International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 4 Issue 4, April-2015, pp: (74-81), Impact Factor: 1.252, Available online at: www.erpublications.com

Improve Performance of Fletcher-Reeves (FR) Method

Salah Gazi Shareef[1], Ahmed Anwer Mustafa[2]
[1]Department of Mathematics, Faculty of Science, University of Zakho
[2]Department of Mathematics, Faculty of Science, University of Zakho
Kurdistan Region-Iraq

Abstract: Conjugate gradient (CG) methods are famous for solving nonlinear unconstrained optimization problems because they required low computational memory. In this paper, we propose a new conjugate gradient (β_k^{New1}) which possesses global convergence properties using exact line search and inexact line search. The given method satisfies sufficient descent condition under strong Wolfe line search. Numerical results based on the number of iterations (NOI) and number of function (NOF), have shown that the new β_k^{New1} performs better than Flecher-Reeves (FR) CG methods.

Keywords: Unconstrained optimizations, Conjugate gradient method, Sufficient Descent Condition, Global Convergent.

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:

$$\min_{x \in R^n} f(x) \quad (1.1)$$

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

$$x_{k+1} = x_k + \alpha_k d_k, k = 0,1,2,\dots \quad (1.2)$$

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

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

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:

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

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

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

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

$$\beta_k^{BA2} = \frac{y_k^T y_k}{g_k^T g_k} \quad (1.8)$$

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

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

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

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

Where, g_k and g_{k+1} g 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 [7], FR is Fletcher and Reeves [9], PR is Polak and Ribiere [4], DX is Dixon[3],BA3 is AL - Bayati, A.Y. and AL-Assady[2],

LS is Liu and Storey[11], DY is Dai and Yuan [10] , RMIL is Rivaie, Mustafa, Ismail and Leong[8] and lastly AMRI denotes Abdelrhaman Abashar, Mustafa Mamat, Mohd Rivaie and Ismail Mohd[1].

In this paper, we propose our new $\beta_k^{New\ 1}$and compared its performance with standard formulas of (FR) method .
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 sufficient descent condition and the global convergence proof of our new method. In section 4 numerical results, percentages, graphics and discussion. Lastly, In section 5 conclusion.

2. New proposed method and algorithm

In this algorithm, we modification the numerator in the proposed by Fletcher and Reeves method in 1964, where he proposed that:

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} (2.1)$$

Our proposal is

$$g_{k+1} = g_{k+1} - \gamma \frac{{g_{k+1}}^T v_k}{{v_k}^T y_k} y_k (2.2)$$

where, $\gamma \in (0,1]$

The new method is as follows:

$$\beta_k^{New\ 1} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} (2.3)$$

We programmed the new method and compared with the numerical results of the method Fletcher and Reeves and we noticed superiority of the new method that proposed on the method of Fletcher and Reeves.

2.1 Algorithm of the New1 Method

Step (1): Given $x_0 \in R^n$,$\varepsilon > 0, 0 < \gamma \leq 1$
Set k = 0, Compute f(x_0), g_0, $d_k = -g_k$

Step (2): If $\|g_{k+1}\| < \varepsilon$ stop.

Step (3): Compute $\alpha_k > 0$ satisfying the strong Wolfe condition

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

Step (4): Compute $d_{k+1} = -g_{k+1} + \beta_k^{New\ 1} d_k$.

$$g_{k+1} = g_{k+1} - \gamma \frac{{g_{k+1}}^T v_k}{{v_k}^T y_k} y_k$$

$$\beta_k^{New\ 1} = \frac{{g_{k+1}}^T g_{k+1}}{{g_k}^T g_k}$$

Step (5): If $|{g_{k+1}}^T g_k| \geq \|g_{k+1}\|^2$ go to step (1) else continue.
Step (6): Set $k = k + 1$, go to step (2)

3. The Global convergent Analysis of the New Method

The convergence properties of $\beta_k^{New\,1}$ will be studied. For an algorithm to converge, it is necessary to show that the sufficient descent condition and the global convergence properties.

3.1 Sufficient Descent Condition

For the sufficient condition to hold, then
$g_k^T d_k \le -C\|g_k\|^2 \text{for} k \ge 0 \quad \text{and} \quad C > 0 \quad (3.1)$

Theorem 3.1

Consider a CG method with search direction (1.3) and $\beta_k^{New\,1}$defined as (2.3), assume that $\alpha_k$is satisfies strong Wolfe condition then, condition (3.1) will holds for all $k \ge 0$ in both cases exact line search and inexact line search.

Proof

By using induction mathematical
If $k = 0$, then we will have $g_0^T d_0 \le -C\|g_0\|^2$. Hence condition (3.1) hold.
We need to show that for $k \ge 1$, condition (3.1), we also holds.
Now we prove the current search direction satisfies (3.1) at the iteration $(k+1)$. From (1.3) , multiply by g_{k+1} then

$$\begin{aligned} g_{k+1}^T d_{k+1} &= g_{k+1}^T(-g_{k+1} + \beta_k^{New\,1} d_{k+1}^T) \\ &= -\|g_{k+1}\|^2 + \beta_k^{New\,1} g_{k+1}^T d_k \end{aligned}$$

The proof is compete if the line search is exact, then $g_{k+1}^T d_k = 0$, and thus,

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

Which implies that d_{k+1} is a sufficient descent condition.
Now, if the line search is an inexact line search which requires $g_{k+1}^T d_k \ne 0$.
$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_k^{New\,1} g_{k+1}^T d_k (3.2)$
Put (2.2) and (2.3) in (3.2), we get

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

$$\Rightarrow \quad g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \left(d_k^T g_{k+1} - \gamma \frac{g_{k+1}{}^T v_k}{v_k{}^T y_k} d_k^T y_k \right)$$

$$\Rightarrow \quad g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \left(d_k^T g_{k+1} - \gamma \frac{\alpha_k d_k^T g_{k+1}}{\alpha_k d_k^T y_k} d_k^T y_k \right)$$

Since $d_k^T y_k$ and α_k are scalars, then
$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} d_k^T g_{k+1}(1-\gamma) \quad (3.4)$
By strong Wolfe condition, we have

$$\begin{aligned} g_{k+1}^T d_{k+1} &\le -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\|g_k\|^2}(-c_2 d_k^T g_k(1-\gamma)) \\ &= -\|g_{k+1}\|^2 + \|g_{k+1}\|^2(1-\gamma) \\ &= (c_2 - c_2\gamma - 1)\|g_{k+1}\|^2 \end{aligned}$$

Since $0 < c_2 < 1$ and $\gamma \in (0,1]$, then $(c_2 - c_2\gamma - 1) < 0$
$g_{k+1}^T d_{k+1} \le -C\|g_{k+1}\|^2 \text{where} C = c_2 - c_2\gamma - 1$ ■

Lemma 3.1

The norm of consecutive search direction are given by below expression
$\|d_{k+1}\| \le |\beta_k^{New\,1}|\|d_k\|$, for all k

Proof

From (1.3), we have
$d_{k+1} + g_{k+1} = \beta_k^{New\,1} d_k$, By take norm both sides, we have
$\|d_{k+1} + g_{k+1}\| = |\beta_k^{New\,1}|\|d_k\|$, By using triangular inequality, we get

$\|d_{k+1}\| \leq \|d_{k+1} + g_{k+1}\| = |\beta_k^{New\,1}|\|d_k\|$, Hence, we get
$\|d_{k+1}\| \leq |\beta_k^{New\,1}|\|d_k\|$, for all k

■

Lemma 3.2
The norm of search direction and the norm of gradient are the same that is
$\|d_k\|^2 = \|d_k\|^2$ (3.5)

Proof
Multiply this equation $d_k = -g_k$by g_k^T, we get
$g_k^T d_k = -\|g_k\|^2$(3.6)
By square (3.6), we have

$$(g_k^T d_k)^2 = -\|g_k\|^4 \Rightarrow \|g_k\|^2\|d_k\|^2 = \|g_k\|^4$$

Since $g_k \neq 0$, we get (3.5)

■

Lemma 3.3
The following relation holds for $k \geq 0$ in exact line search.
$\|g_{k+1} - d_k\|^2 = \|g_{k+1}\|^2 + \|d_k\|^2$(3.7)

Proof

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

Since $g_{k+1}{}^T d_k = 0$, we get (3.7)

■

3.2 Global Convergent
The following assumption are often needed to prove the convergence of the nonlinear conjugate gradient method, see [6]

Assumption1:
- **(i)** f is bounded below on the level set R^n continuous and differentiable in a neighborhood N of the level set $L = \{x \in R^n : f(x) \leq f(x_0)\}$ at the initial point x_0.
- **(ii)** The gradient $g(x)$ is Lipschitz continuous in N, so there exists a constant $L > 0$ such that $\|g(x) - g(y)\| \leq L\|x - y\|$ for any $x, y \in N$.

Based on this assumption, we have the below theorem that was proved by Zoutendijk [5]

Theorem 3.1
Suppose that assumption1 holds. Consider any conjugate gradient of the from (1.3) where d_k is a descent search direction and we take α_k in both cases exact line search and inexact line search. Then the following condition known as Zoutendijk condition holds

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty$$

From the previous information, we can obtain the following convergence theorem of the conjugate gradient methods.

Theorem 3.2
Suppose that assumption1 is true. Consider any conjugate gradient method of the form (1.3), where, α_k is obtained by both cases exact line search and inexact line search and d_k is a descent search direction then either,
$lim_{k\to\infty}\|g_k\| = 0Or\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty$

Proof
To prove Theorem3.2, we use contradiction. If Theorem 3.2 is not true, then there exists a constant $\mu > 0$, such that
$\|g_i\| \geq \mu, \forall i \geq 0.$ (3.8)
Rewrite (1.3), we get
$d_{k+1} + g_{k+1} = \beta_k^{New\,1} d_k$(3.9)
Squaring the above equation, we get

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

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

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

Substitute $\beta_k^{New\,1}$, we have

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

From Lemma 3.2, it gives us

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

Hence fork $= 0$ the above inequality yield

$$\frac{\|d_1\|^2}{(g_1^T d_1)^2} \le \frac{1}{\|g_0\|^2} + \frac{1}{\|g_1\|^2}$$

Hence for allk, we conclude that

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \le \frac{1}{\|g_0\|^2} + \frac{1}{\|g_k\|^2}$$

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)

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

We take summation both sides, we get

$$\sum_{k=0}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} \ge \mu^2 \sum_{k=0}^{\infty} \frac{1}{k} = \infty$$
$$\sum_{k=0}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} \ge \infty$$

Which contradicts Zountendijk condition in Theorem 3.1 The proof is then complete. ■

4. Numerical Results and Discussions

This section is devoted to test the implement of the new method. We compare the new conjugate gradient algorithm (New1) and standard (F/R). The comparative tests involve well known nonlinear problems (classical test function) with different function $4 \le N \le 5000$. all programs are written in FORTRAN 95 language and for all cases the stopping condition $\|g_{k+1}\|_\infty \le 1 \times 10^{-5}$
and restart using Powell condition $|g_k{}^T g_{k+1}| \ge 0.2\|g_{k+1}\|^2$.The line search routine was a cubic interpolation which uses function and gradient values. The results given in tables (4.1) and (4.2) specifically quote the number of iteration NOI and the number of function NOF. Experimental results in tables (4.1) and (4,2) confirm that the new conjugate gradient algorithm (New1) is superior to standard algorithm (F/R)
with respect to the number of iterations NOI and the number of functions NOF.

International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 4 Issue 4, April-2015, pp: (74-81), Impact Factor: 1.252, Available online at: www.erpublications.com

Comparative Performance of Two Algorithm Standard F/R and New1

Table (4.1)

No. of test	Test function	N	Standard Formula (FR)		New Formula (New1)	
			NOI	NOF	NOI	NOF
1	Rosen	4 100 500 1000 5000	30 30 30 30 30	85 85 85 85 85	29 29 29 29 30	80 80 80 80 82
2	Cubic	4 100 500 1000 5000	13 14 15 15 15	38 40 44 44 44	12 13 13 13 13	35 37 37 37 37
3	Powell	4 100 500 1000 5000	40 42 43 43 43	109 123 125 125 125	27 29 30 36 41	76 89 91 110 128
4	Wolfe	4 100 500 1000 5000	11 45 46 52 141	23 91 93 105 293	11 45 49 49 105	23 91 99 99 224
5	Wood	4 100 500 1000 5000	27 27 27 27 29	61 61 61 61 66	25 26 26 26 26	57 59 59 59 59
6	Non-diagonal	4 100 500 1000 5000	23 27 27 27 27	61 73 73 73 73	23 27 27 27 27	61 73 73 73 73

Table (4.2)

No. of test	Test function	N	Standard Formula (FR)		New Formula (New1)	
			NOI	NOF	NOI	NOF
9	G-centeral	4 100 500 1000 5000	18 24 28 28 28	123 194 251 251 251	12 16 17 17 23	65 118 131 131 213

10	Beal	4 100 500 1000 5000	11 12 12 12 12	28 30 30 30 30	11 11 11 11 11	28 28 28 28 28
11	G-full	4 100 500 1000 5000	3 141 299 392 891	7 283 599 785 1783	3 114 263 381 895	7 229 527 763 1791
12	Powell3	4 100 500 1000 5000	F F F F F	F F F F F	15 16 16 16 16	33 35 35 35 35
7	OSP	4 100 500 1000 5000	8 52 134 199 481	44 189 406 614 1572	8 48 129 185 470	42 160 410 582 1538
8	Recip	4 100 500 1000 5000	3 14 20 26 26	15 81 118 148 127	3 13 19 23 23	15 72 102 113 109
Total			**4024**	**10845**	**3688**	**9692**

Comparing the rate of improvement between the new algorithm (New1) and the standard algorithm (F/R)

Table (4.3)

Tools	**Standard algorithm (F/R)**	**New algorithm (New1)**
NOI	100%	91.6501%
NOF	100%	89.3648%

Table (4.3) shows the rate of improvement in the new algorithm (New1)with the standard algorithm (F/R), The numerical results of the new algorithm is better than the standard algorithm, As we notice that (NOI), (NOF) of the standard algorithm are about 100%, That means the new algorithm has improvement on standard algorithm prorate (8.3499%) in (NOI) and prorate (10.6352%) in (NOF), In general the new algorithm (New1) has been improved prorate (9.49256%) compared with standard algorithm (F/R).

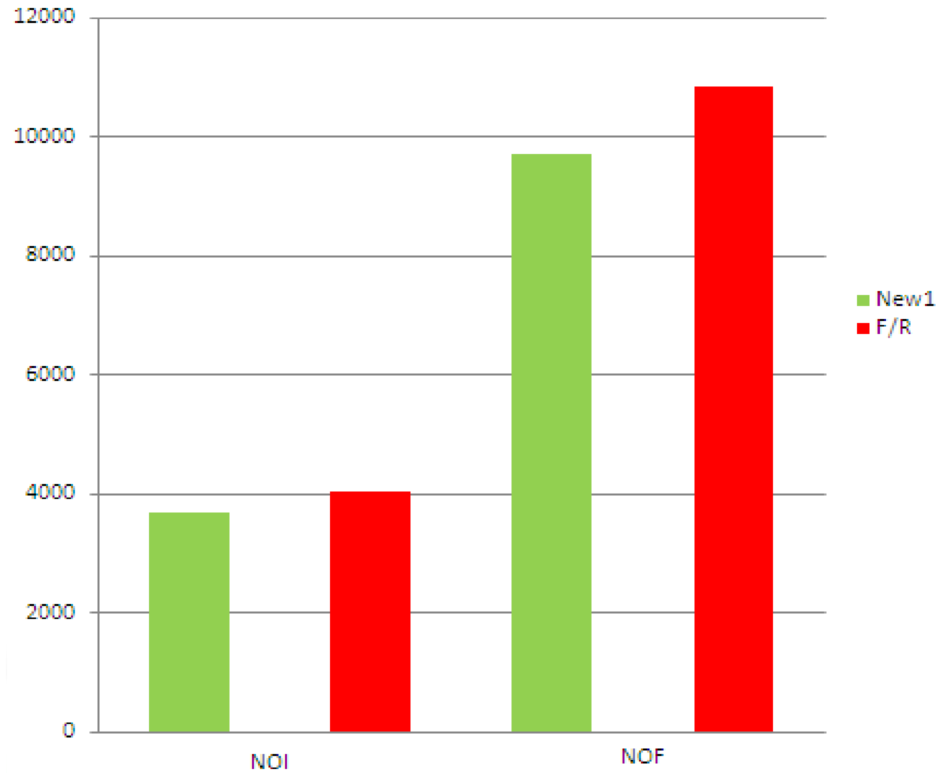

Figure (4.1): shows the comparison between new algorithm (New1) and the standard algorithm (R/F) according to the total number of iterations (NOI) and the total number of functions (NOF).

Conclusion

In this paper, we proposed a new and simple $\beta_k^{New\ 1}$ that has global convergence properties. Numerical results have shown that this new $\beta_k^{New\ 1}$performs better than FR .In the future we can improve the method to HS, PR, DX, DY, LS and other method.

References

[1]. A. Abashar, M. Mamat, M. Rivaieand I. Mohd, Global Convergence Properties of a New Class of Conjugate Gradient Method forUnconstrained Optimization, International Journal of Mathematical Analysis, 8 (2014), 3307-3319.

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

[3]. Dixon, L. G. W.,Conjugate Gradient Algorithms Quadratic Termination Without Linear Searches, Journal of Inst. Of Mathematics And its Applications,15(1975).

[4]. E. Polak and G. Ribiere, Note sur la convergence de directionsconjugees, Rev. FrancaiseInformatRechercheOperationelle. 3(16) (1969), pp. 35-43.

[5]. G. Zoutendijk, Nonlinear programming computational methods, Abadie J. (Ed.) Integer and nonlinear programming, (1970), 37-86.

[6]. M. Rivaie, A. Abashar, M. Mamat and I. Mohd, The Convergence Properties of a New Type of Conjugate Gradient Methods, Applied Mathematical Sciences, 8 (2014),33-44.

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

[8]. 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, 6 (2012), 1131-1146.

[9]. R. Fletcher, and C. Reeves, Function minimization by conjugate gradients, Comput. J. 7(1964), pp. 149-154.

[10]. Y. H. Dai and Y. Yuan, A note on the nonlinear conjugate gradient method, J. Compt. Appl. Math., 18(2002), 575-582..

[11]. Y. Liu and C. Storey, Efficient generalized conjugate gradient algorithms part 1: Theory, J. Comput. Appl. Math. 69 (1992), pp. 129-137.