An efficient hybrid conjugate gradient method with descent properties under strong Wolfe line search

I M Sulaiman¹, M Mamat², Basim A. Hassan³, P L Ghazali⁴, M Z Ahmad⁵

¹,²Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, 21300, Terengganu, Malaysia.
²,³Institute of Engineering Mathematics, Universiti Malaysia Perlis, 02600 Arau Perlis, Malaysia
³Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, Iraq
⁴Faculty of Business and Management Sciences, Universiti Sultan Zainal Abidin, Terengganu, Malaysia

Abstract. The hybrid conjugate gradient parameters are among the efficient variants of conjugate gradient (CG) methods for solving large-scale unconstrained optimization problems. This is due to their nice convergence properties and low memory requirements. In this paper, we present a new hybrid conjugate gradient method based on famous CG algorithms for large-scale unconstrained optimization. The proposed hybrid CG method can generate a descent search direction at each iteration provided the strong Wolfe line search is employed. Numerical results have been presented which show that the proposed method is efficient and promising.

1. Introduction

In this paper, we consider the following unconstrained optimization problem

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

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a smooth function whose gradient $g_k$ is available. Problems of these form often arise in applied mathematics, economics, engineering, and many more. Some of the classical methods available for solving equation (1) are the Steepest Descent method, Newton method, and Conjugate Gradient (CG) method [1]. However, researchers are more interested in the conjugate gradient method because of its simplicity, global convergence properties, low memory requirement, ability to solve large-scale unconstrained optimization problems, and practical applications [2-9]. The CG algorithm computes its sequence of iterates using the following iteration formula

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

where the step-size $\alpha_k > 0$ is computed using a line search method and $d_k$ is the search direction defined as follows

$$d_k = \begin{cases} -g_k, & k = 0 \\ -g_k + \beta_k d_{k-1}, & k \geq 1 \end{cases}$$
The parameter $\beta_k \in R$ represent the CG coefficient and it differentiate the types of CG algorithm. In computing for the step-size, $\alpha_k$ is said to satisfy any of the line search conditions such as the exact line search

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

(4)

The weak Wolfe line search

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

(5)

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k$$

(6)

or strong Wolfe line search, equation (5) and

$$|g_{k+1}^T d_k| \leq \sigma |g_k^T d_k|$$

(7)

where the often used constant $0 < \delta < \sigma < 1$. The classical conjugate gradient formulas are categorized into two groups. The first group include Hestenes and Steifel (HS) [10], Polak Ribiere and Polyak (PRP) [11,12], and Liu-Storey (LS) [13].

$$\beta^H_k = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}}, \beta^RP_k = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}, \beta^LS_k = \frac{g_k^T (g_k - g_{k-1})}{-d_{k-1}^T g_{k-1}}$$

(8)

This group are classified as the most efficient CG algorithm for solving large scale optimization problems. Their efficiency is attributed to the in-built automatic restart feature they possess which prevent them from jamming. However, the convergence result of some of these methods is yet to be established under certain line search methods. Also, these group may fail to converge for certain benchmark function [4]. The second group includes Fletcher and Reeves (FR) [14], Dai and Yuan (DY) [15] and Conjugate Descent (CD) [16]:

$$\beta^FR_k = \frac{g_k^T g_k}{\|g_{k-1}\|^2}, \beta^{DY}_k = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}}, \beta^{CD}_k = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}$$

(9)

These methods have poor numerical performance because they do not possess the in-built automatic restart feature which makes them vulnerable to jamming phenomena [17].

The above stated drawbacks and advantages motivated researchers to study many variants of the conjugate gradient method such as the spectral CG method, the three-term CG method and the hybrid CG algorithms. The aim is to overcome the lapses and maximize in the advantages of the existing CG methods. It is worthy to know that most of these researchers focussed more on the hybrid conjugate gradient methods that combine two or more CG coefficient $\beta_k$ [18]. The FR method possess the sufficient descent condition and converge globally under exact minimization condition. Al-Baali [19] extended the FR method in equation (9) as follows:

$$\beta_k \in \{0, \beta^FR_k\}$$

(10)

With the above derivation, the author show that the FR method satisfy the descent property and further established the global convergence proof using inexact line search under certain standard conditions. Some numerical results have been presented to support the theoretical proof. Based on equation (10), Touati-Ahmed and Storey [20] proposed a hybrid parameter with formula defined as follows

$$\beta^TS_k = \begin{cases} \beta^{PRP}_k & \text{if } 0 \leq \beta^{PRP}_k \leq \beta^FR_k \\ \beta^FR_k & \text{else where} \end{cases}$$

(11)
This method produces an efficient numerical performance which is better that the classical FR and PRP method and the author further show that $\beta_k^{HS}$ has good convergence properties. Dai and Yuan \cite{21} further extended the HS and DY method to present the following algorithms

$$\beta_k^{HYZ} = \max \{0, \min\{ \beta_k^{HS}, \beta_k^{DY} \} \}$$  \hspace{1cm} (12)

$$\beta_k^{HDY} = \max \{-c\beta_k^{DY}, \min\{ \beta_k^{HS}, \beta_k^{DY} \} \}$$

where $c = \frac{1-\sigma_1}{1+\sigma_1}$. An interesting feature these methods is their ability to possess the restart property when a bad direction is generated.

In this paper, we present a new hybrid CG method for unconstrained minimization functions and further show that the proposed method satisfies the sufficient descent condition under some line search condition. The motivation and method formulation is presented in the next section. The rest part of the manuscript is structured as follows: Section 3 present the sufficient descent condition of the proposed method and section 4 present the numerical results with respect to performance profile by Dolan and More \cite{27}. Finally, the conclusion is presented in section 5.

2. A New Hybrid Conjugate Gradient Method

Based on the nice convergence properties and efficient numerical performance of $\beta_k^{PRP}$ method, Wei et al. \cite{28} extended this method by introducing $\frac{||g_k||}{||g_{k-1}||}$ to the numerator as follows

$$\beta_k^{WYL} = \frac{g_k^T(g_k - \frac{||g_k||}{||g_{k-1}||}g_{k-1})}{||g_{k-1}||^2}.$$  \hspace{1cm} (13)

This method not only inherited the good convergence behaviour of the classical PRP method but also the nice numerical performance. The authors prove the global convergence of $\beta_k^{WYL}$ under certain conditions. However, Zhang \cite{29} further extended the WYL method by expanding the numerator as follows:

$$\beta_k^{NPRP} = \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} |g_k^T g_{k-1}|}{||g_{k-1}||^2}. $$  \hspace{1cm} (14)

The author show that the method satisfies the sufficient descent condition and further proved the convergence under strong Wolfe line search.

Motivated the idea of Wei et al. \cite{28}, Zhang \cite{29} and the structure of Touati-Ahmed and Storey \cite{20}, this paper propose a hybrid conjugate gradient method whose coefficient is given as

$$\beta_k^{SIM} = \begin{cases} 
\frac{g_k^T(g_k - \theta g_{k-1})}{||g_{k-1}||^2} & \text{if } ||g_k||^2 > \theta g_k^T g_{k-1} \\
\frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} |g_k^T g_{k-1}|}{||g_{k-1}||^2} & \text{Otherwise} 
\end{cases}$$  \hspace{1cm} (15)

where $\theta = \frac{||g_k||}{||g_{k-1}||}$. Now, we present the algorithm of our proposed hybrid conjugate gradient method as follows.

Algorithm 2.1. Algorithm for $\beta_k^{SIM}$

Stage 1. Initialization. Given $x_0 \in \mathbb{R}^n$, $d_k = -g_k$, set $k = 0$. 


To study the global convergence of the proposed hybrid CG method, the following assumptions are often needed on the objective function.

Assumptions
A. \( f(x) \) is bounded from below on the level set \( \Omega = \{ x \in \mathbb{R}^n / f(x) \leq f(x_0) \} \).
B. In some neighbourhood \( N \) of \( \Omega \), \( f \) is smooth and \( g(x) \) is Lipchitz continuous in \( N \), such that, there exist \( L > 0 \) (constant) satisfying:

\[
\|g(x) - g(y)\| \leq L\|x - y\| \quad \forall x, y \in N. \tag{16}
\]

3. Convergence Analysis
In this section, we shall study the convergence analysis of the proposed hybrid method. We need to show that the proposed method satisfies the sufficient descent condition.

3.1. Sufficient Descent Condition
The sufficient descent is given by

\[
g_k^T d_k \leq -c \|g_k\|^2, \quad c \in (0, 1) \tag{17}
\]

This condition is very essential when studying different conjugate gradient methods. From equation (15), it is obvious that

\[
0 \leq \beta_k^S IM \leq \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \tag{18}
\]

We need the following theorem to show that the proposed method satisfies the sufficient descent condition defined in equation (17) when strong Wolfe Powell line search is considered.

Theorem 3.1: Consider the sequences \( \{g_k\} \) and \( \{d_k\} \) generated by equations (2), (3), and (18), with the step-size \( \alpha_k \) calculated using the SWP line search given by equations (5) and (7). If \( \sigma \in \left(0, \frac{1}{2}\right) \), then, we say equation (17) holds.

Proof: The proof of this theorem is by induction. From equation (3), we have

\[
d_k = -g_k + \beta_k d_{k-1}. \tag{19}
\]

Multiplying through equation (19) by \( g_k^T \), we have

\[
g_k^T d_k = -\|g_k\|^2 + \beta_k g_k^T d_{k-1}. \tag{20}
\]

Dividing through equation (20) by \( \|g_k\|^2 \) and applying equations (7) and (17), we obtain

\[
-1 + \sigma \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -1 - \sigma \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2}. \tag{21}
\]
Considering equation (3), we have \( g^T_i d_i = -\|g_i\|^2 \). Next, suppose that similar expressions hold true for \( k - 1 \), that is \( g^T_i d_i < 0 \), \( i = 1, 2, 3, ..., k - 1 \). If we repeat the same process for equation (17), we have

\[
- \sum_{j=0}^{k-1} \sigma^j \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -2 + \sum_{j=0}^{k-1} \sigma^j. \tag{22}
\]

From equation (22), it follows that

\[
\sum_{j=0}^{k-1} \sigma^j < \frac{1 - \sigma^k}{1 - \sigma},
\]

\[
- \frac{1 - \sigma^k}{1 - \sigma} \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -2 + \frac{1 - \sigma^k}{1 - \sigma}.
\]

However, since \( \sigma \in \left(0, \frac{1}{2}\right) \), then, when \( \sigma \leq m \), where \( m = \frac{1}{2} \), we get \( \frac{1 - \sigma^k}{1 - \sigma} < 2 \). Letting \( c = 2 - \frac{(1 - \sigma^k)}{(1 - \sigma)} \), it follows that

\[
c - 2 \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq c.
\]

And this completes the proof.

4. Numerical Results

In this section, we present the numerical experiments of the proposed method using some classical unconstrained optimization problems and compare the performance with that of existing methods of WYL and NPRP. The comparison is based on number of iteration and CPU. All algorithms are coded in MATLAB R2018a subroutine programming and run on a CoreI5 PC with a 2.50 GHz CPU, and 8.00 GB RAM. All problems have been implement using dimensions ranging from \( 2 \leq x \leq 1000 \) to ascertain the efficiency of the proposed algorithm as presented in table 1.

| No. | Functions         | N | Initial points                                                  |
|-----|-------------------|---|-----------------------------------------------------------------|
| 1   | Booth             | 2 | (-8,-8)(25,25),(50,50),(100,100)                                  |
| 2   | TRECCANI          | 2 | (5,5)(10,10),(20,20),(50,50)                                     |
| 3   | Zettl             | 2 | (5,5)(10,10),(20,20),(30,30)                                     |
| 4   | Raydan            | 2 | (1,1),(3,3),(5,5),(7,7)                                          |
| 5   | Three-Hump        | 2 | (-10,-10),(10,10),(20,20),(25,25)                                |
| 6   | Six-Hump          | 2 | (-8,-8),(8,8),(-10,-10),(10,10)                                 |
| 7   | Colville function | 2,4| (2,2),(4,4),(7,7),(10,10)                                        |
| 8   | Extended Wood     | 2,4| (3,3),(5,5),(20,20),(30,30)                                     |
| 9   | Extended penalty  | 2,4,10| (80,80),(100,100),(111,111),(120,120)                             |
| 10  | Fletcher          | 2,4,10,100| (3,3,3),(5,5,5),(8,8,8),(9,9,9)                                |
| 11  | Gen Tridiagonal   | 2,4,10,100| (7,7,7),(10,10,10),(13,13,13),(20,20,20)                        |
| 12  | Extended Beale    | 2,4,10,100| (-19,-19,-19),(1,1,1),(13,13,13),(23,23,23)                      |
| 13  | Extended Denschnb | 2,4,10,100| (8,8,8),(13,13,13),(30,30,30),(50,50,50)                        |
| 14  | Extended Tridiagonal 1 | 2,4,10,100,500| (12,12,12),(17,17,17),(20,20,20),(30,30,30)                  |
| 15  | Generalized Quartic 1 | 2,4,10,100,500| (5,5,5),(10,10,10),(15,15,15),(20,20,20)                      |
| 16  | Shallow           | 2,4,10,100,500| (10,10,10),(25,25,25),(50,50,50),(70,70,70)                    |
More so, we analyze the performance using the performance profile tool introduced by Dolan and More [27] and the results are presented figures 1 and 2.

This performance profile is used to assesses the performance of the set of the solvers $S$ on every test problem used. The figure 1 represent the performance analysis based on the number of iteration and figure 2 present the performance based on CPU time.

![Performance Profile](image)

**Figure 1:** Performance Profile Based on Number of Iteration
From both figures 1 and 2, it can be obviously seen that the proposed SIM algorithm possess a similar performance to NPRP algorithms. However, the method performed better than both WYL and NPRP method both in term of number of iteration and CPU time. This indicate that the hybrid algorithm is efficient and promising.

5. Conclusion

In this paper, we present a new hybrid CG algorithm for solving unconstrained optimization problems. The proposed method is based on the structure of Touti-Ahmad and Storey hybrid CG method and combined the classical methods of WYL and NPRP. Under some mild condition, the sufficient descent condition was discussed. Furthermore, the performance of the proposed method was evaluated on twenty-three unconstrained optimization test problem using various initial points. The comparison was done using the performance profile introduced by Dolan and More and was based on number of iteration and CPU time. The obtained result show that our proposed SIM coefficient is promising and can thus, be used as alternatives for unconstrained optimization problems.

References

[1] Yulun Wu, Mengxiang Zhang, Yan Li 2020. Modified Three-Term Liu-Storey Conjugate Gradient Method for Solving Unconstrained Optimization Problems and Image Restoration Problems, Mathematical Problems in Engineering, vol. 2020, Article ID7859286, 20 pages. https://doi.org/10.1155/2020/7859286
[2] Ibrahim S M, Mamat M, Abdelrahman A, Rivaie M., Fatma S M, and Zabidin S 2015. The Global Convergence Properties of an Improved Conjugate Gradient Method. Applied Mathematical Sciences. 9, 1857–1868.
[3] Li W, Yang Y 2015. A nonmonotone hybrid conjugate gradient method for unconstrained optimization. J. Inequal. Appl. 2015, 124. https://doi.org/10.1186/s13660-015-0644-1
[4] Kaelo, P., Magulwa, P. & Thuto, M.V 2020. A globally convergent hybrid conjugate gradient method with strong Wolfe conditions for unconstrained optimization. Math Sci 14, 1–9. https://doi.org/10.1007/s40096-019-00310-y
[5] Sulaiman I M, Mamat M, Waziri M Y, Yakubu A U, Malik M 2021. The convergence properties of a new hybrid conjugate gradient parameter for unconstrained optimization models. Journal of physics: Conference series. 1734, 012012. doi:10.1088/1742-6596/1734/1/012012.
[6] Dai, Y.H., Yuan, Y 1999. A nonlinear conjugate gradient method with a strong global convergence property.
[7] Sulaiman I M, Yakubu A U, Mamat M 2020. Application of spectral conjugate gradient methods for solving unconstrained optimization problems. *International Journal of Optimization and Control: Theories and Applications*, 10(2), 198-205.

[8] Mtagulwa P, Kaelo P 2019. An efficient modified PRP-FR hybrid conjugate gradient method for solving unconstrained optimization problems. Appl. Numer. Math. 145, 111–120.

[9] Yakubu A U, Sulaiman I M, Mamat M, Puspa L G, Khalid K 2020. The global convergence properties of a descent conjugate gradient method. *Journal of Advanced Research in Dynamical and Control Systems*, 12(2), 1011-1016.

[10] Hestenes, M.R., Steifel, E 1952. Methods of conjugate gradients for solving linear systems. J. Res. Natl. Bureau Stand. 49(6), 409–436.

[11] Polak, E., Ribiere, G 1969. Note sur la convergence de directions conjugées, Rev. Francaise Informat Recherche Opérationnelle, 3e Année, 16, 35–43.

[12] Polyak B. T. 1969. The conjugate gradient method in extremum problems, USSR Comp. Math. Math. Phys., 9, pp. 94–112.

[13] Liu, Y., Storey, C 1991. Efficient generalized conjugate gradient algorithms, Part 1: theory. J. Optim. Theory Appl. 69(1), 129–137.

[14] Fletcher, R., Reeves, C.M 1964. Function minimization by conjugate gradients. Comput. J.

[15] Dai, Y.H., Yuan, Y 1999. A nonlinear conjugate gradient method with a strong global convergence property. SIAM J. Optim. 10(1), 177–182.

[16] Fletcher, R 1987. Practical Methods of Optimization, Unconstrained Optimization. Wiley, New York.

[17] Babaie-Kafaki, S 2012. A quadratic hybridization of Polak-Ribier’e re-Polyak and Fletcher-Reeves conjugate gradient methods. J. Optim. Theory Appl. 154(3), 916–932.

[18] Mamat M, Sulaiman I M, Maulana M, Sukono, Zakaria Z A 2020. An efficient spectral conjugate gradient parameter with descent condition for unconstrained optimization. *Journal of Advanced Research in Dynamical and Control Systems*, 12(2), 2487-2493.

[19] M. Al-Baali. 1985. Descent property and global convergence of the Fletcher-Reeves method with inexact line search, IMA J. Numerical Analysis., vol.5, pp. 121–124.

[20] Touati-Ahmed, D., Storey, C. 1990. Efficient hybrid conjugate gradient techniques. J. Optim. Theory Appl. 64(2), 379–397.

[21] Y. H. Dai, Y. Yuan. 2001. An efficient hybrid conjugate gradient method for unconstrained optimization, Annals of Operations Research, vol.103, pp. 33-47.

[22] M. Jiangtao, G. Nengzhu, W. Zengxin. 2007. Hybrid conjugate gradient methods for unconstrained optimization. Optimization method and Software. vol.22(2), pp.297-307.

[23] Malik M, Abas S S, Mamat M, Sukono, Sulaiman I M 2020. A new hybrid conjugate gradient method with global convergence properties. *International Journal of Advanced Science and Technology*, 29(5), 199-210.

[24] Omesa, A U., Sulaiman I M., Mamat, M., Waziri, M Y., Foziah, M U., Rindengan, A.J., Salaki, D.T. 2020. New hybrid conjugate gradient method for solving fuzzy nonlinear equations. Journal of Advanced Research in Dynamical and Control Systems,12(2), 2020, 585-590.

[25] Kamfa, K, Waziri, M.Y., Sulaiman, I.M, Ibrahim, M.A.H, Mamat, M, Abas, S.S. 2020. An efficient hybrid bfgs-cg search direction for solving unconstrained optimization problems. Journal of Advanced Research in Dynamical and Control Systems, 12(2), 1035-1041.

[26] Liu J K, Li S J 2014. New hybrid conjugate gradient method for unconstrained optimization. Appl. Math. Comput. 245, 36-43.

[27] E. D. Dolan and J. J. More. 2002. Benchmarking optimization software with performance profiles," *Mathematical Programming*, vol. 91, pp. 201-213.

[28] Wei, Z., Yao, S., and Liu, L. 2006. The convergence properties of some new conjugate gradient methods. *Applied Mathematics and computation*, 183(2), 1341-1350.

[29] Zhang, L. 2009. An improved Wei–Yao–Liu nonlinear conjugate gradient method for optimization computation. *Applied Mathematics and computation*, 215(6), 2269-2274.