Performance Analysis of a Modified Conjugate Gradient Algorithm for Optimization Models

S E Olowo¹, I M Sulaiman²*, M Mamat², A E Owoyemi³, M A Zaini⁴, Kalfin⁵, and S H Yuningsih⁶

¹Department of Mathematics, Kano University of science and Technology, Kano, Nigeria.
²Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Campus Kuala Nerus, 22200, Terengganu, Malaysia.
³Department of General Studies, Federal College of Agricultural Produce Technology, Kano, Nigeria.
⁴Institute of Engineering Mathematics, Universiti Malaysia Perlis, Kangar, Malaysia.
⁵Doctor Program of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia.
⁶Master Program of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Indonesia.

*Corresponding author: kademi1985@gmail.com

Abstract. The Conjugate gradient (CG) algorithms is very important and widely used in solving optimization models. This is due to its simplicity as well as global convergence properties. Various line search procedures as usually employ in the analysis of the CG methods. Recently, many studies have been done aimed at improving the CG method. In this paper, an alternative formula for conjugate gradient coefficient has been proposed which possesses the global convergence properties under exact minimization condition. The result of the numerical computation has shown that this new coefficient performs better than the existing CG methods.

1. Introduction
The CG algorithms play an essential role when obtaining the solution of optimization models, particularly, when the problems involve large-dimension. This is as a result of its low memory requirement. The general formula of this method is given as

\[ \min f(x), \quad x \in \mathbb{R}^n, \]

with \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) been smooth & \( g \equiv \nabla f(x) \) denotes the gradient. The CG approaches generates an iterative sequence \( \{x_k\} \) via a method given as

\[ x_k = x_{k-1} + \alpha_{k-1}d_{k-1}, \quad \forall k \geq 1 \]  

where \( \alpha_{k-1} \geq 1 \) is obtained with any line search method. One of the commonly used line search scheme is defined as

\[ f(x_k + \alpha_k d_k) = \min f(x_k + \alpha d_k), \]

and referred to the exact minimization condition which is computed along the direction of search \( d_k \) with formula.
\[ d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geq 1, \end{cases} \tag{4} \]

The scalar \( \beta_k \in \mathbb{R} \) is known as the coefficient of the CG algorithm with different parameter characterizing the CG algorithms. Some of the classical formulas for CG coefficients \( \beta_k \) are defined below

\[
\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}, \quad \beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1})}, \quad \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T (g_k - g_{k-1})}, \quad \beta_k^{CD} = -\frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}}.
\]

\( g_k \) is known as the gradients with the Euclidean norm denoted by \( \|\| \). The algorithms define above are Liu-Storey (LS) [21], Conjugate Descent (CD) [15], Hestenes-Stiefel (HS) [9], Polak-Ribiere-Polyak (PRP) [1, 4], Dai-Yuan (DY) [20], and Fletcher-Reeves (FR) [16]. These methods have been shown to be equivalent if the quadratic functions \( f(x) \) is strongly convex. However, their behaviours differ for non-convex functions [22].

Several studies of conjugate gradient methods have been focusing on its convergence behaviour under several line search procedures. These include Zoutendijk [5], who prove the global converges of FR using exact minimization condition. However, Powell [8], illustrates through a counter example that FR performs poorly in experimentation. Further studies on the convergence of methods linked to FR algorithm were presented by Touati-Ahmed and Storey [2], Al-Baali [7] via the strong Wolfe conditions. For more references on current findings of the CG algorithms, we refer researchers [11,13,14,17,18,19].

In this study, the authors present an efficient type of the CG parameter followed by its algorithm in the next section. The converge analysis of the proposed method is discussed under exact minimization condition in section 3. Numerical results and discussion of our new formula with other well-known formulas followed in section 4 and concluding in section 5.

2. New modification with its algorithm

Lately, Rivaie et al. [10], developed an alternative coefficient of the CG method by replacing the denominator of PRP [1, 4], and HS [9] while retaining the numerator as follows

\[
\beta_k^{RMIL} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (d_{k-1} - g_k)}.
\]

Motivated by this idea, we construct a simple variant of RMIL coefficient by introducing a new numerator of Rivaie et al. [10] as follows.

\[
\beta_k^{SMO} = \frac{\|g_k\|^2 - g_k^T d_{k-1}}{d_{k-1}^T (d_{k-1} - g_k)}.
\]

and present its algorithms below

**Algorithm 1. Variant of RMIL**

Step 1. With \( x_0 \in \mathbb{R}^n \), \( d_0 = -g_k \), \( k = 0 \), check if \( \|g_k\| \leq \varepsilon \), terminate.

Step 2. Calculate \( \beta_k^{SMO} \) using (6).

Step 3. Obtain \( d_k \) via (4). If \( \|g_k\| \leq \varepsilon \), terminate
Step 4. Calculate $\alpha_k$ via (3).
Step 5. Update $x_k$ via (2).
Step 6. Check if $\|g_k\| \leq \varepsilon$, terminate, else, set $k = k + 1$ and start from the beginning.

3. Global convergence analysis

For any method to be considered effective and robust, it must satisfy the descent condition and also the convergence criteria. Here, the convergence of the proposed $\beta_k^{SMO}$ scheme will be studied.

3.1 Sufficient descent condition

This condition is true if

$$g_k^T d_k \leq -C \|g_k\|^2 \text{ for } k \geq 0 \text{ and } C > 0$$  \hspace{1cm} (7)

The Deduction that follows would be used to show the proposed method possess (7) under exact minimization criterion.

Theorem 1
Given a CG algorithm 1, then the sufficient descent property (7) is true $\forall k \geq 0$.

Proof: For $k = 0$, it’s obvious $g_0^T d_0 = -C \|g_0\|^2$. Thus, (7) holds. Next is to prove (7) also holds for $k \geq 1$. From search direction in equation (4), it implies that

$$d_{k+1} = -g_{k+1} + \beta_{k+1}^{SMO} d_k$$

Multiplying through (4) by $g_{k+1}^T$ gives

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

But $g_{k+1}^T d_k = 0$ \hspace{1cm} (8)

and thus, (7) holds true. This completes the proof. \hspace{1cm} ■

3.2 Global convergence properties

We need the simplification of our coefficient $\beta_k^{SMO}$ to make the global convergence proof easier

$$\beta_k^{SMO} = \frac{\|g_k\|^2}{d_{k-1}^T (d_{k-1}^T - g_k)} = \|g_k\|^2 \frac{g_k^T d_{k-1}}{d_{k-1}^T - d_{k-1}^T g_k}$$  \hspace{1cm} (9)

Since $g_k^T d_{k-1} = 0$, then (9) reduces to

$$0 \leq \beta_k^{SMO} \leq \frac{\|g_k\|^2}{\|d_{k-1}\|^2}$$  \hspace{1cm} (10)

Now, we need the assumptions below for the convergence analysis of these types of algorithms.

Assumption A. On the level set $\Omega = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$, the function $f(x)$ is bounded from below.
**Assumption B.** The function $f$ is smooth and its gradient $g(x)$ is Lipschitz continuous in some neighbourhood $N$ of $\Omega$, such that $\exists L > 0$ (constant) which satisfy:

$$\|g(x) - g(y)\| \leq L\|x - y\| \text{ for } \forall x, y \in N.$$ 

The following important lemma follows from Assumptions A and B, whose proof can be found in Zoutendijk [5].

**Lemma 1**
Assume Assumption A and B holds, for a given algorithm defined by algorithm 1. Then, the Zoutendijk assertion [5] that follow is true

$$\sum_{k=0}^{\infty} \left( g_k^T d_k \right)^2 < \infty \quad (11)$$

Based on the above lemma and Assumption B, we have the convergence theorem as follows

**Theorem 2**
For any given CG process defined using algorithm 1, let Assumptions A and B holds, at this point

$$\lim_{k \to \infty} \|g_k\| = 0 \text{ or } \sum_{k=0}^{\infty} \left( g_k^T d_k \right)^2 < \infty \quad (12)$$

**Proof.**
Rewriting search direction equation in (4), gives

$$d_{k+1} + g_{k+1} = \beta_{k+1}^{SMO} d_k,$$

If we square through the above equation, we have

$$\|d_{k+1}\|^2 = \left( \beta_{k+1}^{SMO} \right)^2 \|d_k\|^2 - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2 \quad (13)$$

Substitute (10) in (13), will give

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

$$\|d_{k+1}\|^2 = \frac{\|g_{k+1}\|^4}{\|d_k\|^2} - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2 \quad (14)$$

Considering (7), it follows

$$g_{k+1}^T d_{k+1} = -C \|g_{k+1}\|^2$$

and thus, (14) becomes

$$\|d_{k+1}\|^2 = \frac{\|g_{k+1}\|^4}{\|d_k\|^2} + 2C \|g_{k+1}\|^2 - \|g_{k+1}\|^2$$

$$\|d_{k+1}\|^2 = \frac{\|g_{k+1}\|^4}{\|d_k\|^2} - \|g_{k+1}\|^2 (1 - 2C) \quad (15)$$

Multiplying through (15) by $\|d_{k+1}\|^2$ gives
\[ \frac{\|d_{k+1}\|}{\|d_{k+1}\|} = \frac{\|g_{k+1}\|}{\|d_{k}\|} \left( \frac{\|g_{k+1}\|^2}{\|d_{k}\|^2} - \frac{\|g_{k+1}\|^2}{\|d_{k}\|^2} (1 - 2c) \right) \]

\[ \frac{\|d_{k+1}\|}{\|g_{k+1}\|} = \frac{\|g_{k+1}\|}{\|d_{k}\|} \left( (2c - 1) + \frac{\|g_{k+1}\|^2}{\|d_{k}\|^2} \right) \]

\[ \|d_{k+1}\|^2 \|g_{k+1}\|^2 \leq \|d_{k+1}\|^2 \]

which implies \( \lim_{k \to \infty} \left( \frac{g_{k+1}^T d_{k+1}}{\|d_{k+1}\|} \right)^2 < 0 \) from Lemma 1.

If Theorem 2 is not true, then it will imply that \( \lim_{k \to \infty} \left( \frac{g_{k+1}^T d_{k+1}}{\|d_{k+1}\|} \right)^2 = \infty \), using (16), which gives

\[ \infty \leq \|g_{k+1}\|^2 \]

Therefore, for sufficiently large \( k \), Theorem 2 holds true. ■

4. Numerical results

This subdivision is the outcome of the computational experiment from iteration number and CPU time. Most of the benchmark problems used in the analysis are from Andrei [12], as shown in Table 1. The terminating condition is set as \( \|g_k\| < 10^{-6} \). Four initial guesses are considered for each problem starting with guess close to the solution point to points far away. All problems are solved on MATLAB (R 2015b) programming. The results are analysed by Dolan and More [3] performance profile as presented in Figure 1 and Figure 2. In the performance profile, they evaluated the set of solvers \( S \) on performance \( P \) which is the test problems.

| No | Functions                  | DIM | Initial guess                      |
|----|----------------------------|-----|-----------------------------------|
| 1  | Booth                      | 2   | (50, 50), (100, 100), (10, 10), (25, 25) |
| 2  | Three-Hump Camel           | 2   | (21, 21), (23, 23), (13, 13), (17, 17) |
| 3  | Six Hump Camel             | 2   | (10, 10), (-10, -10), (8, 8), (-8, -8) |
| 4  | Treccani                   | 2   | (50, 50), (100, 100), (5, 5), (7, 7). |
| 5  | Ext Freud and Roth         | 2, 4| (20, -20, -20), (33, -33, 33), (0, 0), (10,10) |
| 6  | Dixon and Price            | 2, 4| (5, 5), (13, 13), (30, 30), (33, 33) |
| 7  | Generalized Trig 1         | 2, 4| (33, 33), (135,135), (15, 15), (23,23). |
| 8  | Ext Maratos                | 2, 4| (25,25), (55,55), (15, 15), (16, 16 ...16), |
| 9  | Hager                      | 2, 4| (22, 22, ...), (22, -23, 23, ...23), (-1, -1, ...,-1), (21, 21, ...21). |
| 10 | Ext Penalty                | 2, 4| (135,135,...135),(200,200,...,200), (100, 100,...,100),(105,105,...,105). |
| 11 | Quadratic Penalty QP2      | 2, 4| (-15, 15, ...), (-25, 25, ... 25), (-11, 11, ... 11), (13, 13, ... 13) |
From the above Fig 1 and Fig 2, it has been clearly revealed the new SMO algorithm performed efficiently both in terms of CPU time and iteration number by solving the whole test functions successful with 100% success. On the other hand, FR, PRP, and RMIL methods respectively solved 76%, 86%, and 91% of the functions. Thus, deduce that the proposed SMO algorithm is more efficient that the existing methods used in the analysis.

5. Conclusion

Recently, several researches have modified the CG algorithms. This paper presented an improved modification of the CG method base on modification of RMIL coefficient. The obtained computational results illustrate that the new algorithm is more reliable and competitive compare to methods of FR, PRP, and AMRI. The exact minimization condition was applied for the study the global convergence of SMO. In future, we intend to prove that the method also converge globally under other line searches.

References

[1] Polyak B T 1969 The conjugate gradient method in extreme problems, USSR Comp. Math. Phys. 9 94-112
[2] Touati-Ahmed D and Storey C 1990 Efficient hybrid conjugate gradient techniques J. Optim. Theory Appl 64 379–397
[3] Dolan E and More J J 2002 Benchmarking optimization software with performance profile, *Math. Prog.* 91, 201–213.

[4] Polak E and Ribière G 1969 Note sur la convergence de directions conjuguée, *Rev. Francaise Informat Recherche Operationelle* 16 35-43

[5] Zoutendijk G 1970 Nonlinear programming, computational methods, in Integer and Nonlinear Programming J. Abadie, ed., North- Holland, Amsterdam 37-86

[6] Gilbert J C and Nocedal J 1992 Global convergence properties of conjugate gradient methods for optimization *SIAM. J. Optim.* 2 21-42

[7] Al-Baali M 1985 Descent property and global convergence of the Fletcher-Reeves method with inexact line search *IMA J. Numer. Anal.* 5 121-124

[8] Powell M J D 1986 Convergence properties of algorithm for nonlinear optimization *SIAM Rev.* 28 487-500

[9] Hestenes M R and Stiefel E L 1952 Methods of conjugate gradients for solving linear systems *J. Research Nat. Bur. Standard* 49 409-436

[10] Rivaie M, Mamat M, Leong W J and Ismail M 2012 A new Conjugate gradient coefficient for large scale Nonlinear unconstrained optimization *Int. Journal of Math. Analysis* 6 (23) 1131-1146

[11] Rivaie M, Abashar A, Mamat M and Mohd I 2014 The convergence properties of a new type of conjugate gradient methods *Appl. Math. Sci.* 8 33-44

[12] Andrei N 2008 An unconstrained optimization test functions collection *Advanced Modelling and optimization* 10 (1) 147-161

[13] Andrei N 2009 Accelerated conjugate gradient algorithm with finite difference Hessian / vector product approximation for unconstrained optimization *J. comput. Appl. Math* 30 361–369

[14] Sulaiman I M, Mamat M, Abashar A, Rivaie M, Mohamad F S and Saleh Z 2015 The global convergent properties of an improved conjugate gradient method *Applied Mathematical Science* 9 (38) 1857-1868

[15] Fletcher R 1987 *Practical Method of Optimization Unconstrained Optimization* New York: John Wiley & Sons

[16] Fletcher R and Reeves C 1964 Function minimization by conjugate gradients *Comput. J.* 7 149-154

[17] Olowo S E, Abudulhadi A A and Sulaiman I M 2016 An improved version of conjugate gradient Algorithms with global convergence properties for unconstrained optimization *International Journal of Engineering and Research Technology* 75-86

[18] Sulaiman I M and Mamat M 2020 A New Conjugate Gradient Methods with Descent Properties and its Application to Regression analysis *Journal of Numerical Analysis, Industrial and Applied Mathematics* 14 (3-4) 1-15

[19] Hager W W and Zhang H C 2005 A new conjugate gradient method with guaranteed descent and efficient line search *SIAM J. Optim.* 16 170–192

[20] Dai Y H and Yuan Y 1999 A nonlinear conjugate gradient with a strong global convergence properties *SIAM J. Optim.* 10 177-182

[21] Liu Y and Storey C 1992 Efficient generalized conjugate gradient algorithms part 1: Theory *J. Comput. Appl. Math.* 69 129-137

[22] Yuan Y and Sun W 1999 *Theory and methods of optimization* Beijing: Science Press