New Conjugacy Coefficient for Conjugate Gradient Method for Unconstrained Optimization

Hamsa TH. Chilmeran **Huda Y. Najm**
hamsathrot@uomosul.edu.iq *huda.najm@uod.ac*
College of Computer Sciences and Mathematics
University of Mosul, Mosul, Iraq

Received on: 18/04/2011 **Accepted on: 21/06/2011**

ABSTRACT

In this paper, we derived a new conjugacy coefficient of conjugate gradient method which is based on non-linear function using inexact line searches. This method satisfied sufficient descent condition and the converges globally is provided. The numerical results indicate that the new approach yields very effective depending on number of iterations and number of functions evaluation .

Keywords: unconstrained optimization, conjugate gradient method, inexact line search, global convergence, and strong wolf condition.

معامل ترافق جديد لطريقة التدرج المترافق للأمثلية غير المقيدة

همسة ثروت سعيد **هدى يونس نجم**
كلية علوم الحاسوب والرياضيات، جامعة الموصل

تاريخ استلام البحث: 2011/04/18 **تاريخ قبول البحث: 2011/06/21**

الملخص

تم في هذا البحث اشتقاق معامل ترافق جديد لطريقة التدرج المترافق التي تعتمد على الدوال غير الخطية باستخدام خط بحث غير مضبوط. هذه الطريقة حققت شرط الانحدار الكافي كما أن التقارب الشامل لهذه الطريقة تم برهانه. دلت النتائج العددية على أن الطريقة الجديدة تحقق نتائج عددية ذات كفاءة جيدة بالاعتماد على عدد التكرارات وعدد حسابات الدالة.

الكلمات المفتاحية: أمثلية غير مقيدة، طريقة التدرج المترافق، خط بحث غير تام، تقارب شامل، شرط وولف القوي.

1. Introduction

The history of conjugate gradient method began with seminar paper of Hestenes and Stiefele in [10] who presented an algorithm for solving symmetric, positive definite linear algebraic systems. In [8] Fletcher and Reeves extended the domain of application of CG method to non-linear problems, thus starting the non-linear conjugate gradient research direction.

The conjugate gradient method represents a major contribution to the panoply of methods for solving large-scale unconstrained optimization problems. They are characterized by low memory requirements and have strong global convergence properties. The popularity of these methods is remarkable partially due to their simplicity both in their algebraic expression and in their implementation in computer codes, and partially due to their efficiency in solving large-scale unconstrained optimization problems.[4]

Let function $f : R^n \to R$ be continuously differentiable, Consider the unconstrained optimization problem

$$\min\{ f(x) : x \in R^n \} \qquad \ldots(1)$$

we denote by a conjugate gradient method which generates a sequence of iterates by letting:

$x_{k+1} = x_k + \lambda_k d_k$ …(2)

Where, λ_k is a step length which is computed by carrying out a line search and d_k is the search direction defined by:

$$\begin{bmatrix} d_1 = -g_1 & k = 1 \\ d_{k+1} = -g_{k+1} + \beta_k d_k & k \geq 1 \end{bmatrix}$$ …(3)

Where, $g(x)$ denotes the gradient of $f(x)$ at x_k ,and $\beta_k \in R$ is known as a conjugate gradient coefficient , some well-known formulas are given as follows:

$\beta_k^{HS} = \dfrac{g_{k+1}^T y_k}{d_k^T y_k}$ (Hestenes-Stiefel, [10],(1952))

$\beta_k^{FR} = \dfrac{g_{k+1}^T g_{k+1}}{g_k^T g_k}$ (Fletcher-Reeves (FR),[8] ,(1964))

$\beta_k^{PR} = \dfrac{g_{k+1}^T y_k}{g_k^T g_k}$ (Polak- Ribière (PR) ,[14],(1969))

$\beta_k^{BA} = \dfrac{-y_k^T y_k}{d_k^T g_k}$ (Al-Bayati & Al-Assady ,[2], 1986)

$\beta_k^{CD} = \dfrac{g_{k+1}^T g_{k+1}}{-d_k^T g_k}$ (Fletcher (CD),[8] ,(1987))

$\beta_k^{LS} = \dfrac{g_{k+1}^T y_k}{-d_k^T g_k}$ (Liu-Storey (LS),[12],(1991))

$\beta_k^{DY} = \dfrac{g_{k+1}^T g_{k+1}}{d_k^T y_k}$ (Dai-Yuan (DY),[5],(1999))

Where, $y_k = g_{k+1} - g_k$ and $\|.\|$ stands for the Euclidean norm. There are numerous research on convergence properties of these methods. The corresponding conjugate gradient methods can be abbreviated as HS, FR, PR, CD, LS, and DY methods. Although these methods are identical when f is a strong convex quadratic function and line search is exact, they have different performances when applied to minimizing general nonlinear functions with inexact line searches.

The most studied properties of CG are its global convergence properties. Zoutendijk [19] proved the global convergence of FR method. Al-Baali [1], Touati-Ahmed and Storey [17], Gilbert and Nocedal [9] has further analyzed the global convergence of algorithms related to the FR method with strong Wolfe condition. Powell [15] also proved that FR is a superior method compared to others.

There are several line search rules for choosing step-length λ_k, (see [16]) for example, exact minimization rule, Armijo rule, Goldstein rule, Wolfe rule, etc. In this paper we analyze the general results on convergence of line search methods with the following two line search rules:

The weak Wolfe-conditions:

$f(x_k + \lambda_k d_k) - f(x_k) \leq \delta \lambda_k g_k^T d_k$ …(4)

$g(x_k + \lambda_k d_k)^T d_k \geq \sigma g_k^T d_k$ …(5)

the strong Wolfe-conditions:

$f(x_k + \lambda_k d_k) - f(x_k) \leq \delta \lambda_k g_k^T d_k$ …(6)

$$\left|g(x_k+\lambda_k d_k)^T d_k\right| \le -\sigma g_k^T d_k \qquad \ldots(7)$$

where $\delta \in (0,1)$ and $\sigma \in (\delta, \frac{1}{2})$

In [5] Dai and Yuan proposed a conjugate gradient method which generates a descent search direction at every iteration and converges globally to the solution if the Wolfe conditions are satisfied within the line search strategy. In this paper, we give a new conjugate gradient method and show that our method always produces a descent search direction and global converges if the Wolfe conditions are satisfied.

2. Extension of Conjugacy Coefficient for Conjugate Gradient Method:

In [18], Yabe and Sakaiwa extended the Dai-Yuan Method by Supposing that the current search direction d_k is a descent direction, namely $g_k^T d_k < 0$ at the kth iteration. Now, then needs to find a β_{k+1} that produces a descent search direction d_{k+1}. This requires that

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_{k+1} g_{k+1}^T d_k < 0 \qquad \ldots(8)$$

Letting τ_{k+1} be a positive parameter, then define

$$\beta_{k+1} = \frac{\|g_{k+1}\|^2}{\tau_{k+1}} \qquad \ldots(9)$$

Equation (8) is equivalent to

$$\tau_{k+1} > g_{k+1}^T d_k \qquad \ldots(10)$$

Taking the positivity of τ_{k+1} into consideration, they have

$$\tau_{k+1} > \max\{g_{k+1}^T d_k, 0\} \qquad \ldots(11)$$

Therefore, if condition (11) is satisfied for all *k*, the conjugate gradient method with (9) produces a descent search direction at every iteration. From (9), we can get various kinds of conjugate gradient methods by choosing various τ_{k+1}, where τ_{k+1} satisfying (11) and prove global convergence of the proposed method. We note that the Wolfe condition (5) guarantees $d_k^T y_k > 0$ and that

$$d_k^T y_k = d_k^T g_{k+1} - d_k^T g_k > d_k^T g_{k+1}$$

This implies that

$$d_k^T y_k > \max\{g_{k+1}^T d_k, 0\} \qquad \ldots(12)$$

By setting $\tau_{k+1} = d_k^T y_k$ formula (9) reduces to this DY method:

$$\beta_{k+1}^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k}$$

It follows from (3) and (9) that

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

The above relation can be rewritten as

$$\beta_{k+1} = \frac{g_{k+1}^T d_{k+1}}{-\tau_{k+1} + g_{k+1}^T d_k} \qquad \ldots(13)$$

Recall that if we set $\tau_{k+1} = d_k^T y_k$, this method reduces to the DY method

3. New Conjugacy Coefficient for Conjugate Gradient Method:

In this section, we are going to study the development of a new CG-method based on non-linear function taking the idea of inexact line searches.

let $\beta_k^{HS} = \dfrac{g_{k+1}^T y_k^*}{d_k^T y_k^*}$ (Hestenes –Stiefel) using $y_k^* = y_k + m s_k$ where $y_k = g_{k+1} - g_k$ and $m \le 10^{-6}$ (see[3]).

letting τ_k be a positive parameter we define:

$$\beta_k = \frac{g_{k+1}^T y_k^*}{\tau_k}$$

$$d_{k+1} = -g_{k+1} + \beta_k d_k$$

$$d_{k+1}^T y_k = -g_{k+1}^T y_k + \beta_k d_k^T y_k$$

In [14] suggested the following Perry's conjugacy condition:

$d_{k+1}^T y_k = -t\, g_{k+1}^T s_k$, $t > 0$ is a scalar

where, $s_k = x_{k+1} - x_k = \lambda_k d_k$.

$$-t\, g_{k+1}^T s_k = -g_{k+1}^T y_k + \beta_k d_k^T y_k$$

Submit every y_k by y_k^* , we get:

$$-t\, g_{k+1}^T s_k = -g_{k+1}^T y_k^* + \beta_k d_k^T y_k^*$$

$$-t\, g_{k+1}^T s_k = -\tau_k\, \beta_k + \beta_k d_k^T y_k^*$$

$$-t\, g_{k+1}^T s_k = \beta_k (-\tau_k + d_k^T (g_{k+1} - g_k + m s_k))$$

$$\therefore \beta_k^{New} = \frac{-t\, g_{k+1}^T s_k}{-\tau_k + d_k^T g_{k+1} - d_k^T g_k + m\lambda_k \|d_k\|^2} \qquad \ldots(14)$$

where, τ_k is a positive parameter.

since any β_k must be positive for this reason, we suppose the formula such as:

$$\beta_k^{New} = \begin{Bmatrix} \beta_k^{New} & if\ \beta_k^{New} > 0 \\ 1 - \beta_k^{New} & if\ \beta_k^{New} < 0 \end{Bmatrix} \qquad \ldots(15)$$

3.1 The New Algorithm :

Step 1: For the initial point τ_0 , $x_1 \in R^n$, ε , Set $d_1 = -g_1$,$k = 1$, if $\|g_1\| \le \varepsilon$, then stop.

Step 2: Set $d_k = -g_k$

Step 3: Find $\lambda_k > 0$ satisfying the wolf conditions.

Step 4: Let $x_{k+1} = x_k + \lambda_k d_k$ and If $\|g_{k+1}\| \le \varepsilon$ then stop .

Step 5: Compute β_k by the formula (15), then generate d_{k+1} by (3), and set $\tau_k = \tau_k + \varepsilon$, if ($\tau_k \ge 1$) then set $\tau_k = 0.5$.

Step 6 : If $k = n \ or \ \frac{|g_k^T g_{k+1}|}{\|g_{k+1}\|^2} \geq 0.2$,then go to step 1.

Step 7: Set *k*: = *k+1*, go to Step 2.

4. Global Convergence Properties of New Methods:

In this section, the convergence properties of new algorithm with the inexact line search analyze and in order to ensure the sufficient descent condition, using wolf condition line search.

In the global convergence analysis of many iterative methods, the following assumption is often needed:

Assumption(A) :

(i) f is bounded below on the level set $\Omega = \{x \in R^n : f(x) \leq f(x_0)\}$.

(ii) In some neighborhood Ω_0 of Ω , f is differentiable and it is gradient $g(x)$ is Lipschitz continuous, namely, there exists a constant $L \succ 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\| \ , \ \forall \, x, y \in N \qquad \ldots(16)$$

Under these assumptions on *f* there exists a constant $\varepsilon > 0$ such that

$$\|g_k\| \leq \varepsilon \qquad \forall \ k \qquad \ldots(17)$$

Lemma (1) :

Suppose the assumption (A) hold , let the sequence $\{ x_k \}$ generated by new algorithm and the step length λ_k satisfies wolf conditions, then

$$g_{k+1}^T d_{k+1} \leq -\delta \|g_{k+1}\|^2 \qquad \ldots(18)$$

where, δ is a positive constant.

Proof:

We prove the theorem with Wolfe conditions, by induction, For initial direction $(k = 1)$ we have

$$d_1 = -g_1 \ \rightarrow \ d_1^T g_1 = -\|g_1\|^2 < 0$$

Suppose $d_k^T g_k < 0 \qquad \forall k$

Now we prove if $k = k + 1$ then:

$$d_{k+1} = -g_{k+1} + \beta_k d_k \qquad \ldots(19)$$

multiply both sides of the above relation by g_{k+1}^T we get :

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

dividing both sides of the above relation by $\|g_{k+1}^T\|^2$ we get :

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} + 1 = \beta_k \, g_{k+1}^T d_k$$

submit β_k from (14) in above relation, we have :

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} + 1 = \frac{-t \, g_{k+1}^T s_k}{-\tau_k + d_k^T g_{k+1} - d_k^T g_k + m\lambda_k \|d_k\|^2} \ \frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2}$$

put $s_k = \lambda_k d_k$ we get :

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} + 1 = \frac{-t\lambda_k g_{k+1}^T d_k}{-\tau_k + d_k^T g_{k+1} - d_k^T g_k + m\lambda_k \|d_k\|^2} \frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2}$$

By using strong Wolfe condition, we get:

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \le \frac{-t\sigma^2 \lambda_k g_k^T d_k}{-\tau_k + d_k^T g_{k+1} - d_k^T g_k + m\lambda_k \|d_k\|^2} \frac{g_k^T d_k}{\|g_{k+1}\|^2}$$

since, $g_k^T d_k \le 0$ and $\|g_{k+1}\|^2 \ge 0$ then, there exists a positive constant c such that $g_k^T d_k \le -c\|g_{k+1}\|^2$ where $0 \le c \le 1$.

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \le \frac{t c\sigma^2 \lambda_k g_k^T d_k}{-\tau_k + d_k^T g_{k+1} - d_k^T g_k + m\lambda_k \|d_k\|^2}$$

$$\frac{\|g_{k+1}\|^2}{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2} \ge \frac{-\tau_k + d_k^T g_{k+1} - d_k^T g_k + m\lambda_k \|d_k\|^2}{t c\sigma^2 \lambda_k g_k^T d_k}$$

$$\frac{\|g_{k+1}\|^2}{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2} \ge \frac{-\tau_k}{t c\sigma^2 \lambda_k g_k^T d_k} + \frac{\sigma d_k^T g_k}{t c\sigma^2 \lambda_k g_k^T d_k} - \frac{d_k^T g_k}{t c\sigma^2 \lambda_k g_k^T d_k} + \frac{m\lambda_k \|d_k\|^2}{t c\sigma^2 \lambda_k g_k^T d_k}$$

since, $d_k = -g_k$ then:

$$\frac{\|g_{k+1}\|^2}{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2} \ge \frac{\tau_k}{t c\sigma^2 \lambda_k \|g_k\|^2} + \frac{\sigma}{t c\sigma^2 \lambda_k} - \frac{1}{t c\sigma^2 \lambda_k} - \frac{m\lambda_k \|d_k\|^2}{t c\sigma^2 \lambda_k \|d_k\|^2}$$

$$\frac{\|g_{k+1}\|^2}{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2} \ge \left(\frac{\tau_k}{t c\sigma^2 \lambda_k \|g_k\|^2} + \frac{1}{t c\sigma \lambda_k}\right) - \left(\frac{1}{t c\sigma^2 \lambda_k} + \frac{m}{t c\sigma^2}\right)$$

$$\frac{\|g_{k+1}\|^2}{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2} \ge \left(\frac{\tau_k + \sigma\|g_k\|^2}{t c\sigma^2 \lambda_k \|g_k\|^2}\right) - \left(\frac{1 + \lambda_k m}{t c\sigma^2 \lambda_k}\right) \ge -\left(\frac{1 + \lambda_k m}{t c\sigma^2 \lambda_k}\right)$$

$$\frac{\|g_{k+1}\|^2}{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2} \ge -\left(\frac{1 + \lambda_k m}{t c\sigma^2 \lambda_k}\right)$$

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \le -\left(\frac{t c\sigma^2 \lambda_k}{1 + \lambda_k m}\right)$$

Let $c = \left(\dfrac{t c\sigma^2 \lambda_k}{1 + \lambda_k m}\right)$

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \le -c$$

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

where, $\delta = c + 1$

$\therefore g_{k+1}^T d_{k+1} \le -\delta \|g_{k+1}\|^2$

The following lemma, called the Zoutendijk condition, which is often used to prove global convergence of conjugate gradient methods. It was originally given by Zoutendijk.

Lemma (2) : Suppose that assumptions (i) and (ii) hold. Consider the methods in the form of (2) and (19), where d_k satisfies $g_k d_k < 0$ for all k, and λ_k is obtained by (4)-(5) or (6)-(7) then have:

$$\sum_{k\ge1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty \qquad \ldots(20)$$

For proof see [19] or [6].

To ensure that an algorithm converges to a point x where $g(x) \neq 0$, we need not only a well-chosen step lengths but also, a well-chosen search directions d_k. We focus in this section, on a key parameter. The angle θ_k between d_k and the SD direction $-g_k$ defined by :

$$d_k^T\ g_k = -\cos\theta_k \ \|g_k\|_2\ \ \|d_k\|_2 \qquad \ldots(21)$$

Zoutendijk theorem is the main tool to analyze the convergence properties of the various descent methods.

Theorem (1) (Zoutendijk):

Consider any iteration of the from (2) where,

* d_k is descent direction
* λ_k satisfies Wolfe conditions (4) and (5)
* f is bounded below in R^n
* That the gradient g is Lipschitz Continuous in an open set N containing the level set $\delta = \{x : f(x) \le f(x_1)\}$ where x_1 is the starting point i.e. there exists a constant L such that

$$\|g(x) - g(y)\| \le L\ \|x - y\| \qquad \forall \quad x, y \in N \qquad \ldots(22)$$

(this implies that $f \in C^1$ on N)

Then,

$$\sum_{k\ge1} \cos\ \theta_k \|g_k\|^2 < \infty \qquad \ldots(23)$$

For proof see [11].

Theorem (2) (Global convergence for new coefficient conjugacy):

Consider the iteration method $x_{k+1} = x_k + \alpha_k d_k$ where d_k defined by (19) and (15) and suppose the assumption A holds. Then, the new algorithm either stops at stationary point i.e. $\|g_k\| = 0$ or $\underset{k\to\infty}{Lim\ inf}\|g_k\| = 0$

Proof:

The proof is by contradiction i.e. if theorem is not true then, $\|g_k\| \neq 0$ then, there exists a positive scalar ε such that :

$\|g_k\| \geq \varepsilon \,, \qquad \forall\ k$

since, $d_{k+1} = -g_{k+1} + \beta_k d_k$,

multiply both side the above relation by g_{k+1}^T we get

$$g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2 = \frac{-t\,\lambda_k g_{k+1}^T d_k}{-\tau_k + d_k^T g_{k+1} - d_k^T g_k + m\lambda_k \|d_k\|^2}\, g_{k+1}^T d_k$$

from strong Wolfe –conditions, we get:

$$g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2 \geq \frac{t\sigma^2 \lambda_k g_k^T d_k}{-\tau_k + d_k^T g_{k+1} - d_k^T g_k + m\lambda_k \|d_k\|^2}(-g_k^T d_k)$$

since $-g_k^T d_k = \|g_k\|^2 = c$ where $0 \leq c \leq 1$

$$g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2 \geq \frac{t c\sigma^2 \lambda_k g_k^T d_k}{-\tau_k + d_k^T g_{k+1} - d_k^T g_k + m\lambda_k \|d_k\|^2}$$

$$\frac{1}{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2} \leq \frac{-\tau_k + d_k^T g_{k+1} - d_k^T g_k + m\lambda_k \|d_k\|^2}{t c\sigma^2 \lambda_k g_k^T d_k}$$

$$\frac{1}{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2} \leq \frac{-\tau_k}{t c\sigma^2 \lambda_k g_k^T d_k} - \frac{\sigma d_k^T g_k}{t c\sigma^2 \lambda_k g_k^T d_k} - \frac{d_k^T g_k}{t c\sigma^2 \lambda_k g_k^T d_k} + \frac{m\lambda_k \|d_k\|^2}{t c\sigma^2 \lambda_k g_k^T d_k}$$

$$\frac{1}{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2} \leq \left(\frac{-\tau_k}{t c\sigma^2 \lambda_k g_k^T d_k}\right) - \left(\frac{1}{t c\sigma\lambda_k} + \frac{1}{t c\sigma^2\, \lambda_k} + \frac{m}{t c\sigma^2}\right)$$

since, $d_k = -g_k$ then:

$$\therefore \frac{1}{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2} \leq \left(\frac{-\tau_k}{t c\sigma^2 \lambda_k g_k^T d_k}\right)$$

$$g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2 \geq \frac{t c\sigma^2 \lambda_k g_k^T d_k}{-\tau_k}$$

let $\omega = \dfrac{t c\sigma^2}{\tau_k}$

$$g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2 \geq \omega\, g_k^T s_k \;,$$

square both sides of the above relation we get:

$$\left(g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2\right)^2 \geq \left(\omega\, g_k^T s_k\right)^2,$$

use the fact $(g_k^T s_k)^2 = \|s_k\|^2\ \|g_k\|^2\ \cos^2\theta_k$ and divide both sides of the above relation by $\omega^2 \|s_k\|^2$ then:

$$\frac{1}{\omega^2 \|s_k\|^2}\left(g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2\right)^2 \geq \frac{(g_k^T s_k)^2}{\|s_k\|^2} = \frac{\|g_k\|^2 \|s_k\|^2 \cos^2\theta_k}{\|s_k\|^2} = \|g_k\|^2 \cos^2\theta_k$$

and since $\|g_k\| \geq \varepsilon$ then we get:

$$\frac{1}{\omega^2 \|s_k\|^2}\left(g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2\right)^2 \geq \frac{(g_k^T s_k)^2}{\|s_k\|^2} = \|g_k\|^2 \cos^2\theta_k \geq \varepsilon^2 \cos^2\theta_k$$

taking the sum for $k \geq 1$, we get :

$$\frac{1}{\omega^2 \|s_k\|^2}\sum_{k=1}^{\infty}\left(g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2\right)^2 \geq \sum_{k=1}^{\infty}\frac{(g_k^T s_k)^2}{\|s_k\|^2} = \sum_{k=1}^{\infty}\|g_k\|^2 \cos^2\theta_k \geq \varepsilon^2 \cos^2\theta_k = \infty$$

$\therefore$ Contradiction with Zoutendijk condition therefore, $\|g_k\|^2 = 0$ or

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

4. Numerical Results:

In this paper, we have proposed a new algorithm for solving over (10) non-linear unconstrained test functions (see appendix).

These computational experiments show that the new approach given in this study is successful. We claim that the new algorithm (1.3) is better than the standard CG-algorithm namely, there is about (21.83 %) improvement in number of function evaluations (NOF) , there is about (8.32%) improvement in number of iterations (NOI), overall the calculation and for different dimension for ($100 \leq n \leq 5000$), all the algorithms in this paper use the same ILS strategy.

All the results are obtained by using (Pentium 4 computer). All programs are written in FORTRAN 90 language and for all cases the stopping criterion taken to be:

$$\|g_{k+1}\| \leq 10^{-5}$$

The comparative performance for all of these algorithms is evaluated by considering a number of function Evaluations NOF and a number of iterations NOI .

Table (1) Comparison of our new algorithm with standard H/S CG-algorithm.

Table (2) Performance percentage of our new algorithm compared with H/S CG –algorithm.

Table (1) Comparison of our new Algorithm with Standard CG – algorithm

Test fun.	Dim	HS ALGORITHEM		New algorithm	
		NOI	NOF	NOI	NOF
Powell	100	40	107	49	102
Wolfe	100	49	99	24	49
Shallow	100	10	25	10	25
Sum	100	12	63	20	70
Recip	100	5	16	2	6
Strait	100	6	14	4	9
Miele	100	34	110	15	37

Test fun.	Dim	HS ALGORITHEM		New algorithm	
		NOI	NOF	NOI	NOF
Powell3	100	14	31	16	34
Osp	100	45	143	25	63
Edger	100	6	15	5	12
Powell	500	40	107	51	106
Wolfe	500	52	105	22	45
Shallow	500	10	25	13	31
Sum	500	24	103	18	61
Recip	500	5	16	5	13
Strait	500	6	14	11	23
Miele	500	40	139	18	43
Powell3	500	14	31	16	34
Osp	500	105	305	103	290
Edger	500	6	15	5	12
Powell	1000	41	109	60	124
Wolfe	1000	70	141	22	45
Shallow	1000	10	25	17	39
Sum	1000	18	80	19	60
Recip	1000	5	16	7	17
Strait	1000	6	14	11	23
Miele	1000	47	173	21	49
Powell3	1000	14	31	17	36
Edger	1000	6	15	5	12
Osp	1000	173	480	130	376
Powell	5000	41	109	77	159
Wolfe	5000	110	229	39	79
Shallow	5000	10	25	23	52
Sum	5000	29	131	30	92
Recip	5000	6	18	11	27
Strait	5000	11	23	6	14
Miele	5000	47	173	25	58
Powell3	5000	20	42	14	31
Osp	5000	335	1016	430	1029
Edger	5000	6	15	5	12
Total		**1528**	**4348**	**1401**	**3399**

Table (2) Performance Percentage of new Algorithm Compared with Standard H/S CG-algorithm

Tools	H/S- **algorithm**	NEW algorithm
NOF	100%	78.14%
NOI	100%	91.68%

REFERENCES

[1] Al-Baali, M., (1985) ," Descent property and global convergence of the Fletcher–Reeves method with inexact line search", IMA J.Numer. Anal. 5,121–124.

[2] AL - Bayati, A.Y. and AL-Assady, N.H. (1986). “Conjugate gradient method” Technical Research report, No (1), school of computer studies, Leeds university.

[3] AL-Bayati, A.Y. and and Hassan, B.,A..,(2010) "Modified Variable Metric algorithms with exact line searches based on the quadratic models", ph .D Thesis University Mosul.

[4] Andrei, N., (2009),"open problems in nonlinear conjugate gradient algorithms for unconstrained optimization"

[5] Dai,Y. and Yuan, Y. (1999), "A Nonlinear conjugate gradient method with a strong global convergence property",SIAM J. Optim.,10. ,177– 182.

[6] Dai,Z., (2011) ," Two modified HS type conjugate gradient methods for unconstrained optimization problems", J. Nonlinear Analysis 74, 927–936.

[7] Fletcher, R. (1987) ,"Practical Methods of Optimization", John Wiley & Sons, New York

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

[9] Gilbert, J.C. and Nocedal, J. (1992). "Global convergence properties of conjugate gradient methods for optimization". SIAM J. Optim.,2(1), 21- 42.

[10] Hestenes,M.R and Stiefel, E.,(1952),"Method of conjugate gradient for solving linear equations", Journal of Research of the National Bureau of Standards , 5(49), 409-436.

[11] Kinsella. J. (2011).,"Course Note for MS4327 Optimization". http://jkcray Mths.ul.ie/ms4327/slides.pdf

[12] Liu,Y. and Storey,C.,(1991) ," Efficient generalized conjugate gradient Algorithms",part1:theory,J.Optimizat.Theor.Appl.69,129–137.

[13] Luksan L.and Vlcek J.(2005), "Shifted limited-memory VM Methods for unconstrained optimization",J.of Computational and Applied Math., 186, 365-390.

[14] Polak, E. and Ribiere,G.(1969)," Not sur la convergence de Directions conjugate". Rev. Franaisse Informants. Research operational, 3e Anne.", 16, 35-43

[15] Powell, M.J.D. (1986)."Convergence properties of algorithm for nonlinear optimization". SIAM Review., 28(4), 487-500.

[16] SHI .Z J., (2004),"Convergence of line search methods for unconstrained optimization",[J]. Applied Mathematics and Computation, 157,393-405.

[17] Touati-Ahmed, D.and Storey,C. (1990). "Efficient hybrid conjugate gradient techniques", J.Optim.Theory Appl.,64, 379- 397.

[18] Yabe,H. and Sakaiwa, N.,(2005),"A New Nonlinear Gradient Method For Unconstrained Optimization",.Journal of the Operations Research Society of Japan,Vol. 48, No. 4, 284-296

[19] Zoutendijk, G., (1970), "Nonlinear programming computational methods",in:J.Abadie (Ed.),Integer and Nonlinear Programming,North-Holland, Amsterdam,pp. 37–86.

Appendix
(The Test Functions for Unconstrained Optimization)

1. Generalized Powell Function:

$$f(x)=\sum_{i=1}^{n/4}\begin{bmatrix}(x_{4i-3}+10x_{4i-2})^2+5(x_{4i-1}-x_{4i})^2+\\(x_{4i-2}-2x_{4i-1})^4+10(x_{4i-3}-x_{4i})^4\end{bmatrix}$$

$$x_0=(3,-1,0,1,\ldots,3,-1,0,1)^T$$

2. Wolfe Function:

$$f(x)=[-x_1(3-x_1/2)+2x_2-1]^2+\sum_{i=1}^{n-1}[x_{i-1}-x_i(3-x_i/2+2x_{i+1}-1)]^2+[x_{n-1}-x_n(3-x_n/2)-1]^2$$

$$x_0=(-1,\ldots,-1)^T$$

3. Generalized Shallow Function:

$$f(x)=\sum_{i=1}^{n/2}\left[x_{2i-1}^2-x_{2i}\right]^2+(1-x_{2i-1})^2$$

$$x_0=(-2,-2,\ldots,-2,-2)^T.$$

4. Sum of Quatrics (SUM) function:

$$f(x)=\sum_{i=1}^{n}(x_i-i)^4 \quad .$$

$$x_0=(1,\ldots,1)^T$$

5. Generalized Recip Function:

$$f(x)=\sum_{i=1}^{n/3}\left[(x_{3i-1}-5)^2+x_{9i-1}^2+\frac{x_{3i}^2}{(x_{3i-1}-x_{3i-2})^2}\right],$$

$$X_0=(2,5,1,\ldots,2,5,1)^T$$

6. Generalized Strait Function:

$$f(x)=\sum_{i=1}^{n/2}(x_{2i-1}^2-x_{2i})^2+100(1-x_{2i-1})^2,$$

$$X_0=(2,-2,\ldots,2,-2)^T$$

7. Miele Function:

$$f(x)=\sum_{i=1}^{n/4}\left(\exp(x_{4i-3})+10x_{4i-2}\right)^2+100\left(x_{4i-2}+x_{4i-1}\right)^6$$
$$+\left(\tan(x_{4i-1}-x_{4i})\right)^4+\left(x_{4i-3}\right)^8+\left(x_{4i}-1\right)^2,$$

$$x_0=(1,2,2,2,\ldots,1,2,2,2)^T$$

8 .Generalized Powell 3 Function:

$$f(x)=\sum_{i=1}^{n/3}\left\{3-\left[\frac{1}{1+(x_i-x_{2i})^2}\right]-sin(\frac{\pi\, x_{2i}x_{3i}}{2})-exp\left[-(\frac{x_i+x_{3i}}{x_{2i}}-2)^2\right]\right\},$$

$$x_0=(0,1,2;\ldots;0,1,2)^T$$

9.Oren & Spedicato OSP Function:

$$f(x) = \left(\sum_{i=1}^{n} i(x_i)^2 \right)^2 ,$$

$$x_0 = [1,...,1]^T$$

10. Generalized Edger Function:

$$f(x) = \sum_{i=1}^{n/2} (x_{2i-1} - 2)^4 + (x_{2i-1} - 2)^2 x_{2i}^2 + (x_{2i} + 1)^2$$

$$x_0 = (1.,0.,...,1.,0.)^T$$