AN IMPROVED AND FAST CONJUGATE GRADIENT COEFFICIENT FOR LARGE-SCALE OPTIMIZATION WITH REDUCED COMPUTATION TIME

by

Ibrahim Abdullahi[1,], Usman Sani[1], Baba Galadima Agaie[1], and P.L. Ndukum[2]*
[1]Federal University Dutse (FUD), P.M.B 7156 Dutse, Jigawa State, Nigeria
[2] Department of Mathematics and Computer Science, P.O. Box 67, Dchang. Western Cameroon
[*]Corresponding author: iabdullahi94@gmail.com, ibrahim.abdullahi@fud.edu.ng

Abstract

Nonlinear conjugate gradient methods (CGMs) are widely used for solving unconstrained optimization problems. These methods are among the earliest known techniques for solving large-scale unconstrained optimization problems. In this paper, we propose a modified conjugate gradient coefficient (β_k). The new method possesses sufficient descent properties with Wolfe-Powell line search condition. The proposed method is globally convergent while the simulation results are obtained with strong Wolfe-Powell line search for the purpose of comparison. We employed performance profile to show the strength of the proposed method against some CGMs using some test problems. It is observed that the proposed method is effective as compared to some CGMs.

Keywords *Unconstrained optimization; Conjugate gradient method; Global convergence; Conjugate gradient coefficient.*

1 Introduction

Today's sophisticated societies require minimum cost with maximum benefit possible. Several problems in various fields of study are formulated as optimization problems and solved with the help of various optimization algorithms. The CGMs are a family of well received local and global searches to date for solving unconstrained optimization problems arising both in the academic realm and the real world. The methods enjoy wide acceptance because of their reliability for finding solutions to the optimization problems. Obtaining optimal solutions within the shortest possible time is one of an indicator of the efficiency of the method. Therefore, identifying some shortcomings and rectifying them in the form of a new modified algorithm for the betterment of this family of methods' is worthwhile.

Let the function $f: R^n \to R$ be continuously differentiable. Given the following unconstrained optimization problem

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

and its gradient denoted by $g(x)$, the solution to Eq. 1, given an initial guess $x_0 \in R^n$, the sequence $\{x_k\}$ generated by the conjugate gradient(CG) method is given as

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

and the direction d_k is defined by

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

where x_k is the current iterate, β_k is the CG coefficient and $\alpha_k > 0$ is the step-length obtained by a line search. In the implementation and analysis of the CG methods, step-length (α_k) need to satisfy some line search conditions either the exact line searches or the inexact line searches. In this paper, we compute α_k using the following inexact line search given by

$$f(x_k + \alpha d_k) \leq \rho \alpha {g_k}^T d_k, \tag{4}$$

and $\quad |{d_k}^T g(x_k + \alpha_k d_k)| \leq \sigma |{g_k}^T d_k|$

where d_k is the descent direction and $0 < \delta < \sigma < 1$. Recently, various CG methods came into existence where the parameter β_k is the main difference among them. For details of some CG methods with their global convergence refer to the work by Hager and Zhang [8]. The summary of the pioneer CG methods is given in the Table 1

Table 1: The classical formulas for parameter β_k

No.	β_k	**Method name**	**References**
1	$\frac{\|\|g_{k+1}\|\|^2}{\|\|g_k\|\|^2}$	Fletcher-Reeves(FR) method	Fletcher and Reeves [7]
2	$-\frac{\|\|g_{k+1}\|\|^2}{{d_k}^T g_k}$	Conjugate Descent(CD) method	Fletcher [6]
3	$\frac{\|\|g_{k+1}\|\|^2}{{d_k}^T y_k}$	Dai-Yuan(DY) method	Dai and Yuan [4]
4	$\frac{{g_{k+1}}^T y_k}{\|\|g_k\|\|^2}$	Polak-Rebiere-Polyak(PRP) method	Polyak [13]
5	$-\frac{{g_{k+1}}^T y_k}{{d_k}^T g_k}$	Liu-Storey(LS) method	Liu and Storey [11]
6	$\frac{{g_{k+1}}^T y_k}{{d_k}^T y_k}$	Hestenes-Stiefel(HS) method	Hestenes and Stiefel [10]

The methods in Table 1 are equivalent using exact line search when the objective functions are convex quadratic but the methods behave differently if the general objective function is non-convex $||.||$ denotes the Euclidean norm. Some classical methods such as FR, DY and CD are known for their strong convergence properties but they are usually not computationally powerful. On the other hand, methods like PRP, HS, and LS perform better computationally but may not always converge. Some of these weaknesses associated with the classical methods create a gap for improvement on the existing methods through modification and hybridization to address some of these predicaments. The main objective is to establish the global convergence of these modified or hybrid methods as well as achieve better experimentation results in terms of the number of iteration, function evaluations and CPU time as compared to some existing methods. Researches carried out by the researchers such as Hager and Zhang [9], Hager and Zhang [8], Liu and Feng [12], Wei *et al.* [14], Wei *et al.* [15], Ibrahim and Rohanin [3], Ibrahim and Rohanin [31] Zhang and Zheng [16], Du and Liu [5], Rivaie *et al.* [18], Ibrahim and Rohanin [27] among others focused on modified CG methods. In line with this, Rivaie *et al.* [18] gave their CG coefficient as

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

The works of [19,20] motivated Jiang and Jian [21] to proposed modified CG method called modified Dai-Yuan (MDY) whose aim was to improve the numerical performance of DY method while retaining its good property. Also, the same idea was extended to FR method called modified Fletcher-Reeves (MFR), where the parameters β_k were given by

$$\beta_k^{MDY} = \frac{||g_{k+1}||^2}{\max\{d_k{}^T y_k, \mu|g_{k+1}{}^T d_k|\}}, \tag{7}$$

and

$$\beta_k^{MFR} = \frac{||g_{k+1}||^2}{\max\{||g_{k+1}||^2, \mu|g_{k+1}{}^T d_k|\}}, \tag{8}$$

where $y_k = g_{k+1} - g_k$ and $\mu > 1$. Yao *et al.* [23] extended the idea of the work by Wei *et al.* [22] to the HS method and proposed a CG method as

$$\beta_k^{YWH} = \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{d_{k-1}{}^T (g_{k-1} - g_k)}, \tag{9}$$

under strong Wolfe line search with parameter $\sigma < \frac{1}{3}$. It has been shown that the YWH method can generate sufficient descent directions and converges globally for general objective functions. Wei *et al.* [26] gave a new version of PRP method, WYL for short form, where the parameter β_k is given as

$$\beta_k^{WYL} = \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{||g_{k-1}||^2}. \tag{10}$$

This method possesses the good properties of the PRP method, such as excellent numerical results.

Rivaie *et al.* [28] presented an extension of the CGM proposed by Rivaie *et al.* [29] named $\beta_k^{\mathrm{RMIL+}}$. This method, by using exact line search obviously reduces to β_k^{RMIL} where $g_k{}^T d_{k-1} = 0$. Furthermore, global convergence properties of a new class of CGM for Unconstrained Optimization was another modified CGM presented by Abdelrahman *et al.* [30] on the bedrock of a modified CGM proposed by Rivaie *et al.* [29] which has some good properties such as sufficient descent conditions, global convergence, linear convergence rate and angle conditions. The method presented in [30] used the denominator of the method in [29] with the modification on the numerator. The global convergence of this method was established with exact line search.

The organization of this paper is as follows. In Section 2, we present our proposed β_k called β_k^{IR2} and prove that our method can always generate descent direction. Section 3 presents the global convergence of our method. Section 4 covers the numerical experiments as well as representation of our method against other CG methods using performance profiles by Dolan and More [25].

2 New CG coefficient and its descent property

In contrast to some existing modified methods, we propose a new modified CG method, IR2 for short form, where parameter β_k is given as

$$\beta_k^{\mathrm{IR2}} = \begin{cases} \dfrac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{\mu|g_k{}^T d_{k-1}| + ||g_{k-1}||^2}, & \text{if } \left|1 - \dfrac{g_k{}^T g_{k-1}}{||g_{k-1}||\,||g_k||}\right| < \mu \\[2ex] \dfrac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{d_{k-1}{}^T (d_{k-1} - g_k)}, & \text{otheerwise} \end{cases} \tag{11}$$

The method shows that $0 \le \beta_k^{\mathrm{IR2}} \le \beta_k^{\mathrm{FR}}$. In a nutshell, the parameter μ plays a critical role in ensuring the descent of the method aside the role of $|g_k{}^T d_{k-1}|$. We will show that the proposed formula Equation (11) possesses the sufficient descent property in this section.

Algorithm 2.1. Execution phase of algorithm with β_k^{IR2} as the CG coefficient

Initialization. Select $x_0 \epsilon \boldsymbol{R^n}, \varepsilon > 0, \rho \in (0,1), \sigma \in (\rho, 1), \mu > 1,$ set $k = 0, d_0 = -g_0$;
while *maxiterate* **do**
  Test for convergence;
  if $||g_k|| \leq \varepsilon$ **then**
    Converged;
  end if
  Compute α_k using Inequality (4) and Inequality (5);
  $f(x_k + \alpha d_k) \leq f(x_k) + \rho \alpha {g_k}^T d_k$ and;
  $|{d_k}^T g(x_k + \alpha_k d_k)| \leq {g_k}^T d_k$;
  Variable update;
  $x_{k+1} = x_k + \alpha_k d_k$;
  Compute $f(x_{k+1})$ and $g(x_{k+1})$;
  Computation of CG coefficient β_k

$$\beta_k^{\mathrm{IR2}} = \begin{cases} \dfrac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} {g_k}^T g_{k-1}}{\mu |{g_k}^T d_{k-1}| + ||g_{k-1}||^2}, & \text{if } \left|1 - \dfrac{{g_k}^T g_{k-1}}{||g_{k-1}|| \, ||g_k||}\right| < \mu \\ \dfrac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} {g_k}^T g_{k-1}}{{d_{k-1}}^T (d_{k-1} - g_k)}, & \text{otheerwise} \end{cases}$$

  Generate d_k using Equation (3);
  Set $k = k + 1$;
end while

Lemma 1 *Let the sequences* $\{x_k\}$ *and* $\{d_k\}$ *be generated by Algorithm 2.1 for* β_k^{IR2}*. Then,* ${g_k}^T d_k < 0$ *holds true.*

Proof We proceed by induction to arrive at the conclusion. It is obvious to have ${g_1}^T d_1 = -||g_1||^2 < 0,$ if $k = 1$. Assume that ${g_{k-1}}^T d_{k-1} < 0$ holds true for $k - 1$, to obtain ${g_k}^T d_k < 0$ particularly for our method (β_k^{IR2}).
From the search direction, we have

$$g_k^T d_k = -||g_k||2 + \beta k \, g_k^T d_{k-1} \leq -||g_k||^2 + |\beta k||g_k^T d_{k-1}|. \tag{12}$$

Case (i) If $\beta_k^{\mathrm{IR2}} = \dfrac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} {g_k}^T g_{k-1}}{\mu |{g_k}^T d_{k-1}| + ||g_{k-1}||^2}$. Note that ${d_{k-1}}^T (d_{k-1} - g_k) = {d_{k-1}}^T g_k - {d_{k-1}}^T g_{k-1}$ which implies ${d_{k-1}}^T g_k + ||g_{k-1}||^2 < ||g_{k-1}||^2$, that is, $||g_{k-1}||^2 > 0$. It follows from Eq. (3)

$$\begin{aligned} {d_k}^T g_k &= -||g_k||^2 + \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} {g_k}^T g_{k-1}}{\mu |{g_k}^T d_{k-1}| + ||g_{k-1}||^2} {g_k}^T d_{k-1} \\ &\leq -\left||g_k|\right|^2 + \frac{\left||g_k|\right|^2 - \frac{\left||g_k|\right|}{\left||g_{k-1}|\right|} {g_k}^T g_{k-1}}{\mu |{g_k}^T d_{k-1}| + \left||g_{k-1}|\right|^2} |{g_k}^T d_{k-1}| \qquad (13) \\ &\leq -\left||g_k|\right|^2 + \frac{\left||g_k|\right|^2 - \frac{\left||g_k|\right|}{\left||g_{k-1}|\right|} {g_k}^T g_{k-1}}{\mu |{g_k}^T d_{k-1}|} |{g_k}^T d_{k-1}| \\ &= -\left||g_k|\right|^2 + \frac{\left||g_k|\right|^2 - \frac{\left||g_k|\right|}{\left||g_{k-1}|\right|} {g_k}^T g_{k-1}}{\mu} \qquad (14) \end{aligned}$$

Since $\beta_k^{\text{IR2}} \neq 0\ \beta_k$ and $g_k{}^T g_{k-1} > 0$, we have $0 < cos\theta_k < 1$ where θ_k is the angle between g_k and g_{k-1}

$$\frac{-\mu||g_k||^2 + ||g_k||^2 - ||g_k||^2 cos\theta_k}{\mu} = \frac{-(\mu - 1 + cos\theta_k)||g_k||^2}{\mu} < 0 \tag{15}$$

For $\mu \geq 1$ then $g_k{}^T d_k < 0$ holds $\forall\, k \geq 1$.

Case(ii) If $\beta_k^{\text{IR2}} = \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{d_{k-1}{}^T (d_{k-1} - g_k)}$. Note that $d_{k-1}{}^T (d_{k-1} - g_k) = ||d_{k-1}||^2 - g_k{}^T d_{k-1,}$, that is, $\|d_{k-1}\|^2 > 0$. It follows from Eq. (3)

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

$$\leq -||g_k||^2 + \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{-g_k{}^T d_{k-1}} g_k{}^T d_{k-1} = -||g_k||^2 - ||g_k||^2 + \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}. \tag{16}$$

Since $\beta_k^{\text{IR2}} \neq 0$ and $g_k^T g_{k-1}$, we have $0 < cos\theta_k < 1$ where θ_k is the angle between g_k and g_{k-1}, we have

$$= -2||g_k||^2 + ||g_k||^2 cos\theta_k = -(2 - cos\theta_k)||g_k||^2 < 0. \tag{17}$$

Lemma 2 *The relation* $0 \leq \beta_k^{\text{IR2}} \leq \frac{g_k{}^T d_k}{g_{k-1}{}^T d_{k-1}}$ *holds for any* $k \geq 1$.

Proof from Eq. (10),

$$0 \leq \beta_k^{\text{IR2}} = \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{\mu|g_k{}^T d_{k-1}| + ||g_{k-1}||^2} \leq \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{||g_{k-1}||^2} = \beta_k^{\text{WYL}}..$$

If $\beta_k = \beta_k^{\text{WYL}}$, It follows from Eq. (3),

$$g_k{}^T d_k = -||g_k||^2 + \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{||g_{k-1}||^2} g_k{}^T d_{k-1}.$$

Since $\beta_k^{\text{WYL}} \neq 0$ and $g_k^T g_{k-1} > 0$, we have $0 < cos\theta_k < 1$ where θ_k is the angle between g_k and g_{k-1}. From Eq. (3), we have

$$g_k{}^T d_k \leq -||g_k||^2 + \frac{||g_k||^2 g_{k-1}{}^T d_{k-1} - ||g_k||^2 g_{k-1}{}^T d_{k-1} cos\theta_k}{||g_{k-1}||^2}$$

$$= \frac{-||g_k||^2 ||g_{k-1}||^2 + ||g_k||^2 g_{k-1}{}^T d_{k-1} - ||g_k||^2 g_{k-1}{}^T d_{k-1} cos\theta_k}{||g_{k-1}||^2}$$

$$= \frac{||g_k||^2 g_{k-1}{}^T d_{k-1} + ||g_k||^2 g_{k-1}{}^T d_{k-1} - ||g_k||^2 g_{k-1}{}^T d_{k-1} cos\theta_k}{||g_{k-1}||^2} = \frac{(2 - cos\theta_k)||g_k||^2 g_{k-1}{}^T d_{k-1}}{||g_{k-1}||^2} < 0. \tag{18}$$

Therefore, from Eq. (18) we have

$$\beta_k^{\text{IR2}} \leq \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{||g_{k-1}||^2} = \frac{(2 - cos\theta_k)||g_k||^2}{||g_{k-1}||^2} < \frac{g_k{}^T d_k}{g_{k-1}{}^T d_{k-1}}. \tag{19}$$

Case(ii) $\beta_k^{\text{IR2}} = \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{d_{k-1}{}^T (d_{k-1} - g_k)} \leq \frac{g_k{}^T d_k}{g_{k-1}{}^T d_{k-1}}$, from Eq. (3), we have

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

Let θ_k be the angle between g_k and g_{k-1} where $0 < cos\theta_k < 1$, then

$$\leq -||g_k||^2 + \frac{||g_k||^2 g_{k-1}{}^T d_{k-1} - ||g_k||^2 cos\theta_k d_{k-1}{}^T g_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)}$$

$$= \frac{-||g_k||^2||d_{k-1}||^2 + ||g_k||^2 d_{k-1}{}^T g_k + ||g_k||^2 g_{k-1}{}^T d_{k-1} - ||g_k||^2 cos\theta_k d_{k-1}{}^T g_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)}$$

$$\leq \frac{-||g_k||^2||g_{k-1}||^2 + ||g_k||^2 d_{k-1}{}^T g_{k-1} + ||g_k||^2 g_{k-1}{}^T d_{k-1} - ||g_k||^2 cos\theta_k d_{k-1}{}^T g_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)}$$

$$= \frac{||g_k||^2 g_{k-1}{}^T d_{k-1} + 2||g_k||^2 g_{k-1}{}^T d_{k-1} - ||g_k||^2 cos\theta_k d_{k-1}{}^T g_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)}$$

$$\frac{(3 - cos\theta_k)||g_k||^2 d_{k-1}{}^T g_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)} < 0. \tag{20}$$

Therefore, from Eq. (20), we have

$$\beta_k^{\mathrm{IR2}} = \frac{||g_k||^2 - \frac{||g_k||}{||g_{k-1}||} g_k{}^T g_{k-1}}{d_{k-1}{}^T(d_{k-1} - g_k)} = \frac{(3 - cos\theta_k)||g_k||^2}{d_{k-1}{}^T(d_{k-1} - g_k)} \leq \frac{g_k{}^T d_k}{g_{k-1}{}^T d_{k-1}}. \tag{21}$$

Thus, the proof is complete.

3 Global Convergence

In this section, we highlight the following basic assumptions that aid the global convergence of the proposed method.

Assumption (3.1):

(i) The level set $\mathbf{M} = \{x \in R^n : f(x) \leq f(x_0)\}$ is bounded.

(ii) In some neighborhood **N** of **M**, the function $f(x)$ is continuously differentiable and its gradient is Lipchitz continuous. i.e., there exist a constant $L > 0$ such that

$$\|\nabla f(x) - \nabla f(y)|| \leq L||x - y||, \forall\, x, y \in N. \tag{22}$$

The implication of these assumptions on the function f , there exist a constant $\gamma \geq 0$ such that

$$\|\nabla f(x)|| \leq \gamma, \forall\, x \in N. \tag{23}$$

To prove the global convergence of the proposed method, the result of the following lemma, usually called Zoutendijk conditions are required. For proof, refer to Zoutendijk [17], Dai *et al.* [24].

Lemma 3 *Supposed* **Assumption** (**3.1**) *holds and consider any CG method of the form* $x_{k+1} = x_k + \alpha_k d_k$ *and the direction* $d_{k+1} = -g_{k+1} + \beta_k^{\mathrm{IR2}} d_k$ $d_0 = -g_0$, *where* α_k *satisfies Eq.(4) and Eq.(5). Then,*

$$\sum_{k \geq 1} \frac{(g_k{}^T d_k)^2}{||d_k||^2} < +\infty$$

From Lemma 3, we have the following theorem which present the global convergence of the proposed method.

Theorem 1 *Let* **Assumption** (**3.1**) *holds and the sequence* $\{x_k\}$ *and* $\{d_k\}$ *be generated by*

Algorithm 2.1 with β_k^{IR2}, α_k is obtained by Eq.(4) and Eq.(5). Then

$$\lim_{k\to\infty} inf\, ||g_k|| = 0\,. \tag{25}$$

Proof We proceed by contradiction to arrive at the conclusion.

Suppose that $\lim_{k\to\infty} inf\, ||g_k|| \neq 0$, it implies that there exists $m > 0$ such that

$$||g_k|| \geq m, \forall\, k \geq 0. \tag{26}$$

From Eq. 3, we have

$$(\beta_k^{\mathrm{IR2}} d_{k-1})^2 = (d_k + g_k)^2 \tag{27}$$

it follows from Eq. (27) and Lemma 2

$$\begin{aligned} ||d_k||^2 &= (\beta_k^{\mathrm{IR2}})^2 ||d_{k-1}||^2 - 2{g_k}^T d_k - \left||g_k|\right|^2 \\ &\leq \left(\frac{{g_k}^T d_k}{{g_{k-1}}^T d_{k-1}}\right)^2 \left||d_{k-1}|\right|^2 - 2{g_k}^T d_k - \left||g_k|\right|^2. \end{aligned} \tag{28}$$

Dividing both side of Eq. (28) by $({g_k}^T d_k)^2$ to get

$$\begin{aligned} \frac{||d_k||^2}{({g_k}^T d_k)^2} &\leq \frac{\left||d_{k-1}|\right|^2}{({g_{k-1}}^T d_{k-1})^2} - \frac{2}{{g_k}^T d_k} - \frac{\left||g_k|\right|^2}{({g_k}^T d_k)^2} = \frac{\left||d_{k-1}|\right|^2}{({g_{k-1}}^T d_{k-1})^2} - \left(\frac{1}{||g_k||} + \frac{||g_k||}{{g_k}^T d_k}\right)^2 + \frac{1}{||g_k||^2} \\ &\leq \frac{\left||d_{k-1}|\right|^2}{\left({g_{k-1}}^T d_{k-1}\right)^2} + \frac{1}{\left||g_k|\right|^2}. \end{aligned} \tag{29}$$

Hence

$$\frac{\left||d_k|\right|^2}{\left({g_k}^T d_k\right)^2} \leq \sum_{i=1}^{k} \frac{1}{||g_i||^2} \leq \frac{k}{m^2}, \tag{30}$$

furthermore,

$$\frac{({g_k}^T d_k)^2}{||d_k||^2} \geq m^2 \sum_{k\geq 1} \frac{1}{k} = +\infty\,.$$

$\sum_{k\geq 1}\frac{1}{k}$ is divergent and $\sum_{k\geq 1}\frac{({g_k}^T d_k)^2}{||d_k||^2}$. This contradicts Inequality (24). Thus, the proof is complete.

4 Numerical results

In this section, to compare the numerical strength of our method where $\beta_k = \beta_k^{\mathrm{IR2}}$, we present the experiments of our proposed method against some existing methods in the literature. We consider some compiled test functions by Andrei [1] and Andrei [2] for small, medium and large scales with the specified initial points in those papers and other various initial points to validate the numerical strength of our method versus some methods in existence, using inexact line search conditions (4) and (5) for all methods in this paper for easy comparison with δ = 0.0001 and σ = 0.01.

In carrying out the simulations, the number of iterations, the number of function evaluations and CPU time (t) were put into consideration as parameters to determine the numerical strength of the proposed formula IR2 as compared with FR, DY, and CG-descent. The stopping criterion for all the methods is taken as $||g_k|| \leq \varepsilon$, where $\varepsilon = 10^{-5}$. We implemented all the methods using MATLAB R2015b (8.6.0.267246) in double precision arithmetic on CP computer, intel(R) Core (TM) i7-4790 CPU 3.60 GHz, 2TB HDD and 16.00GB RAM. Tables 2 and 3 contain the test problems with different initial points, a total of sixty-four different functions were considered and tested with different variables from 2 to as much as 100,000 as the case may be which give rise to a set of 870 problems. The computations were forced to stopped either when the number of iterations or function evaluations exceeded the maximum limit set or line search failed to find the next positive step size, that is, failure is recorded and the formula in question cannot locate the optimum solution for a given problem based on the criteria. For iterations, we set 5000 as the maximum while 20000 is the maximum for the number of function evaluations.

Table 2: List of test functions

No	Function	Dim	Initial points
1	Extended Rosenbrock	2	(-1.2,1),2,5,8,10
2	Extended White Holst	2	(-1.2,1),3,20
3	Extended Beale	2,5000	(1,0.8,...,0.8),7
4	Extended Penalty	10,50,10000,20000	9,12,17
5	Perturbed Quadratic	2,4,32,500,1500,100000	-2, 0.5,5,8,13
6	Raydan 1	2,10,100,10000,50000	1,11
7	Raydan 2	20,200,20000,70000	1,3,6,9
8	Diagonal 1	2,8,80,800,8000	1,4,7,10,17
9	Diagonal 2	2,4,12,200,2000	1,10,20
10	Diagonal 3	2,20,500	1,12,14
11	Hager	8,200,500,5000	1,1.5,3.5,4.5,6.5
12	Extended Tridiagonal 1	2,7000,70000	2,4,6,8,10
13	Extended 3 Exponential Terms	20,5000,100000	0.1,0.5
14	Generalized Tridiagonal 2	4,12,12000	-1,15
15	Diagonal 4	2,200,500,5000	1,3,7,11,18
16	Diagonal 5	2,4,200,10000	1.1,1.2,3.1,4.1,5.1
17	Generalized PSC 1	12000	(3,0.1...0.1),6,12,18
18	Extended Block Diagonal BD 1	12000,40000	1.1,3.1
19	Extended Maratos	4	(1.1,0.1...,0.1)
20	Extended Cliff	2,4,12,120,200	-3,3,6
21	Quadratic Diagonal Perturbed	2,4,120,200,10000	0.5,2.5,4.5,6.5,8.5
22	Extended Hiebert	2	0,1,13
23	Quadratic	2,10,200,2000	1,13,15,17
24	Extended Quadratic Penalty	400,4000,7000	1,2,3,4,5
25	Extended Quadratic Penalty 2	100	1,3,7,13,18
26	A Quadratic 2	2,12,40,4000,8000	0.5,3.5,7.5,9.5,12.5
27	Extended EP1	4,40,800,1000	1.5,3.5,5.5,6.5,8.5
28	Extended Tridiagonal 2	10,200,2000,100000	1,4,8,12,20
29	TRIDIA	4,8,100,400,4000	1,2,3,4,5
30	ARWHEAD	4,40,400,4000	1
31	NONDIA	4,12,40,120,12000	-2,-1,1,2,3
32	NONDQUAR	5,50,500,5000	(1,-1,...1),-1,1,3
33	DQDRTIC	5,50,5000	(1,-1,...1),3,5,6
34	EG2	4	7
35	DIXMAANA	4,12,32,4000,10000	2,3,4,8
36	DIXMAANB	4,12,400,4000,10000	2,3,4,8,13
37	DIXMAANC	4,12,32,400,10000	2,3,4,8
38	DIXMAAND	4,12,32,4000,10000	2,3,4,8,13
39	DIXMAANL	4,12,400,4000,10000	2,3,4,8,13
40	Partial Perturbed Quadratic	4,12,120,12000	0.5,1.5,5.5,7.5
41	Broyden Tridiagonal	4,400,4000,40000	-3,-1,1,3
42	Almost Perturbed Quadratic	2,10,2000,8000	0.5,2.5,4.5,6.5

43	Tridiagonal Perturbed Quadratic	6,12,600,6000	0.5,2.5,6.5,9.5
44	HIMMELBHA	4,1200	(0,2,...,2)
45	STAIRCASE	4,32,400,4000,40000	1,2,4,7
46	LIARWHD	4,40,400,4000,40000	4,5,6,7
47	DIAGONAL 6	2,10,100,1000,10000	1,3,5,9
48	DIXON3DQ	4,40,400	-1,1,2,3

Table 3: Continue...

No	Function	Dim	Initial points
49	DENSCHNA	400	3,6
50	DENSCHNB	4000	1,3,4
51	DENSCHNC	40,4000	2
52	SINQUAD	4,40,400,4000	0.1,0.3,0.5,0.7
53	BIGGSB 1	4,32,400,4000	0,2,3,5
54	Extended Block DiagonalBD 2	32,40,4000	2
55	Generalized quartic GQ1	4,40,400,4000	1,2,3,5
56	Diagonal 7	2,10,200,2000	1
57	Diagonal 8	2,20,200,2000	1
58	Full Hessian	2	1
59	Generalized quartic GQ2	4,32,400,4000	1,2,3,4
60	EXTROSN B	4,16,40,4000,40000	3,5,7
61	ARGLINB	4,40,400,4000	(0.01,0.001,...,0.001),1.5,2.5,3.5
62	FLETCHCR	4,32,40,400,4000	0.5,1.5,2.5,3.5
63	HIMMELB G	4,16,400,4000	1.5,2.5,5.5,7.5
64	DIAGONAL 9	2,10,200,500,1000	1,2,4,6

4.1 Performance comparisons

We report the numerical results of IR2 versus FR, DY, and CG-descent methods and show clearly through graphs the performance difference between the IR2 versus FR, DY, and CG-descent methods, we employ the Performance Profile by Dolan and More [25] to compare the performance base on the number of iterations, number of function evaluations and CPU time(seconds). The performance of the methods is produced in Figures 1 through 9.

Suppose S is the set of solvers on the test set P of problems. Assume, S consists of n_s solvers, P consists of n_p problems. For every problem $p \in P$ and solver $s \in S$, denotes $t_{p,s}$ as the computing time (or number of iteration or number of function evaluation etc.)required to solve problem $p \in P$ by $s \in S$. Then, the comparison between different solvers is based on the performance ratio given by

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s}: s \in S\}}$$

Assuming that a parameter $r_M \geq r_{p,s}$ for all p, s is chosen if and only if solver s does not solve problem p. Define

$$\rho_s(t) = \frac{1}{n_p} \text{size}\{p \in P: \log_{p,s} r \leq t\}.$$

where size A refers to the number of elements in set A, then $\rho_s(t)$ is the probability for the solver $s \in S$ that a performance ratio $r_{s,p}$ is within a factor $t \in R^n$. ρ_s is the cumulative distribution function for the performance ratio. The value of $\rho_s(1)$ is the probability that the solver will win over the rest of the solvers.

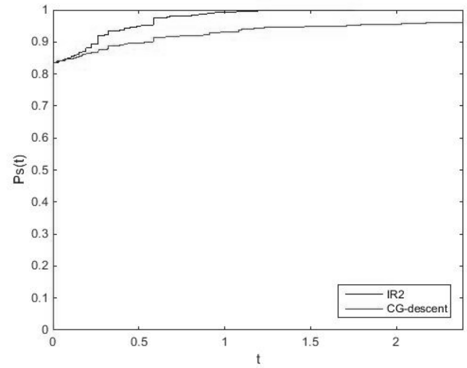

Figure 1: Performance Profile based on Iteration for IR2 versus CG-descent

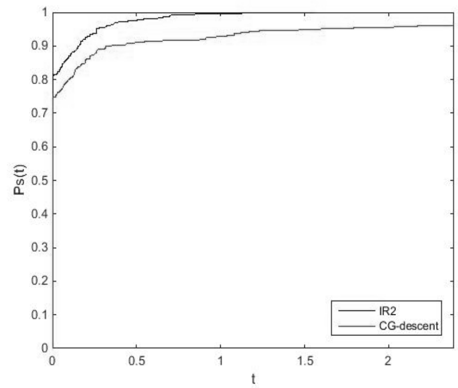

Figure 2: Performance Profile based on Function evaluation for IR2 versus CG-descent

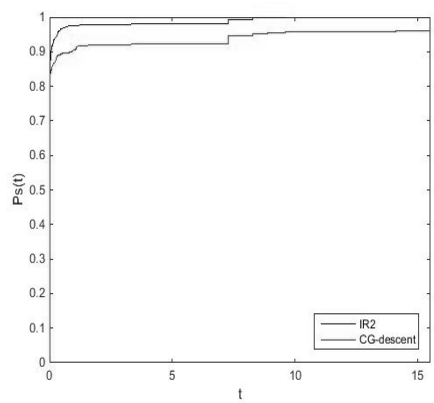

Figure 3: Performance Profile based on CPU time for IR2 versus CG-descent

The performance of proposed CG coefficient (IR2) against CG-descent as presented in the Figures 1 through 3 show that the proposed method is effective and efficient bearing in mind that CG-descent method is a recent well known and promising CG coefficient proposed by Hager and Zhang [9]. Considering the number of iterations from Figure 1, it is clear that

IR2 outperforms CG-descent method for the test problems since the graph is at the topmost all through, thereby making it to more preferred solver since it has large probability $\rho_s(1)$. Observed that from Figure 2, the IR2 is the best solver because it performs few number of function evaluations as compared to the other method to attain the optimum solutions for the test problems. Furthermore, the estimated time of the proposed method is clearly better, that is, the proposed formula finds solutions to the test problems within shortest possible time as compared to CG-descent formula and therefore the solver IR2 is effective and efficient.

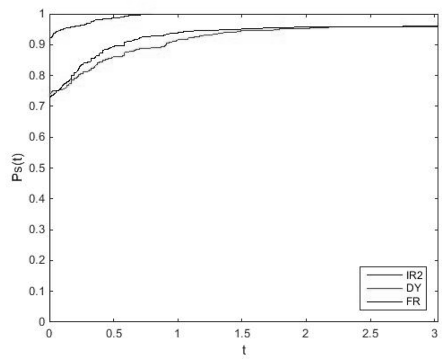

Figure 4: Performance Profile based on Iteration for IR2 versus DY and FR

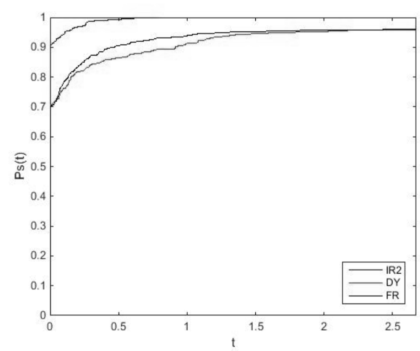

Figure 5: Performance Profile based on Function evaluation. IR2 versus DY and FR

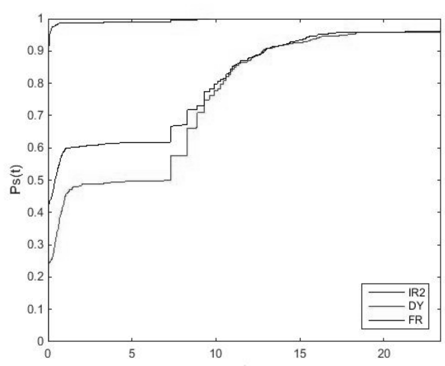

Figure 6: Performance Profile based on CPU time. IR2 versus DY and FR

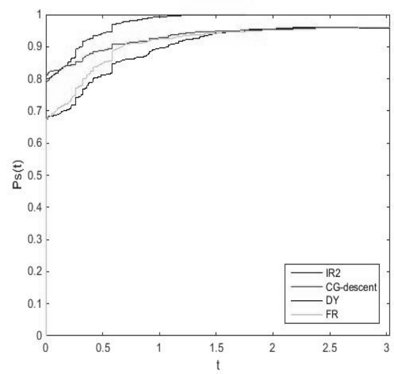

Figure 7: Performance Profile based on Iteration for IR2 versus CG-descent, DY and FR

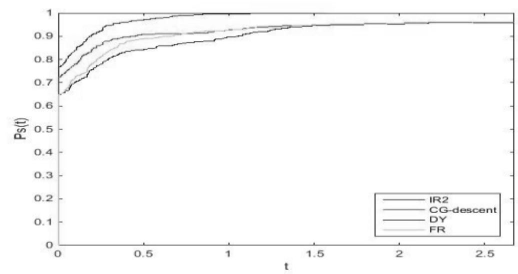

Figure 8: Performance Profile based on Function evaluation. IR2 versus CG-descent, DY and FR

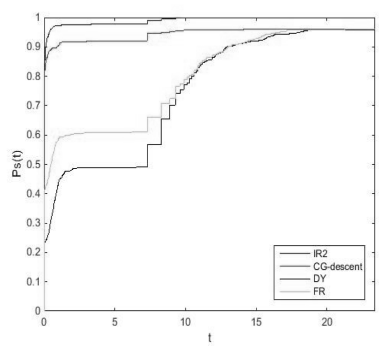

Figure 9: Performance Profile based on CPU time. IR2 versus CG-descent, DY and FR.

The Figures 4, 5 and 6 represent the performance of the proposed formula IR2 compared with the FR and DY methods. All the three figures from iteration to the CPU time, it is clear that the proposed formula is efficient since it outperforms all the two methods. Meanwhile, Figures 7,8 and 9 show the combine performance of all the four formulas under consideration, that is, the proposed formula versus the FR, DY, and CG-descent. The graphs demonstrate the strength of the IR2.

The proposed CG coefficient has advantage over the methods compared with since for the test problems been considered, some methods such as FR, DY and CG could not solve all problems due to line search failure, that is, step length recorded was not significance enough for search direction and therefore for such methods we record not a number (NAN) for the purpose graphical comparison of the methods while at times failure was recorded if the method could not achieve the desired result within the maximum iterations and function evaluations set.

5 Conclusion

In this paper, we proposed another type of modified CG method for solving unconstrained optimization problems. The proposed method generates descent condition at every iteration under Wolfe line search condition. Under line search condition (4) and (5), we established the global convergence of the proposed method. For the purpose of the experiment in this paper, we take $\mu = 9.5$. The simulation results of the proposed method shown to be effective when compared to some CG methods (FR, DY, and CG-descent). We employed Performance Profiles by Dalon and More' to show the effectiveness of our proposed method against some existing methods.

References

Andrei, N. An unconstrained optimization test functions collection, Adv. Model. Optim. 2008. 10(1): 147-161.

Andrei, N. Test functions for unconstrained optimization, Research Institute for informatics, Center for Adv. Model. Optim. 2004. 1-10.

Ibrahim, A. and Rohanin, A. Convergence Analysis of a New Conjugate Gradient Method for Unconstrained Optimization, AMS. 2015. 9(140): 6969-6984.

Dai, Y.H. and Yuan, Y. A nonlinear conjugate gradient method with a strong global convergence property, SIAM J. Optim. 1999. 10(1): 177-182.

Du, X. and Liu, J. Global convergence of a spectral HS conjugate gradient method, Proc. Eng. 2011. 15: 1487-1492.

Fletcher, R. Practical Method of Optimization, Vol. 1, Unconstrained Optimization, John Wiley and Sons, New York. 1987.

Fletcher, R. and Reeves, C. M. Function minimization by conjugate gradients, comput. J. 1964. 7(2): 149-154.

Hager, W. W. and Zhang, H. A survey of nonlinear conjugate gradient method, Pac. J. Optim. 2006. 2(1): 35-58.

Hager, W. W. and Zhang, H. A new conjugate gradient method with guaranteed descent and an efficient line search, SIAM J. Optim.2005. 16(1): 170-192.

Hestenes, M. R. and Stiefel, E. Methods of conjugate gradients for solving linear systems, Journal of Research of the National Bureau of Standards. 1952. 49: 409-436.

Liu, Y. and Storey, C. Efficient generalized conjugate gradient algorithms part 1: Theory, J. Comput. Appl. Math. 1992. 69: 129137.

Liu, J. and Feng, Y. Global convergence of a modified ls nonlinear conjugate gradient method, Proc. Eng. 2011. 15: 4357-4361.

Polyak, B. T. The conjugate gradient method in extremal problems, USSR Comput. Math. Math. Phys. 1969. 9(4): 94-112.

Wei, Z. Li, G. and Qi, L. New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems, Appl. Math. Comput. 2006. 179(2): 407-430.

Wei, Z. Huang, H. and Tao, Y. A modified Hestenes-Stiefel conjugate gradient method and its

convergence, J. Math. Res. Exposition. 2010. 30(2): 297-308.
Zhang, Y. Zheng, H. and Zhang, C. Global convergence of a modified PRP conjugate gradient method, Proc. Eng.2012. 31: 986-995.
Zoutendijk, G. Nonlinear programming, Integer and Nonlinear Programming. 1970. 143(1): 37-86.
Rivaie, M., Abashar, A., Mamat, M., and Mohd, I. The convergence properties of a new type of conjugate gradient methods, AMS. 2015. 9(54): 2671-2682.
Dai, Z. and Wen, F. Another improved Wei–Yao–Liu nonlinear conjugate gradient method with sufficient descent property, Appl. Math. Comput. 2014. 1(8): 33-44.
Dai, Y. A nonmonotone conjugate gradient algorithm for unconstrained optimization, J. Syst. Sci. Complex. 2002. 2(15): 139-145.
Jiang, X.Z. and Jian, J.B. A sufficient descent Dai–Yuan type nonlinear conjugate gradient method for unconstrained optimization problems, Nonlinear Dynam. 2013. 1(72): 101-112.
Wei, Z., Yao, S. and Liu, L. The convergence properties of some new conjugate gradient methods, Appl. Math. Comput. 2006. 2(183): 1341-1350.
Yao, S., Wei, Z. and Hai, H. A note about WYLs conjugate gradient method and its applications, Appl. Math. Comput. 2007. 2(191): 381-388.
Dai, Y., Han, J. , Liu, G. , Sun, D., Yin, H. and Yuan, Y.X. Convergence properties of nonlinear conjugate gradient methods, SIAM J. Optim. 2000. 2(10): 345-358.
Dolan, E.D. and More, J.J. Benchmarking optimization software with performance profiles, Math. Program. 2002. 2(91): 201-213.
Wei, Z., Yao, S. and L. Liu, The convergence properties of some new conjugate gradient methods, Appl. Math. Comput. 2006. 2(183): 1341-1350.
Ibrahim, A. and Rohanin, A. Global Convergence Analysis of a Nonlinear Conjugate Gradient Method for Unconstrained Optimization Problems, IJST. 2016. 9(41), DOI: 10.17485/ijst/2016/v9i41/90175.
Rivaie, M., Mamat, M., and Abashar, A. A new class of nonlinear conjugate gradient coefficients with exact and inexact line searches, Appl. Math. Comput. 2015. 268: 11521163.
Rivaie, M., Mamat, M., June, L.W. and Mohd, I. A new class of nonlinear conjugate gradient coefficients with global convergence properties, Appl. Math. Comput. 2012. 218:11323-11332.
Abashar, A., Mamat, M., Rivaie M. and Mohd, I. Global Convergence Properties of a New Class of Conjugate Gradient Method for Unconstrained Optimization, AMS. 2014. 8(67): 3307 - 3319.
Ibrahim, A. and Rohanin, A. Global convergence analysis of a new hybrid conjugate gradient method for unconstrained optimization problems, MJFAS. 2017. 2(13): 40-48.