International Journal of Engineering & Technology, 7 (3.28) (2018) 354-357

International Journal of Engineering & Technology

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

Research paper

A New Hybrid of Conjugate Gradient Method with Descent Properties

Syazni Shoid[1], Norhaslinda Zull[1], Mohd Rivaie[2], Mustafa Mamat[1]*, Puspa Liza Ghazali[3], Mohamad Afendee Mohamed[1]

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

Abstract

Many researchers are interested in developing and improving the conjugate gradient (CG) method because of its convergence properties and efficiency in solving large-scale problems. This work introduces new CG coefficient (β_k) will be presented in such a way to improve the performance of the previous CG methods. The new method is the hybrid between HS and SYRM methods. This method always produces a descent search direction at each iteration. The preliminary numerical comparisons with some others CG methods have shown that this new method is efficient in solving all given problems under Strong Wolfe Powell (SWP) line search condition.

Keywords: *Conjugate gradient method; descent direction; sufficient descent condition; Strong Wolfe Powell line search; hybrid method.*

1. Introduction

In optimization, Conjugate Gradient (CG) is the useful iterative method in finding the minimum function of n variables due to their Q-conjugacy direction. Our problem can be written as,

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

where the objective function, $f: R^n \to R$ is continuously differentiable function. The decision variable $x \in R^n$ while the constraint set $X \in R^n$. In case of $X = R^n$, this problem is categorized as unconstrained optimization.
When applied to solve the problem, this method generates the sequence of vector x_k and the $x_0 \in R^n$ become an initial point by the iteration,

$$x_{k+1} = x_k + \alpha_k d_k \tag{1}$$

where the $\alpha_k > 0$, d_k are the stepsize and search direction, respectively. We can use any line search method to find the α_k. However, most line search methods in used in practice is that of inexact category. The value of α_k is estimated such a way that it gives sufficient decrease in the objective function and no longer searching in the direction when x_k is far from the solution. Thus, it will give inexpensive and low computational cost. By and large, strong Wolfe-Powell (SWP) line search is very popular, introduced by Wolfe [9] that introduced two conditions which are given as follows,

$$f(x_k + \alpha_k d_k) \le f(x_k) + \mu \alpha_k g_k^T d_k \tag{2}$$

$$\left| g_{k+1}^T d_k \right| \le -\sigma g_k^T d_k \tag{3}$$

where $0 < \mu < \sigma < 1$.
The d_k assumed to be a descent direction at each iterations and generated as

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

From equation (4), we get $g_k = \nabla f(x_k)$ with β_k known as the coefficient. There are several types of CG methods such as classical, modification and hybrid methods. Some well-known classical CG methods were introduced by Fletcher and Reeves (FR) [1], Polak, Ribiere and Polyak (PRP) [2], Hestenes and Stiefel (HS) [3] and Rivaie et al. (RMIL) [4], to name a few.

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

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}} \qquad \beta_k^{RMIL} = \frac{g_k^T (g_k - g_{k-1})}{d_k^T (d_k - g_{k+1})}$$

where $\|.\|$ denotes the Euclidian norm of vectors. Many researchers attempt to improve the classical CG methods by making some modifications on them. For examples, Wei et al. (WYL) [5] proposed a new CG formula variant of the PRP method, Sui Qidi [7] makes some modification on Zhang's method, Abashar [6, 22] and Rivaie et al. (RAMI) [8] has modified RMIL and proposed AMRI and RAMI methods respectively. All formulas mention is shown below.

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

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

$$\beta_k^{SUIQIDI} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} \left|g_k^T g_{k-1}\right|}{\left|g_k^T d_{k-1}\right| + \|g_k\|^2},$$

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

The hybrid CG method is formed by combining two or more CG methods, usually from classical CG-type. Touati-Ahmed and Storey (TS) [12], Salih et al. [13] and Alhawarat et al. (PRP**) [14] are the examples of hybrid CG.

$$\beta_k^{TS} = \begin{cases} \beta_k^{PRP} = \dfrac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} & \text{if} \quad 0 \le \beta_k^{PRP} \le \beta_k^{FR} \\ \beta_k^{FR} = \dfrac{\|g_k\|^2}{\|g_{k-1}\|^2} & \text{, otherwise} \end{cases}$$

$$\beta_k^{PRP**} = \begin{cases} \beta_k^{PRP} = \dfrac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} & \text{if} \quad \|g_{k-1}\|^2 \ge \left|g_k^T g_{k-1}\right| \\ \beta_k^{NPRP} = \dfrac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|}\left|g_k^T g_{k-1}\right|}{\|g_{k-1}\|^2} & \text{otherwise} \end{cases}$$

There are many others researches on CG methods that can be refer in [17-21, 23]. This paper is structured as the following. In the second section, we present a new types of CG coefficient. The third section exposes the idea of sufficient descent condition, followed by fourth section which layout some experimental results corresponding to this newly proposed β_k . Finally, we dicuss the result and conclude the findings in fifth and sixth section respectively.

2. New type of conjugate gradient coefficient

The main contribution of this paper is the introduction of a new β_k formula namely as β_k^{SRM} (SRM). It is categorized as hybrid CG method obtained by combination of HS and SYRM [16]. The performances of hybrid CG method is better than that of classical CG method, since hybrid method can exploit the attractive features of the original methods used to form the new parameter. The following formula shows the SRM method,

$$\beta_k^{SRM} = \begin{cases} \beta_k^{HS} = \dfrac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}}, & g_k^T g_k > \left|g_k^T g_{k-1}\right| \\ \beta_k^{SYRM} = \dfrac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_k - g_{k-1}\|}\left|g_k^T g_{k-1}\right|}{d_{k-1}^T (d_{k-1} - g_k)}, & else \end{cases} \tag{5}$$

There are cases where HS and SYRM fail to solve the unconstrained optimization function. Thus, the contribution of the SRM method is to improve the performance of SYRM and HS method by combining the good features of both methods.
Any CG algorithm has the general structure as illustrated below.

Algorithm 1: Conjugate Gradient algorithm
S1: Let $k = 0$, set an initial point x_0
S2: Calculate β_k from (5).
S3: Calculate search direction from (4). If $g_k = 0$, stop .
S4: Calculate α_k using inexact line search from (2) and (3).
S5: Recalculate the new point. Let $x_{k+1} = x_k + \alpha_k d_k$
S6: Convergent test and stopping criteria.
If $f(x_{k+1}) < f(x_k)$ and $\|g_k\| < \varepsilon$, then stop.
Otherwise, let $k = k+1$, jump to S1.

3. Convergence analysis

This section examines the convergent properties of SRM. For all $k \ge 0$, if there exists a constant $C > 0$ then, the search directions is said to satisfy sufficient descent condition.

$$g_k^T d_k \le -C\|g_k\|^2 \tag{6}$$

The theorem below shows that SRM possess the sufficient decent condition.

Theorem 1
Suppose that the sequence g_k and d_k are generated by Algorithm 1, β^{SRM} is given as (5) and the step length is determined by the SWP line search (2) and (3). If $\sigma < \frac{1}{5}$, then the sequence d_k satisfies the sufficient condition (6).

Proof
The proof of Theorem 1 can be divided in two cases as follow:

Case 1: $\beta_k^{SRM} = \beta^{HS}$

Simplify the formula HS in (5), to obtain

$$\beta_{k+1}^{HS} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{(g_{k+1} - g_k)^T d_k} \le \frac{\|g_{k+1}\|^2}{d_k^T g_{k+1} - d_k^T g_k} \tag{7}$$

From (4), multiply it by g_{k+1}^T and divide by $\|g_{k+1}\|^2$, then

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} = -1 + \beta_{k+1}^{HS} \frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2}. \tag{8}$$

The descent property of d_k is proved by induction. When $k=0$, $g_0^T d_0 = -\|g_0\|^2 < 0$ if $g_0 \neq 0$. Now suppose that $d_i, i=1,2,3,..., k$, are all descent direction. From line search conditions, it follows that $d_k^T y_k \geq (\sigma-1) d_k^T g_k > 0$. Applying inequality (3) and (7), then

$$\beta_{k+1}^{HS} \left| g_{k+1}^T d_k \right| \leq -\sigma \left(g_k^T d_k \right) \frac{\|g_{k+1}\|^2}{g_k^T d_k (-1-\sigma)}, \quad (9)$$

Apply (8) into (9) to get,

$$-1 + \frac{\sigma}{(1-\sigma)} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -1 + \frac{\sigma}{(1+\sigma)}. \quad (10)$$

Thus, by induction, $g_k^T d_k < 0$ holds for all $k \geq 0$. Therefore, the sufficient descent condition (6) holds. The proof is complete.

Case 2: $\beta_k^{SRM} = \beta^{SYRM}$

Firstly, simplify the formula SYRM in (5), to obtain

$$\beta_{k+1}^{SYRM} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_{k-1} - g_k\|} \left| g_{k+1}^T g_k \right|}{d_k^T (d_k - g_{k+1})} \leq \frac{\|g_{k+1}\|^2}{\|d_k\|^2} \quad (11)$$

Multiply (4) by g_{k+1}^T and divide by $\|g_{k+1}\|^2$, then

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} = -1 + \beta_{k+1}^{SYRM} \frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2} \quad (12)$$

Proof by induction is used for the descent property of d_k. At $k=0$, $g_0^T d_0 = -\|g_0\|^2 < 0$ if $g_0 \neq 0$. Now, suppose that $d_i, i=1,2,3,..., k$, are all descent direction. Using inequalities (3) and (7), then

$$\beta_{k+1}^{SYRM} \left| g_{k+1}^T d_k \right| \leq -\sigma (g_k^T d_k) \frac{\|g_{k+1}\|^2}{\|d_k\|^2} \quad (13)$$

From (12), assume that

$$-1 + \sigma \frac{g_k^T d_k}{\|d_k\|^2} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -1 - \sigma \frac{g_k^T d_k}{\|d_k\|^2} \quad (14)$$

Repeating this process and taking in the fact that $g_0^T d_0 = -\|g_0\|^2$ yields,

$$-\sum_{i=0}^{k} \sigma^i \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -2 + \sum_{i=0}^{k} \sigma^i. \quad (15)$$

Note that,

$$\sum_{i=0}^{k} [\sigma]^i < \sum_{i=0}^{\infty} [\sigma]^i = \frac{1}{1-\sigma}. \quad (16)$$

Then, (15) can be written as

$$-\frac{1}{1-\sigma} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -2 + \frac{1}{1-\sigma}. \quad (17)$$

Thus, $g_k^T d_k < 0$ hold for all $k \geq 0$. Let $C = 2 - \frac{1}{1-\sigma}$, where $C \in (0,1)$, then

$$C - 2 \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -C. \quad (18)$$

Therefore, the sufficient descent condition (6) holds and thus complete the proof.

4. Results and discussion

This section deals with the analysis of the newly proposed method which is SRM. We only test it under SWP line search and compare the performance with RAMI, RMIL, SYRM and HS methods. For each method, we examine two parameters namely the iteration number and the processor time needed for the method to achieve minimal value of function. Elevent different standard test problems with some ranges number of variables were used. As suggested by Hilstrom [15], we selected initial point that is not very near to the solution point. As such, random number generator was used for this purpose. However, we believe that this way it will require more complex coding which in turn may increase the CPU time. Therefore, we just have chosen four initial points from a point that is close to the solution point to the point that is distant away. Besides that, the different initial points serve as a mean to evaluate the convergence properties of the proposed method and its robustness.

We use $\varepsilon = 10^{-6}$ and the gradient value as stopping criteria. Matlab 12 subroutine program was used for solving all the problem functions. The list of problems functions are based on Andrei [10] and can be found in Table 1.

Table 1: List of Problems Functions

No.	Functions	Variables	Initial Points
1	Six hump	2	(1,…,1),(5,…,5),(10,…,10),(15,…,15)
2	Three hump	2	(10,…,10),(50,…,50),(100,…,100),(150,…,150)
3	Booth	2	(10,…,10),(50,…,50),(100,…,100),(150,…,150)
4	Extended Quadratic Penalty QP2	2,100,500,1000,5000	(1,…,1),(5,…,5),(10,…,10),(15,…,15)
5	Diagonal 2	2,100,500,1000,5000	(1,…,1),(0.5,…,0.5),(0.25,…,0.25),(0.2,…,0.2)
6	Extended White & Holst	2,100,500,1000,5000	(30,…,30),(35,…,35),(40,…,40),(73,…,73)
7	Extended Rosenbrock	2,100,500,1000,5000	(13,...,13),(16,...,16),(20,...,20),(30,...,30)
8	Extended Beale	2,100,500,1000,5000	(-1,…,-1),(2,…,2),(19,…,19),(24,…,24)
9	Extended Himmelblau	2,100,500,1000,5000	(13,...,13),(30,...,30),(100,...,100),(180,...,180)
10	Extended DENSCHNB	2,100,500,1000,5000	(5,…,5),(8,…,8),(12,…,12),(30,…,30)
11	Shallow	2,100,500,1000,5000	(10,…,10),(25,…,25),(50,…,50),(100,…,100)

The performance result is shown in Fig. 1 and 2, measured based on two different parameters, the number of iteration and the CPU time respectively. Dolan and More [11] were the author whom the concept of performance profile indebted to.

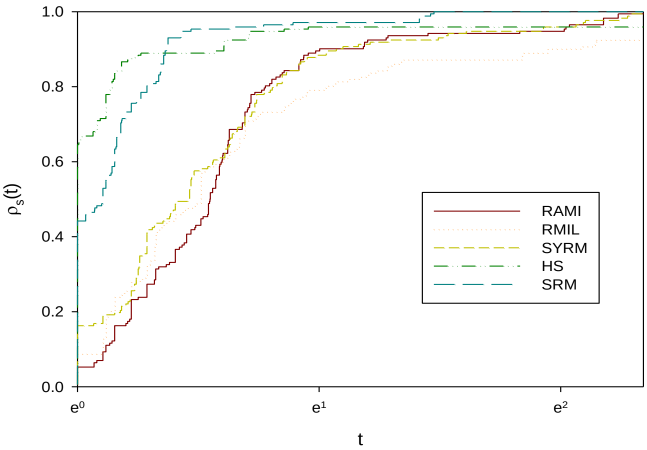

Fig. 1: Performance profile based on number of iteration

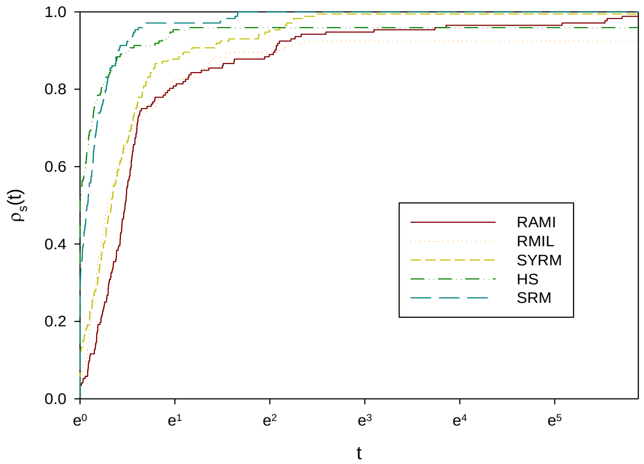

Fig. 2: Performance profile based on CPU time

5. Discussion

On the left side of Fig. 1, the curve located at the top of the figure represents the methods that solve the problems with the least number of iteration. It appears that the top left of the figure when $\tau = e^0$ is dominated by HS and followed by SRM methods, indicating that they are both efficient in executing the test. In the left side of Fig. 2, the top curve in the figures signifies the method that solves all the test problems with most minimum CPU time. So that, it is clear that the top curves are also the SRM and HS methods.

The right side of both Fig.1 and Fig. 2 represents the total test problems that have been successfully solved by each method. The higher the line plot, the more test problems is solved by the algorithms. In both figures, only the SRM method achieved $\rho_s(1)$.

6. Conclusion

In this paper, our new β_k named as SRM has been presented. Based on the result, SRM is the second efficient method after HS. However, when it comes to robustness, SRM is the most robust method, which makes it a better choice than HS. The best method to consider is the one with the high value of $\rho_s(\tau)$. Therefore, as the final conclusion, SRM is considered to be the best method.

Acknowledgement

This research is partially funded by the Malaysian Ministry of Higher Education via MyPhD sponsorship.

References

[1] Fletcher R & Reeves C (1964), Function minimization by conjugate gradients, Comput. J., 7, 149-154.
[2] Polak E & Ribiere G (1969), Note sur la convergence de directions conjugees, Rev.Francaise Informat Recherche Operationelle, 3E Annee, 16, 35-43.
[3] Hestenes MR & Steifel E (1952), Method of conjugate gradient for solving linear equations, J. Res. Nat. Bur. Stand., 49, 409-436.
[4] Rivaie M, Mamat M, Leong WJ & Ismail M (2012), A new conjugate gradient coefficient for large scale nonlinear unconstrained optimization, Int. Journal of Math. Analysis, 6(23), 1131-1146.
[5] Wei Z X, Yao S & Liu L (2006), The convergence properties of some new conjugate gradient methods, Applied Mathematics and Computation,183, 1341-1350.
[6] Abashar A, Mamat M, Rivaie M, Mohd I & Omer O (2014), The proof of sufficient descent condition for a new type of conjugate gradient methods, AIP Conf. Proc., 1602(1), 296-303.
[7] Sui Q (2011), An improved conjugate gradient method and it's global convergence, Applied Mathematical Sciences, 5(50), 2483-2490.
[8] Rivaie M, Abashar A, Mamat M & Mohd I (2014), The convergence properties of a new type of conjugate gradient methods, Applied Mathematical Sciences, 8, 1, 33-44.
[9] Wolfe P (1971), Convergence conditions for ascent method. II: some corrections, SIAM Rev., 13,2, 185-188.
[10] Andrei N (2008), An unconstrained optimization test function collection, J. Adv. Modeling and Optimization, 10, 147-161.
[11] Dolan E & More JJ (2002), Benchmarking optimization software with performance profile, Math. Prog., 91(2), 201-213.
[12] Touati-Ahmed D & Storey C (1990), Efficient hybrid conjugate gradient techniques, J. Optim. Theory Appl., 64(2), 379-397.
[13] Salih Y, Hamoda MA & Rivaie M (2018), New hybrid conjugate gradient method with global convergence properties for unconstrained optimization, Malaysian Journal of Computing and Applied Mathematics, 1(1), 29-38.
[14] Alhawarat A, Mamat M, Rivaie M & Salleh Z (2015), An efficient hybrid conjugate gradient method with the strong wolfe-powell line search, Mathematical Problems in Engineering, 2015, 1-7.
[15] Hilstrom KE (1977), A simulation test approach to the evaluation of nonlinear optimization algorithms, ACM Trans. Maths. Softw, 3, 305-315.
[16] Shoid S, Rivaie M & Mamat M (2016), A modification of classical conjugate gradient method using strong wolfe line search, AIP Conf. Proc., 1739, 1-8.
[17] Shapiee N, Rivaie M & Mamat M (2016), A new classical conjugate coefficient with exact line search, AIP Conf. Proc., 1739, 1-8.
[18] Ghani ANH, Rivaie M & Mamat M (2016), A modified form of conjugate gradient method for unconstrained optimization problems, AIP Conf. Proc., 1739, 1-8.
[19] Hajar N, Mamat M, Rivaie M & Jusuh I (2016), A new of descent conjugate gradient method with exact line search, AIP Conf. Proc., 1739, 1-8.
[20] Khadijah W, Rivaie M, Mamat M & Jusuh I (2016), A spectral KRMI conjugate gradient method under the strong-Wolfe line search, AIP Conf. Proc., 1739, 1-8
[21] Abidin ZZ, Mamat M & Rivaie M (2016), A new steepest descent method with global convergence properties, AIP Conf. Proc., 1739, 1-8.
[22] Abashar A, Mamat M, Rivaie M, Mohd I & Omer O (2014), Global convergence properties of a new class of conjugate gradient method for unconstrained optimization, Applied Mathematical Sciences, 8(67), 3307-3319.
[23] Abidin NZ, Mamat M, Dangerfield B, Zulkepli JH, Baten MA & Wibowo A (2014), Combating obesity through healthy eating behavior: A call for system dynamics optimization, Plos One, 9(12), 1-17.