HYBRID CONJUGATE GRADIENT METHOD UNDER EXACT AND INEXACT LINE SEARCH

[1]NUR SYARAFINA MOHAMED, [2]MUSTAFA MAMAT, [3]MOHD RIVAIE, [4]SHAZLYN MILLEANA SHAHARUDDIN

[1]Universiti Kuala Lumpur, Malaysian Institute of Industrial Technology, Johor Bharu, Malaysia(Authora)
[2]Faculty of Informatics and Computing Universiti Sultan Zainal Abidin (Unisza), Campus Tembila Besut, Terengganu, Malaysia(Authorb)
[3]Department of Computer Sciences and Mathematics, Universiti Teknologi MARA (UiTM) Terengganu(Authorc)
[4]Department of Mathematics, Faculty of Science and Mathematics, Universiti Pendidikan Sultan Idris, Malaysia(Authord)

Abstract - Hybridization is one of the frequent approaches in modifying the Conjugate Gradient (CG) method. In this paper, a new hybrid CG is introduced and its performance is compared under two different line searches; exact and inexact line searches. The proposed hybrid CG is known as HSMR, is evaluated as a convex combination of RMIL and SMR methods. The comparisons was tested with twenty one different test functions under various dimensions. Both results are compared between number of iterations and total number of CPU time. The comparison performance showed that the hybrid CG is promising under exact line search compared to inexact line search in both of number of iterations and central CPU time.

Keywords - (Conjugate Gradient Method; Exact Line Search; Hybrid Conjugate Gradient Method)

I. INTRODUCTION

Consider the objective function of an unconstrained optimization;

$$\min\left\{f(x): x \in R^n\right\} \tag{1}$$

where $f: R^n \to R$ is a continuously differentiable function which is bounded from below. R^n is denoted as $n-$ dimensional Euclidean space [1]. With any $x_0 \in R^n$ as an initial guess, generally, a sequence of $\{x_k\}$ is generated by employing the CG iterative method given by;

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

where x_k is the k^{th} iterative point and $\alpha_k > 0$ is a step size while the search direction d_k is defined by;

$$d_k = \begin{cases} -g_k & \text{if} \quad k = 0 \\ -g_k + \beta_k d_{k-1} & \text{if} \quad k \geq 1 \end{cases}. \tag{3}$$

The scalar β_k is called the CG coefficient and g_k is the gradient of f at point x_k. Step size $\alpha_k > 0$ is the stepsize determined either by using exact line search and inexact line search, given as;

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha d_k). \tag{4}$$

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k,$$

$$\left| g(x_k + \alpha_k d_k)^T \leq \sigma \left| g_k^T d_k \right| \right| \tag{5}$$

In CG method, the step size, α_k which obtained by certain line search method gives influence in possessing the global convergence properties. In this paper, the inexact line search used is called as a strong Wolfe-Powell (SWF), where $\alpha_k > 0$ satisfies (5)with $0 < \delta < \sigma < 1$ are both constants. Due to the well-known ability to provide the optimal step size, exact line search is used [2]. Recently, studies have shown that the newer technologies with faster processors and better equipment have successfully eliminated the speed problems often suffered by exact line search as demonstrated by Rivaie et al [1]. This motivates plenty of its applications for solving unconstrained optimization problems. From [3-4], a classical CG method introduced were;

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

$$\beta_k^{SMR} = \max\left\{0, \frac{\|g_k\|^2 - \left|g_k^T g_{k-1}\right|}{\|d_{k-1}\|^2}\right\} \tag{7}$$

The g_k and g_{k-1} are the abbreviations of $g(x_k)$ and $g(x_{k-1})$. They are the gradients of $f(x)$ at points x_k and x_{k-1} respectively. Euclidean norm of the vectors is denoted by $\|\|$. The above corresponding methods are known as RMIL(Rivaie-Mustafa-Ismail-Leong) [3] and SMR (Syarafina-Mustafa-Rivaie) [4]. Different CG method yiled different performance of CG algorithm due to the different choices for calculating the CG coefficients. Meanwhile, the hybrid CG proposed in this paper is compared to some well-known hybrid CG proposed by the previous researchers, which are;

$$\beta_k^{HDY} = \max\{0, \min\{\beta_k^{HS}, \beta_k^{DY}\}\} \tag{8}$$

$$\beta_k^{HHUS} = \max\{0, \min\{\beta_k^{PRP}, \beta_k^{FR}\}\} \tag{9}$$

$$\beta_k^{HLSCD} = \max\{0, \min\{\beta_k^{LS}, \beta_k^{CD}\}\} \tag{10}$$

$$\beta_k^{HJHJ} = \frac{\|g_k\|^2 - \max\{0, \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}\}}{\max\{\|g_{k-1}\|^2, d_{k-1}^T (g_k - g_{k-1})\}}$$

(11)

From (8)-(11) HDY is a modification effort from Dai and Yuan [5] combining its algorithm with Hestenes and Steifel [6], HHUS was introduced by Hu and Storey [7], HLSCD is the combination of LS [8] and CD [9] and HJHJ is a hybrid CG from Jinbao-Han-Jiang [10]. In this paper, the main objective is to compare the performance of hybrid CG methods under exact and inexact line searches. The idea is to test different hybrid methods on different test functions and the comparisons are on their number of iterations and central processing per unit (CPU) time. In this paper, section one present the overview of the study discussing all important fundamental used for the research. Section two discusses the new hybrid CG method proposed with its algorithm and motivations. Then, the next section elaborates the proving process of the hybrid CG proposed in order to show the global convergence properties criterion. Finally, the last two sections analyses about the results and conclusion for this study.

II. HSMR METHOD

HSMR method is a result of combining SMR and RMIL.Integrating SMR which has good computational properties and RMIL which has strong convergence properties [1-2] give HSMR better performance compared to the original classical CG under both exact and inexact line searches. The combinations of all the good criteria of SMR and RMIL are used in order to obtain a better practical algorithm both in numerical and convergence analysis. The HSMR method is known as;

$$\beta_k^{HSMR} = \max\{0, \min\{\beta_k^{SMR}, \beta_k^{RMIL}\}\} \qquad (12)$$

The idea was initiated by using the restart strategy proposed in [7]. If the value tend to be negative values, it is preferable to set $\beta_k = 0$, which implies a restart along g_k. If $\beta_k < 0$, the search direction d_k from (3) tend to almost opposite to d_{k-1}. By restricting the CG coefficient to $\beta_k \geq 0$ the two consecutive search directions are prevented from tending to be almost opposite [11]. The new algorithm of β_k^{HSMR} is given as follows;

Algorithm 2.1: Conjugate Gradient Algorithm

A complete algorithm of CG method could be generated as follows:

Step 1: Initialization. Set $k = 0$ and select $x_0 \in \Re^n$, $d_0 = -g_0$, if $g_0 = 0$, stop.

Step 2: Compute β_k^{HSMR} based on (12).

Step 3: Compute search directions d_k based on (3). If $\|g_k\| \leq \varepsilon$, then stop. Otherwise, go to the next step.

Step 4: Compute for α_k based on (4) and (5).

Step 5: Updating new initial point using (2).

Step 6: Convergence test and stopping criteria. If $f(x_{k+1}) < f(x_k)$ and $\|g_k\| \leq \varepsilon$ then, stop. Otherwise go to Step 2 with $k = k+1$.

III. CONVERGENCE ANALYSIS

In this section, the convergence analysis for HSMR based on exact and inexact line search in (4) is analyzed. An algorithm has to possess both sufficient descent condition and global convergence properties for a method to be converged in order to have a good practical algorithm. The convergence analysis for SMR and RMIL can be reached out in [3-4].

3.1 Sufficient Descent Condition

Sufficient descent condition holds when

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

Theorem 1: Consider a CG method with search direction (3) and β_k^{HSMR} defined as (12), then, condition (13) will hold for all $k \geq 0$.

Proof: From (3), know that $g_0^T d_0 = -C\|g_0\|^2$. Hence, condition (13) holds. In order to show condition (13) also hold for $k \geq 1$, multiply (3) by g_k^T. Then,

$$g_k^T d_k = -g_k^T g_k + \beta_k^{HSMR} g_k^T d_{k-1}$$
$$= -\|g_k\|^2 + \beta_k^{HSMR} g_k^T d_{k-1}. \qquad (14)$$

For $k = 1$, it is easy to know that $g_1^T d_1 = -\|g_k\|^2 < 0$. Assume that $g_{k-1}^T d_{k-1} < 0$ holds for $k-1$ and $k \geq 2$. To prove that $g_k^T d_k < 0$ for all k, consider these two cases:

CASE I: If $\beta_k^{RMIL} = \frac{g_k^T (g_k - g_{k-1})}{\|d_{k-1}\|^2} < 0$, automatically β_k^{HSMR} return to 0 such that $\beta_k^{SMR} \geq 0$. When β_k^{HSMR} =0, it is known that $g_k^T d_k = -\|g_k\|^2 < 0$.

CASE II: If $\beta_k^{RMIL} = \frac{g_k^T (g_k - g_{k-1})}{\|d_{k-1}\|^2} \geq 0$, then, $\beta_k^{HSMR} \neq 0$ and can either be β_k^{SMR} or β_k^{RMIL}.

Since the line search is exact, it is known that $g_k^T d_{k-1} = 0$. Thus, $g_k^T d_k = -\|g_k\|^2$ implying d_k is a sufficient descent direction. Hence, the descent

condition holds i.e., $g_k^T d_k \le -C\|g_k\|^2$. The proof is completed.

3.2 Global Convergence Properties

From [3] and [4], it is known that β_k^{SMR} and β_k^{RMIL} can be simplified to

$$0 \le \beta_{k+1} \le \frac{\|g_{k+1}\|^2}{\|d_k\|^2}. \qquad (15)$$

Case II stated that β_k^{HSMR} and can either be β_k^{SMR} or β_k^{RMIL} if $\beta_k^{RMIL} = \frac{g_k^T(g_k - g_{k-1})}{\|d_{k-1}\|^2} \ge 0$. Then, from [3],[4] and [12], in the analysis of global convergence properties, the following assumption is needed.

Assumption 1

1. f is bounded below on the level set R^n and is continuous and differentiable in a neighbourhood N of the level set $\ell = \{x \in R^n \mid f(x) \le f(x_0)\}$ at the initial point x_0.
2. The gradient $g(x)$ is Lipschitz continuous in N, so, there exists a constant $L>0$ such that; $\|g(x) - g(y)\| \le L\|x - y\|$ for any $x, y \in N$.

Under this assumption, the following lemma is obtained, which was proved by Zoutendijk [11].

Lemma 1: Suppose that Assumption 1 holds. Consider any CG methods of the form (3) where d_k is a descent search direction and α_k satisfies the exact minimization rules. Then the following conditions known as Zoutendijk conditions hold;

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

The proof of this lemma can be seen from [13].By using this lemma; the following convergence theorem of the conjugate gradient method can be obtained by using (15).

Theorem 1: Suppose that Assumption 1 holds. Consider any CG methods in the form of (3) and (2) where α_k is obtained by the exact minimization rules. Also, suppose that Assumption 1 and the descent condition hold. Then either

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

Proof: By induction, if Theorem 1 is not true then, there exists a constant $c > 0$ such that

$$\|g_k\| \ge c \qquad (16)$$

Rewriting (3) and squaring both sides, we get

$$\|d_{k+1}\|^2 = (\beta_{k+1})^2\|d_k\|^2 - 2g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2 \qquad (17)$$

Using $(g_{k+1}^T d_{k+1})^2$ and dividing both sides,

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

By completing the square, the equation becomes

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

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

Applying (15), yields

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

$$= \frac{\|g_{k+1}\|^2}{\|d_k\|^2\|d_{k+1}\|^2} + \frac{1}{\|g_{k+1}\|^2} \le \frac{1}{\|d_k\|^2} + \frac{1}{\|g_{k+1}\|^2}$$

(18)

From (18), noting that

$$\frac{1}{\|d_0\|^2} = \frac{1}{\|g_0\|^2},$$

then,

$$\frac{\|d_k\|^2}{(g_k d_k)^2} = \frac{1}{\|g_0\|^2} + \frac{1}{\|g_k\|^2}.$$

Hence,

$$\frac{\|d_k\|^2}{(g_k d_k)^2} \le \sum_{i=0}^{k} \frac{1}{\|g_i\|^2} \text{ and } \frac{(g_k d_k)^2}{\|d_k\|^2} \ge \frac{c^2}{k}.$$

Then, from (16) and (18), it follows; $\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \infty$.

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

Corollary 1: If $\sum_{k=0}^{\infty}\|d_k\|^2 = 0$, then $\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty$ holds.

Proof: By using contradiction, assume that $\|g_k\| \ge c$ and $\sum_{k=0}^{\infty}\|d_k\|^2 = \infty$.For $\|g_k\| \to \infty$, then, $\frac{1}{\|g_k\|} \to 0$.

From (18),

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \le \frac{1}{\|d_k\|^2}$$

then,

$$\|d_k\|^2 \le \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2}.$$

Which leads to, $\sum_{k=0}^{\infty}\|d_k\|^2 \le \sum_{k=0}^{\infty}\frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2}$ and $\infty \le \sum_{k=0}^{\infty}\frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2}$.

This contradicts the Zoutendijk conditions. Hence the corollary holds.The proof is completed.Referring to Case I and Case II, the proving process of β_k^{HSMR} follows either β_k^{SMR} or β_k^{RMIL} in [3-4]. However, the convergence properties of β_k^{RMIL} under inexact line search are not yet proven, so its numerical performances under that type of line search approach is still unknown. Since HSMR sometimes follows the RMIL method, if the hybrid method performs poorly under inexact line search, it is possible that happens due to the characteristics of RMIL.

IV. RESULTS AND FINDINGS

This section shows the numerical performance of the new coefficient HSMR compared under two line searches.In this research, artificial problems are chosen from the small to large-scale twenty-one test functions as listed in Table I. The list is considered from Andrei [14]. A large-scaled problem is chosen in order to detect a cynical observer preventing the algorithm being tuned in particular functions [15]. For each of the test functions selected, random initial points are chosen in order to test the global convergence properties along with the robustness of the methods. On the other hand, for each of the random initial points chosen, they are tested on various dimensions from two to thousand dimensions [16-17]. Points chosen can also be used to test the global convergence properties and the robustness of the methods.

No	Function
1	Camel Function- Three Hump
2	Camel Function- Six Hump
3	Booth Function
4	Dixon and Price Function
5	Treccani Function
6	Zettl Function
7	Leon Function
8	Quartic Function
9	Extended Freudstein and Roth Function
10	Extended Himmelblau Function
11	Extended Rosenbrock Function
12	Extended Denschnb Function
13	Extended White and Holst Function
14	Extended Beale Function
15	Cube Function
16	Extended Tridiagonal 1 Function
17	Shallow Function
18	Generalized Quartic Function
19	Fletchcr Function
20	Extended Maratos Function
21	Diagonal 4

Table1: List of Test Functions

Results analyses are based on MATLAB subroutine programme on workstation Intel Core i7, 2.2 GHz tested on number of iterations and central processing time per unit. The stopping criterion is set to $\|g_k\| \le 10^{-6}$, where $\varepsilon = 10^{-6}$. Performance profiles based on Dolan and More [18] are shown graphically in figures below.

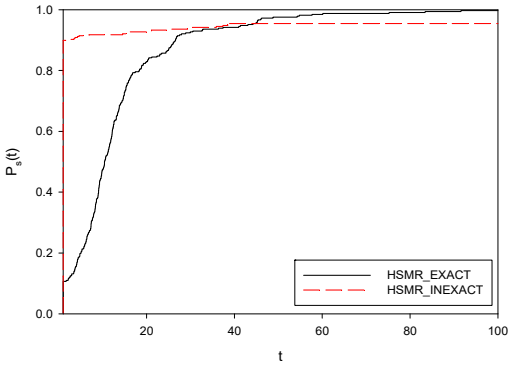


Figure1.Performance Profile for HSMR under Exact and Inexact Line Search

V. CONCLUSION

In this paper, a new hybrid of CG algorithm β_k^{HSMR} is proposed and examined for solving unconstrained optimization problem under to line searches; exact and inexact line search. Proofs showed that the algorithm satisfies the convergence properties globally under descent conditions. Numerical results have shown that the new algorithm proposed performs better under exact line search. In the future, the new algorithm is going to be tested under different search rules.

ACKNOWLEDGEMENT

The authors are very grateful to the editors and the anonymous reviewers for their valuable comments and suggestions which improved this paper substantially. The author is very grateful to Universiti Kuala Lumpur to make the publication possible.

REFERENCES

[1] M. Rivaie, M. Mamat and A. Abashar, A new class of nonlinear conjugate gradient coefficients with exact and inexact line searches, Appl. Math. Comp. 268:1152-1163. (2015).

[2] M. A. H. Ibrahim, M. Mamat and L. W. June,A New Search Direction, Sains Malaysia. 43(10):1591- 1597. (2014).

[3] M. Rivaie, M. Mamat, W. J. Leong, and I. Mohd, A New Class of Nonlinear Conjugate Gradient Coefficient with Global Convergence Properties. Appl. Math. Comput. 2012. 218: 11323-11332.

[4] N. S. Mohamed, M. Mamat, and M. Rivaie, New Coefficient of Conjugate Gradient Method for Unconstrained Optimization. Jurnal Teknologi. 2016. 78:6-4,131-136.
[5] Y. H. Dai, and Y. Yuan, A Nonlinear Conjugate Gradient with a Strong Global Convergence Properties,SIAM Journal Optimization, 10: 177-182. (2000).
[6] M. R. Hestenes and E. Steifel, Method of Conjugate Gradient for Solving Linear Equations J. Res. Nat. Bur. Stand, 49:409-436. (1952).
[7] J. Jinbao, H. Lin, and J. Xianzhen, A Hybrid Conjugate Gradient Method With Descent Property for Unconstrained Optimization, Applied Mathematical Modelling. (2014).
[8] Y. Liu, C. Storey, Efficient Generalized Conjugate Gradient Algorithms,Part 1: Theory. J. Optim. Theory Appl. 69: 129–137. (1991).
[9] R. Fletcher, Practical Methods of Unconstrained Optimization, (J. Wiley and Sons, New York, 1987).
[10] N. Aini, M. Mamat, and M. Rivaie, A modified Conjugate Gradient Coefficient with Inexact Line Searchfor Unconstrained Optimization, AIP Conference Proceeding 1787, (American Institute of Physics, Melville, NY, 2016).
[11] M. J. D. Powell, Nonconvex minimization calculations and the conjugate gradient method, in LectureNotes in Mathematics 1066: 122–141, (Springer, Berlin, 1984).
[12] The Convergence Properties of a New Type of Conjugate Gradient Methods, Applied Mathematical Sciences, 8(1):33-34. (2014).
[13] G. Zoutendijk, Nonlinear Programming Computational Methods. J. Abadie (Ed.), Integer and Nonlinear Programming, 37-86. (1970).
[14] D. Touati-Ahmed, C. Storey, Global Convergent Hybrid ConjugateGradient Methods,C. J. Optim. Theory Appl. 64: 379–397. (1990).
[15] N. Andrei, An Unconstrained Optimization Test Function Collection, Advanced Modelling and Optim. 10(1). (2008).
[16] J. J. More, B. Garbow, and K. E. Hillstrom, Testing Unconstrained Optimization Software, Journal ACM Transaction on Mathematical Software, 7(1): 17-41. (1981).
[17] K.E. Hilstrom, Test Approach to the Evaluation of Nonlinear Optimization Algorithms, A.C.M. Trans. Maths. Softw. 3: 305–315. (1977).
[18] E. Dolan, J.J. More,Benchmarking Optimization Software with Performance Profiles, Maths. Prog. Ser. 91: 201–213. (2002).

★★★