Applied Mathematical Sciences, Vol. 9, 2015, no. 38, 1857 - 1868
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ams.2015.411998

The Global Convergence Properties of an Improved Conjugate Gradient Method

Ibrahim S. Mohammed[1*], Mustafa Mamat[1], Abdelrahman Abashar[1, 2], Mohd Rivaie[3], Fatma Susilawati bt. Mohamad[1], and Zabidin Salleh[4]

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

[2]Faculty of Engineering
Red Sea University, Sudan

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

[4]Pusat Pangajian Informatik dan Matematik Gunaan
Universiti Malaysia Terengganu (UMT), Malaysia

Abstract

Conjugate gradient (CG) methods have played a significant role in solving large scale unconstrained optimization. This is due to its simplicity, low memory requirement, and global convergence properties. Various studies and modifications have been done recently to improve this method. In this paper, we proposed a new conjugate gradient parameter (β_k) which possesses global convergence properties under the exact line search. Numerical result shows that our new formula performs better when compared to other classical conjugate gradient methods.

Keywords: Conjugate gradient method; exact line search; global convergence

1. Introduction

Conjugate gradient method (CG) is designed to solve large scale unconstrained optimization problem. In general, the method has the following form.

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

where $f : R^n \to R$ is a continuously differentiable function. The CG methods are iterative methods of the form,

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

where $\alpha_k \geq 0$ is the step length computed using exact line search by the formula

$$f(x_k + \alpha_k d_k) = \min f(x_k + \alpha d_k), \tag{3}$$

and d_k is the search direction computed as follow

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

where g_k denotes the gradient of $f(x)$ at x_k. β_k is a coefficient that characterizes the conjugate gradient methods. Some well-known methods are given as follows,

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \tag{5}$$

$$\beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}, \tag{6}$$

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1})}, \tag{7}$$

$$\beta_k^{LS} = \frac{g_k^T (g_k - g_{k-1})}{-d_{k-1}^T g_{k-1}}, \tag{8}$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T (g_k - g_{k-1})}, \tag{9}$$

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

where g_{k-1} and g_k are the gradients of $f(x)$ at the point x_{k-1} and x_k respectively. Also, $\|.\|$ denotes Euclidean norm of vectors. The methods stated above are known as Fletcher-Reeves (FR) [18], Polak-Ribiere-Polyak (PRP) [5], Hestenes-Steifel (HS) [13], Liu-Storey (LS) [23], Dai-Yuan (DY) [21], and Conjugate Descent (CD) [17], respectively. It shows that if $f(x)$ is a strongly convex quadratic function, then in theory, all these methods are equal with the use of exact line search. However, for non-quadratic functions, their behaviors differ [22, 24].

The global convergence properties of the conjugate gradient methods under different line searches have been studied by many researchers. Zoutendijk [7], proves that FR method converges globally under exact line search. Powell [12], however, shows that the performance of the FR method is poor by giving a counter example. Al-Baali [10], Touati-Ahmed and Storey [3], Gilbert and Nocedal [8], have further analyzed the global convergence of algorithm related to the FR method with the strong Wolfe condition. Powell [11], further showed that FR method is superior when compared with other CG methods. For recent findings and further studies of the CG methods, refer to Andrei [16], Rivaie et al. [14], Sun and Zhang [9], Wei et al. [25], Hager and Zhang [19], Abdelrhaman et al. [1].

In this paper, we investigate the convergence and the efficiency of a new conjugate gradient method under the exact line search. In the next section, we propose a new formula for the coefficient β_k and algorithm. In section 3, we show that this new method satisfies the sufficient descent properties and the global convergence proof under the exact line search. In section 4, we carried out some numerical comparisons of our new method with FR, PRP and AMRI methods and discuss the results. Finally, section 5 present the concluding part of the work.

2. New Formula for the Cg Coefficient

Recently, Abdelrhaman et al. [1] modified a work done by Rivaie et al. [14]. He proposed a new coefficient β_k by modifying the numerator. This coefficient is known as β_k^{AMRI} and is expressed as

$$\beta_k^{AMRI} = \frac{g_k^T\left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{\|d_{k-1}\|^2} \tag{11}$$

where ||.|| denotes to the two norm of vectors.

Motivated by the idea of AMRI, we constructed a new β_k known as β_k^{SMAR} where SMAR denotes Sulaiman, Mustafa, Abdelrhaman, and Rivaie. The new method is given as follows,

$$\beta_k^{SMAR} = \frac{g_k^T \left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} d_{k-1} \right)}{\|d_{k-1}\|^2} \tag{12}$$

The algorithm is given below:

Algorithm 2.1

Step1. Given an initial point $x_0 \in R^n$, $\varepsilon \in (0,1)$, Set $d_0 = -g_0$, $k = 0$, if $\| g_0 \| \leq \varepsilon$, then stop.

Step2. Compute β_k^{SMAR} based on (12).

Step3. Compute d_k based on (4). If $\| g_k \| \leq \varepsilon$, then terminate, else go to step 4

Step4. Compute step size based on (3).

Step5. Update new point base on (2).

Step6. If $f(x_k) \leq f(x_{k+1})$ and $\| g_k \| \leq \varepsilon$, then terminate,

Otherwise, set $k = k + 1$ and go to Step 1.

3. Convergence analysis

In this section, the convergence properties of β_k^{SMAR} will be studied. First of all, we need to establish the sufficient descent condition properties.

3.1. Sufficient descent condition

For sufficient descent condition to hold, then

$$g_k^T d_k \leq -C \| g_k \|^2 \text{ for } k \geq 0 \text{ and } C > 0 \tag{13}$$

The following Theorem shows that our new formula with exact line search will possess the sufficient descent condition.

Theorem 1

Consider a CG method with the search direction (4) and β_k^{SMAR} given as (12), then condition (13) holds for all $k \geq 0$.

Proof:

If $k = 0$, then it is clear that $g_0^T d_0 = -C \| g_0 \|^2$. Hence condition (13) holds true. We also need to show that for $k \geq 1$, condition (13) will also hold true.

From (4), we have

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

By multiplying both sides of (4) by g_{k+1}, we obtain

$$g_{k+1}^T d_{k+1} = g_{k+1}^T(-g_{k+1} + \beta_{k+1}^{SMAR} d_K) = -\| g_{k+1} \|^2 + \beta_{k+1}^{SMAR} g_{k+1}^T d_k . \tag{14}$$

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

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

Therefore, it implies that d_{k+1} is a sufficient descent direction. Hence, $g_k^T d_k \le -C \| g_k \|^2$ holds true. The proof is completed. ■

3.2. Global convergence properties

In this section, we present some basic Assumptions which are often needed for global convergence analysis of CG methods.

Assumption 1

(1) f is bounded below on the level set R^n and is continuous and differentiable in a neighborhood N of the set $N = \{x \in R^n \mid f(x) \le f(x_0)\}$ at the initial point x_0.

(2) The gradient $g(x)$ is Lipschitz continuous in N, namely, so there exist a constant $L > 0$ such that $\| g(x) - g(y) \| \le L \| x - y \|, \quad \forall x, y \in N.$

From (12), we know that

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

Hence we obtained a simplification of SMAR as

$$0 \le \beta_k^{SMAR} \le \frac{\| g_{k+1} \|^2}{\| d_k \|^2} . \tag{16}$$

Under Assumption 1, we have the following useful Lemma, which was proved by Zoutendijk [7].

Lemma 1. Suppose Assumption 1 hold true. Consider any CG method of the form [4], where d_k is a descent search direction and α_k satisfies the one-dimensional search direction condition. Then, the Zoutendijk condition holds, which is given by

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\| d_k \|^2} < \infty . \tag{17}$$

By applying Lemma 1, we obtain the convergent theorem of the CG method given below using (16).

Theorem 2

Suppose assumption 1 holds true. Consider any CG method of the form (2) and (4), where α_k is obtained using exact line search. Then

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

Proof:

We prove by contradiction. That is, if Theorem 2 is not true, then, there exist a constant $c > 0$ such that

$$\|g_k\| \geq c \tag{18}$$

We rewrite (4) as

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

Squaring both side of the equation gives

$$\|d_k\|^2 = (\beta_k)^2\|d_{k-1}\|^2 - 2g_k^T d_k - \|g_k\|^2 \tag{19}$$

Dividing both side of (19) by $\left(g_k^T d_k\right)^2$, then we get

$$\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} = \frac{(\beta_k)^2\|d_{k-1}\|^2}{\left(g_k^T d_k\right)^2} - \frac{2g_k^T d_k}{\left(g_k^T d_k\right)^2} - \frac{\|g_k\|^2}{\left(g_k^T d_k\right)^2}$$

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

$$\leq \frac{(\beta_k)^2\|d_{k-1}\|^2}{\left(g_k^T d_k\right)^2} + \frac{1}{\|g_k\|^2} \tag{20}$$

From (16), it becomes

$$\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} \leq \frac{1}{\|g_k\|^2}$$

Hence

$$\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} \leq \sum_{i=0}^{k}\frac{1}{\|g_i\|^2}$$

$$\frac{g_k^T d_k}{\|d_k\|^2} \geq \frac{c^2}{k} \tag{21}$$

Therefore, from (21) and (18), it shows that

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

This, however, contradicts the Zoutendijk condition in Lemma 1. Therefore, the proof is completed. ■

4. Numerical Results

In this section, we present some numerical performance based on the comparisons of our proposed algorithm with FR, PRP and AMRI algorithms. The comparisons are based on the number of iterations and CPU time. We considered $\| g_k \| < 10^{-6}$ to be stopping criteria. For each of the test problems, four initial points are used. These four initial points will lead us to test the global convergence and the robustness of our method. All algorithms were implemented under exact line search to avoid complexity and to obtain the actual value of the step length. All codes of the test problems considered in Table 1 were written on MATLAB 7.6.0 (R 2008a) subroutine programming. This was run on Intel® Core™ i5-2410M CPU @ 2.30 GHz processor, 4GB for RAM memory and Windows 7 Professional operating system. Most of the test functions considered are from Andrei [15]. The numerical results are shown in Table 1. The performance results are presented in Figure 1 and Figure 2 respectively, based on the performance profile introduced by Dolan and More [4]. In the performance profile, they introduced the notion of a process used to evaluate and compare the performance of the set of solvers S on a test P. Suppose there exist n_s solvers and n_p problems, for each problem p and solver s, they define

$t_{p,s}$ = computing time needed to solve problem by solver (the number of iteration or CPU time).

Requiring a baseline for comparisons, they compared the performance on problem p by solver s with the best performance by any solver on this problem using the performance ratio

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s} : s \in S\}}$$

We suppose that parameter $r_m \geq r_{p,s}$ for all p, s is chosen, and $r_{p,s} = r_M$ if and only if solver s does not solve problem p. The performance of solvers s on any given problem might be of interest, but because we would prefer obtaining the overall assessment of the performance of the solver, then it was defined as

$$p_s(t) = \frac{1}{n_p} size\{p \in P : r_{p,s} \leq t\}.$$

Thus $p_s(t)$ was the probability for solver $s \in S$ that a performance ratio $r_{p,s}$ was within a factor $t \in R$ of the best possible ration. Then, function p_s was the cumulative distribution function for the performance ratio. The performance profile $p_s : R \to [0,1]$ for a solver was a non-decreasing, piecewise, and continuous from right. The value of $p_s(1)$ is the probability that the solver will win over the rest of the solvers. In general, a solver with high value of $p(\tau)$ or at the top right of the figure are preferable or represent the best solver

Table 1: List of problem functions

No	Function	Dimension	Initial Points
1	Six hump	2	(8, 8), (-8, -8), (10, 10), (-10, -10)
2	Three hump	2	(5, 5), (7, 7), (11, 11), (15, 15)
3	Booth	2	(10, 10), (25, 25), (50, 50), (100, 100)
4	Treccani	2	(5, 5), (7, 7), (50, 50), (100, 100)
5	Leon	2	(-0.5, 1), (5, -5), (5, 15), (5, -15)
6	Matyas	2	(2, 2), (5, 5), (10, 10), (20, 20)
7	Dixon and Price	2, 4	(5, 5), (13, 13), (30, 30), (33, 33)
8	Fletcher	2, 4, 10	(5, 5,…, 5), (7, 7, …, 7), (10, 10, …,10), (15, 15, …,15)
9	Extended Maratos	2, 4, 10	(15, 15,…,15), (20, 20,…20), (50, 50,…,50), (100, 100,…,100)
10	Extended Penalty	2, 4, 10, 100	(100,100,…,100),(105,105,…,105), (135,135,…,135),(200,200,…,200)
11	Generalized Trig	2, 4, 10, 100	(1, -1 ,…,-1), (5, 5, …, 5), (12, 12, …, 12), (23, 23, …23)
12	Raydan1	2, 4, 10, 100	(1, 1, …, 1), (3, 3, …, 3), (-0.5, 1, …, 1), (-10, -10, …,-10)
13	Hager	2, 4, 10, 100	(-1, -1, …, -1), (21, 21, …, 21), (23, 23, …, 23), (-23, 23, …,23)
14	Perturbed Quadratic	2, 4, 10, 100	(0.5,0.5, …,0.5),(1, 1, …,1), (101,101, …, 101), (105, 105, …, 105)
15	Quadratic Penalty QP2	2, 4, 10, 100, 500	(-11, 11, …, 11), (13, 13, …, 13), (-15, 15, …, 15), (-25, 25, …, 25)
16	Feudenstein & Roth	2, 4, 10, 100, 500, 1000	(1, 3, …, 3), (7, 7, …, 7), (23, 23, …, 23),(33, 33, …, 33)
17	Rosenbrock	2, 4, 10, 100, 500, 1000, 10000	(13, 13, …, 13), (24, 24, …,24), (33, 33, …, 33), (35, 35 , …, 35)
18	Shallow	2, 4, 10, 100, 500, 1000, 10000	(11, 11, …, 11), (25, 23, …,23), (25, 25, …, 25), (35, 35, …, 35)
19	Extended Tridiagonal	2, 4, 10, 100, 500, 1000, 10000	(12, 12, …, 12), (17, 17, …,17), (20, 20, …, 20),(30, 30, …, 30)
20	Diagonal 4	2, 4, 10, 100, 500, 1000, 10000	(2, 2, …, 2), (5, 5, …,5), (10, 10, …, 10), (15, 15, …, 15)
21	Extended Denschnb	2, 4, 10, 100, 500, 1000, 10000	(3, 3, …, 3), (8, 8, …,8), (-15, 15, …, 15), (-25, 25, …, 25)
22	Extended Beale	2, 4, 10, 100, 500, 1000, 10000	(-0.5, 0.5, …, 0.5), (0.5, 1, …,1), (-2, -0.5, …, -0.5), (7, -9, …, -9)
23	Himmelblau	2, 4, 10, 100, 500, 1000, 10000	(0, 1, …, 1), (0.5, 5, …,5), (-6, -6, …, -6), (-15, 5, …, 5)
24	Generalized Quartic	2, 4, 10, 100, 500, 1000, 10000	(0.5, 0.5, …, 0.5), (1, 1, …, 1), (3, 3, …, 3), (5, 5, …, 5)
25	Ext White and Holst	2, 4, 10, 100, 500, 1000, 10000	(0, 0, …, 0), (-0.5, 1, …,1), (-5, 10, …, 10), (-7, 7, …, 7)

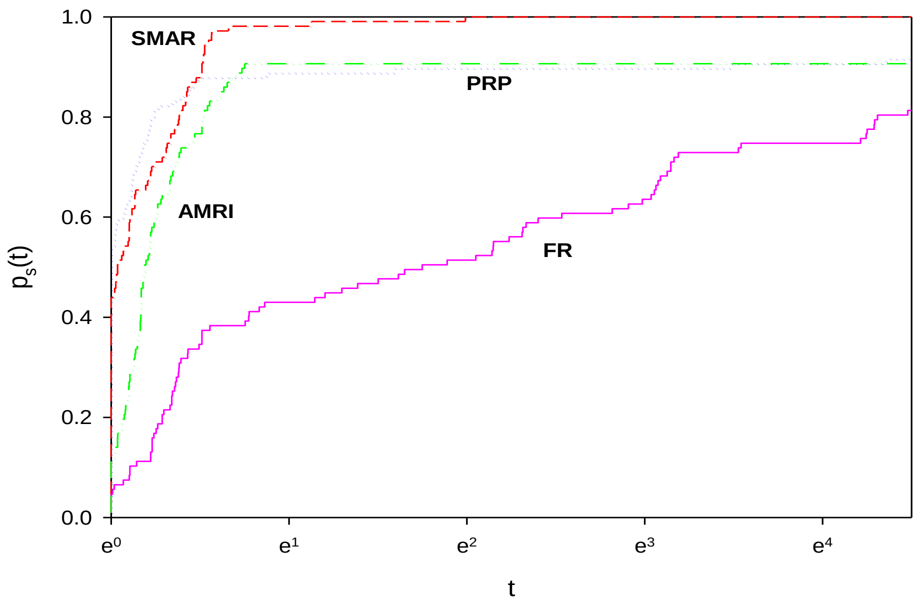

Fig.1: Performance profile based on the number of iterations.

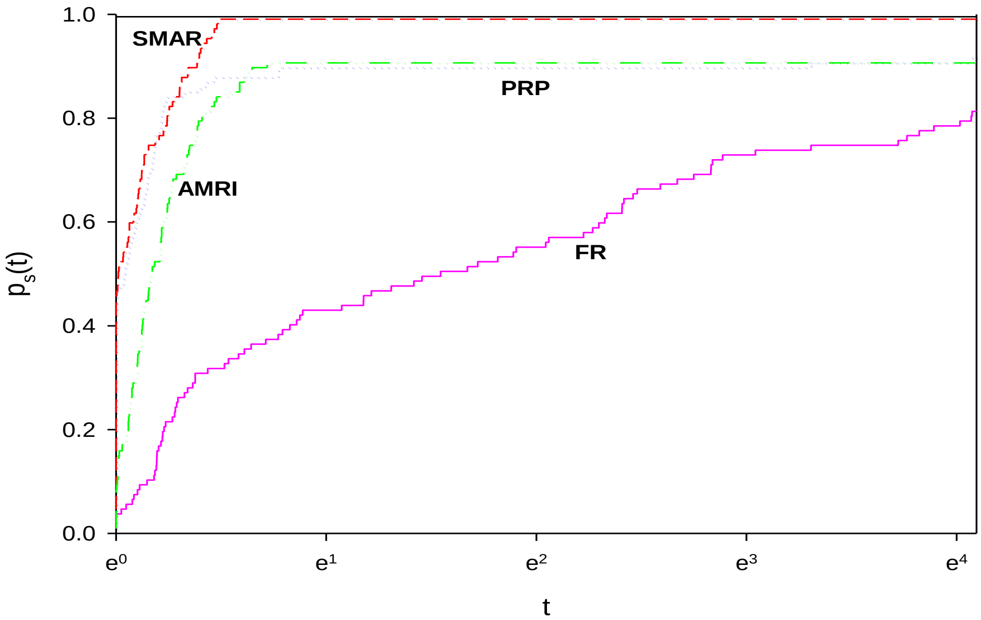

Fig.2: Performance profile based on the CPU time.

Clearly, from Figures 1 and 2, we realize that our proposed algorithm was able to solve 100% of all the test problems, whereas, the FR method was only able to solve 81% of the test problems, AMRI solve 90% of the problems, and PRP was only able to solve 91% of the test problems. Therefore, we conclude that SMAR method performs better than FR, PRP and AMRI methods.

5. Conclusion

In this study, we have examined the conjugate gradient method with a new formula (12) under the exact line search and have proved that it converges globally under some assumptions. We have also shown that the sufficient descent condition holds for all search directions if we use the exact minimization rule. Numerical results show that our proposed method performs better than FR, PRP and AMRI. In future, we hope to test this new β_k^{SMAR} using inexact line search.

Acknowledgments. The authors would like to thank the government of Malaysia for the funding of this research under the Fundamental Research Grant Scheme (Grant no. 59256) and also the government of Kano State, Nigeria.

References

[1] A. Abashar, M. Mustafa, M Rivaie and M. Ismail, The proof of sufficient descent condition for a new type of conjugate gradient method, *AIP Conf. Proc.* **296** 1602 (2014); http://dx.doi.org/10.1063/1.4882502

[2] B. T. Polyak, The conjugate gradient method in extreme problems, *USSR Comp. Math. Phys*. **9** (1969), 94-112. http://dx.doi.org/10.1016/0041-5553(69)90035-4

[3] D. Touati-Ahmed, C. Storey, Efficient hybrid conjugate gradient techniques, *J. Optim. Theory Appl*. **64** (1990), 379–397. http://dx.doi.org/10.1007/bf00939455

[4] E. Dolan, J.J. More, Benchmarking optimization software with performance profile, *Math. Prog.* **91** (2002), 201–213. http://dx.doi.org/10.1007/s101070100263

[5] E. Polak and G. Ribière, Note sur la convergence de directions conjuguée, *Rev. Francaise Informat Recherche Operationelle, 3e Année* 1**6** (1969), 35-43.

[6] G. H. Liu, J. Y. Han, and H. X. Yin, Global convergence of the Fletcher-Reeves algorithm with an inexact line search, Report, *Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing*, 1993.

[7] G. Zoutendijk, Nonlinear programming, computational methods, in Integer and Nonlinear Programming, *J. Abadie, ed., North- Holland, Amsterdam*, (1970), 37-86.

[8] J.C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM. J. Optim*. **2** (1992), 21-42. http://dx.doi.org/10.1137/0802003

[9] J. Sun, J Zhang, Global convergence of conjugate gradient methods without line search, *Ann. Oper. Res*. **103** (2001), 161-173.

[10] M. Al-Baali, Descent property and global convergence of the Fletcher-Reeves method with inexact line search, *IMA J. Numer. Anal*. **5** (1985), 121-124. http://dx.doi.org/10.1093/imanum/5.1.121

[11] M.J.D. Powell, Convergence properties of algorithm for nonlinear optimization, *SIAM Rev*. **28** (1986), 487-500. http://dx.doi.org/10.1137/1028154

[12] M.J.D. Powell, Restart procedures for the conjugate gradient method, *Math. Program*. **12** (1977), 241-254. http://dx.doi.org/10.1007/bf01593790

[13] M. R. Hestenes and E. L. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Research Nat. Bur. Standard.* **49** (1952), 409-436. http://dx.doi.org/10.6028/jres.049.044

[14] M. Rivaie, M. Mamat, W.J. Leong and M. Ismail, A new class of nonlinear conjugate gradient coefficients with global convergence properties, *Applied Mathematics and Computation.* **218** (2012), 11323-11332. http://dx.doi.org/10.1016/j.amc.2012.05.030

[15] N. Andrei, An unconstrained optimization test functions collection, *Advanced Modelling and optimization* **10(1)** (2008), 147-161.

[16] N. Andrei, Accelerated conjugate gradient algorithm with finite difference Hessian / vector product approximation for unconstrained optimization. *J. comput. Appl. Math*. **230** (2009), 570–582. http://dx.doi.org/10.1016/j.cam.2008.12.024

[17] R. Fletcher, Practical Method of Optimization, Vol. 1, Unconstrained Optimization, John Wiley & Sons, New York, 1987.

[18] R. Fletcher and C. Reeves, Function minimization by conjugate gradients, *Comput. J.* **7** (1964), 149-154. http://dx.doi.org/10.1093/comjnl/7.2.149

[19] W.W. Hager, and H.C. Zhang, A new conjugate gradient method with guaranteed descent and efficient line search. *SIAM J. Optim.* **16** (2005), 170–192.

http://dx.doi.org/10.1137/030601880

[20] Y. F. Hu and C. Storey, Global convergence result for conjugate gradient methods, *J. Optim. Theory Appl.* **71** (1991)**,** 399–405.
http://dx.doi.org/10.1007/bf00939927

[21] Y.H. Dai, Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence properties, *SIAM J. Optim*. **10** (1999), 177-182.
http://dx.doi.org/10.1137/s1052623497318992

[22] Y.H Dai, Y. Yuan, Nonlinear Conjugate Gradient Method, Shanghai Scientific and Technical Publisher, Beijing, 1998.

[23] Y. Liu, C. Storey, Efficient generalized conjugate gradient algorithms part 1: Theory, *J. Comput. Appl. Math*. **69** (1991), 129-137.
http://dx.doi.org/10.1007/bf00940464

[24] Y. Yuan, and W. Sun, Theory and Methods of Optimization. Science press of China, Beijing, 1999.

[25] Z. Wei, Y. Shengwei, and L. Linging, The convergence properties of some new conjugate gradient methods. *Appl. Math. Comput*. **183** (2006), 1341-1350.
http://dx.doi.org/10.1016/j.amc.2006.05.150

Received: December 10, 2014; Published: March 9, 2015