

# Using a modified secant equation for unconstrained optimization

July 5, 2017

R. Dehghani, M. M. Hosseini

Faculty of Mathematics, yazd University, PO Box 89195-74, Yazd, Iran.  
r.dehghani9080@yahoo.com, hosse\_m@yazd.ac.ir.

## Abstract

We make some efficient modifications on the modified secant equation proposed by Zhang and Xu (2001). Then we introduce modified BFGS method using propose secant equation, and obtain some attractive results in theory and practice. We establish the global convergence property of the proposed method without convexity assumption on the objective function. Numerical results on some testing problems from CUTer collection show the priority of the proposed method to some existing modified secant methods in practice

**Keywords:** Unconstrained optimization. Modified secant equation. BFGS method,

**AMS subject Classification:** 90C30, 90C53, 49M37, 65K05.

## 1 Introduction

Consider the unconstrained nonlinear optimization problem

$$\min f(x), \quad x \in \mathbb{R}^n, \quad (1)$$

where  $f$  is twice continuously differentiable. Secant methods are popular iterative methods for solving (1), with the iterates being constructed as follows:

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

where  $\alpha_k$  is a step size and  $d_k$  is a searchdirection obtained by solving  $B_k d_k = -g_k$ , with  $g_k = \nabla f(x_k)$  and  $B_k$  an approximation of the Hessian matrix of  $f$  at  $x_k$  satisfying the secant equation.

The standard secan equation can be established as follows (see [7]). We have

$$g_{k+1} - g_k = \int_0^1 \nabla^2 f(x_k + ts_k) dt s_k, \quad (2)$$

where  $s_k = x_{k+1} - x_k$ . Since  $B_{k+1}$  is to approximate  $G(x_{k+1}) = \nabla^2 f(x_{k+1})$ , the secant equation is defined to be

$$B_{k+1} s_k = y_k, \quad (3)$$

where  $y_k = g_{k+1} - g_k$ . The relation (3) is sometimes called the standard secant equation. (see Dennis and Moré [6] for a comprehensive treatment of quasi-Newton methods particularly the secant methods).

A family of secant methods is Broyden family [1] in which the updates are defined by

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} + \mu w_k w_k^T, \quad w_k = (s_k^T B_k s_k)^{1/2} \left[ \frac{y_k}{s_k^T y_k} - \frac{B_k s_k}{s_k^T B_k s_k} \right] \quad (4)$$

where  $\mu$  is a scale parameter. The BFGS, DFP and SR1 updates are obtained by setting  $\mu = 0$ ,  $\mu = 1$  and  $\mu = 1/(1 - s_k^T B_k s_k / s_k^T B_k s_k)$ , respectively.

Among the secant methods, the most efficient quasi-Newton method is the BFGS method which was proposed by Broyden [2], Fletcher [9], Goldfarb [14] and Shanno [21].

When  $f$  is convex, the global convergence of the BFGS method have been studied by some authors (see [3, 4, 16, 20, 22]). Dai [5] have constructed an example to show that the standard BFGS method may fail for non-convex functions with inexact line search. Mascarenhas [19] showed that the nonconvergence of the standard BFGS method even with exact line search. Li and Fukushima [17, 18] made a modification on the standard BFGS method and developed a modified BFGS method that is globally convergent without a convexity assumption on the objective function  $f$ .

The usual Secant equation employs only the gradients and the available function values are ignored. In order to get a higher order accuracy of approximating the Hessian matrix of the objective function, several researchers have modified the usual Secant equation (3) to make full use of both the gradient and function values (see [23]-[26]).

Zhang and Xu [26] using Taylor's Series modified (3) as follows:

$$B_{k+1} s_k = y_k^\neq, \quad (5)$$

where

$$y_k^\neq = y_k + \frac{\vartheta_k}{\|s_k\|^2} s_k, \quad \vartheta_k = 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k. \quad (6)$$

Using modified secant equation (5), they proposed the following BFGS update formula:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k^\neq y_k^{\neq T}}{s_k^T y_k^\neq}. \quad (7)$$

They showed that the modified BFGS method is locally and superlinearly converge with the assumption  $f$  is uniformly convex function. However, if  $f$  is a general function may this method nonconvergence.

This motivated us to modification on modified secant equation (5). Then, we make use of the new secant equation in a BFGS update formula. Under some proper assumptions, we prove the global convergence property for general functions.

The rest of our work is organized as follows: In Section 2, we introduce an alternative approximation of the secant equation. In Section 3, we investigate the convergence of our proposed method. In Section 4, we report the numerical results.

## 2 Modified secant equation

In this section, we first describe a modified secant equation in [26] that utilize both the available gradient and function values. Moreover, this method has a better theoretical feature than the usual secant equation and the secant equation introduced in [23].

Using the Taylor's series for  $f(x)$ , we have

$$f_k = f_{k+1} - g_{k+1}^T s_k + \frac{1}{2} s_k^T G_{k+1} s_k - \frac{1}{3!} s_k^T (T_{k+1} s_k) s_k + O(\|s_k\|^4), \quad (8)$$

and

$$s_k^T g_k = s_k^T g_{k+1} - s_k^T G_{k+1} s_k + \frac{1}{2} s_k^T (T_{k+1} s_k) s_k + O(\|s_k\|^4), \quad (9)$$

where

$$s_k^T (T_{k+1} s_k) s_k = \sum_{i,j,l=1}^n \frac{\partial^3 f(x_{k+1})}{\partial x^i \partial x^j \partial x^l} \partial s^i \partial s^j \partial s^l. \quad (10)$$

Cancellation of the terms which include the tensor yields

$$s_k^T G_{k+1} s_k \simeq (g_{k+1} - g_k)^T s_k + 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k \quad (11)$$

Then using a new approximation  $B_{k+1}$ , we have

$$s_k^T B_{k+1} s_k = y_k^T s_k + \vartheta_k \quad (12)$$

where

$$\vartheta_k = 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k \quad (13)$$

This suggests the following new secant equation

$$B_{k+1} s_k = y_k^\neq, \quad (14)$$

where  $y_k^\neq = y_k + \frac{\vartheta_k}{\|s_k\|^2} s_k$  and  $\vartheta_k = 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k$ .

One theoretical advantages of the new modified secant method can be seen from the following theorem [25].

**Theorem 2.1.** Suppose that the function  $f$  is sufficiently smooth. If  $\|s_k\|$  is small enough, then we have:

$$s_k^T (G_{k+1} s_k - y_k^\neq) = O(\|s_k\|^4), \quad (15)$$

and

$$s_k^T (G_{k+1} s_k - y_k) = O(\|s_k\|^3). \quad (16)$$

Notice thatt if the objective function  $f$  is uniformly convex, then

$$y_k^{\neq T} s_k = y_k^T s_k + 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^T s_k = 6(f_k - f_{k+1}) + 2(g_k + 2g_{k+1})^T s_k > 0.$$

which guarantees the positive definite of the matrix  $B_k$  for uniformly convex function. However, if  $f$  is a general function, may happen  $y_k^{\neq T} s_k \leq 0$ . Hence the positive definiteness of the matrix  $B_k$  can not be guarantee for general function.

In addition, Theorem 2.1, demonstrate if  $\|s_k\| > 1$ , the standard secant equation is expected to be more accurate than the modified secant equation (14). In this case, the use of (14) does not seem to be suitable.

To overcome these problems, we introduce some modification on (14) as follows:

$$B_{k+1}s_k = y_k^*, \quad (17)$$

with

$$y_k^* = \bar{y}_k + \gamma \|g_k\|^2 s_k + \max\left(-\frac{\bar{y}_k^T s_k}{\|s_k\|^2}, 0\right) s_k \quad (18)$$

where  $\bar{y}_k = y_k + \rho_k \frac{\vartheta_k}{\|s_k\|^2} s_k$  and  $\gamma$  is a positive constant and

$$\rho_k = \begin{cases} e^{-\|s_k\|} & \text{for } \|s_k\| \leq 1, \\ 0, & \text{otherwise.} \end{cases} \quad (19)$$

Clearly, if  $\|s_k\| \rightarrow 0$  then  $\rho_k \rightarrow 1$  and if  $\|s_k\| > 1$  then  $\rho_k = 0$ .

It is easy to see

$$y_k^{*T} s_k \geq \rho_k \gamma \|g_k\|^2 \|s_k\|^2 > 0, \quad (20)$$

which guarantees that positive definite of the update matrix  $B_k$ , for the general function.

We can now give a new BFGS algorithm using our new secant relation for solving (1) as follows.

**Algorithm 1: New modified BFGS method:**

**Step 1 :** given  $\varepsilon$  as a tolerance for convergence,  $\sigma_1 \in (0, 1)$ ,  $\sigma_2 \in (\sigma_1, 1)$ , a starting point  $x_0 \in \mathbb{R}^n$ , and a positive definite matrix  $B_0$ . Set  $k = 0$ .

**Step 2 :** If  $\|g_k\| < \varepsilon$  then stop.

**Step 3 :** Compute a search direction  $d_k$  : Solve  $B_k d_k = -g_k$ .

**Step 4:** Compute  $\alpha_k$  by using the following Wolfe conditions:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \sigma_1 \alpha_k g_k^T d_k, \quad (21)$$

and

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma_2 g_k^T d_k. \quad (22)$$

**Step 5 :** Set  $x_{k+1} = x_k + \alpha_k d_k$ . Compute  $y_k^*$  by (18) . Update  $B_{k+1}$  by

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k^* y_k^{*T}}{s_k^T y_k^*}. \quad (23)$$

**Step 6:** Set  $k = k + 1$  and go to Step 2.

Next, we will investigate the global and superlinear convergence of the proposed Algorithm without convexity assumption on the objective function  $f$ .

### 3 Convergence analysis

Here, we establish the convergence of Algorithm 1. We need the following usual assumptions.

**Assumption A.** The level set  $D = \{x \mid f(x) \leq f(x_0)\}$  is bounded.

**Assumption B.** The function  $f$  is continuously differentiable on  $D$ , and there is a constant  $L \geq 0$  such that, for all  $x, y \in D$ ,  $\|g(x) - g(y)\| \leq L \|x - y\|$ .

Clearly, these assumptions imply that there exists a constant  $m > 0$  such that

$$\|g(x)\| \leq m, \quad \forall x \in D. \quad (24)$$

From assumption A and the Wolfe conditions we deduce that  $\{f(x_k)\}$  is a nonincreasing sequence, which ensures  $\{x_k\} \subset D$  and the existence of  $x^*$  such that

$$\lim_{k \rightarrow \infty} f(x_k) = f(x^*). \quad (25)$$

In order to establish the global convergence of Algorithm 1, we present the following useful Lemmas.

**Lemma 3.1.** Suppose that Assumption B holds and  $y_k^*$  define by (14). Then, there exist have

$$\|y_k^*\| \leq M \|s_k\| \quad (26)$$

where  $M$  is positive constant.

**Proof:** Considering Assumption B and definition of  $\vartheta_k$ , we have

$$|\vartheta_k| \leq 3L \|s_k\| \quad (\text{See the relations leading to (5.10) of [23]}). \quad (27)$$

Therefore, since  $\rho_k \in [0, 1]$  we can give

$$\|\bar{y}_k\| = \|y_k + \rho_k \frac{\vartheta_k}{\|s_k\|^2} s_k\| \leq \|y_k\| + \frac{|\vartheta_k|}{\|s_k\|} \leq L \|s_k\| + 3L \|s_k\| = 4L \|s_k\|. \quad (28)$$

Hence, from (24) and (28) we have

$$\|y_k^*\| \leq 2\|\bar{y}_k\| + \gamma \|g_k\|^2 \|s_k\| \leq (2L + \gamma m^2) \|s_k\|. \quad (29)$$

This equation imply the (26) with  $M = 2L + \gamma m^2$

**Lemma 3.2.** Let  $f$  satisfy assumptions A and B, and  $\{x_k\}$  be generated by Algorithm 1 and there exist constants  $a_1$  and  $a_2$  such that

$$\|B_k s_k\| \leq a_1 \|s_k\|, \quad s_k^T B_k s_k \geq a_2 \|s_k\|^2, \quad (30)$$

for infinitely many  $k$ . Then, we have

$$\liminf_{k \rightarrow \infty} g(x_k) = 0. \quad (31)$$

**Proof:** Since  $s_k = \alpha_k d_k$ , it is clear that (30) holds true if  $s_k$  is replaced by  $d_k$ . From (30) and the relation  $g_k = -B_k d_k$ , we have

$$d_k^T B_k d_k \geq a_2 \|d_k\|^2, \quad a_2 \|d_k\| \leq \|g_k\| \leq a_1 \|d_k\|. \quad (32)$$

Let  $\Lambda$  be the set of indices  $k$  for which (30) hold. By using (22) and Assumption B, we have

$$L\alpha_k \|d_k\|^2 \geq (g_{k+1} - g_k)^T d_k \geq -(1 - \sigma_2) g_k^T d_k. \quad (33)$$

This implies that, for any  $k \in \Lambda$ ,

$$\alpha_k \geq \frac{-(1 - \sigma_2) g_k^T d_k}{L \|d_k\|^2} = \frac{(1 - \sigma_2) d_k^T B_k d_k}{L \|d_k\|^2} \geq \frac{(1 - \sigma_2) a_2}{L}. \quad (34)$$

Moreover, by (25), we obtain

$$\sum_{k=1}^{\infty} (f_k - f_{k+1}) = \lim_{N \rightarrow \infty} \sum_{k=1}^N (f_k - f_{k+1}) = \lim_{N \rightarrow \infty} (f(x_1) - f(x_N)) = f(x_1) - f(x^*),$$

which yields

$$\sum_{k=1}^{\infty} (f_k - f_{k+1}) < \infty.$$

Using (21), we obtain

$$\sum_{k=1}^{\infty} \alpha_k g_k^T d_k < \infty,$$

which ensures

$$\lim_{k \rightarrow \infty} \alpha_k g_k^T d_k = 0.$$

This together with (34) lead to

$$\lim_{k \in \Lambda, k \rightarrow \infty} d_k^T B_k d_k = \lim_{k \in \Lambda, k \rightarrow \infty} -g_k^T d_k = 0.$$

which long with (32), yields (31).  $\square$

The following Lemma is taken from [3], we represent it here but omit the proof.

**Lemma 3.3.** (Theorem 2.1 of [3]) If there are positive constants  $M_1$  and  $M_2$  such that for all  $k \geq 0$ ,

$$\frac{\|y_k^*\|^2}{s_k^T y_k^*} \leq M_1 \quad \text{and} \quad \frac{s_k^T y_k^*}{\|s_k\|^2} \geq M_2, \quad (35)$$

then there exist constants  $a_1$  and  $a_2$  such that, for any positive integer  $t$ , (30) holds for at least  $\lfloor \frac{t}{2} \rfloor$  values of  $k \in \{1, 2, \dots, t\}$ .

Now, we prove the global convergence for Algorithm 1.

**Theorem 3.1.** Let  $f$  satisfy the assumptions A and B, and  $\{x_k\}$  be generated by Algorithm 1. Then (31) holds.

**Proof:** Assume, to the contrary that the conclusion is not true. Then, there exists a positive constant  $\delta$  such that, for all  $k$ ,

$$\|g_k\| > \delta, \quad (36)$$

Hence, (20) imply that

$$y_k^{*T} s_k \geq \delta^2 \|s_k\|^2, \quad (37)$$

Therefore, using Lemma 3.1 and (37), we obtain

$$\frac{\|y_k^*\|^2}{s_k^{*T} y_k^*} \leq M, \quad \forall k \geq 0.$$

Hence, (35) holds for all  $k$ . Using Lemma 3.3 to the subsequence  $\{B_k\}_{k \in K}$ , clearly there exist constants  $a_1 > 0$  and  $a_2 > 0$  such that (30) holds for infinitely many  $k$ . Then Lemma 3.2 completes the proof.  $\square$

The above theorem shows a global convergence property of MBFGS method without convexity assumption on the objective function.

To establish the superlinear convergence of Algorithm 1, we need further the following assumption.

**Assumption C.** Suppose  $x_k \rightarrow x^*$  at which  $g(x^*) = 0$ ,  $G(x)$  is positive definite, and  $G(x)$  is Holder continuous at  $x^*$ , i.e., there exist constants  $v \in (0, 1)$  and  $L_2 > 0$  such that

$$\|G(x) - G(x^*)\| \leq L_2 \|x - x^*\|^v$$

for all  $x$  in the neighborhood of  $x^*$ .

Similar to the proof of Theorem 3.8 in [18], it is not difficult to prove the superlinear convergence of Algorithm 1. Hence, we only present the theorem without giving the proof.

**Theorem 3.2.** Suppose that assumptions A, B and C hold. Let the sequence  $\{x_k\}$  be generated by the Algorithm 1. Then  $\{x_k\}$  is superlinearly convergent.

## 4 Numerical results

We are to compare the performance of the following four methods on some unconstrained optimization problems:

**BFGS:** the usual BFGS method.

**MBFGS:** proposed method (Algorithm 1).

**ZXMBFGS:** the modified BFGS of Zhang and Xu using (5) [26].

**BFGSA<sub>k</sub>(2):** using the following Modified secant equation to update  $B_k$  which suggest by Wei et.al in [23]

$$B_{k+1}s_k = y_k + \frac{\theta_k}{\|s_k\|^2}s_k \quad (38)$$

where  $\theta_k = 2(f_k - f_{k+1}) + (g_k + g_{k+1})^T s_k$ .

We have tested all the considered algorithms on 120 test problems from CUTer library [15]. A summary of these problems are given in Table 1. All codes were written in Matlab 2012 and run on PC with CPU Intel(R) Core(TM) i5-4200 3.6 GHz, 4 GB of RAM memory and Centos 6.2 server Linux operation system.

In all algorithms, the initial matrix is  $B_0 = I$  and the steplengths  $\alpha_k$  satisfies the Wolfe conditions, with  $\sigma_1 = 0.001$ , and  $\sigma_2 = 0.1$ .

For all the test problems, the termination condition is  $\|g_k\| \leq 10^{-5}$ , or  $\|f_{k+1} - f_k\| \leq 10^{-20} \max(1, f_k)$ .

We use the profiles of Dolan and Moré (see [8]) to evaluate performance of these four algorithms with respect to number of iteration and the total number of function and gradient evaluations being equal to  $N_f + nN_g$ , where  $N_f$  and  $N_g$  denote the number of function and gradient evaluations respectively. Figs 1 and 2 demonstrate the results of the comparisons. From Figs. 1 and 2, it is easy to observe that the MBFGS method is the most efficient for solving these 120 test problems among the four methods. We see that MBFGS method solves about 75% and 74% of the test problems with the fewest number of iterations and function evaluations, respectively.



Table 1: Test problems taken from CUTEr library.

No	test function	dim	No	test function	dim	No	test function	dim
1	AIRCRFTB	10	41	GULF	3	81	ARWHEAD	500
2	ALLINITU	20	42	HATFLDD	d3	82	BOX	100
3	ARGLINA	200	43	HEART8LS	d10	83	BRKMCC	2
4	BARD	100	44	HELIX	3	84	BROYDN7D	1000
5	BIGGS3B	100	45	HILBERTA	2	85	CHAINWOO	3000
6	BIGGS5	100	46	HILBERTB	10	86	COSINE	1000
7	BIGGS6	100	47	HILBERTF	5	87	CUBE	2
8	BOX2	100	48	HILBERTG	2	88	CURLY10	3000
9	BOX3	200	49	HILBERTH	2	89	CURLY20	3000
10	BROWNAL	200	50	JENSMP	2	90	CURLY30	3000
11	BROWNBS	200	51	KOWOSB	5	91	DENSCHNE	3
12	BRYBND	5000	52	LIARWHD	3000	92	DENSCHNF	2
13	CHNROSNB	50	53	LMINSURF	3000	93	EDENSCH	50
14	DECONVU	60	54	test LMINSURY	2	94	EG2	1000
15	DENSCHNA	10	55	MANCINO	100	95	ENGVAL1	100
16	DENSCHNB	10	56	MOREBV	3000	96	FLETGBV2	3000
17	DENSCHNC	10	57	NLMSURF	3000	97	FLETCHCR	1000
18	DIXMAANA	2000	58	NONDIA	d3000	98	FMINSURF	961
19	DIXMAANB	2000	59	NONDQUAR	3000	99	GENHUMPS	3000
20	DIXMAANC	2000	60	OSBORNEB	10	100	GENROSE	500
21	DIXMAAND	2000	61	PALMER5C	6	101	HAIRY	2
2	DIXMAANE	2000	62	POWELLSG	1000	102	HATFLDFL	3
3	DIXMAANF	2000	63	QUARTC	3000	103	HUMPS	2
4	DIXMAANG	2000	64	ROSENBR	2	104	JIMACK	3000
5	DIXMAANH	2000	65	S308	d2	105	MARATOSB	2
6	DIXMAANI	2000	66	SCHMVETT	3	106	MSQRTALS	500
7	DIXMAANJ	2000	67	SCHMVETT	2	107	MSQRTBLS	500
8	DIXMAANK	20	68	SNAIL	2	108	NCB20	120
29	DIXMAANL	2000	69	SPARSQR	3000	109	NCB20B	20
30	DIXON3DQ	3000	70	SPMSRTLS	2000	110	NONCVXU2	1000
31	DQDRTIC	3000	71	SROSENBR	2000	111	PENALTY1	50
32	DQRTIC	20	72	TESTQUAD	2000	112	PENALTY2	50
33	EIGENALS	200	73	TOINTTGSS	3000	113	PENALTY3	100
34	EIGENBLS	1000	74	TQUARTIC	3000	114	POWER	dim
35	EIGENCLS	500	75	TRIDIA	dim	115	SINEVAL	2
36	ENGVAL2	6	76	VAREIGVL	50	116	SINQUAD	5
37	EXPFIT	16	77	WATSON	10	117	SPARSINE	3000
38	EXTROSNB	1000	78	WOODS	3000	118	TOINTQOR	50
39	FMINSRF2	3000	79	YFITU	5	119	VARDIM	10
40	GROWTHLS	10	80	ZANGWIL2	2	120	VIBRBEAM	10

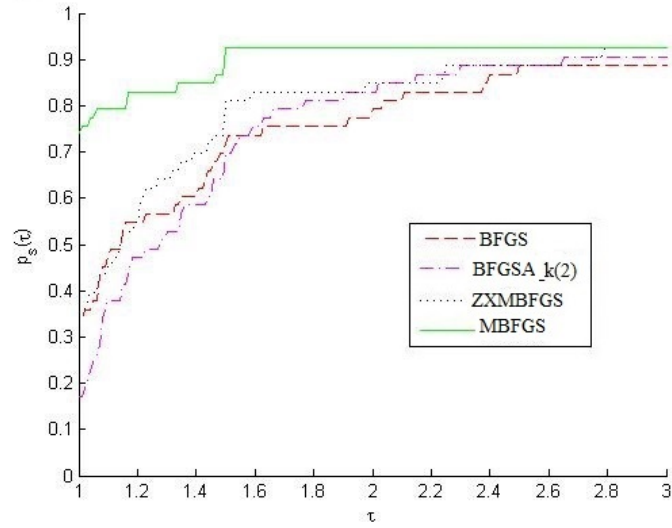


Figure 1: The Dolan-Moré performance profiles using number of function evaluations.

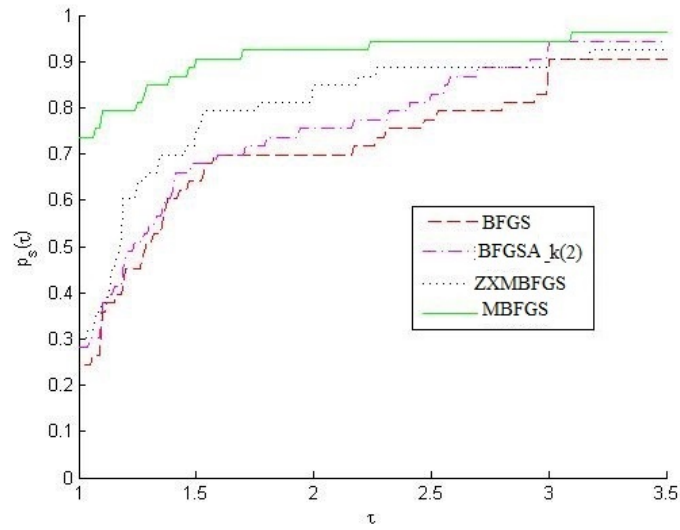


Figure 2: The Dolan-Moré performance profiles using number of iterations.

## 5 Conclusion

We introduced a modified BFGS (MBFGS) method using a new secant equation. An interesting feature of the proposed method is to take both the gradient and function values into account. Under suitable assumptions, we established the global convergence of the proposed method for the general functions. Numerical results on the collection of problems in CUTER show that the proposed method is efficient as compared to several proposed BFGS methods.

## References

- [1] Broyden, C.G., A class of methods for solving nonlinear simultaneous equations, *Math. Comp.* 19: 577-593 (1965).
- [2] Broyden, C.G., The convergence of a class of double rank minimization algorithms-Part 2: The new algorithm, *J. Inst. Math. Appl.* 6: 222-231 (1970).
- [3] Byrd, R., Nocedal, J., A tool for the analysis of quasi-Newton methods with application to unconstrained minimization, *SIAM J. Numerical Anal.* 26: 727-739 (1989).
- [4] Byrd, R., Nocedal, J., Yuan, Y., Global convergence of a class of quasi-Newton methods on convex problems, *SIAM J. Numer. Analysis* 24: 1171-1189 (1987).
- [5] Dai, Y., Convergence properties of the BFGS algorithm, *SIAM J. Optim.* 13: 693-701 (2003).
- [6] Dennis, J.E., More, J.J., A characterization of superlinear convergence and its application to quasi-Newton methods, *Math. Comp.* 28: 549-560 (1974).
- [7] Dennis, J.E., Schnabel, R.B., *Numerical Methods for Unconstrained Optimization and Non-linear Equations*, Prentice-Hall, Englewood Cliffs, New Jersey, 1983.
- [8] Dolan, E., Moré, J.J., Benchmarking optimization software with performance profiles, *Math. Program.* 91: 201-213 (2002).
- [9] Fletcher, R., A new approach to variable metric algorithms, *Comput.* 13: 317-322 (1970).
- [13] Ford, J.A., Saadallah, A.F., On the construction of minimisation methods of quasi-Newton type, *J. Comput. Appl. Math.* 20: 239-246 (1987).
- [14] Goldfarb, D., A family of variable metric methods derived by variational means, *Math. Comp.* 24: 23-26 (1970).
- [15] Gould, N.I.M., Orban, D. and Toint, Ph.L., CUTER, a constrained and unconstrained testing environment, revisited, *ACM Trans. Math. Softw.* 29: 373-394 (2003).
- [16] Griewank, A., The global convergence of partitioned BFGS on problems with convex decompositions and Lipschitzian gradients, *Math. Program.* 50: 141-175 (1991).
- [17] Li, D.H., Fukushima, M., On the global convergence of BFGS method for nonconvex unconstrained optimization problems, *SIAM Journal on Optimization*. 11: 1054-1064 (2001).
- [18] Li, D.H. and Fukushima, M., A modified BFGS method and its global convergence in non-convex minimization, *J. Comput. Appl. Math.* 129: 15-35 (2001).
- [19] Mascarenhas, W.F., The BFGS method with exact line searches fails for non-convex objective functions, *Math. Program.* 99: 49-61 (2004).
- [20] Powell, M.J.D., Some global convergence properties of a variable metric algorithm for minimization without exact line searches. In: Cottle, R.W., Lemke, C.E. (eds.), *Nonlinear Programming*, SIAMAMS Proceedings, vol. IX. SIAM, Philadelphia (1976).
- [21] Shanno, D.F., Conditioning of quasi-Newton methods for function minimization, *Math. Comp.* 24: 647-656 (1970).
- [22] Toint, Ph.L., Global convergence of the partitioned BFGS algorithm for convex partially separable optimization. *Math. Program.* 36: 290-306 (1986).
- [23] Wei, Z., Li, G., Qi, L., New quasi-Newton methods for unconstrained optimization problems, *Appl. Math. Comput.* 175: 1156-1188 (2006).
- [24] Yuan, G. and Wei, Z., Convergence analysis of a modified BFGS method on convex minimizations, *Computational Optimization and Applications*, 47: 237-255 (2010).

- [25] Zhang, J.Z., Deng, N.Y., Chen, L.H., New quasi-Newton equation and related methods for unconstrained optimization, *J. Optim. Theory Appl.* 102: 147-167 (1999).
- [26] Zhang, J.Z., Xu, C.X., Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equation, *J. Comput. Appl. Math.* 137: 269-278 (2001).
- [27] Zhu, D.T., Nonmonotone backtracking inexact quasi-Newton algorithms for solving smooth nonlinear equations, *Appl. Math. Comput.* 161: 875-895 (2005).