B_0^{-1}\|\,\|F(x_0)-F(x^*)-BF(x^*)(x_0-x^*)\|\\ &\quad+\|[I-B_*^{-1}\,BF(x^*)](x_0-x^*)\|\\ &\quad+\|B_*^{-1}-B_0^{-1}\|\,\|BF(x^*)(x_0-x^*)\|\\ &\leq \|B_0^{-1}\|\,\|F(x_0)-F(x^*)-BF(x^*)(x_0-x^*)\|\\ &\quad+\|[I-B_*^{-1}\,BF(x^*)](x_0-x^*)\|\\ &\quad+\|B_0^{-1}\|\,\|B_*^{-1}\|\,\|B_0-B_*\|\,\|BF(x^*)(x_0-x^*)\|.\end{aligned}$$

It follows from the Banach perturbation lemma, since

$$\|B_*^{-1}\|\,\|B_0-B_*\|\leq\beta\eta|||B_0-B_*|||<\beta\eta\delta<2\beta\eta\delta<1,$$

that B_0^{-1} exists and $\|B_0^{-1}\|\leq\beta/(1-2\beta\eta\delta)$. Thus, from the previous inequality

$$\|x_1-x^*\|\leq\left[\frac{\beta}{1-2\beta\eta\delta}(\rho+2\beta\eta\delta\psi)+r_*\right]\|x_0-x^*\|\leq r\|x_0-x^*\|.$$

Assume now by way of induction that for $k=0,1,\ldots,m-1$, $|||B_k-B_*|||\leq 2\delta$ and $\|x_{k+1}-x^*\|\leq r\|x_k-x^*\|$. Then,

$$\begin{aligned}|||B_{k+1}-B_*|||-|||B_k-B_*|||&\leq\alpha_1\sigma(x_k,x_{k+1})^p|||B_k-B_*|||+\alpha_2\sigma(x_k,x_{k+1})^p\\ &\leq(2\alpha_1\delta+\alpha_2)\|x_k-x^*\|^p<(2\alpha_1\delta+\alpha_2)\varepsilon_r^p r^{pk}.\end{aligned}$$

Summing both sides from $k=0$ to $m-1$ we obtain

$$|||B_m-B_*|||\leq|||B_0-B_*|||+(2\alpha_1\delta+\alpha_2)\varepsilon_r^p/(1-r^p)\leq 2\delta \tag{3.5}$$

so $\|B_m-B_*\|\leq 2\eta\delta$ and again by the Banach perturbation lemma, B_m^{-1} exists and $\|B_m^{-1}\|\leq\beta/(1-2\beta\eta\delta)$. To complete the induction we proceed as for $m=0$:

$$\begin{aligned}\|x_{m+1}-x^*\|&\leq\|B_m^{-1}\|\,\|F(x_m)-F(x^*)-BF(x^*)(x_m-x^*)\|\\ &\quad+\|B_m^{-1}\|\,\|B_*^{-1}\|\,\|B_m-B_*\|\,\|BF(x^*)(x_m-x^*)\|\\ &\quad+\|[I-B_*^{-1}\,BF(x^*)](x_m-x^*)\|\\ &\leq[\|B_m^{-1}\|(\rho+\|B_*^{-1}\|\,\|B_m-B_*\|\psi)+r_*]\|x_m-x^*\|\\ &\leq\left[\frac{\beta}{1-2\beta\eta\delta}(\rho+2\beta\eta\delta\psi)+r_*\right]\|x_m-x^*\|\leq r\|x_m-x^*\|.\end{aligned}$$

Note that we have $\|B_k^{-1}\| \leq \beta/(1-2\beta\eta\delta)$ and that $\|B_k\| \leq 2\eta\delta + \|B_*\|$ which completes the proof. □

Theorem 3.1 is valid, in particular, when F is Fréchet differentiable at x^* which shows that the Hölder continuity property of $\nabla F(\cdot)$ at x^*, assumed in Theorem A2.1 in Dennis and Walker [6], can be considerably relaxed. The following corollary of Theorem 3.1 demonstrates that the classical result of Broyden, Dennis, and Moré [2, Theorem 3.2] can be extended in the same manner.

Corollary 3.1. *If $F: D \subset \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz continuous and Fréchet differentiable at $x^* \in D$, where D is an open, convex subset of $\mathbb{R}^n$, $F(x^*) = 0$, $\nabla F(x^*)^{-1}$ exists and the property* (3.3) *in Theorem* 3.1 *holds with $B_* = \nabla F(x^*)$, then the conclusions of Theorem* 3.1 *hold.* □

The next theorem extends Theorem A3.1 in Dennis and Walker [6] to B-differentiable functions.

Theorem 3.2. *Let $F: D \subset \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitz continuous and B-differentiable at $x^* \in D$, where D is an open, convex subset of $\mathbb{R}^n$ and $F(x^*) = 0$, let $\{x_k\}$ be a sequence generated by* (3.1) *which converges to x^* with $x_k \neq x^*$ for all but finitely many k and such that for some norm $\|\cdot\|_1$ and some $r \in (0, 1)$,*

$$\|x_{k+1} - x^*\|_1 \leq r\|x_k - x^*\|_1, \quad k = 0, 1, \ldots. \tag{3.6}$$

If $s_k = x_{k+1} - x_k$ and $B_* \in \mathbb{R}^{n\times n}$ is any nonsingular matrix, then the norm-independent condition

$$\lim_{k\to\infty} \frac{\|(B_k - B_*)s_k\|}{\|s_k\|} = 0 \tag{3.7}$$

holds if and only if the norm-independent condition

$$\lim_{k\to\infty} \left\| [I - B_*^{-1}\, BF(x^*)] \frac{(x_k - x^*)}{\|x_k - x^*\|} - \frac{(x_{k+1} - x^*)}{\|x_k - x^*\|} \right\| = 0$$

holds. In particular, if the condition (3.7) *holds in some norm, then for any vector norm $\|\cdot\|$, $\{x_k\}$ converges q-superlinearly to x^* if and only if*

$$\lim_{k\to\infty} \left\| [I - B_*^{-1}\, BF(x^*)] \frac{(x_k - x^*)}{\|x_k - x^*\|} \right\| = 0.$$

Proof. One has

$$(B_k - B_*)s_k = [B_* - BF(x^*)](x_k - x^*) - B_*(x_{k+1} - x^*) + BF(x^*)(x_k - x^*) - F(x_k).$$

Since $\|x_{k+1} - x^*\|_1 \leq r\|x_k - x^*\|_1$, $k = 0, 1, \ldots$, and $s_k = x_{k+1} - x_k$,

$$(1 - r)\|x_k - x^*\|_1 \leq \|s_k\|_1 \leq (1 + r)\|x_k - x^*\|_1.$$

By the B-differentiability of F at x^*,

$$\lim_{k\to\infty} \frac{\|BF(x^*)(x_k-x^*)-F(x_k)\|_1}{\|s_k\|_1} \leq \lim_{k\to\infty} \frac{\|F(x_k)-F(x^*)-BF(x^*)(x_k-x^*)\|_1}{(1-r)\|x_k-x^*\|_1}=0.$$

Thus, $\lim_{k\to\infty}\|(B_k-B_*)s_k\|_1/\|s_k\|_1=0$ if and only if

$$\lim_{k\to\infty} \frac{\|[B_*-BF(x^*)](x_k-x^*)-B_*(x_{k+1}-x^*)\|_1}{\|s_k\|_1}=0.$$

Further, one has

$$\frac{\|[B_*-BF(x^*)](x_k-x^*)-B_*(x_{k+1}-x^*)\|_1}{\|s_k\|_1} \leq \frac{\|B_*\|_1}{(1-r)} \left\| [I-B_*^{-1}\,BF(x^*)]\frac{(x_k-x^*)}{\|x_k-x^*\|_1} - \frac{(x_{k+1}-x^*)}{\|x_k-x^*\|_1} \right\|_1,$$

and

$$\frac{\|[B_*-BF(x^*)](x_k-x^*)-B_*(x_{k+1}-x^*)\|_1}{\|s_k\|_1} \geq \frac{1}{\|B_*^{-1}\|_1(1+r)} \left\| [I-B_*^{-1}\,BF(x^*)]\frac{(x_k-x^*)}{\|x_k-x^*\|_1} - \frac{(x_{k+1}-x^*)}{\|x_k-x^*\|_1} \right\|_1.$$

Thus, the first conclusion of the theorem holds and the second conclusion follows trivially since only norm-independent "zero" limits have been used. □

Similarly to Theorem 3.1, Theorem 3.2 is true, in particular, when F is Fréchet differentiable at x^* which shows that the Hölder continuity property of $\nabla F(\cdot)$ at x^*, assumed in Theorem A3.1 in Dennis and Walker [6], can be relaxed.

In the case where $B_*=\nabla F(x^*)$, Theorem 3.2 can be strengthened by replacing the linear convergence assumption (3.6) by ordinary convergence and imposing a local continuity assumption on the derivative of F. This result is given next and it extends Theorem 2.2 in Dennis and Moré [3] to B-differentiable functions.

Theorem 3.3. *Suppose that $F: D\subset \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous and B-differentiable on D, where D is an open, convex subset of $\mathbb{R}^n$, that for some $x^*\in D$, $BF(\cdot)$ is continuous at x^*, and that $\nabla F(x^*)$ is nonsingular ($\nabla F(x^*)$ exists by Theorem 2.1). Let $\{B_k\}$ be a sequence of nonsingular matrices in $\mathbb{R}^{n\times n}$ and suppose that for some $x_0\in D$ the sequence $\{x_k\}$ generated by (3.1) remains in D and converges to x^*, where $x_k\neq x^*$ for all but finitely many k.*

Then, if $s_k=x_{k+1}-x_k$, $\{x_k\}$ converges q-superlinearly to x^* and $F(x^*)=0$ if and only if

$$\lim_{k\to\infty} \frac{\|(B_k-\nabla F(x^*))s_k\|}{\|s_k\|}=0. \tag{3.8}$$

Proof. Suppose that (3.8) holds. Since

$$(B_k - \nabla F(x^*))s_k = F(x_{k+1}) - F(x_k) - \nabla F(x^*)(x_{k+1} - x_k) - F(x_{k+1}), \quad (3.9)$$

the continuity of $BF(\cdot)$ at x^* and (3.8) imply, in view of Theorem 2.1, that

$$\lim_{k\to\infty} \|F(x_{k+1})\| / \|x_{k+1} - x_k\| = 0. \quad (3.10)$$

Thus, $F(x^*) = 0$, and since $\nabla F(x^*)$ is nonsingular, by Lemma 2.1 there exists $\alpha > 0$ such that

$$\|F(x_{k+1})\| = \|F(x_{k+1}) - F(x^*)\| \geq \alpha \|x_{k+1} - x^*\|.$$

Therefore,

$$\frac{\|F(x_{k+1})\|}{\|x_{k+1} - x_k\|} \geq \frac{\alpha \|x_{k+1} - x^*\|}{\|x_{k+1} - x^*\| + \|x_k - x^*\|} = \alpha \frac{\rho_k}{1 + \rho_k},$$

where $\rho_k = \|x_{k+1} - x^*\| / \|x_k - x^*\|$. Thus, (3.10) implies that $\rho_k/(1+\rho_k)$ converges to zero and hence $\{\rho_k\}$ also converges to zero as desired.

Suppose now that $\{x_k\}$ converges q-superlinearly to x^* and $F(x^*) = 0$. Observe that

$$\frac{\|F(x_{k+1})\|}{\|x_{k+1} - x_k\|} = \frac{\|F(x_{k+1}) - F(x^*)\|}{\|x_k - x^*\|} \cdot \frac{\|x_k - x^*\|}{\|x_{k+1} - x_k\|}$$

$$= \frac{\|F(x_{k+1}) - F(x^*)\|}{\|x_{k+1} - x^*\|} \cdot \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \cdot \frac{\|x_k - x^*\|}{\|x_{k+1} - x_k\|}.$$

Since F is Lipschitz continuous on D and $\lim_{k\to\infty} \|x_{k+1} - x_k\| / \|x_k - x^*\| = 1$ (see [3]), (3.10) holds. It then follows from (3.9) that (3.8) is satisfied. □

4. Broyden's method

In this section we study local convergence properties of Broyden's method for solving a system of nonlinear equations $F(x) = 0$ where $F: \mathbb{R}^n \to \mathbb{R}^n$. We generalize classical results in [2, 5] by relaxing standard Fréchet differentiability assumptions on the function F.

Broyden's method for solving the system $F(x) = 0$ has the form [1, 2, 5]

$$x_{k+1} = x_k + s_k, \quad s_k = -B_k^{-1} F(x_k), \qquad k = 0, 1, \ldots, \quad (4.1)$$

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k) s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} s_k}, \quad y_k = F(x_{k+1}) - F(x_k). \quad (4.2)$$

Broyden's method is one of the most efficient and robust quasi-Newton methods and it has the key q-superlinear convergence property under appropriate assumptions.

Denote by $\|v\|_2$ the l_2 vector norm of $v \in \mathbb{R}^n$ and by $\|A\|_2$ the l_2 matrix norm of $A \in \mathbb{R}^{n \times n}$, respectively. We shall also refer to the Frobenius norm of matrix A which is defined as the l_2 norm of A written as a vector [5].

The following lemma extends Lemma 8.2.1 in Dennis and Schnabel [5] (see also Broyden et al. [2]) to B-differentiable functions.

Lemma 4.1. *Let $F:\mathbb{R}^n\to\mathbb{R}^n$ be Lipschitz continuous and B-differentiable on an open, convex set $D\subset\mathbb{R}^n$. Suppose also that $BF(\cdot)$ is Lipschitz continuous at $x^*\in D$ (with Lipschitz constant $L>0$) with respect to the l_2 vector norm (this implies, by Theorem 2.1, that $\nabla F(x^*)$ exists). Define*

$$A_+ = A_c + \frac{(y_c - A_c s_c)s_c^T}{s_c^T s_c}, \quad \textit{where}$$

$$A_c s_c = -F(x_c), \quad y_c = F(x_+) - F(x_c), \quad x_+ = x_c + s_c,$$

and assume that $x_c, x_+\in D$, $x_c\neq x_+$.

Then,

$$\|A_+ - \nabla F(x^*)\| \leq \|A_c - \nabla F(x^*)\| + L\max\{\|x_+ - x^*\|_2, \|x_c - x^*\|_2\} \tag{4.3}$$

where $\|\cdot\|$ is either the Frobenius or the l_2 matrix norm.

Proof. By the definition of A_+,

$$A_+ - \nabla F(x^*) = A_c - \nabla F(x^*) + \frac{(y_c - A_c s_c)s_c^T}{s_c^T s_c}$$

$$= (A_c - \nabla F(x^*))\left[I - \frac{s_c s_c^T}{s_c^T s_c}\right] + \frac{(y_c - \nabla F(x^*)s_c)s_c^T}{s_c^T s_c}.$$

For either the Frobenius or the l_2 matrix norm, it follows [5] that

$$\|A_+ - \nabla F(x^*)\| \leq \|A_c - \nabla F(x^*)\| \left\| I - \frac{s_c s_c^T}{s_c^T s_c}\right\|_2 + \frac{\|y_c - \nabla F(x^*)s_c\|_2}{\|s_c\|_2}.$$

Using $\|I - s_c s_c^T/s_c^T s_c\|_2 = 1$ [5] and

$$\|y_c - \nabla F(x^*)s_c\|_2 \leq L\max\{\|x_+ - x^*\|_2, \|x_c - x^*\|_2\}\|s_c\|_2$$

concludes the proof. The last inequality follows from Lemma 2.2 by setting $x = x_+$, $y = x_c$, and $z = x^*$. □

Observe that formula (4.3), called the bounded deterioration property [5], is a special case of formula (3.3) in Theorem 3.1.

The next theorem extends Theorem 8.2.2 in Dennis and Schnabel [5] (see also Broyden et al. [2, Theorem 4.3]) to B-differentiable functions.

Theorem 4.1. *Let $F:\mathbb{R}^n\to\mathbb{R}^n$ be Lipschitz continuous and B-differentiable on an open, convex set $D\subset\mathbb{R}^n$, and suppose that $BF(\cdot)$ is Lipschitz continuous at $x^*\in D$ (with respect to the l_2 vector norm), that $F(x^*)=0$, and $\nabla F(x^*)^{-1}$ exists.*

Then, there exist $\varepsilon>0$ and $\delta>0$ such that, if $\|x_0 - x^*\|_2\leq\varepsilon$ and $\|B_0 - \nabla F(x^*)\|_2\leq\delta$, then the sequence $\{x_k\}$ generated by (4.1)–(4.2) is well defined and converges q-superlinearly to x^*.

Proof. By Lemma 4.1, formulas (3.2) and (3.3) in Theorem 3.1 hold with $B_* = \nabla F(x^*)$. Thus, the conclusions of Theorem 3.1 hold and, in particular, $\{x_k\}$ is well defined and converges q-linearly to x^* and formula (3.5) holds. The conclusion now follows from Theorem 3.3 with the proof proceeding along the same lines as the proof of q-superlinear convergence in Theorem 8.2.2 in [5]. □

Another extension of Theorem 8.2.2 in Dennis and Schnabel [5] was recently proposed by Kojima and Shindo [11]. They considered a system of piecewise continuously differentiable equations and modified Broyden's method in such way that the formulas (4.1)–(4.2) were applied as long as the points x_k stayed within a given C^1-piece (where the equation functions were continuously differentiable) and an appropriate initial starting matrix B_0 was used when the point x_k moved to a new piece. Their local convergence results states that q-superlinear convergence of $\{x_k\}$ is preserved if the assumptions of the classical theorem hold for each C^1-piece. Note that while they do not require the existence of $\nabla F(x^*)$, their method could be computationally less efficient since it requires storing a potentially large number of matrices (up to 2^n, the number of possible C^1-pieces) in addition to other bookkeeping chores.

Another Newton-type method which does not require the existence of $\nabla F(x^*)$ was proposed by Qi and Sun [15]. This method generalizes the Newton-type method of Pang [13] to nonlinear equations defined by semismooth functions and utilizes the generalized Jacobian $\partial F(x^*)$ instead of $\nabla F(x^*)$. The authors obtain local q-superlinear convergence under the assumption that $\partial F(x^*)$ is nonsingular. While this method bears a certain resemblance to quasi-Newton methods studied here, it is fundamentally different since its iterations require Jacobian matrix information in contrast to simple updates used by quasi-Newton methods.

5. Application to nonlinear complementarity problems

In this section we propose a new method for solving nonlinear complementarity problems. The method applies Broyden's algorithm to two related formulations of the nonlinear complementarity problem as a system of B-differentiable equations. The results of Section 4 are used to analyze its local convergence properties.

A nonlinear complementarity problem has the form

NCP: find x such that $x \geq 0,\ F(x) \geq 0,\ x^{\mathrm{T}}F(x) = 0,$

where $F: \mathbb{R}^n \to \mathbb{R}^n$. Pang [13] defined the following function $H: \mathbb{R}^n \to \mathbb{R}^n$,

$$H(x) = \min(x, F(x)), \tag{5.1}$$

where the "min" operator denotes the componentwise minimum of two vectors, and observed that x solves NCP if and only if

$$H(x) = 0. \tag{5.2}$$

Thus, the nonlinear complementarity problem NCP can be converted to an equivalent system of nonlinear equations (5.2).

Pang [13] recently developed a novel Newton method for solving a B-differentiable nonlinear system of equations and applied it to nonlinear complementarity problems via system (5.2). His method requires a solution of a mixed linear complementarity problem in each iteration and, under appropriate assumptions, is shown to have the q-quadratic local convergence property. A globalized version of the method with line-search is also shown in [13] to have the global convergence property under certain assumptions. We propose to apply Broyden's method described in Section 4 to the system (5.2) and show that, under assumptions similar to those in [13], it exhibits q-superlinear local convergence to the solution point. Our method has the advantage of requiring only the rank one update of the inverse of the matrix B_k in each iteration which is $O(n^2)$. In contrast, the Newton method in Pang [13] solves in each iteration either a mixed linear complementarity problem which potentially requires an exponential number of computations (in case of degeneracy) or a system of linear equations requiring $O(n^3)$ computations (in case of nondegeneracy). Harker and Pang [7] and Pang [13] also survey and discuss earlier Newton and quasi-Newton methods for solving the problem NCP. In general, all those methods have similar local convergence properties and they all solve a certain mixed linear complementarity problem in each iteration.

The following theorem proved in Pang [13] summarizes the properties of the function H.

Theorem 5.1 [13]. *Let $F:\mathbb{R}^n \to \mathbb{R}^n$ be Fréchet differentiable. Then:*

(1) *The function H, defined in (5.1), is everywhere B-differentiable with the B-derivative given by*

$$(BH(z)v)_i = \begin{cases} \nabla F_i(z)^{\mathrm{T}} v & \text{if } i \in \alpha(z), \\ \min(\nabla F_i(z)^{\mathrm{T}} v,\, e_i^{\mathrm{T}} v) & \text{if } i \in \beta(z), \\ e_i^{\mathrm{T}} v & \text{if } i \in \gamma(z), \end{cases}$$

where $\alpha(z) = \{i \mid F_i(z) < z_i\}$, $\beta(z) = \{i \mid F_i(z) = z_i\}$, $\gamma(z) = \{i \mid F_i(z) > z_i\}$, and e_i is the ith unit vector.

(2) *H is Fréchet differentiable at a point z if and only if for each index $i \in \beta(z)$, $\nabla F_i(z) = e_i$. In particular, this holds if $\beta(z)$ is empty.*

(3) *If $\nabla F(\cdot)$ is Lipschitz continuous at z, and if H is Fréchet differentiable at z, then the B-derivative $BH(\cdot)$ is Lipschitz continuous at z.* □

Suppose that x^* solves the problem NCP and define the index sets

$$I = \{i \mid x_i^* = 0,\ F_i(x^*) > 0\}, \qquad J = \{i \mid x_i^* > 0,\ F_i(x^*) = 0\},$$

$$L = \{i \mid x_i^* = 0,\ F_i(x^*) = 0\}.$$

For any index sets T and Z, define $\nabla_T F_Z(x^*)$ to be the matrix $[\partial F_i(x^*)/\partial x_j]$, $i \in Z$, $j \in T$. Note that $I = \gamma(x^*)$, $J = \alpha(x^*)$, $L = \beta(x^*)$ and that if H is Fréchet differentiable at x^*, then, by Theorem 5.1,

$$\nabla H(x^*) = \begin{bmatrix} \nabla_J F_J(x^*) & \nabla_{I\cup L} F_J(x^*) \\ 0 & I_m \end{bmatrix}, \tag{5.3}$$

where the identity matrix I_m is $m \times m$ with $m = |I| + |L|$ ($|Z|$ denotes the cardinality of set Z).

The next theorem shows that Broyden's method (4.1)-(4.2) applied to the system (5.2) is locally q-superlinearly convergent under appropriate assumptions on the Jacobian $\nabla F(x^*)$.

Theorem 5.2. *Let $F: \mathbb{R}^n \to \mathbb{R}^n$ be Fréchet differentiable in a neighborhood of x^* and let $\nabla F(\cdot)$ be Lipschitz continuous at x^* (with respect to the l_2 vector norm). Suppose that x^* solves the nonlinear complementarity problem* NCP *and that the following assumptions hold*:

(i) *$\nabla_J F_J(x^*)$ is nonsingular*;

(ii) *$\nabla F_i(x^*) = e_i$ for all $i \in L$.*

Then, there exist $\varepsilon > 0$ and $\delta > 0$ such that, if $\|x_0 - x^*\|_2 \leq \varepsilon$ and $\|B_0 - \nabla H(x^*)\|_2 \leq \delta$, then the sequence $\{x_k\}$ generated by (4.1)-(4.2) *applied to the system $H(x) = 0$ is well defined and converges q-superlinearly to x^*.*

Proof. The result will follow from Theorem 4.1 if we can show that the assumptions of Theorem 4.1 hold for H. By Theorem 5.1, assumption (ii) implies that H is Fréchet differentiable at x^* and $BH(\cdot)$ is Lipschitz continuous at x^*. Finally, in view of formula (5.3), assumption (i) implies that $\nabla H(x^*)^{-1}$ exists. □

Observe that if x^* is a nondegenerate solution of NCP, i.e., $L = \emptyset$, then assumption (ii) holds automatically. For degenerate solutions, however, assumption (ii) imposes a certain condition on the degenerate part of $\nabla F(x^*)$ which substantially restricts the class of functions F to which Theorem 5.2 applies. The same restrictive assumption (ii) is imposed by Pang [13] to prove quadratic convergence of the generalized Newton method applied to the system $H(x) = 0$. However, under a strengthened assumption (i), Pang was able to prove in a subsequent paper [14] quadratic convergence rate of an arbitrary locally convergent sequence for a modified Newton method applied to $H(x) = 0$ without assuming condition (ii).

Another reformulation of NCP as a system of nonlinear equations uses the concept of a Minty-map [11], i.e., the function $G: \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$G(z) = F(z^+) + z^-, \tag{5.4}$$

where $z_i^+ = \max(z_i, 0)$, $z_i^- = \min(z_i, 0)$, $z^+ = (z_1^+, \ldots, z_n^+)^T$, and $z^- = (z_1^-, \ldots, z_n^-)^T$. It is easy to verify that there is a one-to-one correspondence between a solution z of the system of equations

$$G(z) = F(z^+) + z^- = 0 \tag{5.5}$$

and a solution x of NCP where $x = z^+$. Thus, the problem NCP can be converted to an equivalent system of nonlinear equations different from (5.2).

Harker and Xiao [8] applied the Newton method of Pang [13] for solving a B-differentiable nonlinear system of equations to nonlinear complementarity problems via system (5.5). Their method requires a solution of the same mixed linear complementarity problem in each iteration as Pang's method, which uses the system (5.2), and has similar local convergence properties (a globalized version of the method is also shown in [8] to converge globally to nondegenerate solutions). We again apply Broyden's method from Section 4 to the system (5.5) and show that it exhibits q-superlinear local convergence to the solution point.

The next theorem summarizes the properties of the function G. Conclusions (1)-(2) were proved in Harker and Xiao [8] while conclusion (3) is new.

Theorem 5.3 [8]. *Let $F:\mathbb{R}^n \to \mathbb{R}^n$ be Fréchet differentiable. Then:*

(1) *The function G, defined in (5.4), is everywhere B-differentiable with the B-derivative given by*

$$(BG(z)v)_i = \sum_{j=1}^{n} BG_i^j(z)v_j,$$

$$BG_i^j(z)v_j = \begin{cases} F_{ij}(z^+)v_j & \text{if } j \in \alpha_0(z), \\ F_{ij}(z^+)v_j^+ + I_{ij}v_j^- & \text{if } j \in \beta_0(z), \\ I_{ij}v_j & \text{if } j \in \gamma_0(z), \end{cases} \tag{5.6}$$

where $\alpha_0(z) = \{i \mid z_i > 0\}$, $\beta_0(z) = \{i \mid z_i = 0\}$, $\gamma_0(z) = \{i \mid z_i < 0\}$, $F_{ij}(z^+) = \partial F_i/\partial x_j(z^+)$, and $I_{ij} = 1$ if $i = j$, $I_{ij} = 0$ if $i \neq j$.

(2) *G is Fréchet differentiable at a point z if and only if for each index $j \in \beta_0(z)$, $F_{ij}(z^+) = I_{ij}$, $i = 1, \ldots, n$. In particular, this holds if $\beta_0(z)$ is empty.*

(3) *If $\nabla F(\cdot)$ is Lipschitz continuous at z^+, and if G is Fréchet differentiable at z, then the B-derivative $BG(\cdot)$ is Lipschitz continuous at z.*

Proof. We only need to prove conclusion (3). Observe that for an arbitrary point y near z one has

$$\alpha_0(y) = \alpha_0(z) \cup (\alpha_0(y) \cap \beta_0(z)),$$
$$\gamma_0(y) = \gamma_0(z) \cup (\gamma_0(y) \cap \beta_0(z)), \qquad \beta_0(y) \subset \beta_0(z).$$

In view of (5.6) and the Fréchet differentiability of G at z one may write

$$\begin{aligned} &BG_i(y)v - BG_i(z)v \\ &= \sum_{j\in\alpha_0(y)} F_{ij}(y^+)v_j + \sum_{j\in\beta_0(y)} [F_{ij}(y^+)v_j^+ + I_{ij}v_j^-] + \sum_{j\in\gamma_0(y)} I_{ij}v_j \\ &\quad - \sum_{j\in\alpha_0(z)} F_{ij}(z^+)v_j - \sum_{j\in\alpha_0(y)\cap\beta_0(z)} F_{ij}(z^+)v_j \\ &\quad - \sum_{j\in\beta_0(y)\cap\beta_0(z)} [F_{ij}(z^+)v_j^+ + I_{ij}v_j^-] - \sum_{j\in\gamma_0(y)\cap\beta_0(z)} I_{ij}v_j - \sum_{j\in\gamma_0(z)} I_{ij}v_j \\ &= \sum_{j\in\alpha_0(y)} [F_{ij}(y^+) - F_{ij}(z^+)]v_j + \sum_{j\in\beta_0(y)} [F_{ij}(y^+) - F_{ij}(z^+)]v_j^+. \end{aligned}$$

Lipschitz continuity of $BG_i(\cdot)$ at z now follows since $\|y^+ - z^+\| \leq \|y - z\|$ and $\|v^+\| \leq \|v\|$. □

Remark. Since Lipschitz continuity of $BG(\cdot)$ at z implies in particular that the B-derivative $BG(z)$ is strong in the sense of Theorem 2.1 (3), conclusion (3) extends part (c) of Theorem 1 in Harker and Xiao [8] where the index set $\beta_0(z)$ was assumed empty. Thus, the key global convergence result for the damped-Newton method in [8, Theorem 3] can be generalized to the class of functions for which the assumption in conclusion (2) holds (the same is true for the corresponding local convergence result for their Newton method which follows from the results in [13]).

As before, suppose that x^* solves the problem NCP and define the index sets I, J, and L. If we define z^* by

$$z_I^* = -F_I(x^*), \quad z_J^* = x_J^*, \quad z_L^* = x_L^*, \tag{5.7}$$

then it is easy to see that z^* solves the system $G(z) = 0$. Note also that $I = \gamma_0(z^*)$, $J = \alpha_0(z^*)$, $L = \beta_0(z^*)$ and that if G is Fréchet differentiable at z^*, then, by Theorem 5.3,

$$\nabla G(z^*) = \begin{bmatrix} \nabla_J F_J(x^*) & 0 \\ \nabla_J F_{I \cup L}(x^*) & I_m \end{bmatrix},$$

where $x^* = (z^*)^+$ and the identity matrix I_m is $m \times m$ with $m = |I| + |L|$.

The next theorem shows that Broyden's method (4.1)–(4.2) applied to the system (5.5) is locally q-superlinearly convergent under appropriate assumptions on the Jacobian $\nabla F(x^*)$.

Theorem 5.4. *Let $F: \mathbb{R}^n \to \mathbb{R}^n$ be Fréchet differentiable in a neighborhood of x^* and let $\nabla F(\cdot)$ be Lipschitz continuous at x^* (with respect to the l_2 vector norm). Suppose that x^* solves the nonlinear complementarity problem* NCP *and that the following assumptions hold:*

(i) *$\nabla_J F_J(x^*)$ is nonsingular;*

(ii) *For all $j \in L$, $i = 1, \ldots, n$, $\partial F_i / \partial x_j(x^*) = 1$ if $i = j$ and $\partial F_i / \partial x_j(x^*) = 0$ if $i \neq j$.*

Then, there exist $\varepsilon > 0$ and $\delta > 0$ such that, if $\|z_0 - z^*\|_2 \leq \varepsilon$ and $\|B_0 - \nabla G(z^*)\|_2 \leq \delta$, where z^* is defined in (5.7), *then the sequence $\{z_k\}$ generated by* (4.1)–(4.2) *applied to the system $G(z) = 0$ is well defined and converges q-superlinearly to z^* which in particular implies that the sequence $\{x_k = z_k^+\}$ is well defined and converges q-superlinearly to $x^* = (z^*)^+$.*

Proof. The proof of this result is the same as that of Theorem 5.2 except that Theorem 5.3 is used here instead of Theorem 5.1. □

The remarks that follow Theorem 5.2 apply to Theorem 5.4 as well.

6. Computational examples

This section contains computational results obtained for two small nonlinear complementarity problems using Broyden's method applied to two formulations of these problems discussed in the previous section.

Problem 1 (A Nondegenerate Nonlinear Complementarity Problem, [9]). Consider the following problem: find $x \in \mathbb{R}^4$ such that $x \geq 0$, $F(x) \geq 0$ and $x^T F(x) = 0$ where $F: \mathbb{R}^4 \to \mathbb{R}^4$ is given by

$$F_1(x) = 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6,$$

$$F_2(x) = 2x_1^2 + x_1 + x_2^2 + 3x_3 + 2x_4 - 2,$$

$$F_3(x) = 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 3x_4 - 1,$$

$$F_4(x) = x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3.$$

This problem has a solution

$$x^* = (\tfrac{1}{2}\sqrt{6} \approx 1.2247, 0, 0, 0.5), \qquad F(x^*) = (0, 2 + \tfrac{1}{2}\sqrt{6} \approx 3.2247, 5, 0).$$

Since $L = \emptyset$, the solution x^* is nondegenerate and it is easy to check that the assumptions of Theorems 5.2 and 5.4 hold at x^*.

Problem 2 (A Degenerate Nonlinear Complementarity Problem, [11]). Consider the following problem: find $x \in \mathbb{R}^4$ such that $x \geq 0$, $F(x) \geq 0$ and $x^T F(x) = 0$ where $F: \mathbb{R}^4 \to \mathbb{R}^4$ is given by

$$F_1(x) = 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6,$$

$$F_2(x) = 2x_1^2 + x_1 + x_2^2 + 10x_3 + 2x_4 - 2,$$

$$F_3(x) = 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 9x_4 - 9,$$

$$F_4(x) = x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3.$$

This problem has two solutions

$$x_D^* = (\tfrac{1}{2}\sqrt{6} \approx 1.2247, 0, 0, 0.5), \qquad F(x_D^*) = (0, 2 + \tfrac{1}{2}\sqrt{6} \approx 3.2247, 0, 0),$$

and

$$x_{ND}^* = (1, 0, 3, 0), \qquad F(x_{ND}^*) = (0, 31, 0, 4).$$

Since $L = \emptyset$ for the solution x_{ND}^*, it is a nondegenerate solution and it is easy to check that the assumptions of Theorems 5.2 and 5.4 hold at x_{ND}^*. On the other hand, $L = \{3\} \neq \emptyset$ for the solution x_D^*, so it is a degenerate solution and it can be verified that the assumptions (ii) of Theorems 5.2 and 5.4 do not hold at x_D^*.

Table 1

Results for problems 1 and 2

Algorithm	Starting point	Number of iterations	
		Problem 1	Problem 2
MIN	(0, 0, 0, 0)	15	27 (ND)
MINTY	(0, 0, 0, 0)	16	20 (ND)
MIN	(1, 1, 1, 1)	15	14 (ND)
MINTY	(1, 1, 1, 1)	16	67 (D)
MIN	(3, 3, 3, 3)	17	52 (ND)
MINTY	(3, 3, 3, 3)	38	45 (ND)
MIN	(5, 5, 5, 5)	17	21 (ND)
MINTY	(5, 5, 5, 5)	41	45 (ND)
MIN	(100, 100, 100, 100)	16	25 (ND)
MINTY	(100, 100, 100, 100)	44	49 (ND)
MIN	(−2, −2, −2, −2)	24	19 (ND)
MINTY	(−2, −2, −2, −2)	19	failed
MIN	(1, 0, 1, 0)	14	37 (ND)
MINTY	(1, 0, 1, 0)	339	126 (D)
MIN	(10, 0, 10, 0)	18	43 (ND)
MINTY	(10, 0, 10, 0)	43	51 (ND)
MIN	(1, 0, 0, 0)	20	36 (ND)
MINTY	(1, 0, 0, 0)	15	17 (ND)
MIN	(10, 0, 0, 0)	18	19 (ND)
MINTY	(10, 0, 0, 0)	46	47 (ND)
MIN	(0, 1, 1, 0)	12	30 (ND)
MINTY	(0, 1, 1, 0)	failed	491 (D)
MIN	(0, 10, 10, 0)	17	19 (ND)
MINTY	(0, 10, 10, 0)	44	46 (ND)
MIN	(1, −1, −1, 1)	15	35 (ND)
MINTY	(1, −1, −1, 1)	12	27 (D)
MIN	(−1, 1, 1, −1)	16	44 (ND)
MINTY	(−1, 1, 1, −1)	36	failed
MIN	(10, −10, −10, 10)	17	38 (ND)
MINTY	(10, −10, −10, 10)	43	failed
MIN	(1, −1, 1, −1)	16	38 (ND)
MINTY	(1, −1, 1, −1)	failed	71 (D)
MIN	(10, −10, 10, −10)	17	68 (ND)
MINTY	(10, −10, 10, −10)	43	49 (ND)

D = degenerate solution, ND = nondegenerate solution.

The above two problems were tested using Broyden's method applied to Pang's formulation [13] of NCP given in (5.2), which is called "MIN", and Harker–Xiao formulation [8] of NCP given in (5.5), which is called "MINTY". Table 1 exhibits numerical results obtained using both versions of the method with the same 17 starting points and the initial matrix B_0 set to be the identity matrix.

As the results in Table 1 show, Broyden's method is surprisingly robust in that it almost always converges to a solution of the problem regardless of the starting point chosen. They also indicate that for both Problems 1 and 2 the "MIN" method typically converges faster to a solution than the "MINTY" method. In fact, the "MIN" method has successfully found a solution in all cases while the "MINTY" method failed in several instances. It is also interesting to note that the "MIN" method always found the nondegenerate solution of Problem 2 in contrast to the "MINTY" method which was able to find the degenerate solution of Problem 2 in several cases. These results, of course, do not necessarily imply superiority of the "MIN" method over the "MINTY" method since they are restricted to two small problems. In fact, the results in Harker and Xiao [8] for the modified Newton method suggest that the "MINTY" approach may have some computational advantages over the "MIN" approach.

In general, our results compare favorably with those reported in Josephy [9] and Harker and Xiao [8] for Problem 1 and the results in Kojima and Shindo [11] for Problem 2. This suggests that the method holds promise and that more extensive tests on larger problems should be performed to establish its efficiency.

References

[1] C.G. Broyden, "A class of methods for solving nonlinear simultaneous equations," *Mathematics of Computation* 19 (1965) 577–593.

[2] C.G. Broyden, J.E. Dennis Jr. and J.J. Moré, "On the local and superlinear convergence of quasi-Newton methods," *Journal of the Institute of Mathematics and its Applications* 12 (1973) 223–245.

[3] J.E. Dennis Jr. and J.J. Moré, "A characterization of superlinear convergence and its application to quasi-Newton methods," *Mathematics of Computation* 28 (1974) 549–560.

[4] J.E. Dennis Jr. and J.J. Moré, "Quasi-Newton methods, motivation and theory" *SIAM Review* 19 (1977) 46–89.

[5] J.E. Dennis Jr. and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Prentice-Hall, Englewood Cliffs, NJ, 1983).

[6] J.E. Dennis Jr. and H.F. Walker, "Convergence theorems for least-change secant update methods," *SIAM Journal on Numerical Analysis* 18 (1981) 949–987.

[7] P.T. Harker and J.S. Pang, "Finite-dimensional variational inequality and non-linear complementarity problems: A survey of theory, algorithms and applications," *Mathematical Programming* (*Series B*) 48 (1990) 161–220.

[8] P.T. Harker and B. Xiao, "Newton's method for the nonlinear complementarity problem: A B-differentiable equation approach," *Mathematical Programming* (*Series B*) 48 (1990) 339–357.

[9] N.H. Josephy, "Newton's method for generalized equations," Technical Summary Report No. 1965, Mathematics Research Center, University of Wisconsin (Madison, WI, 1979).

[10] N.H. Josephy, "Quasi-Newton methods for generalized equations," Technical Summary Report No. 1966, Mathematics Research Center, University of Wisconsin (Madison, WI, 1979).

[11] M. Kojima and S. Shindo, "Extensions of Newton and quasi-Newton methods to systems of PC^1 equations," *Journal of Operations Research Society of Japan* 29 (1986) 352–374.

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

[13] J.S. Pang, "Newton's method for B-differentiable equations," *Mathematics of Operations Research* 15 (1990) 311–341.

[14] J.S. Pang, "A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems," *Mathematical Programming* 51 (1991) 101–131.

[15] L. Qi and J. Sun, "A nonsmooth version of Newton's method and an interior point algorithm for convex programming," Technical Report 90-01, Department of Industrial Engineering and Management Sciences, Northwestern University (Evanston, IL, 1990).

[16] S.M. Robinson, "Local structure of feasible sets in nonlinear programming, part III: stability and sensitivity," *Mathematical Programming Study* 30 (1987) 45–66.

[17] S.M. Robinson, "Newton's method for a class of nonsmooth functions," Working paper, Department of Industrial Engineering, University of Wisconsin (Madison, WI, 1988).