International Journal of Engineering & Technology, 7 (4.33) (2018) 521-525.

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A New Conjugate Gradient Method with Exact Line Search

Mouiyad Bani Yousef[1], Mustafa Mamat[1]*, Mohd Rivaie[2]

[1]*Department of Mathematics, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Kuala Terengganu, Malaysia*
[2]*Department of Computer Sciences and Mathematics, Universiti Technology MARA (UiTM) Terengganu, Kuala Terengganu Campus, Malaysia*
**Corresponding author E-mail: must@unisza.edu.my*

Abstract

The nonlinear conjugate gradient (CG) method is a widely used approach for solving large-scale optimization problems in many fields, such as physics, engineering, economics, and design. The efficiency of this method is mainly attributable to its global convergence properties and low memory requirement. In this paper, a new conjugate gradient coefficient is proposed based on the Aini-Rivaie-Mustafa (ARM) method. Furthermore, the proposed method is proved globally convergent under exact line search. This is supported by the results of the numerical tests. The numerical performance of the new CG method better than other related and more efficient compared with previous CG methods.

***Keywords**: Conjugate gradient method; exact line search; sufficient descent property; Global convergence; performance profile.*

1. Introduction

In general, an unconstrained minimization problem has the following form:

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

where $f: R^n \to R$ is a continuously differentiable function The iterative formula commonly used for solving (1) is given as follows:

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

where x_k is the current iteration point, and $\alpha_k > 0$ is the step-size obtained using the exact line search formula :

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

The search direction d_k, is calculated by using the CG formula which is defined by

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

The parameter $\beta_k \in R$ is the CG coefficient that characterizes different CG methods, while the gradient of $f(x)$ at point x_k is denoted by g_k Some examples of well-known formulas for β_k are Polak-Ribiere-Polyak (PRP) [1, 2], Fletcher and Reeves (FR) [3], Wei et al. [4], the 'Aini-Rivaie-Mustafa (ARM) method [5], Hestenes and Stiefel (HS) [6], Conjugate Descent (CD) by Fletcher [7] and Dai-Yuan (DY) [8]. Their formulas are given as follows :

$$\beta_K^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} \tag{5}$$

$$\beta_K^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \tag{6}$$

$$\beta_K^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} \tag{7}$$

$$\beta_k^{ARM} = -\frac{m_k \|g_k\|^2 - |g_k^T g_{k-1}|}{m_k g_{k-1}^T d_{k-1}},$$

where

$$m_k = \frac{\|d_{k-1} + g_k\|}{\|d_{k-1}\|} \tag{8}$$

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

$$\beta_K^{CD} = \frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}} \quad (10)$$

$$\beta_K^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T (g_k - g_{k-1})} \quad (11)$$

The convergence properties of CG method had been studied in various researches. Polak and Ribiere proved the global convergence of PRP method with the exact line search in [2]. Zoutendijk [9] determined the convergence properties of FR method under exact line search. Later, Al-Baali [10] and Touti-Ahmed and Storey [11] extended the proof of FR for Wolfe line search. In [12], Powell gave out a counterexample clarifying that there exists a non-convex function for which the PRP method fails even under exact line search. Afterwards, Gilbert and Nocedal [13] suggested the PRP+ method and proved its global convergence under the Wolfe-Powell line search. Wei et al. [4] presented a positive CG method that looks like the original PRP method.

Lately, new designs for CG coefficient are more dedicated on ensuring good numerical performance and global convergence under various types of line search. Current references on nonlinear CG method can be found in Powell [14], Hager and Zhang [15], Andrei [16], Wei et al. [4], Sofi et al. [17]. Abashar et al. [18], Jusoh et al. [19] and Rivaie et al. [20, 21].

In this paper, we propose a modified CG method where its formula is presented in Section 2. Section 3 highlights the convergence analysis of the proposed CG method under exact line search. In the next section, we evaluate the numerical efficiency of the new method by comparing its performance with other CG methods. Finally, Section 5 presents the conclusion of the study.

2. The new CG coefficient formula

β_k^{MMR} is extended to β_k^{ARM} method, that is MMR denotes the researchers; Mouiyad, Mustafa, and Rivaie. The formula is as the follow,

$$\beta_k^{MMR} = \frac{m_k \|g_k\|^2 - \left(g_k^T g_{K-1}\right)}{m_k \|g_{k-1}\|^2}$$

where

$$m_k = \frac{\|d_{k-1} + g_k\|}{\|d_{k-1}\|} \quad (12)$$

Now algorithm is givens as follow:

Algorithm 2.1

1st Step: Initialization. Given $x_0 \in R^n$, $d_0 = -g_0, k = 0$.

2nd Step: Compute β_k based on (12).

3rd Step: Compute d_k based on (4). If $\|g_k\| = 0$, then stop.

4th Step: Compute stepsize based on (3).

5th Step: Update the new point based on (2)

6th Step: If $f(x_k) < f(x_{k+1})$ and $\|g_k\| \le \varepsilon$, then stop.

Otherwise, set $k = k+1$ and return to Step 2.

3. The Convergence analysis of MMR method

This section shows that, the convergence properties of our new CG coefficient β_k^{MMR}, is studied. And for the above algorithm to be convergent, it should fulfil the sufficient descent condition and possess global convergence properties.

3.1. The sufficient descent condition

The following formula gives the sufficient descent condition:

$$g_k^T d_k \le -c\|g_k\|^2 \quad \text{for } k \ge 0 \text{ and } c > 0. \quad (13)$$

Theorem A

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

Proof: If k = 0, then $g_0^T d_0 \le -c\|g_0\|^2$. Hence, condition (13) true. We also need to show that for $k \ge 0$, condition (13) will also true.

From (4), we have

$$d_{k+1} = -g_{k+1} + \beta_{k+1}^{MMR} d_k \quad (14)$$

Now multiply both sides of (14) by g_{k+1}, then we have,

$$g_{k+1}^T d_{k+1} = g_{k+1}^T(-g_{k+1} + \beta_{k+1}^{MMR} d_k)$$
$$= -\|g_{k+1}\|^2 + (\beta_{k+1}^{MMR} g_{k+1}^T d_k) \quad (15)$$

using exact line search $g_{k+1}^T d_{k+1} = 0$. Then,

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

this implies that d_{k+1} is a sufficient descent direction.

Hence, $g_{k+1}^T d_k \le -c\|g_k\|^2$ holds. The proof is completed.

3.2. The Global convergence properties

The analysis on global convergence of the proposed CG method requires the following assumptions.

Assumption A $f(x)$ is bounded from below on the level set $\Omega = \left\{ x \in R^n / f(x) \le f(x_0) \right\}$, where x_0 is the starting point and $f(x)$ is a continuously differentiable function in a neighborhood N of the level set Ω .

Assumption B The gradient $g(x)$ is Lipschitz continuous in N , hence there exists a constant $L > 0$ such that $\|g(x) - g(y)\| \le L\|x - y\|$ for any $x, y \in N$.

To make proof of our convergence easier, we first simplify our new β_k^{MMR} from (12),

This implies that

$$0 \le \beta_{k+1}^{MMR} \le \frac{\|g_k\|^2}{\|g_{k-1}\|^2} = \beta_k^{FR} \tag{16}$$

In [9], Zoutendijk proves the following lemma. Also this lemma holds true for the exact minimization rule, the Wolfe rules and the Goldstein rules as shown in [22].

Lemma A: Suppose that assumption A and B holds and the CG method is defined as (4), where d_k is a descent search direction and α_k satisfies the one-dimensional search direction condition. Then, the following condition holds.

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

By using Lemma A, we can get the following convergence theorem of the MMR CG method.

Theorem 2

Assume that Assumption A and B holds, and the MMR CG method is defined as (2) and (4) with (12) as the β_k **.** If the α_k is obtained by exact minimization rule and the descent condition holds true, then,

$$\lim_{K\to\infty} \|g_k\| = 0 \tag{18}$$

Proof: Theorem 2 will be proved by contradiction. Suppose that Theorem 2 is not true, then a constant a > 0 exists such that

$$\|g_k\| \ge a \tag{19}$$

Rewriting search direction (4) as

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

and then squaring both sides of the equation, we obtain

$$\|d_{k+1}\|^2 = (\beta_{k+1}^{MMR})^2 \|d_k\|^2 - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2 \tag{20}$$

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

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

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

$$\le \frac{(\beta_{k+1}^{MMR})^2 \|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2} \tag{21}$$

Applying (16) yields

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

$$= \frac{\|d_k\|^2}{\|g_{k-1}\|^4} + \frac{1}{\|g_{k+1}\|^2}$$

$$\frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \le \frac{\|d_k\|^2}{\|g_{k-1}\|^4} + \frac{1}{\|g_{k+1}\|^2} \tag{22}$$

Note that, $g_k^T d_k = -\|g_k\|^2$ and $(g_{k+1}^T d_{k+1})^2 = \|g_k\|^4$. Hence,

$$\frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} = \frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \le \frac{\|d_k\|}{\|g_{k-1}\|^4} + \frac{1}{\|g_{k+1}\|^2}$$

By noting that $\|d_0\|^2 = \|g_0\|^2$ and using recursive, we obtain

$$\frac{\|d_k\|^2}{\|g_{k+1}\|^2} \le \sum_{i=0}^{k} \frac{1}{\|g_i\|^2} \tag{23}$$

Therefore, from (19) and (23), it shows that

$$\frac{\|g_k\|^4}{\|d_k\|^2} \ge \frac{a^2}{k},$$

Hence,

$$\sum_{k=0}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} = \sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} = \infty$$

This contradiction the Zoutendijk condition in Lemma A. Thus, a proof is completed.

4. The Numerical results and discussion

section four is centered on analyzing the efficiency of β_k^{MMR}**,** starting by testing it on a set of unconstrained optimization test problems and comparing its performance with four CG methods (PRP, FR, WYL, and ARM). These comparisons are based on number of iteration and CPU time. The stopping criterion is set at $\|g_k\| \le 10^{-6}$. The test functions used are selected from [23] and tested with different dimensions $(2 \le n \le 10000)$. For each of the test problems, three initial points of varying distances from the solution point are used in order to test the global convergence properties of the new formula.

The list of tests and initial points are presented in Table 1.

Table 1: A list of problem functions

No.	Functions	n	Initial Points
1	Six Hump	2	(4,4)(-21,-21)(43,43)
2	Three Huump	2	(3,3)(21,21)(62,62)
3	Booth	2	(-8,-8)(49,49)(80,80)
4	TRECCANI	2	(20,20)(79,79) (-2.1 ,2)
5	Zettl	2	(6,6)(20,20)(-100,-100)
6	Colville	4	(99,…,99)(-20,..,-20) (-150,...,-150)
7	Raydan 2	2,4	(1,3)(-17,16)(2,24)
8	Dixon and Price	2,4	(101,…,101)(1,…,1) (23,…,23)
9	ARrwhead	2,4,10	(3,…,3)(23,…,23) (81,…,81)
10	Hager	2,4,10,100	(6,…,6)(-17,…,-17) (-78,…,-78)
11	Raydan 1	2,4,10,100	(20,…,20) (1,…,1) (10,…,10)
12	Freudestein and Roth	2,4,10,100	(4.5,...,4.5)(31,…,31) (-19,…,-19)
13	Extended Maratos	2,4,10,100 500,1000	(18,…,18) (-84,-106) (-4.5,…,-4.5)
14	DQDRTIC	2,4,10,100 500,1000	(10,...,10)(50,...,50) (100,…,100)
15	Generalized Quartic 1	2,4,10,100 500,1000	(10,…,10)(20,…,20) (80,…,80)
16	Extended Beale	2,4,10,100 500,1000	(11,…,11)(30,…,30) (-1.3,…,-1.3)
17	Extended Denschnb	2,4,10,100 500,1000	(3,…,3)(23,…,23) (200,…,200)
18	Extended Himmelblau	2,4,10,100 500,1000	(1,5)(10,…,10) (41,…,41)
19	Extended Shallow	2,4,10,100 500,1000	(11,…,11)(-1,…,-1) (-50,110)
20	Extended Strait	2,4,10,100 500,1000	(10,...,10)(50,…,50) (100,…, 100)
21	Extended Tridiagonal1	2,4,10,100 500,1000	(1.9,…,136)(2,…,2) (17,...,17)
22	Sum Squares	2,4,10,100 500,1000	(3.7,...,3.7)(15,…,15) (35,…,35)
23	Quartic	2,4,10,100 500,1000	(3,…,3)(19,…,19) (29,...29)
24	White and Holst	2,4,10,100 500,1000	(10,…,100)(11,…,11) (-1.3,…,-1.3)
25	FLETCHCR	2,4,10,100 500,1000	(-1.3,5) (-110,-111) (37,…,37)
26	Qing	2,4,10,100 500,1000	(7,…,7)(80,...,80) (-2,…,-2)
27	Styblinski-Tank	2,4,10,100 500,1000	(14,…,14)(80,..,80) (-0.1,…,0.2)
28	Extended Quadratic penalty	2,4,10,100 500,1000	(2,…,2)(19,…,19) (59,…,59)
29	Qudratic 2	2,4,10,100 500,1000	(0.5,…,0.5)(20,...,20) (80,…,80)
30	Diagonal 4	2,4,10,100 500,1000	(10,…,10)(50,…,50) (100,…,100)
31	Diagonal 2	2,4,10,100 500,1000	(1,…,1)(5,…,5) (15,…,15)
32	Extended Quadratic penalty2	2,4,10,100 500,1000	(0.5,…,0.5)(21,…,21) (50,…,50)
33	Extended Rosenbrock	2,4,10,100 500,1000	(2,…,2)(20,…,20) (80,…,80)
34	Extended Block Diagonal1	2,4,10,100 500,1000	(0.1,..,0.1)(10,…,10) (10,100)
35	TRIDIA	2,4,10,100 500,1000	(1,…,1)(10,…,10) (50,…,50)

We employ MATLAB version R2015a subroutine programming to execute the algorithms on a PC computer with Intel(R) Core™ i3-4005U CPU @ 1.70GHz processor, 4GB RAM, and Windows 10 Professional operating system. To graph the data we used the Sigma Plot 10 programme as shown in Figures 1 and 2. The performance of all tested CG methods is studied by using performance profile proposed by Dolan and More [24]. This approach evaluates and compares the performance of set of interested solvers *S* on a whole set of test problems *P*.

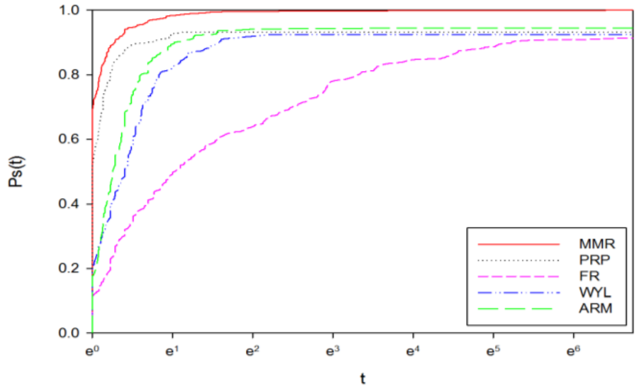

Fig. 1: Performance profile based on the number of iterations.

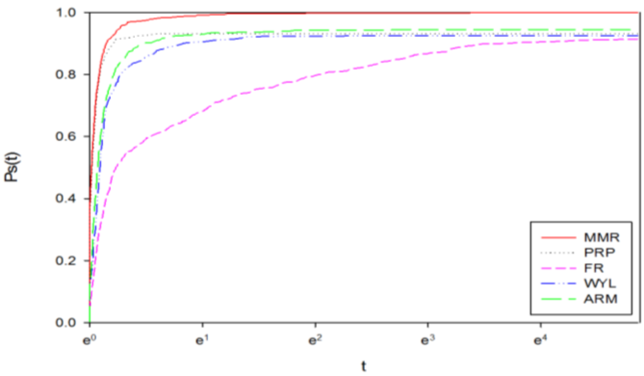

Fig. 2: Performance profile based on the CPU time.

The value of $P_S(1)$ in the performance profile is the probability that the method manages to solve all the tested problems, thus winning through the rest of the solvers. In general, a solver with high value of $P_s(t)$ or positioned at the top right of the figure is regarded as the better solver. Figures 1 and 2 plot the performance of our new method in comparison with the PRP, FR, WYL, and ARM methods. By observing the top position in both figures, which are occupied by MMR, it is clear that MMR has the better performance in terms of number of iteration and CPU time. In regards to number of problems solved, the PRP, FR, WYL, and ARM methods are only able to solve 93%, 91 %, 92 %, and 94% of all the test problems, respectively. On the contrary, our new formula solves 100% of the test problem functions and reaches $P_s(1)$, which makes it the best solver for this test.

5. Conclusion

In this paper, we proposed a new conjugate gradient CG method and proved its global convergence and sufficient descent properties under exact line search. Also, the numerical results prove that our new modification is better than other tested CG parameters.

Acknowledgement

The authors would like to thank the editors and the referees for their suggestions on our paper; and we would like to thank Universiti Sultan Zainal Abidin, for financial support under the fundamental research Grant Scheme (FRGS)/1/2017/STG06/UniSZA/01/1

References

[1] Polyak, Boris T. "The conjugate gradient method in extremal problems." *USSR Computational Mathematics and Mathematical Physics* 9.4 (1969): 94-112.
[2] Polak, Elijah, and Gerard Ribiere. "Note sur la convergence de méthodes de directions conjuguées." *Revue française d'informatique et de recherche opérationnelle. Série rouge*3.16 (1969): 35-43.
[3] Fletcher, Reeves, and Colin M. Reeves. "Function minimization by conjugate gradients." *The computer journal*7.2 (1964): 149-154.
[4] Wei, Zengxin, Shengwei Yao, and Liying Liu. "The convergence properties of some new conjugate gradient methods." *Applied Mathematics and computation* 183.2 (2006): 1341-1350.
[5] Shapiee, Norrlaili, Mohd Rivaie, and Mustafa Mamat. "A new classical conjugate gradient coefficient with exact line search." *AIP Conference Proceedings*. Vol. 1739. No. 1. AIP Publishing, 2016.
[6] Hestenes, Magnus Rudolph, and Eduard Stiefel. *Methods of conjugate gradients for solving linear systems*. Vol. 49. No. 1. Washington, DC: NBS, 1952.
[7] Fletcher, Roger. *Practical methods of optimization*. John Wiley & Sons, 2013.
[8] Dai, Yu-Hong, and Yaxiang Yuan. "A nonlinear conjugate gradient method with a strong global convergence property." *SIAM Journal on optimization* 10.1 (1999): 177-182.
[9] Zoutendijk, G. "Nonlinear programming, computational methods." *Integer and nonlinear programming* (1970): 37-86.
[10] Al-Baali, Mehiddin. "Descent property and global convergence of the Fletcher—Reeves method with inexact line search." *IMA Journal of Numerical Analysis* 5.1 (1985): 121-124.
[11] Touati-Ahmed, D., and C. Storey. "Efficient hybrid conjugate gradient techniques." *Journal of optimization theory and applications* 64.2 (1990): 379-397.
[12] Powell, Michael JD. "Nonconvex minimization calculations and the conjugate gradient method." *Numerical analysis*. Springer, Berlin, Heidelberg, 1984. 122-141.
[13] Gilbert, Jean Charles, and Jorge Nocedal. "Global convergence properties of conjugate gradient methods for optimization." *SIAM Journal on optimization* 2.1 (1992): 21-42.
[14] Powell, Michael James David. "Restart procedures for the conjugate gradient method." *Mathematical programming* 12.1 (1977): 241-254.
[15] Hager, William W., and Hongchao Zhang. "A new conjugate gradient method with guaranteed descent and an efficient line search." *SIAM Journal on optimization* 16.1 (2005): 170-192.
[16] Andrei, Neculai. "Accelerated conjugate gradient algorithm with finite difference Hessian/vector product approximation for unconstrained optimization." *Journal of Computational and Applied Mathematics* 230.2 (2009): 570-582.
[17] Wei, Zengxin, Shengwei Yao, and Liying Liu. "The convergence properties of some new conjugate gradient methods." *Applied Mathematics and computation* 183.2 (2006): 1341-1350.
[18] Abashar, Abdelrahman, et al. "Global convergence properties of a new class of conjugate gradient method for unconstrained optimization." (2017).
[19] Jusoh, I., Mamat, M., & Rivaie, M. (2014). A new edition of conjugate gradient methods for large-scale unconstrained optimization. International Journal of Mathematical Analysis, 8, 2277-2291.
[20] Rivaie, Mohd, et al. "The convergence properties of a new type of conjugate gradient methods." (2016).
[21] Rivaie, Mohd, et al. "A new class of nonlinear conjugate gradient coefficients with global convergence properties." *Applied Mathematics and Computation* 218.22 (2012): 11323-11332.
[22] Yuan, G., Lu, S., & Wei, Z. (2010). A line search algorithm for unconstrained optimization. Journal of Software Engineering and Applications, 3, 503-509.
[23] Andrei, Neculai. "An unconstrained optimization test functions collection." *Adv. Model. Optim* 10.1 (2008): 147-161.
[24] Dolan, Elizabeth D., and Jorge J. Moré. "Benchmarking optimization software with performance profiles." *Mathematical programming* 91.2 (2002): 201-213.