SIAM J. OPTIM.
Vol. 11, No. 4, pp. 1054–1064

ON THE GLOBAL CONVERGENCE OF THE BFGS METHOD FOR NONCONVEX UNCONSTRAINED OPTIMIZATION PROBLEMS*

DONG-HUI LI† AND MASAO FUKUSHIMA‡

Abstract. This paper is concerned with the open problem of whether the BFGS method with inexact line search converges globally when applied to nonconvex unconstrained optimization problems. We propose a cautious BFGS update and prove that the method with either a Wolfe-type or an Armijo-type line search converges globally if the function to be minimized has Lipschitz continuous gradients.

Key words. unconstrained optimization, BFGS method, global convergence

AMS subject classifications. 90C30, 65K05

PII. S1052623499354242

1. Introduction.

The BFGS method is a well-known quasi-Newton method for solving unconstrained optimization problems [5, 7]. Because of favorable numerical experience and fast theoretical convergence, it has become a method of choice for engineers and mathematicians who are interested in solving optimization problems.

Local convergence of the BFGS method has been well established [3, 4]. The study on global convergence of the BFGS method has also made good progress. In particular, for convex minimization problems, it has been shown that the iterates generated by the BFGS method are globally convergent if the exact line search or some special inexact line search is used [1, 2, 6, 9, 16, 17, 18]. On the other hand, little is known concerning global convergence of the BFGS method for nonconvex minimization problems. Indeed, so far, no one has proved global convergence of the BFGS method for nonconvex minimization problems or has given a counter example that shows nonconvergence of the BFGS method. Whether the BFGS method converges globally for a nonconvex function remains unanswered. This open problem has been mentioned many times and is currently regarded as one of the most fundamental open problems in the theory of quasi-Newton methods [8, 15].

Recently, the authors [12] proposed a modified BFGS method and established its global convergence for nonconvex unconstrained optimization problems. The authors [11] also proposed a globally convergent Gauss–Newton-based BFGS method for symmetric nonlinear equations that contain unconstrained optimization problems as a special case. The results obtained in [11] and [12] positively support the open problem. However, the original question still remains unanswered.

The purpose of this paper is to study this problem further. We introduce a *cautious update* in the BFGS method and prove that the method with a Wolfe-type or an Armijo-type line search converges globally if the function to be minimized has Lipschitz continuous gradients. Moreover, under appropriate conditions, we show that the cautious update eventually reduces to the ordinary update.

*Received by the editors April 11, 1999; accepted for publication (in revised form) November 6, 2000; published electronically May 10, 2001.
http://www.siam.org/journals/siopt/11-4/35424.html

†Department of Applied Mathematics, Hunan University, Changsha, China 410082 (dhli@mail.hunu.edu.cn). This work was done while this author was visiting Kyoto University.

‡Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan (fuku@amp.i.kyoto-u.ac.jp).

In the next section, we present the BFGS method with a cautious update. In section 3, we prove global convergence and, under additional assumptions, superlinear convergence of the algorithm. In section 4, we report some numerical results with the algorithm.

We introduce some notation: For a real-valued function $f : R^n \to R$, $g(x)$ and $G(x)$ denote the gradient and Hessian matrix of f at x, respectively. For simplicity, $g(x_k)$ and $G(x_k)$ are often denoted by g_k and G_k, respectively. For a vector $x \in R^n$, $\|x\|$ denotes its Euclidean norm.

2. Algorithm. Let $f : R^n \to R$ be continuously differentiable. Consider the following unconstrained optimization problem:

$$\min f(x), \quad x \in R^n. \tag{2.1}$$

The ordinary BFGS method for (2.1) generates a sequence $\{x_k\}$ by the iterative scheme:

$$x_{k+1} = x_k + \lambda_k p_k, \qquad k = 0, 1, 2, \ldots,$$

where p_k is the BFGS direction obtained by solving the linear equation

$$B_k p + g_k = 0. \tag{2.2}$$

The matrix B_k is updated by the BFGS formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \tag{2.3}$$

where $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$. A good property of formula (2.3) is that B_{k+1} inherits the positive definiteness of B_k as long as $y_k^T s_k > 0$. The condition $y_k^T s_k > 0$ is guaranteed to hold if the stepsize λ_k is determined by the exact line search

$$f(x_k + \lambda_k p_k) = \min_{\lambda > 0} f(x_k + \lambda p_k) \tag{2.4}$$

or the Wolfe-type inexact line search

$$\begin{cases} f(x_k + \lambda_k p_k) \le f(x_k) + \sigma_1 \lambda_k g(x_k)^T p_k, \\ g(x_k + \lambda_k p_k)^T p_k \ge \sigma_2 g(x_k)^T p_k, \end{cases} \tag{2.5}$$

where σ_1 and σ_2 are positive constants satisfying $\sigma_1 < \sigma_2 < 1$. In addition, if $\lambda_k = 1$ satisfies (2.5), we take $\lambda_k = 1$. Global convergence of the BFGS method with the line search (2.4) or (2.5) for convex minimization problems has been studied in [1, 2, 6, 9, 16, 17, 18].

Another important inexact line search is the Armijo-type line search that finds a λ_k that is the largest value in the set $\{\rho^i | i = 0, 1, \ldots\}$ such that the inequality

$$f(x_k + \lambda_k p_k) \le f(x_k) + \sigma \lambda_k g(x_k)^T p_k \tag{2.6}$$

is satisfied, where σ and ρ are constants such that $\sigma, \rho \in (0, 1)$. The Armijo-type line search does not ensure the condition $y_k^T s_k > 0$ and hence B_{k+1} is not necessarily positive definite even if B_k is positive definite. In order to ensure the positive definiteness of B_{k+1}, the condition $y_k^T s_k > 0$ is sometimes used to decide whether or not B_k is updated . More specifically, B_{k+1} is determined by

$$B_{k+1} = \begin{cases} B_k - \dfrac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \dfrac{y_k y_k^T}{y_k^T s_k} & \text{if } y_k^T s_k > 0, \\ B_k & \text{otherwise.} \end{cases} \tag{2.7}$$

Computationally, the condition $y_k^T s_k > 0$ is often replaced by the condition $y_k^T s_k > \eta$, where $\eta > 0$ is a small constant. In this paper, we propose a cautious update rule similar to the above and establish a global convergence theorem for nonconvex problems. For the sake of motivation, we state a lemma due to Powell [17].

LEMMA 2.1 (Powell [17]). *If the BFGS method with the line search* (2.5) *is applied to a continuously differentiable function f that is bounded below, and if there exists a constant $M > 0$ such that the inequality*

$$\frac{\|y_k\|^2}{y_k^T s_k} \leq M \tag{2.8}$$

holds for all k, then

$$\liminf_{k\to\infty} \|g(x_k)\| = 0. \tag{2.9}$$

Notice that if f is twice continuously differentiable and convex, then (2.8) always holds whenever $\{x_k\}$ is bounded. Therefore, global convergence of the BFGS method follows immediately from Lemma 2.1. However, in the case where f is nonconvex, it seems difficult to guarantee (2.8). This is probably the main reason why global convergence of the BFGS method has yet to be proved. In [12], the authors proposed a modified BFGS method by using $\tilde{y}_k = C\|g_k\|s_k + (g_{k+1} - g_k)$ with a constant $C > 0$ instead of y_k in the update formula (2.3). Global convergence of the modified BFGS method in [12] is proved without the convexity assumption on f by means of Lemma 2.1 with a contradictory assumption that $\{\|g_k\|\}$ are bounded away from zero. We further study global convergence of the BFGS method for (2.1). Instead of modifying the method, we introduce a cautious update rule in the ordinary BFGS method. To be precise, we determine B_{k+1} by

$$B_{k+1} = \begin{cases} B_k - \dfrac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \dfrac{y_k y_k^T}{y_k^T s_k} & \text{if } \dfrac{y_k^T s_k}{\|s_k\|^2} \geq \epsilon \|g_k\|^\alpha, \\ B_k & \text{otherwise,} \end{cases} \tag{2.10}$$

where ϵ and α are positive constants.

Now, we state the BFGS method with the cautious update.

ALGORITHM 1.

Step 0 Choose an initial point $x_0 \in R^n$ and an initial symmetric and positive definite matrix $B_0 \in R^{n\times n}$. Choose constants $0 < \sigma_1 < \sigma_2 < 1$, $\alpha > 0$, and $\epsilon > 0$. Let $k := 0$.
Step 1 Solve the linear equation (2.2) to get p_k.
Step 2 Determine a stepsize $\lambda_k > 0$ by (2.5) or (2.6).
Step 3 Let the next iterate be $x_{k+1} := x_k + \lambda_k p_k$.
Step 4 Determine B_{k+1} by (2.10).
Step 5 Let $k := k + 1$ and go to Step 1.

Remark. It is not difficult to see from (2.10) that the matrix B_k generated by Algorithm 1 is symmetric and positive definite for all k, which in turn implies that

$\{f(x_k)\}$ is a decreasing sequence whichever line search (2.5) or (2.6) is used. Moreover, we have from (2.5) or (2.6)

$$-\sum_{k=0}^{\infty} g_k^T s_k < \infty \tag{2.11}$$

if f is bounded below. In particular, we have

$$-\lim_{k\to\infty} \lambda_k g_k^T p_k = -\lim_{k\to\infty} g_k^T s_k = 0. \tag{2.12}$$

3. Global convergence. In this section, we prove global convergence of Algorithm 1 under the following assumption, which we assume throughout this section.

Assumption A. The level set

$$\Omega = \{x \in R^n \mid f(x) \le f(x_0)\}$$

is bounded, the function f is continuously differentiable on Ω, and there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \le L\|x - y\| \qquad \forall x, y \in \Omega. \tag{3.1}$$

Since $\{f(x_k)\}$ is a decreasing sequence, it is clear that the sequence $\{x_k\}$ generated by Algorithm 1 is contained in Ω.

For the sake of convenience, we define the index set

$$\bar{K} = \left\{ i \;\middle|\; \frac{y_i^T s_i}{\|s_i\|^2} \ge \epsilon \|g_i\|^\alpha \right\}. \tag{3.2}$$

By means of $\bar{K}$, we may rewrite (2.10) as

$$B_{k+1} = \begin{cases} B_k - \dfrac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \dfrac{y_k y_k^T}{y_k^T s_k} & \text{if } k \in \bar{K}, \\ B_k & \text{otherwise.} \end{cases} \tag{3.3}$$

Now we proceed to establishing global convergence of Algorithm 1. First, we show the following convergence theorem.

THEOREM 3.1. *Let $\{x_k\}$ be generated by Algorithm 1. If there are positive constants $\beta_1, \beta_2, \beta_3 > 0$ such that the relations*

$$\|B_k s_k\| \le \beta_1 \|s_k\|, \qquad \beta_2 \|s_k\|^2 \le s_k^T B_k s_k \le \beta_3 \|s_k\|^2 \tag{3.4}$$

hold for infinitely many k, then we have

$$\liminf_{k\to\infty} \|g(x_k)\| = 0. \tag{3.5}$$

Proof. Since $s_k = \lambda_k p_k$, it is clear that (3.4) holds true if s_k is replaced by p_k. Let $\mathcal{K}$ be the set of indices k such that (3.4) hold. It is not difficult to deduce from (2.2) and (3.4) that for each $k \in \mathcal{K}$

$$\beta_2 \|p_k\| \le \|g(x_k)\| \le \beta_1 \|p_k\|. \tag{3.6}$$

Consider the case where the Armijo-type line search (2.6) is used with the backtracking parameter ρ. If $\lambda_k \ne 1$, then we have

$$f(x_k + \rho^{-1}\lambda_k p_k) - f(x_k) > \sigma \rho^{-1} \lambda_k g(x_k)^T p_k. \tag{3.7}$$

By the mean-value theorem, there is a $\theta_k \in (0,1)$ such that

$$\begin{aligned} &f(x_k + \rho^{-1}\lambda_k p_k) - f(x_k) \\ &= \rho^{-1}\lambda_k g(x_k + \theta_k \rho^{-1}\lambda_k p_k)^T p_k \\ &= \rho^{-1}\lambda_k g(x_k)^T p_k + \rho^{-1}\lambda_k (g(x_k + \theta_k \rho^{-1}\lambda_k p_k) - g(x_k))^T p_k \\ &\le \rho^{-1}\lambda_k g(x_k)^T p_k + L\rho^{-2}\lambda_k^2 \|p_k\|^2, \end{aligned} \tag{3.8}$$

where $L > 0$ is the Lipschitz constant of g. Substituting (3.8) into (3.7), we get for any $k \in \mathcal{K}$

$$\lambda_k \ge \frac{-(1-\sigma)\rho g(x_k)^T p_k}{L\|p_k\|^2} = \frac{(1-\sigma)\rho p_k^T B_k p_k}{L\|p_k\|^2} \ge (1-\sigma)\beta_2 L^{-1}\rho.$$

This means that for each $k \in \mathcal{K}$, we have

$$\lambda_k \ge \min\{1, (1-\sigma)\beta_2 L^{-1}\rho\} > 0. \tag{3.9}$$

Consider the case where the Wolfe-type line search (2.5) is used. It follows from the second inequality of (2.5) and the Lipschitz continuity of g that

$$L\lambda_k \|p_k\|^2 \ge (g(x_k + \lambda_k p_k) - g(x_k))^T p_k \ge -(1-\sigma_2) g(x_k)^T p_k.$$

This implies

$$\lambda_k \ge \frac{-(1-\sigma_2) g(x_k)^T p_k}{L\|p_k\|^2} = \frac{(1-\sigma_2) p_k^T B_k p_k}{L\|p_k\|^2} \ge (1-\sigma_2)\beta_2 L^{-1}. \tag{3.10}$$

The inequalities (3.10) together with (3.9) show that $\{\lambda_k\}_{k\in\mathcal{K}}$ is bounded away from zero whenever the line search (2.5) or (2.6) is used. It then follows from (2.2) and (2.12) that $p_k^T B_k p_k = -g(x_k)^T p_k \to 0$ as $k \to \infty$ with $k \in \mathcal{K}$. This together with (3.4) and (3.6) implies (3.5). □

Theorem 3.1 indicates that to prove global convergence of Algorithm 1, it suffices to show that there are positive constants $\beta_1, \beta_2, \beta_3$ such that (3.4) holds for infinitely many k. To this end, we quote the following useful result [1, Theorem 2.1].

LEMMA 3.2. *Let B_k be updated by the BFGS formula* (2.3). *Suppose B_0 is symmetric and positive definite and there are positive constants $m \le M$ such that for all $k \ge 0$, y_k and s_k satisfy*

$$\frac{y_k^T s_k}{\|s_k\|^2} \ge m, \quad \frac{\|y_k\|^2}{y_k^T s_k} \le M. \tag{3.11}$$

Then there exist constants $\beta_1, \beta_2, \beta_3 > 0$ such that, for any positive integer t, (3.4) *holds for at least $\lceil t/2 \rceil$ values of $k \in \{1, \dots, t\}$.*

By using Lemma 3.2 and Theorem 3.1, we can establish the following global convergence theorem for Algorithm 1.

THEOREM 3.3. *Let Assumption A hold and $\{x_k\}$ be generated by Algorithm* 1. *Then* (3.5) *holds.*

Proof. By Theorem 3.1, it suffices to show that there are infinitely many indices k satisfying (3.4).

If $\bar{K}$ is finite, then B_k remains constant after a finite number of iterations. Since B_k is symmetric and positive definite for each k, it is obvious that there are constants $\beta_1, \beta_2, \beta_2 > 0$ such that (3.4) holds for all k sufficiently large.

Consider the case where $\bar{K}$ is infinite. For the sake of contradiction, we suppose that (3.5) is not true. That is, there is a constant $\delta > 0$ such that $\|g_k\| \geq \delta$ for all k. It then follows from (3.2) that $y_k^T s_k \geq \epsilon \delta^\alpha \|s_k\|^2$ holds for all $k \in \bar{K}$. This together with (3.1) implies that for any $k \in \bar{K}$, we have

$$\frac{\|y_k\|^2}{y_k^T s_k} \leq \frac{L^2}{\epsilon \delta^\alpha}.$$

Applying Lemma 3.2 to the matrix subsequence $\{B_k\}_{k \in \bar{K}}$, it is clear that there are constants $\beta_1, \beta_2, \beta_3 > 0$ such that (3.4) holds for infinitely many k. The proof is then complete. $\square$

Theorem 3.3 shows that there exists a subsequence of $\{x_k\}$ converging to a stationary point x^* of (2.1). If f is convex, then x^* is a global minimum of f. Since the sequence $\{f(x_k)\}$ converges, it is clear that every accumulation point of $\{x_k\}$ is a global optimal solution of (2.1). That is, we have the following corollary.

COROLLARY 3.4. *Let Assumption* A *hold and* $\{x_k\}$ *be generated by Algorithm* 1. *If* f *is convex, then the whole sequence* $\{g_k\}$ *converges to zero. Consequently, every accumulation point of* $\{x_k\}$ *is a global optimal solution of* (2.1).

In the case where f is nonconvex, Corollary 3.4 is not guaranteed. The following theorem shows that if some additional conditions are assumed, then the whole sequence $\{x_k\}$ converges to a local optimal solution of (2.1).

THEOREM 3.5. *Let* f *be twice continuously differentiable. Suppose that* $s_k \to 0$. *If there exists an accumulation point* x^* *of* $\{x_k\}$ *at which* $g(x^*) = 0$ *and* $G(x^*)$ *is positive definite, then the whole sequence* $\{x_k\}$ *converges to* x^*. *If in addition,* G *is Hölder continuous and the parameters in the line searches satisfy* $\sigma, \sigma_1 \in (0, 1/2)$, *then the convergence rate is superlinear.*

Proof. The assumptions particularly imply that x^* is a strict local optimal solution of (2.1). Since $\{f(x_k)\}$ converges, it follows that x^* is an isolated accumulation point of $\{x_k\}$. Then, by the assumption that $\{s_k\}$ converges to zero, the whole sequence $\{x_k\}$ converges to x^*. Hence $\{g_k\}$ tends to zero and, by the positive definiteness of $G(x^*)$, the matrices

$$A_k \triangleq \int_0^1 G(x_k + \tau s_k) d\tau$$

are uniformly positive definite for all k large enough. Moreover, by the mean-value theorem, we have $y_k = A_k s_k$. Therefore, there is a constant $\bar{m} > 0$ such that $y_k^T s_k \geq \bar{m} \|s_k\|^2$, which implies that when k is sufficiently large, the condition $y_k^T s_k / \|s_k\|^2 \geq \epsilon \|g_k\|^\alpha$ is always satisfied. This means that Algorithm 1 reduces to the ordinary BFGS method when k is sufficiently large. The superlinear convergence of Algorithm 1 then follows from the related theory in [1, 2, 17]. $\square$

4. Numerical experiments. This section reports some numerical experiments with Algorithm 1. We tested the algorithm on some problems [14] taken from MATLAB with given initial points. These problems can be obtained at the website ftp://ftp.mathworks.com/pub/contrib/v4/optim/uncprobs/. We applied Algorithm 1, which will be called the CBFGS method (C stands for cautious), with the Armijo-type or the Wolfe-type line search or polynomial search, to these problems and compared it with the ordinary BFGS method. We used the condition $\|g(x_k)\| \leq 10^{-6}$ as the stopping criterion. For each problem, we chose the initial matrix $B_0 = I$, i.e., the unit matrix. For each problem, the parameters common to the two methods were set identically. Specifically, we set $\sigma_1 = 0.1$ and $\sigma_2 = 0.9$ in the Wolfe-type line search (2.5), and $\sigma = 0.01$ in the Armijo-type line search (2.6). We let $\epsilon = 10^{-6}$ in the cautious update (2.10). We let $\rho = 0.5$ in the Armijo-type line search. Due to roundoff error, sometimes the directions generated by the algorithms may be not descent. We then use the steepest descent direction instead of the related BFGS (or CBFGS) direction if $g_k p_k > -10^{-14}$.

As to the parameter α in the cautious update (2.10), we first let $\alpha = 0.01$ if $\|g_k\| \geq 1$, and $\alpha = 3$ if $\|g_k\| < 1$. We call this choice Rule 1. Rule 1 is intended to make the cautious update closer to the original BFGS update. It is not difficult to see that the convergence theorems in section 3 remain true if we choose α according to this rule. Indeed, even if α varies in an interval $[\mu_1, \mu_2]$ with $\mu_1 > 0$, all the theorems in section 3 hold true. More generally, as an anonymous referee pointed out, the convergence theorems in section 3 remain true if $\epsilon\|g_k\|^\alpha$ is replaced by a general forcing function $\phi(\|g_k\|)$, which is strictly monotone with $\phi(0) = 0$. We also tested the cautious update (2.10) with $\alpha = 1$ always, which we call Rule 2.

Tables 1, 2, and 3 show the computational results, where the columns have the following meanings:

Problem:	the name of the test problem in MATLAB;
Dim:	the dimension of the problem;
CBFGS^1:	the number of iterations for the cautious BFGS method with Rule 1;
CBFGS^2:	the number of iterations for the cautious BFGS method with Rule 2;
BFGS:	the number of iterations for the BFGS method;
L-BFGS:	the number of iterations for the L-BFGS method [13];
off:	the number of k's such that $y_k^T s_k/\|s_k\|^2 < \epsilon\|g_k\|^\alpha$ (CBFGS), the number of k's such that $y_k^T s_k/\|s_k\|^2 < \epsilon$ (BFGS with Armijo search);
SD:	the number of iterations for which the steepest descent direction used;
fnum:	the number of function evaluations;
gnum:	the number of gradient evaluations.

We first tested the CBFGS method and the BFGS method through a MATLAB code we ourselves wrote. The results are given in Tables 1 and 2. Tables 1 and 2 show the performance of the CBFGS method and the BFGS method with the Armijo-type line search and the Wolfe-type line search, respectively. From the results, we see that the CBFGS method is compatible with the ordinary BFGS method. Moreover, for the problem "meyer," the BFGS method did not terminate regularly, while the CBFGS method did. However, this does not mean that the ordinary BFGS method with the Wolfe search fails to converge when applied to solve this problem. If we adjust the parameters appropriately, then starting from the same initial point, the ordinary BFGS method with the Wolfe search also converges to a station point of this "meyer."

We then edited a MATLAB code for the CBFGS method with a polynomial line search based on a standard low storage BFGS code called L-BFGS given by Kelley [13] (http://www.siam.org/catalog/mcc12/fr18.htm). The test results are given in Table 3. Table 3 shows that for the test problems, the CBFGS method converges to a stationary point of the test problem if the L-BFGS method does.

We observed that the condition in the cautious update was usually satisfied, which suggests that the ordinary BFGS method is generally "cautious," and hence it seldom fails in practice. The results also show that the choice of the parameter α affects the performance of the method. Moreover, if we choose it appropriately, then the condition in the cautious update is almost always satisfied and the CBFGS method essentially reduces to the ordinary BFGS method. However, when it was violated too

TABLE 1
Test results for CBFGS/BFGS methods with Armijo search.

		CBFGS[1]				CBFGS[2]				BFGS			
Problem	Dim	Iter	off	SD	fnum	Iter	off	SD	fnum	Iter	off	SD	fnum
badscb	2	42	0	1	91	-	-	-	-	42	0	1	91
badscp	2	193	7	2	332	277	107	3	452	474	259	34	1335
band	10	-	-	-	-	-	-	-	-	-	-	-	-
bard	3	23	0	0	34	23	0	0	34	26	5	0	40
bd	4	-	-	-	-	-	-	-	-	-	-	-	-
beale	2	15	0	0	24	15	0	0	24	15	0	0	24
biggs	6	42	2	0	52	42	0	0	52	42	2	0	52
box	3	30	0	0	40	30	0	0	40	30	0	0	40
bv	10	18	0	0	39	18	0	0	39	18	0	0	38
froth	2	10	0	0	22	10	0	0	22	13	0	0	51
gauss	3	4	0	0	7	4	0	0	7	4	0	0	7
gulf	3	1	0	0	4	1	0	0	4	1	0	0	2
helix	3	27	0	0	54	27	0	0	54	31	0	0	59
ie	10	11	0	0	14	11	0	0	14	10	0	0	12
ie	100	12	0	0	15	12	0	0	15	12	0	0	14
jensam	2	12	0	0	24	12	0	0	24	12	0	0	24
kowosb	4	28	0	0	32	28	0	0	32	29	0	0	32
lin	10	1	0	0	3	1	0	0	3	1	0	0	3
lin	100	1	0	0	3	1	0	0	3	1	0	0	3
lin1	10	2	0	0	21	2	0	0	21	2	0	0	21
lin0	10	2	0	0	19	2	0	0	19	2	0	0	19
meyer	3	6	0	2	57	2	0	0	19	6	0	2	57
osb1	5	50	0	2	121	32	28	29	191	50	0	2	121
osb2	11	53	0	0	80	53	0	0	80	58	0	0	84
pen1	10	152	3	0	215	152	3	0	215	154	3	0	216
pen1	100	279	1	0	406	279	1	0	406	299	2	0	441
pen2	10	917	0	0	1338	917	0	0	1338	571	0	0	851
rose	2	34	0	0	54	34	0	0	54	34	0	0	54
rosex	100	407	0	1	1148	407	0	0	1148	405	0	1	1134
sing	4	29	0	0	52	29	0	0	52	29	0	0	52
singx	400	191	0	0	633	191	0	0	633	182	0	0	590
trid	10	19	0	0	74	19	0	0	74	42	1	0	94
trid	100	112	0	0	637	121	0	0	637	145	0	0	650
trig	10	26	0	0	27	26	0	0	27	26	0	0	27
trig	100	48	0	0	51	48	0	0	51	48	0	0	51
vardim	10	13	0	0	35	13	0	0	35	13	0	0	35
watson	12	58	0	0	87	58	0	0	87	58	0	0	87
watson	20	55	0	0	91	55	0	0	91	55	0	0	91
wood	4	52	0	0	97	52	0	0	97	28	0	1	73

often, the CBFGS method's performance was worse than the BFGS method, even failing to converge.

5. Conclusion. We have proposed a cautious BFGS update and shown that the method converges globally with the Wolfe-type line search or the Armijo-type line search. The method retains the scale-invariance property of the original BFGS method, except for a minor scale dependence of the skipping condition in the cautious update (2.10). Moreover, the cautious update makes B_{k+1} inherit the positive definiteness of B_k no matter what line search is used. The established global convergence theorems do not rely on the convexity assumption on the objective function. The reported numerical results show that the BFGS method with the proposed cautious update is comparable to the ordinary BFGS method. Moreover, the conditions used

TABLE 2
Test results for CBFGS/BFGS methods with Wolfe search.

		CBFGS[1]			CBFGS[2]			BFGS		
Problem	Dim	Iter(off, SD)	fnum	gnum	Iter(off, SD)	fnum	gnum	Iter(SD)	fnum	gnum
badscb	2	5 (5, 0)	34	12	-	-	-	26(1)	114	175
badscp	2	319(199, 8)	1391	2196	180(20, 7)	782	1223	174(10)	587	827
band	10	-	-	-	-	-	-	-	-	-
bard	3	14(0, 0)	46	47	14(0, 0)	44	67	14(0)	46	47
bd	4	-	-	-	-	-	-	-	-	-
beale	2	14(0, 0)	38	55	14(0, 0)	38	55	14(0)	38	55
biggs	6	30(0, 0)	97	159	30(0, 0)	97	159	30(0)	97	159
box	3	16(1, 0)	68	111	20(0, 0)	75	123	20(0)	75	123
bv	10	15(0, 0)	42	54	15(0, 0)	42	54	15(0)	42	54
froth	2	11(2, 0)	27	34	8(0, 0)	24	31	8(0)	24	31
gauss	3	4(0, 0)	13	19	4(0, 0)	13	19	4(0)	13	19
gulf	3	1(0, 0)	2	2	1(0, 0)	2	2	1(0)	2	2
helix	3	30(2, 0)	103	147	28(0, 0)	94	137	28(0)	94	137
ie	10	11(0, 0)	17	20	11(0, 0)	17	20	11(0)	17	20
ie	100	13(0, 0)	19	22	13(0, 0)	19	22	13(0)	19	22
jensam	2	10(0, 0)	30	41	10(0, 0)	30	41	10(0)	30	41
kowosb	4	21(0, 0)	74	122	21(0, 0)	74	122	21(0)	74	122
lin	10	2(0, 0)	9	13	2(0, 0)	9	13	2(0)	9	13
lin	100	2(0, 0)	9	13	2(0, 0)	9	13	2(0)	9	13
lin1	10	5(3, 0)	46	38	2(0, 0)	13	11	2(0)	13	11
lin0	10	5(3, 0)	26	6	2(0, 0)	8	3	2(0)	8	3
meyer	3	11(9, 0)	72	76	-	-	-	-	-	-
osb1	5	44(0,2)	157	221	44(0,2)	157	221	44(0,2)	157	221
osb2	11	52(0, 0)	153	231	52(0, 0)	153	231	52(0)	153	231
pen1	10	104(2, 0)	374	597	34(0, 0)	152	253	34(0)	152	253
pen1	100	36(4, 0)	135	185	70(0, 0)	333	553	70(0)	333	553
pen2	10	852(0, 0)	2594	4135	852(0, 0)	2594	4135	852(0)	2594	4135
rose	2	28(0, 0)	93	141	28(0, 0)	93	141	28(0)	93	141
rosex	100	322(1, 1)	1348	1843	333(0, 2)	1409	1934	333(2)	1409	1934
sing	4	35(0, 0)	134	216	35(0, 0)	134	216	35(0)	134	216
singx	400	186(4, 0)	674	833	196(0, 0)	671	817	196(0)	671	817
trid	10	20(0, 0)	57	57	20(0, 0)	57	57	20(0)	57	57
trid	100	95(0, 0)	546	622	95(0, 0)	546	622	95(0)	546	622
trig	10	24(0, 0)	58	91	24(0, 0)	58	91	24(0)	58	91
trig	100	46(0, 0)	117	185	46(0, 0)	177	185	46(0)	117	185
vardim	10	6(2, 0)	30	25	6(0, 0)	42	65	6(0)	42	65
watson	12	41(0, 0)	172	280	41(0, 0)	172	280	41(0)	172	280
watson	20	47 (0, 0)	210	348	47(0, 0)	210	348	47(0)	210	348
wood	4	48(1, 0)	156	225	49(0, 0)	149	220	49(0)	149	220

in the cautious rule are generally satisfied and hence the cautious update essentially reduces to the ordinary update in most cases. This suggests that the ordinary BFGS method is generally "cautious," and hence the BFGS method seldom fails in practice. We hope that the results established in this paper contribute toward resolving the fundamental open problem of whether the BFGS method converges for nonconvex unconstrained optimization problems.

Acknowledgments. The authors would like to thank two anonymous referees for their helpful comments, which have made the paper clearer and more comprehensive than the earlier version.

TABLE 3
Test results for CBFGS/L-BFGS methods with polynomial search.

		CBFGS[1]			CBFGS[2]			BFGS		
Problem	Dim	Iter	off	fnum	Iter	off	fnum	Iter	off	fnum
badscb	2	-	-	-	-	-	-	-	-	-
badscp	2	-	-	-	-	-	-	-	-	-
band	10	-	-	-	-	-	-	-	-	-
bard	3	28	1	64	28	1	64	28	1	64
bd	4	491	3	1021	491	3	1021	-	-	-
beale	2	19	0	42	19	0	42	19	0	42
biggs	6	52	1	116	52	1	116	52	1	116
box	3	30	0	82	38	0	82	38	0	82
bv	10	17	0	42	17	0	42	17	0	42
froth	2	9	0	23	9	0	23	9	0	23
gauss	3	3	0	8	3	0	8	3	0	8
gulf	3	1	0	3	1	0	3	1	0	3
helix	3	37	0	83	37	0	83	37	0	83
ie	10	9	0	21	9	0	21	9	0	21
ie	100	10	0	23	10	0	23	10	0	23
jensam	2	12	0	31	12	0	31	12	0	31
kowosb	4	28	0	60	28	0	60	28	0	60
lin	10	1	0	4	1	0	4	1	0	4
lin	100	1	0	4	1	0	4	1	0	4
lin1	10	-	-	-	-	-	-	-	-	-
lin0	10	743	4	1501	743	4	1501	-	-	-
meyer	3	-	-	-	-	-	-	-	-	-
osb1	5	-	-	-	-	-	-	-	-	-
osb2	11	58	0	136	58	0	136	72	1	174
pen1	10	188	3	433	188	3	433	199	7	454
pen1	100	796	5	1639	796	5	1639	372	7	773
pen2	10	305	2	721	305	2	721	258	5	592
rose	2	39	0	91	39	0	91	39	0	91
rosex	100	39	0	91	39	0	91	39	0	91
sing	4	50	0	110	50	0	110	50	0	110
singx	400	55	0	116	55	0	116	69	1	148
trid	10	18	0	60	18	0	60	18	0	60
trid	100	127	0	467	127	0	467	131	2	526
trig	10	26	0	53	26	0	53	26	0	53
trig	100	49	0	103	49	0	103	49	0	103
vardim	10	407	2	833	407	2	833	176	3	377
watson	12	61	0	137	61	0	137	87	1	197
watson	20	58	0	133	58	0	133	89	1	204
wood	4	30	0	75	30	0	75	30	0	75

REFERENCES

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

[3] 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. 549–560.

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

[5] J. E. DENNIS, JR. AND R. B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1983.

[6] L. C. W. DIXON, *Variable metric algorithms: Necessary and sufficient conditions for identical behavior on nonquadratic functions*, J. Optim. Theory Appl., 10 (1972), pp. 34–40.

[7] R. Fletcher, *Practical Methods of Optimization,* 2nd ed., John Wiley & Sons, Chichester, 1987.

[8] R. Fletcher, *An overview of unconstrained optimization*, in Algorithms for Continuous Optimization: The State of the Art, E. Spedicato, ed., Kluwer Academic Publishers, Boston, 1994, pp. 109–143.

[9] A. Griewank, *The global convergence of partitioned BFGS on problems with convex decompositions and Lipschitzian gradients*, Math. Program., 50 (1991), pp. 141–175.

[10] A. Griewank and Ph. L. Toint, *Local convergence analysis for partitioned quasi-Newton updates*, Numer. Math., 39 (1982), pp. 429–448.

[11] 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.

[12] D. H. Li and M. Fukushima, *A modified BFGS method and its global convergence in nonconvex minimization*, J. Comput. Appl. Math., to appear.

[13] C. T. Kelley, *Iterative Methods for Optimization*, SIAM, Philadelphia, 1999.

[14] J. J. Moré, B. S. Garbow, and K. E. Hillstrome, *Testing unconstrained optimization software*, ACM Trans. Math. Software, 7 (1981), pp. 17–41.

[15] J. Nocedal, *Theory of algorithms for unconstrained optimization*, Acta Numer., 1 (1992), pp. 199–242.

[16] M. J. D. Powell, *On the convergence of the variable metric algorithm*, J. Inst. Math. Appl., 7 (1971), pp. 21–36.

[17] M. J. D. Powell, *Some global convergence properties of a variable metric algorithm for minimization without exact line searches*, in Nonlinear Programming, SIAM-AMS Proc. IX, R. W. Cottle and C. E. Lemke, eds., AMS, Providence, RI, 1976, pp. 53–72.

[18] Ph. L. Toint, *Global convergence of the partitioned BFGS algorithm for convex partially separable optimization*, Math. Program., 36 (1986), pp. 290–306.