American Journal of Computational Mathematics, **2011, 1, 240-246**
doi:10.4236/ajcm.2011.14028 Published Online December 2011 (http://www.SciRP.org/journal/ajcm)

A Look at the Tool of BYRD and NOCEDAL*

Linghua Huang[1], Guoyin Li[2], Gonglin Yuan[3]
[1]*Department of Mathematics and Statistics*, *Guangxi University of Finance and Economics*, *Nanning*, *China*
[2]*Department of Applied Mathematics*, *The University of New South Wales*, *Sydney*, *Australia*
[3]*Department of Mathematics and Information Science*, *Guangxi University*, *Nanning*, *China*
E-mail: *linghuahuang@163.com*, *g.li@unsw.edu.au*, *glyuan@gxu.edu.cn*
Received July 13, 2011; *revised September* 6, 2011; *accepted September* 15, 2011

Abstract

A power tool for the analysis of quasi-Newton methods has been proposed by Byrd and Nocedal ([1], 1989). The purpose of this paper is to make a study to the basic property (BP) given in [1]. As a result of the BP, a sufficient condition of global convergence for a class of quasi-Newton methods for solving unconstrained minimization problems without convexity assumption is given. A modified BFGS formula is designed to match the requirements of the sufficient condition. The numerical results show that the proposed method is very encouraging.

Keywords: quasi-Newton Method, Unconstrained Minimization, Nonconvex Problem, Global Convergence

1. Introduction

Given a real-valued function $f: R^n \to R$, we are interested in solving the unconstrained optimization problem

$$\min\left\{ f(x) \middle| x \in R^n \right\}$$

by Quasi-Newton methods, which have the form

$$x_{k+1} = x_k + \alpha_k d_k,$$

where α_k is a steplength, and d_k is a search direction given by the following linear equation

$$B_k d + g_k = 0. \tag{1.1}$$

The matrix B_k is updated at every step such that B_k satisfies the so-called secant equation

$$B_{k+1} s_k = \nu_k,$$

where $s_k = x_{k+1} - x_k$, $\nu_k = g_{k+1} - g_k$ and g_j denotes the gradient of f at x_j.

Global convergence of quasi-Newton methods has been widely studied in the past two decades. For convex minimization, Powell [2] showed that, with the weak Wolfe-Powell line search strategies, $\lim\ \inf_{k\to\infty} g_k = 0$. Werner [3] made an extension of Powell's result to some other line searches. Byrd, Nocedal and Yuan [4] made an inside study for the restricted Broyden class of quasi-Newton methods. Byrd and Nocedal (1989) proposed a very useful tool for the analysis of quasi-Newton methods. The basic property (BP) given by Byrd and Nocedal (1989) characterized not only the BFGS formula but also any formula with the structure of the BFGS, some of the examples are the modified BFGS methods given by Li and Fukushima [5-7]. In [5], Li and Fukushima gave a modified BFGS method with good convergence properties for solving symmetric nonlinear equations, while in [6,7], the BFGS type methods with global and superlinear convergence are designed for nonconvex optimization problems. The proofs of some main results given in [5-7] are related closely to the BP. Some modified BFGS methods which possess not only the gradient value but also the function value information have been proposed (see [8,9] etc.). The main purpose of this paper is to give some insight of the BP for a class of quasi-Newton methods.

In the next section we recall some basic concepts and results of [1], and then study a class of quasi-Newton methods. By using the BP, we obtain a natural property of the proposed methods. Moreover, we give a sufficient condition for the quasi-Newton methods, which is motivated also by BP. In section 3, we design a modified BFGS method to match the requirements of the given sufficient condition. As we will see, the proposed method

*This work is supported by Program for Excellent Talents in Guangxi Higher Education Institutions, China NSF grands 10761001, Guangxi Education research project grands 201012MS013, Guangxi SF grands 0991028, and the Scientific Research Foundation of GuangXi University (Grant No. XJZ110632).

is globally and superlinearly convergent for nonconvex unconstrained minimization problems. The numerical results are contained in Section 4. In Section 5, we study the set of good iterates of BFGS and DFP formulas with different step size strategies by empirical analysis. Throughout this paper, the norm is the Euclidean vector norm.

2. Set of Good Iterates

After making an inside study on the tool given by Byrd and Nocedal, we found that the main contribution of [1] are three: 1) gave a power tool for analysis of quasi-Newton methods; 2) showed the BFGS formula possesses the BP that is independent of the algorithmic context of the update for convex minimization proplems and 3) characterized the set of good iterates by using the information of the update matrix $\{B_k\}$ and the iteration point $\{x_k\}$.

For the following BFGS formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{v_k v_k^T}{v_k^T s_k}, \tag{2.1}$$

Byrd and Nocedal [1] proposed a basic property that indicates, under some conditions, the most iterates generated by (2.1) are good iterates. It is more interesting to note that the above conclusion is independent on any line search strategies. The BP given by Byrd and Nocedal (Theorem 2.1 of [1]) is as follows:

Theorem 2.1. Let $\{B_k\}$ be generated by (2.1) with the following properties: $B_1 > 0$ and for all $k \ge 1$,

$$\frac{v_k^T s_k}{s_k^T s_k} \ge \tau_1 \quad \text{for some} \quad \tau_1 > 0, \tag{2.2}$$

$$\frac{\|v_k\|^2}{v_k^T s_k} \le \tau_2 \quad \text{for some} \quad \tau_2 > 0, \tag{2.3}$$

Then for any $p \in (0,1)$ there exist constants $\beta_1, \beta_2, \beta_3 > 0$ such that, for all $k > 1$, the relations

$$\frac{s_j^T B_j s_j}{\|\mathrm{s_j}\| \; \|\mathrm{B_j} d_j\|} \ge \beta_1, \tag{2.4}$$

$$\beta_2 \le \frac{s_j B_j s_j}{s_j^T s_j} \le \beta_3, \tag{2.5}$$

$$\beta_2 \le \frac{\|B_j s_j\|}{\|s_j\|} \le \frac{\beta_3}{\beta_1}, \tag{2.6}$$

hold for at least $\lceil p_k \rceil$ values of $j \in \{1, 2, 3, \cdots, k\}$.

The conclusion of Theorem 2.1 is right for any formula with the form

$$B_{k+1} = B_k - \frac{B_k r_k r_k^T B_k}{r_k^T B_k r_k} + \frac{u_k u_k^T}{u_k^T r_k} \tag{2.7}$$

if $B_1 > 0$ and for all k, $u_k^T r_k > 0$, where $u_k, r_k \in \Re^n$. Moreover, the proof of the conclusion for (2.7) does not need to change anywhere. This makes one can choose s_k and y_k such that $B_k > 0$, (2.2) and (2.3) hold. In fact, Li and Fukushima have followed this way and gave some modified BFGS formula which possess global and superlinear convergence for nonconvex minimization [6] and for symmetric nonlinear equations [5]. In this sense, the tool given in [1] for proving Theorem 2.1 is very powerful.

Let

$$\Re^+ = (0, +\infty), \quad N^+ = \{\text{all nonngative integer}\}.$$

Define four functions SGI, Φ_1, Φ_2 and Φ as follows:

$SGI : \Re^+ \times \Re^+ \times \Re^+ :\to 2^{N^+}$ has the form $SGI(\beta_1, \beta_2, \beta_3)$

$= \{\text{all the index which satisfis } (2.4), (2.5) \text{ and } (2.6) \text{ simultaneeousl}\}$

$\Phi_1 : \Re^+ :\to 2^{N^+}$ has the form $\Phi_1(c) = \{k \,|\, \|\mathrm{B_k} d_k\| \le c \|d_k\|\}$;

$\Phi_2 : \Re^+ :\to 2^{N^+}$ has the form $\Phi_2(c)$

$= \{k \,|\, c \|d_k\|^2 \le d_k^T B_k d_k\}$;

$\Phi : \Re^+ \times \Re^+ :\to 2^{N^+}$ has the form $\Phi(c_1, c_2)$

$= \Phi_1(c_1) \cap \Phi_2(c_2)$.

In [1], the set $SGI(\beta_1, \beta_2, \beta_3)$ was been called the "set of good iterates". From Theorem 2.1 and p can be chosen to be close to 1, we can deduce that, for any line search strategies, if the conditions (2.2) and (2.3) are satisfied, then most of the iterates given by the BFGS formula are good iterates. The meaning of "good" is certified in [1] (Theorem 3.1) by proving, with some certain line search strategies, that any quasi-Newton method is R-linearly convergent for uniform convex minimization problems.

In the remainder of this section, we will give a "set of good iterates" for a general quasi-Newton methods. In order to simplify the presentation, we use $B \ge 0 (> 0)$ to denote any $n \times n$ symmetric and positive semi-definite (definite) matrix B. We state the method, which will be called the GQNLSS: (general quasi-Newton method with some certain line search strategies) as follows.

Algorithm 2.1: GQNLSS

Step 0: Choose an initial point $x_1 \in \Re^n$ and an initial matrix $B_1 \ge 0$. Choose $\delta \in \left(0, \frac{1}{2}\right), \sigma \in (0,1)$ Set $k := 1$.

Step 1: If $g_k = 0$, Stop.

Step 2: Solve (1.1) to obtain a search direction d_k.

Step 3: find α_k by some certain line search strategies.

Step 4: Set $x_{k+1} = x_k + \alpha_k d_k$. Update $B_{k+1} \ge 0$ by some formula.

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

The line search strategies used in the GQNLSS is one of the following three forms:

1) Efficient line search: find α_k satisfies

$$f(x_k + \alpha_k d_k) - f(x_k) \le -\eta_1 \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2},$$

where η_1 is a positive constant.

2) Standard Armijo line search: find $\alpha_k = \rho^{jk}$ such that j_k is the smallest nonnegative integer j satisfying

$$f\left(x_k + \rho^{jk} d_k\right) \le f(x_k) + \eta_2 \rho^{jk} g_k^T d_k, \tag{2.8}$$

where $\eta_2 \in (0,1)$ and $\rho \in (0,1)$ are constants.

3) Weak Wolfe-Powell (WWP) line searches: find α_k satisfies condition:

$$f\left(x_k + \rho^{jk} d_k\right) \le f(x_k) + \eta_2 \rho^{jk} g_k^T d_k, \tag{2.9}$$

and

$$g(x_k + \alpha_k d_k) \ge \sigma g_k^T d_k \tag{2.10}$$

where $\eta_3 \in (0,1)$ and $\sigma \in (\eta_3, 1)$ are constants.

In order to study the convergence behavior of Algorithm 2.1, we will impose the following two assumptions, which have been widely used in the literature to analyze the global convergence of iterative solution methods for minimization problem with inexact line searches (see [10,11] etc.).

Assumption A. The level set is bounded.

$$\Omega = \left\{x \in R^n \;\middle|\; f(x) \le f(x_1)\right\}$$

Assumption B. There exists a constant L such that for any $x, y \in \Omega$,

$$\|g(x) - g(y)\| \le L\|x - y\|.$$

The following natural property, characterizes the updated matrix B_k and the direction d_k generated by GQNLLS method.

Theorem 2.2. Suppose that Assumptions A and B hold, $\{x_k, \alpha_k, B_k, d_k\}$ is generated by GQNLSS. Then

$$\sum_{k=1}^{\infty} \frac{\left(d_k^T B_k d_k\right)^2}{\|d_k\|^2} < +\infty. \tag{2.11}$$

Proof: From Assumption A, we have that $\{x_k\} \subseteq \Omega$. Thus

$$\sum_{k=1}^{\infty} \left(f(x_k + 1) - f(x_k)\right) > -\infty \tag{2.12a}$$

1) For the line searches 1), we have (2.11) by using (1.1), $B_k \ge 0$ and (2.12).

2) For the line searches 2), it suffices to consider the case $\alpha_k \ne 1$. From the definition of α_k, we have

$$f\left(x_k + (\alpha_k/\rho) d_k\right) - f(x_k) > \eta_2 (\alpha_k/\rho) g_k^T d_k.$$

Using the Mean Value Theorem in the above inequality, we obtain $\theta_k \in (0,1)$, such that

$$g\left(x_k + \theta_k (\alpha_k/\rho) d_k\right)^T \left((\alpha_k/\rho)\right) d_k > \eta_2 (\alpha_k/\rho) g_k^T d_k.$$

Dividing the both side of the above inequality by α_k/ρ, we have

$$g\left(x_k + \theta_k (\alpha_k/\rho) d_k\right)^T d_k > \eta_2 g_k^T d_k.$$

Subtracting $g_k^T d_k$ on the both sides of the above inequality, we obtain

$$\left(g\left(x_k + \theta_k (\alpha_k/\rho) d_k\right) - g(x_k)\right)^T d_k > -(1-\eta_2) g_k^T d_k,$$

which combining with Assumption B yields

$$L\theta_k (\alpha_k/\rho) \|d_k\|^2 > -(1-\eta_2) g_k^T d_k.$$

Therefore

$$\alpha_k > \frac{(1-\eta_2)\rho}{L\theta_k} \frac{\left(-g(x_k)^T d_k\right)}{\|d_k\|^2}$$

Hence, we have, by using $\theta_k \in (0,1)$, that

$$\alpha_k > \frac{(1-\eta_2)\rho}{L} \frac{\left|g(x_k)^T d_k\right|}{\|d_k\|^2}$$

Thus

$$\sum_{k=1}^{\infty} \frac{\left(d_k^T B_k d_k\right)^2}{\|d_k\|^2} < +\infty \tag{2.12b}$$

by using (2.8), (1.1) and (2.12).

3) From Assumption B and (2.10), we obtain

$$-(1-\sigma) g_k^T d_k \le (g_{k+1} - g_k)^T d_k \le L\alpha_k \|d_k\|^2,$$

which implies that

$$\alpha_k \ge \frac{1-\sigma}{L} \frac{\left|g_k^T d_k\right|}{\|d_k\|^2}$$

Thus

$$\alpha_k \ge \frac{1-\sigma}{L} \frac{d_k^T B_k d_k}{\|d_k\|^2}$$

by using (1.1). Thus, (2.11) holds by using (2.12).From the results of a)-c), we have (2.11). The proof is complete.

Let $0 \le \lambda_{1k} \le \lambda_{2k} \le \cdots \le \lambda_{nk}$ be the eigenvalues of B_k and $cond(B_k)$ be the condition number of B_k, *i.e.*, $cond\left(B_k\right) = \frac{\lambda_{nk}}{\lambda_{1k}}$

Theorem 2.3. Suppose that Assumptions A and B hold, $\{x_k, \alpha_k, B_k, d_k\}$ is generated by GQNLSS. If there exist a positive constant M and an infinite index set K such that for all $k \in K$,

$$cond\left(B_k\right) \le M, \tag{2.13}$$

then

$$\lim_{k\to\infty} g\left(x_k\right) = 0. \tag{2.14}$$

Proof: From Theorem 2.2, we have

$$\lim_{k\to\infty} \lambda_{1k}^2 \left\|d_k\right\|^2 \le \lim_{k\to\infty} \frac{d_k^T B_k d_k}{\left\|d_k\right\|^2} = 0.$$

Thus,

$$\lim_{k\to\infty} \lambda_{1k} \left\|d_k\right\| = 0.$$

Therefore, from (2.13), we obtain

$$\lim_{k\in K} \left\|B_k d_k\right\| \le \lim_{k\in K} \lambda_{nk} \left\|d_k\right\| = \lim_{k\in K} cond(B_k)\lambda_{1k} \left\|d_k\right\| = 0.$$

which implies that (2.14) holds by using (1.1). The proof is complete.

From Theorem 2.3, we have the following result, which indicates that if GQNLSS fails to converge, then the condition numbers sequence $\left\{cond\left(B_k\right)\right\}$ will tend to infinite.

Corollary 2.1. Suppose that Assumptions A and B hold, $\{x_k, \alpha_k, B_k, d_k\}$ is generated by GQNLSS. If

$$\liminf_{k\to\infty} \left\|g\left(x_k\right)\right\| > 0$$

Then

$$\lim_{k\to\infty} cond\left(B_k\right) = +\infty$$

We call GQNLSS has property SC∞ if

$$\left|\Phi\left(c_1, c_2\right)\right| = +\infty \text{ for some } c_1, c_2 \in \Re^+ \tag{2.15}$$

Theorem 2.4 Suppose that Assumptions A and B hold, $\{x_k, \alpha_k, B_k, d_k\}$ is generated by GQNLSS. If GQNLSS has property SC$^\infty$, then

$$\lim_{k\in\Phi(c_1,c_2)} g\left(x_k\right) = 0$$

Proof: From the definition of Φ, we have

$$\begin{aligned} &\left\|B_k d_k\right\| \le c_1 \left\|d_k\right\| \text{ and } c_2 \left\|d_k\right\|^2 \\ &\le d_k^T B_k d_k, \text{ for } k \in \Phi\left(c_1, c_2\right). \end{aligned} \tag{2.16}$$

From Theorem 2.1, we have

$$0 = \lim_{k\in\Phi_2(c_2)} \frac{\left(d_k^T B_k d_k\right)^2}{\left\|d_k\right\|^2} \ge \lim_{k\in\Phi_2(c_2)} \frac{c_2^2 \left\|d_k\right\|^4}{\left\|d_k\right\|^2} = c_2^2 \lim_{k\in\Phi_2(c_2)} \left\|d_k\right\|^2$$

By using the definitions of Φ_1 and Φ_2, we obtain

$$\lim_{k\in\Phi(c_1,c_2)} \left\|B_k d_k\right\| = 0.$$

Therefore, (2.16) follows (1.1). The proof is complete.

The main contribution of Theorem 2.4 is that it identifies the convergence indices of $\{g_k\}$. It indicates that $\Phi\left(\beta_1, \beta_2\right)$ is the "set of good iterates" in the sense that the method is globally convergent. From Theorem 2.4, we see that (2.15) is a sufficient condition of the global convergence for GQNLSS. The sufficient condition is tight under the sense that the set $\Phi\left(\beta_1, \beta_2\right)$ is as same as the convergence indices of the sequence $\{g_k\}$. From the fact that $\Phi\left(\beta_1, \beta_2\right) \supseteq SCI\left(\beta_1, \beta_2, \beta_3\right)$, we see that, for GQNLSS, the "set of good iterates" is little larger than which given in [1]. Note that the approach here without any convexity assumption.

For BFGS formula, whether (2.2) and (2.3) hold is still open for nonconvex minimization problems. In general, it is very hard to prove them. By Theorem 2.3 and 2.4, we can obtain that if one can prove that for some positive constants c_1 and c_2,

$$\left|\Phi\left(c_1, c_2\right) \cup \left\{k \,\middle|\, cond\left(B_k\right) \le c_1\right\}\right| = +\infty,$$

then the BFGS formula with any one of the three line search strategies mentioned above is globally convergent. This result can be extended to any quasi-Newton formula (see GQNLSS), such as DFP formula.

3. Design Methods with SC$^\infty$

The purpose of this section is to discuss how to design a BFGS-type method (*i.e.*, the update has the form (2.7)) with global convergence (if possible with superlinear convergence). From Theorem 2.4, it suffices to design the methods which possess the property SC∞. Clearly, it is more than enough to find y_k such that for all k, (2.2) and (2.3) hold by using Theorem 2.1. Although following this way may yield some strict conditions (it seems that, for nonconvex minimization problems, it excludes the BFGS update), it is better than nothing at this moment. First we will propose a BFGS-type update to match the requirements of (2.2) and (2.3), and prove the corresponding method possesses the property SC$^\infty$ by using the same way given in the proof of Theorem 2.1 in [1]. Some other new formulas will also be given then in this section.

The first modified BFGS update is as follows:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \tag{3.1}$$

where $s_k = x_{k+1} - x_k = \alpha_k d_k$, $v_k = g_{k+1} - g_k$, $y_k = v_k + m_k s_k$.

The parameter m_k is defined by

$$m_k = \mu_1 + \mu_2 \frac{v_k^T s_k}{s_k^T s_k}$$

where $\mu_1 \in (0, +\infty)$ and $\mu_2 \in [1, +\infty)$ are two constants. For (3.1), we have a corresponding method, which is described as follows.

Algorithm 3.1: A BFGS-Type Method with WWP (BFGSTWWP)

Step 0: Choose an initial point $x_1 \in \Re^n$ and an initial matrix $B_1 > 0$. Set k := 1.

Step 1: If $g_k = 0$, Stop.

Step 2: Solve (1.1) to obtain a search direction d_k.

Step 3: find α_k by WWP (2.9) and (2.10). Moreover, if $\alpha_k = 1$ satisfies WWP, set $\alpha_k = 1$.

Step 4: Set $x_{k+1} = x_k + \alpha_k d_k$. Update B_{k+1} by (3.1).

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

The following Lemma shows that the updated matrix $B_k > 0$, so we can always solve (1.1) uniquely.

Lemma 3.1. Suppose that $\{x_k, \alpha_k, B_k, d_k\}$ is generated by BFGSTWWP. Then for all $k > 1, B_k > 0$.

Proof: Suppose that for a given k, $B_k > 0$. From the definitions of y_k and s_k, we have

$$\begin{aligned} v_k^T s_k &= \alpha_k \left(g_{k+1}^T d_k - g_k^T d_k \right) \ \ge -\alpha_k (1-\sigma) g_k^T d_k \\ &= (1-\sigma) \alpha_k d_k^T B_k d_k > 0 \end{aligned}$$

By using (2.10) and (1.1). Thus

$$\begin{aligned} y_k^T s_k &= v_k^T s_k + \mu_1 s_k^T s_k + \mu_2 v_k^T s_k \\ &\ge \mu_1 s_k^T s_k + (1-\sigma)(1+\mu_2) \alpha_k d_k^T B_k d_k \qquad (3.2) \\ &\ge \mu_1 s_k^T s_k \end{aligned}$$

It is easy to prove that $B_{k+1} > 0$ by using $B_k > 0$ and (3.2). By using $B_1 > 0$ and the induction, we may deduce that for all k, $B_k > 0$.

Theorem 3.1. Suppose that Assumptions A and B hold, $\{x_k, \alpha_k, B_k, d_k\}$ is generated by BFGSTWWP.

Then for any $p \in (0,1)$ there exist constants $\gamma_1, \gamma_2 > 0$ such that, for all $k > 1$, the relations

$$\|B_j s_j\| \le \gamma_1 \|s_j\| \qquad (3.3)$$

$$\gamma_2 s_j^T s_j \le s_j^T B_j s_j, \qquad (3.4)$$

hold for at least $\lceil p_k \rceil$ values of $j \in \{1, 2, \cdots, k\}$.

Proof: For any given k, from the definition of v_k and Assumption B, we have

$$\|v_k\| = \|g_{k+1} - g_k\| \le L \|s_k\|,$$
$$v_k^T s_k > 0$$

and

$$|m_k| \le \mu_1 + |\mu_2| \frac{\|v_k\| \|s_k\|}{s_k^T s_k} \le \mu_1 + |\mu_2| L.$$

Using the above two inequalities and the definitions of y_k, v_k and s_k, we have

$$\begin{aligned} y_k^T y_k &= v_k^T v_k + m_k{}^2 s_k^T s_k \\ &\le \|v_k\|^2 + 2|m_k| \|v_k\| \|s_k\| + m_k^2 \|s_k\|^2 \\ &\le \left(L^2 + 2\left(\mu_1 + |\mu_2| L\right) L + \left(\mu_1 + |\mu_2| L\right)^2 \right) s_k^T s_k. \end{aligned}$$

Therefore, we obtain, by using (3.2), that

$$\frac{y_k^T s_k}{s_k^T s_k} \ge \mu_1 \qquad (3.5)$$

and

$$\frac{y_k^T y_k}{y_k^T s_k} \le \frac{L^2 + 2\left(\mu_1 + L|\mu_2|\right) L + \left(\mu_1 + L|\mu_2|\right)^2}{\mu_1}. \qquad (3.6)$$

Thus, the proof follows from that of Theorem 2.1 in [1].

From Theorem 2.4 and Theorem 3.1, we have the following global convergence result for BFGSTWWP.

Theorem 3.2 Suppose that Assumptions A and B hold, $\{x_k, \alpha_k, B_k, d_k\}$ is generated by BFGSTWWP. Then

$$\lim_{k \in \Phi(\gamma_1, \gamma_2)} g(x_k) = 0.$$

Proof: Omitted.

From the proof of Lemma 3.1, we observe that (3.2) is not independent of the line searches, thus, (3.5) and (3.6) is also not independent of the line searches. Therefore, like the BFGS formula, when the standard Armijo line search is used to the BFGS-type formula (3.1), $B_k > 0$ cannot be guaranteed for some cases because whether the relation $y_k^T s_k > 0$ holds is still open for nonconvex problems. It is possible to give a formula such that $B_k > 0$, (3.5) and (3.6) hold without any line search strategies. For example, the m_k in (3.1) is replaced by

$$m_k^0 = \mu_1 + \ell \frac{v_k^T s_k}{s_k^T s_k}$$

with $\ell \in [1, +\infty)$. In this case, the results of Lemma 3.1, Theorem 3.1, (therefore SC^∞) hold for any $\alpha_k > 0$. This implies clearly, by using Theorem 2.4, that the result of Theorem 3.2 holds if any one of the line search strategies given in Section 2 is used.

From the proof of Theorem 2.1 in [1], it is possible to give the exact values of γ_1 and γ_2. Let

$$\psi(B_1) = tr(B_1) - \ln(\det(B_1)),$$

$$M = \frac{L^2 + 2\left(\mu_1 + L|\mu_2|\right) L + \left(\mu_1 + L|\mu_2|\right)^2}{\mu_1}$$

and

$$\kappa = \frac{1}{1-p}\left(\psi\left(B_1\right) + M - 1 - \ln \mu_1\right).$$

Let $\tau_1 < \tau_2$ denote the two solutions of the following equation

$$1 - t + \ln t = -\kappa.$$

Then $0 < \tau_1 < 1 < \tau_2$. Using the above notation, we obtain

$$\gamma_1 = \frac{\tau_2}{e^{\frac{-\kappa}{2}}}$$

and

$$\gamma_2 = \left(e^{\frac{-\kappa}{2}}\right)^2.$$

In [6], some modified BFGS methods with global convergence and superlinear convergence for non-convex minimization problems have been given. It is easy to check that the methods satisfy (3.4) and (3.3). Thus, the global convergence of the methods given in [6] can be easy to obtained from Theorem 3.2. Under some conditions, we can prove the superlinear convergence of BFGSTWWP by using the similar way of in [6]. We do not repeat the proof here.

We concluded this section by proposing another formula which can be view as a cautious BFGS update. Let $\mu_1 \in (0,1)$ be a very small constant, define

$$\mu_{2k} = \begin{cases} -\mu_1 \frac{s_k^T s_k}{v_k^T s_k} & \text{if } v_k^T s_k \geq \mu_1 s_k^T s_k \\ 0 & \text{otherwise} \end{cases}$$

and

$$m_k = \mu_1 + \mu_{2k} \frac{v_k^T s_k}{s_k^T s_k}$$

Then, $m_k \in \{0, \mu_1\}$ It can be proved for the corresponding BFGS-type formula with any one of the line search strategies, that all the results in this section hold because for any $k, v_k^T s_k \geq \mu_1 s_k^T s_k$. hold even without any line search. Notice that if for some $k, v_k^T s_k \geq \mu_1 s_k^T s_k$, then the corresponding formula deduce to the ordinary BFGS update.

4. Numerical Experiments

In this section, we report the numerical results for the following three methods:

Algorithm 3.1: The Algorithm 3.1 with $\mu_1 = 10^{-3}$ and $\mu_2 = 10^{-10}$. Where $\eta_3 = 0.1, \sigma = 0.9$.

Algorithm 3.2: (3.1) and (3.7) with the Armijor line search, where $\mu_1 = 10^{-3}$ v, $Q = 1$, $\rho = 0.5$ and $\eta_2 = 0.1$. BFGS: The BFGS formula with the WWP. Where $\eta_3 = 0.1$, $\sigma = 0.9$.

For each test problem, the termination is

$$\left\|g\left(x_k\right)\right\| \leq 10^{-6}.$$

For each problem, we choose the initial matrix B1 = I, *i.e.*, the unit matrix. Due to the roundoff error, sometimes the directions generated by the algorithms may be not descent. We then used the steepest descent direction to take place of the related direction if $g_k^T d_k > -10^{-14}$. The detail numerical results are listed at: http://210.36.16.53:8018/publication.asp?id=46065.

In order to rank the iterative numerical methods, one can compute the total number of function and gradient evaluations by the formula

$$N_{\text{total}} = NF + m * NG, \tag{4.1}$$

where m is some integer. According to the results on automatic differentiation [12,13], the value of m can be set to $m = 5$. That is to say, one gradient evaluation is equivalent to m number of function evaluations if automatic differentiation is used. As we all known the BFGS method is considered to be the most efficient quasi-Newton method. Therefore, in this part, we compare the Algorithm 3.1 and Algorithm 3.2 with the BFGS method as follows: for each testing example *i*, compute the total numbers of function evaluations and gradient evaluations required by the evaluated method $j\left(EM\left(j\right)\right)$ and the S method by the formula (4.1), and denote them by $N_{\text{total},i}\left(EM\left(j\right)\right)$ and $N_{\text{total},i}\left(BFGS\right)$; then calculate the ratio

$$r_i\left(EM\left(j\right)\right) = \frac{N_{\text{total},i}\left(EM\left(j\right)\right)}{N_{\text{total},i}\left(BFGS\right)}$$

If $EM\left(j_0\right)$ does not work for example i_0, we replace the $N_{\text{total},i_0}\left(EM\left(j_0\right)\right)$ by a positive constant τ which define as follows

$$\tau = \max\left\{N_{\text{total},i}\left(EM\left(j\right)\right) : \left(i,j\right) \notin S_1\right\}$$

where

$$S_1 = \{\left(i,j\right) : \text{method } j \text{ does not work for example } i\}.$$

The geometric mean of these ratios for method $EM\left(j\right)$ over all the test problems is defined by

$$r\left(EM\left(j\right)\right) = \left(\prod_{i \in S} r_i\left(EM\left(j\right)\right)\right)^{1/|s|},$$

where *S* denotes the set of the test problems and |*S*| the number of elements in *S*. One advantage of the above rule is that, the comparison is relative and hence does not be dominated by a few problems for which the method requires a great deal of function evaluations and gradient

Table 1. Performance of these Algorithms.

Algorithm 3.1	Algorithm 3.2	BFGS
0.9534	1.5763	1

functions.

From **Table 1**, we observe that the Algorithm 3.1 outperforms the BFGS method. Therefore, the Algorithm 3.1 is the most efficient algorithm among quasi-Newton algorithms for solving minimization problems for the chosen tested problems.

5. 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*, Vol. 26, No. 3, 1989, pp. 727-739. doi:10.1137/0726042

[2] M. J. D. Powell, "Some Global Convergence Properties of a variable Metric Algorithm for Minimization without Exact Line Searches," In: R.W. Cottle and C. E. Lemke, Eds., *Nonlinear Programming*, *SIAM-AMS Proceedings*, Vol. 4, American Mathematical Society, Providence, 1976, pp.53-72.

[3] J. Werner, "Über die Globale Knovergenz von Variable-Metric-Verfahre mit Nichtexakter Schrittweitenbestimmung," *Numerische Mathematik*, Vol. 31, No. 3, 1978, pp. 321-334. doi:10.1007/BF01397884

[4] R. Byrd, J. Nocedal and Y. Yuan, "Global Convergence of a Class of Quasi-Newton Methods on Convex Problems," *SIAM Journal on Numerical Analysis*, Vol. 24, No. 5, 1987, pp. 1171-1189. doi:10.1137/0724077

[5] D. Li and M. Fukushima, "A Global and Superlinear Convergent Gauss-Newton-Based BFGS Method for Symmetric Nonlinear Equations," *SIAM Journal on Numerical Analysis*, Vol. 37, No. 1, 1999, pp. 152-172. doi:10.1137/S0036142998335704

[6] D. Li and M. Fukushima, "A Modified BFGS Method and Its Global Convergence in Nonconvex Minimization," *Journal of Computational and Applied Mathematics*, Vol. 129, No. 1-2, 2001, pp. 15-35. doi:10.1016/S0377-0427(00)00540-9

[7] D. Li and M. Fukushima, "On the Global Convergence of the BFGS Method for Nonconvex Unconstrained Optimization Problems," *SIAM Journal on Optimization*, Vol. 11, No. 4, 2001, pp. 1054-1064. doi:10.1137/S1052623499354242

[8] Z. Wei, G. Yu, G. Yuan and Z. Lian, "The Superlinear Convergence of a Modified BFGS-Type Method for Unconstrained Optimization," *Computational Optimization and Applications*, Vol. 29, No. 3, 2004, pp. 315-332. doi:10.1023/B:COAP.0000044184.25410.39

[9] G. Yuan and Z. Wei, "Convergence Analysis of a Modified BFGS Method on Convex Minimizations," *Computational Optimization and Applications*, Vol. 47, No. 2, 2010, pp. 237-255. doi:10.1007/s10589-008-9219-0

[10] E. G. Birgin and J. M. Martínez, "Structured Minimal-Memory Inexact Quasi-Newton Method and Secant Preconditioners for Augmented Lagrangian Optimization," *Computational Optimization and Applications*, Vol. 39, No. 1, 2008, pp. 1-16. doi:10.1007/s10589-007-9050-z

[11] G. Yuan and Z. Wei, "The Superlinear Convergence Analysis of a Nonmonotone BFGS Algorithm on Convex Objective Functions," *Acta Mathematics Sinica*, Vol. 24, No. 1, 2008, pp. 35-42. doi:10.1007/s10114-007-1012-y

[12] A. Griewank, "On Automatic Differentiation," In: M. Iri and K. Tanabe, Eds., *Mathematical Programming*: *Recent Developments and Application*s, Academic Publishers, Cambridge, 1989, pp. 84-108.

[13] Y. Dai and Q. Ni, "Testing Different Conjugate Gradient Methods for Large-Scale Unconstrained Optimization," *Journal of Computational Mathematics*, Vol. 21, No. 3, 2003, pp. 311-320.