The convergence properties of new hybrid conjugate gradient method

Y Salih, M A Hamoda, Sukono, M Mamat

1 Department of Computer Science and Mathematics, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, 22200 Terengganu, Malaysia
2 Department of Mathematics, Faculty of science, AlAsmarya Islamic University, Libya
3 Universitas Padjadjaran, Bandung, Indonesia

Corresponding author: Si2023@putra.unisza.edu.my

Abstract. In this paper, a new Hybrid Conjugate Gradient Methods is presented, which produced sufficient descent search direction at every iteration and global Convergence Properties for solving large-scale nonlinear optimization problem under exact line search. The numerical experiments show that a hybrid method has the best efficiency for the test problems.

1. Introduction

Let us consider the nonlinear unconstrained optimization problem

$$\min \{ f(x) : x \in \mathbb{R}^n \},$$

(1)

Where f is smooth and its gradient g is available. Conjugate gradient methods are very efficient for solving (1), especially when the dimension n is large. The iterates of (CG) method for solving (1) are obtained by

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

(2)

Where $\alpha_k$ is the current iterate, $\alpha_k$ is a positive scalar and called the step length which is determined by some line search, and $d_k$ is the search direction defined by

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

(3)

Where $g_k = \nabla f(x_k)$ and $\beta_k$ is the conjugate gradient coefficient. Well-known formulae for $\beta_k$ include the Hestenes-Steifel (HS) method [1], the Fletcher-Reeves (FR) method [2], the Polak-Ribiere-polyak (PRP) method [3][4], the Conjugate Descent (CD) method [5], Liu-Storey (LS) method [6], and the Dai-Yuan (DY) method [7], which are given, respectively, by

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

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

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

$$\beta_k^{CD} = \frac{\|g_k\|^2}{d_k^2 (g_k - g_{k-1})}$$

$$\beta_k^{LS} = \frac{g_k^T (g_k - g_{k-1})}{-d_k^2 (g_k - g_{k-1})}$$

$$\beta_k^{DY} = \frac{\|g_k\|}{d_k^2 (g_k - g_{k-1})}$$

Many studies have been achieved to explain the poor performance of some methods for some problems, and to establish enough descent property of conjugate gradient techniques [8-17].
2. Hybrid Conjugate Gradient Method and Algorithm

In this section, we present a new Hybrid conjugate gradient method, namely YHM2, where YHM2 denotes Yasir, Hamoda and Mamat. The hybridization manner process happened while two or more classical CG algorithms are blended. The combination of classical CG techniques Used in this paper is referred to PRP and DPRP methods respectively [3][18],

Using the above formula, we proposed

is defined by

The Algorithm is given as follows:

Step1: Initialization
Given
, set
, set
.

Step2: Computing conjugate gradient coefficient
, by formula (6).

Step3: Compute
based on (3) if
then stop.

Step4: Computing step size by

Step5: Update new point base on (2).

Step6: Set
goes to Step 2.

3. Global convergence Analysis

Firstly we need to simplify the
, in order that the proof may be easier. From (6), we know that,

Also,

From [18] we know that

Hence we obtain,

3.1. Sufficient descent condition

For the sufficient descent condition to hold,

The following theorem shows that our new formula with exact line searches possess the sufficient descent condition.
**Theorem 1.** Let \( \{g_k\} \) and \( \{d_k\} \) be sequence generated by the method of the form (2), (3) and the above Algorithm, where the step size \( \alpha_k \) is determined by the exact line search, then (7) holds for all \( k \geq 0 \).

**Proof.**

We prove by induction, that if \( k = 0 \) then \( g_0^T d_0 = -c \|g_0\|^2 \).

Hence, the condition holds true, now we need to prove that:

\( g_k^T d_k \leq -c \|g_k\|^2 \) for \( k \geq 1 \)

From (3) we have \( d_{k+1} = -g_{k+1} + \beta_{k+1} d_k \)

Multiply both sides by \( g_{k+1}^T \)

\( g_{k+1}^T d_{k+1} = g_{k+1}^T (-g_{k+1} + \beta_{k+1} d_k) = -\|g_{k+1}\|^2 + \beta_{k+1} g_{k+1}^T d_k \)  \( \tag{8} \)

For exact line search, we know that \( g_{k+1}^T d_k = 0 \). Thus \( g_{k+1}^T d_{k+1} = 0 \). Therefore, the sufficient descent condition holds.

\( \square \)

### 3.2. Global convergence Properties

In this subsection, we will prove the global convergence properties of the YHM2 under the following assumption.

**Assumption 1**

i) The level set \( \Omega = \{ x \in \mathbb{R}^n, f(x) \leq f(x_0) \} \) is bounded.

ii) Function \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) is continuously differentiable with Lipschitz continuous gradient in \( \mathbb{R}^n \), there exists a constant \( L \) such that,

\[ \|g(x) - g(y)\| \leq L \|x - y\|, \quad \forall \ x, y \in \mathbb{R}^n \]

The following lemma is often used to prove global convergence which was provided by Zoutendijk [19].

**Lemma 1.**

Let \( x_k \) be generated by Algorithm for which Assumption 1 hold. Consider any method in form 1 and 2 and \( d_k \) satisfies \( g_k^T d_k < 0 \) for all \( k \) and \( g_k \) is obtained by \( \alpha_k = \arg \min_{\alpha \geq 0} f(x_k + \alpha d_k) \), then,

\[ \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \]  \( \tag{9} \)

The following theorem is based on Lemma 1.

**Theorem 2**

Suppose that the conditions in Assumption 1 hold. Consider the YHM2 method and step size \( \alpha_k \) is obtained by the exact line search then,

\[ \lim_{k \to \infty} \inf \|g_k\| = 0 \]  \( \tag{10} \)

**Proof. Case 1:**

If \( \|g_k\|^2 \geq \|g_k^T g_{k-1}\| \) then

\[ \beta_k^{YHM2} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2 + g_k^T d_{k-1}} \]

To prove this theorem, we use contradiction. If Theorem 2 is not true, then a constant \( c > 0 \) exists, such that

\[ \|g_k\| \geq c \]  \( \tag{11} \)

From equation (3) we have \( d_k = -g_k + \beta_k^{H} d_{k-1} \) and squaring both sides of the equation, we get

\[ \|d_k\|^2 = \|g_k\|^2 - 2 \beta_k^{H} g_k d_{k-1} + (\beta_k^{H})^2 \|d_{k-1}\|^2 \]  \( \tag{12} \)

Dividing both sides by \( (g_k^T d_k)^2 \), then,
\[ \frac{\|d_k\|^2}{(g_k^T d_k)^2} = \left( \beta_k^{HM2} \right)^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{2}{g_k^T d_k} \frac{\|g_k\|^2}{(g_k^T d_k)^2} \]

\[ = \left( \beta_k^{HM2} \right)^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{1}{\|g_k\|} \frac{\|d_k\|^2}{\|g_k\|^2} + \frac{1}{\|g_k\|^2} \leq \frac{1}{\|g_k\|^2} \]

Because of an exact line search \( g_k^T d_{k-1} = 0 \). Hence,
\[ \frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \sum_{i=0}^{k} \frac{1}{\|g_i\|^2} \]

Therefore,
\[ \frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{c^2}{k} \]

This contradicts the Zoutendijk condition in \( \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \)

The proof is completed.

**Case 2:**
When \( \beta_k^{HM2} = \frac{\|g_k\|^2 - \|g_{k-1}\|^2}{\mu \|g_k\|^2 + \|g_{k-1}\|^2} \), \( \mu > 1 \)
the proof of this theorem can be seen in [18].

### 4. Numerical experiments and discussions

In this section, we selected some preliminary numerical experiments considered in Andrei [20]. The performance of this method is compared with other coefficients such as FR [2], PRP [3][4], WYL [17], and DPRP [18] methods we considered \( \epsilon = 10^{-6} \) and the gradient value as the stopping criteria as suggested by Hillstrom [21]. For every test function used, there are four or five initial points has been tested randomly from it. A list of problem function and the initial points used are shown in table1. All problems mentioned are solved by MATLAB version 8.3.0.532 (R2014a). We used the exact line search to compute the step size, the CPU processor used was Intel(R) Core(TM) i5-M520 (2.40GHz), with RAM 4GB.

Figure 1 and Figure 2 list the performance of the above methods relative to the number of iterations and CPU time respectively, using a performance profile introduced by Dolan and More [22].
Table 1. List of test functions

| NO | Function                   | Dimension | Initial points          |
|----|----------------------------|-----------|-------------------------|
| 1  | Six Hump Camel             | 2         | -10, -8, 8, 10          |
| 2  | Three Hump Function        | 2         | -10, 10, 20, 25         |
| 3  | Booth                      | 2         | 10, 25, 50, 100         |
| 4  | Treccani                   | 2         | 5, 10, 20, 50           |
| 5  | Zettl                      | 2         | 5, 10, 20, 30           |
| 6  | Ex – Rosenbrock            | 2, 4, 10, 100, 500, 1000, 10000 | 13, 25, 30, 50          |
| 7  | Diagonal 4                 | 2, 4, 10, 100, 500, 1000 | 1, 3, 6, 12            |
| 8  | Shallow                    | 2, 4, 10, 100, 500, 1000, 10000 | 10, 25, 50, 70         |
| 9  | Ex - Tridiagonal1          | 2, 4, 10, 100, 500, 1000, 10000 | 12, 17, 20, 30         |
| 10 | Ex- white and Holst        | 2, 4, 10, 100, 500, 1000, 1000 | 3, 5, 7, 10            |
| 11 | Perturbed Quadratic Function | 2, 4, 10, 100, 500, 1000 | 1, 3, 5, 10            |
| 12 | Ex- Denschnb               | 2, 4, 10, 100, 500, 1000, 10000 | 8, 13, 30, 50          |
| 13 | Ex- Beale                  | 2, 4, 10, 100, 500, 1000, 10000 | -1, 3, 7, 10           |
| 14 | Ex – Himmelblau            | 10, 100, 500, 1000, 10000 | 50, 70, 100, 125       |
| 15 | Generalized Quartic        | 2, 4, 10, 100, 500, 1000, 10000 | 1, 2, 5, 7            |
| 16 | Hager                      | 2, 4, 10, 100, 500, 1000, 10000 | 5, 10, 15, 20         |
| 17 | Ex – Penalty               | 2, 4, 10, 100 | 80, 100, 111, 120     |
| 18 | Quadratic QF2              | 2, 4, 10, 100, 500, 1000 | 5, 20, 50, 100        |
| 19 | Ex - Quadratic Penalty qp2 | 2, 4, 10, 100, 500, 1000 | 10, 20, 30, 50        |
| 20 | Diagonal 2                 | 2, 4, 10, 100, 500, 1000 | 1, 5, 10, 15          |
| 21 | Raydan1 Function           | 2, 4, 10, 100 | 1, 3, 5, 7          |
| 22 | Sum Squares Function       | 2, 4, 10, 100, 500, 1000 | 1, 3, 7, 10           |
|    | Generalized Tridiagonal 1  | 2, 4, 10, 100 | 7, 10, 13, 20        |
| 23 | Fletcher                   | 4, 10, 100, 500, 1000, 10000 | 3, 5, 8, 9          |
| 24 | Quadratic QF1              | 2, 4, 10, 100, 500, 1000 | 3, 5, 8, 10          |
| 25 | Generalized Tridiagonal 2  | 2, 4, 10, 100 | 15, 18, 20, 22        |
| 26 | Leon Function              | 2         | 2, 5, 8, 10            |
| 27 | Ex-Wood                    | 4         | 3, 5, 20, 30           |
| 28 | Quartic Function           | 4         | 5, 10, 15, 20          |
| 29 | Matyas Function            | 2         | 1, 5, 10, 15          |
| 30 | Colville Function          | 4         | 2, 4, 7, 10           |
From the above Figures, we can see the YHM2 coefficient better than other coefficients (FR, PRP, DPRP and WYL) because YHM2 solves all of the test problems and achieve 100%.

5. Conclusion
The new hybrid conjugate gradient method have been proven that satisfies the sufficient descent condition and global convergence prosperities under exact line search. The numerical results for this method are more efficient when as compared to other CG methods.
Acknowledgements
The authors would like to thank the Ministry of Higher Education of Malaysia for the funding of this research under the Fundamental Research Grant Scheme (Grant no.FRGS/1/2017/STG06/UniSZA/01/1) and also the Government of Sudan for the funding of this study.

Reference
[1] Hestenes MR and Stiefel E, Methods of conjugate gradients for solving linear systems. Vol. 49. 1952: NBS.
[2] Fletcher R and Reeves C M, Function minimization by conjugate gradients. The computer journal, 1964. 7(2): p. 149-154.
[3] Polak E and Ribiere G, Note sur la convergence de méthodes de directions conjuguées. Revue française d'informatique et de recherche opérationnelle. Série rouge, 1969. 3(16): p. 35-43.
[4] Polyak BT, The conjugate gradient method in extremal problems. USSR Computational Mathematics and Mathematical Physics, 1969. 9(4): p. 94-112.
[5] Fletcher R, Practical methods of optimization. Vol. 1, Unconstrained Optimization. 1980.
[6] Liu Y and Storey C, Efficient generalized conjugate gradient algorithms, part 1: theory. Journal of optimization theory and applications, 1991. 69(1): p. 129-137.
[7] Dai YH and Yuan Y, A nonlinear conjugate gradient method with a strong global convergence property. SIAM Journal on optimization, 1999. 10(1): p. 177-182.
[8] Zhang L and Zhou W, Two descent hybrid conjugate gradient methods for optimization. Journal of Computational and Applied Mathematics, 2008. 216(1): p. 251-264.
[9] Andrei N, A hybrid conjugate gradient algorithm for unconstrained optimization as a convex combination of Hestenes-Stiefel and Dai-Yuan. Studies in Informatics and Control, 2008. 17(1): p. 57.
[10] Andrei N, Hybrid conjugate gradient algorithm for unconstrained optimization. Journal of Optimization Theory and Applications, 2009. 141(2): p. 249-264.
[11] Al-Bayati AY and Al-Kawaz RZ, A new hybrid WC-FR conjugate gradient algorithm with modified secant condition for unconstrained optimization. Journal of Mathematical and Computational Science, 2012. 2(4): p. 937-966.
[12] Mamat M, Rivaie M, Mohd I, and Fauzi M, A new conjugate gradient coefficient for unconstrained optimization. Int. J. Contemp. Math. Sciences, 2010. 5(29): p. 1429-1437.
[13] Rivaie M, Abashar A, Mamat M, The convergence properties of a new type of conjugate gradient methods. 2016.
[14] Hamoda M, Rivaie M and Mamat M, "A New Nonlinear Conjugate Gradient Method With Exact Line Search For Unconstrained Optimization," Journal of Humanities and Applied Science (JHAS), pp. 1-16, 2017.
[15] Aini N, Hajar N, Mamat M, Zull N and Rivaie M, Hybrid quasi-newton and conjugate gradient method for solving unconstrained optimization problems. 2017: Journal of engineering and applied sciences.
[16] Powell M J D, Buhmann M D, and Iserles A, "Approximation theory and optimization: tributes to M. J. D. Powell," ed: Cambridge University Press, 1997, pp. 1-10.
[17] Wei Z, Yao S, and Liu L, The convergence properties of some new conjugate gradient methods. Applied Mathematics and Computation, 2006. 183(2): p. 1341-1350.
[18] Dai Z and Wen F, Another improved Wei–Yao–Liu nonlinear conjugate gradient method with sufficient descent property. Applied Mathematics and Computation, 2012. 218(14): p. 7421-7430.
[19] Zoutendijk G, Nonlinear programming, computational methods. Integer and nonlinear programming, 1970. 143(1): p. 37-86.
[20] Andrei N, An unconstrained optimization test functions collection. Adv. Model. Optim, 2008. 10(1): p. 147-161.
[21] Hillstrom KE, A simulation test approach to the evaluation of nonlinear optimization
algorithms. ACM Transactions on Mathematical Software (TOMS), 1977. 3(4): p. 305-315.

[22] Dolan E.D and Moré JJ, Benchmarking optimization software with performance profiles. Mathematical programming, 2002. 91(2): p. 201-213.