On the Global Convergence of BFGS Method for Nonconvex Unconstrained Optimization Problems

Article *in* SIAM Journal on Optimization · May 2000

DOI: 10.1137/S1052623499354242 · Source: CiteSeer

Masao Fukushima
Nanzan University

On the Global Convergence of BFGS Method for Nonconvex Unconstrained Optimization Problems

Dong-Hui Li [1]
Department of Applied Mathematics
Hunan University
Changsha, China 410082
e-mail: dhli@mail.hunu.edu.cn

Masao Fukushima
Department of Applied Mathematics and Physics
Graduate School of Informatics
Kyoto University
Kyoto 606-8501, Japan
e-mail: fuku@kuamp.kyoto-u.ac.jp

April 7, 1999

Abstract

This paper is concerned with the open problem whether 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 Wolfe-type or Armijo-type line search converges globally if the function to be minimized has Lipschitz continuous gradients.

Key words: unconstrained optimization, BFGS method, global convergence

[1] Present address (available until October, 1999): Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan, e-mail: lidh@kuamp.kyoto-u.ac.jp

1 Introduction

BFGS method is a well-known quasi-Newton method for solving unconstrained optimization problems. 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 theory of BFGS method has been well established [3, 4]. The study on global convergence of BFGS method has also made good progress. In particular, for convex minimization problems, it has been shown that the iterates generated by BFGS are globally convergent, if the exact line search or some special inexact line search is used [1, 2, 5, 8, 13, 14, 15]. On the contrary, however, little is known concerning global convergence of BFGS method for nonconvex minimization problems. Indeed, so far, no one has proved global convergence of BFGS method for nonconvex minimization problems, or has given a counter example that shows nonconvergence of BFGS method. Whether 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 [7, 12].

Recently, the authors [10] proposed a modified BFGS method and established its global convergence for nonconvex unconstrained optimization problems. The authors [9] also proposed a globally convergent Gauss-Newton based BFGS method for symmetric nonlinear equations that particularly contain unconstrained optimization problems as a special case. The results obtained in [9] and [10] 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 BFGS method and prove that the method with Wolfe-type or 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.

In the next section, we present BFGS method with 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.

Some notations: 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 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 BFGS 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 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 particular, if $\lambda_k = 1$ satisfies (2.5), we take $\lambda_k = 1$. Global convergence of BFGS method with line search (2.4) or (2.5) for convex minimization problems has been studied in [1, 2, 5, 8, 13, 14, 15].

Another important inexact line search is Armijo-type line search that finds a λ_k satisfying

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

where σ is a constant such that $\sigma \in (0,1)$. Line search (2.6) is usually implemented by letting λ_k be the maximum value in the set $\{\rho^i \mid i = 0, 1, 2, \ldots\}$, where $\rho \in (0,1)$ is a given constant. 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 B_k is updated or not. 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 [14].

Lemma 2.1 *(Powell [14]) If BFGS method with 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*

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

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 uniformly convex, then (2.8) always holds. Therefore, global convergence of BFGS method follows from

Lemma 2.1 immediately. However, in the case where f is nonconvex, it seems difficult to guarantee (2.8). Maybe this is a reason why global convergence of BFGS method has not been proved. In [10], 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 [10] is proved without convexity assumption on f by means of Lemma 2.1 with a contradictory assumption that $\{\|g_k\|\}$ are bounded away from zero. However, this method lacks the scale-invariance property the original BFGS method enjoys. We now further study global convergence of 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 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 contained in a bounded convex set D. The function f is continuously differentiable on D and there exists a constant $L > 0$ such that

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

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 sets

$$\bar{K} = \{i \mid \frac{y_i^T s_i}{\|s_i\|^2} \ge \epsilon \|g_i\|^\alpha\} \quad \text{and} \quad \bar{K}_k = \{i \in \bar{K} \mid i \le k\}. \tag{3.2}$$

Let i_k be the number of indices $i \in \bar{K}_k$. By means of $\bar{K}$, we may rewrite (2.10) as

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

Taking trace operation on the both sides of (3.3), we get for any k

$$\text{tr}\,(B_{k+1}) = \text{tr}\,(B_0) - \sum_{i\in\bar{K}_k} \frac{\|B_i s_i\|^2}{s_i^T B_i s_i} + \sum_{i\in\bar{K}_k} \frac{\|y_i\|^2}{y_i^T s_i}. \tag{3.4}$$

Now we establish global convergence of Algorithm 1. We first show that Algorithm 1 converges globally if $\bar{K}$ is finite.

Theorem 3.1 *Let Assumption A hold and $\{x_k\}$ be generated by Algorithm 1. If $\bar{K}$ is finite, then we have*

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{3.5}$$

Proof The assumption that $\bar{K}$ is finite implies that there is an index k_0 such that $B_k = B_{k_0} \triangleq B$ holds for all $k \geq k_0$. By the positive definiteness of B, there are positive constants $m_1 \leq M_1$ such that

$$m_1\|p\|^2 \leq p^T Bp \leq M_1\|p\|^2, \quad m_1\|p\|^2 \leq p^T B^{-1}p \leq M_1\|p\|^2, \quad \forall p \in R^n. \tag{3.6}$$

If Wolfe-type line search (2.5) is used, then we get from (3.1) and the second inequality of (2.5)

$$\begin{aligned} L\|s_k\|^2 &\geq y_k^T s_k \geq -(1-\sigma_2)g_k^T s_k \\ &= (1-\sigma_2)\lambda_k^{-1} s_k^T B s_k \\ &\geq (1-\sigma_2)\lambda_k^{-1} m_1 \|s_k\|^2, \quad \forall k \geq k_0, \end{aligned}$$

where the last inequality follows from (3.6). This implies

$$\lambda_k \geq (1-\sigma_2)m_1 L^{-1}, \qquad \forall k \geq k_0.$$

Therefore, we get from (2.12)

$$g_k B^{-1} g_k = g_k^T p_k \to 0.$$

This together with (3.6) implies (3.5).

Next, we consider the case where λ_k is determined by Armijo-type line search (2.6). Let $\bar{\lambda} = \limsup_{k\to\infty} \lambda_k$. If $\bar{\lambda} > 0$, then using an argument similar to the above, we get (3.5). Suppose that $\bar{\lambda} = 0$. This means $\lim_{k\to\infty} \lambda_k = 0$. Let $\bar{x}$ be an arbitrary accumulation point of $\{x_k\}$ and $\{x_k\}_{k\in K}$ be a subsequence converging to $\bar{x}$. Since $p_k = -B^{-1}g_k$ for $k \geq k_0$, it follows that $\{p_k\}_{k\in K} \to \bar{p} \triangleq -B^{-1}g(\bar{x})$. By the line search rule, when k is sufficiently large, $\lambda_k' \triangleq \lambda_k/\rho$ does not satisfy (2.6). So, we have

$$f(x_k + \lambda_k' p_k) - f(x_k) - \sigma\lambda_k' g(x_k)^T p_k \geq 0.$$

Dividing the both sides by λ_k' and then taking the limit yield

$$(1-\sigma)g(\bar{x})^T \bar{p} \geq 0,$$

which implies

$$-g(\bar{x})^T B^{-1} g(\bar{x}) \geq 0.$$

By the positive definiteness of B, we get $g(\bar{x}) = 0$. The proof is then complete. $\square$

We proceed to showing global convergence of Algorithm 1 in the case where $\bar{K}$ is infinite. We will deduce a contradiction with the assumption that there is a constant $\delta > 0$ such that

$$\|g_k\| \geq \delta \tag{3.7}$$

holds for all k. Before establishing a global convergence theorem for Algorithm 1, we show some useful lemmas.

Lemma 3.1 *Let Assumption A hold and $\{x_k\}$ be generated by Algorithm 1. If (3.7) holds for all k, then there exists a constant $M_2 > 0$ such that the inequalities*

$$\text{tr }(B_{k+1}) \leq M_2 i_k \tag{3.8}$$

and

$$\sum_{i \in \bar{K}_k} \frac{\|B_i s_i\|^2}{s_i^T B_i s_i} \leq M_2 i_k \tag{3.9}$$

hold for all k sufficiently large.

Proof It follows from (3.2) and (3.7) that

$$y_i^T s_i \geq \epsilon \delta^\alpha \|s_i\|^2 \tag{3.10}$$

holds for all $i \in \bar{K}$. This together with (3.1) implies that for any $i \in \bar{K}$

$$\frac{\|y_i\|^2}{y_i^T s_i} \leq \frac{L^2}{\epsilon \delta^\alpha} \triangleq M_2'. \tag{3.11}$$

This together with (3.4) yields inequality (3.8) with a suitable constant M_2. Moreover, since tr $(B_{k+1}) > 0$ holds for any k, we get from (3.4) and (3.11)

$$\sum_{i \in \bar{K}_k} \frac{\|B_i s_i\|^2}{s_i^T B_i s_i} \leq \text{ tr }(B_0) + M_2' i_k.$$

This yields inequality (3.9) with a suitable constant M_2. $\square$

Lemma 3.2 *Let Assumption A hold. If (3.7) holds for all k, then there exist positive constants β_1, β_2 and β_3 such that for any $k > 1$ there are at least $\lceil i_k/2 \rceil$ indices $i \in \bar{K}_k$ such that*

$$\|B_i s_i\| \le \beta_1 \|s_i\|, \quad \beta_2 \|s_i\|^2 \le s_i^T B_i s_i \le \beta_3 \|s_i\|^2. \tag{3.12}$$

Proof Notice that (3.7) implies that (3.10) and (3.11) hold for any $i \in \bar{K}$. Therefore, in a way similar to the proof of Theorem 2.1 in [1], we can show the conclusion. □

Lemma 3.2 shows that when $\bar{K}$ is infinite, if (3.7) holds for all k, then there exists an infinite index set $\tilde{K} \subset \bar{K}$ such that

$$\|g_i\| = \|B_i p_i\| \le \beta_1 \|p_i\|, \quad \forall i \in \tilde{K} \tag{3.13}$$

and

$$\|p_i\|^2 \le \beta_2^{-1} p_i^T B_i p_i = -\beta_2^{-1} g_i^T p_i \le \beta_2^{-1} \|g_i\| \, \|p_i\|, \quad \forall i \in \tilde{K}, \tag{3.14}$$

and hence

$$\|p_i\| \le \beta_2^{-1} \|g_i\|, \qquad \forall i \in \tilde{K}. \tag{3.15}$$

Moreover, (3.13) and (3.14) imply

$$\|g_i\|^2 \le \beta_1^2 \|p_i\|^2 \le -\beta_1^2 \beta_2^{-1} g_i^T p_i, \qquad \forall i \in \tilde{K}. \tag{3.16}$$

We prove global convergence of Algorithm 1 with Armijo-type line search.

Theorem 3.2 *Let Assumption A hold and $\{x_k\}$ be generated by Algorithm 1 with λ_k being determined by Armijo-type line search (2.6). Then*

$$\liminf_{k \to \infty} \|g_k\| = 0. \tag{3.17}$$

Proof By Theorem 3.1, it suffices to verify (3.17) when $\bar{K}$ is infinite. Suppose that (3.17) does not hold. Then there is a constant $\delta > 0$ such that (3.7) holds for all k. Let the set $\tilde{K} \subset \bar{K}$ be as specified in the paragraph preceding Theorem 3.2. Then $\tilde{K}$ contains infinitely many indices. Denote $\bar{\lambda} = \limsup_{k \in \tilde{K}, k \to \infty} \lambda_k = \lim_{k \in K', k \to \infty} \lambda_k$, where $K' \subset \tilde{K}$. Since $\{x_k\}_{k \in K'}$ is bounded, it follows from (3.15) that $\{p_k\}_{k \in K'}$ is also bounded. Without loss of generality, we assume that the sequences $\{x_k\}_{k\in K'}$ and $\{p_k\}_{k\in K'}$ converge to some vectors $\bar{x}$ and $\bar{p}$, respectively. It then follows from (2.12) that $\bar{\lambda} g(\bar{x})^T \bar{p} = 0$. By (3.16), it suffices to show that $\bar{\lambda} > 0$. We assume the contrary $\bar{\lambda} = 0$. Then by the line search rule, for all $k \in K'$ sufficiently large, $\lambda_k' \triangleq \lambda_k/\rho$ does not satisfy (2.6). This means

$$f(x_k + \lambda_k' p_k) - f(x_k) \geq \sigma \lambda_k' g_k^T p_k. \tag{3.18}$$

By the mean-value theorem, there is a $\theta_k \in (0,1)$ such that $f(x_k + \lambda_k' p_k) - f(x_k) = \lambda_k' g(x_k + \theta_k \lambda_k' p_k)^T p_k$. Applying this to (3.18), we deduce

$$\begin{aligned} L\lambda_k' \|p_k\|^2 &\geq (g(x_k + \theta_k \lambda_k' p_k) - g(x_k))^T p_k \\ &\geq -(1-\sigma) g_k^T p_k \\ &\geq (1-\sigma)\beta_2 \|p_k\|^2, \end{aligned}$$

where the first inequality follows from (3.1) and the last inequality follows from (3.14). The last inequality contradicts the assumption $\bar{\lambda} = 0$. The proof is then complete. □

We turn to showing global convergence of Algorithm 1 with Wolfe-type line search. To this end, we show a useful lemma similar to Lemma 3.2 in [2].

Lemma 3.3 *Let Assumption A hold and $\{x_k\}$ be generated by Algorithm 1 with λ_k being determined by Wolfe-type line search (2.5). If (3.7) holds for all k, then there exists a constant $m_3 > 0$ such that for all k large enough*

$$\prod_{i \in \bar{K}_k} \lambda_i \geq m_3^{i_k}. \tag{3.19}$$

Proof The formula (3.3) gives the recurrence relation (see e.g. (3.13) in [14])

$$\det B_{i+1} = \frac{y_i^T s_i}{s_i^T B_i s_i} \det B_i, \qquad \forall i \in \bar{K} \tag{3.20}$$

and

$$\det B_{i+1} = \det B_i, \qquad \forall i \notin \bar{K}. \tag{3.21}$$

Let n_k be the largest index in $\bar{K}_k$. Multiplying the inequalities (3.20) for $i \in \bar{K}_k$ and (3.21) for $i \notin \bar{K}_k$ yields

$$\det B_{n_k+1} = \det B_0 \prod_{i \in \bar{K}_k} \frac{y_i^T s_i}{s_i^T B_i s_i}. \tag{3.22}$$

On the other hand, the second inequality of (2.5) implies that for each i

$$y_i^T s_i \geq -(1-\sigma_2) g_i^T s_i = (1-\sigma_2)\lambda_i^{-1} s_i^T B_i s_i.$$

Then in a way similar to the proof of Lemma 3.2 in [2], we get (3.19) by applying the last inequality and (3.8) to (3.22). □

Now we prove global convergence of Algorithm 1 with Wolfe-type line search.

Theorem 3.3 *Let Assumption A hold and $\{x_k\}$ be generated by Algorithm 1 with λ_k being determined by Wolfe-type line search (2.5). Then (3.17) holds.*

Proof By Theorem 3.1, it suffices to verify (3.17) when $\bar{K}$ is infinite. Denote $\bar{K} = \{k_1 < k_2 < \ldots\}$. Notice that (2.11) particularly implies

$$-\sum_{j=1}^{\infty} g_{k_j}^T s_{k_j} < \infty.$$

Since $B_{k_j} s_{k_j} = -\lambda_{k_j} g_{k_j}$, it follows that

$$\sum_{j=1}^{\infty} \|g_{k_j}\|^2 \lambda_{k_j} \frac{s_{k_j}^T B_{k_j} s_{k_j}}{\|B_{k_j} s_{k_j}\|^2} = -\sum_{j=1}^{\infty} g_{k_j}^T s_{k_j} < \infty \tag{3.23}$$

If (3.17) does not hold, then there exists a constant $\delta > 0$ such that (3.7) holds for all k. So, (3.23) implies

$$\sum_{j=1}^{\infty} \lambda_{k_j} \frac{s_{k_j}^T B_{k_j} s_{k_j}}{\|B_{k_j} s_{k_j}\|^2} < \infty. \tag{3.24}$$

Therefore, for any $\zeta > 0$, there exists an integer $j_0 > 0$ such that for any positive integer q,

$$\Big(\prod_{j=j_0+1}^{j_0+q} \lambda_{k_j} \frac{s_{k_j}^T B_{k_j} s_{k_j}}{\|B_{k_j} s_{k_j}\|^2}\Big)^{\frac{1}{q}} \leq \frac{1}{q} \sum_{j=j_0+1}^{j_0+q} \lambda_{k_j} \frac{s_{k_j}^T B_{k_j} s_{k_j}}{\|B_{k_j} s_{k_j}\|^2} \leq \frac{\zeta}{q},$$

where the left-hand inequality follows from the geometric inequality. Thus

$$\begin{aligned}
\Big(\prod_{j=j_0+1}^{j_0+q} \lambda_{k_j}\Big)^{\frac{1}{q}} &\leq \frac{\zeta}{q}\Big(\prod_{j=j_0+1}^{j_0+q} \frac{\|B_{k_j} s_{k_j}\|^2}{s_{k_j}^T B_{k_j} s_{k_j}}\Big)^{\frac{1}{q}} \\
&\leq \frac{\zeta}{q^2} \sum_{j=j_0+1}^{j_0+q} \frac{\|B_{k_j} s_{k_j}\|^2}{s_{k_j}^T B_{k_j} s_{k_j}} \\
&\leq \frac{\zeta}{q^2} \sum_{j=0}^{j_0+q} \frac{\|B_{k_j} s_{k_j}\|^2}{s_{k_j}^T B_{k_j} s_{k_j}} \\
&\leq \frac{\zeta(j_0+q+1)}{q^2} M_2,
\end{aligned}$$

where the last inequality follows from (3.9). Letting $q \to \infty$ yields a contradiction, because Lemma 3.3 ensures that the left-hand side of the above inequality is greater than a positive constant. The proof is complete. $\square$

Theorems 3.1, 3.2 and 3.3 show that there exists a subsequence of $\{x_k\}$ converging to a stationary point of (2.1). The following theorem shows that if additional conditions are assumed, then the whole sequence converges to a local optimal solution of (2.1).

Theorem 3.4 *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 $\sigma, \sigma_2 \in (0, 1/2)$, 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 $\frac{y_k^T s_k}{\|s_k\|^2} \geq \epsilon \|g_k\|^\alpha$ 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, 14]. $\square$

Theorem 3.4 shows a strong convergence property of Algorithm 1. However, it seems difficult in practice to verify the condition $s_k \to 0$ for either of the two line searches used in the algorithm. In the following, we propose an extension of Armijo-type line search to relax this condition.

Let $\sigma_3 \in (0, 1)$ and $\sigma_4 > 0$ be given constants. We determine a stepsize λ_k

satisfying the inequality

$$f(x_k + \lambda_k p_k) \le f(x_k) + \sigma_3 \lambda_k g_k^T p_k - \sigma_4 \|\lambda_k p_k\|^2. \tag{3.25}$$

The only difference between (3.25) and (2.6) lies in the term $-\sigma_4\|\lambda_k p_k\|^2$. Since p_k is a descent direction of f at x_k and $-\sigma_4\|\lambda_k p_k\|^2 = o(\lambda_k)$ as λ_k goes to zero, it is clear that (3.25) holds for all sufficiently small $\lambda_k > 0$. Therefore, we can find a λ_k by a backtracking process similar to Armijo-type line search. In a way similar to Theorems 3.1 and 3.2, it is also not difficult to prove global convergence of Algorithm 1 with line search (3.25). Moreover, (3.25) particularly implies $\sigma_4\|s_k\|^2 \le f(x_k) - f(x_{k+1})$. It then follows from the descent property of $\{f(x_k)\}$ that $s_k \to 0$. Therefore, we can establish a theorem similar to Theorems 3.1 and 3.4. We state the global convergence theorem without proof.

Theorem 3.5 *Let Assumption A hold and $\{x_k\}$ be generated by Algorithm 1 with λ_k satisfying (3.25). Then (3.17) holds. If we further suppose that f is twice continuously differentiable and 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 at x^* and $\sigma_3 \in (0, 1/2)$, then the convergence rate is superlinear.*

4 Numerical Experiments

This section reposts some numerical experience with Algorithm 1. We tested the algorithm on the following three problems taken from [11].

Problem 1: Extended Powell singular function

$$\begin{aligned} f(x) &= \sum_{i=1}^{n/4} \Big\{ (x_{4i-3} + 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 \\ &\quad + (x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-3} - x_{4i})^4 \Big\}, \\ & \qquad x^* = (0, \ldots, 0)^T. \end{aligned}$$

Problem 2: Extended Rosenbrock function

$$f(x) = \sum_{i=1}^{n/2} \Big\{ 100(x_{2i} - x_{2i-1}^2)^2 + (1 - x_{2i-1})^2 \Big\},$$

$$x^* = (1, \ldots, 1)^T.$$

Problem 3: Extended Wood function

$$\begin{aligned} f(x) &= \sum_{i=1}^{n/4} \Big\{ 100(x_{4i-2} - x_{4i-3}^2)^2 + (1 - x_{4i-3})^2 + 90(x_{4i} - x_{4i-1}^2)^2 \\ &\quad + (1 - x_{4i-1})^2 + 10(x_{4i-2} + x_{4i} - 2)^2 + 0.1(x_{4i-2} - x_{4i})^2 \Big\}, \end{aligned}$$

$$x^* = (1, \ldots, 1)^T.$$

We applied Algorithm 1, which will be called CBFGS method (C stands for cautious), with Wolfe-type or Armijo-type line search to these problems and compared it with the ordinary BFGS method. We used the condition $\max\{\|g(x_k)\|,\ \|x_k - x^*\|\} \leq 10^{-5}$ as the stopping criterion. For each problem, we chose different initial starting points but the same 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 chose the parameters as follows. We set the parameters as $\sigma_1 = 0.1$ and $\sigma_2 = 0.49$ in Wolfe-type line search (2.5) and $\sigma = 0.1$ in Armijo-type line search (2.6).

As to the parameters α and ϵ in the cautious update (2.10), we first let

$$\alpha = \begin{cases} 0.01, & \text{if } \|g_k\| \geq 1, \\ 3, & \text{if } \|g_k\| < 1 \end{cases}$$

and $\epsilon = 0.1$. This choice is intended to make the cautious update closer to the original BFGS method. It is not difficult to see that the convergence theorems in Section 3 remain true if we choose α according to this rule. Indeed, more generally, even if α varies in an interval $[\mu_1, \mu_2]$ with $\mu_1 > 0$, all the theorems in Section 3 hold true. The results are shown in Tables 1-4, where P1, P2 and P3 stand for Problems 1,2, and 3, respectively, and k stands for the number of iterations. For CBFGS method, 'off' denotes the number of times the condition in the cautious update was not met, that is, the numbers of k's such that $y_k^T s_k / \|s_k\|^2 < \epsilon \|g_k\|^\alpha$. For BFGS method, 'off' denotes the number of k's such that $y_k^T s_k < 10^{-17}$. Note that, for BFGS method with Wolfe-type line search, 'off' is normally zero since $y_k^T s_k$ is always positive because of the second inequality in (2.5). In the tables, 'Init' stands for the initial point. The tested initial points are $x^0 = (0,0,\ldots,0)^T$, $x^1 = (1,1,\ldots,1)^T$, $x^2 = (10,10,\ldots,10)^T$, $x^3 = (100,100,\ldots,100)^T$, $x^4 = -x^2$, $x^5 = -x^3$, $x^6 = (0,100,0,100,\ldots)^T$ and $x^7 = -x^6$.

Next, to check the influence of α and ϵ on CBFGS method, we solved Problem 1 with various values of α and ϵ starting from the same initial point $x^2 = (10,10,\ldots,10)^T$. The results are shown in Table 5, where

$$\alpha_1 = \begin{cases} 0.0001, & \text{if } \|g_k\| \geq 1, \\ 4, & \text{if } \|g_k\| < 1, \end{cases} \qquad \alpha_2 = \begin{cases} 0.001, & \text{if } \|g_k\| \geq 1, \\ 4, & \text{if } \|g_k\| < 1, \end{cases}$$

$$\alpha_3 = \begin{cases} 0.01, & \text{if } \|g_k\| \geq 1, \\ 4, & \text{if } \|g_k\| < 1, \end{cases} \qquad \alpha_4 = \begin{cases} 0.1, & \text{if } \|g_k\| \geq 1, \\ 4, & \text{if } \|g_k\| < 1, \end{cases}$$

$$\alpha_5 = \begin{cases} 1, & \text{if } \|g_k\| \geq 1, \\ 4, & \text{if } \|g_k\| < 1, \end{cases} \qquad \alpha_6 = 1, \quad \alpha_7 = 2.$$

The results show that in most cases, CBFGS method performs equally well compared with the ordinary BFGS method. We observed that the condition in the cautious update was usually satisfied. However, when it was violated too often, CBFGS method performed worse than BFGS method.

The results also show that the choice of parameters α and ϵ are important computationally. If we choose α and ϵ appropriately, then the condition in the cautious update is almost always satisfied and CBFGS method essentially reduces to the ordinary BFGS method.

Table 1
Test Results for CBFGS Method (Wolfe Search)

Problem	n	Init	k	$\|\nabla f_k\|$	$\|x_k - x^*\|$	off
P1	4	x^1	31	1.0E-7	1.3E-3	0
P1	4	x^2	35	1.0E-6	3.6E-3	0
P1	4	x^3	39	1.0E-6	4.6E-3	0
P1	4	x^4	35	1.0E-6	3.6E-3	0
P1	4	x^5	39	1.0E-6	4.6E-3	0
P1	4	x^6	43	1.0E-6	1.7E-3	0
P1	4	x^7	43	1.0E-6	1.7E-3	0
P2	2	x^0	17	1.1E-4	1.0E-5	0
P2	2	x^2	62	2.0E-6	1.2E-5	1
P2	2	x^4	75	1.0E-7	2.0E-6	1
P2	2	x^6	44	1.1E-5	1.0E-6	0
P2	2	x^7	25	1.0E-7	4.3E-5	0
P2	10	x^0	37	1.7E-4	8.0E-6	0
P2	10	x^2	134	1.0E-6	1.5E-5	0
P2	10	x^4	140	2.9E-5	6.0E-6	1
P2	10	x^6	99	7.7E-5	6.0E-6	0
P2	10	x^7	70	3.0E-6	1.5E-5	0
P2	100	x^0	93	4.7E-5	6.0E-6	0
P2	100	x^2	354	8.0E-6	1.3E-5	0
P2	100	x^6	308	5.1E-5	7.0E-6	0
P2	100	x^7	277	3.6E-5	9.0E-6	0
P3	4	x^0	19	3.4E-5	4.0E-6	0
P3	4	x^2	29	4.9E-5	1.0E-7	0
P3	4	x^3	45	2.4E-4	6.0E-6	0
P3	4	x^6	31	4.0E-4	1.0E-6	0
P3	4	x^7	38	7.2E-5	1.0E-7	0
P3	40	x^0	51	6.9E-5	4.0E-6	0
P3	40	x^2	185	7.0E-4	8.0E-6	0
P3	40	x^3	499	1.3E-4	6.0E-6	0
P3	40	x^4	530	7.4E-5	3.0E-6	0
P3	40	x^5	524	8.0E-6	1.0E-6	0
P3	40	x^6	168	1.8E-4	5.0E-6	0
P3	40	x^7	148	2.0E-5	1.0E-5	0
P3	100	x^0	75	1.8E-4	6.0E-6	0
P3	100	x^2	352	1.0E-4	8.0E-6	0
P3	100	x^3	758	2.1E-5	5.0E-6	0
P3	100	x^4	875	1.6E-4	5.0E-6	0
P3	100	x^5	1239	1.1E-4	7.0E-6	0
P3	100	x^6	289	1.0E-4	5.0E-6	0
P3	100	x^7	233	3.4E-4	7.0E-6	0

Table 2
Test Results for BFGS Method (Wolfe Search)

Problem	n	Init	k	$\|\nabla f_k\|$	$\|x_k - x^*\|$	off
P1	4	x^1	31	1.0E-7	1.3E-3	0
P1	4	x^2	35	1.0E-6	3.6E-3	0
P1	4	x^3	39	1.0E-6	4.6E-3	0
P1	4	x^4	35	1.0E-6	3.6E-3	0
P1	4	x^5	39	1.0E-6	4.6E-3	0
P1	4	x^6	43	1.0E-6	1.7E-3	0
P1	4	x^7	43	1.0E-6	1.7E-3	0
P2	2	x^0	17	1.1E-4	1.0E-5	0
P2	2	x^2	61	1.8E-4	4.0E-6	0
P2	2	x^4	70	1.1E-4	7.0E-6	0
P2	2	x^6	44	1.1E-5	1.0E-6	0
P2	2	x^7	25	1.0E-7	4.3E-5	0
P2	10	x^0	37	1.7E-4	8.0E-6	0
P2	10	x^2	134	1.0E-6	1.5E-5	0
P2	10	x^4	144	1.4E-5	2.0E-6	0
P2	10	x^6	99	7.7E-5	6.0E-6	0
P2	10	x^7	70	3.0E-6	1.5E-5	0
P2	100	x^0	93	4.7E-5	6.0E-6	0
P2	100	x^2	354	8.0E-6	1.3E-5	0
P2	100	x^6	308	5.1E-5	7.0E-6	0
P2	100	x^7	277	3.6E-5	9.0E-6	0
P3	4	x^0	19	3.4E-5	4.0E-6	0
P3	4	x^2	29	4.9E-5	1.0E-7	0
P3	4	x^3	45	2.4E-4	6.0E-6	0
P3	4	x^6	31	4.0E-4	1.0E-6	0
P3	4	x^7	38	7.2E-5	1.0E-7	0
P3	40	x^0	51	6.9E-5	4.0E-6	0
P3	40	x^2	185	7.0E-4	8.0E-6	0
P3	40	x^3	499	1.3E-4	6.0E-6	0
P3	40	x^4	530	7.4E-5	3.0E-6	0
P3	40	x^5	524	8.0E-6	1.0E-6	0
P3	40	x^6	168	1.8E-4	5.0E-6	0
P3	40	x^7	148	2.0E-5	1.0E-5	0
P3	100	x^0	75	1.8E-4	6.0E-6	0
P3	100	x^2	352	1.0E-4	8.0E-6	0
P3	100	x^3	758	2.1E-5	5.0E-6	0
P3	100	x^4	875	1.6E-4	5.0E-6	0
P3	100	x^5	1239	1.1E-4	7.0E-6	0
P3	100	x^6	289	1.0E-4	5.0E-6	0
P3	100	x^7	233	3.4E-4	7.0E-6	0

Table 3
Test Results for CBFGS Method (Armijo Search)

Problem	n	Init	k	$\|\nabla f_k\|$	$\|x_k - x^*\|$	off
P1	4	x^1	41	1.0E-7	1.7E-3	0
P1	4	x^2	43	1.0E-6	2.7E-3	0
P1	4	x^3	61	1.0E-7	2.4E-3	0
P1	4	x^4	43	1.0E-6	2.7E-3	0
P1	4	x^5	61	1.0E-7	2.4E-3	0
P1	4	x^6	70	1.0E-7	3.9E-4	0
P1	4	x^7	70	1.0E-7	3.9E-4	0
P2	2	x^0	21	1.8E-5	1.0E-6	0
P2	2	x^2	45	6.0E-6	2.8E-5	1
P2	2	x^3	65	2.1E-5	1.0E-6	2
P2	2	x^4	42	1.6E-5	2.0E-6	3
P2	2	x^5	68	1.4E-4	5.0E-6	1
P2	2	x^6	36	6.0E-6	1.9E-5	0
P2	2	x^7	70	1.9E-6	2.0E-6	3
P2	10	x^0	53	2.6E-5	5.0E-6	0
P2	10	x^2	118	2.0E-5	4.0E-6	0
P2	10	x^3	168	1.1E-5	1.0E-6	3
P2	10	x^4	124	1.2E-5	4.0E-6	1
P2	100	x^0	127	5.1E-5	9.0E-6	0
P2	100	x^2	364	8.0E-6	2.4E-5	0
P2	100	x^3	468	1.2E-5	1.0E-5	1
P2	100	x^4	322	1.9E-5	1.0E-5	0
P3	4	x^0	45	7.4E-5	2.0E-6	1
P3	4	x^2	68	7.9E-4	8.0E-6	1
P3	4	x^3	93	1.8E-5	1.0E-7	0
P3	4	x^4	34	4.2E-3	8.0E-6	0
P3	4	x^5	75	5.3E-5	1.0E-6	0
P3	4	x^6	49	2.8E-5	1.0E-6	0
P3	4	x^7	71	6.3E-4	4.0E-6	0
P3	40	x^0	158	3.1E-4	9.0E-6	0
P3	40	x^2	467	5.1E-5	3.0E-6	0
P3	40	x^3	293	1.8E-5	8.0E-6	0
P3	40	x^4	163	3.7E-5	9.0E-6	0
P3	40	x^5	671	1.3E-4	9.0E-6	0
P3	40	x^6	209	2.5E-5	9.0E-6	0
P3	40	x^7	288	8.5E-5	9.0E-6	0
P3	100	x^0	280	2.0E-4	1.0E-5	0
P3	100	x^2	612	2.6E-5	9.0E-6	0
P3	100	x^3	520	9.2E-5	7.0E-6	0
P3	100	x^4	303	9.7E-5	7.0E-6	0
P3	100	x^5	960	1.5E-5	9.0E-6	0
P3	100	x^6	416	6.2E-5	9.0E-6	0
P3	100	x^7	482	3.0E-5	8.0E-6	0

Table 4
Test Results for BFGS Method (Armijo Search)

Problem	n	Init	k	$\|\nabla f_k\|$	$\|x_k - x^*\|$	off
P1	4	x^1	41	1.0E-7	1.7E-3	0
P1	4	x^2	43	1.0E-6	2.7E-3	0
P1	4	x^3	61	1.0E-7	2.4E-3	0
P1	4	x^4	43	1.0E-6	2.7E-3	0
P1	4	x^5	61	1.0E-7	2.4E-3	0
P1	4	x^6	70	1.0E-7	3.9E-4	0
P1	4	x^7	70	1.0E-7	3.9E-4	0
P2	2	x^0	21	1.8E-5	1.0E-6	0
P2	2	x^2	45	6.0E-6	2.8E-5	1
P2	2	x^3	62	1.2E-4	2.0E-6	0
P2	2	x^4	43	2.0E-6	1.0E-7	2
P2	2	x^5	68	1.4E-4	5.0E-6	1
P2	2	x^6	36	6.0E-6	1.9E-5	0
P2	2	x^7	71	1.1E-5	1.0E-7	1
P2	10	x^0	53	2.6E-5	5.0E-6	0
P2	10	x^2	118	2.0E-5	4.0E-6	0
P2	10	x^3	169	7.0E-6	2.9E-5	0
P2	10	x^4	124	1.2E-5	4.0E-6	1
P2	100	x^0	127	5.1E-5	9.0E-6	0
P2	100	x^2	364	8.0E-6	2.4E-5	0
P2	100	x^3	435	1.1E-5	9.0E-6	0
P2	100	x^4	322	1.9E-5	1.0E-5	0
P3	4	x^0	45	7.4E-5	2.0E-6	1
P3	4	x^2	68	7.9E-4	8.0E-6	1
P3	4	x^3	93	1.8E-5	1.0E-7	0
P3	4	x^4	34	4.2E-3	8.0E-6	0
P3	4	x^5	75	5.3E-5	1.0E-6	0
P3	4	x^6	49	2.8E-5	1.0E-6	0
P3	4	x^7	71	6.3E-4	4.0E-6	0
P3	40	x^0	158	3.1E-4	9.0E-6	0
P3	40	x^2	467	5.1E-5	3.0E-6	0
P3	40	x^3	293	1.8E-5	8.0E-6	0
P3	40	x^4	163	3.7E-5	9.0E-6	0
P3	40	x^5	671	1.3E-4	9.0E-6	0
P3	40	x^6	209	2.5E-5	9.0E-6	0
P3	40	x^7	288	8.5E-5	9.0E-6	0
P3	100	x^0	280	2.0E-4	1.0E-5	0
P3	100	x^2	612	2.6E-5	9.0E-6	0
P3	100	x^3	520	9.2E-5	7.0E-6	0
P3	100	x^4	303	9.7E-5	7.0E-6	0
P3	100	x^5	960	1.5E-5	9.0E-6	0
P3	100	x^6	416	6.2E-5	9.0E-6	0
P3	100	x^7	482	3.0E-5	8.0E-6	0

Table 5
Various Choices of Parameters in CBFGS (Wolfe Search) for Powell Function

n	α	ϵ	k	$\|\nabla f_k\|$	$\|x_k - x^*\|$	off
4	α_1	1	36	8.0E-6	3.9E-3	11
4	α_1	10^{-1}	32	9.0E-6	7.2E-3	2
4	α_1	10^{-3}	32	9.0E-6	8.3E-3	0
4	α_1	10^{-5}	32	9.0E-6	8.3E-3	0
4	α_2	1	36	8.0E-6	3.9E-3	11
4	α_2	10^{-1}	32	9.0E-6	7.2E-3	2
4	α_2	10^{-3}	32	9.0E-6	8.3E-3	0
4	α_2	10^{-5}	32	9.0E-6	8.3E-3	0
4	α_3	1	36	8.0E-6	3.9E-3	11
4	α_3	10^{-1}	32	9.0E-6	7.2E-3	2
4	α_3	10^{-3}	32	9.0E-6	8.3E-3	0
4	α_3	10^{-5}	32	9.0E-6	8.3E-3	0
4	α_4	1	36	8.0E-6	3.9E-3	11
4	α_4	10^{-1}	32	9.0E-6	7.2E-3	2
4	α_4	10^{-3}	32	9.0E-6	8.3E-3	0
4	α_4	10^{-5}	32	9.0E-6	8.3E-3	0
4	α_5	10^{-1}	36	4.0E-6	5.4E-3	5
4	α_5	10^{-3}	32	9.0E-6	8.3E-3	0
4	α_5	10^{-5}	32	9.0E-6	8.3E-3	0
4	α_6	10^{-1}	1056	1.0E-5	1.1E-2	1043
4	α_6	10^{-3}	187	6.0E-6	4.8E-3	164
4	α_6	10^{-5}	32	9.0E-6	8.3E-3	0
4	α_7	10^{-3}	13	8.0E-6	2.4E-3	0
4	α_7	10^{-5}	2	1.0E-5	2.4E-3	0
40	α_1	10^{-1}	243	9.0E-6	1.7E-2	9
40	α_1	10^{-3}	238	9.0E-6	1.6E-2	0
40	α_1	10^{-5}	238	9.0E-6	1.6E-2	0
40	α_2	10^{-1}	243	9.0E-6	1.7E-2	9
40	α_2	10^{-3}	238	9.0E-6	1.6E-2	0
40	α_2	10^{-5}	238	9.0E-6	1.6E-2	0
40	α_3	10^{-1}	243	9.0E-6	1.7E-2	9
40	α_3	10^{-3}	238	9.0E-6	1.6E-2	0
40	α_3	10^{-5}	238	9.0E-6	1.6E-2	0
40	α_4	10^{-1}	243	9.0E-6	1.7E-2	9
40	α_4	10^{-3}	238	9.0E-6	1.6E-2	0
40	α_4	10^{-5}	238	9.0E-6	1.6E-2	0
40	α_5	10^{-3}	238	9.0E-6	1.6E-2	0
40	α_5	10^{-5}	238	9.0E-6	1.6E-2	0
40	α_7	10^{-3}	82	3.0E-6	3.1E-3	61
40	α_7	10^{-5}	1	7.0E-6	3.1E-3	0

5 Conclusion

We have proposed a cautious BFGS update and shown that the method converges globally with Wolfe-type line search or Armijo-type line search. The method retains the scale-invariance property of the original BFGS method. The established global convergence theorems do not rely on convexity assumption on the objective function. The reported numerical results show that BFGS method with the proposed cautious update performs at least as efficiently as the ordinary BFGS method. Moreover, if we choose the parameters α and ϵ appropriately, then the cautious update essentially reduces to the ordinary update. This suggests that the ordinary BFGS method is already 'cautious' in practice, and hence BFGS method has never failed empirically. We hope that the results established in this paper contribute toward resolving the fundamental open problem whether BFGS method converges for a nonconvex unconstrained optimization problem.

References

[1] R. Byrd and J. Nocedal, A tool for the analysis of quasi-Newton methods with application to unconstrained minimization, *SIAM Journal on Numerical Analysis*, 26 (1989) 727-739.

[2] R. Byrd, J. Nocedal and Y. Yuan, Global convergence of a class of quasi-Newton methods on convex problems, *SIAM Journal on Numerical Analysis*, 24 (1987) 1171-1189.

[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 and J.J. Moré, Quasi-Newton methods, motivation and theory, *SIAM Review*, 19 (1977) 46-89.

[5] L.C.W. Dixon, Variable metric algorithms: necessary and sufficient conditions for identical behavior on nonquadratic functions, *Journal of Optimization Theory and Applications*, 10 (1972) 34-40.

[6] R. Fletcher, *Practical Methods of Optimization,* Second Edition, John Wiley & Sons, Chichester, 1987.

[7] 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.

[8] A. Griewank, The global convergence of partitioned BFGS on problems with convex decompositions and Lipschitzian gradients, *Mathematical Programming*, 50 (1991) 141-175.

[9] D.H. Li and M. Fukushima, A globally and superlinearly convergent Gauss-Newton based BFGS method for symmetric nonlinear equations, to appear in *SIAM Journal on Numerical Analysis.*

[10] D.H. Li and M. Fukushima, A modified BFGS method and its global convergence in nonconvex minimization, Technical Report 98003, Department of Applied Mathematics and Physics, Kyoto University, January 1998.

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

[12] J. Nocedal, Theory of algorithms for unconstrained optimization, *Acta Numerica*, 1 (1992) 199-242.

[13] M.J.D. Powell, On the convergence of the variable metric algorithm, *Journal of the Institute of Mathematics and its Applications*, 7 (1971) 21-36.

[14] 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., SIAM, 1976, pp.53-72.

[15] Ph.L. Toint, Global convergence of the partitioned BFGS algorithm for convex partially separable optimization, *Mathematical Programming*, 36 (1986), 290-306.