

ENHANCE THE EFFICIENCY OF RMIL'S FORMULA FOR MINIMUM PROBLEM

AHMED ANWER MUSTAFA

Dept. of Mathematics, Faculty of Education, University of Zakho, Kurdistan Region – Iraq

(Received: August 17, 2021; Accepted for Publication: October 12, 2021)

ABSTRACT.

In this paper, a new formula of β_k is suggested for conjugate gradient method of solving unconstrained optimization problems based on depends on the creation and update of RMIL'S formula with the inclusion of a parameter and step size of cubic. Our novel proposed CG-method has descent condition and global convergence properties. Numerical comparisons with standard conjugate gradient algorithm of RMIL'S formula show that this algorithm very effective depending on the number of iterations and the number of functions evaluation.

KEYWORDS: RMIL'S formula, Condition of Descent, Sufficient Descent, Global Convergent, Unconstrained Optimizations.

1. INTRODUCTION

The following unconstrained optimization questions is addressed in this study using conjugate gradient methods:

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

where $f : R^n \rightarrow R$ is continuously differentiable. Its gradient is denoted by the notes ∇f or g . Iterative techniques of the kind are commonly employed to solve unconstrained optimization issues.

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

where x_k is the present iteration's starting point, α_k is a positive step length and $d_k \in R^n$ is a search direction. d_k is generally defined by

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

The technique is described by the parameter $\beta_k \in R$. It is very well recognized that

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

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

the choice of β_k has an impact on the method's numerical performance, many researchers are looking into it.

Well-known formulas for β_k are HS is known as Hestenes and Steifel [9], FR is Fletcher and Reeves [11], PR is Polak and Ribiere [3], DX is Dixon [2], BA3 is AL - Bayati, A.Y. and AL-Assady [1], LS is Liu and Storey [12], DY is Dai and Yuan [13], RMIL is Rivaie, Mustafa, Ismail and Leong [7] [8], New by Hussein Ageel and Salah Gazi [4], MIMS is Mamat, Ibrahim and Mohammed Sulaiman [5], hybrid by Zhang, L [15] MMR is Mouiyad, Mustafa and Rivaie [6], and lastly LS+ is the modification of Liu and Storey [14] In this article, RMIL conventional formulae is compared to our novel β_k^{AA3} formula. Here are the remaining portions of the document. It is provided in section 2 as a modern conjugate gradient formula with a new algorithm technique, and in section 3 as a descent condition, sufficient descent condition, and global convergence proof. Figures, percentages, and visuals are presented in section 4. Finally, in section 5, we get to the conclusion.

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

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

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

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

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

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

$$\beta_k^{RMIL} = \frac{g_{k+1}^T y_k}{\|d_k\|^2} \quad (1.12)$$

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

$$\beta_k^{MIMS} = \frac{\frac{g_k^T y_k}{d_{k-1}^T (d_{k-1} - g_k)} + \frac{g_k^T y_k}{\|d_{k-1}\|^2}}{2} \quad (1.14)$$

$$\beta_{k+1}^{hybrid} = \frac{g_{k+1}^T (y_k - t s_k)}{\max \{y_k^T d_k, \|g_k\|^2\}} \quad (1.15)$$

$$\beta_k^{MMR} = \frac{m_k \|g_k\|^2 - (g_k^T g_{k-1})}{m_k \|g_{k-1}\|^2}, \text{ where } m_k = \frac{\|d_{k-1} + g_k\|}{\|d_{k-1}\|} \quad (1.16)$$

$$\beta_k^{LS+} = \left\{ \begin{array}{l} \frac{\|g_k\|^2 - \mu_k |g_k^T g_{k-1}|}{\|g_k\|^2}, \text{ if } \|g_k\|^2 > \mu_k |g_k^T g_{k-1}| \text{ where } \mu_k = \frac{\|x_k - x_{k-1}\|}{\|y_k\|} \\ \beta_k^{DL-HS} = -\mu_k \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}} \quad \text{otherwise} \end{array} \right\} \quad (1.17)$$

2. New proposed method and algorithm

2.1 New CJG Coefficient

In the year 2012, Rivaie, Mamat, Ismail, and Leong proposed this conjugate gradient under exact line search, see [7].

$$\beta_k^{RMIL} = \frac{g_{k+1}^T y_k}{\|d_k\|^2} \quad (2.1)$$

$$\beta_k^{AA3} = \frac{g_{k+1}^T y_k}{\|d_k\|^2} (1 - \eta \frac{g_{k+1}^T y_k}{\|d_k\|^2}) \quad (2.2)$$

where $\eta \in (0,1)$

In this research, we formulated a novel algorithm for conjugate gradient by developing and updating RMIL'S method with the extra of a specific parameter and we formulated the new method under exact line search as follows

The new direction of the search will be as follows

$$d_{k+1} = -g_{k+1} + \beta_k^{AA3} d_k \quad (2.3)$$

2.2 Algorithm of the AA3 Method

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

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

Where $0 < c_1 < c_2 < 1$

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

Step (5): Evaluate equation (2.3) by (2.2)

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)

We programmed the novel algorithm β_k^{AA3} and compared with the numerical results of the algorithm of Rivaie, Mamat, Ismail, and Leong and we noticed superiority of the fresh method (AA3) that suggested on the method of (RMIL).

3. Convergent Analysis of the New Method

The convergence properties of β_k^{AA3} 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: Consider a CJG method with search direction (1.2) and β_k^{AA3} defined as (2.2), Suppose that α_k satisfies strong Wolfe condition then, descent condition will hold for all $k \geq 0$ that is $g_{k+1}^T d_{k+1} \leq 0$.

Proof: - From (2.2) and (2.3) we have

$$d_{k+1} = -g_{k+1} + \left(\frac{g_{k+1}^T y_k}{\|d_k\|^2} \left(1 - \eta \frac{g_{k+1}^T y_k}{\|d_k\|^2} \right) \right) d_k \quad (3.1)$$

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

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{\|d_k\|^2} g_{k+1}^T d_k - \eta \frac{(g_{k+1}^T y_k)^2}{\|d_k\|^4} g_{k+1}^T d_k \quad (3.2)$$

An exact line search that needs $d_k^T g_{k+1} = 0$ can be used to determine the step length α_k . Then the proof is complete.

$$\begin{aligned} d_{k+1}^T g_{k+1} &= -\|g_{k+1}\|^2 \leq 0 \\ d_{k+1}^T g_{k+1} &= -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{d_k^T d_k} g_{k+1}^T d_k - \eta \frac{(g_{k+1}^T y_k)^2}{\|d_k\|^4} g_{k+1}^T d_k \leq 0 \quad \blacksquare \end{aligned}$$

Lemma 3.1

The norm of search direction and the norm of gradient are the same in exact line search that is

$$\|d_k\|^2 = \|g_k\|^2 \tag{3.3}$$

Proof

Multiply this equation $d_k = -g_k$ by g_k^T , we get

$$g_k^T d_k = -\|g_k\|^2 \tag{3.4}$$

By square (3.7), we have $(g_k^T d_k)^2 = -\|g_k\|^4 \Rightarrow \|g_k\|^2 \|d_k\|^2 = \|g_k\|^4 = \|g_k\|^2 \|g_k\|^2$
 Since $g_k \neq 0$, we get (3.6) ■

Global Convergent

Assuming that the following assumptions are frequently required to establish the convergence of the Procedure for nonlinear conjugate gradients.

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 [8]

Theorem 3.2

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

situations exact line searche is used. Then comes the condition called as Zoutendijk condition holds

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

For proof see [10][16]. The following conjugate gradient techniques convergence theorem may be constructed from the above information.

Theorem 3.3

Assume that the assumptions are correct. Consider any conjugate gradient strategy of the sort (1.2) and (1.22) where α_k is acquired through exact line searche, and d_k is the descent search direction than either

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

Proof

Contradiction is used to prove Theorem 3.2. It is false if Theorem 3.2., then there exists a constant $\mu > 0$, such that

$$\|g_k\| \geq \mu \tag{3.5}$$

Rewrite (2.3), we get $d_{k+1} + g_{k+1} = \beta_k^{AA3} d_k$ (3.6)

Squaring the above equation, we get

$$\|d_{k+1}\|^2 = (\beta_k^{AA3})^2 \|d_k\|^2 - 2g_{k+1}^T d_{k+1} - \|g_{k+1}\|^2 \tag{3.7}$$

Divide the two sides of the equation (3.7) by $(g_{k+1}^T d_{k+1})^2$, therefore we end up with

$$\begin{aligned} \frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} &= (\beta_k^{AA3})^2 \frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} - \frac{2}{g_{k+1}^T d_{k+1}} - \frac{\|g_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \\ &= (\beta_k^{AA3})^2 \frac{\|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} - \left(\frac{1}{\|g_{k+1}\|} + \frac{\|g_{k+1}\|}{g_{k+1}^T d_{k+1}} \right)^2 + \frac{1}{\|g_{k+1}\|^2} \\ &\leq \frac{(\beta_k^{AA3})^2 \|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2} \end{aligned}$$

Substitute β_k^{AA3} , we have

$$\begin{aligned} \frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} &\leq \frac{\left(\frac{g_{k+1}^T y_k}{\|d_k\|^2} - \eta \frac{(g_{k+1}^T y_k)^2}{\|d_k\|^4} \right)^2 \|d_k\|^2}{(g_{k+1}^T d_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2} \\ &= \frac{(g_{k+1}^T y_k)^2}{\|d_k\|^2 (g_{k+1}^T d_{k+1})^2} - 2\eta \frac{(g_{k+1}^T y_k)^3}{\|d_k\|^4 (g_{k+1}^T d_{k+1})^2} + \eta^2 \frac{(g_{k+1}^T y_k)^4}{\|d_k\|^6} + \frac{1}{\|g_{k+1}\|^2} \end{aligned}$$

Since $g_{k+1}^T y_k = \|g_{k+1}\|^2 - g_{k+1}^T g_k = c_2 \|g_k\|^2$, we know that $g_{k+1}^T d_k \leq d_k^T y_k$ and by Wolfe condition $c_2 g_k^T d_k \leq d_k^T y_k \Rightarrow -c_2 g_k^T d_k \geq -d_k^T y_k$. This implies that $\|g_k\|^2 \geq \frac{-1}{c_2} d_k^T y_k$ and by lemma 3.1 we get,

$$\begin{aligned} \frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} &\leq -\frac{c_2 (g_{k+1}^T y_k)^2}{d_k^T y_k (g_{k+1}^T d_{k+1})^2} - 2\eta \frac{(g_{k+1}^T y_k)^2 (\|g_{k+1}\|^2 + c_2 \|g_k\|^2)}{\|d_k\|^4 (g_{k+1}^T d_{k+1})^2} - \eta^2 \frac{c_2^3 (g_{k+1}^T y_k)^4}{(d_k^T y_k)^3} \\ &\quad + \frac{1}{\|g_{k+1}\|^2} \end{aligned}$$

Since $(g_{k+1}^T y_k)^2, (g_{k+1}^T d_{k+1})^2, \|g_{k+1}\|^2, \|g_k\|^2, d_k^T y_k, \eta$ and c_2 are greater than or equal zero, so

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

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

Hence for all k , we conclude that $\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{1}{\|g_k\|^2}$

Therefore $\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \sum_{i=0}^k \frac{1}{\|g_i\|^2}$ So, by (3.5)

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

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

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

Which contradicts Zountendijk condition in Theorem 3.2 The proof is then complet■

4. Numerical Results

Test the implementation of the new method in this section. We compare our method with Conjugate Gradient methods (RMIL) the comparative tests involve well-known nonlinear problems (standard test function) with different dimensions $5 \leq N \leq 5000$, 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 given in table (4.1) specifically quote the number of functions NOFS and the number of iterations NOIS. More experimental results and table (4.2) confirm that the new CG is superior to standard (RMIL'S formula) with respect to the NOIS and NOFS.

Table (4.1) :- Comparative Performance of Algorithms Standard RMIL and AA3

No. of Test	Test Functions	N	Standard Formula (RMIL)		New Formula (AA3)	
			NOIS	NOFS	NOIS	NOFS
1	G-Central	5	33	197	18	101
		50	39	265	19	113
		500	48	380	19	113
		1000	51	421	20	126
		5000	56	489	22	156
2	OSP	5	10	56	9	50
		50	39	152	38	146
		500	236	745	178	639
		1000	471	1547	315	1113
		5000	1945	6973	764	3206
3	Cubic	5	16	47	10	31
		50	16	47	10	31
		500	16	47	10	31
		1000	16	47	10	31
		5000	16	47	10	31
4	Miele	5	52	164	56	175
		50	57	229	57	177
		500	90	317	61	198
		1000	90	317	65	219
		5000	106	395	65	219
5	Wood	5	96	199	100	207
		50	103	213	116	239
		500	128	263	117	241
		1000	128	263	121	249
		5000	148	303	126	259
6	Extended PSC1	5	7	18	5	14
		50	6	16	5	14
		500	7	18	5	14
		1000	7	18	5	14
		5000	7	18	5	14
7	G-Biggs	5	126	401	33	97
		50	Fal	Fal	Fal	Fal
		500	Fal	Fal	Fal	Fal
		1000	Fal	Fal	Fal	Fal
		5000	Fal	Fal	Fal	Fal
8	Powel	5	Fal	Fal	107	242
		50	Fal	Fal	823	1799
		500	Fal	Fal	599	1345
		1000	Fal	Fal	146	372
		5000	Fal	Fal	454	1016
9	Rosen	5	Fal	Fal	Fal	Fal
		50	Fal	Fal	30	83
		500	Fal	Fal	30	83
		1000	Fal	Fal	30	83
		5000	Fal	Fal	30	83
10	Shallow	5	8	21	8	21
		50	8	21	8	21
		500	8	21	8	21
		1000	8	21	8	21
		5000	8	21	8	21
11	Non-Diagonal	5	22	58	22	58
		50	27	74	27	74
		500	27	73	27	73
		1000	27	73	27	73
		5000	27	73	27	73
12	Wolfe	5	Fal	Fal	Fal	Fal
		50	Fal	Fal	Fal	Fal
		500	Fal	Fal	Fal	Fal
		1000	Fal	Fal	Fal	Fal
		5000	218	437	218	437
Totals			9052	25717	5031	14267

ahmed.mustafa@uoz.edu.krd

Table (4.2):- Comparing the rate of improvement between the new algorithm (AA3) and the standard algorithm (RMIL)

Tools	Standard (RMIL)	New (AA3)
NOIS	100%	55.5789%
NOFS	100%	55.4769%
Tools	Standard (RMIL)	New (AA3)
NOIS	100%	55.5789%
NOFS	100%	55.4769%

Note, fail that is fail. When failure in both cases, we neglect the results. When success in the other and failure in the other case, we take for fail double the values.

Table (4.2) shows the rate of improvement in the new algorithm (AA3) with the standard algorithms (RMIL), The numerical results of the new algorithm is better than the standard algorithm, as we notice that (NOIS), (NOFS) of the standard algorithm (RMIL) are about 100%, That means the

new algorithm has improvement on standard algorithm (RMIL) prorate (44.4211%) in (NOIS) and prorate (44.5231%) in (NOFS). In general, the new algorithm (AA3) has been improved prorate (44.4721%) compared with standard algorithms (RMIL).

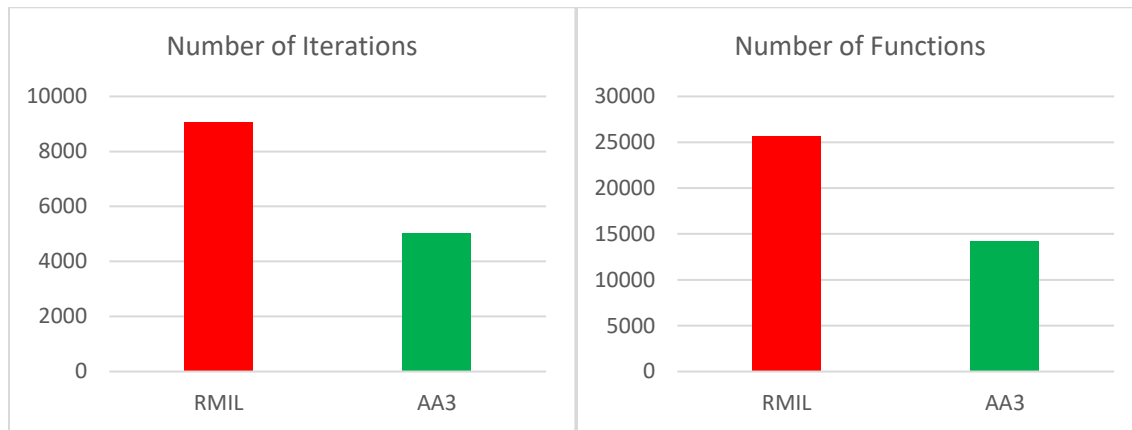


Fig. (4.1):- shows the comparison between new algorithm (AA3) and the standard algorithms (RMIL) according to the total number of iterations (NOIS) and the total number of functions (NOFS).

5. CONCLUSION

In this article, we proposed a new algorithm for CG β_k^{AA3} Which is a development of RMIL's method that has some properties of global convergence. Numerical results have shown that this new β_k^{AA3} performs better than (RMIL'S formula). In the future we can and by same way we proposed many new methods for CG of unconstrained optimization.

REFERENCES

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

Dixon, L. G. W., Conjugate Gradient Algorithms Quadratic Termination Without Linear Searches, Journal of Inst. of Mathematics And its Applications, (15) (1975), 9-15.
 E. Polak and G. Ribiere, Note sur la convergence de directionsconjugees, Rev. Francaise Informat Recherche Operationelle. 3E Annee, (16) (1969), 35-43.
 H. Ageel KH, S. Gazi SH, A New conjugate gradient method for unconstrained optimization problems with descent property, General Letters in Mathematics, 9(2) (2020), 39 -44.
 I. Mohammed S., M. Mamat, K. Kamfa, M. Danlami, A Descent Modification of Conjugate Gradient Method for Optimizatio Models, Iraqi Journal of Science, 61 (7) (2020), 1745-1750.

- M. B. Yousef, M. Mamat and M. Rivaie, A New Modification of Conjugate Gradient Method with Global Convergence and Sufficient Descent Properties for Unconstrained Optimization Problems, *ASM Science Journal*, 12(6) (2019).
- M. Rivaie, M. Mamat, L. W. June and I. Mohd, A new class of nonlinear conjugate gradient coefficients with global convergence properties, *Applied Mathematics and Computation*, 218 (2012), 11323-11332.
- 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.
- M.R. Hestenes and E. Steifel, Method of conjugate gradient for solving linear equations, *J. Res. Nat. Bur. Stand*, 6(49) (1952), 409-436.
- Nocedal J., *Theory of Algorithms for Unconstraint Optimization*, *Acta Numerica*, (1992), 199-242.
- R. Fletcher, and C. Reeves, Function minimization by conjugate gradients, *Comput. J.*, (7) (1964), 149-154.
- Y. Liu and C. Storey, Efficient generalized conjugate gradient algorithms part 1: Theory, *J. Comput. Appl. Math.*, (69) (1991), 129-137.
- Y.H. Dai & Y. Yuan, A nonlinear conjugate gradient with a strong global convergence property, *SIAM J. Optim.*, 10(1) (1999), 177-182.
- Z. Salleh, G. Alhamzi, I. Masmali and A. Alhawarat, A Modified Liu and Storey Conjugate Gradient Method for Large Scale Unconstrained Optimization Problems, *Algorithms*, 277(14) (2021)
- Zhang, L. A derivative-free conjugate residual method using secant condition for general large-scale nonlinear equations. *Numer. Algorithms* (83) (2020), 1277-1293.
- Zoutendijk G., *Nonlinear Programming Computational Methods*, *Integer and Nonlinear Programming*, J. Abadie (Ed), North-Holand. Amsterd, (1970), 37-86.