Malaysian Journal of Mathematical Sciences 10(S) March : 167-178 (2016)
Special Issue: The 10th IMT-GT International Conference on Mathematics, Statistics and its Applications 2014 (ICMSA 2014)

MALAYSIAN JOURNAL OF MATHEMATICAL SCIENCES

Journal homepage: http://einspem.upm.edu.my/journal

An Efficient Modification of Nonlinear Conjugate Gradient Method

Zabidin Salleh [1*], Ahmad Alhawarat [1], Mustafa Mamat [2] and Mohd Rivaie [3]

[1]School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Terengganu, Malaysia

[2]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia

[3]Department of Computer Science and Mathematics, Universiti Technologi MARA (UiTM) Terengganu, Kuala Terengganu Campus, Malaysia

E-mail: zabidin@umt.edu.my

*Corresponding author

ABSTRACT

To solve large and small scale unconstrained optimization problems, conjugate gradient method (CG) is an interesting and active method to find the optimum point for linear and nonlinear optimization functions. Wei et al. (2006) presented an efficient modification of Polak-Ribiere-Polyak (1969) formula, since it passes the global convergence properties under several lines searches with sufficient descent condition. In this paper, we depict a new positive CG method derived from above two coefficients, the new method achieves the global convergence properties with the strong Wolfe-Powel, weak Wolfe Powell, and Modified Armijo line searches. The numerical computations with the strong Wolfe-Powell line search demonstrated the efficiency of the new formula is almost superior to other modern methods.

Keywords: Inexact line search, global convergence, optimization problems, search direction

1. Introduction

The CG method is interesting tool to find the optimum solution/s for nonlinear unconstrained optimization functions which they are bounded below and their gradient is Lipschitz continuous. We consider the problem

$$\min\{f(x): x \in \mathbb{R}^n\}, \tag{1}$$

where f is continuous differentiable function and its gradient is written by $g(x) = \nabla f(x)$. The CG method is iterative method starting from $x_0 \in \mathbb{R}^n$ which is given as follows,

$$x_{k+1} = x_k + \alpha_k d_k \ , k = 0, 1, 2, 3, \ldots, \tag{2}$$

and d_k denote the search direction defined by,

$$d_k = \begin{cases} -g_k & if \quad k = 0 \\ -g_k + \beta_k d_{k-1} & if \quad k \geq 1 \end{cases} \tag{3}$$

where $g_k = g(x_k)$, β_k is a scalar (CG coefficient or CG method) and $\alpha_k > 0$ is the steplength obtained to implementation of CG method. In order to find α_k there are several line searches, the following line searches are used to find the step size in order to get the global convergence properties:

- Strong Wolfe-Powell

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \tag{4}$$
$$|g(x_k + \alpha_k d_k)^T d_k| \leq \sigma \left| g_k^T d_k \right| \tag{5}$$

where $0 < \delta < \sigma < 1$.

- Weak Wolfe-Powell ,

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \tag{6}$$
$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k. \tag{7}$$

where $0 < \delta < \sigma < 1$.

- Modified Wolfe-Powell (MWL) which proposed by (Yu et al., 2009).

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \min\{\delta t_k g_k^T d_k, -\gamma t_k^2 \|d_k\|^2\} \tag{8}$$
$$g(x_k + t_k d_k)^T d_k \geq \sigma g_k^T d_k \tag{9}$$

where $\delta \in (0,1), \sigma \in (\delta, 1)$ and $\gamma > 0$.

In particular strong Wolfe-Powell line search is used to find the numerical results in Section 4, since it does not need expensive computational work to find the best steplength.

The most famous formulas for CG method formulas are (Hestenses and Stiefel, 1952) (HS), (Fletcher and Reeves, 1964) (FR), and (Polak and Ribiere, 1969) (PRP), which they are given as follows respectively,

$$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{(g_{k-1}^T g_{k-1})d_{k-1}}, \quad \beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T g_{k-1}}, \quad \beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}.$$

The convergence properties for FR, and PRP formulas with exact line search was studied by Zoutendijk, 1970, and Polak and Ribiere, 1969. Gilbert and Nocedal, 1992, proved that $\max\{0, \text{PRP}\}$ method is globally convergent by using more than one line search. There are many researches interested with the CG methods. We suggest the following references (Alhawarat et al., 2014; Rivaie et al., 2012; Mamat et al., 2010).

Wei et al. (2006) presented one of the best CG formulas which is similar to PRP method. In this paper, we refer it to WYL formula, one of the advantages for WYL coefficient is non-negative method. Many modifications have appeared, as follows, see (Wei et al., 2006; Shengwei et al., 2007; Zhang, 2009).

$$\beta_k^{NPRP} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|}\left\|g_k^T g_{k-1}\right\|}{\|g_{k-1}\|^2}, \quad \beta_k^{WYL} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{\|g_{k-1}\|^2}$$

$$\beta_k^{VHS} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{d_{k-1}^T(g_k - g_{k-1})}.$$

2. The Modified Formula

The new CG formula in this paper is given as follows

$$\beta_k^{AMZR} = \frac{g_k^T(\tau_k . g_k - g_{k-1})}{\tau_k . (g_{k-1}^T g_{k-1})} \tag{10}$$

where $\tau_k = \frac{\|g_{k-1}\|}{\|g_k\|}$ and $\|.\|$ is the Euclidean norm. β_k^{AMZR} is similar to β_k^{WYL} and reveals the problem where PRP formula failed. The following simplifications are important and useful for proving the next results. The β_k^{AMZR} could be written as follows,

$$\beta_k^{AMZR} = \frac{g_k^T(\tau_k . g_k - g_{k-1})}{\tau_k . (g_{k-1}^T g_{k-1})} = \frac{\|g_k\|^2 \|g_{k-1}\| - \|g_k\| g_k^T g_{k-1}}{\|g_{k-1}\|^3}.$$

Using Cauchy-Schwartz inequality,

$$\beta_k^{AMZR} = \frac{g_k^T(\tau_k . g_k - g_{k-1})}{\tau_k . (g_{k-1}^T g_{k-1})} \geq \frac{\|g_k\|^2 \|g_{k-1}\| - \|g_k\|^2 \|g_{k-1}\|}{\|g_{k-1}\|^3} = 0 \quad (11)$$

and

$$\beta_k^{AMZR} = \frac{g_k^T(\tau_k . g_k - g_{k-1})}{\tau_k . (g_{k-1}^T g_{k-1})} \leq \frac{\|g_k\|^2 \|g_{k-1}\| + \|g_k\|^2 \|g_{k-1}\|}{\|g_{k-1}\|^3} = \frac{2\|g_k\|^2}{\|g_{k-1}\|^2}. \quad (12)$$

3. Global Convergence Properties

The following assumption is required for the following theorems.

Assumption 1.

A. $f(x)$ is bounded on the level set $A = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$ where x_1 is an initial point.

B. In some neighborhood W of A, f is continuous and differentiable function, and its gradient is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that for any $x, y \in W$ we have,

$$\|g(x) - g(y)\| \leq L\|x - y\|.$$

One of the important rules in CG method is the descent property, namely

$$g_k^T d_k \leq -\|g_k\|^2, \text{ where } k \geq 0.$$

To establish the global convergence of the CG methods we need the following lemma.

Lemma 1. (Zoutendijk, 1970). Suppose Assumptions 1 with the descent property are hold. Let any form of (3), and α_k computed by Wolfe-Powell line search. Then

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \quad (13)$$

holds, which is called as Zoutendijk condition.

The sufficient descent condition is given as follows,

$$g_k^T d_k \leq -c\|g_k\|^2, \text{ where } k \geq 0 \text{ and } c \in (0,1). \tag{14}$$

Theorem 1. Consider the sequences g_k and d_k are constructed by Algorithm 1 and α_k is computed by (4) and (5) if $\sigma \in \left(0, {}^1/_4\right)$. Then (14) holds.

Proof. From (3), for $k = 0$, we have $g_0^T d_0 = -\|g_0\|^2 \leq -c\|g_0\|^2$. Suppose that (14) is true until $k-1$ for $k \geq 1$. Multiply (3) by g_k^T, it becomes

$$\begin{aligned} g_k^T d_k &= g_k^T(-g_k + \beta_k d_{k-1}) = -\|g_k\|^2 + \beta_k g_k^T d_{k-1} \\ &= -\|g_k\|^2 + g_k^T d_{k-1}\left(\frac{\|g_{k-1}\|\|g_k\|^2 - \|g_k\| g_k^T g_{k-1}}{\|g_{k-1}\|^3}\right). \end{aligned}$$

Divided both side by $\|g_k\|^2$ yield,

$$\frac{g_k^T d_k}{\|g_k\|^2} = -1 + \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2}\left(\frac{\|g_{k-1}\|\|g_k\|^2 - \|g_k\| g_k^T g_{k-1}}{\|g_k\|^2\|g_{k-1}\|}\right). \tag{15}$$

Using Cauchy-Schwarz inequality, we have

$$0 \leq \frac{\|g_{k-1}\|\|g_k\|^2 - \|g_k\| g_k^T g_{k-1}}{\|g_k\|^2\|g_{k-1}\|} \leq 2. \tag{16}$$

Using (5), (15) and (16), we have

$$-1 + 2\sigma\frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -1 - 2\sigma\frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2}.$$

By induction hypothesis, $g_{k-1}^T d_{k-1} \leq -c\|g_{k-1}\|^2$ for $k \geq 1$, this implies that

$$-\sum_{j=0}^{k}(2\sigma)^j \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -2 + \sum_{j=0}^{k}(2\sigma)^j.$$

Since $\sum_{j=0}^{k}(2\sigma)^j \leq \frac{1}{1-2\sigma}$, then

$$-\frac{1}{1-2\sigma} \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -2 + \frac{1}{1-2\sigma}.$$

Since $0 < 2\sigma < {}^1\!/_2$ when $0 < \sigma < {}^1\!/_4$, we have $1 < \frac{1}{1-2\sigma} < 2$. Let $c = 2 - \frac{1}{1-2\sigma} \in (0,1)$, then $c - 2 \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -c$. So

$$g_k^T d_k \leq -c\|g_k\|^2 \text{ where } c \in (0,1). \blacksquare$$

3.1 Global Convergence with the Wolfe-Powell Line Search

Gilbert and Nocedal, 1992, present an important theorem to find the global convergent for a positive part of PRP method, it summarized by three conditions, one of them called property star, which plays strong roles in the studies of CG methods.

Property $*$ (Gilbert and Nocedal, 1992). Let a CG method of type (1) and (2), and suppose $0 < \gamma \leq \|g_k\| \leq \bar{\gamma}$ then the CG method possesses property $*$ if there exists $b > 1$ and $\lambda > 0$ such that for all $k \geq 1$, we get $|\beta_k| \leq b$, and if $\|x_k - x_{k-1}\| \leq \lambda$, then $|\beta_k| \leq \frac{1}{2b}$.

Theorem 2. (Gilbert and Nocedal, 1992). Consider CG method of form (2) and (3) with β_k satisfies Property $*$, (14) holds, the line search match Zoutendijk condition, and Assumption 1 holds. Then the iteration are globally convergent.

Lemma 2. Consider β_k^{AMZR} in forms (2) and (3). If Assumption 1 holds, then it is satisfy Property $*$.

Proof. Let $b = 2\frac{\bar{\gamma}^2}{\gamma^2} > 1$, and $\lambda \leq \frac{\gamma^3}{4L\bar{\gamma}^2 b}$. Then

$$|\beta_k^{AMZR}| = \left|\frac{g_k^T(\tau_k . g_k - g_{k-1})}{\tau_k . (g_{k-1}^T g_{k-1})}\right| \leq 2\frac{\|g_k\|^2}{\|g_{k-1}\|^2} \leq 2\frac{\bar{\gamma}^2}{\gamma^2} = b.$$

Using $\|x_k - x_{k-1}\| \leq \lambda$, we have

$$|\beta_k^{AMZR}| = \left|\frac{g_k^T(\tau_k g_k - g_{k-1})}{\tau_k \|g_{k-1}\|^2}\right| \leq \frac{\|g_k\|(\|\tau_k g_k - g_k\| + \|g_k - g_{k-1}\|)}{\tau_k \|g_{k-1}\|^2}$$
$$\leq \frac{2L\lambda\bar{\gamma}^2}{\gamma^3} \leq \frac{1}{2b}.$$

The proof is completed. ■

Since the CG formula in (10) satisfies the descent property and Property $*$, and the strong Wolfe-Powell satisfies Zoutendijk condition, we can present the following theorem without proof based on Theorem 2.

Theorem 3. Suppose that Assumption 1 holds. Consider any form of (2), (3) with CG formula as in (10) where α_k computed by (4) and (5), then $\lim_{k\to\infty} \inf\|g_k\| = 0$.

To match the global convergence properties for the new formula with the weak-Wolfe-Powell line search, we need the following lemmas. For the proof the reader could refer to (Dai and Yuan, 1998).

Lemma 3. Assume Assumptions 1 holds true and the sequences g_k and d_k are constructed by Algorithm 1, and α_k is computed by (6) and (7). Suppose (14) holds with Property $*$. Then $d_k \neq 0$, and $\sum_{k=0}^{\infty}\|u_{k+1} - u_k\|^2 < \infty$, where $u_k = \frac{d_k}{\|d_k\|}$.

Lemma 4. Consider Assumption 1 holds and the sequences g_k and d_k are constructed by Algorithm 1 and (14) holds. If $\beta_k \geq 0$ and Property $*$ satisfied, then there exist $\lambda > 0$ such that for any $\Delta \in \mathbb{N}$ and any index k_0, there is an index $k > k_0$ satisfying $\left|\mathcal{K}_{k,\Delta}^{\lambda}\right| > \frac{\lambda}{2}$, where $\mathcal{K}_{k,\Delta}^{\lambda} = \{i \in \mathbb{N}: k \leq i \leq k + \Delta - 1, \|x_i - x_{i-1}\| > \lambda\}$, $\mathbb{N}$ define the set of all positive integers, $\left|\mathcal{K}_{k,\Delta}^{\lambda}\right|$ defines the numbers of elements in $\mathcal{K}_{k,\Delta}^{\lambda}$.

By using Property $*$ and Lemmas 2 and 3, the globally convergence for (10) with (6) and (7) can be established similar to Theorem 3.3.3 in (Dai and Yuan, 1998).

Theorem 4. Assume the sequences g_k and d_k are constructed by Algorithm 1 with WWP line search and Lemmas 3 and 4 are true, then $\lim_{k\to\infty} \inf\|g_k\| = 0$.

3.2. Global Convergence with MWL Line Search

Yu et al., 2009, designed modified Wolfe line search to establish the global convergence for PRP method. In this section we show that our method is globally convergent under MWL line search. The following lemmas are results from (8) and (9) which the proof can be found in Yu et al., 2009.

Lemma 5. Consider Assumption 1 is true. Suppose any iteration method of the form (2) and (3), and α_k is obtained by (8) and (9). If for all k, $g_k^T d_k \leq 0$, then $\lim_{k\to\infty} \alpha_k \|d_k\| = 0$.

Lemma 6. Consider that Assumption 1 holds. Suppose any iteration method of the form (2), (3), and α_k is obtained by (8) and (9). If for all k, $g_k^T d_k \leq 0$, then there exist a constant $M_1 > 0$ such that $\alpha_k > M_1 \frac{|g_k^T d_k|}{\|d_k\|^2}$, and by using Assumption 1 there exist a positive constant γ such that $\|g_k\| < \gamma$ for all $x \in A$.

Lemma 7. Assume Assumption 1 holds, and x_k is constructed by Algorithm 1. If for all k, $\|g_k\| \geq \varepsilon$, there exists $M_2 > 0$ such that for all k we have $\|d_k\| \leq M_2$.

Theorem 5. Assume the sequences g_k and d_k are constructed by Algorithm 1 and α_k exists by (8) and (9) if $g_k^T d_k \leq 0$. Then either $g_k = 0$ for some k or $\lim_{k\to\infty} \inf\|g_k\| = 0$.

Proof. Assume that $g_k \neq 0$ for all k. By the above lemmas we have

$$\lim_{k\to\infty} \alpha_k \|d_k\|^2 = 0, \tag{17}$$
$$\lim_{k\to\infty} |g_k^T d_k| = 0. \tag{18}$$

Use the forms (3) and (10) with Assumption 1, then

$$g_k^T d_k = -\|g_k\|^2 + \beta_k g_k^T d_{k-1}.$$

This implies

$$\begin{aligned}\|g_k\|^2 \leq |g_k^T d_k| + |\beta_k g_k^T d_{k-1}| &\leq |g_k^T d_k| + \frac{|g_k^T(\tau_k g_k - g_{k-1})|}{\tau_k \|g_{k-1}\|^2} |g_k^T d_{k-1}| \\ &\leq |g_k^T d_k| + \frac{\|g_k\| |\tau_k g_k - g_{k-1}|}{\tau_k \|g_{k-1}\|^2} \|g_k\| \|d_{k-1}\| \\ &\leq |g_k^T d_k| + 2L \frac{\|g_k\|^3 \alpha_{k-1} \|d_{k-1}\|^2}{\|g_{k-1}\|^3}.\end{aligned}$$

Taking the limit for both side and using (17) and (18), then $\lim_{k\to\infty} \inf\|g_k\| = 0$. ■

4. Numerical Results and Discussion

Here, some of standard optimization test functions are taken from (Bongartz et al., 1995), (Andrei, 2008), and (Adorio and Diliman, 2005) to evaluate the efficiency of the new coefficient, we compare the new coefficient with the other conventional and recently CG methods; we choose WYL, VHS, NPRP, and FR formulas.

TABLE 1: Example functions which used with the strong wolf condition

Function	Number of variables	Initial points
Ex. White & Holst fun.	500,1000,5000,10000	(-1.2,1,-1.2,1…),(5,5,…,5), (10,10,…,10),(15,15,…15)
Ext. Rosenbrock fun.	500,1000,5000,10000	(-1.2,1,-1.2,1…),(5,5,…,5), (10,10,…,10),(15,15,…15)
Six hump fun.	500,1000,5000,10000	(0.5,…,.5) (1,1,…,1) ,(2,2,…,2),
Ex. Beale fun.	500,1000,5000,10000	(1,1,…,1), (5,5,…,5), (10,10…,10),(15,15,…,15)
Three hump fun.	500,1000,5000,10000	(.5,.5,…,.5), (1,1,…,1), (2,2…,2), (10,10,…,10)
Ext Himmelblau fun.	500,1000,5000,10000	(1,1,…,1), (5,5,…,5), (10,10…,10),(15,15,…,15)
Diagonal 2 fun.	500,1000,5000,10000	(.2,.2,…,.2), (.25,.25,…,.25), (.5,.5,…,.5),(1,1,…,1)
NONSCOMP fun.	500,1000,5000,10000	(1,1,…,1), (-1,-1,…,-1), (-2,-2…,-2), (-5,-5,…,-5)
Ext. DENSCHNB fun.	500,1000,5000,10000	(1,1,…,1), (5,5,…,5), (10,10…,10),(15,15,…,15)
Shallow fun.	500,1000,5000,10000	(-2,-2,…,-2), (2,2,…,2), (5,5…,5), (10,10,…,10)
Booth fun.	100, 200,300, 400, 500	(.5,.5,…,.5), (2,2,…,2), (5,5…,5), (10,10,…,10)
Ex. quadratic penalty fun.	500,1000,5000,10000	(2,2,…,2), (5,5,…,5), (10,10…,10), (15,15,…,15)
DIXMAANA fun.	500,1000,5000,10000	(2,2,…,.2), (5,5,…,5), (10,10…,10), (15,15,…,15)
DIXMAANB fun.	500,1000,5000,10000	(-2,-2,…,-2), (-1,-1,…,-1), (0,0…,0), (1,1,…,1)
NONDIA fun.	10,20,30,40,50	(-2,-2,…,-2), (-1,-1,…,-1), (0,0…,0), (1,1,…,1)
Ex. Tridiagonal 1 fun.	500,1000,5000,10000	(1,1,…,1), (-1,-1,…,-1), (2,2…,2), (5,5,…,5)
DQDRTIC fun.	500,1000,5000,10000	(-1,-1,…,-1), (1,1,…,1), (2,2…,2), (3,3,…,3)
Diagonal 4 fun.	500,1000,5000,10000	(1,1,…,1), (5,5,…,5), (10,10…,10),(15,15,…,15)
Raydan 2 fun.	500, 1000, 5000,10000	(1,1,…,1), (5,5,…,5), (10,10…,10), (15,15,…,15)

The Table reports some of standard test function which is used to test the efficiency of the new method and comparing with the other CG methods, the first column present the number of function from above list, the second column present the dimension/s which is used in the programming code, and the last column present the initials values. Notice that for every dimension we test four initial points i.e. a large number of data which lead us to more accuracy. Notice that, Ex. mean Extended, and fun. mean function.

The tolerance ε is chosen to be 10^{-6} for all formulas to study how the efficiency of these formulas towards the optimal solution, we choose the gradient value as the stopping criteria. We used Matlab 7.9 program, with hosted computer Intel® Core™ i3 CPU and 2GB of DDR2 RAM, and 4GB DDR2 RAM. Figures 1 and 2 are constructed by using a performance profile presented by Dolan and Moré (2002).

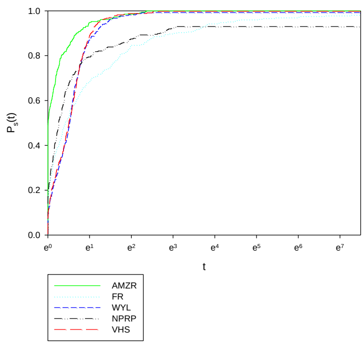

Figure 1: Performance profile with the CPU time.

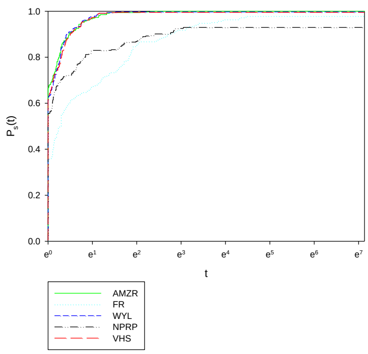

Figure 2: Performance profile based with the No. of iterations

From Figure 1 and Figure 2, it is clearly demonstrated that the new coefficient method is better than other methods. The best method has a curve that is top and right of the graph. In addition the new coefficient solve all above test functions where the other methods cannot.

5. Conclusion

We present a new CG formula that similar to PRP method and the global convergence properties are presented with several line searches. Our numerical results had shown that the new coefficient has efficiency when it compared to the other modern CG formulas. In the future work, we will focus at speed, accuracy, and memory space, using hydride or new nonlinear CG methods. In addition we will try to compare several line searches under new or traditional CG methods.

Acknowledgments

The authors gratefully acknowledge the Ministry of Education Malaysia, and Universiti Malaysia Terengganu that this research was partially supported under the Fundamental Research Grant Scheme (FRGS) vot. 59256.

References

Alhawarat, A., Mamat, M., Rivaie, M., and Mohd, I. (2014). A new modification of nonlinear conjugate gradient coefficients with global convergence properties. *International Journal of Mathematical, Computational, Physical and Quantum Engineering*. **8**: 54-60.

Adorio, E. P., and Diliman, U. P. (2005). Mvf-multivariate test functions library in c for unconstrained global optimization. http://www. geocities. ws/ eadorio/mvf.pdf

Al-Baali, M. (1985). Descent property and global convergence of the Fletcher—Reeves method with inexact line search. *IMA Journal of Numerical Analysis*. **5**: 121-124.

Andrei, N. (2008). An unconstrained optimization test functions, collection. *Advanced Modeling and Optimization*. **10**: 147-161.

Bongartz, I., Conn, A. R., Gould, N., and Toint, P. L. (1995). CUTE: Constrained and unconstrained testing environment. *ACM Transactions on Mathematical Software (TOMS)*. **21**: 123-160.

Dai, Y. H. and Yuan, Y. (1998). *Nonlinear Conjugate Gradient Methods*. Shanghai Scientific and Technical, Shanghai, China.

Dolan, E. D., and Moré, J. J. (2002). Benchmarking optimization software with performance profiles. *Mathematical programming*. **91**: 201-213.

Fletcher, R., and Reeves, C. M. (1964). *Function minimization by conjugate gradients. The computer journal*. **7**: 149-154.

Gilbert, J. C., and Nocedal, J. (1992). Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on optimization*. **2**: 21-42.

Polak, E., and Ribiere, G. (1969). *Note sur la convergence de méthodes de directions conjuguées*. ESAIM:MathematicalModelling and Numerical Analysis-Modélisation Mathématique Et Analyse Numérique. **3**: 35-43.

Shengwei, Y., Wei, Z. and Huang, H. (2007). A note about WYL's conjugate gradient method and its applications. *Applied Mathematics and computation.* **191**: 381-388.

Wei, Z., Yao, S., and Liu, L. (2006). The convergence properties of some new conjugate gradient methods. *Applied Mathematics and Computation.* **183**: 1341-1350.

Yu, G., Guan, L., and Wei, Z. (2009). Globally convergent Polak–Ribière–Polyak conjugate gradient methods under a modified Wolfe line search. *Applied Mathematics and Computation.* **215**: 3082-3090.

Zhang, L. (2009). An improved Wei–Yao–Liu nonlinear conjugate gradient method for optimization computation. *Applied Mathematics and computation.* **215**: 2269-2274.

Zoutendijk, G. (1970). Nonlinear programming, computational methods. *Integer and nonlinear programming.* **143**: 37-86.