Journal of Physics: Conference Series

PAPER • OPEN ACCESS

New hybrid BFGS-CG method for solving unconstrained optimization

To cite this article: Wan Farah Hanan Wan Osman *et al* 2021 *J. Phys.: Conf. Ser.* **1874** 012080

View the article online for updates and enhancements.

240th ECS Meeting
Digital Meeting, Oct 10-14, 2021
Register early and save
up to 20% on registration costs
Early registration deadline Sep 13
REGISTER NOW

This content was downloaded from IP address 102.90.5.10 on 31/08/2021 at 10:35

New hybrid BFGS-CG method for solving unconstrained optimization

Wan Farah Hanan Wan Osman[1*], Mustafa Mamat[2], Mohd Asrul Hery Ibrahim[3]

[1]University College Tati, 24000 Kemaman, Terengganu. Email: wfarahanan@uctati.edu.my
[2]Universiti Sultan Zainal Abidin, Kuala Terengganu, Terengganu. Email: must@unisza.edu.my
[3]Universiti Malaysia Kelantan, Kota Bharu, Kelantan. Email: hery.i@umk.edu.my

Abstract. Conjugate gradient method and quasi-Newton (QN) method are both well known solvers for solving unconstrained optimization problems. In this paper, we proposed a new conjugate gradient method denoted as Wan, Asrul and Mustafa (WAM) method. This WAM method is then combined with the QN method to produce a new hybrid search direction which is QN-WAM. Based on numerical results, the proposed hybrid method proved to be more efficient compared to the original quasi-Newton method and other hybrid methods.

1. Introduction

One of the most efficient methods for solving unconstrained optimization problems is quasi-Newton (QN) method. Consider the following unconstrained optimization problems:

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

where $x\in\mathbb{R}^n$ is a decision variable and $f(x)$ is an objective function from $\mathbb{R}^n$ to $\mathbb{R}$. Iterative method is used for solving unconstrained problems and known as line search method. Line search is a method for determining the stepsize of an optimization algorithm as follows:

$$x_{k+1} = x_k + \alpha_k d_k \quad (2)$$

where α_k is a stepsize and d_k is the search direction. The search direction, d_k is calculated using:

$$d_k = -H_k g_k \quad (3)$$

where g_k is a gradient of $f(x)$ and H_k is a Hessian update formula for QN method. Then the step size, α_k in (2) is computed using strong Wolfe line search that is:

$$f(x_k+\alpha_k d_k) \le f(x_k) + c_1\alpha_k g_k^T d_k \quad (4)$$

$$\left| g(x_k+\alpha_k d_k)^T d_k \right| \le -c_2 g_k^T d_k \quad (5)$$

where $0<c_1<c_2<1$.

During the last few decades, the convergence properties of QN methods have received much study. Many researchers tried to improve the performance of the QN method by combining it with the search direction of other methods, such as the conjugate gradient (CG) method, and the steepest descent (SD) method.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

In this paper, Broyden-Fletcher-Goldfarb-Shanno (BFGS) inverse Hessian update formula is used. Many researchers had proved that the BFGS method is the most effective method and converge fastest amongst the current QN methods [1],[2],[3]. The BFGS update formula is derived as follows:

$$H_{k+1} = H_k - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} \tag{6}$$

with $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$ and H_{k+1} must satisfy the secant equation $H_{k+1} s_k = y_k$.

In order to have an algorithm with good convergence properties, a new search direction which is a hybrid of QN and CG are developed. At first, a new CG method is proposed and named as WAM method. Then, the BFGS update formula is used to approximate the inverse Hessian in new hybrid BFGS-WAM method combining BFGS and WAM.

2. New CG method

The search direction for CG method is defined as:

$$d_k = \begin{cases} -g_k, & \text{if } k = 0. \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geq 1. \end{cases} \tag{7}$$

where g_k is the gradient of $f(x)$ at the point x_k and β_k is known as the CG coefficient. Hestenenes-Stiefel (HS), Fletcher and Reeves (FR), Polak, Ribiere and Polyak (PRP), and Liu and Storey (LS) are some examples of known β_k.

Wei et al. (2006b) had proposed a variation of the PRP method and denoted as WYL (Wei-Yao-Liu). The formula is defined as:

$$\beta_k^{WYL} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|d_{k-1}\|} g_k^T g_{k-1}}{\|g_{k-1}\|^2} \tag{8}$$

They proved that their method performed better compared to PRP method.

Then, Zhang (2009) had modified WYL method and defined his method as:

$$\beta_k^{NPPRP} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} \left|g_k^T g_{k-1}\right|}{\|g_{k-1}\|^2} \tag{9}$$

Zhang (2009) proved that the modified method has as better convergence properties compared to WYL method and fulfil the sufficient descent condition with strong Wolfe line search.

Recently, Jiang et al. (2014) proposed their variation of the PRP method and WYL method denoted as JMJ method or β_k^{JMJ} and defined as:

$$\beta_k^{JMJ} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|d_{k-1}\|} \left|g_k^T d_{k-1}\right|}{\|g_{k-1}\|^2} \tag{10}$$

At each iteration, the method proved to be descent and converges globally for general nonconvex function by using Wolfe line search.

Motivated by (9) and (10), a new CG method is suggested. This new method is named WAM or referred as β_k^{WAM}, which is based on the researchers' names: - Wan, Asrul and Mustafa. Based on the idea, instead of using $\left|g_k^T g_{k-1}\right|$ or $\left|g_k^T d_{k-1}\right|$ in (9) and (10) respectively, the CG coefficient for WAM algorithm uses $\left|d_{k-1}^T g_{k-1}\right|$ and the denominator of β_k^{WAM} is similar to FR and PRP methods. Hence, the new method is as follows:

$$\beta_k^{WAM} = \frac{g_k^T g_k - \frac{\|g_k\|}{\|d_{k-1}\|}\left|d_{k-1}^T g_{k-1}\right|}{g_{k-1}^T g_{k-1}} \tag{11}$$

In the formula, g_k is the current gradient and g_{k-1} is the previous gradient of the function, and d_{k-1} is the search direction.

3. Hybrid BFGS-WAM method

The search direction for hybrid BFGS-WAM method is given as:

$$d_k = \begin{cases} -H_k g_k, & k = 0 \\ -H_k g_k + \eta \beta_k^{WAM} d_{k-1}, & k \geq 1 \end{cases} \tag{12}$$

where $\beta_k^{WAM} = \frac{g_k^T g_k - \frac{\|g_k\|}{\|d_{k-1}\|}\left|d_{k-1}^T g_{k-1}\right|}{g_{k-1}^T g_{k-1}}$ and $\eta \in (0,1]$.

In this study, the BFGS update formula is used to approximate the inverse Hessian in the new QN-WAM method and is referred as BFGS-WAM method. Algorithm 1 below is for hybrid BFGS-WAM method under strong Wolfe line search.

Algorithm 1: BFGS-WAM method

Step 1: Given a starting point x_0 and $H_0 = I$.
Step 2: Stop if $\|g(x_{k+1})\| < 10^{-6}$.
Step 3: Compute d_k based on (12).
Step 4: Calculate α_k based on (4) and (5).
Step 5: Compute $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$.
Step 6: Update inverse Hessian matrix H_{k+1}.
Step 7: Set $k = k+1$ and return to Step 2.

4. Convergence analysis

In this section, the convergence properties of the hybrid BFGS-WAM method are proved under strong Wolfe line search. The new method must satisfy the sufficient descent condition and possesses global convergence properties. Some assumptions of the objective function are needed in order to prove the convergence.

The search direction, d_k of an optimization algorithm is said to possess sufficient descent property if the following sufficient descent condition is satisfied.

$$g_k^T d_k \leq -C\|g_k\|^2 \tag{13}$$

for $k > 0$ and $C > 0$. The following assumption is needed to prove the global convergence of the new methods.

Assumption 1.1

Consider the following.

i. The objective function f is twice continuously differentiable.
ii. The level set L is convex and there exist positive constants c_2 and c_3, satisfying $c_2\|z\|^2 \leq z^T F(x) z \leq c_3\|z\|^2$ for all $z \in \mathbb{R}^n$ and $x \in L$, where $F(x)$ is the Hessian matrix for f.
iii. The Hessian matrix is Lipschitz continuous at the point x^*, that is, there exists a positive constant c_4 satisfying $\|g(x) - g(x^*)\| \leq c_4\|x - x^*\|$ for all x in a neighbourhood of x^*.

Theorem 1.1
Consider a hybrid QN-WAM algorithm under strong Wolfe line search with search direction (12) and suppose that Assumption 1.1 hold. Then, the sufficient condition (13) of the algorithm holds for all $k \geq 0$ and $C > 0$.

In order to prove the convergence of the new hybrid method, some basic assumptions of the objective function are needed as defined in Assumption 1.1 and the following lemmas are desired.

Lemma 1.1 (Byrd et al., 1987 and Byrd and Nocedal, 1989)
Let $\{B_k\}$ a symmetric and positive definite BFGS formula, and $y_k^T s_k > 0$ for all k. Moreover, assume that $\{s_k\}$ and $\{y_k\}$ are such that

$$\frac{\|(y_k - G)s_k\|}{\|s_k\|} \leq \varepsilon_k$$

for some symmetric and positive definite G and for some sequence $\{\varepsilon_k\}$ with the property $\sum_{k=1}^{\infty} \varepsilon_k < \infty$. Then,

$$\lim_{k\to\infty} \frac{\|(B_k - G)d_k\|}{\|d_k\|} = 0$$

and the sequence $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ are bounded.

Lemma 1.2 (Nocedal and Wright, 2006)
Suppose that Assumption 1.1 hold true, then the stepsize can be determined by the strong Wolfe line search satisfies

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \frac{-c_4(g_k^T d_k)^2}{\|d_k\|^2}$$

where $c_4 > 0$.

Lemma 1.3
Suppose that Assumption 1.1 hold true. Consider any CG method and search direction of the hybrid method, d_k and stepsize α_k satisfies the strong Wolfe line search. Then, the following condition holds

$$\sum_{k=0}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} < \infty$$

Theorem 1.2
Suppose that Lemma 1.1, Theorem 1.1 and Assumption 1.1 hold. Consider the hybrid QN-WAM method in the form (12) where α_k is obtained by the strong Wolfe line search and the sufficient descent condition also holds. Then,

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty \tag{14}$$

Proof:
To prove this theorem, contradiction proving is used. Firstly, suppose that there exists a positive constant $a > 0$ such that

$$\|g_k\| \geq a \tag{15}$$

Applying Cauchy Schwartz inequality on equation (12), yields

$$d_k = -H_k g_k + \eta \beta_k d_{k-1}$$

$$\|d_k\| \leq \|H_k\| \|g_k\| + \eta |\beta_k| \|d_{k-1}\|$$

Divide both side by $\|g_k\|^2$, to get

$$\frac{\|d_k\|}{\|g_k\|^2} \le \frac{\|H_k\|\|g_k\|}{\|g_k\|^2} + \frac{|\eta|\|\beta_k\|\|d_{k-1}\|}{\|g_k\|^2}$$

Substitute β_k from equation (11),

$$\frac{\|d_k\|}{\|g_k\|^2} \le \frac{\|H_k\|}{\|g_k\|} + \frac{|\eta|\|g_k\|^2\|d_{k-1}\|}{\|g_{k-1}\|^2\|g_k\|^2}$$

$$= \frac{\|H_k\|}{\|g_k\|} + |\eta|\frac{\|d_{k-1}\|}{\|g_{k-1}\|^2}$$

Since $\|H_k\|$ is bounded, then there exists a constant $u > 0$ such that $\|H_k\| \le u$. From here, the equation becomes

$$\frac{\|d_k\|}{\|g_k\|^2} \le \frac{u}{\|g_k\|} + |\eta|\frac{\|d_{k-1}\|}{\|g_{k-1}\|^2}$$

Note that $d_0 = -H_0 g_0$, then for all k,

$$k = 1,\ \frac{\|d_1\|}{\|g_1\|^2} \le \frac{u}{\|g_1\|} + \eta\frac{\|d_0\|}{\|g_0\|^2} = \frac{u}{\|g_1\|} + \eta\frac{u}{\|g_0\|}$$

$$k = 2,\ \frac{\|d_2\|}{\|g_2\|^2} \le \frac{u}{\|g_2\|} + \eta\frac{\|d_1\|}{\|g_1\|^2} = \frac{u}{\|g_2\|} + \eta\left(\frac{u}{\|g_1\|} + \eta\frac{u}{\|g_0\|}\right)$$

$$= \frac{u}{\|g_2\|} + \frac{\eta u}{\|g_1\|} + \frac{\eta^2 u}{\|g_0\|}$$

$$k = 3,\ \frac{\|d_3\|}{\|g_3\|^2} \le \frac{u}{\|g_3\|} + \eta\frac{\|d_2\|}{\|g_2\|^2} = \frac{u}{\|g_3\|} + \eta\left(\frac{u}{\|g_2\|} + \frac{\eta u}{\|g_1\|} + \frac{\eta u}{\|g_0\|}\right)$$

$$= \frac{u}{\|g_3\|} + \frac{\eta u}{\|g_2\|} + \frac{\eta^2 u}{\|g_1\|} + \frac{\eta^3 u}{\|g_0\|}$$

$$\vdots$$

Hence for all k, it follows that

$$\frac{\|d_k\|}{\|g_k\|^2} \le \sum_{i=0}^{k}\frac{u\eta^{k-i}}{\|g_i\|}$$

Substitute (15) into the equation which then gives

$$\frac{\|d_k\|}{\|g_k\|^2} \le \frac{u}{a}\sum_{i=0}^{k}\eta^{k-i}$$

$$\frac{\|d_k\|}{\|g_k\|^2} \le \frac{u}{a}\sum_{i=0}^{k}\eta^{i}$$

Note that, $\sum_{i=0}^{k}\eta^i = \frac{1-\eta^{k+1}}{1-\eta}$. Hence,

$$\frac{\|d_k\|}{\|g_k\|^2} \le \frac{u}{a}\left(\frac{1-\eta^{k+1}}{1-\eta}\right)$$

$$\frac{\|d_k\|}{\|g_k\|^2} \leq \frac{u - u\eta^{k+1}}{a - a\eta}$$

$$\frac{\|d_k\|}{\|g_k\|^2} \leq \frac{u}{a - a\eta}$$

$$\frac{\|g_k\|^4}{\|d_k\|^2} \geq \left(\frac{a - a\eta}{u}\right)^2$$

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} \geq \sum_{k=0}^{\infty} \left(\frac{a - a\eta}{u}\right)^2 = \infty$$

From Theorem 1.2, it is implied that $\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} \geq \sum_{k=0}^{\infty} \left(\frac{a - a\eta}{u}\right)^2 = \infty$. This contradicts the Lemma 1.3.

Hence the proof is completed.

5. Numerical results and discussion

In this section, the numerical results for hybrid BFGS-WAM methods is presented and analyzed. The hybrid BFGS-WAM method is compared with the standard BFGS method and WAM method. All the methods are tested by using twenty-six different test functions obtained from Jamil and Yang (2013), Andrei (2008) and Issam (2005). For each test function, four different initial points are taken starting from a point near the solution to a point far from it, as suggested by Hiltsrom (1977). The value of parameter η in (5.1) is set to be 10^{-5} based on Ibrahim et al. (2014).

All problems mentioned above are solved by using Matlab R2014b, via a computer with Intel Core i5 processor and 4GB RAM. The stopping criterion is set to $\|g_k\| \leq \varepsilon$ where $\varepsilon = 10^{-6}$. For inexact line search, the stepsize α_k satisfied the strong Wolfe conditions as defined by (4) and (5), with the values used for parameter c_1 and c_2 are 0.0001 and 0.1 respectively. The tests are conducted with different variables, starting from smaller scales to the larger ones $2 \leq n \leq 1000$.

The numerical results are then analyzed by using performance profile introduced by Dolan and More (2002).

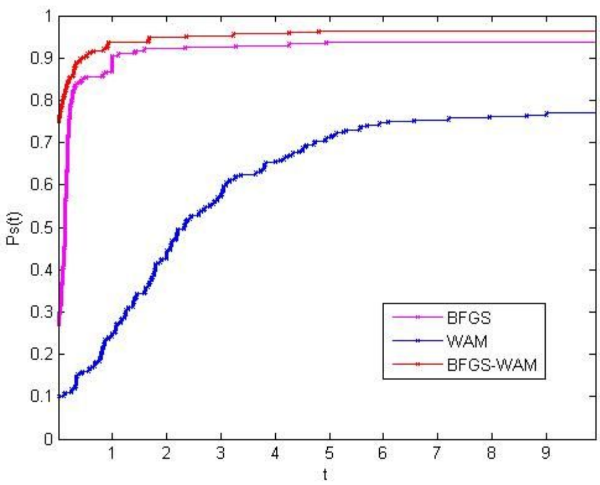

Figure 1 Performance profile of BFGS, WAM and BFGS-WAM methods based on number of iteration

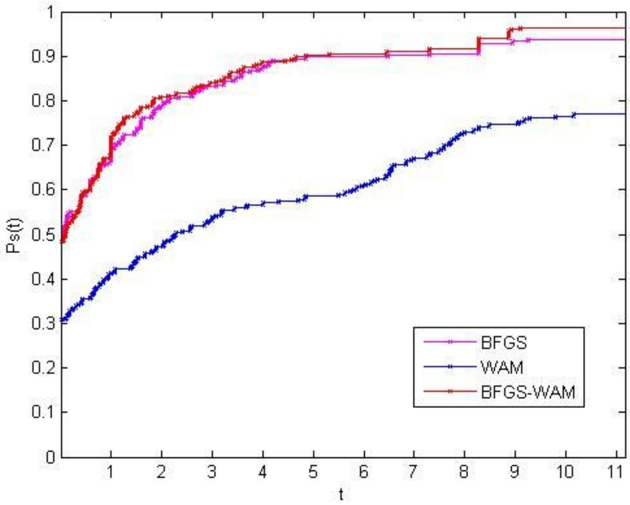

Figure 2 Performance profile of BFGS, WAM and BFGS-WAM methods based on CPU time

The left side of the Figure 1 and Figure 2 indicate that the lowest curve belongs to WAM which implies that WAM possesses the lowest efficiency. The top curve is belongs to BFGS-WAM, showing that it has the highest efficiency, followed by BFGS and WAM methods. The right side of Figure 1 and 2 show that BFGS-WAM method solves 97.27%, while BFGS method solves 93.75% of the test problems.

6. Conclusion

For conclusion, a new hybrid BFGS-WAM method had been proposed. The new method is theoretically proved the sufficient descent property and global convergence with strong Wolfe line search. The numerical results proved that the BFGS-WAM method is more efficient and robust in solving standard optimization test functions compared to the original BFGS method and WAM method.

References

[1] Al-Baali, M., and Khalfan, H. 2007. An Overview of Some Practical Quasi- Newton Methods for Unconstrained Optimization. *Sultan Qaboos University Journal For Science*, *12*(2), 199–209.

[2] Babaie-Kafaki, S. 2011. A modified BFGS algorithm based on a hybrid secant equation. *Science China Mathematics*, *54*(9), 2019.

[3] Fletcher, R. and Reeves, C. 1964. Function minimization by conjugate gradient. *The computer Journal*, 7: 149-154.

[4] Hestenes, M. R. and Steifel, E. 1952. Method of conjugate gradient for solving linear equations. *Journal of Research of the National Bureau Standards,* 49: 409–436.

[5] Polyak, B. T. 1969. The conjugate gradient method in extremal, *USSR Computational Mathematics and Mathematical Physics,* 9(4), 94–112.

[6] Polak, E. and Ribiere, G. (1969). Note on the convergence of methods of conjugate directions. *Rev. Francaise Informat Recherche Operationelle,* 3: 35–43.

[7] Andrei, N. 2008. A hybrid conjugate gradient algorithm with modified secant condition for unconstrained optimization as a convex combination of Hestenes-Stiefel and Dai-Yuan algorithms. *Studies in Informatics and Control, ISSN 1220-1766*, 17(4): 373–392.

[8] Buckley, A. 1978. A combined conjugate-gradient quasi-Newton minimization algorithm. *Mathematical Programming*, 15(1), 200–210.

[9] Byrd, R. H., Nocedal, J., and Yuan, Y.-X. 1987. Global convergence of a class of quasi-Newton methods on convex problems. *SIAM Journal on Numerical Analysis*. Vol 24 (5), 1171–1190.

[10] Chong, E. K. P. and Zak, S. H. 2013. An introduction to optimization (4th ed). *John Wiley and Sons, New Yor*k.

[11] Dai, Y. 2012. A Perfect Example for The BFGS Method*, *Mathematical Programming*, 138, 50–530.

[12] Dai, Y. H. and Yuan, Y. 1999. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization*. 10(1), 177–182.

[13] Dai, Z. and Wen, F. 2011. A modified CG-DESCENT method for unconstrained optimization. *Journal of Computational and Applied Mathematics.,* 235: 3332–3341.

[14] Fletcher, R., and Xu, C. 1987. Hybrid methods for nonlinear least squares. *IMA Journal of Numerical Analysis*, *7*(3), 371–389.

[15] Han, L. and Neumann, M. 2003. Combining Quasi-Newton and Steepest Descent Directions. *International Journal of Applied Mathematics*, 12: 167–171.

[16] Hassan, M. A., Mansi, M., and June, L. W. 1999. A Switching Criterion in Hybrid Quasi-Newton BFGS - Steepest Descent Direction. *Pertanika Journal of Science and Technology*, 7(2), 111–123.

[17] Himmeblau, D. M. 1972. Applied nonlinear programming. *McGraw-Hills, New York*.

[18] Ibrahim, M. A. H., Mamat, M., Sofi, A. Z. M., Mohd, I., and Ahmad, W. M. A. W. 2010 Alternative algorithms of broyden family AMI: for unconstrained optimization. *AIP Conference Proceedings* , 1309(1), 670–680.

[19] Ibrahim, M. A. H., Mamat, M., and Leong, W. J. 2014. BFGS Method: A new search direction. *Sains Malaysiana*, 43(10), 1591–1597.

[20] Issam, A. R. M. 2005. Minimization of extended quadratic functions with inexact line searchs. *Journal of Korean Society of Industrial and Applied Mathematics*, 9(1), 55–61.

[21] Jamil, M. and Yang, X. 2013. A Literature Survey of Benchmark Functions For Global Optimization Problems. *International Journal of Mathematical Modelling and Numerical Optimisation*, 4(2), 150–194.

[22] Nocedal, J. 1992. Theory of algorithms for unconstrained optimization. *Acta Numerica 1*, 199–242.

[23] Osman, W. F. H. W., Ibrahim, M. A. H., & Mamat, M. 2017. Hybrid DFP-CG method for solving unconstrained optimization problems. In *Journal of Physics: Conference Series* (Vol. 890, No. 1, p. 012033). IOP Publishing.

[24] Osman, W. F. H. W., Ibrahim, M. A. H., & Mamat, M. 2018. A new search direction of DFP-CG method for solving unconstrained optimization problems. In *AIP Conference Proceedings* (Vol. 1974, No. 1, p. 020024). AIP Publishing.