International Journal of Engineering & Technology, 7 (3.28) (2018) 84-88

International Journal of Engineering & Technology

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

Research paper

A modified conjugate gradient coefficient under exact line search for unconstrained optimization

Nurul 'Aini[1], Nurul Hajar[1], Mohd Rivaie[2], Mustafa Mamat[1*]

[1]*Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UNISZA), 21300 Kuala Terengganu, Malaysia*
[2]*Department of Computer Sciences and Mathematics, Universiti Teknologi MARA (UiTM), 21080 Kuala Terengganu, Terengganu, Malaysia*
**Corresponding author E-mail: must@unisza.edu.my*

Abstract

The conjugate gradient (CG) method is a well-known solver for large-scale unconstrained optimization problems. In this paper, a modified CG method based on AMR* and CD method is presented. The resulting algorithm for the new CG method is proved to be globally convergent under exact line search both under some mild conditions. Comparisons of numerical performance are made involving the new method and four other CG methods. The results show that the proposed method is more efficient.

Keywords: *Conjugate gradient method; unconstrained optimization; sufficient descent; global convergence; exact line search.*

1. Introduction

In this study, the unconstrained optimization function is formulated by

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

where $f : R^n \to R$ is a continuously differentiable function and its gradient at point x_k is denoted as g_k. The iterative method used to solve (1) is defined by

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

where $\alpha_k > 0$ is the stepsize, x_k is the k^{th} iterative point and d_k is the search direction. In this paper, the exact line search is used to calculate the stepsize which is given by

$$f(x_k + \alpha_k d_k) = \min_{\alpha \ge 0} f(x_k + \alpha d_k). \qquad (3)$$

The exact line search is also known as the optimal line search. As the name suggests, it calculates the optimal stepsize which gives the best possible reduction of the objective function. However, it is slow and can be ineffective when the initial point chosen is far from the solution point [7, 26]. As a result, most researchers prefer to use inexact line search [11, 6, 29] which is easier to implement and faster to converge to a solution. In recent years, new computers are equipped with faster processors thus successfully eliminating the speed inefficiency of exact line search as demonstrated by [12, 22-25]. This encourages more applications of this type of line search in unconstrained optimization [3, 17-19].
For CG method, the search direction is generally defined by

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

The scalar parameter β_k is the CG coefficient. To date, there are several choices of β_k, each showing varying results when applied on unconstrained optimization functions. Some examples of the well-known formulas of β_k are the Fletcher-Reeves (FR) [9], Polak-Ribière (PR) [20], Hestenes-Stiefel (HS) [13] and Conjugate Descent (CD) [10] methods. Their equations are as follow:

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

It is generally accepted that the CG method is very effective for solving optimization problems, especially when involving problems of large scale. Its usefulness makes it an interesting subject of study for many researchers. However, the efficiency of each CG method is highly dependent on the accuracy of the line search method used to determine the stepsize. In [30] established the global convergence of FR under exact line search. The same was proven by [1, 16], but with strong Wolfe line search. Computation wise, PR and HS are both superior to FR, though some counter example by [21] showed that these two methods can cycle infinitely without converging to a solution. Under exact line search, the CD method follow the same properties as the FR method [5].
In recent years, more variations of CG method have been developed. Some interesting examples are the RMIL [23] and WYL [27] methods, which are defined by

$$\beta_k^{RMIL} = \frac{g_k^T\left(g_k - g_{k-1}\right)}{d_{k-1}^T d_{k-1}},$$

$$\beta_k^{WYL} = \frac{g_k^T\left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{g_{k-1}^T g_{k-1}}.$$

The CG algorithm of WYL method has been proven to be globally convergent with exact line search and Grippo-Lucidi line search. The numerical results also show that it is competitive with PRP method. Under some conditions, in [14] proved that WYL method satisfies the sufficient descent condition under strong Wolfe line search. Later, several new CG methods based on WYL method are introduced. One of them is proposed by [2], referred as AMR* where the equation of the CG coefficient is written as

$$\beta_k^{AMR*} = \frac{g_k^T\left(m.g_k - g_{k-1}\right)}{m\left(g_{k-1}^T g_{k-1}\right)}, \qquad m = \frac{\|g_{k-1}\|}{\|g_k\|}.$$

Our main focus in this paper is to propose a new modification of CG method and establishes it global convergence under exact line search. The contents of this paper are divided into five sections with the introduction as the first one. The second section contains the new formula of β_k and its CG algorithm. In the third section, we analyze the convergence of the new CG method under exact line search. Next, we present some numerical results and discussion in Section 4 plus a short conclusion in Section 5.

2. New Conjugate gradient method

In this section, we present a new CG coefficient which we will refer as β_k^{ARM} or ARM method. It is an extension of AMR*, a CG coefficient introduced in [2] which had been shown to possess sufficient descent and global convergence properties. Based on the idea, we propose a new equation of β_k where we apply a new parameter for m_k and the denominator is similar to that of CD method. The equation of the new CG coefficient is as follows:

$$\beta_k^{ARM} = -\frac{m_k\|g_k\|^2 - \left|g_k^T g_{k-1}\right|}{m_k g_{k-1}^T d_{k-1}}, \quad \text{where} \quad m_k = \frac{\|d_{k-1}+g_k\|}{\|d_{k-1}\|} \tag{5}$$

Note that

$$\begin{aligned}\beta_k^{ARM} &= -\frac{\frac{\|d_{k-1}+g_k\|}{\|d_{k-1}\|}\|g_k\|^2 - \left|g_k^T g_{k-1}\right|}{\frac{\|d_{k-1}+g_k\|}{\|d_{k-1}\|}g_{k-1}^T d_{k-1}},\\ &\le \frac{\frac{\|d_{k-1}+g_k\|}{\|d_{k-1}\|}\|g_k\|^2}{\frac{\|d_{k-1}+g_k\|}{\|d_{k-1}\|}\left(-g_{k-1}^T d_{k-1}\right)},\\ &\le -\frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} = \beta_k^{CD}\end{aligned} \tag{6}$$

From above, it is shown that β_k^{ARM} is improved from β_k^{CD} in terms of its denominator. The following algorithm will implement the ARM method.

Step 1: Initialization. Given x_0, set $k = 0$.

Step 2: Compute β_k based on (5).

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

Step 4: Compute α_k based on (3).

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

Step 6: Convergence test and stopping criteria. If $f(x_{k+1}) < f(x_k)$ and $\|g_k\| \le 10^{-6}$, then stop. Otherwise, go to Step 1 with $k := k+1$.

3. Convergence analysis

In this section, we present the convergence analysis of ARM method under exact line search. A convergent algorithm has to satisfy the sufficient descent and global convergence properties. In order to establish the convergence of the new CG method, we need the following assumptions of the objective function.

Assumption 1

(i) $f(x)$ is bounded below on the level set $\ell = \left\{x \middle| f(x) \le f(x_0)\right\}$ where x_0 is the initial point.

(ii) In some neighbourhood N of ℓ, $f(x)$ is continuously differentiable and its gradient is Lipschitz continuous; then there exists a constant $L > 0$ such that $\|g(x) - g(y)\| \le L\|x - y\|$, for all $x, y \in N$.

3.1. Sufficient descent condition

If the search direction of an algorithm satisfies sufficient descent condition, then we may say the algorithm generates descent direction at every iteration. The sufficient descent condition is written as

$$g_k^T d_k \le -c\|g_k\|^2 \text{ for } k \ge 0 \text{ and } c > 0. \tag{7}$$

Theorem 3.1: *Suppose that Assumption 1 holds true. Consider a CG method where the search direction is given as (4) and the step size α_k is determined by using exact line search, then the sufficient descent condition (7) holds for all $k \ge 0$.*

Proof: If $k = 0$, then $g_0^T d_0 = -\|g_0\|^2$. Hence, condition (7) holds true. We also need to show that for $k \ge 1$, condition (7) will also hold true.
Multiplying (4) by g_k, we then have

$$\begin{aligned} g_k^T d_k &= g_k^T(-g_k + \beta_k d_{k-1}) \\ &= -\|g_k\|^2 + \beta_k g_k^T d_{k-1}. \end{aligned} \tag{8}$$

For exact line search, we know that $g_k^T d_{k-1} = 0$. Thus,

$$g_k^T d_k = -\|g_k\|^2,$$

which implies that d_k 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

We need to show that the CG algorithm with the new β_k is globally convergent under exact line search. From (6) and Theorem 3.1, the β_k can be simplified to

$$\beta_k^{ARM} \le \beta_k^{CD} = \beta_k^{FR}. \quad (9)$$

Lemma 3.1: *Suppose that Assumption 1 holds true, consider any d_k of the form (4) and α_k obtained under exact minimization rule for all $k \ge 0$. Then, the following Zoutendijk condition holds where*

$$\sum_{k\ge 1} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \quad (10)$$

Proof: The proof can be referred in [30].

By using Lemma 3.1, we establish the following theorem to prove the global convergence of the new CG method with exact line search.

Theorem 3.2: *Suppose that Assumption 1 and Theorem 3.1 hold true. Consider the CG method in the form of (2) and (4) with (5) as the β_k in addition to the stepsize α_k obtained by exact line search. Then,*

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \quad (11)$$

Proof: The proof is done by contradiction. Now, assuming that Theorem 3.2 is not true, then there exists a constant $\delta > 0$ such that

$$\|g_k\| \ge \delta. \quad (12)$$

From (4), by squaring both sides of the equation, we have

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

Note that for exact line search $g_k^T d_{k-1} = 0$. Hence

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

Substituting (9) into the equation, we have

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

Now, we divide both sides by $\|g_k\|^4$ which then gives us

$$\frac{\|d_k\|^2}{\|g_k\|^4} \le \frac{1}{\|g_k\|^2} + \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4}.$$

By using recursive formula and noting that $d_0 = g_0$, we have

$$\frac{\|d_k\|^2}{\|g_k\|^4} \le \sum_{i=0}^{k} \frac{1}{\|g_i\|^2},$$

$$\frac{\|g_k\|^4}{\|d_k\|^2} \ge \frac{\delta^2}{k}.$$

Hence,

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} \ge \infty.$$

The last inequality contradicts the Zoutendijk condition in Lemma 3.1. The proof is thus completed. ■

4. Results and discussion

This section discusses the numerical tests of our new β_k and the results obtained. The same tests are also applied on some selected CG methods (AMR*, WYL, HS and CD) in order to compare their performance with our method. We pick out 17 standard test problems from [4, 15, 28]. The capabilities of each method under exact line search are measured in terms of number of iterations and CPU time taken to solve the test problems. We set the algorithm to stop when $\|g_k\| \le 10^{-6}$ or when the number of iterations exceed 10000. The codes are written in MATLAB 2015 subroutine programme and all the tests are performed by using a laptop with CPU processor Intel(R) Core(TM) i5 and 8GB RAM memory. The lists of functions used are displayed in Table 1.

Table 1: A list of problem functions.

No.	Function	Number of Variables	Initial Points
1	Six Hump	2	(3,3), (13,13), (37,37)
2	Three Hump	2	(-2,-2), (18,-18), (57,57)
3	Zettl	2	(6,6), (14,14), (64,64)
4	Colville	4	(4.4,…,4.4), (24,…,24), (71,…,71)
5	Dixon and Price	2,4	(12,12,…,12), (23,23,…,23), (69,69,…,69)
6	Hager	2,4	(6,…,6), (12,…,12), (19.5,19.9, 19.5,19.9)
7	Raydan 1	2,4	(7,…,7), (12,…,12), (22,…,22)
8	Raydan 2	2,4	(6,…,6), (11,…,11), (18,…,18)
9	Powell	4,8	(3.5,…,3.5), (15,…,15), (40,…,40)
10	Extended White and Holst	2, 4, 10, 100, 500, 1000	(-1,-1.5,…,-1,-1.5), (5,6,…,5,6), (11.2,11,…,11.2,11)
11	Extended Rosenbrock	2, 4, 10, 100, 500, 1000	(-10,…,-10), (18,…,18), (68,…,68)
12	Shallow	2, 4, 10, 100, 500, 1000	(11,…,11), (23,…,23), (80.5)
13	Extended Strait	2, 4, 10, 100, 500, 1000	(4,…,4), (11,…,11), (38,…,38)
14	Extended Himmelblau	2, 4, 10, 100, 500, 1000	(17.8,…,17.8), (40,…,40), (115,106,…,115,106)
15	DENSCHNB	2, 4, 10, 100,	(5,…,5), (25,…,25),

		500, 1000	(225,…,225)
16	Generalized Quartic	2, 4, 10, 100, 500, 1000	(11,…,11), (28,…,28), (87,-80,…,87,-80)
17	Extended Tridiagonal 1	2, 4, 10, 100, 500, 1000	(13,…,13), (24.7,…,24.7), (60,…,60)

The results are compiled in two graphs of performance profile based on [8] to show the improvements of the ARM method in comparison to some other CG methods. In the performance profile, we plot the fraction ρ of problems for which the method is within a factor τ of the least iterations or the best time. From the left side of the figure, we can determine the percentage of the test problems for which a method is the fastest. While, the right side indicates the percentage of the test problems that are successfully solved by each solver. Generally, the method that has the highest value of $\rho(\tau)$ and is located at the top right of figure is preferable.

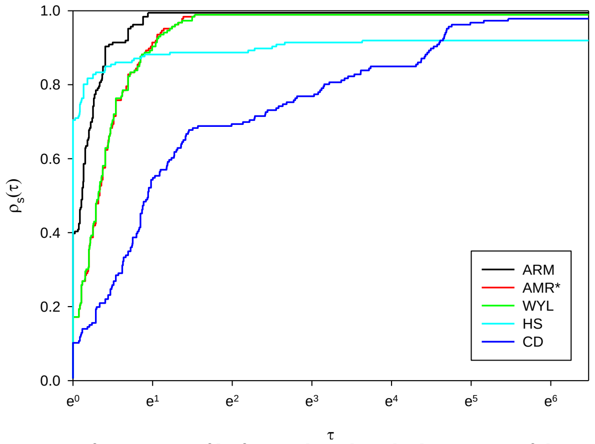

Fig. 1: Performance profile for CG-based methods in terms of the number of iterations

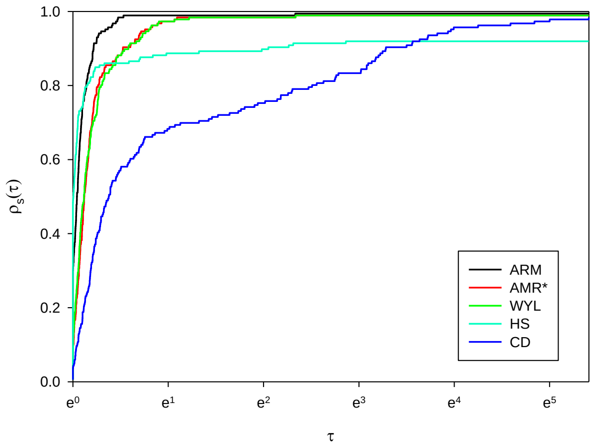

Fig. 2: Performance profile for CG-based methods in terms of the CPU time

Fig. 1 and 2 show the performance of the considered CG methods based on number of iterations and CPU time respectively. From both figures, ARM solves 99.46% of the test problems. On the other hand, AMR* and WYL methods equally solves 98.92% of the problems, CD method solves 98.39%, while HS method only manage to reach 91.93% despite being the fastest solver for most of the tests. Overall, ARM have the highest number of successful tests. It can also be regarded as the fastest solver after HS. This is based on the location of its curve at the left side of Fig. 1 and 2 which are both higher than the other solvers except HS. Thus, we can say that ARM is the most efficient method among all the tested solvers.

5. Conclusion

In this paper, we have proposed a new CG coefficient based on the AMR* and CD method. The search direction of the proposed method satisfies the sufficient descent condition when used with exact line search. Convergence analysis on the new algorithm also shows that it is globally convergent. In addition, the effectiveness of the ARM method has been investigated via numerical tests on a set of 17 unconstrained optimization test problems. Analyses on the numerical outcomes show that the new algorithm perform better than some CG methods both in number of iterations and time taken for solving each test functions. Based on these results, we conclude that ARM is a good alternative for solving unconstrained optimization functions.

Acknowledgement

The authors would like to thank The Ministry of Higher Education of Malaysia (MOHE) for the funding of this research under MyPhD KPM.

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] Alhawarat A, Mamat M, Rivaie M & Mohd I (2014), A new modification of nonlinear conjugate gradient coefficients with global convergence properties, International Journal of Mathematical, *Computational, Statistical, Natural and Physical Engineering* 8, 54-60.
[3] Abidin, N.Z., Mamat, M., Dangerfield, B., Zulkefli, J.H., Baten, Md.A., Wibowo, A (2014) Combating obesity through healthy eating behavior: A call for system dynamics optimization, Plos One, 9(12), 0114135.
[4] Andrei N (2008), An unconstrained optimization test functions collection, *Advanced Modelling and Optimization* 10, 147-161.
[5] Andrei N (2011), 40 conjugate gradients algorithms for unconstrained optimization, *Bull. Malays. Math. Sci. Soc.* 34, 319–330.
[6] Armijo L (1966), Minimization of functions having Lipschitz continuous partial derivatives, *Pacific J. Mathematics* 16, 1-3.
[7] Chong EKP & Żak SH (2002), An introduction to optimization, *John Wiley and Sons*, New York.
[8] Dolan E & More JJ (2002), Benchmarking optimization software with performance profile, *Maths. Prog*. 91, 201-213.
[9] Fletcher R & Reeves C (1964), Function minimization by conjugate gradients, *Comput. J.* 7, 149-154.
[10] Fletcher R (1987), Practical method of optimization, Unconstrained optimization, *John Wiley and Sons*, New York.
[11] Goldstein AA (1965), On steepest descent, *SIAM J. Control.* 3, 147-151.
[12] Hajar, N., Mamat, M., Rivaie, M., Jusoh, I (2016). A new type of descent conjugate gradient method with exact line search, *AIP Conference Proceedings*, 1739, 020089
[13] Hestenes MR & Steifel E (1952), Method of conjugate gradient for solving linear equations, *J. Res. Nat. Bur. Stand*. 49, 409-436.
[14] Huang H, Wei Z & Yao S (2007), The proof of the sufficient descent condition of the Wei-Yao-Liu conjugate gradient method under the strong Wolfe-Powell line search, *Applied Mathematics and Computation* 189, 1241-1245.
[15] Jamil M & Yang XS (2013), A literature survey of benchmark functions for global optimisation problems, *International Journal of Mathematical Modelling and Numerical Optimisation* 4, 150–194.
[16] Liu GH, Han JY & Yin HX (1995), Global convergence of the Fletcher-Reeves algorithm with inexact line search, *Applied Mathematics*, 75-82.
[17] Mamat, M., Subiyanto, Kartono, A (2013). Mathematical model of cancer treatments using immunotheraphy, chemotheraphy and biochemotheraphy, *Applied Mathematical Sciences*, 7(5-8), 247-261
[18] Mamat, M., Deraman, S.K., Noor, N.M.M., Rokhayati, Y (2012, Diet problem and nutrient requirement using fuzzy linear programming approach, *Asian Journal of Applied Sciences*, 5(1), 52-59
[19] Mohamed, N.S., Mamat, M., Mohamad, F.S.., Rivaie, M.M. (2016), A new coefficient of conjugate gradient methods for nonlinear unconstrained optimization, *Jurnal Teknologi*, 78(6-4), 131-136.
[20] Polak E & Ribiere G (1969), Note sur la convergence de directions conjugees, *Rev. Francaise Informat Recherche Operationelle* 3, 35-43.

[21] Powell MJD (1984), Nonconvex minimization calculations and the conjugate gradient method, Lecture Notes in Mathematics, 1066, *Springer-Verlag*, Berlin, 122-141

[22] Rivaie M, Mamat M & Abashar A (2015), A new class of nonlinear conjugate gradient coefficients with exact and inexact line searches, *Applied Mathematics and Computation* 268, 1152-1163.

[23] Rivaie M, Mamat M, June LW & Mohd I (2012), A new class of nonlinear conjugate gradient coefficient with global convergence properties, *Applied Mathematics and Computation* 218, 11323-11332.

[24] Shapiee, N., Rivaie,M., Mamat, M (2016). A new classical conjugate gradient coefficient with exact line search, *AIP Conference Proceedings*, 1739, 020082.

[25] Shoid, S., Rivaie,M., Mamat, M (2016). A modification of classical conjugate gradient method using strong Wolfe line search, *AIP Conference Proceedings*, 1739, art.no. 020071.

[26] Sun W & Yuan YX (2006), Optimization theory and methods: Nonlinear programming, *Springer Science and Business Media*, New York.

[27] Wei Z, Yao S & Liu L (2006), The convergence properties of some new conjugate gradient methods, *App-l. Math. Comput*. 183, 1341-1350.

[28] Witte BF & Holst WR (1964), Two new direct minimum search procedures for functions of several variables, *Proceedings of the* ACM *Spring Joint Computer Conference, pp.* 195-209.

[29] Wolfe P (1969), Convergence conditions for ascent method, *SIAM Rev.* 11, 226-235.

[30] Zoutendijk G (1970), Nonlinear programming computational methods, in Abadie J. (Ed.) *Integer and nonlinear programming*, Amsterdam.