Journal of Basics and Applied Sciences Research (JOBASR)
ISSN (print): 3026-9091, ISSN (online): 1597-9962
Volume 2(3) September 2024
DOI: https://doi.org/10.33003/jobasr-2024-v2i3-61

A Sufficient Descent Modified Nonlinear Conjugate Gradient Method for Solving Large Scale Unconstrained Optimization Problems

Moyi, A. U.[1*], Abdullahi, N.[2] and Aliyu, N.[2]
[1]Department of Mathematical Sciences, Faculty of Science, Federal University Gusau Zamfara State, Nigeria.
[2]Department of Mathematical Sciences, Faculty of Science, Nigerian Defence Academy (NDA) Kaduna, Nigeria.
*Corresponding Author Email: aliyumoyik@yahoo.com

ABSTRACT
Nonlinear conjugate gradient methods (CG) are prominent iterative techniques widely used for solving large-scale unconstrained optimization problems. Their wide application in many fields is due to their simplicity, low memory requirements, computationally less costs and global convergence properties. However, some of the CG methods do not possess the sufficient descent conditions, global convergence properties and good numerical result. To overcome these drawbacks, numerous studies and modification have been conducted to improve on these methods. In this research, a modification of a new Conjugate gradient parameter that posses sufficient descent conditions and global convergence properties is presented. The global convergence result is established using the Strong Wolf Powell condition (SWP). Extensive numerical experiment was conducted on a set of standard unconstrained optimization test functions. The results show that the method outperforms some well-known methods in terms of efficiency and robustness.

Keywords:
Unconstrained optimization,
Conjugate gradient method,
Line search,
Global convergence.

INTRODUCTION

Consider the unconstrained minimization problem

$$\min_{x \in R^n} \; f(x) \quad (1)$$

that is widely used for solving large-scale unconstrained optimization problem, where $f: \mathbb{R}^{\boldsymbol{n}} \rightarrow \mathbb{R}$ is continuously differentiable function and n is the dimension of x, which is assume to be large. The iterate of the conjugate gradient methods are obtain by

$$x_{k+1} = x_k + \alpha_k d_k \quad (2)$$

where $\alpha_k > 0$, is a step length and usually computed via different line search methods. The search direction d_k is computed as follows:

$$d_k = \begin{cases} -g_k & if\ k = 0 \\ -g_k + \beta_k d_{k-1} & if\ k > 0 \end{cases} \quad (3)$$

Where g_k is the gradient of $f(x)$ at the point x_k, $\beta_k \epsilon R$ is a scalar known as CG parameter.

There are several choices of the conjugate gradient parameter β_k in the literature each yielding to a different CG method. Some of the well known CG coefficients are as follows:

$$\beta_k^{HS} = \frac{g_{k+1}^T(g_{k+1}-g_k)}{(g_{k+1}-g_k)d_k} \quad (4)$$

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{||g_k||^2} \quad (5)$$

$$\beta_k^{PRP} = \frac{g_{k+1}^T(g_{k+1}-g_k)}{||g_k||^2} \quad (6)$$

$$\beta_k^{CD} = \frac{-g_{k+1}^T g_k}{d_k^T g_k} \quad (7)$$

$$\beta_k^{LS} = \frac{g_{k+1}^T(g_{k+1}-g_k)}{-d_k g_k} \quad (8)$$

$$\beta_k^{DY} = \frac{g_{k+1}^T g_k}{(g_{k+1}-g_k)d_k} \quad (9)$$

The above formulae are known as Hestenes-Stiefel (1952) (HS), Fletcher-Reeves (1964) (FR), Polak-Ribiere-Polayak (1969) (PRP), Conjugate Descent (1987) (CD), Liu-Storey (1991) (LS) and Dai-Yuan (1999) (DY). These formulas for conjugate gradient method have been categorized into two groups : The first group include PRP (1969), HS (1952), and LS (1991), they are considered among the most efficient CG algorithms for the solution of unconstrained optimization problems particularly the large scale, because they possess an inbuilt automatic restart feature that helps prevent them from jamming. But yet, their convergence properties are still now not established under certain inexact line search condition. The other group includes the FR (1964), CD (1987), and DY (1999) Methods. Though these methods possesses strong convergence properties, but are often considered very poor in terms of numerical performance due to their jamming phenomena (Hager and Zhang 2006).

The above mentioned drawbacks inspired researchers to revise and propose numerous modification of the conjugate gradient method with the aim of overcoming these drawbacks.

The global convergence properties of CG techniques are the most researched and well-known. The referenced study by zotendjik (1970), proved the global convergence of FR method using the exact line search. This was later refuted by Powell (1977) by giving a counter example. Many researchers believe that the PRP method is the most reliable CG method, but it is also known not to possess global convergence properties, as shown by Powell (1984). Powell (1986) has also shown that FR method is superior method when compared with the others. In addition, other researchers such as A-lBaali (1985), Touati-Ahmed and Storey (1990), Gilbert and Nocedal (1992), have further analysed the global convergence of algorithm related to the FR method using the inexact line search with a strong wolf condition.

Andrei (2011) has classified the CG method in three different groups, the classical CG method, the scaled CG method and lastly the hybrid CG method . The classical CG method is the most simplest and yet easy to applied. Formulas (4) to (9) are regarded as classical CG formulas. However to find and produce a new CG method of this type is quite difficult. Therefore, researchers usually come out with the scaled or hybrid CG method.

Among the earliest hybrid CG algorithm is β_k^{TS} developed by Touati-Ahmed and Storey (1990) with the formula defined as follows:

$$\beta_k^{TS} = \begin{cases} \beta_k^{PRP} & if\ 0 < \beta_k^{PRP} \le \beta_k^{FR} \\ \beta_k^{FR} & Otherwise \end{cases} \qquad (10)$$

Also some of the best research on this type of CG method has been conducted are by Wei et al. (2006), Yuan et al. (2010). However it is obvious that to find the formula for these types is quite difficult and complicated to understand. Thus base on this reason, this paper only concentrated on the classical formulas of CG methods.

In recent years much effort has been place on developing and constructing a new and simple formula for CG methods with good numerical performance and global convergence properties (see refs. Alhawarat et al (2017), Salih Y.et al (2018), Aini N. et al (2019), Saleh N.A et al (2020), Auwal A.M et al (2023) and Kabiru A. et al (2024)).

New conjugate gradient parameter

In an attempt to overcome some of the drawbacks discussed above, Rivaie et al, (2012) proposed a new CG coefficient known as β_k^{RMIL}. This coefficient is known to fulfill the sufficient descent condition and also possess global convergence properties.

This β_k^{RMIL} is denoted as:

$$\beta_k^{RMIL} = \frac{g_{k+1}^T(g_{k+1} - g_k)}{||d_k||^2} \qquad (11)$$

Based on Rivaie et al (2012), Revaie et al (2015) also has proposed $RMIL^+$ conjugate gradient method whose coefficient β_k is defined as

$$\beta_k^{RMIL^+} = \frac{g_{k+1}^T(g_{k+1} - g_k - d_k)}{||d_k||^2} \qquad (12)$$

And has shown that $RMIL^+$ is globally convergent and gives good numerical results under the exact line search.

Motivated by these findings we develop a new conjugate gradient parameter β_k known as ANN (Aliyu, Nasiru and Najib) This formula retained the original numerator of $RMIL^+$ but maintained the Fletcher and Reeves (FR) denominator. The formula is defined as

$$\beta_k^{ANN} = \frac{g_{k+1}^T(g_{k+1} - g_k - d_k)}{||g_k||^2} \qquad (13)$$

where $||\cdot||$ is the Euclidean norm of vectors.
We employed the inexact line search which include wolf and strong wolf line search to compute the step length (α_k) defined below:

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k{}^T d_k \quad (14)$$

and

$$|g(x_k + \alpha_k d_k)^T d_k| \le -\sigma g_k{}^T d_k \quad (15)$$

where $0 < \delta < \sigma < 1$ are two constant and $g_{k+1} = g(x_k + \alpha_k d_k)$

Algorithm
Step 1: Initialization. Given $x_0 \in R^n$ select some positive value for δ and set $k = 0$
Step 2: Compute β_k based on (13)
Step 3: Compute d_k based on (3). If $g_k = 0$, then stop.
Step 4: Compute α_k based on (14)-(15)
Step 5: Updating new point based on (2)
Step 6: Convergent test and stopping criteria.
If $f(x_{k+1}) < f(x_k)$ and $||g_k|| < \varepsilon$ then stop.
Otherwise go to step 1 with $k = k + 1$

MATERIALS AND METHODS

Convergence Analysis

In this section, we will provide the proofs of the sufficient descent properties and the global convergence of ANN method when it applied under strong wolfe line search of equation (14) and (15).

Sufficient descent condition

This condition is defined as:
$g_k{}^T d_k \le -c||g_k||^2$ for $k \ge 0$ and $c > 0$. (16)
Also the inequality below would be use for the proofs of the sufficient descent property (16) and the global convergence properties when it applied under the strong wolf e line search.

$$0 \le \beta_k{}^{ANN} \le \frac{||g_{k+1}||^2}{||g_k||^2}, \quad \forall k \ge 1$$

Lemma 1: Assume that g_k and d_k are obtained by Algorithm with $\sigma \le \frac{3}{2}$ then $\forall k \ge 1$, we get

$$||g_k|| < 2\,||d_k||$$
$$||g_k||^2 < 4\,||d_k||^2$$
$$\Rightarrow \quad -||d_k|| < -\frac{||g_k||}{2} \quad (17)$$

Proof. The proof of this lemma would be by induction. It is obvious that for k = 0, the result holds true. Suppose (16) holds true for some k > 0. Then we have

$$||g_{k+1} + d_{k+1}||^2 = (g_{k+1} + d_{k+1})^T (g_{k+1} + d_{k+1})$$
$$= ||g_{k+1}||^2 + ||d_{k+1}||^2 + 2 g_{k+1}{}^T d_{k+1} \quad (18)$$

From (3) we have that

$$g_{k+1}{}^T d_{k+1} = -||g_{k+1}||^2 + \beta_{k+1}{}^{ANN} g_{k+1}{}^T d_k, \quad (19)$$

Substituting (19) into (18) we get

$$||g_{k+1} + d_{k+1}||^2 = ||g_{k+1}||^2 + ||d_{k+1}||^2 - 2||g_{k+1}||^2 + 2\beta_{k+1}{}^{ANN} g_{k+1}{}^T d_k$$
$$||g_{k+1} + d_{k+1}||^2 \le ||d_{k+1}||^2 - ||g_{k+1}||^2 + 2|\beta_{k+1}{}^{ANN}||g_{k+1}{}^T d_k|,$$

Applying (14) and (15) conditions and noting that $\beta_{k+1}{}^{ANN} \ge 0$

$$||g_{k+1} + d_{k+1}||^2 \le ||d_{k+1}||^2 - ||g_{k+1}||^2 + 2\sigma \beta_{k+1}{}^{ANN}|g_{k+1}{}^T d_k|,$$

Since $0 \le \beta_{k+1}{}^{ANN} \le \frac{||g_{k+!}||^2}{||g_k||^2}$ and applying Cauchy-Schwartz inequality we get

$$||g_{k+1} + d_{k+1}||^2 \le ||d_{k+1}||^2 - ||g_{k+1}||^2 + 2\sigma \frac{||g_{k+1}||^2}{||g_k||^2} ||g_k||\,||d_k||$$
$$\Rightarrow \quad ||g_{k+1} + d_{k+1}||^2 \le ||d_{k+1}||^2 - ||g_{k+1}||^2 + 2\sigma ||g_{k+1}||^2 \frac{||d_k||}{||g_k||}$$

Applying (17) we get

$$||g_{k+1} + d_{k+1}||^2 \le ||d_{k+1}||^2 - ||g_{k+1}||^2 - 2\sigma \left(\frac{-||g_k||}{2||d_k||}\right) ||g_{k+1}||^2$$
$$\Rightarrow \quad ||g_{k+1} + d_{k+1}||^2 \le ||d_{k+1}||^2 + (\sigma - 1)||g_{k+1}||^2$$

Hence

$$||g_{k+1} + d_{k+1}||^2 + (1 - \sigma)||g_{k+1}||^2 \le ||d_{k+1}||^2,$$

Since $(1 - \sigma) > 0$ we obtain

$$(1 - \sigma)||g_{k+1}||^2 \le ||d_{k+1}||^2,$$
$$\Rightarrow \quad ||g_{k+1}||^2 < \left(\frac{1}{1-\sigma}\right) ||d_{k+1}||^2$$

Hence $||g_{k+1}||^2 \le 4\,||d_{k+1}||^2$ whenever $\sigma \le \frac{3}{4}$ then $||g_k|| < 2\,||d_k||$
Which implies that (16) holds for k+1. Therefore the proof is completed.

Using **Lemma 1**. We get the relationship between g_k and d_k below, when d_k and g_k are generated by our Algorithm with $\sigma \le \frac{3}{4}$, that is

$$\frac{||g_k||^2}{4} < \frac{||d_k||^2}{1} \text{ or } \frac{1}{||d_k||^2} < \frac{4}{||g_k||^2} \quad \text{for all } k \ge 0 \quad (20)$$

The following theorem establishes the sufficient descent property and will be used to prove the global convergence.

Theorem 1. Let the sequence $\{g_k\}$ and $\{d_k\}$ be generated by our algorithm with $\sigma \le \frac{3}{4}$ then for all

$k \geq 0$ Then (i)

$$\frac{-1}{1-\sigma} < \frac{g_k{}^T d_k}{||g_k||^2} < \frac{2\sigma-1}{1-\sigma} \quad (21)$$

Hence the condition (16) holds

Proof: From (16) the result is obvious for k = 0.
Now consider k > 0. From (3), we have

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

$$\Rightarrow \quad \frac{g_{k+1}{}^T d_{k+1}}{||g_{k+1}||^2} = -1 + \beta_{k+1}{}^{ANN} \frac{g_{k+1}{}^T d_k}{||g_{k+1}||^2}$$

$$= -1 + \beta_{k+1}{}^{ANN} \frac{||g_k||^2}{||g_{k+1}||^2} \cdot \frac{g_{k+1}{}^T d_k}{||g_k||^2} \quad (22)$$

From Strong Wolf Powell condition, we have

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

Which together with that $\beta_k{}^{ANN} \geq 0$

$$\sigma \beta_{k+1}{}^{ANN} g_k{}^T d_k \leq \beta_{k+1}{}^{ANN} g_k{}^T d_k \leq -\sigma \beta_{k+1}{}^{ANN} g_k{}^T d_k \quad (23)$$

Dividing through equation (23) by $||g_{k+1}||^2$

$$\sigma \beta_{k+1}{}^{ANN} \frac{g_k{}^T d_k}{||g_{k+1}||^2} \leq \frac{g_{k+1}{}^T d_k}{||g_{k+1}||^2} \leq \sigma \beta_{k+1}{}^{ANN} \frac{g_k{}^T d_k}{||g_{k+1}||^2}$$

is the same as

$$\sigma \beta_{k+1}{}^{ANN} \frac{||g_k||^2}{||g_{k+1}||^2} \cdot \frac{g_{k-1}{}^T d_{k-1}}{||g_k||^2} \leq \frac{g_{k+1}{}^T d_k}{||g_{k+1}||^2} \leq -\sigma \beta_{k+1}{}^{ANN} \frac{||g_k||^2}{||g_{k+1}||^2} \cdot \frac{g_k{}^T d_k}{||g_k||^2}$$

and

$$-1 + \sigma \beta_k{}^{ANN} \frac{||g_k||^2}{||g_{k+1}||^2} \cdot \frac{g_k{}^T d_k}{||g_k||^2} \leq \frac{g_{k+1}{}^T d_k}{||g_{+1k}||^2} \leq -1 - \sigma \beta_{k+1}{}^{ANN} \frac{||g_k||^2}{||g_{k+1}||^2} \cdot \frac{g_k{}^T d_k}{||g_k||^2}$$

By (21) and the condition $\beta_k{}^{ANN} \geq 0$

$$-1 - \sigma \beta_{k+1}{}^{ANN} \left(\frac{\sigma}{1-\sigma}\right) \frac{||g_k||^2}{||g_{k+1}||^2} < \frac{g_{k+1}{}^T d_k}{||g_{k+1}||^2} < -1 + \sigma \beta_{k+1}{}^{ANN} \left(\frac{\sigma}{1-\sigma}\right) \frac{||g_k||^2}{||g_{k+1}||^2}$$

Substituting (21) in to the above we get

$$\frac{-1}{1-\sigma} < \frac{g_k{}^T d_k}{||g_k||^2} < \frac{2\sigma-1}{1-\sigma}$$

Hence the result holds for k and the proof is complete.

Global convergence analysis

The assumptions defined below are very necessary in the study of global convergence of the CG algorithm.

Assumption 1

(i) $f(x)$is bounded from below on the level set and is continuous and differentiable in a neighborhood N of the level set $\tau = \{x \in R^n \backslash f(x) \leq f(x_o)\}$ where x_0 is the starting point and f is smooth in a neighborhood N of the level set τ.

(ii) $g(x)$ is Lipchitz continuous in N, so $\exists\, l > 0$ (constant) such that

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

The following lemma is necessary in the study of global convergence of the CG algorithm.

Lemma 2: Suppose that assumption 1 holds. For any CG algorithm defined on (2)-(4) where the step length α_k is computed by the strong wolf line search. Then the following condition known as the zountendijk condition will be satisfied

$$\sum_{k=0}^{\infty} ||g_k||^2 cos^2\theta_k < \infty. \quad (24)$$

Where θ_k is the angle between d_k and $-g_k$ which is given by

$$cos\theta_k = \frac{-g_k{}^T d_k}{||g_k||||d_k||} \quad (25)$$

Lemma 3: Let the sequence $\{x_k\}$ be produce by our Algorithm with $\sigma \leq \frac{3}{4}$ .Then for all $k \geq 0$ we have

$$\sum_{k=0}^{\infty} \frac{||g_k||^4}{||d_k||^2} < \infty.$$

Proof. Multiplying (21) by $\frac{||g_k||}{||d_k||}$ and using (25) we get

$$c_2 \frac{||g_k||}{||d_k||} < cos\theta_k < c_1 \frac{||g_k||}{||d_k||} \text{ for all } k \geq 0 \quad (26)$$

Where $c_1 = \frac{1}{1-\sigma}$ and $c_2 = \frac{2\sigma-1}{1-\sigma}$

Since $c_2 > 0$ when $0 < \sigma < \frac{3}{4}$, then $cos\theta_k > 0$. Hence,

$$c_2{}^2 \frac{||g_k||^2}{||d_k||^2} < cos^2\theta_k.$$

This implies

$$c_2{}^2 \sum_{k=0}^{\infty} \frac{||g_k||^4}{||d_k||^2} < \sum_{k=0}^{\infty} ||g_k||^2 cos^2\theta_k$$

From (24) and (25) together, it follows that

$$\sum_{k=0}^{\infty} \frac{||g_k||^4}{||d_k||^2} < \infty.$$

Hence the proof is complete.

Theorem 2: Let the sequence $\{x_k\}$ be produce by our algorithm where $\sigma \leq \frac{3}{4}$ and suppose Assumption 1 holds. Then,

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

Proof. Let the opposite of the theorem be true, hence $\exists \mu > 0$ and the integer k, such that

$$\|g_k\| \geq \mu, \ \ \forall k > k_1$$

Hence

$$\frac{1}{||g_k||^2} < \frac{1}{\mu^2} \text{ for all } k > k_1 \text{ and } ||g_k|| \neq 0 \quad (27)$$

Now rewrite (3) as $d_k + g_k = \beta_k{}^{ANN} d_{k-1}$, then square it both side

$$||d_k||^2 + ||g_k||^2 + 2g_k{}^T d_k = (\beta_k{}^{ANN})^2 ||d_{k-1}||^2$$

$$||d_k||^2 = -||g_k||^2 - 2g_k{}^T d_k + (\beta_k{}^{ANN})^2 ||d_{k-1}||^2$$

Applying theorem 1 we obtain

$$||d_k||^2 < -||g_k||^2 + \left(\frac{2}{1-\sigma}\right)||g_k||^2 + (\beta_k^{ANN})^2||d_{k-1}||^2$$

This leads to

$$||d_k||^2 < \left(\frac{1+\sigma}{1-\sigma}\right)||g_k||^2 + (\beta_k^{ANN})^2||d_{k-1}||^2 \quad (28)$$

Now since $\beta_k^{ANN} \leq \frac{||g_k||^2}{||g_{k-1}||^2}$, then we get

$$||d_k||^2 < \left(\frac{1+\sigma}{1-\sigma}\right)||g_k||^2 + \frac{||g_k||^2}{||g_{k-1}||^2}||d_{k-1}||^2 \quad (29)$$

Now we multiply both side of (29) by $\frac{1}{||g_k||^4}$ to get

$$\frac{||d_k||^2}{||g_k||^4} < \left(\frac{1+\sigma}{1-\sigma}\right)\frac{1}{||g_k||^2} + \frac{||d_{k-1}||^2}{||g_{k-1}||^4}$$

$$= \left(\frac{1+\sigma}{1-\sigma}\right)\frac{1}{||g_k||^2} + \frac{||d_{k-1}||^2}{||g_{k-1}||^4}\left(\frac{1}{||g_{k-1}||^2}\right) \quad (30)$$

By substituting (17) into (30) we get

$$\frac{||d_k||^2}{||g_k||^4} < \left(\frac{1+\sigma}{1-\sigma}\right)\frac{1}{||g_k||^2} + \left(\frac{1}{4\,||g_{k-1}||^2}\right) \quad (31)$$

Combining (27) and (31) together we have

$$\frac{||d_k||^2}{||g_k||^4} < \left(\frac{1+\sigma}{1-\sigma} + \frac{1}{4}\right)\frac{1}{\mu^2},\ \forall\, k \geq k_1 + 1$$

This means that

$$\frac{||d_k||^2}{||g_k||^4} > \frac{(4-4\sigma)\mu^2}{5-3\sigma},\quad \forall k > k_1 + 1 \quad (32)$$

Since (32) holds $\forall k > k_1 + 1$, then

$$\sum_{k=0}^{n}\frac{||g_k||^4}{||d_k||^2} > \sum_{k=k_1+1}^{n}\frac{||g_k||^4}{||d_k||^2} > \sum_{k=k_1+1}^{n}\frac{(4-4\sigma)\mu^2}{(5-3\sigma)} = \left[\left(\frac{(4-4\sigma)\mu^2}{(5-3\sigma)}\right)\right] \quad (33)$$

From (33) we get

$$\sum_{k=0}^{\infty}\frac{||g_k||^4}{||d_k||^2} > \sum_{k=k_1+1}^{\infty}\frac{||g_k||^4}{||d_k||^2}$$

$$\lim_{n\to\infty}\sum_{k=k_1+1}^{n}\frac{||g_k||^4}{||d_k||^2} > \lim_{n\to\infty}\sum_{k=k_{1+1}}^{n}\left[\left(\frac{(4-4\sigma)\mu^2}{(5-3\sigma)}\right)\right](n-k_1) = \infty$$

And thus, contradict lemma 3. So the proof is complete.

RESULTS AND DISCUSSION

Results

In this section we present the numerical result from experiments conducted applied in testing the performance and proficiency of β_k^{New}. A comparison was made with the other CG methods, which involve FR and RMIL+ Some standard benchmark problems were considered with different dimension as summarized in Table 1. The performance was based on CPU time and the number of iteration (NI) and under Strong Wolf Powell line search.

All problems and formulas are coded and run on the same Matlab programs and for all algorithm the stopping criterion was set as $\| g_k \| \leq 10^{-6}$, different initial guess are used in the computations with the variable dimension $2 \leq n \leq 2000$. The performance was also analyzed via the performance profile developed by Dolan and More (2002) as can be seen in figure 1 and figure 2.

Table 1. Numerical Results of ANN, FR and RMIL+

Fletcher function (cute)	(3,3…)	100	838	2.2903	2631	6.713	47	0.2257
	(5,5…)	200	2906	12.0284	-	-	10	0.5183
	(-5,-5…)	500	164	1.3434	86	0.7152	11	0.7663
	(8,8…)	1000	8525	82.2137	-	-	9	0.9057
Extended Rosenbrock	(-0.4,-0.4)	50	-	-	-	-	12	0.258
	(1,1…)	100	1	8.01E-04	1	8.01E-04	1	9.82E-04
Extended Himmelblau	(0.2,0.2...)	40	57	0.1567	18	0.0378	13	0.0637
	(4,4…)	100	329	1.0406	10	0.0378	20	0.1388
Extended Tridiagonal 1	(8,8…)	4	248	0.4359	28	0.0591	28	0.0718
	(0.3,0.3...)	10	325	0.5747	10	0.0309	10	0.0964
BIGGSB1 function (cute)	(0.1,0.1...)	10	-	-	-	-	18	0.4059
	(1.2,1.2...)	50	-	-	-	-	18	0.4983
	(0.5,0.5...)	500	-	-	-	-	19	1.2885
Extended Maratos	(1.4,1.4...)	80	-	-	-	-	10	0.3228
	(0.2,0.2...)	800	-	-	-	-	13	1.0335
Shallow function	(1,1…)	50	1	7.45E-04	1	7.40E-04	1	6.42E-04
	(1,1…)	500	1	0.002	1	0.0019	1	0.002
	(6,6…)	1000	446	4.722	32	0.3911	13	1.1706
Quadratic QF2 function	(1.5,1.5...)	50	459	1.2506	99	0.199	79	0.307
	(0.1,0.1...)	100	1158	4.0945	163	0.541	151	0.6341
Edensch function	(1.5,1.5...)	60	12	0.4126	31	0.1912	12	0.3699
	(1.6,1.6...)	100	-	-	46	0.3454	11	0.4869
Dixon and price function	(1.2,1.2...)	10	228	0.4597	102	0.2187	92	0.2714
	(0.6,0.6...)	20	211	0.4963	197	0.4205	197	0.4077
	(0.2,0.2...)	100	866	3.1405	296	0.9303	289	0.8991
Sphere function	(1.3,1.3...)	200	1	0.0073	1	0.0071	1	0.0068
	(2,2…)	500	1	0.0104	1	0.0103	1	0.0104
Extended freudenstein	(0.1,0.1...)	10	-	-	37	0.1223	11	0.1714
NONSCOMP function	(8,8…)	500	-	-	-	-	9	0.5669
	(-2,-2…)	1000	2058	21.3426	63	0.6995	62	0.6765
	(8,8…)	2000	-	-	-	-	11	1.6616
Generlized Tridiagonal 1	(1.5,1.5...)	200	43	0.3506	23	0.1591	35	0.3313
	(8,8…)	600	-	-	-	-	67	1.9293
Generlized Tridigonal 2	(-1,-1…)	50	1	7.34E-04	1	8.06E-04	1	7.03E-04
	(-3.4,-3.4.)	100	40	0.161	-	-	10	0.3008
Extended quadratic penalty (QP1)	(-1.7,-1.7)	1000	-	-	21	0.5052	8	0.8947
	(3.2,3.2...)	2000	-	-	-	-	8	1.5447

Cube function	(-5.5,-5.5) (-9.6,-9.6)	50 500	59 -	0.1738 -	47 -	0.1143 -	47 104	0.145 1.9434
Extended powell function	(2.1,2.1...) (3,3…)	50 100	4 4	0.1137 0.1729	4 4	0.1149 0.1678	4 4	0.1637 0.0964
Extended wood	(1,1…)	20	1	3.84E-04	1	5.94E-04	1	4.58E-04
	(1,1…)	40	1	5.05E-04	1	4.51E-04 -	1	3.84E-04
Extended Beal function	(3.8,3.8...) (4.8,4.8...)	40 400	- -	- -	- -	-	7 7	0.1783 0.7173
Extended DESCHNB (cute)	(3.5,3.5...)	1000	23	0.2808	10	0.1386	17	0.151
	(1,1…)	50	1	7.83E-04	1	6.22E-04	1	4.66E-04
Test								
	(3.5,3.5...)	100	-	-	-	-	7	0.2721
Extended white & Holst function	(1,1…)	1000	1	0.0035	1	8.62E-04	1	0.0032
Liarwhd function	(1,1…)	20	1	5.02E-04	1	5.54E-04	1	5.35E-04

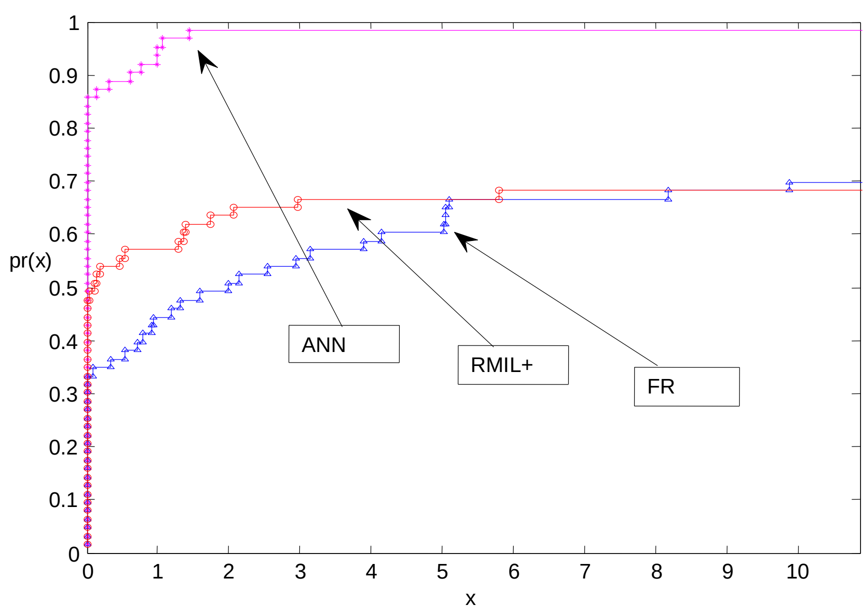

Figure 1: Performance profile base on the number of iteration (NI)

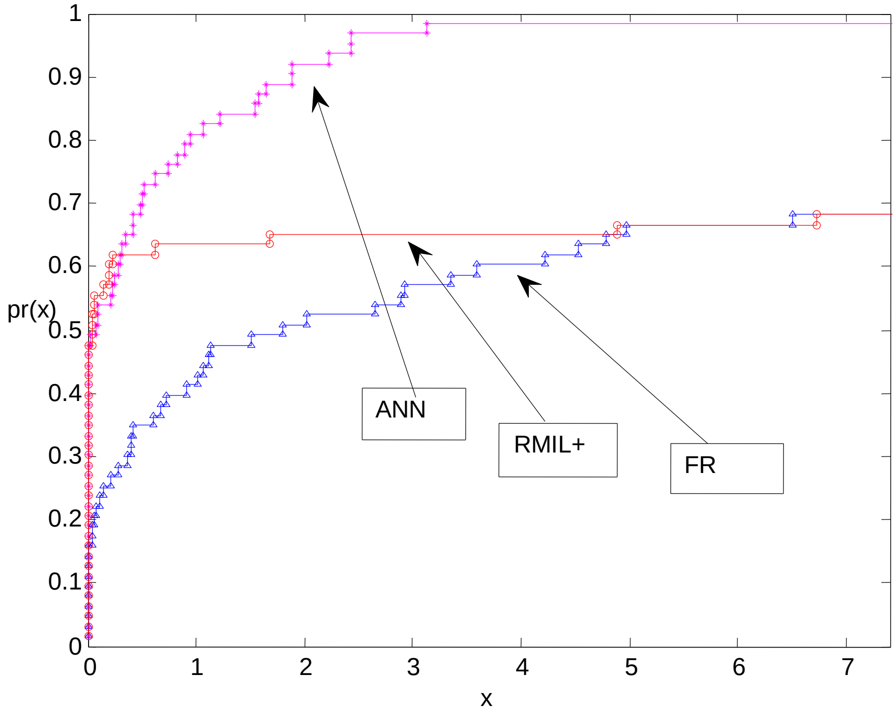

Figure 2: Performance profile base on the CPU time in seconds

Discussion of the Result

The performance profile is employed to evaluate and compares the performance of the methods on a set of standard test problems. All of the test problems are from Andrei (2009) . To show the robustness, test problems have been implemented under low, medium, and high dimensions, as, 2, 4, 10, 50, 100, 500, 1000, and 2000. Furthermore, for each dimension two or more different initial points are used one of which is the initial point suggested by Andrei (2011). The comparison is based on the number of iteration (NI) and the cpu time (in seconds). In Table 1 a method is considered to fail if and we report "-" if the number of iterations exceed 2000 and CPU time exceeded 10 min (600s).

In Table 1, FR and RMIL+ failed to sole 17 problems while ANN solves all the problems. In figure 1 and 2 a solver with the high values of pr(x) or at the top right of the figure are considered to be the best solver. ANN has the best performance since it can solve all the test problems. Considering both figures we can see that ANN outperforms both of the two methods as we can observe on the arrow indicating each solver on the both figures.

These show that the ANN algorithm is efficient and robust in solving large scale unconstrained optimization problems.

CONCLUSION

In this paper, we have modified a new and simple nonlinear conjugate gradient co-efficient, and also we have provide proof for the sufficient descent property and the global convergence properties and lastly we present the numerical result which suggest that ANN algorithm has the best performance when compared with FR and RMIL+.

REFERENCES

Al-Nasser, A.D.; Al-Omari, A.I. (2013) Acceptance Aini, N. Rivaie, M. Mamat, M., Sulaiman, IM. (2019) A Hybrid of Quasi-Newton Method with CG Method for Unconstrained optimization. *Journal of Physics conf. Ser.* 1366(012079)

Al-Baali, M. (1985). Descent property and global convergence of the Fletcher-Reeves method with inexact line search. IMA .*Journal of Numerical analysis* 5(1): 121-124.

Alhawarat, A., Salleh, Z., Mamat, M., Rivaie, M. (2017). An efficient modified Polak-Ribiere-Polayak conjugate gradient method with global convergence properties. *Optimization Methods Software*. 32(6), 1299-1312

Andrei, N. (2009) Accelerated conjugate gradient algorithm with finite difference Hessian / vector product approximation for unconstrained optimization. *Journal Computation and Applied Mathematics*. 230 : 570-582.

Andrei, N. (2011) 40 Conjugate gradient algorithms for unconstrained optimization. *Bull. Malays. Math. Sci*. Soc. 34 : 319-330

A.M. Awwal, L.Wang, P.Kumam,M.I. Sulaiman, S. Salisu, N. Salihu, and P.Yodjai, (2023) Generalized RMIL conjugate gradient method under the strong Wolfe line search with application in image processing, *Math. Methods Appl. Sci.* 46(16), pp. 17544–17556

Dai, Y., Yuan, (1999) Y. A nonlinear conjugate gradient with strong global convergence properties SIAM *Journal on optimization* 10(1): 177-182,

Dolan, E. D and More J.J. (2002) Benchmarking optimization software with performance profile, *Mathematical programming* 91(2): 201-213

Fletcher, R. and Reeves, C. M (1964) Function minimization by conjugate gradients, *The Computer Journal* 7(2): 149-154

Fletcher R., (1989) *Practical methods of optimization*, Chichester, England, John wiley & Sons.

Gilbert, J. C. and Nocedal, J. (1992). Global convergence properties of conjugate gradient methods for optimization. SIAM *journal on optimization* 2(1): 21-42.

Hager, W.W and Zhang, H. (2006). A survey of nonlinear conjugate gradient methods. *Pacific journal of optimization* 2(1): 35-58.

Hestenes, M.R.and Stiefel E. (1952) methods of conjugate gradient for solving linear systems, *Journal of Research of the National. Bureu of Standards* 49: 409-436.

Kabiru, A., Mohammed, Y.W., Abubakar, S.H., and Salisu M., (2024) Two RMIL- Type scheme with compressed sensing applications. *Optimization Methods and Software* https://doi.org/10.1080/10556788.2024.2425001

Liu, Y., storey, C. (1991) Efficient generalized conjugate gradient algorithms, part 1: theory, *Journal of Optimization Theory Applications* 69 (1): 129-137.

Polak. E., Ribiere, G. (1969) Note Sur la convergence de methods de direction conjugues. ESAIM: *Mathematical Modeling and Numerical Analysis Modlisation Mathematique of Analyse Numrique* 3 (RI): 35-43

Powell, M.J.D. (1977). Restart procedures for the conjugate gradient method. *Mathematical programming* 12(1): 241-254.

Powell, M.J.D. (1984) Nonconvex minimization calculation and the conjugate gradient method *lecture notes in mathematics*. 1066 Springer-Verlag. Berlin. pp. 122-141.

Powell, M.J.D. (1986) Convergence properties of algorithm for nonlinear optimization, SIAM Rev. 28: 487-500

Rivaie, M., Mamat, M., June, Ismail, M. (2012) A new class of nonlinear conjugate gradient coefficient with global convergence properties. *Applied Mathematics and Computation* 218: 0096-3003.

Rivaie, M.,Mamat, M., Abashar, A. (2015) A new class of nonlinear conjugate gradient coefficient with exact and inexact line search. *Applied Mathematics and Computation*, 268: 1152-1163.

Saleh, N.A, Sulaiman , I.M., Mamat, M., Deiby, T.S., Nelson, N. (2020) A new Hestene-Steifel and Fletcher-Reeves conjugate gradient method with Descent properties for optimization models. *International journal of Supply and operation Management*, 7(4) 344-349

Salih, Y., Hamoda, M.A., Rivaie, M. (2018) New hybrid conjugate gradient method with global convergence properties for unconstrained optimization. *Malays. Journal of Computer and Applied mathematics* 1(1), 29-38

Touati-Ahmed, D., Storey, C. (1990) Efficient hybrid conjugate gradient techniques *Journal of Optimization Theory Applications* 64(2): 379-397.

Wei, Z.G. Li. L. (2006) New nonlinear conjugate gradient method formulas for large- scaled unconstrained optimizations problems, *Applied Mathematics*, 179: 407-430.

Yuan, G. S. Lu. Z. Wei. (2010) A line search algorithm for unconstrained optimization. *Journal of Software Engineering*. 3: 503-509

Zoutendijk, G. (1970). Nonlinear programming computational methods *Integer and Nonlinear programming* 143(1): 37-86.