A Modified BFGS Method and Its Global Convergence in Nonconvex Minimization

January 26, 1998

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

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

Abstract

In this paper, we propose a modification of the BFGS method for unconstrained optimization. A remarkable feature of the proposed method is that it possesses a global convergence property even without convexity assumption on the objective function. Under certain conditions, we also establish superlinear convergence of the method.

Key words: BFGS method, global convergence, superlinear convergence

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

1 Introduction

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

$$\min f(x), \quad x \in R^n. \tag{1.1}$$

Among numerous iterative methods for solving (1.1), quasi-Newton methods constitute particularly important class.

Throughout the paper, we assume that f in (1.1) has Lipschitz continuous gradients, i.e. there is a constant $L > 0$ such

$$\|g(x) - g(y)\| \leq L\|x - y\|, \qquad \forall x, y \in R^n, \tag{1.2}$$

where $g(x)$ denotes the gradient of f at x and $\|\cdot\|$ denotes the Euclidean norm of a vector. We will often abbreviate $g(x_k)$, $f(x_k)$, etc. as g_k, f_k, etc. respectively.

During the past two decades, global convergence of quasi-Newton methods has received growing interests. When f is convex, it was shown that Broyden's class of quasi-Newton methods converges globally and superlinearly if exact line search was used (see [12] and [6]). When inexact line search is used, Byrd, Nocedal and Yuan [3] proved the global and superlinear convergence of the convex Broyden's class with Wolfe-type line search except for DFP method. Byrd and Nocedal [2] obtained global convergence of BFGS method with backtracking line search. Zhang and Tewarson [15] obtained global convergence of the preconvex Broyden's class of methods. Li [11] proved that when f is a convex quadratic functin, DFP method still retains global convergence. There are many other studies on global convergence of quasi-Newton methods (see e.g. [10],[12], [14]). We refer to [7] for a good recent review.

We note that these studies focused on the convex problem. What will happen if quasi-Newton methods are applied to (1.1) where f is not convex? In what conditions does BFGS method converge globally and superlinearly for nonconvex minimization problems? These questions have remained open for many years (see

[8]). The purpose of this paper is to study these problems. We will show that if BFGS method is slightly modified, then under certain conditions, the modified method also converges globally and superlinearly for nonconvex unconstrained optimization problems.

We organize the paper as follows. In the next section, we give the motivation of our study and describe the modified method. In Section 3, we discuss global convergence and superlinear convergence of the modified method. In Section 4, we describe a practicable modified BFGS method and establish its global and superlinear convergence.

2 Motivation and Algorithm

In this section, we present a modified BFGS after describing our motivation.

First, we briefly review Newton's method. Suppose for the moment that f is twice differentiable. Then Newton's method for solving (1.1) takes the following iterative process. Given the k-th iterate x_k, we determine the Newton direction p_k by

$$G_k p_k + g_k = 0, \tag{2.1}$$

where $G_k = G(x_k)$ denotes the Hessian matrix of f at x_k. Once p_k is obtained, the next iterate is generated by $x_{k+1} = x_k + p_k$. Under suitable conditions, Newton's method converges locally and quadratically. To enlarge the convergence domain of Newton's method, a globalization strategy may be employed. In particular, when G_k is positive definite, the vector p_k determined by (2.1) is a descent direction of f at x_k. Thus, one can choose a step length $\lambda_k > 0$ satisfying

$$f(x_k + \lambda_k p_k) \leq f(x_k) + \sigma \lambda_k g(x_k)^T p_k, \tag{2.2}$$

where $\sigma \in (0, 1)$ is a given constant. Line search (2.2) can be fulfilled by a backtracking process of Armijo-type, i.e., $\lambda_k = \rho^{i_k}$, where $\rho \in (0, 1)$ is a given constant and i_k is the smallest nonnegative integer i for which $\lambda_k = \rho^i$ satisfies (2.2). Then

take $x_{k+1} = x_k + \lambda_k p_k$. Newton's method with line search process is called a damped Newton method. It converges globally and quadratically under some conditions.

Quasi-Newton methods were developed based on Newton's method in which G_k is substituted by some matrix B_k to avoid the calculation of a Hessian matrix. That is, Newton's direction is approximated by the so called quasi-Newton direction p_k generated by

$$B_k p_k + g_k = 0, \tag{2.3}$$

where B_k is an approximation of $G(x_k)$. Among quasi-Newton methods, BFGS method is currently regarded as the most effective method. In this method, B_k is updated by the following formula:

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

where $s_k = x_{k+1} - x_k$ and $\delta_k = g_{k+1} - g_k$. The update formula (2.4) has the property that if B_k is symmetric positive definite and $\delta_k^T s_k > 0$, then B_{k+1} is also symmetric positive definite. Therefore the quasi-Newton direction p_k generated by (2.3) is a descent direction of f at x_k no matter whether G_k is positive definite or not.

Convergence properties of BFGS methods on convex minimization problems have been well studied (see e.g. [2], [3], [12], [13], [14],[15]). Yet it is not known whether this method converges globally when it is applied to solve nonconvex minimization problems even if exact line search is used. Moreover, since $G(x)$ may not be positive definite in the case where f is nonconvex, it is doubtful that a positive definite B_k still affords a good approximation of G_k. So, it would be reasonable to expect that a proper modification of the method is effective for nonconvex problems.

Recall that in the case where f is nonconvex, the Newton direction p_k generated by (2.1) may not be a descent direction of f at x_k since G_k is not necessary positive definite. Therefore the line search (2.2) may not be well-defined. To overcome this difficulty, some modified Newton's methods have been proposed. For example, we

may generate a direction p_k from (2.1) in which G_k is replaced by the matrix

$$\bar{G}_k \triangleq G_k + r_{k-1} I, \tag{2.5}$$

where I is the unit matrix and a positive constant r_{k-1} is chosen so that $\bar{G}_k$ is positive definite. For this modified Newton's method, we have the following result.

Theorem 2.1 *Let the level set*

$$\Omega = \{x : f(x) \leq f(x_0)\}$$

be bounded and f be twice continuously differentiable on a convex set containing Ω. Assume that $\{x_k\}$ is generated by the modified Newton's method using matrices $\bar{G}_k$ given by (2.5) and Armijo line search satisfying (2.2). Suppose that $\{r_k\}$ is bounded above and that there exists an accumulation point $\bar{x}$ of $\{x_k\}$ at which $G(\bar{x})$ is positive definite. Then

(i)

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{2.6}$$

(ii) If we assume further that $\{x_k\} \to \bar{x}$, $r_k \to 0$, and $\sigma \in (0, 1/2)$, then the convergence rate is at least superlinear.

(iii) If in addition, $G(x)$ is Lipschitz around $\bar{x}$ and there exists a constant $C > 0$ such that $r_{k-1} \leq C\|g_k\|$ for all k, then the convergence rate is quadratic.

Proof Mention that the line search (2.2) is well-defined. Summing inequalities (2.2) we obtain that

$$-\sum_{k=0}^{\infty} \sigma \lambda_k g_k^T p_k \leq f(x_0) - \lim_{k\to\infty} f(x_k) < \infty,$$

where the limit exists because of the descent property of $\{f(x_k)\}$ and the boundness of Ω. In particular, $\lambda_k g_k^T p_k \to 0$. Let $\{x_k\}_{k\in K} \to \bar{x}$. Denote $\bar{\lambda} = \liminf_{k\to\infty k\in K} \lambda_k \geq 0$. If $\bar{\lambda} > 0$, then $g_k^T p_k \to 0$ as $k \in K \to \infty$ or equivalently $-g_k^T \bar{G}_k^{-1} g_k \to 0$ as $k \in K \to \infty$. Since $G(\bar{x})$ is positive definite and $r_k \geq 0$ for each k, $\{\bar{G}(x_k)\}_{k\in K}$

is uniformly positive definite when k is sufficiently large. So, we get $\|g(\bar{x})\| = \lim_{k\in K, k\to\infty} \|g_k\| = 0$.

If $\bar{\lambda} = 0$, without loss of generality, we assume that $\lim_{k\in K,k\to\infty} \lambda_k = 0$ and $\bar{G}_k \to \bar{G}(\bar{x})$ as $k \in K \to \infty$, where the later particularly follows from the boundness of $\{r_k\}$. By the line search criterion, when $k \in K$ is sufficiently large, $\lambda'_k \triangleq \lambda_k/\rho$ does not satisfy (2.2). In other words, we have

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

Dividing by λ'_k and taking limits in both sides as $k \in K \to \infty$ we get

$$g(\bar{x})^T \bar{p} \geq \sigma g(\bar{x})^T \bar{p}, \tag{2.7}$$

where $\bar{p} = -\bar{G}(\bar{x})^{-1} g(\bar{x})$. As $G(\bar{x})$ is positive definite and r_k is nonnegative, it is clear that both $\bar{G}(\bar{x})$ and $\bar{G}(\bar{x})^{-1}$ is positive definite. Since $\sigma \in (0,1)$, (2.7) implies that $g(\bar{x})^T \bar{p} = -g(\bar{x})^T \bar{G}(\bar{x})^{-1} g(\bar{x}) = 0$. So, $g(\bar{x}) = 0$, i.e., (2.6) holds.

(ii) By assumption, we have $p_k = -\bar{G}(x_k)^{-1} g_k \to 0$. Since $r_k \to 0$, we have

$$\begin{aligned}
& f(x_k + p_k) - f(x_k) - \sigma g_k^T p_k \\
&= (1-\sigma) g_k^T p_k + \frac{1}{2} p_k^T G(x_k) p_k + o(\|p_k\|^2) \\
&= -(1-\sigma) p_k^T \bar{G}(x_k) p_k + \frac{1}{2} p_k^T G(x_k) p_k + o(\|p_k\|^2) \\
&= -(1-\sigma-\frac{1}{2}) p_k^T \bar{G}(x_k) p_k - \frac{1}{2} p_k^T (\bar{G}(x_k) - G(x_k)) p_k + o(\|p_k\|^2) \\
&= -(\frac{1}{2} - \sigma) p_k^T \bar{G}(x_k) p_k + o(\|p_k\|^2) \\
&= -(\frac{1}{2} - \sigma) p_k^T \bar{G}(x_k) p_k + o(\|p_k\|^2),
\end{aligned}$$

where the last two equalities follow from the assumption $r_k \to 0$. Since $\sigma \in (0, 1/2)$ and $\bar{G}(x_k)$ is positive definite, it follows that when k is sufficiently large, the unit stepsize $\lambda_k \equiv 1$ is accepted by the line search criterion. Thus $x_{k+1} = x_k + p_k$ for all k large enough. Moreover, we have

$$0 = \bar{G}_k p_k + g_k$$

$$
\begin{aligned}
&= \bar{G}_k(x_k + p_k - \bar{x}) + g_k - \bar{G}_k(x_k - \bar{x}) \\
&= \bar{G}_k(x_{k+1} - \bar{x}) + g(x_k) - g(\bar{x}) - G_k(x_k - \bar{x}) - r_{k-1}(x_k - \bar{x}) \\
&= \bar{G}_k(x_{k+1} - \bar{x}) + [\int_0^1 G(\bar{x} + t(x_k - \bar{x}))dt - G_k](x_k - \bar{x}) - r_{k-1}(x_k - \bar{x}).
\end{aligned}
$$

That is

$$
(x_{k+1} - \bar{x}) = -\bar{G}_k^{-1}[\int_0^1 G(\bar{x} + t(x_k - \bar{x}))dt - G_k](x_k - \bar{x}) - r_{k-1}\bar{G}_k^{-1}(x_k - \bar{x}).
$$

Since $G(\bar{x})$ is positive definite and $r_k \geq 0$, $\bar{G}_k$ is uniformly positive definite for sufficiently large k. Thus, we deduce from above equality that there exists a positive constant M such that

$$
\|x_{k+1} - \bar{x}\| \leq M\{\int_0^1 \|G(\bar{x} + t(x_k - \bar{x})) - G_k\|dt + r_{k-1}\}\|x_k - \bar{x}\| \tag{2.8}
$$

which proves (ii)

(iii) From (2.8) we have that

$$
\begin{aligned}
\|x_{k+1} - \bar{x}\| &\leq M\{\int_0^1 \|G(\bar{x} + t(x_k - \bar{x})) - G_k\|dt + r_{k-1}\}\|x_k - \bar{x}\|\} \\
&\leq M\{\int_0^1 \|G(\bar{x} + t(x_k - \bar{x})) - G_k\|dt + C\|g_k\|\}\|x_k - \bar{x}\| \\
&= M\{\int_0^1 \|G(\bar{x} + t(x_k - \bar{x})) - G_k\|dt + C\|g(x_k) - g(\bar{x})\|\}\|x_k - \bar{x}\| \\
&\leq M(L_1 + CL)\|x_k - \bar{x}\|^2,
\end{aligned}
$$

where $L > 0$ and $L_1 > 0$ are Lipschitz constants of g and G respectively. □

The above theorem shows that if we choose r_k in a suitable way, then the modified damped Newton's method still retains fast local convergence as well as global convergence. This fact prompts us to consider the question: Is it possible to modify quasi-Newton methods in such a way that B_k approximates $\bar{G}_k$, thereby ensuring global and superlinear convergence without convex assumption on the problem?

First observe that since G_{k+1} satisfies

$$
G_{k+1}(x_{k+1} - x_k) \approx g_{k+1} - g_k,
$$

the matrix $\bar{G}_{k+1}$ will satisfy the relation

$$\bar{G}_{k+1}(x_{k+1} - x_k) = (G_{k+1} + r_k I)(x_{k+1} - x_k) \approx (g_{k+1} - g_k) + r_k(x_{k+1} - x_k),$$

or equivalently

$$\bar{G}_{k+1} s_k \approx \delta_k + r_k s_k, \tag{2.9}$$

where $\delta_k = g_{k+1} - g_k$. Therefore, it would be reasonable to require the matrix B_{k+1} to satisfy (2.9) exactly, i.e.,

$$B_{k+1} s_k = y_k, \tag{2.10}$$

where $y_k = \delta_k + r_k s_k$. If r_k is small, (2.10) can be regarded as an approximation of the ordinary secant equation $B_{k+1} s_k = \delta_k$.

Now, we propose a modified BFGS method based on the above consideration.

Algorithm 1 (Modified BFGS Method : MBFGS)

Initial Choose an initial point $x_0 \in R^n$ and an initial positive definite matrix B_0. Let $k := 0$.

Step 1 Solve the following linear equations to get p_k:

$$B_k p + g_k = 0.$$

Step 2 Find a stepsize $\lambda_k > 0$ satisfying the Wolfe type line search conditions:

$$\begin{cases} f(x_k + \lambda_k p_k) \le f(x_k) + \sigma_1 \lambda_k g_k^T p_k, \\ g(x_k + \lambda_k p_k)^T p_k \ge \sigma_2 g_k^T p_k, \end{cases} \tag{2.11}$$

where σ_1 and σ_2 are constants satisfying $0 < \sigma_1 < \sigma_2 < 1$. Moreover, if $\lambda_k = 1$ satisfies (2.11), we take $\lambda_k = 1$.

Step 3 Let the next iterate be $x_{k+1} = x_k + \lambda_k p_k$.

Step 4 Let $s_k = x_{k+1} - x_k = \lambda_k p_k$, $\delta_k = g_{k+1} - g_k$, $y_k = \delta_k + r_k s_k$ and $\eta_k = \dfrac{\delta_k^T s_k}{\|s_k\|^2}$. Choose any $r_k \ge 0$ satisfying

$$\epsilon \le \eta_k + r_k \le r, \tag{2.12}$$

where ϵ and r are positive constants.

Step 5 Update B_k using the formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \tag{2.13}$$

Step 6 Let $k := k + 1$ and go to step 1.

3 General Convergence Analysis

We beginin with the global convergence analysis of MBFGS method. It is easy to see that if λ_k satisfies (2.11), then B_{k+1} is symmetric positive definite provided that B_k is symmetric positive definite. We also notice that the Lipschitz continuouity of g ensures that

$$|\eta_k| = \frac{|\delta_k^T s_k|}{\|s_k\|^2} = \frac{\|\delta_k\|}{\|s_k\|} \leq L.$$

In the later of the paper, we always assume that the level set

$$\Omega = \{x \mid f(x) \leq f(x_0)\}$$

is bounded.

It is clear from the first inequality of (2.11) that $\{f(x_k)\}$ is a nonincreasing sequence, which ensures that $\lim_{k\to\infty} f(x_k)$ exists and $\{x_k\} \subset \Omega$.

To establish the global convergence of MBFGS method method, we first prove some useful lemmas.

Lemma 3.1 *Let $\{x_k\}$ be generated by MBFGS method. Then we have*

$$\sum_{k=0}^{\infty} (-g_k^T s_k) < \infty \tag{3.1}$$

$$\epsilon \|s_k\|^2 \leq y_k^T s_k \leq r \|s_k\|^2, \quad k = 1, 2, \ldots, \tag{3.2}$$

$$\|y_k\| \leq (L + r)\|s_k\|, \quad k = 1, 2, \ldots, \tag{3.3}$$

where ϵ and r is specified in the algorithm and L is the Lipschitz constant of g given by (1.2).

Proof (3.1) can be obtained by summing the first inequality of (2.11) directly. (3.2) and (3.3) follow from (2.12) and the Lipschitz continuity of g. □

The following two lemmas are similar to those in [3].

Lemma 3.2 *There is a constant $M_1 > 0$ such that*

$$\text{tr } B_{k+1} \leq M_1(k+1) \quad \textit{and} \quad \sum_{i=0}^{k} \frac{\|B_i s_i\|^2}{s_i^T B_i s_i} \leq M_1(k+1). \tag{3.4}$$

Proof By Lemma 3.1, taking the trace operation in both sides of (2.13) yields

$$\begin{aligned} \text{tr } B_{k+1} &= \text{tr } B_k - \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \frac{\|y_k\|^2}{y_k^T s_k} \\ &\leq \text{tr } B_k - \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \frac{(L+r)^2}{\epsilon} \\ &\leq \vdots \\ &\leq \text{tr } B_0 - \sum_{i=0}^{k} \frac{\|B_i s_i\|^2}{s_i^T B_i s_i} + \frac{(L+r)^2}{\epsilon}(k+1). \end{aligned}$$

Since B_{k+1} is positive definite, tr $B_{k+1} > 0$. The last inequality implies (3.4). □

Lemma 3.3 *There is a constant $c_1 > 0$ such that for sufficiently large k,*

$$\prod_{i=0}^{k} \lambda_i \geq c_1^k. \tag{3.5}$$

Proof By means of the formula (see e.g., Lemma 7.6 of [5])

$$\det(I + u_1 u_2^T + u_3 u_4^T) = (1 + u_1^T u_2)(1 + u_3^T u_4) - (u_1^T u_4)(u_2^T u_3),$$

taking determinant in (2.13) we get

$$\begin{aligned} \det B_{k+1} &= \det\{B_k(I - \frac{s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{B_k^{-1} y_k y_k^T}{y_k^T s_k})\} \\ &= \det B_k \det(I - s_k \frac{(B_k s_k)^T}{s_k^T B_k s_k} + B_k^{-1} y_k \frac{y_k^T}{y_k^T s_k}) \end{aligned}$$

$$
\begin{aligned}
&= \det B_k[(1 - s_k^T \frac{B_k s_k}{s_k^T B_k s_k})(1 + (B_k^{-1} y_k)^T \frac{y_k}{y_k^T s_k}) \\
&\quad -(-s_k^T \frac{y_k}{y_k^T s_k})(\frac{(B_k s_k)^T}{s_k^T B_k s_k} B_k^{-1} y_k)] \\
&= \det B_k \frac{y_k^T s_k}{s_k^T B_k s_k} \\
&\geq \det B_k \frac{\delta_k^T s_k}{s_k^T B_k s_k} \\
&\geq \det B_k \frac{-(1-\sigma_2) g_k^T s_k}{-\lambda_k g_k^T s_k} \\
&= \frac{1-\sigma_2}{\lambda_k} \det B_k \\
&\geq \vdots \\
&\geq [\prod_{i=0}^{k} \frac{1}{\lambda_k}](1-\sigma_2)^{k+1} \det B_0,
\end{aligned}
$$

where the second inequality follows from the second inequality of (2.11).

Since $\det B_{k+1} \leq [\frac{1}{n} \text{tr } B_{k+1}]^n \leq [\frac{M_2(k+1)}{n}]^n$ by Lemma 3.2, we deduce from the above inequality that

$$
\prod_{i=0}^{k} \lambda_k \geq \frac{(1-\sigma_2)^{k+1} \det B_0}{\det B_{k+1}} \geq \frac{(1-\sigma_2)^{k+1} \det B_0}{[\frac{M_2(k+1)}{n}]^n}.
$$

When k is sufficiently large, this implies (3.5). □

Now, we establish a global convergence theorem for MBFGS method. The proof is similar to the one given in [14].

Theorem 3.1 *Let $\{x_k\}$ be generated by MBFGS method with B_k being updated by (2.13). Then we have*

$$
\liminf_{k \to \infty} \|g_k\| = 0. \tag{3.6}
$$

Proof For the purpose of contradiction we assume that $\|g_k\| \geq \gamma > 0$ for all k.

Since $B_k s_k = \lambda_k B_k p_k = -\lambda_k g_k$ and hence $\lambda_k = \frac{\|B_k s_k\|}{\|g_k\|}$, it follows from (3.1)

$$
\infty \quad > \quad \sum_{k=0}^{\infty} (-g_k^T s_k)
$$

$$
\begin{aligned}
&= \sum_{k=0}^{\infty} \frac{1}{\lambda_k} s_k^T B_k s_k \\
&= \sum_{k=0}^{\infty} \frac{\|g_k\|}{\|B_k s_k\|} s_k^T B_k s_k \\
&= \sum_{k=0}^{\infty} \|g_k\|^2 \lambda_k \frac{s_k^T B_k s_k}{\|B_k s_k\|^2} \\
&\geq \gamma^2 \sum_{k=0}^{\infty} \lambda_k \frac{s_k^T B_k s_k}{\|B_k s_k\|^2}.
\end{aligned}
$$

Therefore, for any $\zeta > 0$ there exists constant $k_0 > 0$ such that for any positive integer q,

$$
q\{\prod_{k=k_0+1}^{k_0+q} \lambda_k \frac{s_k^T B_k s_k}{\|B_k s_k\|^2}\}^{\frac{1}{q}} \leq \sum_{k=k_0+1}^{k_0+q} \lambda_k \frac{s_k^T B_k s_k}{\|B_k s_k\|^2} \leq \zeta,
$$

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

$$
\begin{aligned}
(\prod_{k=k_0+1}^{k_0+q} \lambda_k)^{\frac{1}{q}} &\leq \frac{\zeta}{q}[\prod_{k=k_0+1}^{k_0+q} \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}]^{\frac{1}{q}} \\
&\leq \frac{\zeta}{q^2} \sum_{k=k_0+1}^{k_0+q} \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} \\
&\leq \frac{\zeta}{q^2} \sum_{k=0}^{k_0+q} \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} \\
&\leq \frac{\zeta(k_0+q+1)}{q^2} M_1.
\end{aligned}
$$

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. Thus, we get (3.6). □

The above theorem shows the global convergence property of MBFGS method without convex assumption on f. It only relies on the assumption that f has Lipschitz continuous gradients.

Now, we turn to prove the locally superlinear convergence property of MBFGS method. To do this, we need the following set of assumptions.

Assumption (A) (1) f is twice continuously differentiable near x^*.

(2) $\{x_k\} \to x^*$ at which $g(x^*) = 0$ and $G(x^*)$ is positive definite.

(3) G is Hölder continuous at x^*, i.e., there exist constants $\nu \in (0,1)$ and $M_2 > 0$ such that

$$\|G(x) - G(x^*)\| \leq M_2 \|x - x^*\|^\nu \tag{3.7}$$

for all x in a neighborhood of x^*.

(4) $\{r_k\}$ satisfies

$$\sum_{i=0}^{\infty} r_k < \infty. \tag{3.8}$$

Under conditions (1) and (2) in Assumption (A), there is a neighbourhood $U(x^*)$ of x^* such that $G(x)$ is uniformly positive definite for all $x \in U(x^*)$. Therefore, there is a constant $m > 0$ such that for all $d \in R^n$ and $x \in U(x^*)$

$$\|g(x)\| = \|g(x) - g(x^*)\| \geq m\|x - x^*\| \tag{3.9}$$

and

$$d^T G(x) d \geq m\|d\|^2. \tag{3.10}$$

In particular, (3.9) and (3.10) hold with $x = x_k$ for all k sufficiently large.

Let ϕ_k denote the angle between s_k and $B_k s_k$, i.e.,

$$\cos \phi_k = \frac{s_k^T B_k s_k}{\|s_k\| \|B_k s_k\|} = -\frac{g_k^T s_k}{\|g_k\| \|s_k\|}.$$

We will establish superlinear convergence of MBFGS method after proving several lemmas similar to those in [3].

Lemma 3.4 *Under Assumption (A), there are constants $a_1 > 0, a_1' > 0$ and $a_2 > 0$ and an index k' such that when $k \geq k'$,*

$$a_1' \|g_k\| \cos \phi_k \leq \|s_k\| \leq a_1 \|g_k\| \cos \phi_k, \tag{3.11}$$

$$\prod_{i=0}^{k} \cos^2 \phi_i \geq a_2^{k+1}. \tag{3.12}$$

Proof From (2.11) and Taylor's expansion with (3.10), there is an index k_1 such that when $k \geq k_1$

$$\sigma_1 g_k^T s_k \geq f_{k+1} - f_k \geq g_k^T s_k + \frac{m}{2}\|s_k\|^2,$$

which implies that

$$\|s_k\|^2 \leq -\frac{2(1-\sigma_1)}{m} g_k^T s_k = \frac{2(1-\sigma_1)}{m}\|g_k\|\|s_k\|\cos\phi_k.$$

Letting $a_1 = \frac{2(1-\sigma_1)}{m}$ we get the right-hand inequality of (3.11).

Again by (2.11) and (3.2) we have that

$$r\|s_k\|^2 \geq y_k^T s_k \geq \delta_k^T s_k \geq -(1-\sigma_2) g_k^T s_k = (1-\sigma_2)\|g_k\|\|s_k\|\cos\phi_k.$$

This yields the left-hand inequality of (3.11).

Since $\|B_k s_k\| = \lambda_k \|g_k\|$, (3.11) implies that when $k \geq k_1$

$$\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} = \frac{\|B_k s_k\|}{\|s_k\|\cos\phi_k} = \lambda_k \frac{\|g_k\|}{\|s_k\|\cos\phi_k} \geq \frac{\lambda_k}{a_1 \cos^2\phi_k}.$$

Multiplying these inequalities from k_1 to $k > k_1$ and using the geometric inequality we get

$$\begin{aligned}
\{\prod_{i=k_1}^{k} \frac{\lambda_i}{a_1\cos^2\phi_i}\}^{\frac{1}{k-k_1+1}} &\leq \{\prod_{i=k_1}^{k} \frac{\|B_i s_i\|^2}{s_i^T B_i s_i}\}^{\frac{1}{k-k_1+1}} \\
&\leq \frac{1}{k-k_1+1}\sum_{i=k_1}^{k} \frac{\|B_i s_i\|^2}{s_i^T B_i s_i} \\
&\leq \frac{1}{k-k_1+1}\sum_{i=0}^{k} \frac{\|B_i s_i\|^2}{s_i^T B_i s_i} \\
&\leq M_1 \frac{k+1}{k-k_1+1},
\end{aligned}$$

where the last inequality follows from (3.4). Therefore we obtain for any k large enough

$$\prod_{i=k_1}^{k} \cos^2\phi_i \geq [\prod_{i=k_1}^{k} \lambda_i] a_1^{-(k-k_1+1)} M_1^{-(k-k_1+1)} [\frac{k-k_1+1}{k+1}]^{k-k_1+1}$$

$$
\begin{aligned}
&= [\prod_{i=0}^{k} \lambda_i] a_1^{-(k-k_1+1)} M_1^{-(k-k_1+1)} [\frac{k-k_1+1}{k+1}]^{k-k_1+1} \prod_{i=0}^{k_1-1} \lambda_i^{-1} \\
&\geq c_1^k (a_1^{-1} M_1^{-1} \frac{k-k_1+1}{k+1})^{k-k_1+1} \prod_{i=0}^{k_1-1} \lambda_i^{-1},
\end{aligned}
$$

where the last inequality follows from (3.5). Since k_1 is a fixed index, the above inequality implies that when k is sufficiently large, say $k \geq k'$, (3.12) holds. □

Lemma 3.5 *Under Assumption (A), we have*

$$\sum_{k=0}^{\infty} \|x_{k+1} - x^*\|^{\nu} < \infty. \tag{3.13}$$

Moreover, if denote $\tau_k = \max\{\|x_k - x^*\|^{\nu}, \|x_{k+1} - x^*\|^{\nu}\}$, *then*

$$\sum_{k=0}^{\infty} \tau_k < \infty. \tag{3.14}$$

Proof Denote $f_* = f(x^*)$. First we show that there is a constant $a_3 > 0$ such that

$$f_{k+1} - f_* \leq (1 - a_3 \cos^2 \phi_k)(f_k - f_*) \tag{3.15}$$

for all k sufficiently large. Indeed, (2.11) implies

$$f_{k+1} - f_* \leq (f_k - f_*) + \sigma_1 g_k^T s_k.$$

To show (3.15), it therefore suffices to verify that there is a constant $a_4 > 0$ such that when k is sufficiently large

$$g_k^T s_k \leq -a_4 \cos^2 \phi_k (f_k - f_*). \tag{3.16}$$

Since x^* satisfies $g(x^*) = 0$, by Taylor's expansion, there is a constant $M_3 > 0$ such that $f_k - f_* \leq M_3 \|x_k - x^*\|^2$. Thus from (3.9) and (3.11) we deduce that when k is sufficiently large

$$
\begin{aligned}
g_k^T s_k &= -\|g_k\| \|s_k\| \cos \phi_k \leq -a_1' \|g_k\|^2 \cos^2 \phi_k \\
&\leq -a_1' m^2 \|x_k - x^*\|^2 \cos^2 \phi_k \leq -a_1' \frac{m^2}{M_3} \cos^2 \phi_k (f_k - f_*).
\end{aligned}
$$

This inequality means that (3.16) holds. So we get (3.15).

Now we prove (3.13). From Lemma 3.2 and the above discussion, we see that there exists an index $k_1' > 0$ such that when $k \geq k_1'$, (3.12) and (3.15) hold. Therefore

$$\begin{aligned}
f_{k+1} - f_* &\leq (1 - a_3 \cos^2 \phi_k)(f_k - f_*) \\
&\leq \vdots \\
&\leq (f_{k_1'} - f_*) \prod_{i=k_1'}^{k} (1 - a_3 \cos^2 \phi_i) \\
&\leq (f_{k_1'} - f_*)[\frac{1}{k - k_1' + 1} \sum_{i=k_1'}^{k} (1 - a_3 \cos^2 \phi_i)]^{k-k_1'+1} \\
&= (f_{k_1'} - f_*)[1 - a_3 \frac{1}{k - k_1' + 1} \sum_{i=k_1'}^{k} \cos^2 \phi_i]^{k-k_1'+1} \\
&\leq (f_{k_1'} - f_*)[1 - a_3 (\prod_{i=k_1'}^{k} \cos^2 \phi_i)^{\frac{1}{k-k_1'+1}}]^{k-k_1'+1} \\
&= (f_{k_1'} - f_*)[1 - a_3 (\prod_{i=0}^{k} \cos^2 \phi_i)^{\frac{1}{k-k_1'+1}} (\prod_{i=0}^{k_1'-1} \cos^2 \phi_i)^{-\frac{1}{k-k_1'+1}}]^{k-k_1'+1} \\
&\leq (f_{k_1'} - f_*) \left(1 - a_3 a_2^{\frac{k+1}{k-k_1'+1}} (\prod_{i=0}^{k_1'-1} \cos^2 \phi_i)^{-\frac{1}{k-k_1'+1}}\right)^{k-k_1'+1},
\end{aligned}$$

where the last inequality follows from (3.12) in Lemma 3.2. Since k_1' is a constant, the second term in the circle bracket goes to a positive constant as k goes to infinity. Therefore, the above inequality shows that there is a constant $\rho_1 \in (0, 1)$ such that when k is sufficiently large, $f_{k+1} - f_* \leq (f_{k_1'} - f_*)\rho_1^{k-k_1'+1}$. Again by Taylor's expansion and (3.10) we have $f_{k+1} - f_* \geq \frac{m}{2}\|x_{k+1} - x^*\|^2$ for all sufficiently large k. Therefore, we get $\|x_{k+1} - x^*\| \leq [\frac{2}{m}(f_{k+1} - f_*)]^{\frac{1}{2}} \sqrt{\rho_1}^{k-k'+1} = \sqrt{\rho_1}^{k-k'+1} [\frac{2}{m}(f_{k+1} - f_*)]^{\frac{1}{2}} \sqrt{\rho_1}^{k} \triangleq a_5 \sqrt{\rho_1}^{k}$. This implies (3.13).

At last, notice that $\tau_k \leq \|x_k - x^*\|^\nu + \|x_{k+1} - x^*\|^\nu$, (3.14) follows from (3.13) directly. □

Lemma 3.6 *Let Assumption (A) hold. Denote* $Q = G(x^*)^{-1/2}$, $H_k = B_k^{-1}$ $H_{k+1} =$

B_{k+1}^{-1}. *Then there are positive constants* $b_i, i = 1, 2, \cdots, 7$ *and* $\alpha \in (0, 1)$ *such that for all* k *sufficiently large*

$$\|B_{k+1} - G(x^*)\|_{Q,F} \leq (1 + b_1\tau_k)\|B_k - G(x^*)\|_{Q,F} + b_2\tau_k + b_3 r_k. \tag{3.17}$$

and

$$\begin{aligned} &\|H_{k+1} - G(x^*)^{-1}\|_{Q^{-1},F} \\ &\quad \leq \quad (\sqrt{1 - \alpha\mu_k^2} + b_4\tau_k + b_5 r_k)\|H_k - G(x^*)^{-1}\|_{Q^{-1},F} + b_6\tau_k + b_7 r_k, \end{aligned} \tag{3.18}$$

where τ_k *is defined by Lemma 3.5.* $\|A\|_{Q,F} = \|Q^T A Q\|_F$, $\|\cdot\|_F$ *is the Frobenius norm of a matrix and*

$$\mu_k = \frac{\|Q^{-1}[H_k - G(x^*)^{-1}]y_k\|}{\|H_k - G(x^*)^{-1}\|_{Q^{-1},F}\|Qy_k\|}. \tag{3.19}$$

In particular, $\{\|B_k\|\}$ *and* $\{\|H_k\|\}$ *are bounded.*

Proof From the update formula (2.13) we get

$$\begin{aligned} &\|B_{k+1} - G(x^*)\|_{Q,F} \\ &\quad = \quad \|B_k - G(x^*) - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}\|_{Q,F} \\ &\quad \leq \quad \|B_k - G(x^*) - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{\delta_k \delta_k^T}{\delta_k^T s_k}\|_{Q,F} + \|\frac{y_k y_k^T}{y_k^T s_k} - \frac{\delta_k \delta_k^T}{\delta_k^T s_k}\|_{Q,F} \\ &\quad \leq \quad (1 + b_1\tau_k)\|B_k - G(x^*)\|_{Q,F} + b_2\tau_k + \|\frac{y_k y_k^T}{y_k^T s_k} - \frac{\delta_k \delta_k^T}{\delta_k^T s_k}\|_{Q,F} \end{aligned} \tag{3.20}$$

where the last inequality follows from the inequality (49) of [9]. Moreover,

$$\begin{aligned} &\|\frac{y_k y_k^T}{y_k^T s_k} - \frac{\delta_k \delta_k^T}{\delta_k^T s_k}\|_{Q,F} \\ &\quad = \quad \|\frac{(\delta_k + r_k s_k)(\delta_k + r_k s_k)^T}{(\delta_k + r_k s_k)^T s_k} - \frac{\delta_k \delta_k^T}{\delta_k^T s_k}\|_{Q,F} \\ &\quad = \quad \|\frac{\delta_k^T s_k (\delta_k + r_k s_k)(\delta_k + r_k s_k)^T - (\delta_k + r_k s_k)^T s_k \cdot \delta_k \delta_k^T}{(\delta_k + r_k s_k)^T s_k (\delta_k^T s_k)}\|_{Q,F} \end{aligned}$$

$$
\begin{aligned}
&= r_k \| \frac{\delta_k^T s_k (\delta_k s_k^T + s_k \delta_k^T + r_k s_k s_k^T) - \|s_k\|^2 \delta_k \delta_k^T}{(\eta_k + r_k)\|s_k\|^2 (\delta_k^T s_k)} \|_{Q,F} \\
&\leq r_k \|Q\|_F^2 \| \frac{\delta_k^T s_k (\delta_k s_k^T + s_k \delta_k^T + r_k s_k s_k^T) - \|s_k\|^2 \delta_k \delta_k^T}{(\eta_k + r_k)\|s_k\|^2 (\delta_k^T s_k)} \|_F \\
&\leq r_k \|Q\|_F^2 \frac{\delta_k^T s_k (\|\delta_k s_k^T\|_F + \|s_k \delta_k^T\|_F + r_k \|s_k s_k^T\|_F) + \|s_k\|^2 \|\delta_k \delta_k^T\|_F}{(\eta_k + r_k)\|s_k\|^2 (\delta_k^T s_k)} \\
&\leq r_k \frac{\|\delta_k\| \|s_k\| (\|\delta_k\| \|s_k\| + \|s_k\| \|\delta_k\| + r_k \|s_k\|^2) + \|s_k\|^2 \|\delta_k\|^2}{(\eta_k + r_k)\|s_k\|^2 (\delta_k^T s_k)} \|Q\|_F^2 \\
&= r_k \frac{3\|\delta_k\|^2 + r_k \|\delta_k\| \|s_k\|}{(\eta_k + r_k)(\delta_k^T s_k)} \|Q\|_F^2. \qquad (3.21)
\end{aligned}
$$

Again, by means of the positive definiteness of $G(x^*)$, we see from (3.10) that when k is sufficiently large

$$
\delta_k^T s_k = (g(x_{k+1}) - g(x_k))^T (x_{k+1} - x_k) \geq m \|x_{k+1} - x_k\|^2 = m \|s_k\|^2.
$$

Notice that $r_k \to 0$ by (3.8), $\eta_k + r_k \geq \epsilon$ by (2.12) and $\|\delta_k\| = \|g(x_{k+1}) - g(x_k)\| \leq L\|s_k\|$ by (1.2), it follows from (3.21) that there is a positive constant c_3 such that for all k sufficiently large

$$
\| \frac{y_k y_k^T}{y_k^T s_k} - \frac{\delta_k \delta_k^T}{\delta_k^T s_k} \|_{Q,F} \leq c_3 r_k.
$$

Combining this and (3.20) we get (3.17).

Now, we prove (3.18). We will use Lemma 3.1 of [4] to do this. First, it is well-known that the inverse update formula of (2.13) is

$$
\begin{aligned}
H_{k+1} &= H_k + \frac{(s_k - H_k y_k) s_k^T + s_k (s_k - H_k y_k)^T}{y_k^T s_k} - \frac{y_k^T (s_k - H_k y_k) s_k s_k^T}{(y_k^T s_k)^2} \\
&= (I - \frac{s_k y_k^T}{y_k^T s_k}) H_k (I - \frac{y_k s_k^T}{y_k^T s_k}) + \frac{s_k s_k^T}{y_k^T s_k}.
\end{aligned}
$$

It is the dual form of DFP update formula in the sense that $H_k \leftrightarrow B_k$, $H_{k+1} \leftrightarrow B_{k+1}$ and $s_k \leftrightarrow y_k$. We also have

$$
\|Q y_k - Q^{-1} s_k\| \leq \|Q\| \|y_k - Q^{-2} s_k\|
$$

$$
\begin{aligned}
&= \|Q\|\|y_k - G(x^*)s_k\| \\
&= \|Q\|\|\delta_k - G(x^*)s_k + r_k s_k\| \\
&\leq \|Q\|[\| \int_0^1 G(x_k + \tau s_k)s_k d\tau - G(x^*)s_k\| + r_k\|s_k\|] \\
&\leq \|Q\|[\int_0^1 \|G(x_k + \tau s_k) - G(x^*)\|d\tau\|s_k\| + r_k\|s_k\|] \\
&\leq \|Q\|[M_2 \int_0^1 \|x_k - x^* + \tau s_k\|^\nu d\tau + r_k]\|s_k\| \\
&\leq \|Q\|[M_2 \int_0^1 (\tau\|x_{k+1} - x^*\| + (1-\tau)\|x_k - x^*\|)^\nu d\tau + r_k]\|s_k\| \\
&\leq \|Q\|(M_2\tau_k + r_k)\|s_k\|, \qquad (3.22)
\end{aligned}
$$

where M_2 is the Hölder constant for G in (3.7). Since $\tau_k \to 0$ by (3.14) and $r_k \to 0$ by (3.8), it is clear that when k is sufficiently large, $\|Qy_k - Q^{-1}s_k\| \leq \beta\|Q^{-1}s_k\|$ for some constant $\beta \in (0, \frac{1}{3})$. Therefore, from Lemma 3.1 of [4] (with the identification $s \leftrightarrow y_k$, $y \leftrightarrow s_k$, $B \leftrightarrow H_k$, $A \leftrightarrow G(x^*)^{-1}$ and $M \leftrightarrow Q^{-1}$) (see also e.g. [1]), there are constants $\alpha \in (0,1)$ and $b_4', b_5' > 0$, such that

$$
\begin{aligned}
\|H_{k+1} - G(x^*)^{-1}\|_{Q^{-1},F} \leq\ & (\sqrt{1 - \alpha\mu_k^2} + b_4'\frac{\|Q^{-1}s_k - Qy_k\|}{\|Qy_k\|})\|H_k - G(x^*)\|_{Q^{-1},F} \\
& + b_5'\frac{\|s_k - G(x^*)^{-1}y_k\|}{\|Qy_k\|}, \qquad (3.23)
\end{aligned}
$$

where μ_k is defined by (3.19). Moreover, there exists a constant $b_6' > 0$ such that for all k large enough,

$$
\begin{aligned}
\|Qy_k\| &= \|Q(g(x_{k+1}) - g(x_k) + r_k s_k)\| \\
&\geq \|Q(g(x_{k+1}) - g(x_k))\| - r_k\|Qs_k\| \\
&\geq b_6'\|x_{k+1} - x_k\| - r_k\|Q\|\|s_k\| \\
&= (b_6' - r_k\|Q\|)\|s_k\|,
\end{aligned}
$$

where the second inequality follows from the positive definiteness of $G(x^*)$ and the fact that $x_k \to x^*$. Since $r_k \to 0$, the above inequality implies that there is a constant $c' > 0$ such that when k is sufficiently large, $\|Qy_k\| \geq c'\|s_k\|$. So, we get

from (3.22)

$$\frac{\|Q^{-1}s_k - Qy_k\|}{\|Qy_k\|} \leq (c')^{-1}\|Q\|(M_2\tau_k + r_k) \tag{3.24}$$

and

$$\begin{aligned}
\frac{\|s_k - G(x^*)^{-1}y_k\|}{\|Qy_k\|} &= \frac{\|s_k - Q^2y_k\|}{\|Qy_k\|} \\
&= \frac{\|Q(Qy_k - Q^{-1}s_k)\|}{\|Qy_k\|} \\
&\leq \frac{\|Q\|\|Qy_k - Q^{-1}s_k\|}{c'\|s_k\|} \\
&\leq (c')^{-1}\|Q\|^2(K\tau_k + r_k).
\end{aligned}$$

Combining this, (3.24) and (3.22) with (3.23), we get (3.18).

Finally, since $\sum_{k=0}^{\infty}\tau_k < \infty$ and $\sum_{k=0}^{\infty} r_k < \infty$, by means of Lemma 3.1 of [4], we see from (3.17) and (3.18) that both $\|B_k - G(x^*)\|_{Q,F}$ and $\|H_k - G(x^*)^{-1}\|_{Q^{-1},F}$ converge. In particular, $\{\|B_k\|_F\}$ and $\{\|H_k\|_F\}$ are bounded. □

Now we establish superlinear convergence of MBFGS method.

Theorem 3.2 *Let Assumption (A) hold. If in (2.11), we choose σ_1 to satisfy $\sigma_1 \in (0, \frac{1}{2})$, then $\{x_k\}$ converges to x^* superlinearly.*

Proof We first show the Dennis-Moré condition

$$\lim_{k\to\infty} \frac{\|(B_k - G(x^*))s_k\|}{\|s_k\|} = 0. \tag{3.25}$$

Since $\sqrt{1-t} \leq 1 - \frac{1}{2}t \ \forall t \in (0,1)$, $\|s_k\| \to 0$, $\tau_k \to 0$ and $\{\|H_k\|\}$ is bounded, it follows from (3.18) that there are positive constants M_4 and M_5 such that for all k sufficiently large

$$\|H_{k+1} - G(x^*)^{-1}\|_{Q^{-1},F} \leq (1 - \frac{1}{2}\alpha\mu_k^2)\|H_k - G(x^*)^{-1}\|_{Q^{-1},F} + M_4\tau_k + M_5 r_k,$$

i.e.

$$\begin{aligned}
&\frac{1}{2}\alpha\mu_k^2\|H_k - G(x^*)^{-1}\|_{Q^{-1},F} \\
&\quad\leq \|H_k - G(x^*)^{-1}\|_{Q^{-1},F} - \|H_{k+1} - G(x^*)^{-1}\|_{Q^{-1},F} + M_4\nu_k + M_5 r_k.
\end{aligned}$$

Summing these inequalities we get

$$\frac{1}{2}\alpha \sum_{k=\tilde{k}}^{\infty} \mu_k^2 \|H_k - G(x^*)^{-1}\|_{Q^{-1},F} < \infty,$$

where $\tilde{k}$ is a sufficiently large index such that (3.18) holds for all $k \geq \tilde{k}$. In particular, we have $\lim_{k\to\infty} \mu_k^2 \|H_k - G(x^*)^{-1}\|_{Q^{-1},F} = 0$, i.e.

$$\lim_{k\to\infty} \frac{\|Q^{-1}(H_k - G(x^*)^{-1})y_k\|^2}{\|H_k - G(x^*)^{-1}\|_{Q^{-1},F}\|Qy_k\|^2} = 0.$$

However, since $\|H_k - G(x^*)^{-1}\|_{Q^{-1},F}$ is bounded, the above inequality implies that

$$\lim_{k\to\infty} \frac{\|Q^{-1}(H_k - G(x^*)^{-1})y_k\|}{\|Qy_k\|} = 0. \tag{3.26}$$

Moreover we have

$$\begin{aligned}
&\|Q^{-1}(H_k - G(x^*)^{-1})y_k\| \\
&\quad = \|Q^{-1}H_k(G(x^*) - B_k)G(x^*)^{-1}y_k\| \\
&\quad \geq \|Q^{-1}H_k(G(x^*) - B_k)s_k\| - \|Q^{-1}H_k(G(x^*) - B_k)(s_k - G(x^*)^{-1}y_k)\| \\
&\quad \geq \|Q^{-1}H_k(G(x^*) - B_k)s_k\| - \|Q^{-1}H_k(G(x^*) - B_k)(s_k - G(x^*)^{-1}\delta_k)\| \\
&\qquad -r_k\|Q^{-1}H_k(G(x^*) - B_k)G(x^*)^{-1}s_k\|.
\end{aligned} \tag{3.27}$$

Note that

$$\begin{aligned}
&\|Q^{-1}H_k(G(x^*) - B_k)(s_k - G(x^*)^{-1}\delta_k)\| \\
&\quad = \|Q^{-1}H_k(G(x^*) - B_k)G(x^*)^{-1}(G(x^*)s_k - \delta_k)\| \\
&\quad = \|Q^{-1}H_k(G(x^*) - B_k)G(x^*)^{-1}[(G(x^*) - G(x_k))s_k + (G(x_k)s_k - \delta_k)\| \\
&\quad \leq \|Q^{-1}H_k(G(x^*) - B_k)G(x^*)^{-1}\|\{\|G(x^*) - G(x_k)\|\|s_k\| + \|G(x_k)s_k - \delta_k\|\} \\
&\quad = o(\|s_k\|)
\end{aligned}$$

and

$$\begin{aligned}
&r_k\|Q^{-1}H_k(G(x^*) - B_k)G(x^*)^{-1}s_k\| \\
&\quad \leq r_k\|Q^{-1}H_k(G(x^*) - B_k)G(x^*)^{-1}\|\|s_k\| = o(\|s_k\|),
\end{aligned}$$

where we used the fact that $\{\|B_k\|\}$ and $\{\|H_k\|\}$ are bounded (and therefore uniformly nonsingular), $G(x)$ is continuous and $r_k \to 0$. Therefore, it follows from (3.27) that for some constant $m_1 > 0$

$$\|Q^{-1}(H_k - G(x^*)^{-1})y_k\| \geq m_1\|(G(x^*) - B_k)s_k\| - o(\|s_k\|). \tag{3.28}$$

On the other hand, we have from (3.3) in Lemma 3.1

$$\|Qy_k\| \leq \|Q\|\|y_k\| \leq (L + r)\|Q\|\|s_k\|. \tag{3.29}$$

From (3.28), (3.29) and (3.26) we conclude that the Dennis-Moré condition (3.25) holds.

Next we verify that $\lambda \equiv 1$ for all k sufficiently large. Since $\|p_k\| = \|H_k g_k\| \to 0$, by Taylor's expansion we obtain

$$\begin{aligned}
& f(x_k + p_k) - f(x_k) - \sigma_1 g(x_k)^T p_k \\
&= (1 - \sigma_1)g(x_k)^T p_k + \frac{1}{2}p_k^T G(x_k + \bar{\theta}_k p_k)p_k \\
&= -(1 - \sigma_1)p_k^T B_k p_k + \frac{1}{2}p_k^T G(x_k + \bar{\theta}_k p_k)p_k \\
&= -(\frac{1}{2} - \sigma_1)p_k^T B_k p_k - \frac{1}{2}p_k^T (B_k - G(x_k + \bar{\theta}_k p_k))p_k \\
&= -(\frac{1}{2} - \sigma_1)p_k^T G(x^*)p_k + o(\|p_k\|^2),
\end{aligned}$$

where $\bar{\theta}_k \in (0, 1)$ and the last equality follows from (3.25). Thus $f(x_k + p_k) - f(x_k) - \sigma_1 g(x_k)^T p_k \leq 0$ for all k sufficiently large. In other words, $\lambda_k = 1$ satisfies the first inequality of (2.11) for all k sufficiently large.

On the other hand, we have

$$\begin{aligned}
& g(x_k + p_k)^T p_k - \sigma_2 g(x_k)^T p_k \\
&= (g(x_k + p_k) - g(x_k))^T p_k + (1 - \sigma_2)g(x_k)^T p_k \\
&= p_k^T G(x_k + \bar{\theta}'_k p_k)p_k - (1 - \sigma_2)p_k^T B_k p_k \\
&= p_k^T G(x_k + \bar{\theta}'_k p_k)p_k - (1 - \sigma_2)p_k^T G(x_k)p_k + o(\|p_k\|^2) \\
&= \sigma_2 p_k^T G(x_k)p_k + o(\|p_k\|^2),
\end{aligned}$$

where $\bar{\theta}'_k \in (0,1)$. So we have $g(x_k+p_k)^T p_k \geq \sigma_2 g_k^T p_k$, which means that $\lambda_k = 1$ satisfies the second inequality of (2.11) for all k sufficiently large. Therefore, we assert that the unit stepsize is accepted when k is sufficiently large. Consequently we can deduce that $\{x_k\}$ converges superlinearly. □

4 A Practicable MBFGS Method

In Section 3, we have proposed a general modification of BFGS method. We have also shown that if r_k is chosen to satisfy (2.12) then under appropriate conditions, the modified BFGS method converges globally. If in addition (3.8) holds, then the convergence rate is superlinear. In other words, the global and superlinear convergence of MBFGS method depends on the choice of $\{r_k\}$. Therefore, it is important to select $\{r_k\}$ in a suitable way so that it is practicable and satisfies (2.12) and (3.8). The following theorem shows that $r_k = \|g_k\|$ is a suitable choice.

Theorem 4.1 *Let the level set Ω is bounded. $\{x_k\}$ is generated by MBFGS method with $r_k = \|g_k\|$ and B_k being updated by (2.13). Then*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{4.1}$$

If in addition, the condition (1)– (3) in Assumption (A) hold and $\sigma_1 \in (0, \frac{1}{2})$, then $\{x_k\}$ converges superlinearly.

Proof We assume to be the contrary that $\{\|g_k\|\}$ is bounded away from zero. Then (2.12) holds. Therefore, Theorem 3.1 implies that (4.2) holds, which is a contradiction.

If in addition, the conditions (1) – (3) in Assumption (A) hold, then since

$$r_k = \|g_k\| = \|g(x_k) - g(x^*)\| \leq L\|x_k - x^*\|,$$

the proof of Lemma 3.5 indicates that $\sum_{k=0}^{\infty} \|x_k - x^*\| < \infty$. In other words, $\{r_k\}$ satisfies (3.8). On the other hand, from (3.10), the conditions (1) and (2) in

Assumption (A) imply that g is strongly monotone around x^*, so that there exists a constant $m_2 > 0$ such that

$$y_k^T s_k \geq \delta_k s_k = (g_{k+1} - g_k)^T s_k \geq m_2 \|s_k\|^2$$

for all k sufficiently large. Therefore, from Theorem 3.2 we get the superlinear convergence of $\{x_k\}$ and complete the proof. □

5 Backtracking Line Search

In this section, we extend the results obtained in the former sections for MBFGS method with backtracking line search. In other words, we will consider the algorithm in which line search condition (2.11) is replaced by the following inequality:

$$f(x_k + \lambda_k p_k) \leq f(x_k) + \sigma \lambda_k g_k^T p_k, \tag{5.1}$$

where $\sigma \in (0, 1)$ is a given constant.

However, in this case, the condition that $y_k^T s_k \geq 0$ may fail to be fulfilled, which leads B_{k+1} to be possibly not positive definite even if B_k is positive definite. Therefore, further modification on y_k is necessary. To do this, we choose $y_k = \delta_k + t_k \|g_k\| s_k$. The constant $t_k > 0$ is determined by

$$t_k = 1 + \max\{-\frac{\delta_k^T s_k}{\|s_k\|^2}, 0\}. \tag{5.2}$$

Following this choice of t_k we always have that $y_k^T s_k \geq \|g_k\| \|s_k\|^2$.

By doing this, we establish the following algorithm.

Algorithm 2 (Modified BFGS Method with Backtracking Line Search)

Initial Take initial point $x_0 \in R^n$ and initial positive definite matrix B_0 and condtant $\sigma \in (0, 1)$. $k := 0$.

Step 1 Solve the following linear equations on p to get p_k.

$$B_k p + g_k = 0. \tag{5.3}$$

Step 2 For a given $\rho \in (0,1)$, find the smallest integer of i, say i_k, satisfying that

$$f(x_k + \rho^i p_k) \leq f(x_k) + \sigma \rho^i g_k^T p_k. \tag{5.4}$$

Take $\lambda_k = \rho^{i_k}$.

Step 3 Get the next iterate by $x_{k+1} = x_k + \lambda_k p_k$.

Step 4 Update B_k to get B_{k+1} by BFGS formula

$$B_k - \frac{B_k s_k s_k^T B_k}{B_k s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \tag{5.5}$$

where $s_k = x_{k+1} - x_k = \lambda_k p_k$ and $y_k = \delta_k + t_k \|g_k\| s_k$.

Step 5 $k := k + 1$ and go to step 1.

Remark The differences between Algorithm 1 and 2 are the line search step 2 and the expression of y_k.

Lemma 5.1 *Let the level set Ω is bounded. $\{x_k\}$ is generated by Algorithm 2. If $\|g_k\| \geq \zeta$ for all k and some constant $\zeta > 0$, then there are positive constants β_j $j = 1,2,3$ such that for all k the following inequalities*

$$\|B_i s_i\| \leq \beta_1 \|s_i\|, \quad \text{and} \quad \beta_2 \|s_i\|^2 \leq s_i^T B_i s_i \leq \beta_3 \|s_i\|^2 \tag{5.6}$$

hold for at least $[\frac{k}{2}]$ number of i.

Proof It is easy to see that $\{x_k\} \subset \Omega$ and $\|y_k\| \leq C\|s_k\|$ for some positive constant C. By the choice of t_k we already have that $y_k^T s_k \geq \|g_k\| \|s_k\|^2 \geq \zeta \|s_k\|^2$. Therefore, by means of Theorem 2.1 of [2] we complete the proof. □

Denote

$$J = \{i \mid (5.6) \text{ holds } \}. \tag{5.7}$$

Lemma 5.2 *Let the level set Ω is bounded. If $\|g_k\| \geq \zeta > 0$ for all k, then there is a positive constant $\bar{\lambda}$ such that $\lambda_i \geq \bar{\lambda}$ for all $i \in J$.*

Proof If $\lambda_i \neq 1$, then by the line search process we see that $\lambda_k' \equiv \lambda_k/\rho$ does not satisfy the line search inequality (5.1), i.e.

$$f(x_i + \lambda_i' p_i) - f(x_i) > \sigma \lambda_i' g_i^T p_i. \tag{5.8}$$

By means of mean-value theorem and (1.2) we have

$$\begin{aligned} f(x_i + \lambda_i' p_i) &= f(x_i) + \lambda_i' g(x_i)^T p_i + \lambda_i' (g(x_k + \theta_i \lambda_i' p_i) - g(x_i))^T p_i \\ &\leq f(x_i) + \lambda_i' g(x_i)^T p_i + (\lambda_i')^2 L \|p_i\|^2, \end{aligned}$$

where $\theta_i \in (0,1)$. Substitute this to (5.8) we get that for all $i \in J$,

$$\begin{aligned} \lambda_i' L \|p_i\|^2 &> -(1-\sigma) g(x_i)^T p_i \\ &= (1-\sigma) p_i^T B_i p_i \geq (1-\sigma)\beta_2 \|p_i\|^2. \end{aligned}$$

This means that $\lambda_i' \geq \frac{1-\sigma}{L}$ and $\lambda_i = \lambda_i'/\rho \geq \frac{1-\sigma}{L}\rho^{-1} \triangleq \bar{\lambda} > 0$. □

The global convergence theorem of Algorithm 2 is as follows.

Theorem 5.1 *Let the level set Ω be bounded. $\{x_k\}$ is generated by Algorithm 2. Then*

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

Proof If $\|g_k\| \geq \zeta > 0$ for all k. Then summing (5.1) we have

$$\begin{aligned} \infty &> \sum_{i=0}^{\infty} (-g_i^T s_i) = \sum_{i=0}^{\infty} \|g_k\|^2 \lambda_i \frac{s_i^T B_i s_i}{\|B_i s_i\|^2} \\ &\geq \delta^2 \sum_{i=0}^{\infty} \lambda_i \frac{s_i^T B_i s_i}{\|B_i s_i\|^2} \geq \delta^2 \sum_{\substack{i=0 \\ i\in J}}^{\infty} \lambda_i \frac{s_i^T B_i s_i}{\|B_i s_i\|^2} \\ &\geq \delta^2 \bar{\lambda} \sum_{\substack{i=0 \\ i\in J}}^{\infty} \frac{s_i^T B_i s_i}{\|B_i s_i\|^2}. \end{aligned}$$

At last, repeat the proof of Theorem 3.1 we conclude the proof. □

For the superlinear convergence of Algorithm 2, it is easy to see that if Assumption (A) (2) holds, then $t_k \equiv 1$ for all sufficiently large k. Therefore, we can get the following theorem by a similar way to the proof of Theorem 3.2.

Theorem 5.2 *Let the conditions (1) – (3) in Assumption (A) hold. If σ is chosen to satisfy $\sigma_1 \in (0, \frac{1}{2})$, then $\{x_k\}$ converges to x^* superlinearly.*

References

[1] C.G. Broyden, J.E. Dennis, Jr. and J.J. Moré, On the locall and superlinear convergence of quasi-Newton methods, *J. Inst. Math. Appl.*, 12 (1973) 223-246.

[2] 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) 727-739.

[3] 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) 1171-1189.

[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) 549-560.

[5] J.E. Dennis and J.J. Moré, Quasi-Newton methods, motivation and theory, *SIAM Rev.* 19 (1977) 46-89.

[6] L.C.W. Dixon, Variable metric algorithms: necessary and sufficient conditions for identical behavior on nonquadratic functions, *J. Opt. Theory Appl.* 10 (1972) 34-40.

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

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

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

[10] A. Griewank, The global convergence of partitioned BFGS on problems with convex decompositions and Lipschitz gradients, *Math. Prog.* 50 (1991) 141-175.

[11] D. Li, On the global convergence of DFP method, *Acta. Hunan Univ.*, 1993, No.5.

[12] M.J.D. Powell, On the convergence of the variable metric algorithms, *J. Inst. Math. Appl.* 7 (1971) 21-36.

[13] M.J.D. Powell, Some global convergence properties of variable metric algorithms for minimization without exact line searches, in *Nonlinear Programming*, SIAM-AMS Proceedings, **Vol.IX**, R.W. Cottle, and C.E. Lemke, eds., SIAM 1976.

[14] Ph.L. Toint, Global convergence of the partitioned BFGS algorithm for partially separable optimization, *Math. Prog.* 36 (1986) 290-306.

[15] Y. Zhang and R.P. Tewarson, Quasi-Newton algorithm with updates from preconvex part of Broyden's family, *J. Numer. Anal.* 8 (1988) 487-509.