



A NEW MODIFIED STRUCTURED BFGS ALGORITHM BY UTILIZING A NEW MODIFIED SECANT EQUATION

ZAHRA KHOSHGAM, ALI ASHRAFI*

Department of Mathematics, Faculty of Mathematics, Statistics and Computer Science, Semnan University, Semnan, Iran

Abstract. In this paper, we consider a modified structured Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm to solve an unconstrained optimization problem with partial information on the Hessian. To approximate the Hessian matrix, we propose a modified secant (quasi-Newton) equation. Under appropriate conditions, we show that the proposed algorithm has local and super-linear convergence. Numerical experiments are done on a set of the unconstrained non-linear least squares problems. The results demonstrate the practical effectiveness of the proposed algorithm.

Keywords. Structured BFGS method; Least squares problems; Secant equation.

2010 Mathematics Subject Classification. 90C53, 49M37.

1. INTRODUCTION

A basic problem in engineering science is to fit a model to observations subject to errors. Least squares problems appear very naturally when one would like to estimate values of parameters of a mathematical model from measured data, which are subject to errors [1]. Least squares problems may be regarded as a special case for unconstrained minimization with a special structure and hence can be solved by unconstrained optimization algorithms. Least squares problems are in the form of

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} r(x)^T r(x) = \frac{1}{2} \sum_{i=1}^m r_i^2(x), \quad (1.1)$$

where $m \geq n$, the residual function $r(x) = (r_1(x), \dots, r_m(x))^T$ is non-linear, and $r_i : \mathbb{R}^n \rightarrow \mathbb{R}$ are twice continuously differentiable for $i = 1, 2, \dots, m$. The derivatives of f can be expressed in terms of the Jacobian matrix of $r(x)$, which is the $m \times n$ matrix of the first partial derivatives defined by

$$J(x) = \left[\frac{\partial r_j}{\partial x_i} \right]_{j,i=1,2,\dots,m}. \quad (1.2)$$

*Corresponding author.

E-mail addresses: z.khoshgam@semnan.ac.ir (Z. Khoshgam), a.ashrafi@semnan.ac.ir (A. Ashrafi).

Received February 4, 2018; Accepted June 29, 2019.

The gradient vector and Hessian of objective function (1.1) have special forms, which are given by

$$\nabla f(x) = \sum_{i=1}^m r_i(x) \nabla r_i(x) = J(x)^T r(x), \quad (1.3)$$

$$\nabla^2 f(x) = \sum_{i=1}^m \nabla r_i(x) \nabla r_i(x) + \sum_{i=1}^m r_i(x) \nabla^2 r_i(x) = J(x)^T J(x) + \sum_{i=1}^m r_i(x) \nabla^2 r_i(x), \quad (1.4)$$

where $\nabla^2 r_i(x)$ is the Hessian of r_i at x . Throughout this paper, we denote

$$\begin{aligned} C(x) &= J(x)^T J(x), \quad S(x) = \sum_{i=1}^m r_i(x) \nabla^2 r_i(x), \quad g_k = \nabla f(x_k), \\ J_k &= J(x_k), \quad r_k = r(x_k), \quad s_k = x_{k+1} - x_k, \quad y_k = g_{k+1} - g_k. \end{aligned}$$

It is interesting that $C(x)$ contains only the first-order information which is readily available, while $S(x)$ contains second-order information which usually expensive to be obtained. Among the iterative methods to solve (1.1), there are the structured quasi-Newton methods. Standard quasi-Newton methods approximate the whole Hessian matrix of $f(x)$. Many kinds of updating formulae, and convergence properties of these methods were proposed [2, 3]. If the information of the Hessian matrix is partially known, it is desirable to use this information to obtain more efficient methods. Dennis, Martínez and Tapia presented some quasi-Newton methods with partially known Hessian [4]. The structure principle is an important concept of structured quasi-Newton methods for these problems [4, 5].

◆*Structure principle:* Assume that $\nabla^2 f(x_k) = C(x_k) + S(x_k)$. Given $B_k = C(x_k) + \bar{A}_k$ as an approximation to $\nabla^2 f(x_k)$, we want $B_{k+1} = C(x_{k+1}) + \bar{A}_{k+1}$ to be an approximation of $\nabla^2 f(x_{k+1})$, where $x_{k+1} = x_k + s_k$.

By the structure principle, \bar{A}_k and \bar{A}_{k+1} are approximations of $S(x_k)$ and $S(x_{k+1})$, respectively. Note that if $\nabla^2 f(x)$ and $C(x)$ are nonsingular, we can write

$$\nabla^2 f(x_k)^{-1} \approx C^{-1}(x_k) + A_k,$$

where \bar{A}_k, A_k are unknown matrices containing the second order information of Hessian $\nabla^2 f(x_k)$ and $\nabla^2 f(x_k)^{-1}$, respectively. Now we discuss how to choose A_k . Consider the iterative formula of the structured quasi-Newton method, i.e.,

$$x_0 \in \mathbb{R}^n, \quad x_{k+1} = x_k + s_k, \quad k = 0, 1, \dots, \quad s_k = \alpha_k d_k,$$

where α_k is the step lenght to be computed by the line search procedure along the search direction d_k , and often determined to satisfy the Wolfe conditions [2, 3], i.e.,

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k \nabla f(x_k)^T d_k, \quad (1.5)$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \sigma \nabla f(x_k)^T d_k, \quad (1.6)$$

with $0 < \delta < \sigma < 1$. The search direction d_k is computed by

$$d_k = -H_k g_k,$$

in which $H_k \in \mathbb{R}^{n \times n}$ is an approximation of the inverse Hessian; more exactly, $H_k \approx \nabla^2 f(x_k)^{-1}$ or equivalently $H_k = C(x_k)^{-1} + A_k$. The matrix H_{k+1} is imposed with the scope of satisfying a particular equation,

namely, the secant equation, which implicitly includes the Hessian information. Among the most popular secant equations, there is the standard secant equation, that is,

$$H_{k+1}y_k = s_k. \quad (1.7)$$

Main drawback of usual secant equation (1.7) is that it utilizes only the gradient information and ignores the available function values. Therefore this issue was considered by researchers recently. In order to get a better approximation of the Hessian matrix, Wei, Li and Qi [6], Biglari, Hassan and Leong [7] and Zhang, Deng and Chen [8], respectively proposed the following modified secant equations using the Taylor expansion for the objective function;

$$B_{k+1}s_k = \hat{y}_k, \quad \hat{y}_k = y_k + \frac{\gamma_k}{s_k^T u_k} u_k, \quad (1.8)$$

$$B_{k+1}s_k = \bar{y}_k, \quad \bar{y}_k = y_k + \frac{2\gamma_k}{s_k^T u_k} u_k, \quad (1.9)$$

$$B_{k+1}s_k = \tilde{y}_k, \quad \tilde{y}_k = y_k + \frac{3\gamma_k}{s_k^T u_k} u_k, \quad (1.10)$$

in which

$$\gamma_k = 2(f_k - f_{k+1}) + s_k^T(g_k + g_{k+1}), \quad (1.11)$$

where $u_k \in \mathbb{R}^n$ is any vector parameter satisfying $s_k^T u_k \neq 0$. Note that these secant equations contain both gradient and function value information at the current and the previous step.

Recently, in order to get a high-order accuracy in approximating the second curvature of the objective function, Khoshgam and Ashrafi made a new modification of the SCALCG using a revised form of the modified secant equation suggested by Wei and Li and Qi [6] which was proposed in [9], that is,

$$\nabla^2 f(x_{k+1})s_k = \tilde{y}_k, \quad \tilde{y}_k = y_k + \xi_k \frac{\max\{0, \gamma_k\}}{s_k^T u_k} u_k, \quad (1.12)$$

where ξ is a non-negative constant such that $\xi \in [0, 3]$ and $u_k \in \mathbb{R}^n$ is any vector parameter satisfying $s_k^T u_k \neq 0$ and γ_k is defined by (1.11). Also, Khoshgam and Ashrafi obtained a special scalar $\frac{5}{3}$ for ξ using the fifth order Teylor expansion of the objective function about the iterate x_{k+1} . They proved that \tilde{y}_k is better than y_k due to \tilde{y}_k approximates $\nabla^2 f(x_{k+1})s_k$ with higher accuracy [9]. Main idea in [4] is to update A_k based on the standard secant equation (1.7), i.e.,

$$(C^{-1}(x_{k+1}) + A_{k+1})y_k = s_k.$$

This is equivalent to

$$A_{k+1}y_k = s_k^\#, \quad (1.13)$$

where

$$s_k^\# = s_k - C^{-1}(x_{k+1})y_k.$$

To solve problem (1.13), Dennis, Gay and Welsch [10] proposed the structured Davidon-Fletcher-Powell (DFP) update, and Dennis and Walker [11] showed the local and q-super linear convergence of this method. Dennis, Martínez and Tapia [4] proved the local and q-super linear convergence of the structured BFGS update, which was proposed in [12]. In 2006, Chen, Deng and Zhang developed the new modified quasi-Newton methods for structured unconstrained optimization based on the modified

secant (quasi-Newton) equation (1.10), which is suggested by Zhang, Deng and Chen [8]. Kobayashi, Narushima and Yabe [13] proposed conjugate gradient methods to solve the non-linear least squares problems by using the structure of the Hessian of the objective function.

In this paper, we consider the Khoshgam and Ashrafi's secant equation to approximate the Hessian matrix of the objective function. Then, according to the mentioned secant equation, a progressive modified structured BFGS algorithm is suggested to solve problem (1.1) in Section 2 and also, a local and q-super linear convergence property of the new algorithm is shown. The method is numerically compared with the four structured BFGS methods in Section 3. The conclusions are presented in the last section.

2. A NEW MODIFIED STRUCTURED BFGS ALGORITHM

In order to get a more accurate approximation of the Hessian matrix of the objective function f in problem (1.1), we consider the modified secant equation suggested by Khoshgam and Ashrafi [9],

$$\nabla^2 f(x_{k+1})s_k = \tilde{y}_k, \tilde{y}_k = y_k + \frac{5 \max\{\gamma_k, 0\}}{3 s_k^T u_k} u_k, \quad (2.1)$$

where γ_k is defined by (1.11) and $u_k \in \mathbb{R}^n$ is any vector parameter satisfying $s_k^T u_k \neq 0$.

Now, we present a modified structured BFGS method to solve least squares problems based on the approximation \tilde{y}_k to $\nabla^2 f(x_{k+1})s_k$. It is clear that the new secant equation (2.1) is equivalent to

$$H_{k+1}\tilde{y}_k = s_k. \quad (2.2)$$

So

$$A_{k+1}\tilde{y}_k = s_k^\#, \quad (2.3)$$

where $s_k^\# = s_k - C(x_k)^{-1}\tilde{y}_k$. According to the scheme of least change secant update formula proposed by Dennis, Martínez and Tapia in [4], a double rank update for A_k satisfying (2.3) is

$$A_{k+1} = A_k + \frac{(s_k^\# - A_k\tilde{y}_k)s_k^{\#T} + s_k^\#(s_k^\# - A_k\tilde{y}_k)^T}{s_k^{\#T}\tilde{y}_k} - \frac{\tilde{y}_k^T(s_k^\# - A_k\tilde{y}_k)s_k^{\#T}}{(s_k^{\#T}\tilde{y}_k)^2}. \quad (2.4)$$

Based on the above discussion, the following algorithm is obtained.

Algorithm 2.1. New Modified Structured BFGS Method (*NM – SBFGS*)

Step 1. Set $k := 0$; select a initial point $x_0 \in \mathbb{R}^n$, a symmetric and positive definite matrix $A_0 \in \mathbb{R}^{n \times n}$ and constants $0 < \delta < \frac{1}{2} < \sigma < 1$. Go to step 2.

Step 2. If the stopping criteria is satisfied, then, stop. Otherwise compute $H_k = C^{-1}(x_k) + A_k$ and $d_k = -H_k g_k$.

Step 3. Determine a step size $\alpha_k > 0$ satisfying the Wolfe line search conditions (1.5) and (1.6).

Step 4. Let

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

Step 5. Set Compute

$$\begin{aligned} y_k &= g_{k+1} - g_k, \\ \gamma_k &= (g_{k+1} + g_k)^T s_k - 2(f_{k+1} - f_k), \\ \tilde{y}_k &= y_k + \xi \frac{\max\{0, \gamma_k\}}{s_k^T u_k} u_k, \\ s_k^\# &= s_k - C^{-1}(x_{k+1})\tilde{y}_k, \end{aligned}$$

and update A_k by (2.4) and set $k := k + 1$ and go to Step 2.

In the following, we turn to the convergence properties of the *NM-SBFGS* algorithm. First, we prove $s_k^T \tilde{y}_k > 0$, which implies that A_{k+1} is always computed by (2.4). Next, we show that the bounded deterioration of $\{A_k\}$ and the Dennis-Moré condition of $\{B_k\}$. Consequently, the local and super-linear convergence is achieved by the modified structured BFGS method. The following assumptions for problem (1.1) will be used in the rest of the paper.

Denote $D = \{x : \|x - x^*\| < \varepsilon_1\}$, where $\varepsilon_1 > 0$ is a sufficiently small constant and $x^* \in \mathbb{R}^n$ is a local solution of the problem (1.1) in D .

Assumption 2.2. (I) $C(x^*)$ is non-singular.

(II) In D , $f(x)$ is twice continuously differentiable and there exists a constant $L > 0$ such that, for all $x, y \in D$,

$$\begin{aligned} \|\nabla^2 f(x) - \nabla^2 f(y)\| &\leq L\|x - y\|, \\ \|C(x) - C(y)\| &\leq L\|x - y\|, \\ \|f(x) - f(y)\| &\leq L\|x - y\|, \\ \|g(x) - g(y)\| &\leq L\|x - y\|, \\ \|C^{-1}(x) - C^{-1}(y)\| &\leq L\|x - y\|. \end{aligned}$$

(III) The matrix $\nabla^2 f(x^*)$ is positive definite, i.e., there exist positive constants m and M such that, for all $x \in D$ and $z \in \mathbb{R}^n$,

$$m\|z\|^2 \leq \frac{1}{2}z^T \nabla^2 f(x^*)z \leq M\|z\|^2.$$

(IV) For a given $u_k \in \mathbb{R}^n$ in (2.1), there exists a constant $\tau \in (0, 1]$ such that

$$|u_k^T s_k| \geq \tau \|u_k\| \|s_k\|; k = 0, 1, 2, \dots \quad (2.5)$$

It follows easily from Assumption 2.2 (II) that for any $x, \hat{x} \in D$,

$$\|g(x) - g(\hat{x}) - \nabla^2 f(x^*)(x - \hat{x})\| \leq L\sigma(x, \hat{x})\|x - \hat{x}\|,$$

where $\sigma(x, \hat{x}) = \max\{\|x - x^*\|, \|\hat{x} - x^*\|\}$.

Now we give the following useful lemmas for our global convergence analysis.

Lemma 2.3. *Let Assumption 2.2 hold. If the sequence $\{s_k\}$ is generated by the NM-SBFGS algorithm converging to x^* and $x_k \neq x^*$ for $k = 1, 2, \dots$, then there exists a positive constant K such that*

$$\|\tilde{y}_k - \nabla^2 f(x^*)s_k\| \leq K\sigma(x_{k+1}, x_k)\|s_k\|, \quad \forall x_k, x_{k+1} \in D.$$

Proof. By (2.1), we have

$$\begin{aligned} \|\tilde{y}_k - \nabla^2 f(x^*)s_k\| &= \|y_k + \xi \frac{\{0, \gamma_k\}}{s_k^T u_k} u_k - \nabla^2 f(x^*)s_k\| \\ &\leq \|y_k - \nabla^2 f(x^*)s_k\| + \xi \frac{|\gamma_k|}{|s_k^T u_k|} \|u_k\| \\ &\leq \|g_{k+1} - g_k - \nabla^2 f(x^*)s_k\| + \frac{\xi}{\tau \|s_k\|} |(g_{k+1} + g_k)^T s_k - 2(f_{k+1} - f_k)|. \end{aligned}$$

Thus, there exists $\xi_1, \xi_2, \xi_3 \in [x_{k+1}, x_k]$, which is the line segment between x_{k+1} and x_k such that

$$\begin{aligned} \|\tilde{y}_k - \nabla^2 f(x^*)s_k\| &\leq \|\nabla^2 f(\xi_1) - \nabla^2 f(x^*)\| \|s_k\| + \frac{\xi}{\tau \|s_k\|} |s_k^T [\nabla^2 f(\xi_2) - \nabla^2 f(\xi_3)] s_k| \\ &\leq L \|\xi_1 - x^*\| \|s_k\| + \frac{\xi}{\tau} L \|\xi_2 - \xi_3\| \|s_k\| \\ &\leq L(1 + \frac{\xi}{\tau}) \sigma(x_{k+1}, x_k) \|s_k\|. \end{aligned}$$

Set $K = L(1 + \frac{\xi}{\tau})$. The proof is completed. \square

Theorem 2.4. *Let Assumption 2.2 (I) hold. If the sequence $\{x_k\}$ is generated by the NM-SBFGS algorithm converging to x^* and $x_k \neq x^*$ for all $k = 0, 1, 2, \dots$, then $s_k^T \tilde{y}_k > 0$. So, the matrix A_{k+1} , which is computed using (2.4), is positive definite.*

Proof. From (1.6), for all k , we have

$$s_k^T \tilde{y}_k = s_k^T y_k + \xi \frac{\max\{0, \gamma_k\}}{s_k^T u_k} s_k^T u_k \geq s_k^T y_k = s_k^T g_{k+1} - s_k^T g_k \geq -(1 + \sigma) s_k^T g_k > 0.$$

\square

Lemma 2.5. *If the sequence $\{x_k\}$ is generated by the NM-SBFGS algorithm converging to x^* , then there exists $K_1 > 0$ such that*

$$\|A^* \tilde{y}_k - s_k^\# \| \leq K_1 \sigma(x_{k+1}, x_k) \|s_k\|, \forall x_k, x_{k+1} \in D, \quad (2.6)$$

where $A^* = \nabla^2 f(x^*)^{-1} - C^{-1}(x^*)$.

Proof. It is obvious that

$$\begin{aligned} \|A^* \tilde{y}_k - s_k^\# \| &= \|s_k - C^{-1}(x_{k+1}) \tilde{y}_k - \nabla^2 f(x^*)^{-1} \tilde{y}_k + C^{-1}(x^*) \tilde{y}_k\| \\ &\leq \|(\nabla^2 f(x^*)^{-1})^{-1}\| \|\tilde{y}_k - \nabla^2 f(x^*) s_k\| + \|\tilde{y}_k\| \|C(x_{k+1})^{-1} - C^{-1}(x^*)\|. \end{aligned}$$

By Lemma 2.3, we have

$$\|\tilde{y}_k\| \leq K \sigma(x_{k+1}, x_k) \|s_k\| + \|\nabla^2 f(x^*)\| \|s_k\|.$$

Thus, by Assumption 2.2 and the convergence of $\{x_k\}$, we can claim that there exists $K_1 > 0$ such that

$$\|A^* \tilde{y}_k - s_k^\# \| \leq K_1 \sigma(x_{k+1}, x_k) \|s_k\|, \quad \forall x_k, x_{k+1} \in D.$$

\square

The following lemma shows that the Dennis-Moré condition holds.

Lemma 2.6. *Under the assumptions of Lemma 2.3, we have*

$$\lim_{k \rightarrow \infty} \frac{\|(B_k - \nabla^2 f(x^*))s_k\|}{\|s_k\|} = 0. \quad (2.7)$$

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

Proof. Since the $B_k \approx H_k^{-1}$ and $B_{k+1}s_k = \tilde{y}_k$, we see that the proof is obvious. So we omit it. \square

The following theorem shows that Algorithm 2.1 converges superlinearly.

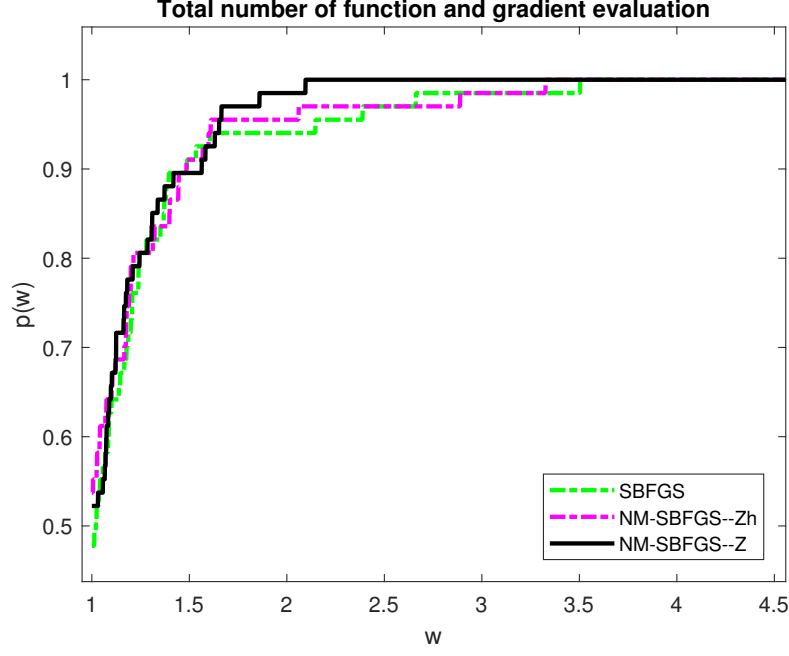


FIGURE 1. Total number of function and gradient evaluations performance profiles

Theorem 2.7. *If the sequence $\{x_k\}$ is generated by the NM-SBFGS algorithm converging to x^* , then there exists $\varepsilon, \delta > 0$ such that, for any initial point x_0 with $\|x_0 - x^*\| < \varepsilon$ and any $A_0 \in \mathbb{R}^{n \times n}$ with $\|A_0 - A^*\| < \delta$, the sequence $\{x_k\}$ q -superlinearly converges to x^* .*

Proof. The proof is from [4, Theorem 3.1]. Therefore, according to the well-known characterization result of Dennis and Moré [14], we conclude that the proposed method converges superlinearly. \square

3. NUMERICAL EXPERIMENTS

Here, we present some numerical results obtained by applying MATLAB 7. 14.0.739 (R 2012a) using a computer Intel (R) Core(TM) i7-6700HQ CPU @ 2.60 GHz with 8.00 GB RAM with implementation of the Algorithm 2.1. To test efficiency of the new modified secant equation (2.1) in the structured BFGS method, we compare the performance of the following three methods.

- (1) SBFGS: Consider Algorithm 2.1 with $\xi = 0$, corresponding to original secant equation which was presented in [4];
- (2) NM-SBFGS-ZH: Consider Algorithm 2.1 with $\xi = 3$, corresponding to Zhang secant equation (1.10) which was presented in [15];
- (3) NM-SBFGS-Z: Consider Algorithm 2.1 with $\xi = \frac{5}{3}$, corresponding to Khoshgam secant equation (2.1).

The runs were performed on a set of 18 unconstrained least squares problems provided in [16] with the minimum dimension being equal to 2. In Table 1 these 18 test problems are listed, where "No.", "n", "m" and "Start point" stand respectively for the serial number of the problem, the dimension of the problem, the number of residual functions and the initial point x_0 . Each function was minimized for four different starting x_0 , $0.1x_0$, $0.01x_0$ and $2x_0$, resulting in a total 72 test problems.

TABLE 1. List of Least Squares Problems

No.	Problem Name	n	m	Start point (x_0)
1	Chained Rosenbrock	4	7	-1.2, 1, ..., -1.2, 1
2	Chained Powell Singular	4	4	3, -1, 0, 1, ..., 3, -1, 0, 1
3	Generalized Broyden tridiagonal (p=7/3)	30	30	-1, -1, ..., -1
4	Generalized Broyden banded	6	6	-1, -1, ..., -1
5	Chained Freudenstein and Roth	4	6	0.5, 0.5, ..., 0.5, -2
6	Wright and Holt zero residual	2	10	$\sin^2 1, \sin^2 2, \dots, \sin^2 n$
7	Toint quadratic merging	4	6	0.5, 0.5, ..., 0.5
8	Chained serpentine	4	6	0.8, 0.8, ..., 0.8
9	Chained and modified problem (HS48)	5	7	-1, -1, ..., -1
10	Countercurrent reactors	5	5	0.1, 0.2, 0.3, 0.4, 0.5, 0.4, 0.3, 0.2, ..., ...,
11	Structured Jacobian Problem	5	5	-1, -1, ..., -1
12	Shifted discrete boundary value	4	4	$\frac{n}{(n+1)^2}, \frac{2(n-1)}{(n+1)^2}, \dots, \frac{n}{(n+1)^2}$
13	Modified discrete boundary value	4	4	$-\frac{n}{(n+1)^2}, -\frac{2(n-1)}{(n+1)^2}, \dots, -\frac{n}{(n+1)^2}$
14	Generalized Broyden tridiagonal (p=2)	30	30	-1, -1, ..., -1
15	Chained and modified problem (HS48)	8	14	-1, -1, ..., -1
16	Broyden tridiagonal problem	4	4	-1, -1, ..., -1
17	Extended Wood problem	4	4	-3, -1, ..., -3, -1
18	Countercurrent reactors	7	7	0.1, 0.2, 0.3, 0.4, 0.5, 0.4, 0.3, 0.2, ..., ...,

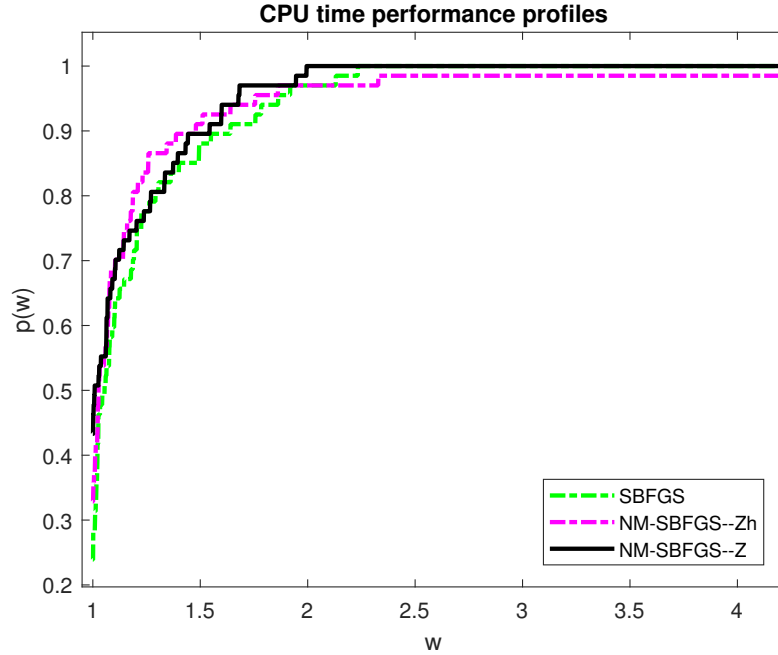


FIGURE 2. CPU time performance profiles

In our implementations, if

$$g_k^T d_k > -10^{-7} \|g_k\| \|d_k\|, \quad (3.1)$$

then d_k is considered as an improper search direction. Although (3.1) is rare, when it happens, the algorithm is restarted with $d_k = -g_k$. Also, we set $u_k = s_k$ in (2.1). All algorithms implement the Wolfe conditions (1.5) and (1.6) with $\delta = 0.0001$ and $\sigma = 0.9$. The steplength α_k is computed using the interpolation scheme proposed in [2]. For the first inner iteration of the line search procedure, the initial trial value being equal to 1. Moreover, all attempts to find an approximation of the solution of the test problems are terminated if one of the following conditions is satisfied (i) $\|g_k\| \leq 10^{-5}$;

(ii) The total number of iterations exceed 1000.

Efficiency comparisons were drawn using the Dolan–Moré performance profile [17], on the running time and the total number of function and gradient evaluations being equal to $N_f + 3N_g$, where N_f and N_g respectively denote the number of function and gradient evaluations. Performance profile gives, for every $\omega \geq 1$, the proportion $p(\omega)$ of the test problems that each considered algorithmic variant has a performance within a factor of ω of the best. Figures 1 and 2 demonstrate the results of comparisons.

Figures show, NM-SBFGS–Z is competitive and it outperforms the other algorithms with respect to the total number of function and gradient evaluations and the running time. These observations imply that our modified NM-SBFGS–Z algorithm with $\xi = \frac{5}{3}$ turns out to be practically effective.

4. CONCLUSIONS

In this paper, we applied the new modified secant equation of Khoshgam and Ashrafi [9] to develop the new modified Structured BFGS algorithm for solving non-linear least squares problems. The method has been shown to be locally and q-super linearly convergent. Preliminary numerical results showed the efficiency of the proposed method in the sense of the Dolan–Moré performance profile.

Acknowledgements

This research was supported by Research Council of Semnan University, Iran.

REFERENCES

- [1] M. Gander, M.J. Gander, F. Kwok, Scientific Computing-An Introduction using Maple and MATLAB, Springer, New York, 2014.
- [2] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, New York, 2006.
- [3] W. Sun, Y.X. Yuan, Optimization Theory and Methods: Nonlinear Programming, Springer, New York, 2006.
- [4] J.E. Dennis, H.J. Martínez, R.A. Tapia, Convergence theory for the structured BFGS secant method with an application to nonlinear least squares, J. Optimiz. Theory App. 61 (1989), 161-178.
- [5] W. Zhou, X. Chen, Global Convergence of a new hybrid Gauss-Newton structured BFGS method for nonlinear least squares problems, SIAM J. Optim. 20 (2010), 2422-2441.
- [6] Z. Wei, G. Li, L. Qi, New quasi-Newton methods for unconstrained optimization problems, Appl. Math. Comput. 175 (2006), 1156-1188.
- [7] F. Biglari, M.A. Hassan, W.J. Leong, New quasi-Newton methods via higher order tensor models, J. Comput. Appl. Math. 235 (2011), 2412-2422.
- [8] J.Z. Zhang, N.Y. Deng, L.H. Chen, New quasi-Newton equation and related methods for unconstrained optimization, J. Optim. Theory App. 102 (1999), 147-167.
- [9] Z. Khoshgam, A. Ashrafi, A new modified scaled conjugate gradient method for large-scale unconstrained optimization with non-convex objective function, Optim. Method Softw. 34 (2019), 783-796.
- [10] J.E.Jr. Dennis, D.M. Gay, R.E. Welsch, An adaptive nonlinear least-squares algorithm, ACM Trans. Math. Softw. 7 (1981), 348-368.

- [11] J.E. Jr. Dennis, H.F. Walker, Convergence theorems for least change secant update methods, *SIAM J. Numer. Anal.* 18 (1981), 949-987.
- [12] M. Al-Baali, R. Fletcher, Variational methods for non-linear least-squares, *J. Oper. Res. Soc.* 36 (1985), 405–421.
- [13] M. Kobayashi, Y. Narushima, H. Yabe, Nonlinear conjugate gradient methods with structured secant condition for non-linear least squares problems, *J. Comput. Appl. Math.* 234 (2010), 375-397.
- [14] J.E.Jr. Dennis, J.J. Moré, A characterization of superlinear convergence and its application to quasi-newton methods, *Math. Comput.* 28 (1974), 54-460.
- [15] L.H. Chen, N.Y. Deng, J.Z. Zhang, A modified quasi-Newton method for structured optimization with partial information on the Hessian, *Comput. Optim. Appl.* 35 (2006), 5-18.
- [16] L. Lukšan, C. Matonoha, Vlček, Sparse Test Problems for Unconstrained Optimization, Institute of Computer Science Academy of Sciences of the Czech Republic, Technical report No. 1064 (2010), 1-23.
- [17] E.D. Dolan, J.J. Moré, Benchmarking optimization software with performance profiles, *Math. Program.* 91 (2002), 201-213.