An improved quasi-Newton equation on the quasi-Newton methods for unconstrained optimizations

Basim A. Hassan¹, Kanikar Muangchoo², Fadhil Alfarag³, Abdulkarim Hassan Ibrahim⁴, Auwal Bala Abubakar⁵

¹Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, Iraq
²Faculty of Science and Technology, Rajamangala University of Technology Phra Nakhon (RMUTP), Thailand
³School of Mathematics, University of Birmingham, United Kingdom
⁴Department of Mathematics, Faculty of Science, King Mongkut’s University of Technology Thonburi (KMUTT), Bangkok, Thailand
⁵Department of Mathematical Sciences, Faculty of Physical Sciences, Bayero University, Kano, Kano, Nigeria

Article Info

Article history:
Received Dec 4, 2020
Revised Mar 17, 2021
Accepted Apr 6, 2021

Keywords:
Global convergence property
Numerical experiments
Quasi-Newton equation
Unconstrained optimizations

ABSTRACT

Quasi-Newton methods are a class of numerical methods for solving the problem of unconstrained optimization. To improve the overall efficiency of resulting algorithms, we use the quasi-Newton methods which is interesting for quasi-Newton equation. In this manuscript, we present a modified BFGS update formula based on the new quasi-Newton equation, which give a new search direction for solving unconstrained optimizations problems. We analyse the convergence rate of quasi-Newton method under some mild condition. Numerical experiments are conducted to demonstrate the efficiency of new methods using some test problems. The results indicates that the proposed method is competitive compared to the BFGS methods as it yielded fewer iteration and fewer function evaluations.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Kanikar Muangchoo
Faculty of Science and Technology
Rajamangala University of Technology Phra Nakhon (RMUTP)
1381, Pracharat 1 Road, Wongsawang, Bang Sue, Bangkok 10800 Thailand
Email: kanikar.m@rmutp.ac.th

1. INTRODUCTION

Quasi-Newton method is one of the main iterative methods designed to solve unconstrained optimization problem [1]. The method has the (1).

\[ \text{Min } f(x), \quad x \in \mathbb{R}^n \]  

(1)

where \( f : \mathbb{R}^n \to \mathbb{R} \) is smooth function. More details can be found in [2]. The iterative method takes the (2).

\[ x_{k+1} = x_k + \alpha_k d_k \]  

(2)

Journal homepage: http://ijeecs.iaescore.com
where $\alpha_k$ is the parameter determined by exact line-search as.

$$\alpha_k = -\frac{g_k^T d_k}{d_k^T Q d_k} \quad \text{(3)}$$

More details can be found in [3]. The search direction $d_k$ generated by,

$$B_k d_k + g_k = 0 \quad \text{(4)}$$

where $B_k$ is an approximation of the Hessian matrix [4]. In this manuscript, we pay attention to the BFGS method in which $B_k$ is updated by (5).

$$B_{k+1}^{BFGS} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} \quad \text{(5)}$$

where $s_k = x_{k+1} - x_k = \alpha_k d_k$ and $y_k = g_{k+1} - g_k$. Consider $H_k$ be the inverse of $B_k$. Surely, the (5) update can be written as (6).

$$H_{k+1}^{BFGS} = H_k - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k} \left[ 1 + \frac{y_k^T H_k y_k}{s_k^T y_k} \right] s_k s_k^T \quad \text{(6)}$$

To the best of our knowledge the BFGS method may fail to converge by using Wolfe line search for non-convex optimization and also by using with exact line search constructed Dai [2] and Mascarenhas [5]. To defeat (remedy) this case, quasi-Newton equation has been modified in some way or other can be seen some formulas among others as.

| Author(s)          | QN – conditions | References |
|--------------------|-----------------|------------|
| Wei, Li, and Qi     | $B_{k+1} = B_k - \frac{2(f_k - f_{k-1}) + (g_k + g_{k-1})^T s_k}{s_k^T s_k} s_k$ | [6]        |
| Biglari, Hassan, and Leon | $B_{k+1} = B_k - \frac{2(f_k - f_{k-1}) + (g_k + g_{k-1})^T s_k}{s_k^T s_k} s_k$ | [7]        |
| Chen, Deng, and Zhang | $B_{k+1} = B_k - \frac{6(f_k - f_{k-1}) + 3(g_k + g_{k-1})^T s_k}{s_k^T s_k} s_k$ | [8]        |
| Basim              | $B_{k+1} = B_k - \frac{1}{2} y_k + \frac{f_k - f_{k-1}}{s_k^T s_k} s_k$ | [9]        |
| Basim and Mohammed | $B_{k+1} = B_k - \frac{1}{2} y_k + \frac{f_k - f_{k-1}}{s_k^T s_k} s_k$ | [10]       |
| Basim and Ghada    | $B_{k+1} = B_k - \frac{1}{2} y_k + \frac{f_k - f_{k-1}}{s_k^T s_k} s_k$ | [11]       |

where $\nu_k$ is any vector satisfying $s_k^T \nu_k \neq 0$. In [12], a general framework of modified methods is proposed to solve non-convex optimization problems. Thus the modifying methods can guarantee the updated matrix is positive definite which guarantees a descent direction. The convergent properties are also established and generated more accurate second-order curvature approximations than the usual quasi-Newton updates do.

The quasi-Newton equation plays an important role, on studies some of the quasi-Newton methods. In this paper, we propose a modification to the hessian matrix $B_k$ resulting in a new quasi-newton method based on new quasi-Newton equation. The proposed method can guarantee that the updated matrix is positive definite which guarantees a descent direction. The convergent properties are also established. More so the proposed method generates a more accurate second-order curvature approximations than the usual quasi-Newton updates.
2. DERIVING THE NEW QUASI-NEWTON EQUATION

Zahra and Ali [13] propose a type quasi-Newton equation of quasi-Newton methods and generate several QN directions, as (7).

\[ \nabla^2 f(x_{k+1})s_k = -y_k = y_k + \frac{5}{3} \max \left\{ 2(f_k - f_{k+1}) + (g_{k+1} + g_k)^T s_k, 0 \right\} s_k^T u_k \]

(7)

For more details can be found in [13].

So we will deriving quasi-Newton equation motivated by quasi-Newton (7). Now multiplying it by \( s_k^T \), we have.

\[ s_k^T \nabla^2 f(x_{k+1})s_k = s_k^T y_k = s_k^T y_k + \frac{5}{3} \left[ 2(f_k - f_{k+1}) + (g_{k+1} + g_k)^T s_k \right] \]

(8)

we can write:

\[ \frac{3}{5} s_k^T \nabla^2 f(x_{k+1})s_k = \frac{3}{5} s_k^T y_k + 2(f_k - f_{k+1}) + (g_{k+1} + g_k)^T s_k \]

(9)

Now, from (3) and (9), we obtain.

\[ s_k^T \nabla^2 f(x_{k+1})s_k = \frac{3}{5} s_k^T y_k + 2(f_k - f_{k+1}) + (g_{k+1} + g_k)^T s_k + \frac{2}{5} s_k^T \nabla^2 f(x_{k+1})s_k \]

\[ s_k^T \nabla^2 f(x_{k+1})s_k = \frac{3}{5} s_k^T y_k + 2(f_k - f_{k+1}) + (g_{k+1} + g_k)^T s_k - \frac{2}{5} s_k^T s_k \]

(10)

From (10) we get.

\[ s_k^T \nabla^2 f(x_{k+1})s_k = \frac{3}{5} s_k^T y_k + 2(f_k - f_{k+1}) + g_{k+1}^T s_k + \frac{3}{5} g_k^T s_k \]

(11)

Consider \( B_{k+1} \) is an approximate of the Hessian matrix \( \nabla^2 f(x_{k+1}) \) of the \( f(x_k) \).

\[ s_k^T B_{k+1} s_k = \frac{3}{5} s_k^T y_k + 2(f_k - f_{k+1}) + g_{k+1}^T s_k + \frac{3}{5} g_k^T s_k \]

(12)

This gives a new quasi-Newton equation in the form.

\[ B_{k+1} s_k = \frac{3}{5} y_k + \frac{2(f_k - f_{k+1}) + g_{k+1}^T s_k + \frac{3}{5} g_k^T s_k}{s_k^T u_k} u_k \]

(13)

where \( u_k \) is any vector satisfying \( s_k^T u_k \neq 0 \). Motivated by the idea of Zahra and Ali [8], we propose a modified above quasi-Newton equation, we get as.

\[ B_{k+1} s_k = \frac{3}{5} y_k + \frac{\max \left\{ 2(f_k - f_{k+1}) + g_{k+1}^T s_k + \frac{3}{5} g_k^T s_k, 0 \right\}}{s_k^T u_k} u_k \]

(14)

Obviously two choices for \( u_k \) can be computed as.

1. First case \( u_k = g_{k+1} \) gives:
\[ B_{k+1}s_k = y_k = \frac{3}{5} y_k + \max \left\{ \frac{2(f_k - f_{k+1}) + g_{k+1}^T s_k + \frac{3}{5} g_{k+1}^T s_k}{s_k^T g_{k+1}} \right\} \] (15)

2. Second case \( \nu_k = y_k \) gives:
\[ B_{k+1}s_k = y_k = \frac{3}{5} y_k + \max \left\{ \frac{2(f_k - f_{k+1}) + g_{k+1}^T s_k + \frac{3}{5} g_{k+1}^T s_k}{s_k^T y_k} \right\} \] (16)

Using quasi-Newton equation in the BFGS update to give a good result. Now, we are in a situation to explain our algorithm in details.

Stage 1. Give \( \rho, \) a initial point \( x_0 \in \mathbb{R}^n. \) Set \( k = 0. \)
Stage 2. If possible test satisfies then stop.
Stage 3. Solve \( B_k d_k = -g_k. \)
Stage 4. Find \( \alpha_k \) by using the following Wolfe conditions:

\[ f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \] (17A)
\[ d_k^T g(x_k + \alpha_k d_k) \geq \sigma d_k^T g_k \] (17B)

where \( 0 < \delta < \sigma < 1. \)
Stage 5. Set \( x_{k+1} = x_k + \alpha_k d_k. \) Compute \( y_k \) by (14). If \( s_k^T y_k > 0, \) update \( H_{k+1} \) by (6), otherwise let \( H_{k+1} = H_k. \) Set \( k = k + 1 \) and go to Step 2.

The major work, we introduce the positive definite property for the new quasi-Newton equation.

**Theorem 1.**

Suppose that \( s_k^T y_k > 0. \) Then \( B_{k+1} \) is symmetric positive definite.

**Proof.**

Using definition \( y_k \) and multiplying by \( s_k^T, \) we get:
\[ s_k^T y_k = \frac{3}{5} s_k^T y_k + 2(f_k - f_{k+1}) + g_{k+1}^T s_k + \frac{3}{5} g_{k+1}^T s_k \] (18)

From (19), (17) and (18), we obtained.
\[ s_k^T y_k \geq -2\delta \alpha_k^2 s_k + \frac{8}{5} \sigma g_k^T s_k \]
\[ \geq -2\sigma \alpha_k^2 s_k + \frac{8}{5} \sigma g_k^T s_k \]
\[ \geq -\frac{2}{5} \sigma g_k^T s_k \]

As \( s_k^T g_k = \alpha_k d_k^T g_k < 0, \) then
\[ s_k^T y_k > -2/5 \sigma \alpha_k d_k^T g_k \] (20)

Denote the right side part of the inequality by \( \Lambda, \) then one has.
\[ s_k^T y_k > \alpha > 0 \]  \hspace{1cm} (21)

So \( B_{k+1} \) is positive definite.

Though the \( s_k^T y_k > 0 \), does not hold true for non-convex problems, \( B_{k+1} \) may not be positive definite. It serves as a useful condition for positive definiteness of the updates consider.

\[ K = \left\{ k \mid s_k^T y_k \geq \beta \|g_k\|^2 \right\}, \hspace{1cm} (22) \]

where \( \beta > 0 \) is constant and \( \delta > 0 \) is bounded.

### 3. CONVERGENCE ANALYSIS

We will study the global convergence of our method. We need some assumptions. Let the level set \( D = \{ x \mid f(x) \leq f(x_0) \} \), with \( x_0 \) is an initial point of iterative method is restricted.

Using Lipschitz continuous; that is exist constants \( L \) and \( \gamma \), such that.

\[ \|\nabla f(\nu) - \nabla f(\omega)\| \leq L \|\nu - \omega\|, \forall \nu, \omega \in D \]  \hspace{1cm} (23)

and,

\[ \|\nabla f(x)\| \leq \gamma, \forall x \in D \]  \hspace{1cm} (24)

Since \( \{f(x_k)\} \) is a no increasing, which ensures \( \{x_k\} \) is contained in \( D \) and the existence of \( x^* \) we have.

\[ \lim_{k \to \infty} f(x_k) = f(x^*) \]  \hspace{1cm} (25)

Moreover, from the fact that sequence \( x_k \) is bounded, there exists \( \mu \), such that \( \forall k, \)

\[ \|s_k\| = \|x - x_k\| \leq \|x\| + \|x - x_k\| \leq \mu \]  \hspace{1cm} (26)

More details can be found in [14], [15].

Presented the useful theorem to prove that our method is globally convergent.

**Theorem 2.**

If \( \|\nabla f(x)\| \leq \gamma \) is not satisfies for all \( k \). Let \( \{x_k\} \) be generated by modified methods, and the (27) holds.

\[ \|B_k s_k\| \leq \alpha_1 \|s_k\| \text{ and } s_k^T B_k s_k \geq \alpha_2 \|s_k\|^2, \]  \hspace{1cm} (27)

where \( \alpha_1 > 0 \) and \( \alpha_2 > 0 \) are constants. For infinitely \( k \), then we have.

\[ \lim_{k \to \infty} \inf \|g_k\| = 0 \]  \hspace{1cm} (28)
Proof:
By (4) of the Wolfe conditions we obtain.
\[
(g_{k+1} - g_k)^T d_k \geq -(1-\sigma) g_k^T d_k
\] (29)

By using Lipschitz condition we obtain.
\[
(g_{k+1} - g_k)^T d_k \leq L \alpha_k \|d_k\|
\] (30)

This implies that.
\[
\alpha_k \geq \frac{-(1-\sigma) g_k^T d_k}{L \|d_k\|^2} = \frac{(1-\sigma) B_k d_k}{L \|d_k\|^2 \geq \frac{(1-\sigma) a_2}{L}}
\] (31)

using (22) we obtain.
\[
\sum_{k=1}^{\infty} (f_k - f_{k+1}) = \lim_{N \to \infty} \sum_{k=1}^{N} (f_k - f_{k+1}) = \lim_{N \to \infty} (f_1 - f_{k+1}) = f_1 - f
\] (32)

The (32) implies
\[
\sum_{k=1}^{\infty} (f_k - f_{k+1}) \leq + \infty
\] (33)

This together with Wolfe condition (3) yields.
\[
\sum_{k=1}^{\infty} - \alpha_k g_k^T d_k \leq + \infty
\] (34)

Then
\[
\lim_{k \to \infty} \alpha_k g_k^T d_k = 0
\] (35)

in cooperation with (31) provide that.
\[
\lim_{k \to \infty} d_k^T B_k d_k = \lim_{k \to \infty} -(g_k^T d_k) = 0
\] (36)

Combining (33) with (24) we obtain the conclusion (25). The proof is finished.

Nevertheless, we give the lemma for large scale problems to show the convergence property for new algorithm. This lemma was shown by Powell [16].

Lemma 1.
If BFGS method with Wolfe condition is applied to a continuously differentiable function \(f\) that is bounded below, and if there exists a constant \(M\) such that the inequality holds.
\[
\|y_k\|^2 \leq M
\] (37)

Then,
\[ \lim_{k \to \infty} \inf \| g_k \| = 0 \]  

\text{(38)}

**Theorem 3.**

Let \( \{x_k\} \) be generated by the proposed method and \( \nu_k \leq \chi \| s_k \| \). Then we have.

\[ \lim_{k \to \infty} \inf \| g_k \| = 0 \]  

\text{(39)}

**Proof:**

Using a contradiction method with there exists \( \varepsilon > 0 \) such that.

\[ \| g_k \| > \varepsilon \]  

\text{(40)}

Hence, (22) imply that.

\[ s_k^T y_k \geq \beta \| g_k \| \| s_k \| ^2 \geq \beta \gamma \| s_k \| ^2 \]  

\text{(41)}

By the definition of \( y_k \) can be written as.

\[ y_k^* = \frac{3}{5} y_k + \frac{\sigma_k}{s_k^T u_k} v_k, \quad \sigma_k = \max \left\{ 0, \frac{2(f_k - f_{k+1}) + g_k^T s_k + 3/5 g_k^T s_k}{s_k^T u_k} \right\} \]  

\text{(42)}

So, from (24), we get.

\[ y_k^* = \frac{9}{25} (G s_k)^2 + \frac{6}{5} (G s_k) \frac{\sigma_k}{s_k^T u_k} + \frac{(\sigma_k^*)^2}{s_k^T u_k} \]  

\text{(43)}

By simple computation and using \( \nu_k \leq \chi \| s_k \| \), we obtain.

\[ \| y_k^* \| ^2 \leq s_k^T \left( \frac{3}{5} G + \chi \frac{\sigma_k}{s_k^T u_k} \right) s_k \leq v \| s_k \| ^2 \]  

\text{(44)}

Therefore, by dividing the sides of inequality \( \| y_k^* \| ^2 \), inequalities (41) we obtain.

\[ \frac{\| y_k^* \| ^2}{s_k^T y_k} \leq M \]  

\text{(45)}

Using lemma 1, to the sub \( \{B_k\}_{k \in K} \), obviously, clearly there exist \( a_1 \) and \( a_2 \) we get (27) for infinitely many \( k \). Then theorem 2. completes the proof.

### 4. NUMERICAL REPORTS

In this part, we details results of some numerical experiments with the new method and algorithm BFGS. We choose 30 test problems with the different dimension and initial points from literature [17]. Some enlargement was observed for class from test problems [18]-[24]. We set the parameters \( \sigma_1 = 0.1, \sigma_2 = 0.9 \) and \( \varepsilon = 10^{-5} \) in the numerical experiment. By applying a law Himmeblau [25], the stop criterion is “If

An improved quasi-Newton equation on the quasi-Newton methods for unconstrained ... (Basim A. Hassan)
\[ \| f(x_i) \| > 10^{-5}, \text{ let } stop = \| f(x_i) - f(x_{i+1}) \| / \| f(x_i) \| ; \text{ Otherwise, let } stop = \| f(x_i) - f(x_{i+1}) \|. \] For every problem, if \( \| g \| < \varepsilon \) or \( stop < 10^{-5} \) is satisfied, the program will be stopped. The numerical results based on IN and NF, the number of iterations and the number of function evaluations, respectively.

A comparison of the algorithms is given based on the performance profiles of Dolan and More [26]. The comparison is based on NI and NF, respectively. The plots obtained via the Dolan and More performance profiles indicate that the M2 and M3 methods perform better than the M1 method. To facilitate this, we used the following notations to represent the algorithms: M1. This is the standard BFGS method defined by (6). M2. This is the new algorithm with \( \nu_i = \frac{g}{\| g \|} \). Figure 1 and Figure 2, we have found that new proposed methods best BFGS method in about (17-53)% NI and (49-51)% NF. The new methods has good numerical results.

CONCLUSIONS

The paper studies the quasi-Newton method which is popular in solving large-scale unconstrained optimization. It presents a new quasi-Newton equation for gradient methods. Numerical comparisons with the well-known BFGS method are given. Inspired new quasi-Newton very well in practice in comparison with quasi-Newton methods.

ACKNOWLEDGMENT

The first author is grateful to the University of Mosul/College of Computers Sciences and Mathematics for their provided facilities, which helped to improve the quality of this work. The second author acknowledge the financial support of her institution (Rajamangala University of Technology). The last author acknowledge with thanks, the Department of Mathematics and Applied Mathematics at the Sefako Makgatho Health Sciences University.

REFERENCES

[1] Yuan G., Sheng Z., Wang B., Hu W. and, Li C., "The global convergence of a modified BFGS method for nonconvex functions," Journal of Computational and Applied Mathematics, vol. 327, pp. 274-294, 2018, doi: 10.1016/j.cam.2017.05.030.
[2] Dai Y, "Convergence properties of the BFGS algorithm," SIAM J. Optim., vol. 13, no. 3, pp.693-701, 2013, doi: 10.1137/S1052623401383455.
[3] Nocedal, J. and Wright, S. J., "Numerical optimization," Springer Series in Operations Research, 2nd edition, Springer-Verlag, New York, USA, 2006,
[4] Yuan, G, Wei, Z, and Wu, Y, "Modified limited memory BFGS method with nonmonotone line search for unconstrained optimization," J. Korean Math. Soc., vol. 47, no. 4, pp. 767-788, 2010, doi: 10.4134/JKMS.2010.47.4.767.
A new type of quasi-Newton method for unconstrained optimization problems.

The BFGS method with exact line search fails for non-convex objective functions. Some global convergence properties of a variable metric algorithm for minimization without exact line searches, Nonlinear Programming, SIAM, 75, pp. 1156-1186, 2006.

Yuan Y., Wei Z., Li G., and Qi L., "A new nonlinear conjugate gradient method for unconstrained optimization problems," Optimization Methods and Software, 1, pp. 73-84, 1997.

A new modified scaled conjugate gradient method for large-scale unconstrained optimization with non-convex objective function, Optimization Methods and Software, 3, pp. 783-796, 2019.

The superlinear convergence of a modified BFGS-type method for unconstrained optimization, Comput. Optim. Appl., 29, pp. 315-332, 2004.

A modified quasi-Newton method for unconstrained optimization, Journal of Computational and Applied Mathematics, 328, pp. 44-58, 2018.

Some global convergence properties of a variable metric algorithm for minimization without exact line searches, Nonlinear Programming, SIAM-AMS Proceedings, Lemke, eds., SIAM, 1976, pp. 53-72.

Testing unconstrained optimization software, ACM Trans. Math. Software, 7, pp. 17-41, 1981.

A globally convergence spectral conjugate gradient method for solving unconstrained optimization problems, Raf. J. of Comp. & Math., 10, no. 4, pp. 21-28, 2013.

Development a special conjugate gradient algorithm for solving unconstrained minimization problems, Raf. J. of Comp. & Math., 9, no. 1, pp. 73-84, 2012.

A modified class of conjugate gradient algorithms based on quadratic model for nonlinear unconstrained optimization, Raf. J. of Comp. & Math., 9, no. 1, pp. 25-37, 2014.

A new sufficient descent conjugate gradient method for nonlinear, Iraqi Journal of Statistical Sciences, 26, pp. 12-24, 2014.

A new nonlinear conjugate gradient method for unconstrained optimization, Journal of the Operations Research Society of Japan, 48, no. 4, pp. 284-296, 2005.

Conjugate gradient methods using value of objective function for unconstrained optimization, Optimization Letters, 6, no. 5, pp. 941-955, 2011.

Convergence properties of nonlinear conjugate gradient methods, SIAM Journal on Optimization, 10, no. 2, pp. 348-358, 1999.