International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 5 Issue 7, July-2016

Proposed new Scaled conjugate gradient algorithm for Unconstrained Optimization

Khalil K. Abbo[1], Yoksal A. Laylani[2], Hisham M. Khudhur[3]

[1]Department of Mathematics, College of computers sciences and mathematics, University of Mosul , Mosul , Iraq
[2]Department of Mathematics College of sciences, University of Kirkuk, Kirkuk, Iraq
[3]Department of Mathematics, College of computers sciences and mathematics, University of Mosul , Mosul , Iraq

ABSTRACT

In this Research we Proposed A new Scaled conjugate gradient method, The Proposed method becomes converged by assuming some hypothesis. The numerical results show the efficiency of the developed method for solving test Unconstrained Nonlinear Optimization problems.

Keywords: conjugate gradient algorithm, Scaled conjugate gradient, Unconstrained Nonlinear Optimization

INTRODUCTION

The non-linear conjugate gradient (CG) method is a very useful technique for solving large scale unconstrained minimization problems and has wide applications in many fields [10]. This method is an iterative process which requires at each iteration the current gradient and previous direction, which is characterized by low memory requirements and strong local and global convergence properties [3 and 15].

In this paper, we focus on conjugate gradient methods applied to the non-linear unconstrained minimization problem:

$$\min\ f(x),\quad x \in R^n\ . \tag{1}$$

Where $f : R^n \to R$ is continuously differentiable function and bounded below. A conjugate gradient method generates a sequence $x_k,\ k \geq 1$ starting from an initial guess $x_1 \in R^n$, using the recurrence

$$x_{k+1} = x_k + \alpha_k d_k \tag{2}$$

Where the positive step size α_k is obtained by a line search, and the directions d_k are generated by the rule:

$$d_{k+1} = -g_{k+1} + \beta_k d_k\ ,\quad d_1 = -g_1 \tag{3}$$

where $g_k = \nabla f(x_k)$, and let $y_k = g_{k+1} - g_k$ and $s_k = x_{k+1} - x_k$, here β_k is the CG update parameter. Different CG methods corresponding to different choice for the parameter β_k see [1,4 and 12]. The first CG algorithm for non-convex problems was proposed by Fletcher and Revees (FR) in 1964 [13], which defined as

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}. \tag{4}$$

We know that the other equivalents forms for β_k are Polack-Ribeir (PR) and Hestenes- Stiefel (HS) for example

$$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{g_k^T g_k}. \quad \text{and} \quad \beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}. \tag{5}$$

Although all the above formulas are equivalent for convex quadratic functions, but they have different performance for non-quadratic functions, the performance of a non-linear CG algorithm strongly depends on coefficient β_k . Dai and Yuan (DY) in [6] proposed a non-linear CG method (2) and (3) with β_k defined as

$$\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{d_k^T y_k}. \tag{6}$$

Which generates a descent search directions

$$d_k^T g_k < 0. \tag{7}$$

At every iteration k and convergence globally to the solution if the following Wolfe conditions are used to accept the step-size α_k [2]:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k g_k^T d_k \tag{8}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq c_2 g_k^T d_k \tag{9}$$

Where $0 < c_1 < c_2 < 1$. Condition (8) stipulates a decrease of f along d_k if (7) satisfied. Condition (9) is called the curvature condition and it's role is to force α_k to be sufficiently far a way from zero [15]. Which could happen if only condition (8) were to be used. Conditions (8) and (9) are called standard Wolfe conditions (SDWC). Notice that if equation (8) satisfied then always there exists $\bar{\alpha} > 0$ such that for any $\alpha_k \in [0, \bar{\alpha}]$ the conditions (8) and (9) will be satisfied according to the theorem (1) given later. If we wish to find a point α_k, which is closer to a solution of the one dimensional problem

$$\underset{\alpha>0}{Min}\, \phi(\alpha) = \min_{\alpha>0} f\,(\mathrm{x}_k + \alpha d_k) \tag{10}$$

Than a point satisfying (8) and (9) we can impose on α_k the strong Wolfe conditions (STWC):

$$f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k g_k^T d_k \tag{11}$$

$$\left|g(x_k + \alpha_k d_k)^T d_k\right| \leq c_2 \left|g_k^T d_k\right| \tag{12}$$

Where $0 < c_1 < c_2 < 1$. In contrast to (SDWC) $g_{k+1}^T d_k$ cannot be arbitrarily large [15]. The (STWC) with the sufficient descent property

$$d_k^T g_k < -c \left\|g_k\right\|., \; c \in (0,1) \tag{13}$$

Widely used in the convergence analysis for the CG methods.

Theorem (1)

Assume that f is continuously differentiable and that is bounded below along the line $x = x_k + \alpha\, d_k$, $\alpha \in (0, \infty)$. Suppose also that d_k is a direction of descent (7) is satisfied if $0 < c_1 < c_2 < 1$ then there exist nonempty intervals of step lengths satisfying the (SDWC) and (STWC) conditions, For proof see [15].

The Fletcher-Revees (FR) and Dai-Yuan (DY) methods have common numerator $g_{k+1}^T g_{k+1}$. One theoretical difference between these methods and other choices for the update parameter β_k is that the global convergence theorems only require the Lipschitz assumption not the bounded ness assumption [10].

The global convergence for the methods with $g_{k+1}^T g_{k+1}$ in the numerator of β_k established with exact and inexact line searches for general functions [2,7 and 17]. Despite the strong convergence theory that has been developed for methods with $g_{k+1}^T g_{k+1}$ in the numerator of β_k, these methods are all susceptible to jamming, that is they begin to take small steps without making significant progress to the minimum [10]. On the other hand the convergence of the methods with $g_{k+1}^T y_k$ in the numerator (PR) and (HS) for general non-linear function are uncertain, in general the performance of these methods is better than the performance of the methods with $g_{k+1}^T g_{k+1}$ in the numerator of β_k see [10], but they have weaker convergence theorems.

This paper is organized as follows in section 2 Proposed new Scaled conjugate gradient algorithm for Unconstrained Optimization. In section 3 we will show that our algorithm satisfies descent condition for every iteration. Section 4 we will show that our algorithm satisfies Global convergence condition for every iteration. Section 5 presents numerical experiments and comparisons.

2- Proposed new Scaled conjugate gradient algorithm for Unconstrained Optimization, denoted by

$$d_{k+1}=\begin{cases}-g_{k+1}, & \text{for } k=0\\ -(\xi+\beta_{k+1}\dfrac{d_k^T y_k}{g_{k+1}^T y_k})g_{k+1}+\beta_{k+1}d_k, & \text{for } k\ge 1\end{cases} \quad (14)$$

where $\xi\in[0,1]$

New Algorithm

Step 1. Initialization. Select $x_1\in R^n$ and the parameters $0\le\xi\le 1$. Compute $f(x_1)$ and g_1. Consider $d_1=-g_1$ and set the initial guess $\alpha_1=1/\|g_1\|$.

Step 2. Test for continuation of iterations. If $\|g_{k+1}\|\le 10^{-6}$, then stop.

Step 3. Line search. Compute $\alpha_{k+1}>0$ satisfying the Wolfe line search condition (11) and (12) and update the variables $x_{k+1}=x_k+\alpha_k d_k$.

Step 4. Direction new computation, Compute $d_{k+1}=-(\xi+\beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k})g_{k+1}+\beta_{k+1}d_k$. If the restart criterion of Powell $|g_{k+1}^T g_k|\ge 0.2\|g_{k+1}\|^2$, is satisfied, then set $d_{k+1}=-g_{k+1}$ Otherwise define $d_{k+1}=d$. Compute the initial guess $\alpha_k=\alpha_{k-1}\|d_{k-1}\|/\|d_k\|$, set $k=k+1$ and continue with step2 .

3- THE DESCENT PROPERTY OF THE NEW METHOD

Below we have to show the descent property for our proposed new Scaled conjugate gradient algorithm, denoted by

$$d_{k+1}=-(\xi+\beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k})g_{k+1}+\beta_{k+1}d_k .$$

In the following Theorem(2).

Theorem (2)

The search direction d_{k+1} and β_{k+1} given in equation

$$d_{k+1}=-(\xi+\beta_{k+1}\frac{d_k^T y_k}{g_{k+1}^T y_k})g_{k+1}+\beta_{k+1}d_k \quad (**)$$

Will hold for all $k\ge 1$

Proof:-

The proof is by induction.

1- If k=1 then $g_1^T d_1<0$ $d_1=-g_1\rightarrow<0$.

there for $d_k^T y_k>0$ by wolfe conditions.

2- Let the relation $g_k^T d_k<0$ for all k .

3- We prove that the relation is true when $k=k+1$ by multiplying the equation (**) in g_{k+1}^T we obtain

$$g_{k+1}^T d_{k+1} = -(\xi + \beta_{k+1} \frac{d_k^T y_k}{g_{k+1}^T y_k}) g_{k+1}^T g_{k+1} + \beta_{k+1} g_{k+1}^T d_k \tag{15}$$

$$g_{k+1}^T d_{k+1} = -\xi - \beta_{k+1} \frac{d_k^T y_k}{g_{k+1}^T y_k} g_{k+1}^T g_{k+1} + \beta_{k+1} g_{k+1}^T d_k \tag{16}$$

$$g_{k+1}^T d_{k+1} = -\xi + \beta_{k+1} (g_{k+1}^T d_k - \frac{d_k^T y_k}{g_{k+1}^T y_k} g_{k+1}^T g_{k+1}) \tag{17}$$

Suppose that $g_{k+1}^T y_k > 0$

$$g_{k+1}^T d_{k+1} = -\xi + \beta_{k+1} (g_{k+1}^T d_k - \frac{d_k^T y_k}{g_{k+1}^T y_k} g_{k+1}^T g_{k+1}) \tag{18}$$

<0

$$g_{k+1}^T d_{k+1} < 0 \tag{19}$$

4- GLOBAL CONVERGENCE ANALYSIS

Next we will show that CG method with β_{k+1} converges globally. We need the following assumption for the convergence of the proposed new algorithm.

Assumption (1) [11]

1-Assume f is bound below in the level set $S = \{x \in R^n : f(x) \le f(x_\circ)\}$; In some Initial point.

2-f is continuously differentiable and its gradient is Lipshitz continuous, there exist $L > 0$ such that:

$$\|g(x) - g(y)\| \le L \|x - y\| \quad \forall x,y \in N \tag{20}$$

3- f is uniformly convex function, then there exists a constant $\mu > 0$ such that

$$(\nabla f(x) - \nabla f(y))^T (x - y) \ge \mu \|x - y\|^2 \text{, for any } x, y \in S \tag{21}$$

or equivalently

$$y_k^T s_k \ge \mu \|s_k\|^2 \quad \text{and} \quad \mu \|s_k\|^2 \le y_k^T s_k \le L \|s_k\|^2 \tag{22}$$

On the other hand, under Assumption(1), It is clear that there exist positive constants B such

$$\|x\| \le B \quad , \forall x \in S \tag{23}$$

$$\|\nabla f(x)\| \le \overline{\gamma} \quad , \forall x \in S \tag{24}$$

Lemma(1) [11]

Suppose that Assumption (1) and equation (23) hold. Consider any conjugate gradient method in from (2) and (3), where d_k is a descent direction and α_k is obtained by the strong Wolfe line search. If

$$\sum_{k>1} \frac{1}{\|d_{k+1}\|^2} = \infty \tag{25}$$

then we have

$$\liminf_{k \to \infty} \|g_k\| = 0 \tag{27}$$

More details can be found in [1,8 and 14].

Theorem (3)

Suppose that Assumption (1) and equation (23) and the descent condition hold. Consider a conjugate gradient method in the form

$$d_{k+1} = -(\xi + \beta_{k+1} \frac{d_k^T y_k}{g_{k+1}^T y_k}) g_{k+1} + \beta_{k+1} d_k \tag{27}$$

where α_k is computed from Wolfe line search condition (11) and (12), If the objective function is uniformly on set S, then $\lim_{n \to \infty} \inf \| g_k \| = 0$.

Proof:-

Firstly, we need substituting our β_{k+1}, in the direction d_{k+1} there for we obtain

$$d_{k+1} = -(\xi + \beta_{k+1} \frac{d_k^T y_k}{g_{k+1}^T y_k}) g_{k+1} + \beta_{k+1} d_k \tag{28}$$

After simplify above equation we get

$$d_{k+1} = -\xi - \beta_{k+1} \frac{d_k^T y_k}{g_{k+1}^T y_k} g_{k+1} + \beta_{k+1} d_k \tag{29}$$

$$\| d_{k+1} \|^2 = \left\| -\xi g_{k+1} - \beta_{k+1} \frac{d_k^T y_k}{g_{k+1}^T y_k} g_{k+1} + \beta_{k+1} d_k \right\|^2 \tag{30}$$

Suppose that $a = \frac{d_k^T y_k}{g_{k+1}^T y_k}$

$$\| d_{k+1} \|^2 \le \xi \| g_{k+1} \|^2 + a \beta_{k+1} \| g_{k+1} \|^2 + \beta_{k+1} \| d_k \|^2 \tag{31}$$

$$\| d_{k+1} \|^2 \le (\xi + a \beta_{k+1}) \| g_{k+1} \|^2 + \beta_{k+1} \| d_k \|^2 \tag{32}$$

Suppose that $b = \xi + a \beta_{k+1}$

$$\| d_{k+1} \|^2 \le b \| g_{k+1} \|^2 + \beta_{k+1} \| d_k \|^2 \tag{33}$$

$$\| d_{k+1} \|^2 \le b \overline{\gamma}^2 + \beta_{k+1} \| d_k \|^2 \tag{34}$$

$$\| d_{k+1} \|^2 \le \frac{1}{\overline{\gamma}^2} (b (\overline{\gamma}^2)^2 + \overline{\gamma}^2 \beta_{k+1} \| d_k \|^2) \tag{35}$$

Suppose that $c = (b (\overline{\gamma}^2)^2 + \overline{\gamma}^2 \beta_{k+1} \| d_k \|^2)$

$$\| d_{k+1} \|^2 \le c \frac{1}{\overline{\gamma}^2} \tag{36}$$

$$\sum_{k=1}^{\infty}\frac{1}{\|d_{k+1}\|^2}\le\frac{1}{c}\gamma^{-2}\sum_{k\ge1}1=\infty \qquad (37)$$

$$\lim_{k\to\infty}\|g_k\|=0 \qquad (38)$$

5- NUMERICAL RESULTS AND COMPARISONS

In this section, we report some preliminary numerical results. We compare with Classical conjugate gradient direction methods. we compare the performance of new formal d_{k+1} Proposed new Scaled conjugate gradient algorithm for Unconstrained Optimization to classical direction conjugate gradient algorithm by using β_k^{FR} . we have selected (70) large scale unconstrained optimization problem, for each test problems taken from (Andrei, 2008) [5]. For each test function we have considered numerical experiments with the number of variables $n=1000\ ,...,\ 10000$. These new versions are compared with well-known classical direction conjugate gradient algorithm. All these algorithms are implemented with standard Wolfe line search conditions (11) and (12) with. In all these cases, the stopping criteria is the $\|g_k\|=10^{-6}$. All codes are written in doble precision FORTRAN Language with F77 default compiler settings. The test functions usually start point standard initially summary numerical results recorded in the figures (1),(2),(3) by matlab. The performance profile by Dolan and More´ [9] is used to display the performance of the Proposed new Scaled direction conjugate gradient algorithm with classical direction conjugate gradient algorithm by using β_k^{FR}. Define $p=700$ as the whole set of n_p test problems and $S=2$ the set of the interested solvers. Let $l_{p,s}$ be the number of objective function evaluations required by solver s for problem p . Define the performance ratio as

$$r_{p,s}=\frac{l_{p,s}}{l_p^*} \qquad (52)$$

Where $l_p^*=\min\{l_{p,s}:s\in S\}$. It is obvious that $r_{p,s}\ge1$ for all p,s . If a solver fails to solve a problem, the ratio $r_{p,s}$ is assigned to be a large number M . The performance profile for each solver s is defined as the following cumulative distribution function for performance ratio $r_{p,s}$,

$$\rho_s(\tau)=\frac{size\{p\in P:r_{p,s}\le\tau\}}{n_p} \qquad (53)$$

Obviously, $p_s(1)$ represents the percentage of problems for which solver s is the best. See [9] for more details about the performance profile. The performance profile can also be used to analyze the number of iterations, the number of gradient evaluations and the cpu time. Besides, to get a clear observation, we give the horizontal coordinate a log-scale in the following figures.

Notes:-

1- By using Wolfe conditions (11) and (12) to choose α_k [16].

2- $\xi=0.01$.

3- FR is $d_{k+1}=-(\xi+\beta_{k+1}^{FR}\frac{d_k^Ty_k}{g_{k+1}^Ty_k})g_{k+1}+\beta_{k+1}^{FR}d_k$

FRC is $d_{k+1}=-g_{k+1}+\beta_{k+1}^{FR}d_k$.

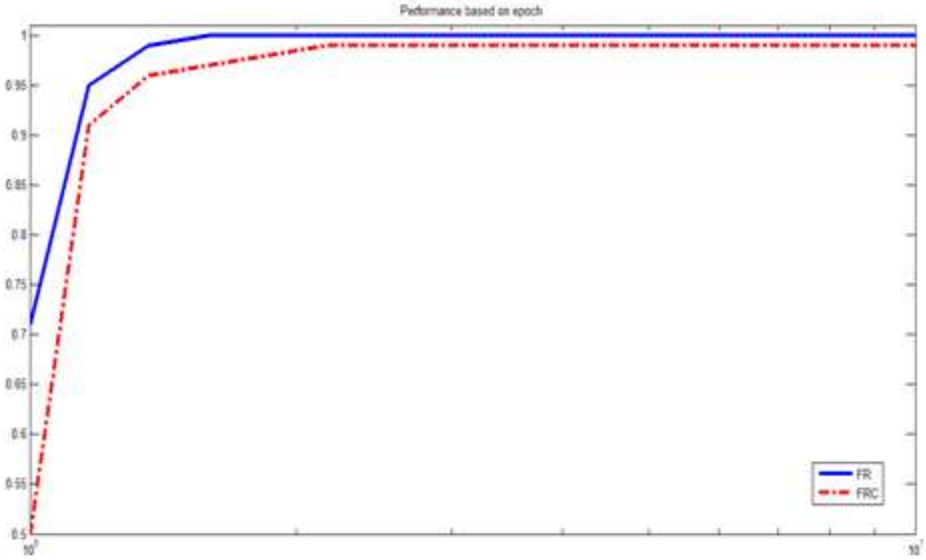

Figure (1): Performance based on iteration

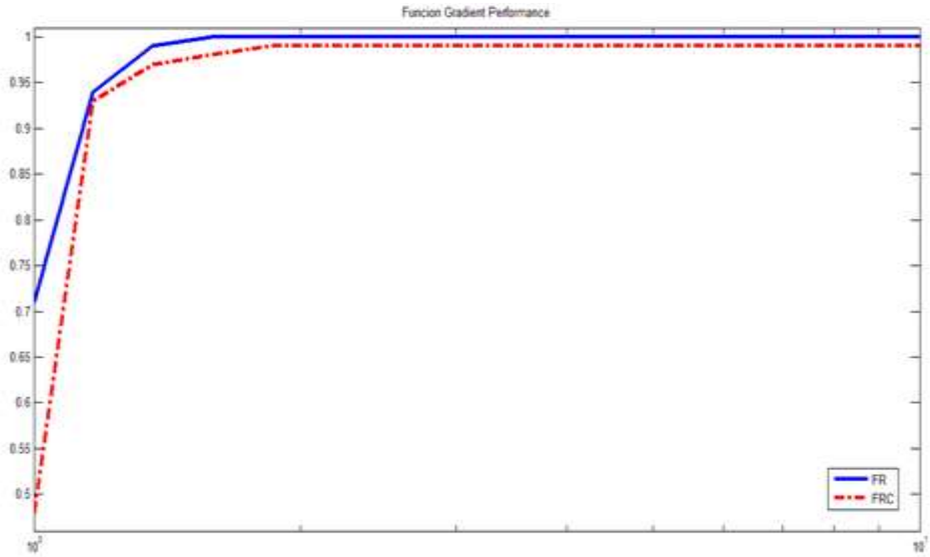

Figure (2): Performance based on Function

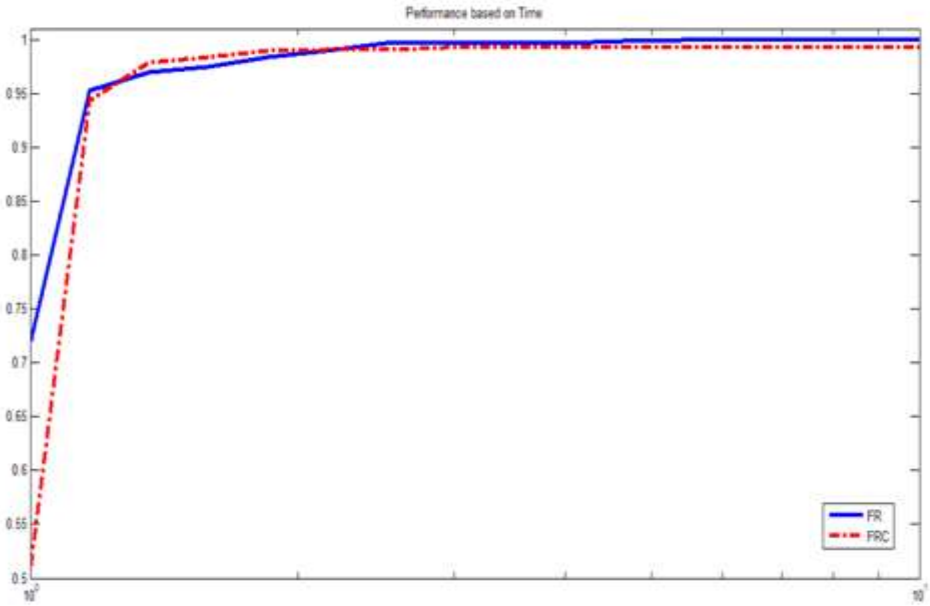

Figure (3) Performance based on Time

REFERENCES

[1]. Abbas Y. Al-Bayati, Abdul-Ghafoor J. and Khalil A. (2009) "Two- Version of conjugate gradient algorithm Based on conjugacy Conditions for unconstrained optimization". American J. of Economics and Business Administration 1, (2).

[2]. Al-Baali M. (1985) "Descent property and global convergence of the Fletcher and Revees Method with inexact line search". SIAM J. Numer. Anal 5.

[3]. Andrei, N. (2010) "Accelerated scaled memory less BFGS preconditioned conjugate Gradient algorithm for unconstrained optimization" European J. Of Operational Research. 204.

[4]. Andrei, N. (2008) "40 conjugate gradient algorithm for unconstrained Optimization" ICI. Technical Report No. 13108.

[5]. Andrei N. (2008), "An Unconstrained Optimization test function collection". Adv. Model. Optimization. 10. pp.147-161.

[6]. Dai, Y. and Yuan, Y. (1999) "A Nonlinear conjugate gradient method with a strong global convergence property". SIAM J. Optim.,10.

[7]. Dai Y. (2001) "New properties of anon-linear conjugate gradient method" Numer. Math. 89.

[8]. Dai,Y. and Liao, L. (2001) "New conjugacy conditions and related non-linear CG methods" Appl. Math optim.,(43). Spring-verlag, New York.

[9]. Dolan. E. D and Mor´e. J. J, "Benchmarking optimization software with performance profiles", Math. Programming, 91 (2002), pp. 201-213.

[10]. Hager, W. and Zhang, H. (2006) "A survey of non-linear conjugate gradient methods" Pacific Journal of Op., 2.

[11]. Hameed, M. Sadiq., (2013)," Modified Conjugate Gradient Algorithms For Solving Unconstraint Minimization Problems" M.Sc. Thesis, College of Computer Sciences And Mathematics, University of Mosul

[12]. Khalil A. (2008) "New CG method for large-scale unconstrained optimization based on Nazareth theorem". Iraq J. of Statistical science, (13).

[13]. Fletcher R. and Reeves G. (1964) "Function minimization by conjugate gradients" computer. Journal, Vol.(7).

[14]. Tomizuka. H and Yabe. H, (2004). "A Hybrid Conjugate Gradient method for unconstrained Optimization".

[15]. Pytlak R. (2009) "Conjugate Gradient Algorithms in Non-convex Optimization" Non-convex Optimization and It's Applications. V. 89, Springer-V.

[16]. Wolfe, M. (1978)."Numerical Methods For Unconstrained Optimization An Introduction". New York: Van Nostraned Reinhold

[17]. Zoutendijk, G. (1970) "Nonlinear programming, computational methods" In Integer and Nonlinear Programming, J. Abadie(Ed), North- Holand.