Available online at www.refaad.com
Gen. Lett. Math., 11(1) (2021), 1- 9
Research Article

General Letters in Mathematics (GLM)

Journal Homepage: http://www.refaad.com/views/GLM/home.aspx

ISSN: 2519-9277 (Online) 2519-9269 (Print)

Global convergence of new three terms conjugate gradient for unconstrained optimization

Ahmed Anwer Mustafa[a*], Salah Gazi Shareef [b]

[a] Department of Mathematics, Faculty of Education, University of Zakho, Kurdistan Region- Iraq
[b] Department of Mathematics, Faculty of Science, University of Zakho, Kurdistan Region- Iraq
Emails: [a]ahmed.mustafa@uoz.edu.krd , [b]salah.gazi2014@gmail.com

Abstract

In this paper, a new formula of β_k is suggested for the conjugate gradient method of solving unconstrained optimization problems based on three terms and step size of cubic. Our new proposed CG method has descent condition, sufficient descent condition, conjugacy condition, and global convergence properties. Numerical comparisons with two standard conjugate gradient algorithms show that this algorithm is very effective depending on the number of iterations and the number of functions evaluated.

Keywords: *Conjugate gradient method; Descent Condition; Sufficient Descent; conjugacy condition; Global Convergent; Unconstrained Optimizations.*

2010 MSC: *42A50, 41A60, 90C52.*

1. Introduction

The conjugate gradient method (CG) plays an important role in solving the unconstrained optimization problem. In general, the method has the following form

$$\text{Min } f(x)\ x \in R^n \tag{1.1}$$

Where $f : R^n \to R$ is continuously differentiable. The CG method is an iterative method of the form,

$$x_{k+1} = x_k + \alpha_k d_k,\ k = 0,1,2,\ldots \tag{1.2}$$

Where x_k is the current iterate point, $\alpha_k > 0$ is a step size and d_k is the search direction. Basically $d_k$is defined by

$$d_k = \begin{cases} -g_k, & k = 0 \\ -g_{k+1} + \beta_k d_k, & k \geq 1 \end{cases} \tag{1.3}$$

Where g_k is the gradient of $f(x)$ at the point x_k. $\beta_k \in R$ is known as conjugate gradient and different β_k will yield different CG methods. Some well-known formulas are given as follows:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \tag{1.4}$$

$$\beta_k^{FR} = \frac{g_{k+1}^T g_k}{g_k^T g_k} \tag{1.5}$$

*Corresponding author
Email address: ahmed.mustafa@uoz.edu.krd (Ahmed Anwer Mustafa)
doi: 10.31559/glm2021.11.1.1
Received: 15 Jul 2021 Revised: 18 Aug 2021 Accepted: 25 Aug 2021

$$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{g_k^T g_k} \tag{1.6}$$

$$\beta_k^{DX} = -\frac{g_{k+1}^T g_k}{d_k^T g_k} \tag{1.7}$$

$$\beta_k^{BA2} = \frac{y_k^T y_k}{g_k^T g_k} \tag{1.8}$$

$$\beta_k^{LS} = \frac{g_{k+1}^T y_k}{-d_k^T g_k} \tag{1.9}$$

$$\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{d_k^T y_k} \tag{1.10}$$

$$\beta_k^{RMIL} = \frac{g_k^T y_k}{d_k^T (d_k - g_{k+1})} \tag{1.11}$$

$$\beta_k^{AMRI} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|}|g_{k+1} g_k|}{\|d_k\|^2} \tag{1.12}$$

$$\beta_k^{New} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - \frac{g_{k+1}^T v_k}{d_k^T y_k} + \mu \frac{g_{k+1}^T d_k}{\|g_k\|^2}, \text{ where } \mu \in (0,1) \tag{1.13}$$

Where g_k and g_{k+1} are the gradients of $f(x)$ at the point x_k and x_{k+1} respectively. The above corresponding methods, HS is known as Hestenes and Steifel [6], FR is Fletcher and Reeves [5], PR is Polak and Ribiere [12], DX is Dixon [4], BA3 is AL - Bayati, A.Y. and AL-Assady [1], LS is Liu and Storey [10], DY is Dai and Yuan [3], RMIL is Rivaie, Mustafa, Ismail and Leong [13], AMRI denotes Abdelrhaman Abashar, Mustafa Mamat, Mohd Rivaie and Ismail Mohd [2] and lastly Hussein Ageel and Salah Gazi [9].

In this paper, we propose our new β_k^{AA1} and compared its performance with standard formulas of (HS) and (PR) methods.

The remaining sections of the paper are arranged as follows. in section 2, the new conjugate gradient formula and algorithm method presented, in section 3, we showed the descent condition, sufficient descent condition, conjugacy condition, and the global convergence proof of our new method. In section 4 numerical results, percentages, graphics, and discussion. Lastly, In section 5 conclusion, In section acknowledgments.

2. New proposed method and algorithm

2.1 Derivation of The New Method

In this section, we will derive our suggestion based on a three-term conjugate gradient method. J. Liu and X. Wu proposed a three-term conjugate gradient method, in which the search direction d_k has the form,

$$d_{k+1} = -g_{k+1} - \beta_k v_k + \delta_k d_k \tag{2.1}$$

Where $\beta_k = \frac{g_{k+1}^T v_k}{y_k{}^T v_k}$ and $\delta_k = \frac{g_{k+1}^T y_k}{y_k{}^T v_k}$, see [11]. Also, there are many three-term conjugate gradient algorithms suggested, for instance:

$$d_{k+1} = -g_{k+1} + \beta_k^{HS} d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k, \text{ see [15]} \tag{2.2}$$

$$d_{k+1} = -g_{k+1} + \beta_k^{PR} d_k - \frac{g_{k+1}^T d_k}{g_k^T g_k} y_k, \text{ see [14]} \tag{2.3}$$

$$d_{k+1} = -g_{k+1} + \left(\beta_k^{PR} + (\delta - 1)\frac{g_{k+1}^T d_k d_k^T y_k}{\|d_k\|^2 \|g_k\|^2}\right) d_k, \text{ where } 0 < \delta < 1 \text{ see [8]} \tag{2.4}$$

We suggest a new three-term conjugate gradient as following:

$$d_{k+1} = -g_{k+1} + \beta_k^{AA1} d_k - \lambda\mu\, y_k \tag{2.5}$$

Where $\lambda < 1$ and $\mu = \frac{g_{k+1}^T d_k}{d_k^T y_k}$

Now, from equation (2.2) and equation (2.5), we have

$$\beta_k^{HS} d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k = \beta_k^{AA1} d_k - \lambda\mu\, y_k$$

$$\Rightarrow \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k = \beta_k^{AA1} d_k - \lambda \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k$$

Multiplying both sides of the above equation by d_k^T , we obtain

$\frac{g_{k+1}^T y_k}{d_k^T y_k} d_k^T d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} d_k^T y_k = \beta_k^{AA1} d_k^T d_k - \lambda \frac{g_{k+1}^T d_k}{d_k^T y_k} d_k^T y_k$ Since $d_k^T y_k$ is a scalar, then

$$\frac{g_{k+1}^T y_k}{d_k^T y_k} d_k^T d_k - g_{k+1}^T d_k = \beta_k^{AA1} d_k^T d_k - \lambda g_{k+1}^T d_k$$

Dividing both sides of the above equation by $d_k^T d_k$, we have

$$\frac{g_{k+1}^T y_k}{d_k^T y_k} - \frac{g_{k+1}^T d_k}{d_k^T d_k} = \beta_k^{AA1} - \lambda \frac{g_{k+1}^T d_k}{d_k^T d_k}$$

This implies that

$$\beta_k^{AA1} = \beta_k^{HS} - \frac{g_{k+1}^T d_k}{d_k^T d_k} + \lambda \frac{g_{k+1}^T d_k}{d_k^T d_k}$$

So

$$\beta_k^{AA1} = \beta_k^{HS} + (\lambda - 1) \frac{g_{k+1}^T d_k}{d_k^T d_k} \tag{2.6}$$

Where $\lambda < 1$.

We programmed the new method β_k^{AA1} and compared with the numerical results of the methods Hestenes and Steifel, Polakand Ribiere and we noticed superiority of the new method (AA1) that proposed on the methods of (HS) and (PR).

2.2 Algorithm of the AA1 Method

Step (1): Given $x_0 \in R^n$,$\varepsilon = 10^{-5}$,$\gamma < 1$
Step (2): Set k = 0, Compute $f(x_0)$, g_0, $d_k = -g_k$
Step (3): Compute $\alpha_k > 0$ satisfying the strong Wolfe condition

$$f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k {g_k}^T d_k$$
$$|\nabla f(x_k + \alpha_k d_k)^T d_k| \leq c_2 |{g_k}^T d_k|$$

Where $0 < c_1 < c_2 < 1$

Step (4): Evaluate $x_{k+1} = x_k + \alpha_k d_k$, $g_{k+1} = \nabla f(g_{k+1})$, If $\|g_{k+1}\| < \varepsilon$ stop.

Step (5): Calculate $d_{k+1} = -g_{k+1} + \beta_k^{AA1} d_k$

$$\beta_k^{AA1} = \beta_k^{HS} + (\lambda - 1) \frac{g_{k+1}^T d_k}{d_k^T d_k}$$

Step (6): If $|{g_{k+1}}^T g_k| > 0.2\|g_{k+1}\|^2$ go to step (2) else $k = k + 1$, go to step (3)

3. The Global convergent Analysis of the New Method

The convergence properties of β_k^{AA1} will be studied. For an algorithm to converge, it is necessary to show that the descent condition, sufficient descent condition, conjugacy condition, and the global convergence properties.

Theorem 3.1: Assume that the sequence $\{x_k\}$ is generated by (1.2), then the search direction in (1.3) with new conjugate gradient method (2.6) satisfy the descent condition, that is $g_{k+1}^T d_{k+1} \leq 0$ with exact and inexact line search.
Proof: - From (1.3) and (2.6) we have

$$d_{k+1} = -g_{k+1} + \left(\beta_k^{HS} + (\lambda - 1) \frac{g_{k+1}^T d_k}{d_k^T d_k}\right) d_k \tag{3.1}$$

Multiply both sides of the above equation by g_{k+1}^T , to obtain

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{d_k^T y_k} g_{k+1}^T d_k + (\lambda - 1) \frac{g_{k+1}^T d_k}{d_k^T d_k} g_{k+1}^T d_k \tag{3.2}$$

This implies that

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{d_k^T y_k} g_{k+1}^T d_k + (\lambda - 1) \frac{\left(g_{k+1}^T d_k\right)^2}{\|d_k\|^2} \tag{3.3}$$

If the step length α_k is chosen by an exact line search which requires $d_k^T g_{k+1} = 0$. Then the proof is complete. If the step length α_k is chosen by inexact line search which requires $d_k^T g_{k+1} \neq 0$ the first two terms of equation (3.3) are less than or equal to zero because the parameter of (HS) satisfies the descent condition, and the third term is less than or equal to zero, since $\frac{\left(g_{k+1}^T d_k\right)^2}{\|d_k\|^2} \geq 0$ and $\lambda < 1$, so,

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{d_k^T y_k} g_{k+1}^T d_k + (\lambda - 1) \frac{\left(g_{k+1}^T d_k\right)^2}{\|d_k\|^2} \leq 0. \blacksquare$$

Theorem 3.2: Assume that the sequence $\{x_k\}$ is generated by (1.2), then the search direction in (1.3) with new conjugate gradient method (2.6) satisfy the sufficient descent condition, that is $g_{k+1}^T d_{k+1} \leq -C\|g_{k+1}\|^2$ with exact and inexact line search.

Proof: From (1.3) and (2.6) we have

$$d_{k+1} = -g_{k+1} + \left(\beta_k^{HS} + (\lambda - 1)\frac{g_{k+1}^T d_k}{d_k^T d_k}\right) d_k \tag{3.4}$$

Multiply both sides of the above equation by g_{k+1}^T , to obtain

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{d_k^T y_k} g_{k+1}^T d_k + (\lambda - 1)\frac{\left(g_{k+1}^T d_k\right)^2}{\|d_k\|^2} \tag{3.5}$$

Now, since the parameter of (HS) is satisfies the descent condition, then the above equation becomes

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

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

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

Let $C = (1 - \lambda)\frac{\left(g_{k+1}^T d_k\right)^2}{\|d_k\|^2\|g_{k+1}\|^2}$ which is positive, then $g_{k+1}^T d_{k+1} \leq -C\|g_{k+1}\|^2$ ■

Theorem 3.3: Assume that the sequence $\{x_k\}$ is generated by (1.2), then the search direction in (1.3) with the new conjugate gradient method (2.6) satisfies the conjugacy condition.

Proof: From (1.3) and (2.6) and multiply both sides by y_k^T , we have

$$d_{k+1}^T y_k = -g_{k+1}^T y_k + \left(\frac{g_{k+1}^T y_k}{d_k^T y_k} + (\lambda - 1)\frac{g_{k+1}^T d_k}{d_k^T d_k}\right) d_k^T y_k \tag{3.6}$$

Implies that

$$d_{k+1}^T y_k = -g_{k+1}^T y_k + \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k^T y_k + (\lambda - 1)\frac{g_{k+1}^T d_k}{d_k^T d_k} d_k^T y_k \tag{3.7}$$

Since $d_k^T y_k$ is Scalar, then

$$d_{k+1}^T y_k = (\lambda - 1)\frac{g_{k+1}^T d_k}{d_k^T d_k} d_k^T y_k \Rightarrow d_{k+1}^T y_k = (\lambda - 1)\frac{g_{k+1}^T v_k}{\alpha_k d_k^T d_k} d_k^T y_k \text{ Since } \alpha_k d_k = v_k \text{ Hence}$$

$$d_{k+1}^T y_k = -(1 - \lambda)\frac{d_k^T y_k g_{k+1}^T v_k}{\alpha_k\|d_k\|^2} \text{ Since } (1 - \lambda)\frac{d_k^T y_k}{\alpha_k\|d_k\|^2} > 0, \text{ Let } t = (1 - \lambda)\frac{d_k^T y_k}{\alpha_k\|d_k\|^2}$$

$$d_{k+1}^T y_k = -t g_{k+1}^T v_k = 0$$ ■

Global Convergent:

Assuming that the following assumptions are frequently required to establish the convergence of the Procedure for nonlinear conjugate gradients, see [6].

Assumptions:

(i) At the beginning point x_0, f is limited below on the level set R^n continuous and differentiable in a neighborhood N of the level set $S = \{x \in R^n : f(x) \leq f(x_0)\}$.

(ii) In N, the gradient $g(x)$ is Lipschitz continuous, hence for any $x, y \in N$, there exists a constant $L > 0$ such that $\|g(x) - g(y)\| \leq L\|x - y\|$.

We have the following theorem when it was shown using these assumptions. [16].

Theorem 3.4: Let us the assumption is correct. Consider any gradient that is conjugated from (1.3) where d_k is a descent search direction and we use α_k in both situations, precise and inexact line searches are used. Then comes the condition called as Zoutendijk condition holds

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

For proof see [16]. The following conjugate gradient techniques convergence theorem may be constructed from the above information.

Theorem 3.5: Assume that the assumptions are correct. Consider any conjugate gradient strategy of the sort (1.2) and (1.3) where α_k is acquired through both exact and inexact line searches, and d_k is the descent search direction than either

$$\lim_{k\to\infty}\|g_k\| = 0 \quad \text{Or} \quad \sum_{k\geq 1}\frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} < \infty$$

Proof:

Contradiction is used to prove theorem 3.5. It is false if theorem 3.5, then there exists a constant μ>0, such that

$$\|g_i\| \geq \mu, \forall i \geq 0. \tag{3.8}$$

From (1.3) and (2.6), we get $d_{k+1} + g_{k+1} = \beta_k^{AA1} d_k$ (3.9)

Squaring the above equation, we get

$$\|d_{k+1}\|^2 = (\beta_k^{AA1})^2 \|d_k\|^2 - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2 \tag{3.10}$$

Divide the two sides of the equation (3.10) by $(g_{k+1}^T d_{k+1})^2$, therefore we end up with

$$\frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} = \frac{(\beta_k^{AA1})^2 \|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} - \frac{2}{g_{k+1}^T d_{k+1}} - \frac{\|g_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2}$$

$$= \frac{(\beta_k^{AA1})^2 \|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} - \left(\frac{1}{\|g_{k+1}\|} + \frac{\|g_{k+1}\|}{g_{k+1}^T d_{k+1}}\right)^2 + \frac{1}{\|g_{k+1}\|^2}$$

$$|\leq \frac{(\beta_k^{AA1})^2 \|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2}$$

Substitute β_k^{AA1}, we have

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

$$\Rightarrow \frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \leq \frac{(g_{k+1}^T y_k)^2}{(d_k^T y_k)^2} \frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + 2(\lambda - 1)\frac{g_{k+1}^T y_k}{d_k^T y_k} \frac{g_{k+1}^T d_k}{\|d_k\|^2} \frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + (\lambda - 1)^2 \frac{(g_{k+1}^T d_k)^2}{\|d_k\|^4} \frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2}$$

$$\Rightarrow \frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \leq \frac{(g_{k+1}^T y_k)^2}{d_k^T y_k d_k^T y_k} \frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2}$$

$$+2(\lambda - 1)\frac{\|g_{k+1}\|^2 g_{k+1}^T d_k + (g_{k+1}^T d_k)^2}{d_k^T y_k \|d_k\|^2} \frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + (\lambda - 1)^2 \frac{(g_{k+1}^T d_k)^2}{\|d_k\|^4} \frac{\|d_k\|^2 d_k^T y_k}{(g_{k+1}^T d_{k+1})^2 d_k^T y_k} + \frac{1}{\|g_{k+1}\|^2}$$

We know that $g_{k+1}^T d_k \leq d_k^T y_k$ and by Wolfe condition $g_{k+1}^T d_k \geq c_2 g_k{}^T d_k \Rightarrow c_2 g_k{}^T d_k \leq d_k^T y_k \Rightarrow -c_2 g_k{}^T d_k \geq -d_k^T y_k$ This implies that $\|g_k\|^2 \geq \frac{-1}{c_2} d_k^T y_k \Rightarrow -c_2\|g_k\|^2 \leq d_k^T y_k$

$$\frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \leq -\frac{(g_{k+1}^T y_k)^2}{c_2\|g_k\|^2 d_k^T y_k} \frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2}$$

$$+2(1 - \lambda)\frac{\|g_{k+1}\|^2 c_2\|g_k\|^2 - (g_{k+1}^T d_k)^2}{d_k^T y_k \|d_k\|^2} \frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} - (\lambda - 1)^2 \frac{(g_{k+1}^T d_k)^2}{\|d_k\|^4} \frac{\|d_k\|^2 d_k^T y_k}{(g_{k+1}^T d_{k+1})^2 c_2\|g_k\|^2} + \frac{1}{\|g_{k+1}\|^2}$$

Therefore c_2, $(\lambda - 1)^2$, $\|g_k\|^2$, $(g_{k+1}^T y_k)^2$, $d_k^T y_k$, $\|d_k\|^2$, $\|d_k\|^4$ and $(g_{k+1}^T d_{k+1})^2$ are greater than zero, then

$$\frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \le 2(1-\lambda)\frac{\|g_{k+1}\|^2 c_2 \|g_k\|^2 - (g_{k+1}^T d_k)^2}{d_k^T y_k \|d_k\|^2}\frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2}$$

$$\Rightarrow \frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \le 2(1-\lambda)\frac{\frac{\|g_{k+1}\|^2 c_2 \|g_k\|^4}{\|g_k\|^2} - (g_{k+1}^T d_k)^2}{d_k^T y_k \|d_k\|^2}\frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2}$$

$$\Rightarrow \frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \le 2(1-\lambda)\frac{\frac{-\|g_{k+1}\|^2 {c_2}^2 \|g_k\|^4}{d_k^T y_k} - (g_{k+1}^T d_k)^2}{d_k^T y_k \|d_k\|^2}\frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2}$$

$$\Rightarrow \frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \le -2(1-\lambda)\frac{\frac{\|g_{k+1}\|^2 {c_2}^2 \|g_k\|^4}{d_k^T y_k} + (g_{k+1}^T d_k)^2}{d_k^T y_k \|d_k\|^2}\frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2}$$

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

Hence $k = 0$ the above inequality yields $\frac{\|d_1\|^2}{(g_1^T d_1)^2} \le \frac{1}{\|g_1\|^2}$

Hence for all k, we conclude that $\frac{\|d_k\|^2}{(g_k^T d_k)^2} \le \frac{1}{\|g_k\|^2}$. Therefore $\frac{\|d_k\|^2}{(g_k^T d_k)^2} \le \sum_{i=0}^{k} \frac{1}{\|g_i\|^2}$ So, by (3.8)

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \le \sum_{i=0}^{k} \frac{1}{\mu^2} \Rightarrow \frac{\|d_k\|^2}{(g_k^T d_k)^2} \le \frac{1}{\mu^2}\sum_{i=0}^{k} 1 \Rightarrow \frac{\|d_k\|^2}{(g_k^T d_k)^2} \le \frac{k}{\mu^2} \Rightarrow \frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \frac{\mu^2}{k}$$

We take summation both sides, we get $\sum_{k\ge1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \mu^2 \sum_{k\ge1} \frac{1}{k} = \infty \Rightarrow \sum_{k\ge1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \infty$

Which contradicts Zountendijk condition in Theorem 3.4 The proof is then complete. ■

4. Numerical Results

This section is devoted to testing the implementation of a new method. We compare our method with Conjugate Gradient methods ((HS) and (PR)) the comparative tests involve well-known nonlinear problems (standard test function) with different dimensions $5 \le N \le 5000$, all programs are written in FORTRAN90 language and for all cases the stopping condition is $|{g_k}^T g_{k+1}| > 0.2\|g_{k+1}\|^2$, the results are given in table (4.1) specifically quote the number of function NOF and the number of iteration NOI. More experimental results and table (4.2) confirm that the new CG is superior to standard CG (PR) and standard CG (HS) concerning the NOI and NOF.

Table (4.1): Comparative Performance of Three Algorithms Standard HS, PR and AA1

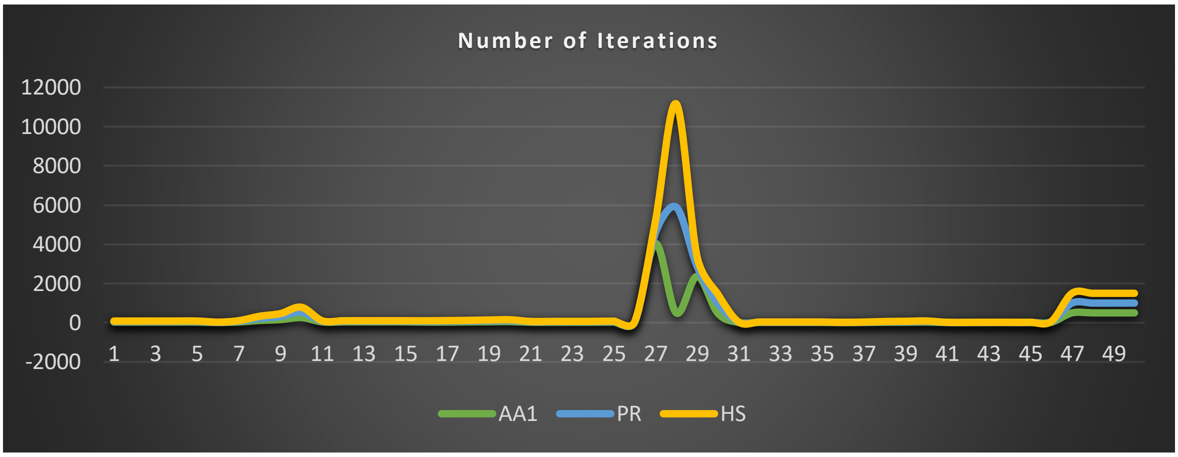
No. of Test	Test Function	N	Standard Formula (HS)		Standard Formula (PR)		New Formula (AA1)	
			NOI	NOF	NOI	NOF	NOI	NOF
	Wood	5	30	68	29	67	25	59
1		50	30	68	29	67	26	61
		500	30	68	30	69	26	61
		1000	30	68	30	69	26	61
		5000	30	68	30	69	26	61
2	OSP	5	10	56	10	56	9	45
		50	36	133	37	147	32	113
		500	112	353	110	341	108	335
		1000	156	473	161	493	148	460
		5000	256	744	276	844	256	765
3	Powell	5	38	108	35	87	26	78
		50	38	108	35	87	26	78
		500	41	124	35	87	26	78
		1000	41	124	35	87	26	78
		5000	41	124	35	87	26	78
4	Miele	5	28	85	37	116	33	103
		50	31	102	44	148	33	103
		500	40	146	44	148	36	121
		1000	46	176	50	180	42	153
		5000	54	211	50	180	47	174
5	G-Central	5	22	159	22	159	19	126
		50	22	159	22	159	21	151
		500	23	171	23	171	22	164
		1000	23	171	23	171	22	164
		5000	28	248	30	270	22	164
	Dixon	5	13	30	13	30	13	30
6		50	4044	8090	532	1188	604	1290
		500	499	1122	5395	10793	5268	10541
		1000	2366	4737	511	1157	491	1103
		5000	476	1068	516	1159	458	1036
7	Beal	5	11	28	11	28	10	22
		50	12	30	12	30	11	26
		500	12	30	12	30	11	26
		1000	12	30	12	30	11	26
		5000	12	30	12	30	11	26
8	Sum	5	6	39	6	39	6	39
		50	11	60	11	60	11	66
		500	21	124	21	123	19	90
		1000	23	128	23	127	26	122
		5000	31	159	31	145	26	112
9	Extended	5	7	18	7	18	6	16
	PSC1	50	6	16	6	16	6	16
		500	7	18	7	18	6	16
		1000	7	18	7	18	6	16
		5000	7	18	7	18	6	16
10	G-Biggs	5	20	64	20	63	19	60
		50	F	F	F	F	F	F
		500	F	F	F	F	F	F
		1000	F	F	F	F	F	F
		5000	F	F	F	F	F	F
	Total		**10839**	**22172**	**10434**	**21479**	**10134**	**20529**

Note: F that is fail, we took F=500.

Table (4.2): Comparing the rate of improvement between the new algorithm (AA1) and the standard algorithm (HS) and (PR)

Tools	Standard (HS)	New (AA1)	Standard (PR)	New (AA1)
NOI	100%	93.4957%	100%	97.2148
NOF	100%	92.5898%	100%	95.5771

Table (4.2) shows the rate of improvement in the new algorithm (AA1)with the standard algorithms (HS) and (PR) , The numerical results of the new algorithm is better than the standard algorithm, As we notice that (NOI), (NOF) of the standard algorithm (HS) are about 100%, That means the new algorithm has improvement on standard algorithm (HS) prorate (6.5043%) in (NOI) and prorate (7.4102%) in (NOF) and standard algorithm (PR) are about 100%, That means the new algorithm has improvement on standard algorithm (PR) prorate (2.7852%) in (NOI) and prorate (4.4229%) in (NOF), In general the new algorithm (AA1) has been improved prorate (5.2807%) compared with standard algorithms (HS) and (PR).



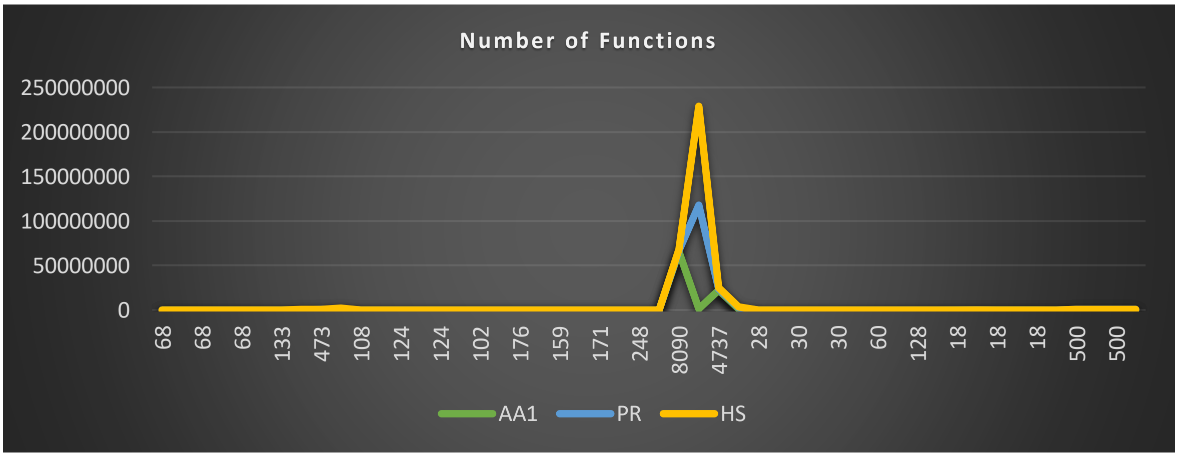


Figure (4.1): shows the comparison between new algorithm (AA1) and the standard algorithms (HS) and (PR) according to the total number of iterations (NOI) and the total number of functions (NOF).

5. Conclusion

In this paper, we proposed a new three term for CG β_k^{AA1} that has some properties of global convergence. Numerical results have shown that this new β_k^{AA1} performs better than (HS) and (PR). In the future we can and by same way we proposed many new three terms for CG of unconstrained optimization.

6. Acknowledgements

I would like to express my thanks Mr. Hussein A. Khatab for his continuous advice and many helpful suggestions during out this work.

References

[1] A.Y. AL - Bayati, and N.H. AL-Assady, Conjugate gradient method, Technical Research. school of computer studies, Leeds University, UK., 1(1986).

[2] A. Abashar, M. Mamat, M. Rivaie and I. Mohd. Global Convergence Properties of a New Class of Conjugate Gradient Method for Unconstrained Optimization. International Journal of Mathematical Analysis, 67(8) (2014), 3307-3319.

[3] Y.H. Dai & Y. Yuan, A nonlinear conjugate gradient with a strong global convergence property, SIAM J. Optim., 10(1) (1999), 177-182. https://doi.org/10.1137/s1052623497318992

[4] L. C. W. Dixon, Conjugate Gradient Algorithms Quadratic Termination Without Linear Searches. Journal of Inst. of Mathematics And its Applications, 15(1) (1975), 9-15. https://doi.org/10.1093/imamat/15.1.9

[5] R. Fletcher, and C. Reeves, Function minimization by conjugate gradients. Comput. J., 7 (2) (1964), 149-154. https://doi.org/10.1093/comjnl/7.2.149

[6] M.R. Hestenes and E. Steifel, Method of conjugate gradient for solving linear equations. J. Res. Nat. Bur. Stand, 6(49) (1952), 409-436.

[7] I. Jusoh, M. Mamat and M. Rivaie, A New Edition of Conjugate Gradient Methods for Large-Scale Unconstrained Optimization. International Journal of Mathematical Analysis, 46(8) (2014), 2277 -2291.

[8] H. A. Khatab and S. G. Shareef, A New conjugate gradient method for unconstrained optimization problems with descent property. General Letters in Mathematics, 9(2) (2020), 39 -44. https://doi.org/10.31559/glm2020.9.2.6

[9] H. A. Khatab and S. G. Shareef, A New Parameter Conjugate Gradient Method Based on Three Terms Unconstrained Optimization. General Letters in Mathematics, 1(7) (2019), 39 -44. https://doi.org/10.31559/glm2019.7.1.5

[10] Y. Liu and C. Storey, Efficient generalized conjugate gradient algorithm's part 1: Theory. J. Comput. Appl. Math., 69 (1) (1991), 129-137. https://doi.org/10.1007/bf00940464

[11] J. Liu and X. Wu, New three-term conjugate gradient method for solving unconstrained optimization problems. Science asia, 40(4) (2014), 295-300. https://doi.org/10.2306/scienceasia1513-1874.2014.40.295

[12] E. Polak and G. Ribiere, Note sur la convergence de directions conjugees, Rev. Francaise Informat Recherche Operationelle. 3E Annee, (16) (1969), 35-43.

[13] M. Rivaie, M. Mamat, L. W. June and I. Mohd, A new conjugate gradient coefficient for large scale nonlinear unconstrained optimization. International Journal of Mathematical Analysis, 23(6) (2012),1131-1146.

[14] L. Zhang, W. Zhou and D.H. Li, A descent modified Polak-Ribi`ere-Polyak conjugate gradient method and its global convergence. IMA Journal of Numerical Analysis, 26(4) (2006),629-640. https://doi.org/10.1093/imanum/drl016

[15] L. Zhang, W. Zhou and D.H. Li, Some descent three-term conjugate gradient methods and their global convergence. Optimization Methods and Software, 22(4) (2007), 697–711. https://doi.org/10.1080/10556780701223293

[16] Zoutendijk G., Nonlinear Programming Computational Methods, Integer and Nonlinear Programming, J. Ababie (Ed), North-Holand. Amsterd, (1970), 37-86.