

Global Convergence Properties of a New Class of Conjugate Gradient Method for Unconstrained Optimization

^{1,*}Abdelrahman Abashar, ²Mustafa Mamat, ³Mohd Rivaie and ¹Ismail Mohd

¹ Department of Mathematics, Faculty of Science and Technology
Universiti Malaysia Terengganu (UMT), Malaysia

² Faculty of Informatics and Computing
Universiti Sultan Zainal Abidin (Unisza)
K. Terengganu, Malaysia

³ Department of Computer Science and Mathematics
Universiti Teknologi MARA (UiTM) Terengganu
K. Terengganu, Malaysia

Copyright © 2014 Abdelrahman Abashar et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Nonlinear conjugate gradient (CG) methods are widely used for solving large scale unconstrained optimization problems. Many studies have been devoted to modified and improve this method. In this paper, a new parameter of CG method that possesses global convergence properties using exact line search is proposed. Numerical results show that the new formula is best and more efficient when compared with the other classical CG methods.

Mathematics Subject Classification: 65K10, 49M37

Keywords: Conjugate gradient, exact line search, global convergence

1. Introduction

Consider the following unconstrained optimization problem

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

where $f : R^n \rightarrow R$ is continuously differentiable function, the CG methods are the best methods for solving (1), especially when the dimension n is large. The iterates of CG methods for solving (1) are obtained by

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \dots, \quad (2)$$

where x_k is current iterate point and $\alpha_k > 0$ is step size. The step size is computed by carrying out some line search, especially the exact line search

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha d_k) \quad (3)$$

The search direction d_k is defined by

$$d_k = \begin{cases} -g_k & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1} & \text{if } k \geq 1, \end{cases} \quad (4)$$

where $\beta_k \in R$ is a scalar. We know that some classical formula's for β_k are the Hestenes-Stiefel (HS) in 1952, the Fletcher-Reeves (FR) in 1964, the Polak-Ribiere- Polyak (PRP) in 1969, the conjugate descent method (CD) in 1987, the Liu-Storey (LS) in 1992, and the Dai-Yuan (DY) in 2000. The parameters of these β_k are given as follows

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}}, \quad \beta_k^{FR} = \frac{g_k^T g_k}{\|g_{k-1}\|^2}, \quad \beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2},$$

$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}, \quad \beta_k^{LS} = \frac{g_k^T (g_k - g_{k-1})}{-d_{k-1}^T g_{k-1}}, \quad \beta_k^{DY} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}}.$$

The convergence properties of the FR conjugate gradient method have been studied by many authors, such as Al-Baali (1985) and Liu et al. (1990). Some researchers have also studied the convergence properties of the other classical formulas. For instance, Powell (1984) and Zoutendijk (1970) proved that the FR method with exact line searches is globally convergent on general functions. Al-Baali (1985) extended this result to inexact line searches, Touati-Ahmed

and Storey(1990), Gilbert and Nocedal (1992) gave another way to discuss the global convergence of the PRP method with the weak Wolfe-Powell line search. In this case, the parameter β_k is not allowed to be negative that is

$$\beta_k = \max\{\beta_k^{PRP}, 0\}$$

Some researchers have also presented good comparative study of some new CG methods such as Andrei (2009), Rivaie et al.(2012), Sun and Zhang (2001), Wei et al. (2006), Farid et al.(2013), Jusoh et al.(2013), and lastly Hager and Zhang (2005).

In this paper, we will show our new β_k and its algorithm in section 2. While in section 3, we show the sufficient descent condition and the global convergence proof of our new method. Later on in section 4, we deal with the Numerical results and discussion and finally we wrapped up everything as a conclusion in section 5.

2. New β_k and Algorithm

In mid of 2012, Rivaie et al. proposed a new nonlinear CG formula which is simple and easy to be used. This formula is given as

$$\beta_k^{RMIL} = \frac{g_k^T (g_k - g_{k-1})}{\|d_{k-1}\|^2}. \tag{5}$$

Even though this formula is simple, they have managed to show that this formula possessed four main requirements of CG formula, the sufficient descent conditions, global convergence properties, angle conditions and linear convergence rate. Intrigued by this new findings, we proposed a new modification of β_k which is defined by

$$\beta_k^{AMRI} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} |g_k^T g_{k-1}|}{\|d_{k-1}\|^2}. \tag{6}$$

The AMRI denotes Abashar, Mustafa, Rivaie and Ismail. It is clear that this new formula possessed the same denominator as RMIL. The following algorithm is a general algorithm for CG methods.

Algorithm 2.1

- Step 1: Initialization. Given x_0 , set $k = 0$.
- Step 2: Compute β_k , based on predetermined formula
- Step 3: Compute d_k based on (4). If $g_k = 0$, then stop.
- Step 4: Compute α_k based on exact line search (3).
- Step 5: Updating new point based on iterative formula (2).
- Step 6: Convergent test and stopping criteria.
 If $f(x_{k+1}) < f(x_k)$ and $\|g_k\| \leq \varepsilon$ then stop.
 Otherwise go to Step 1 with $k = k + 1$.

3. Global Convergence analysis

In this section, we will show the global convergent properties of β_k^{AMRI} . We first begin with the sufficient descent condition.

3.1. Sufficient descent condition

The sufficient descent condition, which is express as,

$$g_k^T d_k \leq -C \|g_k\|^2 \text{ for } k \geq 0, \quad C > 0 \quad (7)$$

The following theorem shows that AMRI with exact line search possess the sufficient descent condition.

Theorem 1.

Suppose that the x_k and d_k are generated by the method of the form (2), (4) and (6), and the step size $\alpha_k > 0$ determined by the exact line search then, condition (7) holds for all $k \geq 0$

Proof:

The proof is by Induction. If $k = 0$ then we already have $g_0^T d_0 = -C \|g_0\|^2$. Hence condition (7) holds true. Then to continue, we need to show that for $k \geq 1$, condition (7) will also holds true. From (4) multiply by g_{k+1}^T then

$$g_{k+1}^T d_{k+1} = g_{k+1}^T (-g_{k+1} + \beta_{k+1} d_k) = -\|g_{k+1}\|^2 + \beta_{k+1} g_{k+1}^T d_k. \quad (8)$$

For exact line search, we know that $g_{k+1}^T d_k = 0$. Thus

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2.$$

Hence, this condition holds true for $k + 1$. The proof is completed. \odot

3.2. Global convergence properties

To study the global convergence properties, first we must prove that β_k^{AMRI} are always not less than zero

$$\beta_{k+1}^{AMRI} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} |g_{k+1}^T g_k|}{\|d_k\|^2} \geq \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} \|g_{k+1}\| \|g_k\|}{\|d_k\|^2} = 0 \quad (9)$$

We can simplify β_{k+1}^{AMRI}

$$\beta_{k+1}^{AMRI} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} |g_{k+1}^T g_k|}{\|d_k\|^2} \leq \frac{\|g_{k+1}\|^2}{\|d_k\|^2} \quad (10)$$

In many theorem proofs, we also needed the following assumption

Assumption 1

- (i) The level set $\Omega = \{x \in R^n \mid f(x) \leq f(x_0)\}$ is bounded, where x_0 is the starting point.
- (ii) In some neighborhood N of, Ω the objective function is continuously differentiable, and its gradient is Lipschitz continuous, namely, there exists a constant $l > 0$ such that $\|g(x) - g(y)\| \leq l\|x - y\|$ for any $x, y \in N$.

Under this assumption, we have the following lemma, which was proved by Zoutendijk(1970).

Lemma 1.

Suppose Assumption 1 holds, let x_k be generated by Algorithm 2.1 and d_k satisfy (7) then the following condition, known as the Zoutendijk condition, holds

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty \quad (11)$$

The following theorem is based on Lemma 1.

Theorem 2.

Suppose that Assumption 1 holds true, x_k is generated by Algorithm 2.1, the α_k is obtained by the exact line search (3) and the sufficient descent condition hold true. Then either

$$\lim_{k \rightarrow \infty} \|g_k\| = 0 \quad \text{or} \quad \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty.$$

Proof:

To prove Theorem 2, we use contradiction. That is, if Theorem 2 is not true, then a constant $c > 0$ exists, such that

$$\|g_k\| \geq c \tag{12}$$

Rewriting (4) as

$$d_{k+1} + g_{k+1} = \beta_{k+1} d_k$$

And squaring both sides of the equation, we get

$$\|d_{k+1}\|^2 = (\beta_{k+1})^2 \|d_k\|^2 - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2 \tag{13}$$

Dividing both side by $(g_{k+1}^T d_{k+1})^2$ then,

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

Applying (9) we get

$$\leq \frac{1}{\|g_{k+1}\|^2} \tag{14}$$

Hence

$$\begin{aligned} \frac{\|d_k\|^2}{(g_k^T d_k)^2} &\leq \sum_{i=0}^k \frac{1}{\|g_i\|^2} \\ \frac{(g_k^T d_k)^2}{\|d_k\|^2} &\geq \frac{c^2}{k} \end{aligned} \tag{15}$$

There from (15) and (12), it follows that

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

This contradicts the Zoutendijk condition in Lemma 1. The proof is completed. \odot

Theorem 3

Suppose that Assumptions 1 holds, consider any CG the methods of form (2) and (4), the α_k obtained by the exact line search and β_k is determined by (6). Then either

$$\lim_{k \rightarrow \infty} \|g_k\| = 0 \text{ or } \sum_{k=0}^{\infty} \frac{(g_k^T)^4}{\|d_k\|^2} < \infty.$$

Proof:

From (13) and (10)

$$\begin{aligned} \|d_{k+1}\|^2 &= \left(\frac{\|g_{k+1}\|^2}{\|d_k\|^2} \right)^2 \|d_k\|^2 - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2 \\ \|d_{k+1}\|^2 &= \frac{\|g_{k+1}\|^4}{\|d_k\|^2} - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2 \end{aligned} \tag{16}$$

We have already proven that sufficient descent condition holds. Therefore, we know that

$$g_{k+1}^T d_{k+1} \leq -C \|g_{k+1}\|^2.$$

Hence from (16),

$$\begin{aligned} \|d_{k+1}\|^2 &= \frac{\|g_{k+1}\|^4}{\|d_k\|^2} + 2c \|g_{k+1}\|^2 - \|g_{k+1}\|^2 \\ \|d_{k+1}\|^2 &= \frac{\|g_{k+1}\|^4}{\|d_k\|^2} - \|g_{k+1}\|^2 (1 - 2c) \end{aligned} \tag{17}$$

Multiply both sides of (17) with $\frac{\|g_{k+1}\|^2}{\|d_{k+1}\|^2}$ then,

$$\begin{aligned} \|d_{k+1}\|^2 \frac{\|g_{k+1}\|^2}{\|d_{k+1}\|^2} &= \frac{\|g_{k+1}\|^2}{\|d_{k+1}\|^2} \left(\frac{\|g_{k+1}\|^4}{\|d_k\|^2} - \|g_{k+1}\|^2 (1 - 2c) \right) \\ \frac{\|d_{k+1}\|^2 \|g_{k+1}\|^2}{\|d_{k+1}\|^2} &= \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \left((2c - 1) + \frac{\|g_{k+1}\|^2}{\|d_k\|^2} \right) \end{aligned}$$

$$\frac{\|d_{k+1}\|^2 \|g_{k+1}\|^2}{\|d_{k+1}\|^2} \leq \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \quad (18)$$

Based on Theorem 2, we know that $\lim_{k \rightarrow \infty} \frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} < 0$.

This will imply that if theorem 3 is not true, then we have $\lim_{k \rightarrow \infty} \frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} = \infty$ and from (18) we

get $\infty \leq \frac{\|g_{k+1}\|^2}{\|d_{k+1}\|^2}$. Hence, Theorem 3 holds true for sufficiently large k .

4. Numerical results and discussions

In this section, we present some numerical results. Twentyone test problems considered in Andrei (2008) have been selected as shown in table 1, which ranges from small scale to large scale to analyze efficiency of AMRI. Based on this, a comparison was evaluated with the other CG methods which include FR, PRP and RMIL. We considered $\varepsilon = 10^{-6}$ and the gradient values as the stopping criteria base on the work of Hillstrom(1977) who suggested that the stopping criteria $\|g_k\| \leq \varepsilon$. For each of the test functions problem, four different initial points are utilized, taking from the one that is closer to the solution and moving to the one that is furthest from it. Therefore, these four initial points will lead us to test to global convergence properties of our method. All problems are tested using MATLAB version 7.10.0 (R 2010a) subroutine programming. We used the exact line search to overcome the complexity of the algorithm as well as how to achieve the exact value of step size value. The CPU processor used was Core™ i3-2328M(2.2GHZ,3MB L3 Cache), with 6GB DDR3 RAM. In some cases failure occurs when the line search were unable to find the positive step size. Numerical results are compared relatively on the number of iteration and CPU time. The performance results shown in Figure1 and 2, respectively, using a performance profile introduced by Dolan and More (2002).

The AMRI method shows the best performance, as it can solve all of the test problems perfectly as shown in Figure 1 and 2. These conjugate gradient coefficients could also be divided into three types. The first type consists of PRP and the second type consists of FR and third type RMIL method and our new method AMRI. We can say that the first type possesses the restart properties, and their performance is much better, the second type does not possess this property, and therefore is slower.

Table 1. A list of problem function

No	Functions	Variables	Initial point
1	Six hump Camel	2	(8,8),(-8,-8),(10,10),(-10,-10)
2	Booth	2	(10,10),(25,25),(50,50),(100,100)
3	Treccani	2	(5,5),(10,10),(50,50),(100,100)
4	Zettl	2	(5,5),(10,10),(20,20),(50,50)
5	Rosenbrock	2,4,10,100,500 ,1000,10000	(13,13,...,13),(16,16,...,16),(20,20,...,20),(30,30,...,30)
6	Extended Penalty	2,4,10,100	(3,3,...,3),(6,6,...,6),(10,10,...,10),(30,30,...,30)
7	Extended Beale	2,4,10,100,500 ,1000,10000	(1,1,...,1),(3,3,...,3),(7,7,...,7),(10,10,...,10)
8	Shallow	2,4,10,100,500 ,1000,10000	(10,10,...,10),(25,25,...,25),(50,50,...,50),(100,100,...,100)
9	Ex-Tridiagonal 1	2,4,10,100,500 ,1000,10000	(30,30,...,30),(12,12,...,12),(17,17,...,17),(20,20,...,20)
10	Raydan 1	2,4,10,100	(1,1,...,1),(3,3,...,3),(5,5,...,5),(-10,-10,...,-10)
11	White and Holst	2,4,10,100,500 ,1000,10000	(3,3,...,3),(6,6,...,6),(9,9,...,9),(-3,-3,...,-3)
12	Quadratic QF2	2,4,10,100,500 ,1000	(5,5,...,5),(7,7,...,7),(10,10,...,10),(-3,-3,...,-3)
13	Diagonal 4	2,4,10,100,500 ,1000,10000	(2,2,...,2),(5,5,...,5),(10,10,...,10),(15,15,...,15)
14	Extended Denschnb	2,4,10,100,500 ,1000,10000	(5,5,...,5),(8,8,...,8),(13,13,...,13),(25,25,...,25)
15	Hager	2,4,10,100	(3,3,...,3),(10,10,...,10),(15,15,...,15),(30,30,...,30)
16	Generalized Tridiagonal 1	2,4,100	(2,2,...,2),(5,5,...,5),(15,15,...,15),(25,25,...,25)
17	Nonscomp	2,4,100,500	(3,3,...,3),(10,10,...,10),(13,13,...,13),(15,15,...,15)
18	Perturbed Quadratic	2,4,10,100	(3,3,...,3),(10,10,...,10),(15,15,...,15),(30,30,...,30)
19	Diagonal 2	2,4,10,100,500 ,1000	(5,5,...,5),(10,10,...,10),(15,15,...,15),(20,20,...,20)
20	Quadratic Penalty QP2	2,4,10,100,500	(10,10,...,10),(20,20,...,20),(50,50,...,50), (100,100,...,100)
21	Himmelbau	2,4,10,100,500 ,1000,10000	(10,10,...,10),(50,50,...,50),(100,100,...,100),(200,200,...,200)

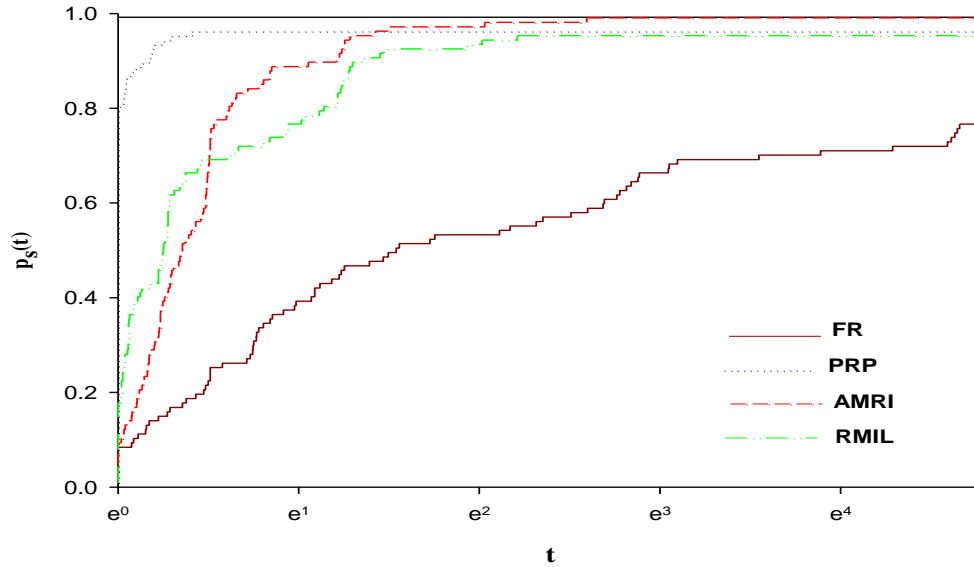


Figure 1 Performance profile based on the number of iterations.

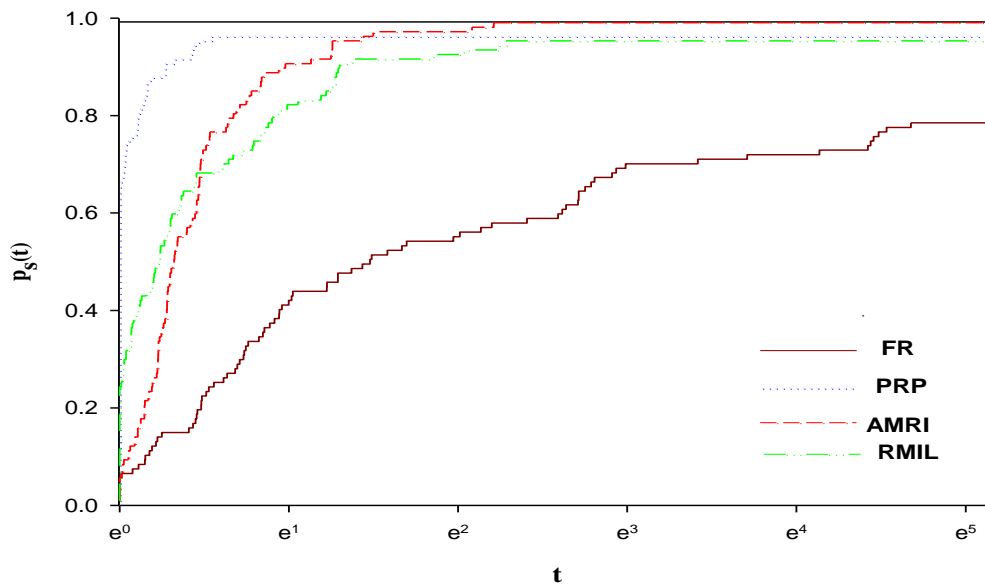


Figure 2 Performance profile based on the CPU time.

Although, the performance of the first type is faster than AMRI, but it can solve only 96% of the test function problems while the second type solves only 78% of the test function problems, and

RMIL method solve 95% of the test function problems. The performance of AMRI therefore lies between FR and PRP much nearer to the PRP method. We show that AMRI is best for other methods because it can solve all of the test function problems.

5. Conclusion

So many researchers have been done previously on CG method which led to the discovery of a variety of CG methods. In the same manner, our paper proposed a new and simple β_k that is easy to be implemented and possesses the global convergence properties. The numerical results have shown that our new method has the best performance when compared to other standard CG methods. For our next research, we are hoping that this new formula could still be improved by the used of inexact line search or by the introduction of new scaling factors.

Acknowledgements

The authors would like to thank the University of Malaysia Terengganu (UMT) for funding of this research under GGP 1/2013 and Ministry of Higher Education under FRGS Vot 59256.

References

- [1] Al-Baali, M. (1985). Descent property and global convergence of Fletcher-Reeves method with inexact line search. *IMA.J.Numer. Anal.*, 5: 121-124.
- [2] Andrei, N. (2008). An unconstrained optimization test functions collection, *Advanced Modeling and Optimization*. 10: 147–161.
- [3] Andrei, N. (2009). Acceleration of conjugate gradient algorithms for unconstrained optimization, *Applied Mathematics and Computation*. 213: 361–369.
- [4] Andrei, N. (2004). A new gradient descent method for unconstrained optimization. *ICI Technical report* (March 2004).
- [5] Dolan, E. and Moré, J.J. (2002). Benchmarking optimization software with performance profiles. *Mathematical Programming* no. 91 (2):201-213.
- [6] Dai, Y.H. and Yuan, Y. (1999). A nonlinear conjugate gradient with a strong global convergence property. *SIAM J.Optim.*,10:177-182.
- [7] Dai, Z.F. (2011). Two modified HS type conjugate gradient methods for unconstrained optimization problems. *Nonlinear Anal.*,74:927-936.

- [8] Fletcher, R. and Reeves, C. (1964). Function minimization by conjugate gradients. *Compute. J.* 7: 149–154.
- [9] Fletcher, R. (1987). *Practical Method of Optimization*, seconded., Unconstrained Optimization. vol.I, Wiley, New York.
- [10] Farid, M., Leong, W. J., Malekmohammadi, N., & Mamat, M. (2013). Scaled diagonal gradient-type method with extra update for large-scale unconstrained optimization. *Abstract and Applied Analysis*, article ID 532041, 5 pages.
- [11] Gilbert, J.C. and Nocedal, J. (1992). Global convergence properties of conjugate gradient methods for optimization. *SIAM J. Optim.* 2: 21–42.
- [12] Hillstom K.E. (1977). A simulation test approach to the evaluation of nonlinear optimization algorithm. *Journal ACM Trans. Mathematics Software* no 3 (4): 305-315.
- [13] Hager, W.W., and Zhang, H.C. (2005). A new conjugate gradient method with guaranteed descent and efficient line search. *SIAM J. Optim.* 16: 170–192.
- [14] Hestenes, M.R., and Stiefel, E. (1952). Method of conjugate gradient for solving linear equations. *J. Res. Nat. Bur. Stand.* 49 : 409–436.
- [15] Jie, S. ,and Jiapan, Z. (2001). Global convergence of conjugate gradient methods without line search. *Ann. Oper. Res.* 103:161–173.
- [16] Jusoh, I., Mamat, M., & Rivaie, M. (2013). A new family of conjugate gradient methods for small-scale unconstrained optimization. Paper presented at the *AIP Conference Proceedings*, 1522:1360-1365
- [17] Liu, Y., and Storey, C. (1991). Efficient generalized conjugate gradient algorithms. Part 1: theory. *J. Optim. Theory Appl.* 69: 129–137.
- [18] Powell, M J D. (1984). Non-convex minimization calculation and the conjugate gradient method. *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, 1066:122–241
- [19] Polak, E. and Ribierre, G. (1969). Note on the convergence of methods of conjugate directions. *Rev. Francaise Informat Recherche Operationelle*, 3:35-43.
- [20] Rivaie, M., Mamat, M., Leong, W. J. and Mohd, I. 2012. "A new class of nonlinear conjugate gradient coefficient with global convergence properties." *Appl. Math. comput.* 218 :11323-11332.

- [21] Touati-Ahmed, D. and Storey, C. (1990). Globally convergent hybrid conjugate gradient methods. *J. Optim. Theory Appl.* 64: 379–397.
- [22] Yuhong, D., Jiye, H., Guanghui L., Defeng S., Hongxia Y., and Yan, X. (1999). Convergence properties of nonlinear conjugate gradient methods. *SIAM J. Optim.* 10: 345–358.
- [23] Wei, Z., Shengwei, Y., and Linging, L. (2006). The convergence properties of some new conjugate gradient methods. *Appl. Math. comput.* 183:1341-1350.
- [24] Zoutendijk, G. (1970). Nonlinear programming computational methods. *J. Abadie (Ed.), Integer and Nonlinear Programming*, North-Holland, Amsterdam: 37–86.

Received: March 11, 2014