Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 9, No. 2, pp. 373–389, April 2013

A GAUSS-NEWTON-BASED BFGS METHOD FOR SYMMETRIC NONLINEAR LEAST SQUARES PROBLEMS*

Weijun Zhou

Abstract: In this paper, based on the special structure of the Hessian of the problem, we propose a derivative-free BFGS method for symmetric nonlinear systems which may be inconsistent (i.e., nonzero residual). It is an extension of the Gauss-Newton-based BFGS method (GN-BFGS) presented by Li and Fukushima [SIAM J. Numer. Anal., 37 (1999), pp. 152-172] for symmetric nonlinear equations (i.e., zero residual case). The proposed method converges globally and superlinearly for both zero residual and nonzero residual problems by utilizing a new approximate norm descent line search. The preliminary numerical results are reported to show its efficiency for nonzero residual problems.

Key words: *symmetric nonlinear least squares, BFGS method, line search, global convergence, superlinear convergence*

Mathematics Subject Classification: *65K05, 90C30*

1 Introduction

The purpose of the paper is to extend the Gauss-Newton-based BFGS (GN-BFGS) method proposed by Li and Fukushima [8] for symmetric nonlinear equations

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

to the following symmetric nonlinear least squares problem

$$\min f(x) = \frac{1}{2}\|F(x)\|^2, \quad x \in R^n, \tag{1.2}$$

where $F : R^n \to R^n$ is a continuously differentiable mapping whose Jacobian $J(x) = F'(x)$ is symmetric, i.e., $J(x) = J(x)^T$. The problem (1.2) is called zero residual if $F(x^*) = 0$ for $x^* \in R^n$. Otherwise, it is called nonzero residual. The model (1.2) covers many practical problems such as data fitting, parameter estimation, the KKT system of unconstrained optimization problem and the saddle point problem [8, 11, 13, 14, 16].

The BFGS method is one of the most efficient quasi-Newton methods for solving optimization, nonlinear equations and nonlinear least squares [4, 6, 11, 16, 17]. In this paper, we mainly investigate derivative-free (i.e., without using the exact derivative) BFGS type methods, please see [7, 6, 10, 15] and references therein for other methods.

*This work was supported by the NSF foundation (10901026) of China, The Open Fund Project of Key Research Institute of Philosophies and Social Sciences in Hunan Universities, and The Key Project of the Scientific Research Fund of the Hunan Provincial Education Department.

ISSN 1348-9151

In the past four decades, there has been significant progress in the theoretical study on BFGS type methods for solving nonlinear equations, nonlinear least squares and nonlinear optimization, especially in the local convergence analysis, see the survey papers [10, 15], the book [12] and references therein.

However, the study on global convergence of derivative-free BFGS type methods with line searches for (1.2) is very fewer [7] (for trust region case, see [3]). To the authors's knowledge, so far, there mainly have been two globally convergent BFGS type methods for nonlinear equations (1.1). One is the Gauss-Newton-based BFGS (GN-BFGS) method proposed by Li and Fukushima [8], which globally converges by using an approximate norm descent line search and was extended to norm descent case by Gu etc. [5]. Another is the hyperplane projection BFGS method presented by Zhou and Li [17], which possesses global convergence under the condition that the system of nonlinear equations is monotone.

Since the GN-BFGS method is based on the Gauss-Newton method, it may inherit some disadvantages of the Gauss-Newton method, for examples, (i) It only uses the first order information of the underlying system, which may lead to its poor performance for large residual problems [1, 12]; (ii) Its global convergence requires the assumption that the Jacobian is uniformly nonsingular; (iii) Its iterative matrix is an approximation to the Gauss-Newton iterative matrix, which easily leads that the direction-finding subproblem is badly conditioned for some problems.

In this paper, by using the special structure of the Hessian of the problem (1.2), we propose a new derivative-free BFGS method with global and superlinear convergence, which can overcome some shortcomings of the GN-BFGS method above mentioned. The proposed method can be regarded as an extension of the GN-BFGS method from zero residual case to nonzero residual case.

The paper is organized as follows. In the next section, we present the new BFGS method and prove its global convergence using a new approximate norm descent line search. In Section 3, we show its superlinear convergence. In Section 4, we discuss some extensions of the proposed method. In Section 5, we report some preliminary numerical results. In Section 6, we discuss some conclusions.

2 Algorithm and Global Convergence

In this section, we first simply recall the GN-BFGS method [8] for nonlinear equations (1.1). Throughout the paper, we denote

$$F_k = F(x_k), \quad J_k = J(x_k), \quad s_k = x_{k+1} - x_k.$$

At each iteration, the GN-BFGS method [8] produces the search direction d_k by the linear equations

$$B_k d = -\tilde{g}_k,$$

where

$$\tilde{g}_k = \frac{F(x_k + \alpha_{k-1} F_k) - F_k}{\alpha_{k-1}}, \tag{2.1}$$

α_{k-1} is the stepsize given by the following approximate norm descent line search, that is, $\alpha_k = \max\{1, r^1, r^2, \cdots\}$ with $\alpha = r^i$ satisfying

$$f(x_k + \alpha d_k) - f(x_k) \le -\sigma_1 \|\alpha d_k\|^2 - \sigma_2 \alpha^2 f(x_k) + \epsilon_k f(x_k), \tag{2.2}$$

where σ_1 and σ_2 are two given positive constants, and $\{\epsilon_k\}$ is given by (2.10) below. Moreover, the matrix B_k is updated by the BFGS formula

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

where

$$\gamma_k = F(x_k + \delta_k) - F_k, \quad \delta_k = F_{k+1} - F_k. \tag{2.3}$$

It is easy to see that if $\|s_k\|$ is small, then by symmetry, we have

$$B_{k+1} s_k = \gamma_k \approx J_{k+1}^T J_{k+1} s_k, \tag{2.4}$$

which is the reason why this method is called the Gauss-Newton-based BFGS method.

It is clear that the gradient and the Hessian of the problem (1.2) are given by

$$\nabla f(x) = J(x)^T F(x), \quad \nabla^2 f(x) = J(x)^T J(x) + \sum_{i=1}^{n} F_i(x) \nabla^2 F_i(x), \tag{2.5}$$

where $F = (F_1, \cdots, F_n)^T$ is twice continuously differentiable. Let x_k be the current iterate, by [12, (7.5.5) on Page 379], we know

$$\nabla^2 f(x_{k+1}) s_k \approx \hat{y}_k \triangleq J_{k+1}^T J_{k+1} s_k + (J_{k+1} - J_k)^T F_{k+1}. \tag{2.6}$$

Denote

$$\bar{g}_{k+1} \triangleq \frac{F(x_{k+1} + \|s_k\|^2 F_{k+1}) - F_{k+1}}{\|s_k\|^2}, \quad \hat{g}_{k+1} \triangleq \frac{F(x_k + \|s_k\|^2 F_{k+1}) - F_k}{\|s_k\|^2}. \tag{2.7}$$

Then if $\|s_k\|$ is small, we have

$$\nabla f(x_{k+1}) = J_{k+1}^T F_{k+1} \approx \bar{g}_{k+1}, \quad J_k^T F_{k+1} \approx \hat{g}_{k+1}. \tag{2.8}$$

Therefore from (2.6), (2.4), (2.7) and (2.8), we obtain

$$\nabla^2 f(x_{k+1}) s_k \approx \hat{y}_k \approx \gamma_k + \bar{g}_{k+1} - \hat{g}_{k+1} \triangleq z_k. \tag{2.9}$$

Let $\{\epsilon_k\}$ and ϵ be a given positive sequence and a positive constant satisfying

$$\sum_{k=0}^{\infty} \epsilon_k \leq \epsilon < \infty. \tag{2.10}$$

Define

$$g_k \triangleq \frac{F(x_k + \epsilon_k F_k) - F_k}{\epsilon_k}. \tag{2.11}$$

Then we have $g_k \approx \nabla f(x_k)$. Let μ be a given positive constant and define

$$y_k \triangleq z_k + \Big(\max\{0, -\frac{z_k^T s_k}{\|s_k\|^2}\} + \mu \|g_k\| \Big) s_k. \tag{2.12}$$

Then we obtain a new BFGS update formula

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

and $B_{k+1}s_k = y_k \approx \nabla^2 f(x_{k+1})s_k$, which shows that B_{k+1} is a good approximation to the Hessian $\nabla^2 f(x_{k+1})$.

The reason of using y_k instead of z_k in (2.13) is to guarantee positive definiteness of the iterative matrix sequence $\{B_k\}$, where we utilize the MBFGS regularized technique proposed by Li and Fukushima [9]. It is clear that $y_k^T s_k > 0$, which is sufficient to guarantee positive definiteness of B_{k+1} as long as B_k is positive definite [12].

Now we present the new derivative-free BFGS method for (1.2) as follows, where we use an approximate norm descent line search to globalize the proposed method.

Algorithm 2.1. [The derivative-free BFGS method with nonmonotone line search]

Step 0. Choose a starting point $x_0 \in R^n$, an initial symmetric positive definite matrix $B_0 \in R^{n\times n}$ and several constants $\sigma \in (0, \frac{1}{2})$, $\mu > 0$ and $r \in (0, 1)$. Let $k := 0$.

Step 1. Compute d_k by solving the following linear equations

$$B_k d = -g_k, \tag{2.14}$$

where g_k is defined by (2.11).

Step 2. Compute $\alpha_k = \max\{1, r^1, r^2, \cdots\}$ with $\alpha = r^i$ satisfying

$$f(x_k + \alpha d_k) - f(x_k) \le \sigma\alpha g_k^T d_k + \epsilon_k f(x_k), \tag{2.15}$$

where $\{\epsilon_k\}$ is given by (2.10).

Step 3. Set $x_{k+1} = x_k + \alpha_k d_k$. Update B_k by the BFGS formula (2.13), where y_k is given by (2.12).

Step 4. Let $k := k + 1$ and go to Step 1.

Remark 2.2. (i) In (2.14), we use g_k instead of $\tilde{g}_k$ since $\tilde{g}_k$ defined by (2.1) is no longer a good approximation to $\nabla f(x_k)$ in the nonzero residual case.

(ii) The line search (2.15) is motivated by (2.2) proposed by Li and Fukushima [8], which is an approximate norm descent line search. Although the search direction d_k may be not a descent direction of f, the line search (2.15) is well defined. In fact, as $\alpha \to 0^+$, the left-hand side of (2.15) goes to zero, while the right-hand side tends to the positive term $\epsilon_k f(x_k)$, which shows that (2.15) is satisfied for all sufficiently small $\alpha > 0$.

(iii) Moreover, the line search (2.15) is different from (2.2) since the latter is only suitable for zero residual problems. For example, the line search (2.2) implies that

$$\lim_{k\to\infty} \alpha_k \|F_k\| = 0. \tag{2.16}$$

Assume that the GN-BFGS method converges superlinearly for nonzero residual problem(that is, $x_k \to x^*$ and $\|F(x^*)\| > 0$), which implies that $\alpha_k \equiv 1$ for large k, then (2.16) yields that $0 < \|F(x^*)\| = \lim_{k\to\infty} \|F_k\| = 0$. This is a contradiction, which shows that (2.2) will undermine superlinear convergence property of this method.

Define the level set

$$\Omega = \{x|\ f(x) \le e^{\epsilon} f(x_0)\}, \tag{2.17}$$

where ϵ is a positive constant satisfying (2.10). Then by the same argument of Lemma 2.1 in [8], we have the following result.

Lemma 2.3. *Let the sequence $\{x_k\}$ be generated by Algorithm 2.1. Then the sequence $\{f(x_k)\}$ converges and $x_k \in \Omega$ for all $k \geq 0$.*

Now we make some assumptions for global convergence of Algorithm 2.1 as follows.

Assumption 2.4. (i) The level set Ω is bounded.

(ii) In some neighbourhood Ω_1 of Ω, $J(x)$ is Lipschitz continuous, that is, there exists a positive constant L such that

$$\|J(x) - J(y)\| \leq L\|x - y\|, \quad \forall x, y \in \Omega_1. \tag{2.18}$$

Assumption 2.4 yields that there exist positive constants M_1, M_2 and L_1 such that

$$\|J(x)\| \leq M_1, \quad \|F(x)\| \leq M_2, \quad \forall x \in \Omega_1, \tag{2.19}$$

$$\|F(x) - F(y)\| \leq L_1\|x - y\|, \quad \forall x, y \in \Omega_1. \tag{2.20}$$

Lemma 2.5. *Let the sequence $\{x_k\}$ be generated by Algorithm 2.1. Then we have*

$$\sum_{k=0}^{\infty} \alpha_k g_k^T d_k < \infty.$$

Proof. It follows from (2.15) and (2.10) directly. □

Lemma 2.6. *Let Assumption 2.4 hold and the sequence $\{x_k\}$ be generated by Algorithm 2.1. Then there exists a positive constant C_1 such that*

$$\mu\|g_k\|\|s_k\|^2 \leq y_k^T s_k \leq C_1\|s_k\|^2.$$

Proof. From (2.12), we have

$$y_k^T s_k = z_k^T s_k + \Big(\max\{0, -\frac{z_k^T s_k}{\|s_k\|^2}\} + \mu\|g_k\|\Big)\|s_k\|^2 \geq \mu\|g_k\|\|s_k\|^2.$$

By (2.3) and (2.20), we know

$$\|\gamma_k\| \leq L_1\|\delta_k\| = L_1\|F_{k+1} - F_k\| \leq L_1^2\|s_k\|.$$

From (2.7), (2.19) and (2.18), we have

$$\begin{aligned}
& \|\bar{g}_{k+1} - \hat{g}_{k+1}\| \\
= \;& \Big\| \int_0^1 J(x_{k+1} + t\|s_k\|^2 F_{k+1})F_{k+1}dt - \int_0^1 J(x_k + t\|s_k\|^2 F_{k+1})F_{k+1}dt \Big\| \\
\leq \;& \|F_{k+1}\| \int_0^1 \big\|J(x_{k+1} + t\|s_k\|^2 F_{k+1}) - J(x_k + t\|s_k\|^2 F_{k+1})\big\|dt \\
\leq \;& M_2 L\|s_k\|.
\end{aligned}$$

Then from (2.12), (2.9), (2.3), (2.7), (2.19) and (2.20), we deduce that

$$\begin{aligned}
y_k^T s_k \;&\leq\; \|y_k\|\|s_k\| \\
&\leq\; \big(2\|z_k\| + \mu\|g_k\|\|s_k\|\big)\|s_k\| \\
&\leq\; \big(2\|\gamma_k\| + 2\|\bar{g}_{k+1} - \hat{g}_{k+1}\| + \mu\|g_k\|\|s_k\|\big)\|s_k\| \\
&\leq\; C_1\|s_k\|^2
\end{aligned}$$

holds for some positive constant C_1. □

If there exists a positive constant τ_1 such that $\|g_k\| \geq \tau_1$ holds for all large k. Then from Lemma 2.6, we have the following important result, which is the key to prove the global convergence of the proposed method.

Lemma 2.7 ([2, Theorem 2.1]). *Let Assumption 2.4 hold. Then there are positive constants $\beta_i, i = 1, 2, 3, 4$ such that*

$$\beta_1\|s_i\| \leq \|B_i s_i\| \leq \beta_2\|s_i\|, \quad \beta_3\|s_i\|^2 \leq s_i^T B_i s_i \leq \beta_4\|s_i\|^2 \tag{2.21}$$

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

Lemma 2.7 implies that

$$\beta_1\|d_i\| \leq \|B_i d_i\| \leq \beta_2\|d_i\|, \quad \beta_3\|d_i\|^2 \leq d_i^T B_i d_i \leq \beta_4\|d_i\|^2 \tag{2.22}$$

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

Lemma 2.8. *Let Assumption 2.4 hold and the sequence $\{x_k\}$ be generated by Algorithm 2.1. If $\alpha_k \neq 1$, then we have*

$$\alpha_k \geq \frac{-(1-\sigma)g_k^T d_k - L\|F_k\|^2\epsilon_k\|d_k\|}{C_2\|d_k\|^2}, \tag{2.23}$$

where C_2 is a positive constant.

Proof. Let

$$G_k = \int_0^1 J(x_k + t\epsilon_k F_k)dt.$$

Then from (2.11) and (2.14), we have

$$g_k = G_k F_k = -B_k d_k. \tag{2.24}$$

From the line search (2.15), we get that $\alpha_k' \triangleq \alpha_k/r$ does not satisfy (2.15), i.e.,

$$f(x_k + \alpha_k' d_k) - f(x_k) > \sigma\alpha_k' g_k^T d_k + \epsilon_k f(x_k) > \sigma\alpha_k' g_k^T d_k,$$

which means that

$$\begin{aligned}
-\sigma\alpha_k' g_k^T d_k &> -\big(f(x_k + \alpha_k' d_k) - f(x_k)\big) \\
&= -\frac{1}{2}\big(F(x_k + \alpha_k' d_k) + F_k\big)^T\big(F(x_k + \alpha_k' d_k) - F_k\big) \\
&= -\frac{1}{2}\Big(2F_k^T\big(F(x_k + \alpha_k' d_k) - F_k\big) + \|F(x_k + \alpha_k' d_k) - F_k\|^2\Big) \\
&\geq -F_k^T\big(F(x_k + \alpha_k' d_k) - F_k\big) - \frac{L_1^2}{2}\|\alpha_k' d_k\|^2,
\end{aligned} \tag{2.25}$$

where the last inequality uses (2.20). From (2.24) and the symmetry of the Jacobian, we have

$$\begin{aligned}
& F_k^T\big(F(x_k + \alpha_k' d_k) - F_k\big) \\
= \;& \alpha_k' F_k^T \int_0^1 J(x_k + t\alpha_k' d_k)d_k dt \\
= \;& \alpha_k' F_k^T G_k d_k + \alpha_k' F_k^T \int_0^1 \Big(J(x_k + t\alpha_k' d_k) - J(x_k + t\epsilon_k F_k)\Big)d_k dt \\
\leq \;& \alpha_k' g_k^T d_k + \alpha_k'\|F_k\|\|d_k\|L\big(\alpha_k'\|d_k\| + \epsilon_k\|F_k\|\big) \\
\leq \;& \alpha_k' g_k^T d_k + LM_2\|\alpha_k' d_k\|^2 + L\alpha_k'\|F_k\|^2\epsilon_k\|d_k\|,
\end{aligned} \tag{2.26}$$

where we use (2.18) in the first inequality, and the last inequality follows from (2.19).

The inequality (2.26) together with (2.25) yield that

$$\alpha_k \geq \frac{r}{L_1^2/2 + LM_2} \frac{-(1-\sigma)g_k^T d_k - L\|F_k\|^2 \epsilon_k \|d_k\|}{\|d_k\|^2}.$$

Set $C_2 = \frac{L_1^2/2+LM_2}{r}$, then we get (2.23). The proof is completed. □

Now it is time for us to give the global convergence result for Algorithm 2.1.

Theorem 2.9. *Let Assumption 2.4 hold and the sequence $\{x_k\}$ be generated by Algorithm 2.1. Then we have*

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{2.27}$$

Proof. We prove the theorem by contradiction. If (2.27) is not true, then there exist an integer $\hat{k}$ and a positive constant τ such that

$$\|\nabla f(x_k)\| = \|J_k^T F_k\| \geq \tau, \forall k \geq \hat{k}, \tag{2.28}$$

which implies that

$$\|F_k\| \geq \tau_1 \quad \text{and} \quad \|g_k\| \geq \tau_1 \tag{2.29}$$

hold for sufficiently large k with some positive constant τ_1. Define the index set

$$T = \{i| \text{ (2.22) holds}\}. \tag{2.30}$$

Then for $k \in T$, from (2.22), (2.14), (2.11) and (2.19), we have

$$\begin{aligned} \|d_k\| &\leq \beta_1^{-1}\|B_k d_k\| = \beta_1^{-1}\|g_k\| \\ &= \beta_1^{-1}\Big\|\int_0^1 J(x_k + t\epsilon_k F_k)F_k dt\Big\| \\ &\leq \beta_1^{-1} M_1 M_2. \end{aligned} \tag{2.31}$$

(i) If $\liminf_{k\in T, k\to\infty} \|d_k\| = 0$, then we deduce from (2.14) and (2.22) that

$$\liminf_{k\in T, k\to\infty} \|g_k\| \leq \beta_2 \liminf_{k\in T, k\to\infty} \|d_k\| = 0,$$

which implies that $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0$ since $J_k = J_k^T$, $\{\|F_k\|\}$ is bounded, $\lim_{k\to\infty} \epsilon_k = 0$, and

$$\begin{aligned} \|\nabla f(x_k)\| &= \|J_k F_k\| = \Big\|g_k - \int_0^1 (J(x_k + t\epsilon_k F_k) - J_k)F_k dt\Big\| \\ &\leq \|g_k\| + L\epsilon_k\|F_k\|^2. \end{aligned}$$

This contradicts to (2.28).

(ii) If $\liminf_{k\in T, k\to\infty} \|d_k\| > 0$. Then there exists a positive constant τ_2 such that

$$\|d_k\| \geq \tau_2, \quad k \in T. \tag{2.32}$$

From Lemma 2.5, we have

$$\lim_{k\to\infty} \alpha_k g_k^T d_k = 0, \tag{2.33}$$

which together with (2.22), (2.32) and (2.33) yields

$$\lim_{k\in T, k\to\infty} \alpha_k = 0. \tag{2.34}$$

From Lemma 2.8, (2.10), (2.32), (2.22) and (2.19), we have, for large $k \in T$,

$$\begin{aligned}
\alpha_k &\geq \frac{-(1-\sigma)g_k^T d_k - L\|F_k\|^2\epsilon_k\|d_k\|}{C_2\|d_k\|^2} \\
&= \frac{(1-\sigma)d_k^T B_k d_k}{C_2\|d_k\|^2} - \frac{L\|F_k\|^2\epsilon_k}{C_2\|d_k\|} \\
&\geq \frac{\beta_3(1-\sigma)}{C_2} - \frac{L\|F_k\|^2\epsilon_k}{C_2\|d_k\|} \\
&\geq \frac{\beta_3(1-\sigma)}{2C_2},
\end{aligned}$$

which contradicts to (2.34). This finishes the proof. □

3 Superlinear Convergence

To prove superlinear convergence of Algorithm 2.1, we need the following assumptions.

Assumption 3.1. (I) The sequence $\{x_k\}$ converges to x^*, where $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite.

(II) In some neighourhood Ω_2 of x^*, $\nabla^2 F_i$, $i = 1, 2, \cdots, n$, are Lipschitz continuous, i.e., there exists a positive constant L_2 satisfying

$$\|\nabla^2 F_i(x) - \nabla^2 F_i(y)\| \leq L_2\|x - y\|, \quad i = 1, 2, \cdots, n, \forall x, y \in \Omega_2, \tag{3.1}$$

where $F = (F_1, \cdots, F_n)^T$ is twice continuously differentiable.

(III) The positive sequence $\{\epsilon_k\}$ satisfies

$$\epsilon_k \leq \rho_0^k(f(x_k) - f(x^*)),$$

where $\rho_0 \in (0, 1)$ is a constant.

Remark 3.2. (i) The condition (I) in Assumption 3.1 implies that x^* is a strongly local minimizer.

(ii) It is clear (see [2, 12]) that conditions (I)-(II) in Assumption 3.1 yield that there exists positive constants L_3, m and M such that

$$\|F(x) - F(y)\| \leq L_3\|x - y\|, \quad \|J(x) - J(y)\| \leq L_3\|x - y\|, \forall x, y \in \Omega_2, \tag{3.2}$$

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L_3\|x - y\|, \quad \forall x, y \in \Omega_2, \tag{3.3}$$

and

$$\frac{1}{2}m\|x_k - x^*\|^2 \leq f(x_k) - f(x^*) \leq \frac{1}{m}\|\nabla f(x_k)\|^2, \tag{3.4}$$

$$\|F(x)\| \leq M, \quad \|J(x)\| \leq M, \quad \forall x \in \Omega_2. \tag{3.5}$$

(iii) The condition (III) in Assumption 3.1 is used to ensure the approximate precision of g_k although it seems a little strong. And this assumption implies that (2.10) still holds.

Now we begin to prove that the Dennis-Moŕe condition holds step by step. We first give the following very useful lemma.

Lemma 3.3. *Let Assumption 3.1 hold. Then we have*

$$\sum_{k=0}^{\infty} \|x_k - x^*\| < \infty. \tag{3.6}$$

Proof. From the line search (2.15) and (2.22), for $k \in T$, we have that

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq -\frac{(1-\sigma)}{C_2}\frac{(g_k^T d_k)^2}{\|d_k\|^2} + \frac{L}{C_2}\frac{|g_k^T d_k|}{\|d_k\|}\epsilon_k\|F_k\|^2 + \epsilon_k f(x_k) \\ &\leq -\frac{(1-\sigma)}{C_2}\frac{\beta_3^2}{\beta_2^2}\|g_k\|^2 + \frac{2LM^2}{C_2}\epsilon_k f(x_k) + \epsilon_k f(x_k) \\ &\triangleq -C_4\|g_k\|^2 + C_5\epsilon_k \end{aligned} \tag{3.7}$$

holds for some positive constants C_4 and C_5, where we use $\|g_k\| = \| \int_0^1 J(x_k + t\epsilon_k F_k)F_k dt\| \leq M^2$ in the second inequality. (3.2) and (3.5) mean that

$$\|\nabla f(x_k) - g_k\| = \left\| \int_0^1 \big(J(x_k + t\epsilon_k F_k) - J_k\big)F_k dt \right\| \leq L\epsilon_k\|F_k\|^2 \leq C_6\epsilon_k \tag{3.8}$$

holds for some positive constant C_6.

Since (3.4) holds for any sufficiently small $m > 0$ and the positive constant C_4 is independent of m, without loss of generality, we can choose $m > 0$ such that $\frac{C_4 m}{2} < 1$. Then (3.8) together with (3.4) yields that

$$f(x_k) - f(x^*) \leq \frac{2}{m}\|g_k\|^2 + \frac{2}{m}\|\nabla f(x_k) - g_k\|^2 \leq \frac{2}{m}\|g_k\|^2 + \frac{2C_6^2}{m}\epsilon_k^2. \tag{3.9}$$

From (3.7) and (3.9), for $k \in T$, we obtain that

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq -\frac{C_4 m}{2}\big(f(x_k) - f(x^*)\big) + 2C_4C_6^2\epsilon_k^2 + C_5\epsilon_k \\ &\leq -\frac{C_4 m}{2}\big(f(x_k) - f(x^*)\big) + C_7\epsilon_k \end{aligned}$$

holds for some positive constant C_7, which implies that

$$f(x_{k+1}) - f(x^*) \leq \rho_1\big(f(x_k) - f(x^*)\big) + C_7\epsilon_k, \tag{3.10}$$

where the positive constant $\rho_1 = 1 - \frac{C_4 m}{2} < 1$. From (III) in Assumption 3.1, without loss of generality, we can assume that $\rho_1 + C_7\rho_0^k \leq \rho_2 < 1$ for some positive constant ρ_2 since $\rho_0, \rho_1 \in (0, 1)$. Then we have from (3.10) and (III) in Assumption 3.1

$$f(x_{k+1}) - f(x^*) \leq (\rho_1 + C_7\rho_0^k)\big(f(x_k) - f(x^*)\big) \leq \rho_2\big(f(x_k) - f(x^*)\big), \tag{3.11}$$

where $\rho_2 < 1$.

Moreover, from (2.15), we have

$$f(x_{k+1}) - f(x^*) \leq (1 + \epsilon_k)\big(f(x_k) - f(x^*)\big) + \epsilon_k f(x^*), \tag{3.12}$$

which together with (III) in Assumption 3.1 implies

$$f(x_{k+1}) - f(x^*) \le (1+\omega_k)\big(f(x_k) - f(x^*)\big), \tag{3.13}$$

where $\omega_k \triangleq \epsilon_k + \rho_0^k f(x^*)$. It is clear that

$$\sum_{k=0}^{\infty} \omega_k \le \omega < \infty$$

holds for some positive constant ω.

From (3.11) and (3.13), we have that

$$\begin{aligned} f(x_{k+1}) - f(x^*) &\le \left(\prod_{i\in\{1,2,\cdots,k\}\setminus T} (1+\omega_i)\right) \prod_{i\in T} \rho_2\big(f(x_0) - f(x^*)\big) \\ &\le \Big(\prod_{i=1}^{k}(1+\omega_i)\Big)\rho_2^{\frac{k}{2}}\big(f(x_0) - f(x^*)\big) \\ &\le e^{\omega}\rho_2^{\frac{k}{2}}\big(f(x_0) - f(x^*)\big). \end{aligned} \tag{3.14}$$

(3.14) and (3.4) yield that

$$\begin{aligned} \sqrt{m/2}\sum_{k=0}^{\infty}\|x_k - x^*\| &\le \sum_{k=0}^{\infty}\sqrt{f(x_{k+1}) - f(x^*)} \\ &\le \sum_{k=0}^{\infty}\sqrt{e^{\omega}(f(x_0) - f(x^*))}(\rho_2^{\frac{1}{4}})^k \\ &< \infty, \end{aligned}$$

which gives (3.6). This finishes the proof. □

From conditions (I) and (II), it is easily to get the following result [12].

Lemma 3.4. *Let conditions (I) and (II) in Assumption 3.1 hold. Then*

$$\frac{\|\big(\nabla f(x_{k+1}) - \nabla f(x_k)\big) - \nabla^2 f(x^*)s_k\|}{\|s_k\|} \le M_3\{\|x_{k+1} - x^*\| + \|x_k - x^*\|\} \tag{3.15}$$

holds for some positive constant M_3.

From (2.9) and the positive definiteness of $\nabla^2 f(x^*)$, we have

$$z_k^T s_k > 0, \tag{3.16}$$

which implies that, for large k,

$$y_k = z_k + \mu\|g_k\|s_k.$$

Then we have

$$\begin{aligned} \frac{\|y_k - \nabla^2 f(x^*)s_k\|}{\|s_k\|} &\le \frac{\|z_k - \big(\nabla f(x_{k+1}) - \nabla f(x_k)\big)\|}{\|s_k\|} \\ &\quad + \frac{\|\big(\nabla f(x_{k+1}) - \nabla f(x_k)\big) - \nabla^2 f(x^*)s_k\|}{\|s_k\|} + \frac{\mu\|g_k\|\|s_k\|}{\|s_k\|} \\ &\triangleq A_1 + A_2 + A_3. \end{aligned}$$

From (3.8) and (3.2), we have

$$A_3 \le \mu\|\nabla f(x_k)\| + \mu\|g_k - \nabla f(x_k)\| \le \mu L_3\|x_k - x^*\| + C_6\epsilon_k. \tag{3.17}$$

Now we estimate A_1. From (2.9), we have

$$\begin{aligned} A_1 &= \frac{\|z_k - (J_{k+1}F_{k+1} - J_kF_k)\|}{\|s_k\|} \\ &= \frac{\|z_k - (J_{k+1}F_{k+1} - J_kF_{k+1}) - J_k(F_{k+1} - F_k)\|}{\|s_k\|} \\ &= \frac{\|\gamma_k - J_k(F_{k+1} - F_k) + (\bar{g}_{k+1} - \hat{g}_{k+1}) - (J_{k+1}F_{k+1} - J_kF_{k+1})\|}{\|s_k\|} \\ &\le \frac{\|\gamma_k - J_k(F_{k+1} - F_k)\|}{\|s_k\|} + \frac{\|\bar{g}_{k+1} - J_{k+1}F_{k+1}\|}{\|s_k\|} + \frac{\|\hat{g}_{k+1} - J_kF_{k+1}\|}{\|s_k\|} \\ &\triangleq B_1 + B_2 + B_3. \end{aligned}$$

Then from (2.3), (3.2) and (3.5), we deduce that

$$\begin{aligned} B_1 &= \frac{\|\big(F(x_k + \delta_k) - F_k\big) - J_k\delta_k\|}{\|s_k\|} \\ &= \frac{\|\int_0^1 \big(J(x_k + t\delta_k) - J_k\big)\delta_k dt\|}{\|s_k\|} \\ &\le \frac{L_3\|\delta_k\|^2}{\|s_k\|} \le L_3^3\|s_k\|, \end{aligned}$$

$$\begin{aligned} B_2 &= \frac{\big\|\int_0^1 \big(J(x_{k+1} + t\|s_k\|^2F_{k+1}) - J_{k+1}\big)F_{k+1}dt\big\|}{\|s_k\|} \\ &\le L_3\|F_{k+1}\|^2\|s_k\| \le L_3M^2\|s_k\|, \end{aligned}$$

and

$$\begin{aligned} B_3 &= \frac{\|\int_0^1 \big(J(x_k + t\|s_k\|^2F_{k+1}) - J_k\big)F_{k+1}dt\|}{\|s_k\|} \\ &\le L\|F_{k+1}\|^2\|s_k\| \le L_3M^2\|s_k\|. \end{aligned}$$

The above three inequalities imply that there exists a positive constant M_4 such that

$$A_1 \le M_4\|s_k\| = M_4\|x_{k+1} - x_k\| \le M_4(\|x_{k+1} - x^*\| + \|x_k - x^*\|).$$

Therefore we have

$$\frac{\|y_k - \nabla^2 f(x^*)s_k\|}{\|s_k\|} \le M_5\eta_k, \tag{3.18}$$

where M_5 is some positive constant and

$$\eta_k = \|x_{k+1} - x^*\| + \|x_k - x^*\| + \epsilon_k.$$

Then from (3.6) and (2.10), we have

$$\sum_{k=0}^{\infty} \eta_k < \infty,$$

which implies that the following Dennis-Móre condition holds.

Lemma 3.5 ([2, Theorem 3.2]). *Let Assumption 3.1 hold. Then we have*

$$\lim_{k\to\infty}\frac{\|\big(B_k-\nabla^2 f(x^*)\big)s_k\|}{\|s_k\|}=0. \tag{3.19}$$

Moreover, the sequences $\{B_k\}$ *and* $\{B_k^{-1}\}$ *are uniformly bounded.*

From (2.14) and Lemma 3.5, we get

$$\|d_k\|=\|B_k^{-1}g_k\|\le\|B_k^{-1}\|\|g_k\|\to 0.$$

This inequality together with the mean value theorem, (2.14) and (3.8) yields that there exists a constant $\theta\in(0,1)$ such that

$$\begin{aligned}
& f(x_k+d_k)-f(x_k)-\sigma g_k^T d_k\\
= \;& \nabla f(x_k)^T d_k+\frac{1}{2}d_k^T\nabla^2 f(x_k+\theta s_k)d_k-\sigma g_k^T d_k\\
= \;& (\frac{1}{2}-\sigma)g_k^T d_k+\big(\nabla f(x_k)-g_k\big)^T d_k+\frac{1}{2}g_k^T d_k+\frac{1}{2}d_k^T\nabla^2 f(x_k+\theta s_k)d_k\\
= \;& (\frac{1}{2}-\sigma)g_k^T d_k+\big(\nabla f(x_k)-g_k\big)^T d_k\\
& +\frac{1}{2}d_k^T\big(\nabla^2 f(x^*)-B_k\big)d_k+\frac{1}{2}d_k^T\big(\nabla^2 f(x_k+\theta s_k)-\nabla^2 f(x^*)\big)d_k\\
= \;& (\frac{1}{2}-\sigma)g_k^T d_k+\big(\nabla f(x_k)-g_k\big)^T d_k+o(\|d_k\|^2)\\
= \;& -(\frac{1}{2}-\sigma)d_k^T B_k d_k+o(\epsilon_k f(x_k))+o(\|d_k\|^2)\\
\le \;& \epsilon_k f(x_k),
\end{aligned}$$

where we use the fact that $\sigma\in(0,\frac{1}{2})$ and Lemma 3.5. From this inequality and (2.15), we have that the unit steplength is always accepted for sufficiently large k, i.e., $\alpha_k\equiv 1$ for large k. This conclusion together with Lemma 3.5 implies the superlinear convergence of Algorithm 2.1.

Theorem 3.6 ([12, Theorem 5.4.6]). *Let Assumption 3.1 hold. Then the sequence* $\{x_k\}$ *be generated by Algorithm 2.1 converges superlinearly, that is,*

$$\lim_{k\to\infty}\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|}=0.$$

4 Extension to Norm Descent Case

From the line search (2.15) in Algorithm 2.1, we see that the sequence $\{f(x_k)\}$ is not necessarily a descent sequence. In this section, we extend Algorithm 2.1 to norm descent case in the sense that $f(x_{k+1})<f(x_k), k=0,1,2,\dots$ by utilizing the technique proposed by Gu eta. [5].

Let $\alpha>0$ be a parameter,

$$g_k(\alpha)\triangleq\frac{F(x_k+\alpha\|s_{k-1}\|F_k)-F_k}{\alpha\|s_{k-1}\|} \tag{4.1}$$

and $d_k(\alpha)$ be the solution of the following nonlinear equations

$$B_k d = -g_k(\alpha). \tag{4.2}$$

Suppose that B_k is positive definite. Now we give a way to determine the search direction d_k as follows, which is motivated by that of [5].

Procedure 1. Let $r \in (0,1)$ and $\sigma > 0$ be given two constants. Let i_k be the smallest nonnegative integer such that the following inequality holds with $\alpha = r^i$, $i = 0, 1, \ldots,$

$$f(x_k + \alpha d_k(\alpha)) - f(x_k) \le -\sigma \|\alpha d_k(\alpha)\|^2. \tag{4.3}$$

Then let

$$d_k = d_k(r^{i_k}).$$

Moreover, if $i_k = 0$, we take

$$\alpha_k = 1.$$

It is clear that (4.3) is well defined, that is, (4.3) terminates finitely. Otherwise

$$0 > -\nabla f(x_k)^T B_k \nabla f(x_k) = \lim_{\alpha \to 0^+} \frac{f(x_k + \alpha d_k(\alpha)) - f(x_k)}{\alpha} \ge -\lim_{\alpha \to 0^+} \sigma \alpha \|d_k(\alpha)\|^2 = 0,$$

which is a contradiction.

If $i_k > 0$, we give the following procedure to compute the stepsize.

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

$$f(x_k + r^{i_k - j} d_k) - f(x_k) \le -\sigma \|r^{i_k - j} d_k\|^2. \tag{4.4}$$

Then let

$$\alpha_k = r^{i_k - j_k}.$$

Procedures 1 and 2 present a way to produce d_k and α_k simultaneously. It is easy to see that if $\alpha_k \neq 1$, then $\alpha'_k = \frac{\alpha_k}{r}$ does not satisfy (4.3), that is,

$$f(x_k + \alpha'_k d_k) - f(x_k) > -\sigma \|\alpha'_k d_k\|^2. \tag{4.5}$$

Now we give the complete norm descent BFGS method as follows.

Algorithm 4.1 (The derivative-free BFGS method with norm descent)**.**

Step 0. Choose a starting point $x_0 \in R^n$, an initial symmetric positive definite matrix $B_0 \in R^{n \times n}$ and several constants $\sigma > 0$, $\|s_{-1}\| > 0$, $\mu > 0$ and $r \in (0,1)$. Let $k := 0$.

Step 1. Compute d_k and α_k by Procedures 1 and 2.

Step 2. Set $x_{k+1} = x_k + \alpha_k d_k$. Update B_k by the BFGS formula (2.13), where y_k is given by (2.12) with $g_k = g_k(r^{i_k})$.

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

Similar to Algorithm 2.1, we can discuss the convergence properties of Algorithm 4.1. For simplicity, we only present the global convergence of Algorithm 4.1 under Assumption 2.4 here.

Theorem 4.2. *Let Assumption 2.4 hold. Then the sequence $\{x_k\}$ generated by Algorithm 4.1 converges globally in the sense that* $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0$.

Proof. We prove the theorem by contradiction. If the conclusion is not true, then (2.28) and (2.29) hold. From Procedures 1 and 2, we have

$$\sum_{k=0}^{\infty} \|\alpha_k d_k\|^2 = \sum_{k=0}^{\infty} \|s_k\|^2 < \infty, \tag{4.6}$$

which implies that

$$\lim_{k\to\infty} \alpha_k d_k = \lim_{k\to\infty} \|s_k\| = 0. \tag{4.7}$$

(i) If $\liminf_{k\in T, k\to\infty} \|d_k\| = 0$, then by (4.2) and (2.22), we have

$$\liminf_{k\in T, k\to\infty} \|g_k\| = \liminf_{k\in T, k\to\infty} \|B_k d_k\| \le \beta_2 \liminf_{k\in T, k\to\infty} \|d_k\| = 0,$$

which implies that $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0$. This leads to a contradiction to (2.28).

(ii) If $\liminf_{k\in T, k\to\infty} \|d_k\| > 0$, then by (4.7), we know

$$\lim_{k\in T, k\to\infty} \alpha_k = 0. \tag{4.8}$$

From (4.5) and similar argument as Lemma 2.8, there exists a positive constant $\bar{C}_1$ such that

$$\alpha_k \ge \frac{-g_k^T d_k - L\|F_k\|^2 \|s_{k-1}\| \|d_k\|}{\bar{C}_1 \|d_k\|^2} = \frac{d_k^T B_k d_k}{\bar{C}_1 \|d_k\|^2} - \frac{L\|F_k\|^2 \|s_{k-1}\|}{\bar{C}_1 \|d_k\|}, \tag{4.9}$$

which together with (4.7) and (2.22) shows that, for sufficiently large $k \in T$,

$$\alpha_k \ge \frac{\beta_3}{2\bar{C}_1}.$$

This contradicts to (4.8). The proof is then finished. □

5 Numerical Experiments

In this section, we only compare the performance of the GN-BFGS method [8] with the parameters $\sigma_1 = \sigma_2 = 0.01$, $r = 0.5$ and Algorithm 2.1 with $\sigma = 0.01$, $r = 0.5$ and $y_k = z_k$ (that is, we use the unmodified BFGS update formula) in (2.13). The code was written in Matlab 7.4. We stopped the iteration if one of the following conditions was satisfied: (i) the total number of iterations(Iter) exceeds 100; (ii) $\|g_k\| \le 10^{-5}$.

Table 1 reports the numerical result for the large residual Trigonometric problem [1], which is given by the following system.

- *Trigonometric problem:*

$$F_i(x) = -d_i + \tilde{F}_i(x)^2, \quad i = 1, 2, \ldots, m,$$

where

$$\tilde{F}_i(x) = -e_i + \sum_{j=1}^{n} \big(a_{ij} \sin x_j + b_{ij} \cos x_j\big), \quad i = 1, 2, \ldots, m,$$

where $x = (x_1, \ldots, x_n)^T$, a_{ij}, b_{ij} are random integers in [-10,10], e_i are random numbers in [0,1], and $d = (d_1, d_2, \ldots, d_n)^T = (1, 2, \ldots, m)^T$. In this section, we set $m = n$.

Table 1: Test results for the methods with different initial points and $\{\epsilon_k\}$ values.

Parameters			The GN-BFGS Method			Algorithm 2.1		
ϵ_k	Initial Point	n	Iter	$\|g_k\|$	$\|F_k\|$	Iter	$\|g_k\|$	$\|F_k\|$
$\frac{1}{k^3}$	0	10	100	7.75e+004	2.40e+002	21	0.00e+000	6.23e+002
$\frac{1}{k^3}$	0	20	100	6.71e+006	7.22e+003	42	0.00e+000	2.43e+003
$\frac{1}{k^3}$	0	30	100	1.37e+007	7.02e+003	22	0.00e+000	3.65e+003
$\frac{1}{k^3}$	0	40	100	3.09e+007	9.07e+003	20	0.00e+000	6.02e+003
$\frac{1}{k^3}$	0	50	100	7.78e+007	2.09e+004	21	0.00e+000	2.08e+004
$\frac{1}{k^3}$	0	100	100	3.94e+008	6.11e+004	22	0.00e+000	4.84e+004
$\frac{1}{k^3}$	$(1,\dots,1)^T$	10	100	1.04e+005	2.91e+002	19	0.00e+000	7.99e+002
$\frac{1}{k^3}$	$(1,\dots,1)^T$	20	100	5.29e+006	5.10e+003	40	0.00e+000	2.14e+003
$\frac{1}{k^3}$	$(1,\dots,1)^T$	30	100	1.98e+007	9.36e+003	22	0.00e+000	4.75e+003
$\frac{1}{k^3}$	$(1,\dots,1)^T$	40	100	5.22e+007	1.90e+004	21	0.00e+000	1.18e+004
$\frac{1}{k^3}$	$(1,\dots,1)^T$	50	100	7.38e+007	1.90e+004	32	0.00e+000	1.84e+004
$\frac{1}{k^3}$	$(1,\dots,1)^T$	100	100	4.49e+008	5.37e+004	45	0.00e+000	5.35e+004
0.6^k	0	10	69	0.00e+000	6.46e+002	27	0.00e+000	7.22e+002
0.6^k	0	20	70	0.00e+000	4.96e+003	21	0.00e+000	4.91e+003
0.6^k	0	30	66	0.00e+000	8.37e+003	25	0.00e+000	6.93e+003
0.6^k	0	40	61	0.00e+000	1.86e+004	21	0.00e+000	1.71e+004
0.6^k	0	50	67	0.00e+000	2.19e+004	25	0.00e+000	1.38e+004
0.6^k	0	100	64	0.00e+000	4.86e+004	35	0.00e+000	4.29e+004
0.6^k	$(1,\dots,1)^T$	10	72	0.00e+000	2.37e+002	14	0.00e+000	8.89e+002
0.6^k	$(1,\dots,1)^T$	20	64	0.00e+000	4.66e+003	23	0.00e+000	2.15e+003
0.6^k	$(1,\dots,1)^T$	30	67	0.00e+000	8.54e+003	43	0.00e+000	8.10e+003
0.6^k	$(1,\dots,1)^T$	40	71	0.00e+000	1.30e+004	27	0.00e+000	1.26e+004
0.6^k	$(1,\dots,1)^T$	50	66	0.00e+000	2.56e+004	31	0.00e+000	2.06e+004
0.6^k	$(1,\dots,1)^T$	100	60	0.00e+000	6.09e+004	36	0.00e+000	5.54e+004

From Table 1, we can see that Algorithm 2.1 performs very well, which always converges to some stationary point successfully. However, the GN-BFGS method diverges in the case $\epsilon_k = \frac{1}{k^3}$. This is not surprising since the GN-BFGS method is based on the Gauss-Newton method and the latter easily performs badly for large residual problems [1, 12]. Moreover, we note that the GN-BFGS method needs more iterations than Algorithm 2.1, which shows that Algorithm 2.1 converges faster than the GN-BFGS method for this problem.

6 Conclusions

In this paper, we propose a new derivative-free BFGS method for symmetric nonlinear least squares with global and superlinear convergence, which is an extension of the GN-BFGS method [8] for symmetric nonlinear equations. The proposed method utilizes the second order information of the problem sufficiently, which makes it perform well for the given large residual problem. How to extend the proposed method to solve more general (non-symmetric) nonlinear least squares and report more numerical results compared with other methods will be our further study.

Acknowledgments

We would like to thank the referees whose very helpful suggestions led to much improvement of this paper.

References

[1] M. Al-Baali and R. Fletcher, Variational methods for nonlinear least squares, *J. Oper. Res. Soc.* 36 (1985) 405–421.

[2] R.H. Byrd and J. Nocedal, A tool for the analysis of quasi-Newton methods with application to unconstrained minimization, *SIAM J. Numer. Anal.* 26 (1989) 727–739.

[3] A.R. Conn, N.I.M. Gould and Ph. L. Toint, *Trust-Region Methods*, SIAM, Philadelphia, 2000.

[4] J.E. Dennis and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[5] G. Gu, D. Li, L. Qi and S. Zhou, Descent directions of quasi-Newton method for symmetric nonlinear equations, *SIAM J. Numer. Anal.* 40 (2002) 1763–1774.

[6] C.T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, SIAM, Philadelphia, 1995.

[7] D. Li and W. Cheng, Recent progress in the global convergence of quasi-Newton methods for nonlinear equations, *Hokkaido Math. J.* 36 (2007) 729–743.

[8] D. Li and M. Fukushima, A globally and superlinearly convergent Gauss-Newton-based BFGS method for symmetric nonlinear equations, *SIAM J. Numer. Anal.* 37 (1999) 152–172.

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

[10] J.M. Martinez, Practical quasi-Newton methods for solving nonlinear systems, *J. Comput. Appl. Math.* 124 (2000) 97–121.

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

[12] W. Sun and Y. Yuan, *Optimization Theory and Methods*, Springer Science and Business Media, LLC, New York, 2006.

[13] F. Wang, D. Li and L. Qi, Global convergence of Gauss-Newton-MBFGS method for solving the nonlinear least squares problem, *Adv. Model. Optim.* 12 (2010) 1–19.

[14] F. Wen and X. Yang, Skewness of return distribution and coeffcient of risk premium, *J. Syst. Sci. Comple.* 22 (2009) 360–371.

[15] Y. Yuan, Recent advances in numerical methods for nonlinear equations and nonlinear least sqaures, *Numer. Alge. Cont. Optim.* 1 (2011) 15–34.

[16] W. Zhou and X. Chen, Global convergence of a new hybrid Gauss-Newton structured BFGS methods for nonlinear least squares problems, *SIAM J. Optim.* 20 (2010) 2422–2441.

[17] W. Zhou and D. Li, A globally convergent BFGS method for nonlinear monotone equations without any merit functions, *Math. Comput.* 77 (2008) 2231–2240.

Manuscript received 7 January 2012
revised 30 August 2012
accepted for publication 10 September 2012

Weijun Zhou
Department of Mathematics, Changsha University of Science and Technology
Changsha 410004, China
E-mail address: weijunzhou@126.com