Hybrid Quasi-Newton with New Conjugate Gradient using Exact Line Search

Ummie Khalthum Mohd Yusof, Mohd Asrul Hery Ibrahim, Mohd Rivaie, Mustafa Mamat, Mohamad Afendee Mohamed, Puspa Liza Ghazali

Abstract: Until now, Quasi-newton (QN) method is the most well-known method for solving unconstrained optimization problem. This method consumes lesser time as compared to Newton method since it is unnecessary to compute Hessian matrices. For QN method, BFGS is the best solver in finding the optimum solutions. Therefore, a new hybrid coefficient which possesses the convergence analysis computed by exact line search is introduced. This new hybrid coefficient is numerically proven by producing the best outcomes with least iteration number and CPU time.

Index Terms: Quasi-Newton method, sufficient descent condition and global convergence, unconstrained optimization.

I. INTRODUCTION

QN methods was developed based on Newton's method, and it is still used for solving unconstrained optimization problems. The most popular quasi-Newton method is BFGS. In 1970, BFGS was suggested independently by Broyden, Fletcher, Goldfard and Shanon after Rank One in 1959 and DFP formula in 1963. Rank one and DFP formula have the tendency in solving the larger non-quadratic problems. The BFGS method was an improvement of both rank one and DFP formula [8]-[11], [13]. This formula is known to robust with the sloppy line search. Next, the conjugate gradient algorithm is which known as an iterative method may produces better approximation to the minimum of a general unconstrained nonlinear function of N variables at each iteration. This algorithm initiated by Hestenes-Stieffel in 1952. With the simple algorithm, the conjugate gradient method is easy to be implemented. The main target of unconstrained optimization is quoted as

$$\min_{m\in R^n} f(m) \quad (1)$$

that is finding the minimum value for a function $f(m)$ such that $f: R^n \to R$ is continuously differentiable and R^n represents n-dimensional Euclidean space and. The iterative methods are used to solve (1). For each q^{th} iteration, an approximation point, m_q and the $(q+1)^{th}$ iteration is related as

$$m_{q+1} = m_q + \alpha_q d_q \quad (2)$$

where m_{q+1}, α_q, d_q each represent the new iterate point, step size and search direction respectively. The d_q considered is as

$$d_q = \begin{cases} -H_q g_q & \text{if } m = 0 \\ -H_q g_q + \eta\left(-g_q + \beta_q d_{q-1}\right) & \text{if } m \geq 1 \end{cases} \quad (3)$$

where g_q is a gradient of $f(m)$ at point m_q and β_q is CG coefficient.

Previous studies gives valuable intentions to latest researchers come out with new CG coefficient that gain ideas from the classical β_q as listed in Table 1.

Table 1: CG coefficients

CG Coefficients	Years
$\beta_q^{HS} = \dfrac{g_q^T\left(g_q - g_{q-1}\right)}{d_{q-1}^T\left(g_q - g_{q-1}\right)}$	Hestenes-Steifel (1952)
$\beta_q^{FR} = \dfrac{g_q^T g_q}{\left\Vert g_{q-1}\right\Vert^2}$	Fletcher-Reeves (1964)
$\beta_q^{PRP} = \dfrac{g_q^T\left(g_q - g_{q-1}\right)}{\left\Vert g_{q-1}\right\Vert^2}$	Polak- Ribiere (1969)

where g_q and g_{q-1} are gradients of $f(m)$ at points m_q and m_{q-1} respectively, while d_{q-1} is search direction for the previous iteration and $\Vert\cdot\Vert$ is a norm of vectors.

In recent times, researchers keep developing new ideas to outcome the disadvantages of previous formula as in [2], [5], [6], [12], [15], [17]-[19], [20], [23].

The search direction for QN method is the solution of linear system

Revised Manuscript Received on February 11, 2019.

Ummie Khalthum Mohd Yusof, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.

Mohd Asrul Hery Ibrahim, Faculty of Entrepneurship and Business, Universiti Malaysia Kelantan, Kelantan, Malaysia.

Mohd Rivaie, Department of Computer Sciences and Mathematics, Universiti Teknologi MARA, Terengganu, Malaysia.

Mustafa Mamat, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.

Mohamad Afendee Mohamed, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.

Puspa Liza Ghazali, Faculty of Economics and Management Sciences, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$$d_q = -H_q g_q \tag{4}$$

where H_q is known as approximation of Hessian, while initial matrix H_0 is determined by the identity matrix, and successively updated by some formula. This research uses a BFGS formula in a classical algorithm. The update formula for BFGS is

$$H_{q+1} = H_q - \frac{H_q s_q s_q^T H_q}{s_q^T H_q s_q} + \frac{y_q y_q^T}{s_q^T y_q} \tag{5}$$

with $s_q = m_q - m_{q-1}$ and $y_q = g_q - g_{q-1}$. The approximation that the Hessian must fulfill is the curvature condition as below

$$H_{q+1} s_q = y_q \tag{6}$$

Instead of using the inexact line search, this paper focuses on BFGS algorithm with exact line search such that

$$\min_{a \geq 0} f\left(m_q + \alpha_q d_q\right) \tag{7}$$

II. THE NEW COEFFICIENT

A. Propose New Coefficient

In solving unconstrained optimization problems, the calculation of the search direction determines the method to be used. The modification of the search direction has already been introduced by previous researchers [21]. By focusing on QN and CG methods, another new hybrid method is introduced by applying the β^{UAM} (Ummie, Asrul and Mustafa) and denoted as BFGS-UAM [8], [9]. The idea on proposing this new β_q is based on previous researchers [7], [22]. The new CG coefficient is

$$\beta_{q+1}^{UAM} = \max\left\{0, \frac{g_{q+1}^T \left(g_{q+1} - \frac{\|g_q\|}{\|g_{q+1}\|} \right) g_q}{\|g_q\|^2} \right\} \tag{8}$$

Hence, the complete algorithms for the HBFGS methods will be arranged below:

Step 0: Given a starting point m_0 and H_0, select the values for s, β and α and then fix $q = 1$.

Step 1: Exit if $\left\| g\left(m_{q+1}\right) < 10^{-6} \, or \, m \geq 1000 \right\|$

Step 2: Calculate the value of search direction by (3)

Step 3: Compute the value of step size α_q by (7)

Step 4: Compute the difference between $s_q = m_q - m_{q-1}$ and $y_q = g_q - g_{q-1}$

Step 5: Update H_{q-1} by (5) to obtain H_q

Step 6: Set $m = m + 1$ and return to Step 1.

III. CONVERGENCE ANALYSIS

The convergent properties of this method clearly presented in this section with assumption and theorem. Sufficient descent condition and global convergence are established to ensure an algorithm is verified for both conditions.

A. Sufficient Descent Condition

For all $q \geq 0$, every search direction, d_q satisfied the descent property

$$g_q^T d_q < 0 \tag{9}$$

The search direction is said to satisfy the sufficient descent condition if there exists a constant $c_1 > 0$ such that

$$g_q^T d_q \leq c_1 \|g_q\|^2 \tag{10}$$

for all $q \geq 0$.

Theorem 1. Suppose a hybrid method having search direction (3) and β^{UAM} given as (8), then condition (11) will holds for all $q \geq 0$.

Proof. Based on (12), we observe that

$$\begin{aligned} g_q^T d_q &= g_q^T \left(-H_q g_q + \eta\left(-g_q + \beta_q d_{q-1}\right)\right) \\ &= -g_q^T H_q g_q - \eta \|g_q\|^2 + \eta \beta_q g_q^T d_{q-1} \end{aligned} \tag{11}$$

since the line is exact, then $g_q^T d_{q-1} = 0, H_q = \lambda$ and thus,

$$\begin{aligned} &= -\lambda \|g_q\|^2 - \eta \|g_q\|^2 \\ &\leq \left(-\lambda - \eta\right) \|g_q\|^2 \\ &\leq -\left(\lambda + \eta\right) \|g_q\|^2 \\ &\leq c_1 \|g_q\|^2 \end{aligned} \tag{12}$$

where $c_1 = -(\lambda + \eta)$ which is bound away from zero. Therefore, $g_q^T d_q \leq c_1 \|g_q\|^2$ holds and thus complete the proof.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

B. Global Convergence Properties

Importantly, to make sure that there are remain not less than zero to examine the global convergence, then able simplify our new β_q^{UAM} as follows:

$$\beta_{q+1}^{UAM}=\begin{cases}0 \quad \text{for } \left\|g_{q+1}\right\|^2<\frac{\left\|g_q\right\|}{\left\|g_{q+1}\right\|}g_{q+1}^T g_q \\ \frac{g_{q+1}^T\left(g_{q+1}-\frac{\left\|g_q\right\|}{\left\|g_{q+1}\right\|}g_q\right)}{\left\|g_q\right\|^2}\le\frac{\left\|g_{q+1}\right\|^2}{\left\|g_q\right\|^2}\end{cases} \quad (13)$$

Hence,

$$0\le\beta_{q+1}^{UAM}\le\frac{\left\|g_{q+1}\right\|^2}{\left\|g_q\right\|^2} \quad (14)$$

A few assumptions are needed to follow the objective function

Assumption 1. Consider the following.

i. The objective function $f(x)$ is twice continuously differentiable.

ii. The level set L satisfies convex property. Let two positive constants c_1 and c_2 satisfying $c_1\|z\|^2\le z^T F(m)z\le c_2\|z\|^2$ for all $z\in R^n$ where $F(m)$ is the Hessian matrix for f.

iii. The Hessian matrix is Lipschitz continuous at the point m^* such that there exists the positive constant c_3 satisfying all m in a neighborhood of m^*.

Under this assumption, the following Lemma are needed, earlier shown true by Zoutendjik.

Lemma 1. Let the Assumption 1 be true. By considering an iterative CG method and search direction of hybrid method, with d_q as a search direction and α_q satisfies the exact line search. Then, the following condition known as the Zoutendjik condition holds

$$\sum_{q=0}^{\infty}\frac{\left(g_q^T d_q\right)^2}{\left\|d_q\right\|^2}<\infty$$

Theorem 1. [24]

Consider a sequence $\{B_q\}$ generated using BFGS formula, where B_q is symmetric and positive definite, and $y_q^T s_q>0$ for all q. In addition, consider $\{s_q\}$ and $\{y_q\}$ such that

$$\frac{\left\|\left(y_q-G_*\right)s_q\right\|}{\left\|s_q\right\|}\le\varepsilon$$

For some symmetric and positive definite matrix $G(m^*)$ and for some sequence $\{\varepsilon_q\}$ with the property $\sum_{q=1}^{\infty}\varepsilon_q<\infty$.

Then, $\lim_{q\to\infty}\frac{\left\|\left(B_q-G_*\right)d_q\right\|}{\left\|d_q\right\|}=0$ and the sequence $\{\|B_q\|\},\{\|B_q^{-1}\|\}$ are bound.

Theorem 2 (global convergence)

Let Assumption 1 and Theorem 1 be true. Consider any β_q^{UAM} method in the form of (3) and (2) where α_q is a result of an exact line search (7) and the sufficient descent condition also holds. Then, (15) is true.

$$\lim_{q\to\infty}\left\|g_q\right\|^2=0 \text{ or } \sum_{q=0}^{\infty}\frac{\left\|g_q\right\|^4}{\left\|d_q\right\|^2}<\infty \quad (15)$$

Proof: Proof by contradiction. In other words, if Theorem 2 does not hold, then the exists a constant $\partial>0$ such that

$$\left\|g_q\right\|\ge\partial \quad (16)$$

Applying Cauchy Schwartz inequality on equation, yields

$$d_q=-H_q g_q+\eta\left(-g_q+\beta_q^{UAM}d_{k-1}\right)$$
$$=-H_q g_q-\eta g_q+\eta\beta_q^{UAM}d_{k-1}$$
$$\left\|d_q\right\|\le\left\|H_q\right\|\left\|g_q\right\|-|\eta|\left\|g_q\right\|+|\eta|\left\|\beta_q^{UAM}\right\|\left\|d_{q-1}\right\| \quad (17)$$

Divide both sides by $\left\|g_q\right\|^2$ to get

$$\frac{\left\|d_q\right\|}{\left\|g_q\right\|^2}\le\frac{\left\|H_q\right\|}{\left\|g_q\right\|}-\frac{|\eta|}{\left\|g_q\right\|}+\frac{|\eta|\left\|\beta_q^{UAM}\right\|\left\|d_{q-1}\right\|}{\left\|g_q\right\|^2} \quad (18)$$

Substitute $\beta_q^{UAM}=\frac{\left\|g_q\right\|^2}{\left\|g_{q-1}\right\|^2}$, then

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

$$\frac{\|d_q\|}{\|g_q\|^2} \le \frac{\|H_q\|}{\|g_q\|} - \frac{|\eta|}{\|g_q\|} + \frac{|\eta|\|d_{q-1}\|}{\|g_{q-1}\|^2} \quad (19)$$

Since $\|H_q\|$ is bounded, then there exists a positive constant $\mu > 0$ such that $\|H_q\| \le \mu$. Hence, the equation become

$$\frac{\|d_q\|}{\|g_q\|^2} \le \frac{\mu}{\|g_q\|} - \frac{|\eta|}{\|g_q\|} + \frac{|\eta|\|d_{q-1}\|}{\|g_{q-1}\|^2} \quad (20)$$

Note that $d_0 = -H_0 g_0$, then for all q can summarize as

$$\frac{\|d_q\|}{\|g_q\|^2} \le \sum_{p=1}^{q} \eta^{q-p}\left(\frac{\mu+\eta}{\|g_q\|}\right) + \frac{\eta^q \mu}{\|g_0\|}$$

Substitute (16) into the equation and gives

$$\frac{\|d_q\|}{\|g_q\|^2} \le \sum_{p=1}^{q} \eta^{q-p}\left(\frac{\mu+\eta}{\partial}\right) + \frac{\eta^q \mu}{\partial}$$

$$\frac{\|d_q\|}{\|g_q\|^2} \le \left(\frac{\mu+\eta}{\partial}\right)\sum_{p=1}^{q} \eta^{q-p} + \frac{\eta^q \mu}{\partial}$$

$$\frac{\|d_q\|}{\|g_q\|^2} \le \left(\frac{\mu+\eta}{\partial}\right)\sum_{p=0}^{q-1} \eta^{p} + \frac{\eta^q \mu}{\partial}$$

Note that, $\sum_{p=0}^{q-1} \eta^p = \frac{1-\eta^q}{1-\eta}$. Then,

$$\frac{\|d_q\|}{\|g_q\|^2} \le \left(\frac{\mu+\eta}{\partial}\right)\left(\frac{1-\eta^q}{1-\eta}\right) + \frac{\eta^q \mu}{\partial}$$

$$\frac{\|d_q\|}{\|g_q\|^2} \le \frac{\mu+\eta}{\partial(1-\eta)}$$

$$\frac{\|g_q\|^4}{\|d_q\|^2} \ge \left(\frac{\partial(1-\eta)}{\mu+\eta}\right)^2$$

$$\sum_{q=0}^{\infty} \frac{\|g_q\|^4}{\|d_q\|^2} \ge \sum_{q=0}^{\infty} \left(\frac{\partial(1-\eta)}{\mu+\eta}\right)^2 = \infty$$

From theorem 2, it is implied that $\sum_{q=0}^{\infty} \frac{\|g_q\|^4}{\|d_q\|^2} \ge \infty$. This contradicts Lemma 1. Hence conclude the proof.

IV. RESULTS AND DISCUSSION

The test problem listed by [3] has been used to analyze the new BFGS-UAM method. Different dimension varying from 2 to 5000 variables [1]. With a total of 200 test problems, the initial point will continuously subtract from the minimum point. The stopping criteria, $\|g_q\| \le 0$ takes place when the iterations exceeds 10,000 [14].

The numerical result of each algorithm for iteration number and CPU time are displayed in Fig. 1 and 2 respectively, indebted to performance profile brought forward by [4]. In Fig.1, BFGS-UAM has the performance dealing with test functions in order to reach optimal solution compared to others. It is because BFGS-UAM has the highest percentage of solving the test functions. This result explained by the fact that this method gives highest percentage of the test problem solved. Similarly behavior in Fig. 2, β_q^{UAM} is superior among other method. It shows that this method is less time consuming. As an overall, this method is the better in both number of iteration and CPU time.

Table 2: Numerical Results for each methods

Coefficient	Number of Iterations	CPU Times
β^{UAM}	72497	3795.7656
β^{WYL}	124217	4952.048
β^{FR}	135161	4889.596

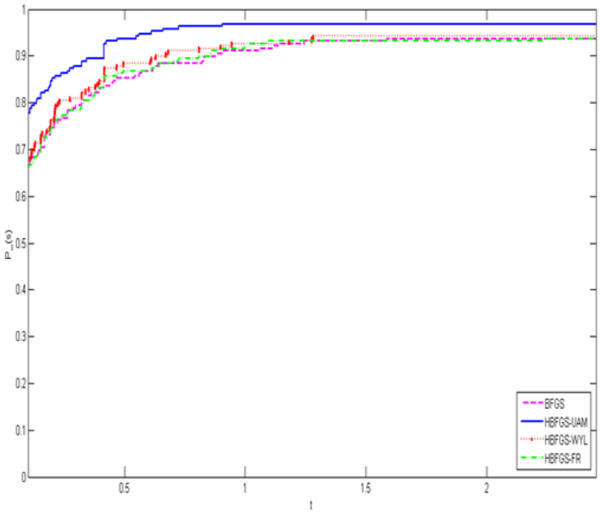

Fig. 1: Performance profile (number of iterations)

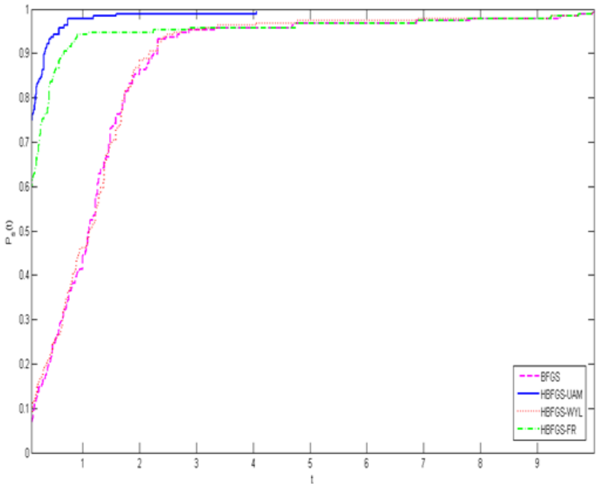

Fig. 2: Performance profile (CPU time)

V. CONCLUSION

A new hybrid coefficient for solving unconstrained optimization problems is clearly presented. The future interest is to test new hybrid coefficient with another quasi newton formula by using the inexact line search.

ACKNOWLEDGMENT

The authors are grateful to Government of Malaysia for funding this research via the Fundamental Research Grant Scheme (FRGS/1/2017/STG06/Unisza/01/1) and also Universiti Sultan Zainal Abidin, Terengganu, Malaysia.

REFERENCES

1. M. Al-Baali, “Numerical experience with a class of self-scaling quasi-newton algorithms,” Journal of Optimization Theory and Applications, 3(96), 1998, pp. 533-553.
2. N. Aini, M. Rivaie, and M. Mamat, “A modified conjugate gradient coefficient with inexact line search for unconstrained optimization,” AIP Conference Proceedings, 1787(1), 2016, pp. 1-6.
3. N. Andrei, “An unconstrained optimization test functions collection,” Advance Modelling and Optimization, 1(10), 2008, pp. 147-161.
4. E. Dolan, and J. J. More, “Benchmarking optimization software with performance profile,” Mathematic Programming, 91(2), 2002 pp. 201-213.
5. N. H. A. Ghani, M. Rivaie, and M. Mamat M, “A modified form of conjugate gradient method for unconstrained optimization problems,” AIP Conference Proceedings, 1739(1), 2016, pp. 1-8.
6. N. Hajar, M. Mamat, M. Rivaie, and I. Jusoh, “A new type of descent conjugate gradient method with exact line search,” AIP Conference Proceedings, 1739(1), 2016, pp. 1-8.
7. M. Hamoda, M. Rivaie, M. Mamat M, and Z. Salleh, “A conjugate gradient method with inexact line search for unconstrained optimization,” Applied Mathematical Sciences, 9(37), 2015, pp. 1823-1832.
8. M. A. H. Ibrahim, M. Mamat, and L. W. June, “The hybrid BFGS-CG method in solving unconstrained optimization problems,” 2014, 2014, pp. 1-6.
9. M. A. H Ibrahim, M. Mamat, and L. W. June, “BFGS method: A new search direction,” Sains Malaysiana, 10(43), 2014, pp. 1591-1597.
10. M. A. H Ibrahim, M. Mamat, L. W. June, and A. Z. M. Sofi, “The algorithms of Broyden-CG for unconstrained optimization problems,” International Journal of Mathematical Analysis, 8(52), 2014, pp. 2591-2600.
11. M. A. H. Ibrahim, M. Mamat, A. Z. M. Sofi, I. Mohd, and W. M. A. W. Ahmad, “Alternative algorithm of Broyden FAMILY for unconstrained optimization,” AIP Conference Proceedings, 1309(1), 2008, pp. 670-680.
12. W. Khadijah, M. Rivaie, M. Mamat, and I. Jusoh, “A spectral KRMI conjugate gradient method under the strong-wolfe line search,” AIP Conference Proceedings, 1739(1), 2016, pp. 1-8.
13. J. Long, X. Hu, and L. Zhang, “Improved Newton’s method with exact line searches to solve quadratic matrix equation,” Journal of Computational and Applied Mathematics, 222(2), 2008, pp. 645-654.
14. Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs. Berlin: Springer, 1996.
15. N. S. Mohamed, M. Mamat, F. S. Mohamad, and M. Rivaie, “A new coefficient of conjugate gradient methods for nonlinear unconstrained optimization,” Jurnal Teknologi, 78(6-4), 2016, pp. 131-136.
16. J. L. More, B. S. Garbow, and K. E. Hillstrom, “Testing unconstrained optimization software,” ACM Transaction on Mathematical Software, 7(1), 1981, pp. 17-41.
17. M. Rivaie, A. Abashar, M. Mamat, and I. Mohd I, “The convergence properties of a new type of conjugate gradient methods,” Applied Mathematical Sciences, 8(1), 2014, pp. 33-44.
18. N. Shapiee, M. R. M. Ali, M. Mamat, and Z. Salleh, “A new simple conjugate gradient coefficient for unconstrained optimization,” Applied Mathematical Sciences, 9(63), 2015, pp. 3119-3130.
19. S. Shoid, M. Rivaie, M. Mamat, and Z. Salleh, “A new conjugate gradient method with exact line search,” Applied Mathematical Sciences, 9(96), 2015, pp. 4799-4812.
20. A. Z. M. Sofi, M. Mamat, I. Mohd, and M. A. H Ibrahim, “Fletcher Reeves Like CG Formula approach on Broyden Family Update,” 3rd International Conference on Mathematical Sciences, 2014, pp. 527-532.
21. A. Z. M. Sofi, M. Mamat, I. Mohd, and Y. Dasril, “An alternative hybrid search direction for unconstrained optimization,” Journal of Interdisciplinary Mathematics, 11(5), 2008, pp. 731-739.
22. Z. Wei, S. Yao and L. Liu, “The convergence properties of some new conjugate gradient method,” Applied Mathematics and Computation, 183(2), 2006, pp. 1341-1350.
23. N. Zull, M. Rivaie, M. Mamat, Z. Salleh, and Z. Amani, “Global convergence of a new spectral conjugate gradient by using strong wolfe line search,” Applied Mathematical Sciences, 9(63), 2015, pp. 3105-3117.
24. 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, 26(3), 1989, pp. 727-739.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication