Indonesian Journal of Electrical Engineering and Computer Science
Vol. 19, No. 2, August 2020, pp. 737~744
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v19.i2.pp737-744

A new quasi-newton equation on the gradient methods for optimization minimization problem

Basim A. Hassan, Ghada M. Al-Naemi
Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, Iraq

Article Info

Article history:

Received Jan 1, 2020
Revised Mar 4, 2020
Accepted Mar 18, 2020

Keywords:

Global convergence
Quasi-newton equation
Quasi-newton method

ABSTRACT

The quasi-Newton equation is the very foundation of an assortment of the quasi-Newton methods for optimization minimization problem. In this paper, we deriving a new quasi-Newton equation based on the second-order Taylor's series expansion. The global convergence is established underneath suitable conditions and numerical results are reported to show that the given algorithm is more effective than those of the normal BFGS method.

Corresponding Author:

Basim A. Hassan,
Department of Mathematics,
College of Computers Sciences and Mathematics,
University of Mosul, Iraq.
Email: basimah@uomosul.edu.iq, basimabas39@gmail.com

1. INTRODUCTION

The quasi-Newton algorithm is one of the more successful algorithms for unconstrained nonlinear programming [1]. These methods, which use the updating formulas for approximation of the Hessian. To minimize a multi-variable nonlinear function this our objective :

$$Min\ \ f(x)\ ,\ \ x \in R^n \tag{1}$$

where f is twice differentiable. Throughout the paper, we define by B_k is a nonnegative definite matrix that estimates the Hessian matrix $Q_k = \nabla^2 f(x_k)$ of $f(x_k)$ and $g_k = \nabla f(x_k)$ is the gradient of $f(x_k)$. More details can be found in [2]. For most optimization algorithms, the search for the minimizer of (1) is carried out by using :

$$x_{k+1} = x_k + \alpha_k d_k \tag{2}$$

where α_k is the step size, and d_k is the search direction. Moreover, the search direction d_k of the quasi-Newton methods often has the form :

$$B_k d_k + g_k = 0 \tag{3}$$

It estimates B_k update formula, we will focus on the BFGS method which has proved to be the most effective of all quasi-Newton methods :

$$B_{k+1}^{BFGS} = B_k - \frac{B_k s_k s_k^T B_k{}^T}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} \tag{4}$$

It's also well known that the matrix B_{k+1} is generated by (4) to satisfy the secant equation :

$$B_{k+1} s_k = y_k \tag{5}$$

where $s_k = x_{k+1} - x_k = \alpha_k d_k$ and $y_k = g_{k+1} - g_k$, for more details see [3, 4].

It is easy to take advantage of the following relationships : $B_{k+1} s_k \Leftrightarrow y_k$, $B_k \Leftrightarrow H_k$ where $B_k^{-1} = H_k$, then the famous BFGS formula is designed by :

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

For details see [5]. When the general function no one has proved the convergence property of the BFGS method. To acquire improved quasi-Newton methods, many modified quasi-Newton equations have been proposed ([6-10], among others) and established its convergence property. Using second-order Taylor's series approximation of the function to derivation the a new quasi-Newton equation and we study convergence property and numerical results.

2. DERIVING QUASI-NEWTON EQUATIONS

We deriving new quasi-Newton equations based on the second-order Taylor's series expansion and is defined by :

$$f(x_k) = f(x_{k+1}) - g_{k+1}^T s_k + \frac{1}{2} s_k^T Q s_k \tag{7}$$

Using exact line search, we get :

$$f(x_k) = f(x_{k+1}) + \frac{1}{2} s_k^T Q s_k \tag{8}$$

which implies that :

$$s_k^T Q s_k = 2(f(x_k) - f(x_{k+1})) \tag{9}$$

Add and subtract from the right tip the $s_k^T y_k$, we get :

$$s_k^T Q s_k = 2(f(x_k) - f(x_{k+1})) + s_k^T y_k - s_k^T y_k \tag{10}$$

The choice B_{k+1} is key to the approximate effect of the Hessian matrix Q :

$$s_k^T B_{k+1} s_k = 2(f(x_k) - f(x_{k+1})) + s_k^T y_k - s_k^T y_k \tag{11}$$

which implies a new QN equation and as follows :

$$B_{k+1}s_k = \tilde{y}_k, \quad \tilde{y}_k = y_k + \frac{2(f(x_k) - f(x_{k+1})) - s_k^T y_k}{s_k^T v_k} v_k \tag{12}$$

where v_k is any vector such that $s_k^T v_k \neq 0$. Now we Applying the a new quasi-Newton equation in the two cases :

Case i : If $v_k = y_k$, we get :

$$\tilde{y}_k = y_k + \frac{2(f(x_k) - f(x_{k+1})) - s_k^T y_k}{s_k^T y_k} y_k \tag{13}$$

Case ii : If $v_k = g_{k+1}$, we get :

$$\tilde{y}_k = y_k + \frac{2(f(x_k) - f(x_{k+1})) - s_k^T y_k}{s_k^T g_{k+1}} g_{k+1} \tag{14}$$

The new algorithm can be staged formally as follows.

New Algorithm :

Stage 1 : Select $x_0 \in R^n$ and $H_0 = I$. Set $k = 0$.

Stage 2 : Stop if $\|g_k\| = 0$. $s_k^T \tilde{y}_k > 0$

Stage 3 : Generate d_k by $d_k = -H_k g_k$.

Stage 4 : Find a α_k based on the Wolfe condition satisfies :

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k \tag{15}$$

$$d_k^T g(x_k + \alpha_k d_k) \ge \sigma d_k^T g_k \tag{16}$$

Stage 5 : Variable update, $x_{k+1} = x_k + \alpha_k d_k$.

Stage 6 : If $s_k^T \tilde{y}_k > 0$, find update H_{k+1} by the formula (6) and (12), otherwise let $H_{k+1} = H_k$.

Stage 7 : Let $k = k+1$. Go to stage 2.

We verify the positive definite property of the update formula for the quasi-Newton method in the next theorem.

Theorem 1.

If and if only $s_k^T \tilde{y}_k > 0$, then the update H_{k+1} is positive definite.

Proof :

The gradient-difference vector y_k, we define by :

$$\tilde{y}_k = y_k + \frac{2(f(x_k) - f(x_{k+1})) - s_k^T y_k}{s_k^T v_k} v_k \tag{17}$$

Multiplying above equation by s_k^T, we obtain :

$$s_k^T \tilde{y}_k = s_k^T y_k + 2(f(x_k) - f(x_{k+1})) - s_k^T y_k \quad (18)$$

From above equation we get :

$$s_k^T \tilde{y}_k = 2(f(x_k) - f(x_{k+1})) \quad (19)$$

Using first Wolfe condition in above equation, we obtain :

$$s_k^T \tilde{y}_k \geq -2\delta\alpha_k g_k^T d_k \quad (20)$$

In fact, $s_k^T g_k = \alpha_k d_k^T g_k < 0$, such that :

$$s_k^T \tilde{y}_k \geq -2\delta\alpha_k d_k^T g_k > 0 \quad (21)$$

The proof is complete.

3. CONVERGENCE ANALYSIS

Now by using under the following assumption, we twist to the convergence result of the new methods.

Assumptions :

a) f is bounded on the set $S = \{x \in R^n | f(x) \leq f(x_0)\}$ and is bounded below on R^n.

b) If there exits a nonnegative constant L such that :

$$\|g(z) - g(w)\| \leq L \|z - w\|, \quad \forall \ z, w \in R \quad (22)$$

and g is called Lipschitz continuous and we get :

$$\|g(x)\| \leq \gamma, \quad \forall \ x \in R \quad (23)$$

If $\{f(x_k)\}$ is a decreasing, then $\{x_k\}$ is contained in R and the existence of x^* we get :

$$\lim_{k \to \infty} f(x_k) = f(x^*) \quad (24)$$

In veracity, that sequence x_k is restricted, there exist some positive constant μ such that :

$$\|s_k\| = \left\|x - \bar{x}\right\| \leq \|x\| + \left\|\bar{x}\right\| \leq \mu. \quad (25)$$

For more details see [11].

Theorem 2.
If the following inequality holds :

$$\|B_k s_k\| \leq a_1 \|s_k\| \quad \text{and} \quad s_k^T B_k s_2 \geq a_2 \|s_k\|^2, \quad (26)$$

where $a_1 > 0$ and $a_2 > 0$ are constants and the $\{x_k\}$ be generated by new methods. For infinitely k, then we get:

$$\lim_{k\to\infty} \inf \ \|g_k\| = 0 \tag{27}$$

Proof :
Using (26) and adding with $g_k = -B_k d_k$, we gives :

$$a_2\|d_k\| \le \|g_k\| \ \le \ a_1\|d_k\| \quad \text{and} \quad d_k^T B_k d_2 \ge a_2\|d_k\|^2 \tag{28}$$

The two conditions form Wolfe rule (15), (16) and (28) to obtain :

$$-(1-\sigma)g_k^T d_k \le (g_{k+1} - g_k)^T d_k \le L\alpha_k\|d_k\|^2 \tag{29}$$

Then (29) implies that :

$$\alpha_k \ \ge \ \frac{-(1-\sigma)g_k^T d_k}{L\|d_k\|^2} = \frac{(1-\sigma)d_k^T B_k d_k}{L\|d_k\|^2} \ge \frac{(1-\sigma)\ a_2}{L} \tag{30}$$

Accordingly (24), we have :

$$\sum_{k=1}^{\infty}(f_k - f_{k+1}) = \lim_{N\to\infty}\sum_{k=1}^{\infty}(f_k - f_{k+1}) = \lim_{N\to\infty}(f_1 - f_{k+1}) = f_1 - f^* \tag{31}$$

from which it follows that :

$$\sum_{k=1}^{\infty}(f_k - f_{k+1}) \le +\infty, \tag{32}$$

merge them with Wolfe law (15) leads :

$$\sum_{k=1}^{\infty} -\alpha_k g_k^T d_k \le +\infty \tag{33}$$

We obtained :

$$\lim_{k\to\infty} \ \alpha_k g_k^T d_k = 0 \tag{34}$$

above equally with (34) give that :

$$\lim_{k\to\infty} \ d_k^T B_k d_k = \lim_{k\to\infty} \ -g_k^T d_k = 0 \tag{35}$$

Merge (35) with (28) we get the finale (27). If f is non-convex function, we need some assumptions on the update, may can lose the positive definiteness. For every k, we define the index set K

$$K = \left\{ k \ : \ \frac{s_k^T \bar{y}_k}{\|s_k\|} \ge \beta\|g_k\|^{\delta} \right\}, \tag{36}$$

where $\beta > 0$ is constant and $\delta > 0$ is bounded. The next lemma quoted in [12] is very significant to study the convergence property.

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 :

$$\frac{\|y_k\|^2}{s_k^T y_k} \le M \tag{37}$$

then :

$$\lim_{k\to\infty} \inf \|g_k\| = 0 \tag{38}$$

A cautious update rule similar to the above lemma.

Theorem 3.

Suppose $\{x_k\}$ be generated by the new method. Then we get :

$$\lim_{k\to\infty} \inf \|g_k\| = 0 \tag{39}$$

Proof :

In view of Theorem 2., sufficiently show that (26) holds for infinitely k. If K is a finite set, then B_k is a constant-matrix, obviously, (26) satisfies. Now, if K is a infinite we will deduce a contradiction with there exists $\varepsilon > 0$ such that :

$$\|g_k\| > \varepsilon \tag{40}$$

It follows from (36), we obtain :

$$s_k^T \tilde{y}_k \ge \beta \varepsilon^{\delta} \|s_k\| \tag{41}$$

Using definition of $\tilde{y}_k$, we have :

$$\begin{aligned} \|\tilde{y}_k\| &= \left\| y_k + \frac{2(f_k - f_{k+1}) - s_k^T y_k}{s_k^T v_k} v_k \right\| \\ &\le \|y_k\| \end{aligned} \tag{42}$$

It follows from (22) and (42) we obtain :

$$\|\tilde{y}_k\| \le \|y_k\| \le L\|s_k\| \tag{43}$$

This, together with (41), lead to :

$$\frac{\|\tilde{y}_k\|^2}{s_k^T \tilde{y}_k} \le M \tag{44}$$

Applying lemma 1, to the $\{B_k\}_{k\in K}$, there exist, a_1 and a_2, we obtain (21) for infinitely many k. Then, proof is finished.

4. NUMERICAL EXPERIMENTS

It has been programmed subroutine Matlab to test the modified BFGS algorithm presented in the previous section. We tested the algorithm on the following problems that have been taken from [13]. Different test functions have been used in different researchs such as [14-24].

All the problems are being resolved successfully, and numbers are given duplicates and job evaluation in Table 1. We have solved these problems through the BFGS algorithm, as shown in Table 1 the numerical results of the new algorithm. The Himmeblau [25], stop rule is used : If $|f(x_k)| > 10^{-5}$, let $stop\ 1 = |f(x_k) - f(x_{k+1})| \ / \ |f(x_k)|$; Otherwise, let $stop\ 1 = |f(x_k) - f(x_{k+1})|$. For every problem, if $\|g_k\| < \varepsilon$ or $stop\ 1 < 10^{-5}$ is satisfied, the program will be stopped. To compare the efficiency of roads in Table 1, we adopt the number of iterations (NI) and the number of evaluations of jobs (NF). Numerically results show that the new algorithm is a little better than the usual BFGS algorithm in this group of test problems.

Table 1. Numerical results

		BFGS algorithm		BFGS with $v_k = y_k$		BFGS with $v_k = g_{k+1}$	
Problems	n	NI	NF	NI	NF	NI	NF
Froth	2	9	26	9	26	11	32
Badscp	2	43	166	34	125	3	31
Badscb	2	3	30	3	30	3	30
Beale	2	15	50	13	43	13	38
Jensam	2	2	27	2	27	2	27
Helix	3	34	113	24	80	17	49
Bard	3	16	54	17	56	12	36
Gauss	3	2	4	2	4	2	4
Gulf	3	2	27	2	27	2	27
Box	3	2	27	2	27	2	27
Sing	4	20	60	14	44	11	35
Wood	4	19	61	19	61	7	22
Kowosb	4	21	65	23	117	10	28
Bd	4	17	54	16	50	8	27
Osb1	5	2	27	2	27	2	27
Biggs	6	25	72	8	50	9	27
Osb2	11	3	31	3	31	3	31
Watson	20	31	102	34	108	6	20
Rosex	100	231	806	46	197	33	765
Singx	400	64	209	126	397	10	32
Pen1	400	2	27	2	27	2	27
Pen2	200	2	5	2	5	2	5
Vardi	100	2	27	2	27	2	27
Trig	500	9	33	8	28	14	42
Bv	500	2	4	2	4	2	4
Ie	500	6	16	7	19	13	37
Band	500	57	281	15	81	5	16
Lin	500	2	4	2	4	2	4
Lin1	500	3	7	3	7	3	7
Lin0	500	3	7	3	7	3	7
Total		649	2422	445	1736	214	1491

Computational results show that a reduction of (31-67)% and (28-38)% in terms of the total number of iterations and function evaluations respectively. Relative efficiency of the new algorithms as shown in Table 2.

Table 2. Relative efficiency of the new algorithms

	BFGS algorithm	BFGS with $v_k = y_k$	BFGS with $v_k = g_{k+1}$
NI	100 %	68.56 %	32.97 %
NF	100 %	71.67 %	61.56 %

5. CONCLUSIONS

In this paper, we deriving a new quasi-Newton equation based on the second-order Taylor's series approximation of the function. We conclude by affirming that the arithmetical findings in this work was effective to solving optimization problems.

ACKNOWLEDGMENT

The authors are very 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.

REFERENCES

[1] D. P. Bertsekas, "Nonlinear Programming," Belmont, Massachusetts, Athena Scientific, 1996.

[2] G. Yuan and Z. Wei, "Convergence analysis of a modified BFGS method on convex minimizations," *Computational Optimization and Applications*, vol. 47, pp. 237-255, 2010.

[3] Y. Dai, "Convergence properties of the BFGS algorithm," *SIAM Journal on Optimization*, vol. 13, no. 3, pp. 693-701, 2002.

[4] Y. Xiao and Z. Wei, "A Modified BFGS Method Without Line Searches for Nonconvex Unconstrained Optimization," *Advances in Theoretical and Applied Mathematics*, vol. 1, pp. 149-162, 2006.

[5] G. Yuan, et al., "The global convergence of a modified BFGS method for nonconvex functions," *Journal of Computational and Applied Mathematics*, vol. 327, pp. 274-294, 2018.

[6] F. Biglari, et al., "New quasi-Newton methods via higher order tensor models," *Journal of Computational and Applied Mathematics*, vol. 235, no. 8, pp. 2412-2422, 2011.

[7] L. H. Chen, et al., "A modified quasi-Newton method for structured optimization with partial information on the Hessian," *Computational Optimization and Applications*, vol. 35, pp. 5-18, 2006.

[8] Z. Wei, et al., "New quasi-Newton methods for unconstrained optimization problems," *Applied Mathematics and Computation*, vol. 175, no. 2, pp. 1156-1188, 2006.

[9] Basim A. H., "A modified quasi-Newton methods for unconstrained Optimization," *Journal of pure and applied mathematics*, vol. 42, pp. 504-511, 2019.

[10] Basim A. H. and Hussein K. K., "A new class of BFGS updating formula based on the new quasi-newton equation," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 3, pp. 945-953, 2019.

[11] R. Dehghani, et al., "The modified quasi-Newton methods for solving unconstrained optimization problems," International Journal of Numerical Modelling: Electronics Networks, Devices and Fields, vol. 32, no. 1, pp. 1-8, 2018.

[12] M. J. D. Powell, "Some global convergence properties of a variable metric algorithm for minimization without exact line searches," in R. W. Cottle and C. E. Lemke, eds, *Nonlinear Programming, SIAM-AMS Proceedings*, vol. 4, pp. 53-72, 1976.

[13] J. J. More, et al., "Testing unconstrained optimization software," *ACM Transactions on Mathematical. Software*, vol. 7, no. 1, pp. 17-41, 1981.

[14] B. A. Hassan and H. N. Jabbar, "A New Transformed Biggs 's Self-Scaling Quasi-Newton Method for Optimization," *Zanco Journal of Pure and Applied Sciences*, vol. 31, no. 2, pp. 1-5, 2019.

[15] R. H. Byrd and J. Nocedal, "A tool for the analysis of quasi-Newton methods with application to unconstrained minimization," *SIAM Journal on Numerical Analysis*, vol. 26, no. 3, pp. 727-739, 1989.

[16] Basim A. H., et al., "A descent extension of the Dai - Yuan conjugate gradient technique," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 2, pp. 661-668, 2019.

[17] Basim A. H., "A new type of quasi-Newton updating formulas based on the new quasi-Newton equation," *Numerical Algebra, Control and Optimization*, vol. 10, no. 2, pp. 227-235, 2020.

[18] Basim A. H., et al., "A new kind of parameter conjugate gradient for unconstrained optimization," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 17, no. 1, pp. 404-411, 2020.

[19] Basim A. H., et al., "A Class of Descent Conjugate Gradient Methods for Solving Optimization Problems," *Applied Mathematical Sciences*, vol. 13, no. 12, pp. 559-567, 2019.

[20] Basim A. H. and Mohammed W. T., "A Modified Quasi-Newton Equation in the Quasi-Newton Methods for Optimization," *Applied Mathematical Sciences*, vol. 13, no. 10, pp. 463-472, 2019.

[21] Basim A. H., "A Globally Convergence Spectral Conjugate Gradient Method for Solving Unconstrained Optimization Problems," *Al-Rafidain Journal of Computer Sciences and Mathematics*, vol. 10, no. 4, pp. 21-28, 2013.

[22] Basim A. H., "Development a Special Conjugate Gradient Algorithm for Solving Unconstrained Minimization Problems," *Al-Rafidain Journal of Computer Sciences and Mathematics*, vol. 9, no. 1, pp. 73-84, 2012.

[23] Basim A. H. and Omer M. E., "A New sufficient descent Conjugate Gradient Method for Nonlinear Optimization," *Iraqi Journal of Statistical Sciences*, vol. 14, no. 26, pp. 12-24, 2014.

[24] Basim A. H., "A new formula for conjugate parameter computation based on the quadratic model," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 13, no. 3, pp. 954-961, 2019.

[25] Y. Yuan and W. Sun, "Theory and Methods of Optimization," Science Press of China, 1999.