International
Journal
Of Advanced Research in Engineering & Management (IJAREM)

A New Conjugate Gradient for Nonlinear Unconstrained Optimization

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— *The conjugate gradient method is a very useful technique for solving minimization problems and has wide applications in many fields. In this paper we propose a new conjugate gradient methods by) for nonlinear unconstrained optimization. The given method satisfies descent condition under strong Wolfe line search andglobal convergence property for uniformly functions.Numerical results based on the number of iterations (NOI)and number of function (NOF), have shown that the new* β_k^{New} *performs better thanas Hestenes-Steifel(HS)CG methods.*

Keywords- *Conjugate gradient method, Descent Condition, Global Convergent, Unconstrained optimizations.*

I. INTRODUCTION

In this study we consider the unconstrained minimization problem

$$\text{Min } f(x) \qquad (1.1)$$

$$x \in R^n$$

where $f : R^n \rightarrow R$ is a smooth function, and R^n denotes an n-dimensional Euclidean space. Generally, a line search method takes the form

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

where $x_k \in R^n$ is the current iterative, d_k is a descent direction of $f(x)$ at x_k and α_k is a step size. For convenience, we denote $\nabla f(x_k)$ by $g_k f(x_k)$ by f_k, $\nabla^2 f(x_k)$ by G_k. If G_k is available and inverse, then

$d_k = -G^{-1}{}_k g_k$ leads to Newton method and $d_k = -g_k$ results in the steepest descent method.

Conjugate gradient method is a very efficient line search method for solving

large unconstrained problems, due to its lower storage and simple computation.

Conjugate gradient method has the form (1.2) in which

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

Various conjugate gradient methods have been proposed, and they are mainly different in the choice of the parameter β_k . Some well-known formulas for β_k given below: β_k

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

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

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

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

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

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

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

$$\beta_k^{HZ} = \left(y_k - 2d_k \frac{\|y_k\|^2}{d_k^T y_k}\right)^T \frac{g_{k+1}}{d_k^T y_k} \tag{1.11}$$

$$\beta_k^{RMIL} = \frac{g_k^T y_k}{d_k^T (d_k - g_{k+1})} \tag{1.12}$$

$$\beta_k^{AMRI} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} |g_{k+1}^T g_k|}{\|d_k\|^2} \tag{1.13}$$

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 [5], FR is Fletcher and Reeves [8], PR is PolakandRibiere [3], CD is Conjugate Descent [9],BA2 is AL - Bayati, A.Y. and AL-Assady[2], LS is Liu and Storey [12], DY is Dai and Yuan [11], HZ is Hager and Zhang [10], RMIL is Rivaie, Mustafa, Ismail and Leong[7] and lastly AMRI denotes AbdelrhamanAbashar, Mustafa Mamat, MohdRivaie and Ismail Mohd[1].

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

II. NEW PROPOSED METHOD OF (CG)

In thissection, we propose our new β_k known as β_k^{new} . This formula is given as

$$\beta_k^{New} = \frac{\|y_k\|^2}{\|g_k\|^2} - \gamma \tag{2.1}$$

We assume that $g_{k+1} = g_{k+1} - \gamma \frac{{g_{k+1}}^T v_k}{{v_k}^T y_k} y_k \ , \ \gamma \in (0,1]$ (2.2)

$$d_{k+1} = -g_{k+1} + \beta_k^{New} d_k \qquad (2.3)$$

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

ALGORITHM OF THE NEW 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 $d_{k+1} = -g_{k+1} + \beta_k^{New} d_k$.

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

$$\beta_k^{New} = \frac{\|y_k\|^2}{\|g_k\|^2} - \gamma$$

Step (5): If $|{g_{k+1}}^T g_k| \geq 0.2\|g_{k+1}\|^2$ go to step (1) else continue.

Step (6): Set $k = k + 1$, go to step (2)

III. GLOBAL CONVERGENCE PROPERTY FOR UNIFORMLY FUNCTIONS

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 $S = \{x \in R^n: f(x) \leq f(x_0)\}$ at the initial point x_0.

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

Based on this assumption, we have the below theorem that was proved by Zoutendijk [4]

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

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

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

The convergence properties of β_k^{New} will be studied. For an algorithm to converge, we show that the descent condition and the global convergence for uniformly Functions.

Under above assumption (i) and (ii), there exists a constant μ such that

$\|\nabla f(x)\| \leq \mu$ for all $x \in S$(3.1)

Descent Condition

For the descent condition to hold, then

$g_{k+1}^T d_{k+1} \leq 0$ for $k \geq 0$ (3.2)

Proof

Multiply both sides of (2.3) by g_{k+1}, we have

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1} + \beta_k^{New} g_{k+1}^T d_k \tag{3.3}$$

Put (2.1) in (2.3), we get

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

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \frac{\|y_k\|^2}{\|g_k\|^2} g_{k+1}^T d_k - \gamma g_{k+1}^T d_k \tag{3.4}$$

We know that $d_k^T g_{k+1} \leq d_k^T y_k$

$$g_{k+1}^T d_{k+1} \leq -\|g_{k+1}\|^2 + \frac{\|y_k\|^2}{\|g_k\|^2} g_{k+1}^T d_k - \gamma d_k^T y_k \tag{3.5}$$

Put (2.2) in (3.5), we have

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

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

Since ${d_k}^T y_k$ is scalar, then

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

By Wolfe condition, we have

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

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

Since $\|g_k\|^2$ is scalar, then

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

By Powell condition, we have

$$g_{k+1}^T d_{k+1} \le -\|g_{k+1}\|^2 + \|y_k\|^2 c_2 (\gamma-1) - 0.2\gamma \|g_{k+1}\|^2 - \gamma \|g_k\|^2$$

$$\Rightarrow g_{k+1}^T d_{k+1} \le -(1+0.2\gamma)\|g_{k+1}\|^2 + \|y_k\|^2 c_2 (\gamma-1) - \gamma \|g_k\|^2$$

Since $0 < c_2 < 1, \gamma \in (0,1]$ and $\|g_{k+1}\|^2, \|y_k\|^2, \|g_k\|^2$ are greater than of zero

$\Rightarrow\ g_{k+1}^T d_{k+1} \le 0$ ■

Lemma 3.1

The norm of consecutive search direction are given by below expression

$\|d_{k+1}\| \le |\beta_k^{New}|\|d_k\|$, for all k

Proof

From (2.3), we have

$d_{k+1} + g_{k+1} = \beta_k^{New} d_k$, By take norm both sides, we have

$\|d_{k+1} + g_{k+1}\| = |\beta_k^{New}|\|d_k\|$, By using triangular inequality, we get

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

$\|d_{k+1}\| \le |\beta_k^{New\,3}|\|d_k\|$, for all k ■

Lemma 3.2

Let assumption (i) and (ii) hold and consider any conjugate gradient method (1.2) and (1.3), where d_k is descent direction and α_k is obtained by the strong Wolfe line search. If

$$\sum_{k\geq 1} \frac{1}{\| d_k \|^2} = \infty \tag{3.6}$$

Then

$$\liminf_{k\to\infty} \| g_k \| = 0 \tag{3.7}$$

For uniformly convex function which satisfies the above assumptions, we can prove that the norm of d_{k+1} given by (2.3) is bounded above. Assume that the function f is uniformly convex function, i.e., there exists a constant $\mu \geq 0$, such that for all $x, x_k \in L$

$$\left(g(x) - g(x_k)\right)^T (x - x_k) \geq \mu \|x - x_k\|^2 \tag{3.8}$$

and the step length α_k is given by strong Wolfe line search.

$$f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k {g_k}^T d_k \tag{3.9}$$

$$|\nabla f(x_k + \alpha_k d_k)^T d_k| \leq -c_2 {g_k}^T d_k \tag{3.10}$$

Using Lemma 3.2 the following result can be proved.

Theorem 3.2

Suppose that the assumptions (i) and (ii) holds. Consider the algorithms (2.1),(2.2) and (2.3) where $\gamma \in (0,1]$ and α_k is obtained by strong Wolfe line search if d_k tends to zero and there exists nonnegative constants η_1 and η_2 such that

$$\|g_k\|^2 \geq \eta_1 \|v_k\|^2 \text{and } \|g_{k+1}\|^2 \leq \eta_2 \|v_k\| \tag{3.11}$$

and f is a uniformly convex function, then

$$\lim_{k\to\infty} g_k = 0 \tag{3.12}$$

Proof

By take the absolute value of equation (2.1)

$|\beta_k^{New}| = \left|\frac{\|y_k\|^2}{\|g_k\|^2} - \gamma\right|$ By triangle inequality, we have

$|\beta_k^{New}| \leq \left|\frac{\|y_k\|^2}{\|g_k\|^2}\right| + |\gamma| \Rightarrow |\beta_k^{New}| \leq \frac{\|y_k\|^2}{\|g_k\|^2} + |\gamma|$ ByLipschitz condition and (3.11), we get

$$|\beta_k^{New}| \leq \frac{K^2\|v_k\|^2}{\eta_1\|v_k\|^2} + \gamma \Rightarrow |\beta_k^{New}| \leq \frac{K^2}{\eta_1} + \gamma \tag{3.13}$$

Since

$$\| d_{k+1}\| \leq \|g_{k+1}\| + |\beta_k^{New}|\|d_k\| \tag{3.14}$$

From (3.11) and (3.13), we have

$$\| d_{k+1}\| \leq \mu + \left(\frac{K^2}{\eta_1} + \gamma\right)\frac{1}{\alpha_k}\|v_k\| \tag{3.15}$$

We square both sides, we have

$$\| d_{k+1}\|^2 \leq \mu^2 + \left(\frac{K^2}{\eta_1} + \gamma\right)^2 \frac{1}{{\alpha_k}^2}\|v_k\|^2 \tag{3.16}$$

From assumption (i), there exist a positive constant such that

$D = max\{\|x - x_k\|, \forall x, x_k \in S\}$, then D is called diagonal of subset N

$$\| d_{k+1}\|^2 \leq \mu^2 + \left(\frac{K^2}{\eta_1} + \gamma\right)^2 \frac{1}{{\alpha_k}^2} D^2 \tag{3.17}$$

$$\| d_{k+1}\|^2 \leq \mu^2 + \left(\frac{K^2}{\eta_1} + \gamma\right)^2 \frac{1}{{\alpha_k}^2} D^2 = \psi^2 \tag{3.18}$$

$$\Rightarrow \| d_{k+1}\|^2 \leq \psi^2 \Rightarrow \frac{1}{\| d_{k+1}\|^2} \geq \frac{1}{\psi^2}$$

By take the summation $\forall k \geq 1$

$$\sum_{k\geq 1} \frac{1}{\| d_{k+1}\|^2} \geq \sum_{k\geq 1} \frac{1}{\psi^2} \Rightarrow \sum_{k\geq 1} \frac{1}{\| d_{k+1}\|^2} \geq \frac{1}{\psi^2}\sum_{k\geq 1} 1 = \infty$$

Which impels that (3.6) is true. Therefore, by lemma 3.2

$\lim \inf_{k\to\infty} \| g_k\| = 0$

Since f is uniformly convex function, then $\lim_{k\to\infty} g_k = 0$

■

IV. NUMERICAL RESULTS AND DISCUSSIONS

This section is devoted to test the implement of the new method. We compare the new conjugate gradient algorithm (New) and standard (H/S). The comparative tests involve well known nonlinear problems

(classical test function) with different function $4 \leq N \leq 5000$. all programs are written in FORTRAN 95

language and for all cases the stopping condition $\|g_{k+1}\|_{\infty} \leq 1 \times 10^{-5}$

and restart using Powell condition $|g_k{}^T g_{k+1}| \geq 0.2\|g_{k+1}\|^2$.The line search routine was a cubic interpolation which uses function and gradient values.

The results given in tables (4.1), (4.2) and (4.3) specifically quote the number of iteration NOI and the number of function NOF. Experimental results in tables (4.1), (4.2) and (4.3) confirm that the new conjugate gradient algorithm (New) is superior to standard algorithm (H/S)

with respect to the number of iterations NOI and the number of functions NOF.

Comparative Performance of Two Algorithm Standard H/S and New Formula

Table (4.1)

No. of test	Test function	N	Standard Formula (HS)		New Formula (New)	
			NOI	NOF	NOI	NOF
1	Rosen	4	27	74	17	44
		100	27	74	18	46
		500	27	74	18	46
		1000	27	74	18	46
		5000	27	74	18	46
2	Cubic	4	13	40	14	41
		100	13	40	16	45
		500	13	40	17	49
		1000	13	40	17	49
		5000	14	42	17	49
3	Powell	4	38	108	35	87
		100	40	122	36	89
		500	41	124	36	89
		1000	41	124	36	89
		5000	41	124	37	91
4	Wolfe	4	17	35	10	21
		100	49	99	47	95
		500	52	105	55	111
		1000	70	141	56	113
		5000	170	349	130	274
5	Wood	4	26	59	26	57
		100	27	61	26	57
		500	28	63	26	57
		1000	28	63	26	57
		5000	28	63	30	65
6	Non-diagonal	4	24	64	24	63
		100	29	79	27	73
		500	F	F	27	73
		1000	29	79	27	73
		5000	30	81	27	73

Table (4.2)						
No. of test	Test function	N	Standard Formula (HS)		New Formula (New)	
			NOI	NOF	NOI	NOF
7	OSP	4	8	45	8	45
		100	49	185	47	165
		500	112	353	95	287
		1000	156	473	140	426
		5000	256	774	257	781
8	Recip	4	3	15	3	15
		100	14	85	13	78
		500	20	104	18	87
		1000	21	98	22	115
		5000	32	146	28	129
9	G-centeral	4	22	159	17	115
		100	22	159	18	128
		500	23	171	18	128
		1000	23	171	18	128
		5000	28	248	21	170
10	Beal	4	11	28	10	27
		100	12	30	10	27
		500	12	30	10	27
		1000	12	30	10	27
		5000	12	30	10	27
11	Mile	4	28	85	31	89
		100	33	114	32	91
		500	40	146	33	93
		1000	46	176	33	93
		5000	54	211	45	134
12	Powell3	4	16	36	10	24
		100	16	36	11	26
		500	16	36	11	26
		1000	16	36	11	26
		5000	16	36	11	26

Table (4.3)						
No. of test	Test function	N	Standard Formula (HS)		New Formula (New)	
			NOI	NOF	NOI	NOF
13	G-full	4	3	7	3	7
		100	134	269	117	235
		500	297	595	273	547
		1000	390	781	388	777
		5000	885	1771	885	1771
Total			**3901**	**10330**	**3581**	**9065**

Comparing the rate of improvement between the new algorithm (New) and the standard algorithm (H/S)

Table (4.4)		
Tools	**Standard algorithm (H/S)**	**New algorithm (New)**
NOI	100%	91.7970%
NOF	100%	87.7541%

Table (4.4) shows the rate of improvement in the new algorithm (New) with the standard algorithm (H/S), 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 (8.203%) in (NOI) and prorate (12.2459%) in (NOF), In general the new algorithm (New) has been improved prorate (10.22445%) compared with standard algorithm (H/S).

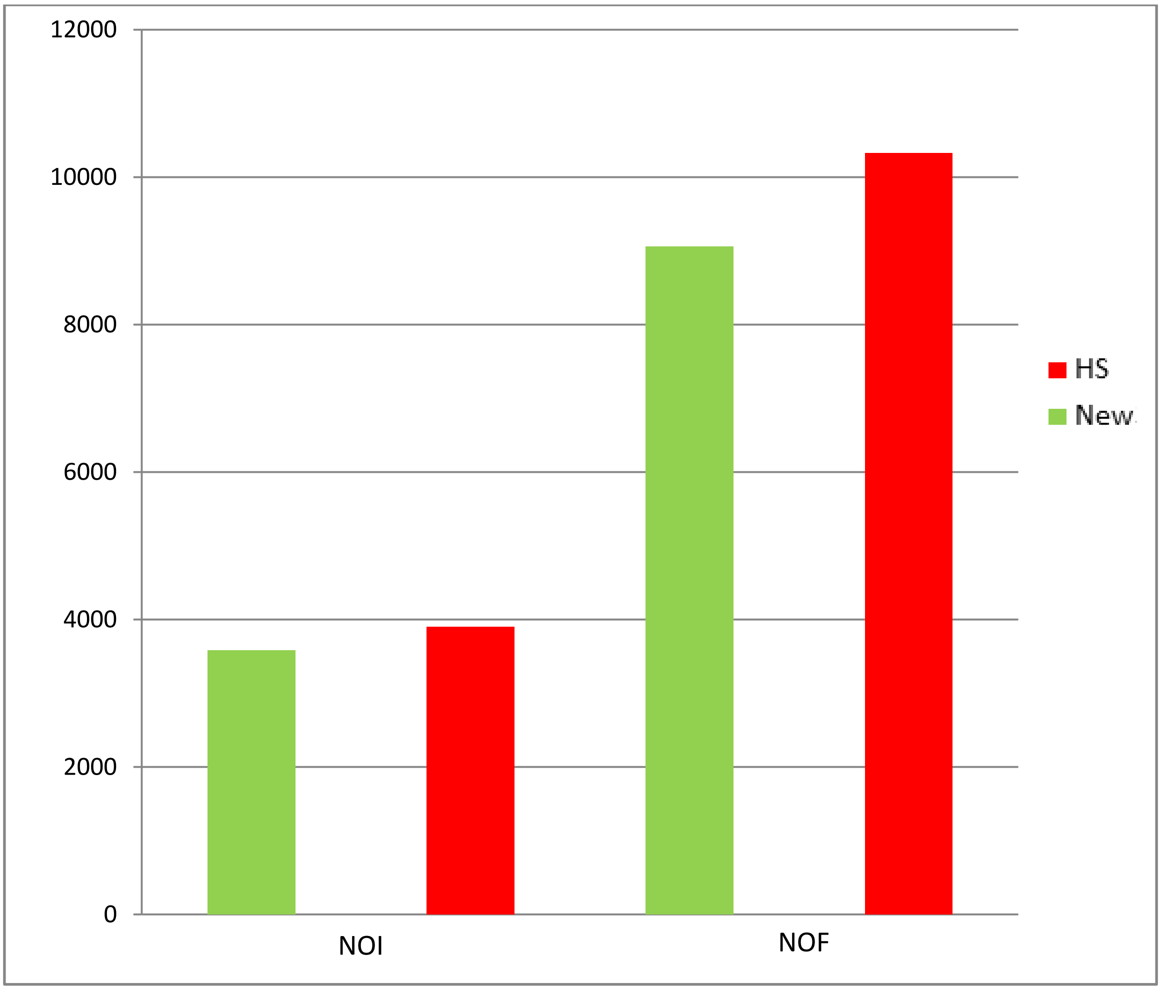

Figure (4.1) shows the comparison between new algorithm (New) and the standard algorithm (H/S) according to the total number of iterations (NOI) and the total number of functions (NOF).

V. CONCLUSION

In this paper, we proposed a new and simple β_k^{New} that has global convergence properties. Numerical results have shown that this new β_k^{New} performs better than Hestenes and E. Steifel (HS) . In the future we can we suggest formulas for γ.

References

[1]. A. Abashar, M. Mamat, M. Rivaieand I. Mohd, Global Convergence Properties of a New Class of Conjugate Gradient Method for Unconstrained 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]. E. Polak and G. Ribiere, Note sur la convergence de directionsconjugees, Rev. FrancaiseInformatRechercheOperationelle. 3(16) (1969), pp. 35-43.

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

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

[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. 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.

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

[9]. R. Fletcher, Practical method of optimization, vol 1, unconstrained optimization, John Wiley & Sons, New York, 1987.

[10]. W.W. Hager, H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, SIAM Journal on Optimization,16 (2005) 170–192

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

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