SIAM J. NUMER. ANAL.
Vol. 40, No. 5, pp. 1763–1774

DESCENT DIRECTIONS OF QUASI-NEWTON METHODS FOR SYMMETRIC NONLINEAR EQUATIONS*

GUANG-ZE GU†, DONG-HUI LI†‡, LIQUN QI‡, AND SHU-ZI ZHOU†

Abstract. In general, when a quasi-Newton method is applied to solve a system of nonlinear equations, the quasi-Newton direction is not necessarily a descent direction for the norm function. In this paper, we show that when applied to solve symmetric nonlinear equations, a quasi-Newton method with positive definite iterative matrices may generate descent directions for the norm function. On the basis of a Gauss–Newton based BFGS method [D. H. Li and M. Fukushima, *SIAM J. Numer. Anal.*, 37 (1999), pp. 152–172], we develop a norm descent BFGS method for solving symmetric nonlinear equations. Under mild conditions, we establish the global and superlinear convergence of the method. The proposed method shares some favorable properties of the BFGS method for solving unconstrained optimization problems: (a) the generated sequence of the quasi-Newton matrices is positive definite; (b) the generated sequence of iterates is norm descent; (c) a global convergence theorem is established without nonsingularity assumption on the Jacobian. Preliminary numerical results are reported, which positively support the method.

Key words. BFGS method, norm descent direction, global convergence, superlinear convergence

AMS subject classifications. 65H10, 90C53

PII. S0036142901397423

1. Introduction.

Let $F : R^n \to R^n$ be continuously differentiable. A general quasi-Newton method for solving the system of nonlinear equations

$$F(x) = 0 \tag{1.1}$$

generates a sequence of iterates $\{x_k\}$ by letting $x_{k+1} = x_k + d_k$, where d_k is a solution of the following system of linear equations:

$$B_k d + F(x_k) = 0. \tag{1.2}$$

If in (1.2), matrix B_k is replaced by $F'(x_k)$, the Jacobian of the function F at x_k, the method reduces to the well-known Newton method. An attractive feature of a quasi-Newton method is its local superlinear convergence property without computation of Jacobians. To enlarge the convergence domain of a quasi-Newton method, line search technique or trust region strategy can be exploited. In this paper, we use a backtracking line search technique to globalize a quasi-Newton method.

A line search step at iteration k of an iterative method determines a scalar $\lambda_k > 0$ which satisfies

$$\|F(x_k + \lambda_k d_k)\| < \|F(x_k)\|. \tag{1.3}$$

The next iterate is then determined by letting $x_{k+1} = x_k + \lambda_k d_k$. The scalar λ_k is called the steplength. Let θ be the norm function defined by

$$\theta(x) = \frac{1}{2}\|F(x)\|^2. \tag{1.4}$$

*Received by the editors November 5, 2001; accepted for publication (in revised form) May 28, 2002; published electronically November 14, 2002. This work was partially supported by the NSF (10171030) of China and the RGC of Hong Kong.

http://www.siam.org/journals/sinum/40-5/39742.html

†Institute of Applied Mathematics, Hunan University, Changsha 410082, China (szzhou@hunu.edu.cn).

‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (madhli@polyu.edu.hk, maqilq@polyu.edu.hk).

Then the nonlinear equation problem (1.1) is equivalent to the following global optimization problem:

$$\min \theta(x), \qquad x \in R^n, \tag{1.5}$$

and condition (1.3) is equivalent to

$$\theta(x_k + \lambda_k d_k) < \theta(x_k). \tag{1.6}$$

An iterative method that generates a sequence $\{x_k\}$ satisfying (1.3) or (1.6) is called a norm descent method. If d_k is a descent direction of θ at x_k, then inequality (1.6) holds for all $\lambda_k > 0$ sufficiently small. Accordingly, the related iterative method is a norm descent method. In particular, Newton's method with line search is norm descent. For a quasi-Newton method, however, d_k may not be a descent direction of θ at x_k even if B_k is symmetric and positive definite. To globalize a quasi-Newton method, Li and Fukushima [6] proposed an approximately norm descent line search technique and established global and superlinear convergence of a Gauss–Newton based BFGS method for solving symmetric nonlinear equations. The method in [6] is not norm descent. In addition, the global convergence theorem is established under the assumption that $F'(x)$ is uniformly nonsingular.

The purpose of this paper is to develop a norm descent Gauss–Newton based BFGS method. We adjust the steplength and the search direction simultaneously so that the generated iterate sequence satisfies (1.6). We update B_k by combining a modified BFGS formula [7] or the cautious BFGS update rule with the Gauss–Newton based BFGS method [6] such that B_{k+1} inherits positive definiteness of B_k no matter whatever line search is used. Under mild conditions, we establish a global convergence theorem which shows that there exists an accumulation point that is a stationary point of problem (1.5) even if $F'(x)$ is singular everywhere. We also get the superlinear convergence of the proposed method.

In the next section, we describe how to generate a quasi-Newton direction that is descent for θ. We also state the steps of the proposed method. In section 3, we establish the global and superlinear convergence of the proposed method. In section 4, we present some numerical results.

2. Descent direction in a quasi-Newton method. In this section, we describe a way to generate a descent quasi-Newton direction for θ and then propose a norm descent BFGS method for solving (1.1). We assume that the function F is continuously differentiable, and its Jacobian $F'(x)$ is symmetric for every $x \in R^n$.

Recall that in Newton's method, the Newton direction is a solution of the Newton equation

$$F'(x_k)d + F(x_k) = 0. \tag{2.1}$$

Equation (2.1) may have no solution if $F'(x_k)$ is singular. In the case where the solution set of (2.1) is empty, instead of solving (2.1), we may solve the least squares problem

$$\min \frac{1}{2}\|F'(x_k)d + F(x_k)\|^2$$

to get a direction d_k, which results in the so-called Gauss–Newton equation

$$F'(x_k)^2 d + F'(x_k)F(x_k) = 0. \tag{2.2}$$

Here we have used the symmetry of $F'(x_k)$. On the other hand, if $F'(x_k)$ is nonsingular, (2.2) is equivalent to (2.1). In [6], a Gauss–Newton based quasi-Newton method was proposed in which the quasi-Newton direction is the solution of the following system of linear equations:

$$B_k d + \bar{q}_k = 0, \tag{2.3}$$

where B_k is an approximation of matrix $F'(x_k)^2$, and $\bar{q}_k$ is an approximation of vector $F'(x_k)F(x_k)$. Specifically, let λ_{k-1} be the steplength used at the previous iteration. Then, vector $\bar{q}_k$ is defined by

$$\bar{q}_k = (F(x_k + \lambda_{k-1}F(x_k)) - F(x_k))/\lambda_{k-1} \approx F'(x_k)F(x_k),$$

and matrix B_k is updated by the BFGS formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \tag{2.4}$$

where $s_k = x_{k+1} - x_k$, $y_k = F(x_k + \delta_k) - F(x_k)$, and $\delta_k = F(x_{k+1}) - F(x_k)$. It is clear that if $\|s_k\|$ is small, then $B_{k+1}s_k = y_k \approx F'(x_{k+1})^2 s_k$. Since the solution d_k of (2.3) may not be a descent direction of θ at x_k when x_k is far away from a solution of (1.1), it is generally not possible to get a steplength $\lambda_k > 0$ satisfying (1.6). Taking this into account, Li and Fukushima [6] proposed a nondescent line search in which the steplength $\lambda_k > 0$ satisfies the following inequality:

$$\theta(x_k + \lambda_k d_k) - \theta(x_k) \le -\sigma_1\|\lambda_k d_k\|^2 - \sigma_2\|\lambda_k F(x_k)\|^2 + \epsilon_k\|F(x_k)\|^2, \tag{2.5}$$

where σ_1 and σ_2 are positive constants, and $\epsilon_k > 0$ satisfies

$$\sum_{k=0}^{\infty} \epsilon_k < \infty.$$

Since ϵ_k is small, $\{x_k\}$ is approximately norm descent.

The purpose of this paper is to develop a norm descent BFGS method. In other words, we want to construct a system of linear equations like (2.3) such that its solution provides a descent direction of θ at x_k.

Observe that

$$\lim_{\lambda_{k-1}\to 0^+} \bar{q}_k = F'(x_k)F(x_k) \overset{\triangle}{=} \tilde{q}_k.$$

Accordingly, the solution of (2.3) with $\tilde{q}_k$ instead of $\bar{q}_k$ is $\tilde{d}_k = -B_k^{-1}F'(x_k)F(x_k)$. If B_k is positive definite and $F'(x_k)$ is symmetric, then $\tilde{d}_k$ is a descent direction of θ at x_k. This observation prompts us to regard λ_{k-1} as a parameter. When this parameter is adjusted to be small enough, the solution of (2.3) is a descent direction of θ at x_k. The following process gives details to realize it.

Let

$$q_k(\lambda) = (F(x_k + \lambda F(x_k)) - F(x_k))/\lambda. \tag{2.6}$$

Consider the system of linear equations with parameter λ:

$$B_k d + q_k(\lambda) = 0. \tag{2.7}$$

Let $d(\lambda)$ be the solution of (2.7). The following lemma shows that when $\lambda > 0$ is sufficiently small, every solution of (2.7) is a descent direction of θ at x_k.

LEMMA 2.1. *Let σ_1 and σ_2 be positive constants and B_k be a symmetric and positive definite matrix. If x_k is not a stationary point of* (1.5), *then there exists a constant $\bar{\lambda} > 0$ depending on k such that when $\lambda \in (0, \bar{\lambda})$, the unique solution $d(\lambda)$ of* (2.7) *satisfies*

$$\nabla\theta(x_k)^T d(\lambda) < 0. \tag{2.8}$$

Moreover, inequality

$$\theta(x_k + \lambda d(\lambda)) - \theta(x_k) \leq -\sigma_1\|\lambda d(\lambda)\|^2 - \sigma_2\|\lambda F(x_k)\|^2 \tag{2.9}$$

holds for all $\lambda > 0$ sufficiently small.

Proof. It is clear that

$$\lim_{\lambda\to 0} q_k(\lambda) = F'(x_k)F(x_k).$$

Therefore, we get from (2.7) that

$$\begin{aligned}\lim_{\lambda\to 0^+} \nabla\theta(x_k)^T d(\lambda) &= -\lim_{\lambda\to 0^+} F(x_k)^T F'(x_k) B_k^{-1} q_k(\lambda)\\ &= -F(x_k)^T F'(x_k) B_k^{-1} F'(x_k) F(x_k).\end{aligned}$$

Since $F'(x_k)$ is symmetric and $F'(x_k)F(x_k) \neq 0$ as x_k is not a stationary point of (1.5), the last equality and the positive definiteness of B_k imply (2.8). We turn to verifying (2.9).

Notice that

$$\begin{aligned}\lim_{\lambda\to 0^+}(\theta(x_k + \lambda d(\lambda)) - \theta(x_k))/\lambda &= \lim_{\lambda\to 0^+} \nabla\theta(x_k)^T d(\lambda)\\ &= -F(x_k)^T F'(x_k) B_k^{-1} F'(x_k) F(x_k) < 0.\end{aligned}$$

However, the right-hand side of (2.9) is $o(\lambda)$. Therefore, inequality (2.9) holds for all $\lambda > 0$ sufficiently small. ∎

Lemma 2.1 motivates us to find a descent quasi-Newton direction by adjusting parameter λ.

Procedure 1. Let constant $\rho \in (0, 1)$ be given. Let i_k be the smallest nonnegative integer such that inequality (2.9) holds with $\lambda = \rho^i$, $i = 0, 1, \ldots$. Let $d_k = d(\rho^{i_k})$, and $q_k = q_k(\rho^{i_k})$.

Procedure 1 ensures that the value of θ at $x_k + \rho^{i_k} d_k$ is less than that of θ at x_k, though d_k may not necessarily be a descent direction of θ at x_k. It is reasonable to let the scalar ρ^{i_k} be the steplength. However, this steplength may be very small if i_k is large. To enlarge steplength, we exploit the following forward procedure.

Procedure 2. Let i_k and d_k be determined by Procedure 1. If $i_k = 0$, let $\lambda_k = 1$. Otherwise, let j_k be the largest positive integer $j \in \{0, 1, 2, \ldots, i_k - 1\}$ satisfying

$$\theta(x_k + \rho^{i_k - j} d_k) - \theta(x_k) \leq -\sigma_1\|\rho^{i_k - j} d_k\|^2 - \sigma_2\|\rho^{i_k - j} F(x_k)\|^2. \tag{2.10}$$

Let $\lambda_k = \rho^{i_k - j_k}$.

Note that (2.10) is satisfied with $j = 0$. Therefore, Procedure 2 is well defined.

Procedures 1 and 2 describe a way to generate d_k and λ_k. It is easy to see from Procedures 1 and 2 that

$$\theta(x_k + \lambda_k d_k) - \theta(x_k) \le -\sigma_1\|\lambda_k d_k\|^2 - \sigma_2\|\lambda_k F(x_k)\|^2, \tag{2.11}$$

which corresponds to (2.5) with $\epsilon_k = 0$. It is also easy to see that if $\lambda_k \neq 1$, then $\lambda'_k = \lambda_k/\rho$ satisfies

$$\theta(x_k + \lambda'_k d_k) - \theta(x_k) > -\sigma_1\|\lambda'_k d_k\|^2 - \sigma_2\|\lambda'_k F(x_k)\|^2. \tag{2.12}$$

Notice that Procedure 1 generates a direction d_k which satisfies

$$B_k d_k + q_k = 0, \tag{2.13}$$

where $q_k = q_k(\rho^{i_k})$. Vector q_k differs from $q_k(\lambda_k)$ if $j_k \neq 0$.

Based on the above process, we propose a norm descent Gauss–Newton based BFGS method as follows.

ALGORITHM 1 (a descent BFGS method).

Initial Let $B_0 \in R^{n\times n}$ be symmetric and positive definite. Let $x_0 \in R^n$. Set $k = 0$.

Step 1 Determine d_k and λ_k by Procedures 1 and 2. Let $x_{k+1} = x_k + \lambda_k d_k$.

Step 2 Update B_k to get B_{k+1} by the modified BFGS formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \tag{2.14}$$

where $s_k = x_{k+1} - x_k$,

$$y_k = \gamma_k + \Big(\max\Big\{0, -\frac{\gamma_k^T s_k}{\|s_k\|^2}\Big\} + \phi(\|F(x_k)\|)\Big)s_k,$$

$\gamma_k = F(x_k + \delta_k) - F(x_k)$, $\delta_k = F(x_{k+1}) - F(x_k)$, and function $\phi : R \to R$ satisfies (i) $\phi(t) > 0$ for all $t > 0$, (ii) $\phi(t) = 0$ if and only if $t = 0$, (iii) $\phi(t)$ is bounded if t is in a bounded set.

Step 3 Let $k := k + 1$ and go to Step 1.

In Step 2 of Algorithm 1, we use a modified BFGS update formula instead of the ordinary BFGS formula. The modified BFGS update formula was proposed by Li and Fukushima [7], where $\phi(t) = \mu t$ with some constant $\mu > 0$. A favorable property for this modification is that B_{k+1} inherits positive definiteness of B_k whatever line search is used [7]. Indeed, it is not difficult to get that

$$y_k^T s_k \ge \max\Big\{\gamma_k^T s_k,\ \phi(\|F(x_k)\|)\|s_k\|^2\Big\} > 0, \tag{2.15}$$

which is sufficient to guarantee positive definiteness of B_{k+1} as long as B_k is positive definite. Suppose that $\{x_k\}$ is contained in a bounded set at which F is continuously differentiable. It is not difficult to deduce that

$$\|y_k\| \le 2\|\gamma_k\| + \phi(\|F(x_k)\|)\ \|s_k\| \le 2L\|\delta_k\| + M\|s_k\| \le (2L^2 + M)\|s_k\|, \tag{2.16}$$

where $M > 0$ is an upper bound of $\phi(\|F(x)\|)$ and $L > 0$ is a Lipschitz constant of F. Inequalities (2.15) and (2.16) imply that

$$\max\Big\{\gamma_k^T s_k,\ \phi(\|F(x_k)\|)\|s_k\|^2\Big\} \le y_k^T s_k \le (2L^2 + M)\|s_k\|^2. \tag{2.17}$$

Another way to develop quasi-Newton methods is to adopt the so-called cautious update rule proposed by Li and Fukushima [8]. The steps of the related BFGS algorithm is stated as follows.

ALGORITHM 2 (a descent cautious BFGS method).

Initial Let $B_0 \in R^{n\times n}$ be symmetric and positive definite. Let $x_0 \in R^n$. Set $k = 0$.

Step 1 Determine d_k and λ_k by Procedures 1 and 2. Let $x_{k+1} = x_k + \lambda_k d_k$.

Step 2 Update B_k to get B_{k+1} by the cautious BFGS formula

$$B_{k+1} = \begin{cases} B_k - \dfrac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \dfrac{\gamma_k \gamma_k^T}{\gamma_k^T s_k} & \text{if } \dfrac{\gamma_k^T s_k}{\|s_k\|^2} \geq \phi(\|F(x_k)\|), \\ B_k & \text{otherwise,} \end{cases} \tag{2.18}$$

where γ_k and ϕ are the same as those in Algorithm 1.

Step 3 Let $k := k + 1$ and go to Step 1.

The only difference between Algorithms 1 and 2 is the update formula. The cautious BFGS method possesses similar properties of the modified BFGS method. For details, we refer to [8].

3. Global and superlinear convergence.

In this section, we prove the global and superlinear convergence of Algorithm 1. The global convergence of Algorithm 2 can be obtained in a similar way. Without specification, we let $\{x_k\}$ and $\{B_k\}$ stand for the sequences of iterates and matrices generated by Algorithm 1, respectively. The following lemma is straightforward from Algorithm 1.

LEMMA 3.1. *The sequence $\{\theta(x_k)\}$ is strictly decreasing. In addition, the following inequalities hold:*

$$\sum_{k=0}^{\infty} \|s_k\|^2 < \infty, \qquad \sum_{k=0}^{\infty} \|\lambda_k F(x_k)\|^2 < \infty. \tag{3.1}$$

We summarize the condition needed for the global convergence of Algorithm 1 as follows.

Assumption A.

(i) The level set

$$\Omega = \{x \in R^n \mid \theta(x) \leq \theta(x_0)\}$$

is bounded.

(ii) Function F is continuously differentiable on Ω, and $F'(x)$ is symmetric for every $x \in \Omega$.

It is clear that under condition (i) in Assumption A, sequence $\{x_k\} \subset \Omega$ is bounded.

We are going to establish a global convergence theorem of Algorithm 1 to show that under Assumption A, there exists an accumulation point of $\{x_k\}$ which is a stationary point of (1.5), namely,

$$\liminf_{k\to\infty} \|\nabla\theta(x_k)\| = 0. \tag{3.2}$$

It is easy to see from Lemma 3.1 that if $\limsup_{k\to\infty} \lambda_k > 0$, then $\liminf_{k\to\infty} \|F(x_k)\| = 0$ and, hence, (3.2) holds. So, we need only to show (3.2) for the case $\lim_{k\to\infty} \lambda_k = 0$. We do it by assuming

$$\liminf_{k\to\infty} \|\nabla\theta(x_k)\| > 0 \tag{3.3}$$

to deduce a contradiction.

Notice that (3.3) particularly implies that there is a constant $\eta > 0$ such that $\|F(x_k)\| \geq \eta$ for all k. It follows from (2.17) and the properties of ϕ that if (3.3) holds, then there are positive constants $c \leq C$ such that

$$c\|s_k\|^2 \leq y_k^T s_k \leq C\|s_k\|^2. \tag{3.4}$$

Therefore, we get the following lemma from (2.16), (3.4), and Theorem 2.1 of [1].

LEMMA 3.2. *If* (3.3) *holds, then there are positive constants* β_i*,* $i = 1, 2, 3$*, such that for any positive integer* k*, inequalities*

$$\|B_i s_i\| \leq \beta_1 \|s_i\|, \qquad \beta_2 \|s_i\|^2 \leq s_i^T B_i s_i \leq \beta_3 \|s_i\|^2 \tag{3.5}$$

hold for at least $\lceil k/2 \rceil$ *many* $i \leq k$.

Inequalities (3.5) together with (2.13) imply that there are at least $\lceil k/2 \rceil$ many $i \leq k$ satisfying

$$\|q_i\| = \|B_i d_i\| \leq \beta_1 \|d_i\|, \qquad \|d_i\| \leq \beta_2^{-1} \|q_i\|. \tag{3.6}$$

We now prove the global convergence of Algorithm 1.

THEOREM 3.3. *Let Assumption* A *hold and* $\{x_k\}$ *be generated by Algorithm* 1*. Then* (3.2) *holds.*

Proof. We need only to show (3.2) for the case $\lim_{k\to\infty} \lambda_k = 0$. In this case, inequality (2.12) holds for all k sufficiently large. Suppose contrarily that (3.2) does not hold or, equivalently, (3.3) holds. Denote by K the set of indices i such that (3.5) holds. Then K is infinite. Since $\{x_k\} \subset \Omega$ is bounded, it is clear that sequences $\{q_k\}_{k\in K}$ and $\{d_k\}_{k\in K}$ are bounded. Let $K_1 \subset K$ and subsequences $\{x_k\}_{k\in K_1}$ and $\{d_k\}_{k\in K_1}$ converge to x^* and d^*, respectively. Then we have

$$\lim_{k\in K_1} q_k = \nabla\theta(x^*). \tag{3.7}$$

Dividing both sides of (2.12) by λ_k' and then taking limits as $k \to \infty$ with $k \in K_1$, we get

$$\nabla\theta(x^*)^T d^* \geq 0. \tag{3.8}$$

On the other hand, taking the inner product with d_k in (2.13), we get

$$0 = d_k^T B_k d_k + q_k^T d_k \geq \beta_2 \|d_k\|^2 + q_k^T d_k.$$

Taking limits in both sides as $k \to \infty$ with $k \in K_1$ yields

$$\nabla\theta(x^*)^T d^* \leq -\beta_2 \|d^*\|^2.$$

This together with (3.8) implies that $d^* = 0$. It then follows from (3.6) that $\lim_{k\in K_1} q_k = 0$, which together with (3.7) yields a contradiction with (3.3). The contradiction proves (3.2). ∎

Remark. In [2] the global convergence of Broyden's class of variable metric methods except for DFP was proved. The proof there depends on the convexity of the objective function. A similar result was obtained by Powell [10] when the BFGS method is applied to convex minimization problems. For nonconvex minimization problems, no theory exists to support the global convergence of the BFGS method.

On the contrary, an example has been constructed [3] recently, which shows that the ordinary BFGS method with the Wolfe line search may fail to converge to a stationary point of a nonconvex unconstrained minimization.

On the other hand, a modified BFGS method was proposed by Li and Fukushima [7]. In the modified BFGS method, the iterative matrix B_k is always positive definite whatever line search is used as long as B_0 is positive definite. Moreover, a liminf result was obtained for nonconvex unconstrained minimization. Besides, another modified BFGS method called the cautious BFGS method was proposed by Li and Fukushima [8]. The cautious BFGS method also possesses global convergence in the sense $\liminf_{k\to\infty} \nabla f(x_k) = 0$ when it is applied to $\min f(x)$. In both papers, the results were obtained without the requirement of nonsingular Hessian. These two papers show the possibility to improve the unconstrained minimization result by Byrd, Nocedal, and Yuan [2] and Powell [10].

This paper adopts a similar updating technique as used in [4] and [5]. Consequently, we established Theorem 3.3, which shows that the iterative sequence has an accumulation point which is a stationary point of problem $\min \theta(x) = \frac{1}{2}\|F(x)\|^2$. It may not be a solution of the nonlinear equation (1.1) if the Jacobian is singular at that point.

The next theorem shows a strong convergence property of Algorithm 1.

THEOREM 3.4. *Let Assumption* A *hold. Suppose that the sequence* $\{x_k\}$ *generated by Algorithm* 1 *has a subsequence converging to a stationary* x^* *at which* $F'(x^*)$ *is nonsingular. Then* x^* *is a solution of* (1.1). *Moreover, the whole sequence* $\{x_k\}$ *converges to* x^*.

Proof. Since x^* satisfies $\nabla\theta(x^*) = F'(x^*)F(x^*) = 0$, we obviously have $F(x^*) = 0$ if $F'(x^*)$ is nonsingular. Since $\{\theta(x_k)\}$ converges, every accumulation point of $\{x_k\}$ is a solution of (1.1). By the nonsingularity of $F'(x^*)$ again, x^* is an isolated limit point of $\{x_k\}$. However, we have from (3.1) that $x_{k+1} - x_k \to 0$ as $k \to \infty$. Therefore, the whole sequence $\{x_k\}$ converges to x^*. ∎

In a way similar to the proof of Theorem 3.8 in [7], it is not difficult to prove the superlinear convergence of Algorithm 1. We state the theorem as follows but omit the proof.

THEOREM 3.5. *Let the conditions of Theorem* 3.4 *hold. Suppose further that* F' *is Lipschitz continuous. Then* $\{x_k\}$ *is superlinearly convergent.*

Similar to the above argument, we can establish the global and superlinear convergence of Algorithm 2. We state the results as follows but omit the proof.

THEOREM 3.6. *Let Assumption* A *hold and* $\{x_k\}$ *be generated by Algorithm* 2. *Then* (3.2) *holds. If the sequence* $\{x_k\}$ *has a subsequence converging to a stationary* x^* *at which* $F'(x^*)$ *is nonsingular, then* x^* *is a solution of* (1.1). *Moreover, the whole sequence* $\{x_k\}$ *converges to* x^*. *If we further suppose that* F' *is Lipschitz continuous, then* $\{x_k\}$ *is superlinearly convergent.*

4. Numerical results. In this section, we test the proposed descent BFGS methods on nonlinear equation problems obtained from [6, 9] and the unconstrained optimization problems obtained from the website ftp://ftp.mathworks.com/pub/contrib/v4/optim/uncprobs/. We call Algorithms 1 and 2 the DBFGS (descent BFGS) method and the CBFGS (cautious BFGS) method, respectively, and call the BFGS method based on the Gauss–Newton approach and the nondescent line search [6] the NBFGS (nondescent BFGS) method. Then we compare their performance.

The parameters are specified as follows. We take $\rho = 0.1$ and $\sigma_1 = \sigma_2 = 10^{-5}$

in (2.9). The initial quasi-Newton matrices are set to be $B_0 = A$ [6] for nonlinear equation problems and $B_0 = I$ for unconstrained optimization problems. The function ϕ is determined by

$$\phi(t) = \begin{cases} Ct^2 & \text{if } t \le 1, \\ Ct^{0.1} & \text{otherwise,} \end{cases}$$

where $C = 10^{-5}$. For the NBFGS method, we update B_k by the BFGS formula [6] if $y_k^T s_k \ge 10^{-5}$. Otherwise, we let $B_{k+1} = B_k$. We stop the iteration process if $\|F(x_k)\| \le 10^{-4}$.

The tested results are listed in Tables 1 and 2. Table 3 gives the average performance of the three methods for solving nonlinear equation problems. The columns of the tables have the following meaning:

Dim: the dimension of the problem.

Method: the name of the algorithm.

Init: the initial point, namely, integer l in Table 1 meaning $x_0 = (l, l, \ldots, l)^T$.

Iter: the total number of iterations.

Inner: for the NBFGS method, the number of iterations at which $y_k^T s_k \ge 10^{-5}$ is satisfied; for the DBFGS method and the CBFGS method, the maximum number of inner iterations to generate the descent direction d_k.

Numf: the number of the function evaluations.

Fnorm: the final value of $\|F(x_k)\|$.

All the three methods terminate at solutions of nonlinear equation problems for all tested starting points. However, for the 33 unconstrained optimization problems, all the three methods fail to converge to a solution for at least 10 problems. The numbers of problems for which the NBFGS method, the DBFGS method, and the CBFGS method fail to converge are 16, 19, and 12, respectively.

The numerical results show that for low dimensional problems, the performance of these three methods is not different very much. For most of the test problems, the DBFGS method and the CBFGS method perform better than the NBFGS method in the iteration number, but worse in the number of the function evaluation. However, for high dimensional problems ($n = 200$ in Tables 1 and 3), both the DBFGS and the CBFGS methods perform much better than the NBFGS method in the iteration number as well as the number of the function evaluation. The maximum numbers of the inner iteration to generate a descent direction of a DBFGS method are generally very small. We also note that the performance of the DBFGS and CBFGS methods is almost the same if the both methods terminate regularly. For unconstrained optimization problems, the DBFGS method fails more frequently than the CBFGS method does.

In summary, the presented numerical results reveal that the DBFGS and CBFGS methods, compared with the NBFGS method, have potential advantages when applied to solve symmetric nonlinear equation whose function is not difficult to compute.

In Tables 1–3, we simply denote the NBFGS method as the BFGS method.

Table 1
Test results for nonlinear equation problems $B_0 = A$.

Dim	Method	Init	Iter	Inner	Numf	Fnorm	Dim	Method	Init	Iter	Inner	Numf	Fnorm
10	BFGS	0	6	0	19	2.9e-06	50	BFGS	0	25	0	76	3.2e-05
	DBFGS		6	1	19	2.9e-06		DBFGS		25	1	76	3.9e-05
	CBFGS		6	1	19	2.9e-06		CBFGS		25	1	76	3.2e-05
	BFGS	1	11	0	35	9.1e-05		BFGS	1	37	0	114	3.3e-05
	DBFGS		10	2	42	3.5e-05		DBFGS		36	2	120	1.6e-05
	CBFGS		10	2	42	3.5e-05		CBFGS		36	2	120	1.6e-05
	BFGS	-1	12	0	38	2.0e-05		BFGS	-1	37	0	114	1.8e-05
	DBFGS		10	2	36	1.1e-05		DBFGS		36	2	120	1.0e-05
	CBFGS		10	2	36	1.1e-05		CBFGS		36	2	120	1.0e-05
	BFGS	10	13	0	41	4.1e-05		BFGS	10	38	0	117	6.0e-05
	DBFGS		13	2	52	2.6e-05		DBFGS		38	2	133	4.9e-05
	CBFGS		13	2	52	2.6e-05		CBFGS		38	2	133	4.8e-05
	BFGS	-10	13	0	41	3.9e-04		BFGS	-10	38	0	117	6.1e-05
	DBFGS		13	2	52	2.5e-05		DBFGS		38	2	133	5.3e-05
	CBFGS		13	2	52	2.5e-05		CBFGS		38	2	133	4.9e-05
	BFGS	10^2	14	0	44	1.4e-05		BFGS	10^2	41	0	127	4.5e-05
	DBFGS		12	2	48	1.3e-05		BFGS		40	2	132	1.5e-05
	CBFGS		12	2	48	1.3e-05		CBFGS		40	2	132	1.5e-05
	BFGS	-10^2	14	0	44	2.0e-05		BFGS	-10^2	41	0	127	5.0e-05
	DBFGS		12	2	48	1.1e-05		DBFGS		38	2	126	8.6e-05
	CBFGS		12	2	48	1.1e-05		CBFGS		38	2	126	8.6e-05
	BFGS	10^3	16	0	50	1.8e-06		BFGS	10^3	44	0	136	4.1e-05
	DBFGS		13	2	51	3.9e-06		DBFGS		40	2	137	9.9e-05
	CBFGS		13	2	51	3.9e-06		CBFGS		40	2	137	9.9e-05
	BFGS	-10^3	16	0	50	7.5e-07		BFGS	-10^3	44	0	136	3.4e-05
	DBFGS		13	2	51	5.0e-06		DBFGS		40	2	137	7.9e-05
	CBFGS		13	2	51	5.0e-06		CBFGS		40	2	137	7.9e-05
	BFGS	10^4	16	0	50	1.1e-05		BFGS	10^4	49	0	152	5.4e-05
	DBFGS		14	2	54	5.7e-06		DBFGS		44	2	160	4.9e-05
	CBFGS		14	2	54	5.7e-06		CBFGS		44	2	160	4.9e-05
	BFGS	-10^4	16	0	50	8.8e-06		BFGS	-10^4	49	0	144	9.6e-05
	DBFGS		14	2	54	5.1e-06		DBFGS		44	2	152	6.9e-05
	CBFGS		14	2	54	5.1e-06		CBFGS		44	2	152	6.9e-05
100	BFGS	0	50	0	151	5.7e-06	200	BFGS	0	501	0	2181	8.9e-05
	DBFGS		50	1	151	5.7e-06		DBFGS		155	3	1142	9.5e-05
	CBFGS		50	1	151	5.7e-06		CBFGS		151	3	1104	9.5e-05
	BFGS	1	63	0	192	1.0e-04		BFGS	1	2191	1	8661	1.0e-04
	DBFGS		62	2	198	6.8e-05		DBFGS		116	3	383	2.2e-05
	CBFGS		62	2	198	6.8e-05		CBFGS		116	3	383	2.2e-05
	BFGS	-1	63	0	192	5.8e-05		BFGS	-1	1971	1	7835	1.0e-04
	DBFGS		62	2	198	5.4e-05		DBFGS		116	3	382	1.8e-05
	CBFGS		62	2	198	5.4e-05		CBFGS		116	3	382	1.8e-05
	BFGS	10	66	0	203	2.0e-05		BFGS	10	4547	1	18085	1.0e-04
	DBFGS		66	2	220	3.7e-05		DBFGS		119	3	405	4.1e-05
	CBFGS		66	2	217	9.0e-05		CBFGS		119	3	405	4.0e-05
	BFGS	-10	66	0	203	1.2e-05		BFGS	-10	4070	1	16177	1.0e-04
	DBFGS		66	2	221	1.8e-05		DBFGS		119	3	406	2.3e-06
	CBFGS		76	2	254	6.9e-05		CBFGS		119	3	407	2.3e-06
	BFGS	10^2	69	0	213	6.7e-05		BFGS	10^2	6095	1	24277	1.0e-04
	DBFGS		67	2	220	5.7e-05		DBFGS		125	4	447	6.9e-05
	CBFGS		66	2	217	9.0e-05		CBFGS		120	4	424	9.7e-05
	BFGS	-10^2	69	0	213	6.2e-05		BFGS	-10^2	6049	1	24093	1.0e-04
	DBFGS		65	2	211	9.8e-05		DBFGS		125	3	433	6.9e-05
	CBFGS		65	2	211	8.5e-05		CBFGS		124	3	439	5.6e-05
	BFGS	10^3	72	0	221	9.2e-05		BFGS	10^3	8330	1	33217	1.0e-04
	DBFGS		70	2	236	5.8e-05		DBFGS		150	3	543	7.3e-05
	CBFGS		67	2	222	2.3e-05		CBFGS		540	3	3704	8.5e-05
	BFGS	-10^3	71	0	220	4.3e-05		BFGS	-10^3	8296	1	33081	1.0e-04
	DBFGS		69	2	232	8.5e-05		DBFGS		145	3	511	9.8e-05
	CBFGS		67	2	222	2.7e-05		CBFGS		533	3	3662	8.3e-05
	BFGS	10^4	73	0	227	6.7e-05		BFGS	10^4	9962	1	39745	1.0e-04
	DBFGS		72	2	246	4.9e-05		DBFGS		185	3	704	8.9e-05
	CBFGS		69	2	233	8.3e-05		CBFGS		1299	2	10043	5.0e-05
	BFGS	-10^4	73	0	227	6.3e-05		BFGS	-10^4	9915	1	39557	1.0e-04
	DBFGS		72	2	242	4.3e-05		DBFGS		590	3	3123	8.5e-05
	CBFGS		69	2	232	5.6e-05		CBFGS		1268	2	9837	4.8e-05

TABLE 2
Test results for unconstrained optimization problems $B_0 = I$.

Method	Prob	Dim	Iter	Inner	Numf	Fnorm	Method	Prob	Dim	Iter	Inner	Numf	Fnorm
BFGS	rose	2	103	0	415	6.3e-005	BFGS	froth	2	-	-	-	-
DBFGS			-	-	-	-	DBFGS			-	-	-	-
CBFGS			668	7	6301	9.1e-05	CBFGS			282	7	3155	9.1e-06
BFGS	beale	2	347	0	1331	9.4e-05	BFGS	jensam	-	-	-	-	-
DBFGS			-	-	-	-	DBFGS			-	-	-	-
CBFGS			155	4	1330	2.6e-05	CBFGS			12	5	65	8.3e-05
BFGS	helix	3	279	0	1205	8.9e-05	BFGS	gulf	3	1	1	4	5.6e-086
DBFGS			-	-	-	-	DBFGS			1	1	4	1.9e-10
CBFGS			156	6	1413	3.1e-05	CBFGS			1	1	4	1.0e-10
BFGS	gauss	3	2	0	8	5.9e-006	BFGS	meyer	3	-	-	-	-
DBFGS			2	2	10	6.0e-06	DBFGS			1	4	14	4.2e-07
CBFGS			2	2	10	6.0e-06	CBFGS			1	4	14	4.2e-07
BFGS	sing	4	218	1	875	8.6e-05	BFGS	wood	4	-	-	-	-
DBFGS			214	9	1847	9.9e-05	DBFGS			-	-	-	-
CBFGS			97	6	650	9.7e-05	CBFGS			617	8	8971	6.9e-05
BFGS	kowosb	5	-	-	-	-	BFGS	biggs	6	59	0	211	4.4e-05
DBFGS			661	4	7031	1.0e-04	DBFGS			101	5	589	6.9e-05
CBFGS			661	4	7028	1.0e-04	CBFGS			101	5	589	6.9e-05
BFGS	osb2	11	225	1	775	4.0e-05	BFGS	watson	2	24	0	90	2.8e-06
DBFGS			-	-	-	-	DBFGS			18	5	124	1.1e-05
CBFGS			-	-	-	-	CBFGS			18	5	124	1.1e-05
BFGS	trid	10	152	0	609	1.8e-05	BFGS	singx	40	-	-	-	-
DBFGS			115	5	682	6.2e-05	DBFGS			-	-	-	-
CBFGS			115	5	682	6.2e-05	CBFGS			741	6	6372	1.0e-04
BFGS	pen1	10	248	0	1048	2.9e-05	BFGS	pen2	10	320	0	1499	5.0e-05
DBFGS			148	7	1235	4.5e-05	DBFGS			-	-	-	-
CBFGS			148	7	1235	4.5e-05	CBFGS			-	-	-	-
BFGS	bv	10	30	0	104	1.0e-05	BFGS	ie	10	5	0	17	2.6e-05
DBFGS			31	3	135	1.8e-05	DBFGS			4	2	18	2.6e-05
CBFGS			31	3	135	1.8e-05	CBFGS			4	2	18	2.6e-05
BFGS	lin	10	1	0	4	1.0e-13	BFGS	lin1	10	2	0	17	7.7e-06
DBFGS			1	1	4	8.9e-16	DBFGS			2	11	28	1.1e-10
CBFGS			1	1	4	8.9e-16	CBFGS			2	11	28	1.1e-10
BFGS	lin0	10	2	0	17	7.7e-07							
DBFGS			2	11	30	1.3e-11							
CBFGS			2	11	30	1.3e-11							

TABLE 3
Average performance for nonlinear equation problems.

Dim	Method	Iter	Inner	Numf	Dim	Method	Iter	Inner	Numf
10	BFGS	13.4	0	42	50	BFGS	46	0	123.6
	DBFGS	11.8	1.9	46.1		DBFGS	38	1.9	129.6
	CBFGS	11.8	1.9	46.1		CBFGS	38	1.9	129.6
100	BFGS	66.8	0	205.6	200	BFGS	5629.7	0.9	22446
	DBFGS	65.5	1.9	215.9		DBFGS	176.8	3.1	770.8
	CBFGS	65.2	1.9	214.1		CBFGS	409.5	2.6	2799.1

REFERENCES

[1] R. H. BYRD AND J. NOCEDAL, *A tool for the analysis of quasi-Newton methods with application to unconstrained minimization*, SIAM J. Numer. Anal., 26 (1989), pp. 727–739.

[2] R. H. BYRD, J. NOCEDAL, AND Y.-X. YUAN, *Global convergence of a class of quasi-Newton methods on convex problems*, SIAM J. Numer. Anal., 24 (1987), pp. 1171–1190.

[3] Y. DAI, *Convergence Properties of the BFGS Algorithm*, Technical report, The State Key Laboratory of Scientific and Engineering Computing, Chinese Academy of Sciences, Beijing, China, 2001.

[4] J. E. DENNIS, JR., AND J. J. MORÉ, *A characterization of superlinear convergence and its application to quasi-Newton methods*, Math. Comp., 28 (1974), pp. 1171–1190.

[5] J. E. DENNIS, JR., AND J. J. MORÉ, *Quasi-Newton methods, motivation and theory*, SIAM Rev., 19 (1977), pp. 46–89.

[6] D. H. LI AND M. FUKUSHIMA, *A globally and superlinearly convergent Gauss–Newton-based BFGS method for symmetric nonlinear equations*, SIAM J. Numer. Anal., 37 (1999), pp. 152–172.

[7] D. H. Li and M. Fukushima, *A modified BFGS method and its global convergence in nonconvex minimization*, J. Comput. Appl. Math., 129 (2001), pp. 15–35.

[8] D. H. Li and M. Fukushima, *On the global convergence of the BFGS method for nonconvex unconstrained optimization problems*, SIAM J. Optim., 11 (2001), pp. 1054–1064.

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

[10] M. J. D. Powell, *Some global convergence properties of a variable metric algorithm for minimization without exact line searches*, in Nonlinear Programming, SIAM-AMS Proceedings, Vol. IX, R. W. Cottle and C. E. Lemke, eds., AMS, Providence, RI, 1976, pp. 55–92.