Indonesian Journal of Electrical Engineering and Computer Science
Vol. 19, No. 2, August 2020, pp. 701~708
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v19.i2.pp701-708

A new variants of quasi-newton equation based on the quadratic function for unconstrained optimization

Basim A. Hassan, Mohammed W. Taha
Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, Iraq

Article Info

Article history:

Received Jan 14, 2020
Revised Feb 23, 2020
Accepted Mar 9, 2020

Keywords:

Global convergence
New quasi-Newton equation
Quasi-Newton methods

ABSTRACT

The focus for quasi-Newton methods is the quasi-Newton equation. A new quasi-Newton equation is derived for quadratic function. Then, based on this new quasi-Newton equation, a new quasi-Newton updating formulas are presented. Under appropriate conditions, it is shown that the proposed method is globally convergent. Finally, some numerical experiments are reported which verifies the effectiveness of the new method.

Corresponding Author:

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

1. INTRODUCTION

The quasi-Newton (QN) method is a willing device to find the minimum value of problem. Let us think over problem for the following form:

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

where $f : R^n \to R$ is continuously differentiable. For more details can be found in [1].

The basic form of a quasi-Newton method for solving (1). and it is written as:

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

where α_k is the step length which satisfies certain Wolfe conditions:

$$\text{“ } f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k \text{ ”} \tag{3}$$

$$\text{“} d_k^T g(x_k + \alpha_k d_k) \ge \sigma \, d_k^T g_k \tag{4}$$

for some parameters $0 < \delta < \sigma < 1$. For details see [2].

The search direction of the quasi-Newton method is defined by:

$$B_k d_k + g_k = 0 \tag{5}$$

where B_k is an approximation matrix of the Hessian, and g_k denotes $g(x_k)$. The matrix $\{B_k\}$ are positive definite and they satisfy the quasi-Newton equation:

$$B_{k+1} s_k = y_k \tag{6}$$

where $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$, is satisfied. For details see [3].

One possibility to obtain B_{k+1} from B_k by an update formula is to have a BFGS update:

$$B_{k+1}^{BFGS} = B_k - \frac{B_k s_k s_k^T B_k^{\ T}}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} \tag{7}$$

Precisely, if $H_k^{-1} = B_k$, $B_{k+1} = U(B_k, y_k, s_k)$ and $H_{k+1} = U(H_k, s_k, y_k)$ then $H_{k+1}^{-1} = B_{k+1}$, this property called duality transformation. Applying this relation to the BFGS method, we will get the dual of the BFGS formula:

$$H_{k+1}^{BFGS} = H_k - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k} + \left[1 + \frac{y_k^T H_k y_k}{s_k^T y_k}\right] \frac{s_k s_k^T}{s_k^T y_k} \tag{8}$$

For details see [4].

Different quasi-Newton methods correspond to different ways of updating the matrix H_{k+1} to include the new curvature information obtained during the k^{th} iteration. Different Improving of these methods are made, in the aim to develop them in [5-8]. We are going to present a new quasi-Newton equation based on quadratic function and show that these methods are the best for solving unconstrained.

2. DERIVING QUASI-NEWTON EQUATION BASED ON THE QUADRATIC FUNCTION:

Before presenting new quasi-Newton equation, we shall derive an estimate of a step size. Assume that the function is defined a quadratic function of the form:

$$f_{k+1} = f_k + \alpha_k g_k^T d_k + \frac{1}{2} \alpha_k^2 d_k^T Q d_k \tag{9}$$

Then the minimum point α_k of above function is given by:

$$\alpha_k = \frac{-g_k^T d_k}{d_k^T Q d_k} \tag{10}$$

Now, we shall derive alternate quasi-Newton equation. Therefore, from above equation we get:

$$\alpha_k d_k^T Q d_k = -g_k^T d_k \tag{11}$$

The basic idea is to approximate either Q by another matrix B_{k+1} to obtain a higher accuracy, can be expressed as:

$$s_k^T B_{k+1} s_k \quad = - s_k^T g_k \tag{12}$$

By the following definition of the different gradient y_k we get:

$$y_k^T s_k = g_{k+1}^T s_k - g_k^T s_k \tag{13}$$

By putting (13) in (12) we get:

$$s_k^T B_{k+1} s_k \quad = y_k^T s_k - g_{k+1}^T s_k \tag{14}$$

One of the possible choices in approximation of $B_{k+1} s_k$ can be given by:

$$B_{k+1} s_k \quad = \bar{y}_k \,, \quad \bar{y}_k = y_k - \frac{g_{k+1}^T s_k}{s_k^T z_k} z_k \tag{15}$$

where $s_k^T z_k \neq 0$ and z_k is any vector.

The value of z_k are not unwavering in a unique technique, but the suitable choice are $z_k = y_k$ and $z_k = g_k$.

If f is non-convex may mislay the positive definiteness. We need for some extra assumptions on the update.

In order to prove the following theorems, we define the index set K as:

$$K = \left\{ k \;:\; \frac{s_k^T \bar{y}_k}{\|s_k\|} \geq \beta \|g_k\|^{\delta} \right\}, \tag{16}$$

where $\beta > 0$ is constant and $\delta > 0$ is bounded.

Now, we give the algorithm of the new method.

Assumptions: $x_0 \in R^n$. Let $k = 0$.

Stage 1: if $\|g_k\| = 0$, then stop.

Stage 2: Solve $H_k^{-1} d_k = - g_k$ for d_k.

Stage 3: Determine the step size α_k such that Wolfe line search rules hold.

Stage 4: Set $x_{k+1} = x_k + \alpha_k d_k$. If $s_k^T \bar{y}_k > 0$, update H_{k+1} by the formula (8) and (15), otherwise let $H_{k+1} = H_k$.

Stage 5: $k = k + 1$, return to stage 1.

The next theorem is very important. We prove the condition $s_k^T y_k > 0$ which is known as the curvature condition

Theorem 1.

The new update (8) and (15) retains positive definiteness if and only if $s_k^T \bar{y}_k > 0$.

Proof:

Having in view that the definition of the different gradient by:

$$\bar{y}_k = y_k - \frac{g_{k+1}^T s_k}{s_k^T z_k} z_k \tag{17}$$

Multiplying above equation by s_k^T, we have:

$$\begin{aligned} s_k^T \bar{y}_k &= s_k^T y_k - g_{k+1}^T s_k \\ &= -s_k^T g_k \end{aligned} \quad (18)$$

From (2) we get $s_k = \alpha_k d_k$. In fact, the search direction of a QN is descent, we noting that $s_k^T g_k = \alpha_k d_k^T g_k < 0$, such that:

$$s_k^T \bar{y}_k = -\alpha_k d_k^T g_k > 0 \quad (19)$$

The proof is complete.

3. CONVERGENCE ANALYSIS

In order to prove the convergence, we consider the following assumptions: The level set "$D = \{x \;/ f(x) \le f(x_0)\}$", with x_0 is an initial point of iterative method is restricted.

Assumption A. Using Lipschitz continuous; that is exist constants L and γ, such that:

$$\|\nabla f(x) - \nabla f(y)\| \le L \|x - y\|, \quad \forall\ x,\ y \in D \quad (20)$$

And

$$\|\nabla f(x)\| \le \gamma, \quad \forall\ x \in D \quad (21)$$

Since $\{f(x_k)\}$ is a decreasing, which ensures $\{x_k\}$ is contained in D and the existence of x^* we get:

$$\lim_{k\to\infty} f(x_k) = f(x^*) \quad (22)$$

In reality, that sequence x_k is restricted, there exist some positive constant μ such that for all k, :

$$\|s_k\| = \|x - \bar{x}\| \le \|x\| + \|\bar{x}\| \le \mu \quad (23)$$

For more details see [9,10].

Existing the handy theorem to explain that our method is globally convergent.

Theorem 2.

If $\|\nabla f(x)\| \le \gamma$ is not holds for all k. Let $\{x_k\}$ be generated by new methods, and the following inequality holds:

$$\|B_k s_k\| \le a_1 \|s_k\| \text{ and } s_k^T B_k s_k \ge a_2 \|s_k\|^2, \quad (24)$$

where $a_1 > 0$ and $a_2 > 0$ are constants. For infinitely k, then we have:

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \quad (25)$$

Proof:

By (4) of the Wolfe conditions we obtain:

$$(g_{k+1}-g_k)^T d_k \geq -(1-\sigma) g_k^T d_k \tag{26}$$

Moreover, from Lipschitz condition we obtain:

$$(g_{k+1}-g_k)^T d_k \leq L\,\alpha_k \|d_k\|^2 \tag{27}$$

Combining above two equation we get:

$$\alpha_k \geq \frac{-(1-\sigma) g_k^T d_k}{L\|d_k\|^2} = \frac{(1-\sigma) d_k^T B_k d_k}{L\|d_k\|^2} \geq \frac{(1-\sigma)\ a_2}{L} \tag{28}$$

using (22), we obtain:

$$\sum_{k=1}^{\infty}(f_k - f_{k+1}) = \lim_{N\to\infty}\sum_{k=1}^{N}(f_k - f_{k+1}) = \lim_{N\to\infty}(f_1 - f_{N+1}) = f_1 - f^* \tag{29}$$

Therefore,

$$\sum_{k=1}^{\infty}(f_k - f_{k+1}) \leq +\infty, \tag{30}$$

which combining with Wolfe condition (3) yields:

$$\sum_{k=1}^{\infty} -\alpha_k g_k^T d_k \leq +\infty \tag{31}$$

Then:

$$\lim_{k\to\infty}\ \alpha_k g_k^T d_k = 0 \tag{32}$$

together with (31) provide that:

$$\lim_{k\to\infty}\ d_k^T B_k d_k = \lim_{k\to\infty}\ -g_k^T d_k = 0 \tag{33}$$

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

For proof a global convergence for non-convex problems, we state a lemma due to Powell [11].

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”:

$$\frac{\|y_k\|^2}{s_k^T\ y_k} \leq M \tag{34}$$

then:

$$\lim_{k\to\infty} \inf \|g_k\| = 0 \quad (35)$$

For more details can be found in [12].

Theorem 3.

Let $\{x_k\}$ be generated by the new method. Then we have:

$$\lim_{k\to\infty} \inf \|g_k\| = 0 \quad (36)$$

Proof:

If K is a finite set, then B_k is a constant-matrix, obviously, (24) satisfies. Now, if K is a infinite we will deduce a contradiction with there exists $\varepsilon > 0$ such that:

$\|g_k\| > \varepsilon$ (37)

It follows from (28) and (16) that:

$$s_k^T \bar{y}_k \geq \beta \varepsilon^{\delta} \|s_k\| \quad (38)$$

By the definition of $\bar{y}_k$, we have:

$$\begin{aligned} \|\bar{y}_k\| &= \left\| y_k - \frac{g_{k+1}^T s_k}{s_k^T z_k} z_k \right\| \\ &\leq \|y_k\| + \frac{\|g_{k+1}\| \|s_k\|}{\|s_k\|} \end{aligned} \quad (39)$$

It follows from (20),(23), (28) and (39) that:

$$\|\bar{y}_k\| \leq L\|s_k\| + \|g_{k+1}\| \leq L\mu + \gamma \quad (40)$$

This, together with (38), lead to:

$$\frac{\|\bar{y}_k\|^2}{s_k^T \bar{y}_k} \leq M \quad (41)$$

Using lemma 1, to the sub $\{B_k\}_{k\in K}$, obviously, existence a_1 and a_2 we get (24) for infinitely many k. By using theorem 2 the proof is Finished.

4. TESTS NUMERICALLY FOR METHODS AND DISCUSSIONS

In this section, some numerical tests are performed in order to illustrate the implementation and efficiency of the proposed method. In our tests, we consider the unconstrained optimization problems from the set provided by [13]. Other test functions have been used in various research such as [14-24]. In solved the problems required the number of iterations (NI) and the number of function evaluations (NF), respectively, which contain in tables. All runs reported in this paper terminate when:

"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". For more details can be found in [25]. Comparison the new methods with the standard BFGS method as shown in Table 1.

Table 1. Comparison the new methods with the standard BFGS method

		BFGS algorithm		BFGS with $z_k = y_k$		BFGS with $z_k = g_k$	
Problems	N	NI	NF	NI	NF	NI	NF
Rose	2	35	140	33	127	24	82
Froth	2	9	26	8	23	8	23
Badscp	2	43	166	34	133	36	129
Badscb	2	3	30	3	30	3	30
Beale	2	15	50	14	47	12	37
Jensam	2	2	27	2	27	2	27
Helix	3	34	113	33	105	29	88
Bard	3	16	54	18	58	16	50
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	19	58	18	53
Wood	4	19	61	20	62	13	41
Kowosb	4	21	65	22	98	21	69
Bd	4	17	54	16	49	15	48
Osb1	5	2	27	2	27	2	27
Biggs	6	25	72	25	77	7	43
Osb2	11	3	31	3	31	3	31
Watson	20	31	102	32	101	26	78
Singx	400	64	209	51	170	16	49
Pen1	400	2	27	2	27	53	164
Pen2	200	2	5	2	5	2	5
Vardim	100	2	27	2	27	2	27
Trig	500	9	33	9	32	8	24
Bv	500	2	4	2	4	2	4
Ie	500	6	16	7	19	7	19
Band	500	57	281	36	180	11	65
Lin	500	2	4	2	4	2	4
Lin1	500	3	7	3	7	3	7
Lino	500	3	7	3	7	3	7
Total		453	1756	409	1593	350	1289

Based on the above comparisons, it indicates our method has improved rate (9-26) % in total number of the iterations and (9-27) % in total number of the function evaluation by compare with the BFGS method. Relative efficiency of the between algorithms as shown in Table 2.

Table 2. Relative efficiency of the between algorithms

	BFGS algorithm	BFGS with $z_k = y_k$	BFGS with $z_k = g_k$
N1	100%	90.28%	77.26%
NF	100%	90.71%	73.40%

5. CONCLUSIONS

Based on a quadratic function we deriving new quasi-Newton equation. The effectiveness of the proposed methods have been shown by some numerical examples. Exciting the remarkable performance of method with choice are $z_k = y_k$ and $z_k = g_k$ on these test problems, it is okay to guess on $z_k = g_{k+1}$.

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] Luksan L. and Vlcek J., "New variable metric methods for unconstrained minimization covering the large-scale case", Technical report No. V 876, 2002.
[2] Wolfe P., "Convergence conditions for ascent methods, (II): some corrections", SIAM Review, vol. 13, pp. 185-188, 1971.
[3] Sun,W., Yuan, Ya-Xiang "Optimization Theory and Methods: Nonlinear Programming", *Springer*, New York 2006.
[4] Jean C. G., "Numerical methods for large-scale minimization", Inria,, Rocquencourt, France, pp.106.
[5] Basim A.H., "A new type of quasi-Newton updating formulas based on the new quasi-Newton equation", *SIAM J. Numerical algebra, control and optimization*. To be apper. 2019.
[6] Biglari F., Hassan M.A., and Leong W. J., "New quasi-Newton methods via higher order tensor models", *J. Comput. Appl. Math*., (8) pp. 2412–2422, 2011.
[7] Chen L.H., Deng N.Y., and. Zhang J.Z, "A modified quasi-Newton method for structured optimization with partial information on the Hessian", *Comput. Optim. Appl*. 35, pp. 5–18, 2006.
[8] Wei Z., Li G., and Qi L., "New quasi-Newton methods for unconstrained optimization problems", *Appl. Math. Comput*.,175, pp. 1156–1188, 2006.
[9] Dehghani R., Hosseini M and Bidabadi N., "The modified quasi-Newton methods for solving unconstrained optimization problems", *WILEY*, pp.1-8, 2017.
[10] Zahra K. and Ali A., "A new modified scaled conjugate gradient method for large-scale unconstrained optimization with non-convex objective function", *Optimization Methods and Software*, pp.1-14, 2018.
[11] Powell M.J.D., "Some global convergence properties of a variable metric algorithm for minimization without exact line searches, in Nonlinear Programming", *SIAM-AMS Proceedings, Lemke, eds., SIAM*, 1976, pp.53-72, 1976.
[12] Byrd R., Nocedal J., "A tool for the analysis of quasi-Newton methods with application to unconstrained minimization", *SIAM J. Numer. Anal*. 26, PP.727–739, 1989.
[13] More J., Garbow B., and Hillstrome K., "Testing unconstrained optimization software", *ACM Trans. Math. Software*, 7,pp. 17-41, 1981.
[14] Basim A. Hassan and Hawraz N. Jabbar, "A New Transformed Biggs 's Self-Scaling Quasi-Newton Method for Optimization", *ZANCO Journal of Pure and Applied Sciences*, 31:1-5, 2018.
[15] Basim A. H., 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*, 3: 945-953, 2019.
[16] Basim A. H., Zeyad M. A. and Hawraz N. J. "A descent extension of the Dai - Yuan conjugate gradient technique", *Indonesian Journal of Electrical Engineering and Computer Science*, 2: 661-668, 2019.
[17] Basim A. H., "A modified quasi-Newton methods for unconstrained Optimization", *Italian journal of pure and applied mathematics*, 42, 504-511, 2019.
[18] Basim A. H. Hussein O. D.and Azzam S. Y. "A new kind of parameter conjugate gradient for unconstrained optimization", *Indonesian Journal of Electrical Engineering and Computer Science*, 17: 404-411, 2020.
[19] Basim A. Hassan, Osama M.T. W. and Ayad A. M. "A Class of Descent Conjugate Gradient Methods for Solving Optimization Problems", HIKARI Ltd, Applied Mathematical Sciences, 12, 559 – 567, 2019.
[20] Basim A. Hassan and Mohammed W. T. "A Modified Quasi-Newton Equation in the Quasi-Newton Methods for Optimization", HIKARI Ltd, Applied Mathematical Sciences, 10, 463 – 472, 2019.
[21] Basim A. Hassan, "A Globally Convergence Spectral Conjugate Gradient Method for Solving Unconstrained Optimization Problems and Mohammed", *Raf. J. of Comp. & Math's*., 10, 21-28, 2013.
[22] Basim A. Hassan, "Development a Special Conjugate Gradient Algorithm for Solving Unconstrained Minimization Problems", *Raf. J. of Comp. & Math's*., 9, 73-84, 2012.
[23] Basim A. Hassan and Omer M. E., "A New sufficient descent Conjugate Gradient Method for Nonlinear Optimization", *Iraqi Journal of Statistical Sciences*, 26, 12-24, 2014.
[24] Basim A. Hassan, "A new formula for conjugate parameter computation based on the quadratic model", *Indonesian Journal of Electrical Engineering and Computer Science*, 3, pp. 954-961, 2019.
[25] Yuan Y. and Sun W., "Theory and Methods of Optimization", Science Press of China, 1999.