A Modified Class of Conjugate Gradient Algorithms Based on Quadratic Model for Nonlinear Unconstrained Optimization

Basim A. Hassan ***Hameed M. Sadiq***

basimabas39@gmial.com

Department of Mathematics,
College of Computer Science and Mathematics
University of Mosul, IRAQ

Received on: 7/10/2012 **Accepted on: 30/1/2013**

ABSTRACT

In this paper, we have investigated a new class of conjugate gradient algorithms for unconstrained non-linear optimization which are based on the quadratic model. Some theoretical results are investigated which are sufficient descent and ensure the local convergence of the new proposed algorithms. Numerical results show that the proposed algorithms are effective by comparing with the Polak and Ribiere algorithm.

Keywords: Conjugate Gradient , Quadratic Model

صنف مطورة من خوارزميات التدرج المترافق المعتمدة على النموذج التربيعي في الامثلية غير الخطية وغير المقيدة

باسم عباس حسن **حميد محمد صادق**

كلية علوم الحاسبات والرياضيات، جامعة الموصل

تاريخ قبول البحث: 2013/1/30 **تاريخ استلام البحث: 2012/7**

الملخص

في هذا البحث تم تطوير صنفا" جديدا" من خوارزميات التدرج المترافق في الأمثلية اللاخطية التي تعتمد على النموذج التربيعي. أعطيت بعض النتائج النظرية التي تمتاز بالانحدار الكافي وتؤكد التقارب الشامل للخوارزميات الجديد. وقد أظهرت النتائج العددية فعالية الخوارزميات الجديدة مقارنة بخوارزمية بولاك و ريبي .

الكلمات المفتاحية: التدرج المترافق المعتمدة ، النموذج التربيعي .

1. Introduction

Our problem is to minimize a function of n variable

minimize $f(x)$, $x \in R^n$ ……………….(1)

where, f is smooth and its gradient $g(x) \equiv \nabla f(x)$ is available. Conjugate gradient methods. For solving (1) are iterative methods of the form

$$x_{k+1} = x_k + \alpha_k d_k \quad \text{……………….(2)}$$

where, $\alpha_k > 0$ is a steplength and d_k is a search direction. Let g_k denotes $g(x_k)$. The search direction at the first iteration is the steepest descent direction, i.e., $d_0 = -g_0$. The consequent search direction can be defined by

$$d_{k+1} = -g_{k+1} + \beta_k d_k \quad \text{……………….(3)}$$

where, β_k is a scalar. If $f(x)$ is a strictly convex quadratic function

$$f(x) = \frac{1}{2} x^T A x \qquad \dots\dots\dots\dots\dots\dots(4)$$

where, $A \in R^{n \times n}$ is a symmetric positive definite matrix, and if α_k is the exact one-dimensional minimizer given by

$$\alpha_k = \frac{-g_k^T d_k}{d_k^T A d_k} \qquad \dots\dots\dots\dots\dots\dots(5)$$

then, the methods (2) and (3) are called the linear conjugate gradient method. Thus, several formulas of β_k where considered, which are equivalent for strictly convex quadratic objective function .Within the framework of linear conjugate gradient method, the conjugacy condition is defined by

$$d_i^T A d_j = 0, \quad i \neq j \qquad \dots\dots\dots\dots\dots\dots(6)$$

for search directions, and this condition guarantees the finite termination of linear conjugate gradient methods.

On the other hand, (2) and (3) are called the non-linear conjugate gradient method for general unconstrained optimization problem. The non-linear conjugate gradient method was first proposed by Fletcher and Reeves [3]. Within the framework of nonlinear conjugate gradient methods, the conjugacy condition is replaced by

$$d_{k+1}^T y_k = 0 \qquad \dots\dots\dots\dots\dots\dots(7)$$

where, $y_k = g_{k+1} - g_k$, because the relations

$$\begin{aligned} d_{k+1}^T A d_k &= \frac{1}{\alpha_k} d_k^T A (x_{k+1} - x_k) \\ &= \frac{1}{\alpha_k} d_{k+1}^T (g_{k+1} - g_k) \\ &= \frac{1}{\alpha_k} d_{k+1}^T y_k \end{aligned} \qquad \dots\dots\dots\dots\dots\dots(8)$$

hold for the strictly convex quadratic objective function, or the mean value theorem yields

$$d_k^T y_k = \alpha_k d_{k+1}^T \nabla^2 f(x_k + \omega \alpha_k d_k) d_k \qquad \dots\dots\dots\dots\dots\dots(9)$$

for some $\omega \in (0,1)$. Thus, condition (7) means that the search directions d_{k+1} and d_k are mutually conjugate with respect to Hessian matrix $\nabla^2 f(x)$ at some point. The extension of conjugacy condition was studied by Perry [4]. He tried to accelerate the conjugate gradient method by incorporating the second-order information into it. specifically, he used the secant condition

$$H_{k+1} y_k = s_k \qquad \dots\dots\dots\dots\dots\dots(10)$$

of quasi-Newton methods, where asymmetric matrix H_{k+1} is an approximation to the inverse Hessian. For quasi-Newton method, the search direction d_{k+1} can be calculated in the form

$$d_{k+1} = -H_{k+1} g_{k+1} \qquad \text{.....................(11)}$$

by (10) and (11), the relation

$$d_{k+1}^T y_k = -(H_{k+1} g_{k+1})^T y_k = -g_{k+1}(H_{k+1} y_k) = -g_{k+1}^T s_k \qquad \text{.....................(12)}$$

holds. By taking this relation into account, Dai and Liao replaced the conjugacy condition (7) by the condition

$$d_{k+1}^T y_k = -g_{k+1}^T s_k \qquad \text{.....................(13)}$$

Recently, Dai and Liao [6] proposed the condition

$$d_{k+1}^T y_k = -t g_{k+1}^T s_k \qquad \text{.....................(14)}$$

where $t \geq 0$ is a scalar.

Well-known formulas for β_k are the Fletcher-Reeves (FR) [5] Polak-Ribiere-Polyak (PRP) [12] and Hestenes-Stiefel (HS) [2] formulas and they are given by

$$\beta_{k+1}^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \qquad \text{.....................(15)}$$

$$\beta_{k+1}^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2} \qquad \text{.....................(16)}$$

$$\beta_{k+1}^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \qquad \text{.....................(17)}$$

The global convergence properties of the FR, PRP and HS methods without regular restarts have been studied by many researchers, including Zoutendijk [7], Al-Baali [8] and Gilbert and Nocedal [9]. The conjugate gradient method with regular restart was stated in [10]. To establish the convergence results of these methods, it is usually required that steplength α_k should satisfy the strong Wolf conditions:

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k \qquad \text{.....................(18)}$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq \sigma \left| g_k^T d_k \right| \qquad \text{.....................(19)}$$

where $0 < \delta < \sigma < 1$. Some convergence analyses even require that α_k be computed by exact line search, that is,

$$f(x_k + \alpha_k d_k) = \min_{\alpha > 0} f(x_k + \alpha d_k) \qquad \text{.....................(20)}$$

On the other hand, many other numerical methods for unconstrained optimization are proved to be convergent under the standard Wolfe conditions:

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k \qquad \text{.....................(21)}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k \qquad \text{.....................(22)}$$

for example, (see Nocedal and Wright [10]).
these line search strategies require the descent condition

$$g_k^T d_k < 0 \ , \quad \text{for all k} \qquad \text{.....................(23)}$$

however most of conjugate gradient methods don`t always generate a descent condition, so condition (23) is usually assumed in the analyses and implementations, (see Hirotaka and Hiroshi [11]).

The structure of the paper is as follows. In section (2), we present the new formulas β_k^{BH1} and β_k^{BH2}. Section (3) shows that the search direction generated by this proposed algorithms at each iteration satisfies the sufficient descent condition and descent algorithm. Section (4) establishes the global convergence property for the new class of CG-methods. Section (5) establishes some numerical results to show the effectiveness of the proposed CG-method and Section (6) gives brief conclusions and discussions.

2. New Formulae for β_k

The quadratic model is obtained from Taylor expansion of the function upto the second order terms, which can be written

$$f(x_k + s) = f(x_k) + g_k^T s + \frac{1}{2} s^T G s \qquad \text{..................(24)}$$

since $s_k = \alpha_k d_k$, then from (24) we get:

$$f(x_k + \alpha_k d_k) = f(x_k) + \alpha_k g_k^T d_k + \frac{1}{2}\alpha_k^2 d_k^T G d_k \qquad \text{..................(25)}$$

where, G is symmetric positive definite matrix.

Now, by using exact line srarch to derive α_k (exact step size), in particular for the quadratic model

$$\alpha_k = -\frac{g_k^T d_k}{d_k^T G d_k} \qquad \text{..................(26)}$$

substituting (26) into (25), we obtain :

$$f(x_k + \alpha_k d_k) - f(x_k) = -\frac{\left(g_k^T d_k\right)^2}{d_k^T G d} + \frac{1}{2}\frac{\left(g_k^T d_k\right)^2}{d_k^T G d_k} \qquad \text{..................(27)}$$

$$f(x_k + \alpha_k d_k) - f(x_k) = -\frac{\left(g_k^T d_k\right)^2}{2\, d_k^T G d_k} \qquad \text{..................(28)}$$

$$d_k^T G d_k = -\frac{\left(g_k^T d_k\right)^2}{2\left(f_{k+1} - f_k\right)} \qquad \text{..................(29)}$$

then, we get

$$G = -\frac{\left(g_k^T d_k\right)^2}{2\left(f_{k+1} - f_k\right) d_k^T d_k} I_{n\times n} \quad \text{, where } I_{n\times n} \text{ is identity matrix} \qquad \text{..................(30)}$$

therefore, from the above equation, we have

$$G^{-1} = -\frac{2\left(f_{k+1} - f_k\right) d_k^T d}{(d_k^T g_k)^2} I_{n\times n} \qquad \text{..................(31)}$$

then, the Newton direction $d_{k+1} = -G^{-1} g_{k+1}$ can be written as follows :

$$d_{k+1} = \left(\frac{2\left(f_{k+1} - f_k\right) d_k^T d_k}{(d_k^T g_k)^2}\right) g_{k+1} \qquad \text{..................(32)}$$

using the conjugacy condition (7), since Newton direction is conjugate gradient with the exact line searches we get :

$$\left(\frac{2(f_{k+1}-f_k)d_k^T d_k}{(d_k^T g_k)^2}\right)g_{k+1}^T y_k = -g_{k+1}^T y_k + \beta_k d_k^T y_k \qquad \text{…………………(33)}$$

then, we have

$$\beta_k = \left(1+\frac{2(f_{k+1}-f_k)d_k^T d}{(d_k^T g_k)^2}\right)\frac{g_{k+1}^T y_k}{d_k^T y_k} \qquad \text{…………………(34)}$$

$$d_{k+1} = -g_{k+1} + \left(1+\frac{2(f_{k+1}-f_k)d_k^T d}{\left(d_k^T g_k\right)^2}\right)\frac{g_{k+1}^T y_k}{d_k^T y_k} d_k \qquad \text{…………………(35)}$$

since, $s_k = \alpha_k d_k$ then:

$$d_{k+1} = -g_{k+1} + \left(1+\frac{2(f_{k+1}-f_k)d_k^T d_k}{(d_k^T g_k)^2}\right)\frac{g_{k+1}^T y_k}{y_k^T s_k} s_k \qquad \text{…………………(36)}$$

where, new formulae denote by β_k^{BH1} is defined by :

$$\beta_k^{BH1} = \left(1+\frac{2(f_{k+1}-f_k)d_k^T d_k}{\left(d_k^T g_k\right)^2}\right)\frac{g_{k+1}^T y_k}{d_k^T y_k} \qquad \text{…………………(37)}$$

or

$$\beta_k^{BH1} = \left(1-\frac{2(f_k-f_{k+1})d_k^T d_k}{\left(d_k^T g_k\right)^2}\right)\frac{g_{k+1}^T y_k}{d_k^T y_k} \qquad \text{…………………(38)}$$

We can therefore modify the Eq. (38) and Eq. (36) by using the idea of Dai and Laio [6] and combining the quasi-Newton direction (32) with conjugace condition (13), we get :

$$-\frac{2(f_k-f_{k+1})d_k^T d_k}{\left(d_k^T g_k\right)^2} g_{k+1}^T y_k + g_{k+1}^T s_k = 0 \qquad \text{…………………(39)}$$

and

$$d_{k+1}^T y_k = -g_{k+1}^T y_k + \beta_k d_k^T y_k = 0 \qquad \text{…………………(40)}$$

from (39) and (40) we get :

$$\beta_k = \left(1-\frac{2(f_k-f_{k+1})d_k^T d}{(d_k^T g_k)^2}\right)\frac{g_{k+1}^T y_k}{d_k^T y_k} + \frac{g_{k+1}^T s_k}{d_k^T y_k} \qquad \text{…………………(41)}$$

since, $s_k = \alpha_k d_k$ then :

$$d_{k+1} = -g_{k+1} + \left\{\left(1-\frac{2(f_k-f_{k+1})d_k^T d_k}{\left(d_k^T g_k\right)^2}\right)\frac{g_{k+1}^T y_k}{s_k^T y_k} + \frac{g_{k+1}^T s_k}{s_k^T y_k}\right\} s_k, \qquad \text{…………………(42)}$$

where new formulae denote by β_k^{BH2} is defined by :

$$\beta_k^{BH2} = \left(1-\frac{2(f_k-f_{k+1})d_k^T d_k}{\left(d_k^T g_k\right)^2}\right)\frac{g_{k+1}^T y_k}{d_k^T y_k} + \frac{g_{k+1}^T s_k}{d_k^T y_k} \qquad \text{…………………(43)}$$

it seems from (43) if exact line search is used ($s_k^T g_{k+1} = 0$) then (43) reduces to (38).

3. The Descent Property and Descent Algorithm

Below we have to show the sufficient descent property for our proposed new conjugate gradient methods, denoted by β_k^{BH1} and β_k^{BH2} . For the sufficient descent property to be hold :

$$g_{k+1}^T d_{k+1} \le -c\|g_{k+1}\|^2 \quad \text{for} \quad k \ge 0 \text{ and } c > 0 \qquad \ldots\ldots\ldots\ldots\ldots\ldots(44)$$

Assumption(1):

Assume f is bounded below in the level set $S = \{x \in R^n : f(x) \le f(x_\circ)\}$; In some neighborhood N of S, f is continuously differentiable and its gradient is Lipshitz continuos, there exist $L > 0$ such that:

$$\|g(x) - g(y)\| \le L\|x - y\| \quad \forall \mathrm{x}, \mathrm{y} \in \mathrm{N} \ . \qquad \ldots\ldots\ldots\ldots\ldots\ldots(45)$$

suppose that Assumption (1) holds and if the line search satisfies the Wolfe condition.
it follows from (22) that

$$d_k^T y_k = d_k^T(g_{k+1} - g_k) \ge (\sigma - 1)d_k^T g_k \qquad \ldots\ldots\ldots\ldots\ldots\ldots(46)$$

on the other hand, the Lipschiz condition (45) implies

$$(g_{k+1} - g_k)^T d_k \le \alpha_k L\|d_k\|^2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots(47)$$

the above two inequalities give

$$\alpha_k \ge \frac{\sigma - 1}{L} \cdot \frac{g_k^T d_k}{\|d_k\|^2} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(48)$$

which (18) implies that

$$f_k - f_{k+1} \ge c\frac{(g_k^T d_k)^2}{\|d_k\|^2} \, . \qquad \ldots\ldots\ldots\ldots\ldots\ldots(49)$$

where, $c = \delta(1-\sigma)/L$, more details can be found in **[13]**.

Theorem (3.1) :

If $\dfrac{2(f_k - f_{k+1})d_k^T d_k}{(d_k^T g_k)^2} = u \ge 1$ then the search direction (3) and β_k^{BH1} given in equation (38), with the condition (44) will hold for all $k \ge 1$.

Proof :

Since $d_0 = -g_0$, we have $g_0^T d_0 = -\|g_0\|^2$, which satisfies (44). Multiplying (35) by g_{k+1}, we have

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \left[1 - \frac{2(f_k - f_{k+1})d_k^T d_k}{(d_k^T g_k)^2}\right]\frac{g_{k+1}^T y_k}{d_k^T y_k} d_k^T g_{k+1} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(50)$$

$$= -\|g_{k+1}\|^2 + [1-u]\frac{g_{k+1}^T y_k}{d_k^T y_k} d_k^T g_{k+1} \qquad \text{…………………(51)}$$

yielding

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + (1-u)\frac{g_{k+1}^T y_k}{v_k^T y_k} v_k^T g_{k+1} \qquad \text{…………………(52)}$$

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + (1-u)\frac{g_{k+1}^T y_k}{(v_k^T y_k)^2}(v_k^T y_k) v_k^T g_{k+1} \qquad \text{…………………(53)}$$

applying the inequality $w^T v \le \frac{1}{2}(\|w\|^2 + \|v\|^2)$ to the second term of the right hand side of the above equality, with $w = (y_k^T v_k) g_{k+1}$ and $v = (g_{k+1}^T v_k) y_k$ we get :

$$d_{k+1}^T g_{k+1} \le -\|g_{k+1}\|^2 + \frac{(1-u)}{(v_k^T y_k)^2}\left(\frac{1}{2}\left[\|g_{k+1}\|^2 (y_k^T v_k)^2 + (g_{k+1}^T v_k)^2 (\|y_k\|^2)\right]\right) \qquad \text{…………………(54)}$$

$$\begin{aligned} d_{k+1}^T g_{k+1} &\le \left[\frac{(1-u)}{2} - 1\right]\|g_{k+1}\|^2 + \frac{(1-u)}{2(v_k^T y_k)^2}(g_{k+1}^T v_k)^2 \|y_k\|^2 \\ &\le \left[\frac{1}{2} - \frac{u}{2} - 1\right]\|g_{k+1}\|^2 + \frac{(1-u)}{2(v_k^T y_k)^2}(g_{k+1}^T v_k)^2 \|y_k\|^2 \end{aligned} \qquad \text{…………………(55)}$$

from (55) we get :

$$\begin{aligned} d_{k+1}^T g_{k+1} &\le \left[-\frac{1}{2} - \frac{u}{2}\right]\|g_{k+1}\|^2 + \frac{(1-u)}{2(v_k^T y_k)^2}(g_{k+1}^T v_k)^2 \|y_k\|^2 \\ &\le -\left[\frac{1}{2} + \frac{u}{2}\right]\|g_{k+1}\|^2 + \frac{(1-u)}{2(v_k^T y_k)^2}(g_{k+1}^T v_k)^2 \|y_k\|^2 \end{aligned} \qquad \text{…………………(56)}$$

$$\le -\|g_{k+1}\|^2 \left(\frac{1}{2} + \frac{u}{2}\right) + \frac{(g_{k+1}^T s_k)^2}{(s_k^T y_k)^2}\left(\frac{1}{2}(1-u)\|y_k\|^2\right) \qquad \text{…………………(57)}$$

therefore, when $\frac{1}{2} + \frac{u}{2} > 0$ and $1-u<0$, we get

$$\begin{aligned} d_{k+1}^T g_{k+1} &\le -\left(\frac{1}{2} + \frac{u}{2}\right)\|g_{k+1}\|^2 \\ &\le -c\|g_{k+1}\|^2 \end{aligned} \qquad \text{…………………(58)}$$

where ,

$$c = \frac{1}{2} + \frac{u}{2} \cdot \qquad \text{…………………(59)}$$

Theorem (3.2) :

For the line search directions defined by

$$d_{k+1}^T = -g_{k+1} + \left(1 - \frac{2(f_k - f_{k+1})d_k^T d_k}{(d_k^T g_k)^2}\right)\frac{g_{k+1}^T y_k}{d_k^T y_k} d_k + \frac{g_{k+1}^T s_k}{d_k^T y_k} d_k \qquad \dots\dots\dots\dots\dots\dots(60)$$

If $\dfrac{2(f_k - f_{k+1})d_k^T d_k}{(d_k^T g_k)^2} = t \geq 1 + \dfrac{2s_k^T y_k}{\|y_k\|^2}$ then

$$d_{k+1}^T g_{k+1} \leq -c\|g_{k+1}\|^2 \ . \qquad \dots\dots\dots\dots\dots\dots(61)$$

Proof :

The inequality (61) holds for $k = 0$, clearly. Now, we let $k \geq 1$. From the following inequality

$$w^T v \leq \frac{1}{2}\left(\|w\|^2 + \|v\|^2\right), \quad \text{and} \quad \mathrm{w, v} \in \mathrm{R}^{\mathrm{n}} \qquad \dots\dots\dots\dots\dots\dots(62)$$

it can be derived that

$$\left((g_{k+1}^T s_k)y_k\right)^T \left((y_k^T s_k)g_{k+1}\right) \leq \frac{1}{2}\left((g_{k+1}^T s_k)^2\|y_k\|^2 + (y_k^T s_k)^2\|g_{k+1}\|^2\right) \qquad \dots\dots\dots\dots\dots\dots(63)$$

so, it follows from (40) , (42) and (63) that

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \left(1 - \frac{2(f_k - f_{k+1})d_k^T d_k}{(d_k^T g_k)^2}\right)\frac{g_{k+1}^T y_k}{s_k^T y_k} s_k^T g_{k+1} + \frac{g_{k+1}^T s_k}{s_k^T y_k} s_k^T g_{k+1} \qquad \dots\dots\dots\dots(64)$$

$$d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + (1-t)\frac{1}{2(s_k^T y_k)^2}\left((g_{k+1}^T s_k)^2\|y_k\|^2 + (y_k^T s_k)^2\|g_{k+1}\|^2\right) + \frac{g_{k+1}^T s_k}{s_k^T y_k} s_k^T g_{k+1}$$

$$\leq -\|g_{k+1}\|^2 + (1-t)\frac{1}{2(s_k^T y_k)^2}\left((g_{k+1}^T s_k)^2\|y_k\|^2 + (y_k^T s_k)^2\|g_{k+1}\|^2\right) + \frac{(g_{k+1}^T s_k)^2}{(s_k^T y_k)^2} s_k^T y_k \qquad \dots\dots\dots\dots(65)$$

$$\leq -\|g_{k+1}\|^2\left(\frac{1}{2} + \frac{t}{2}\right) + \frac{(g_{k+1}^T s_k)^2}{(s_k^T y_k)^2}\left(s_k^T y_k - \frac{t\|y_k\|^2}{2} + \frac{\|y_k\|^2}{2}\right) \qquad \dots\dots\dots\dots\dots\dots(66)$$

therefore, when $\dfrac{1}{2} + \dfrac{t}{2} > 0$ and $s_k^T y_k - \dfrac{t\|y_k\|^2}{2} + \dfrac{\|y_k\|^2}{2} \leq 0$, we get

$$\begin{aligned} d_{k+1}^T g_{k+1} &\leq -\left(\frac{1}{2} + \frac{t}{2}\right)\|g_{k+1}\|^2 \\ &\leq -c\|g_{k+1}\|^2 \end{aligned} \qquad \dots\dots\dots\dots\dots\dots(67)$$

where, $c = \dfrac{1}{2} + \dfrac{t}{2} \cdot$

Now, we can obtain the new descent conjugate gradient algorithms, as follows :

The Descent Algorithm

Step 1. Initialization: Select $x_1 \in R^n$ and the parameters $0 < \delta_1 < \delta_2 < 1$.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. else step 3.

Step 3. Line search: Compute $\alpha_{k+1} > 0$ satisfying the Wolfe line search conditions (18) and (19) and update the variables $x_{k+1} = x_k + \alpha_k d_k$.

Step 4. β_k conjugate gradient parameter which is defined in (38) and (43).

Step 5. Direction computation. Compute $d_{k+1} = -g_{k+1} + \beta_k d_k$. If the restart criterion of Powell $\left|g_{k+1}^T g_k\right| \ge 0.2\,\|g_{k+1}\|^2$, 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 $k = k+1$ go to step2 .

4. Global Convergence Property :

Next, we will show that CG method with β_k^{BH1} and β_k^{BH2} converges globally. We study the convergence of suggested methods by using uniformly convex function, then there exists a constant $\mu > 0$ such that

$$(\nabla f(x) - \nabla f(y))^T (x - y) \ge \mu\|x - y\|^2 \text{ , for any } x, y \in S \qquad \ldots\ldots\ldots\ldots\ldots\ldots(68)$$

or equivalently

$$y_k^T s_k \ge \mu\|s_k\|^2 \quad \text{and} \quad \mu\|s_k\|^2 \le y_k^T s_k \le L\|s_k\|^2 \qquad \ldots\ldots\ldots\ldots\ldots\ldots(69)$$

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

$$\|x\| \le B \;, \forall x \in S \qquad \ldots\ldots\ldots\ldots\ldots\ldots(70)$$

Proposition:

Under Assumption1 and equation (70) on f, there exists a constant $\bar{\gamma} > 0$ such that

$$\|\nabla f(x)\| \le \bar{\gamma} \;,\; \forall x \in S \qquad \ldots\ldots\ldots\ldots\ldots\ldots(71)$$

Lemma(1):

Suppose that Assumption(1) and equation (70) hold. Consider any conjugate gradient method in froms (2) and (3), where d_k is a descent direction and α_k is obtained by the strong Wolfe line search. If

$$\sum_{k>1} \frac{1}{\|d_{k+1}\|^2} = \infty \qquad \ldots\ldots\ldots\ldots\ldots\ldots(72)$$

then we have

$$\lim_{k\to\infty}\left(\inf \|g_k\|\right) = 0 . \qquad \ldots\ldots\ldots\ldots\ldots\ldots(73)$$

more details can be found in [6],[11].

Theorem (4.1):

Suppose that Assumption (1) and equation (70) and the descent condition hold. Consider a conjugate gradient method in the forms $(2)-(3)$ with β_k^{BH1} as in (38), where α_k is computed from Wolf line search conditions (21) and (22). If the objective function is uniformly convex on S, then $\lim_{k\to\infty}\inf\|g_k\| = 0$.

Proof :

Firstly, we need simplify our new β_k^{BH1}, so that our convergence proof will be much easier. Subsisting (49) into (37), we obtain :

$$\beta_k^{BH1} = (1-2c)\frac{g_{k+1}^T y_k}{d_k^T y_k} \qquad \text{.................}(74)$$

Now, we get

$$\|d_{k+1}\| = \left\| -g_{k+1} + (1-2c)\frac{g_{k+1}^T y_k}{d_k^T y_k} d_k \right\| \qquad \text{.................}(75)$$

$$\begin{aligned}
\|d_{k+1}\| &= \left\| -g_{k+1} + \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k - 2c\frac{g_{k+1}^T y_k}{d_k^T y_k} d_k \right\| \\
&= \left\| -g_{k+1} + \frac{g_{k+1}^T y_k}{y_k^T s_k} s_k - 2c\frac{g_{k+1}^T y_k}{y_k^T s_k} s_k \right\| \\
&\le \|g_{k+1}\| + \frac{\|g_{k+1}\|\|L\|\|s\|^2}{\mu\|s\|^2} - 2c\frac{\|g_{k+1}\|\|L\|\|s\|^2}{\mu\|s\|^2} \\
&\le \|g_{k+1}\|\left(1 + \frac{L}{\mu} - 2c\frac{L}{\mu}\right)
\end{aligned} \qquad \text{.................}(76)$$

$$\|d_{k+1}\| \le \left(\frac{\mu + (1-2c)L}{\mu}\right)\bar{\gamma} \qquad \text{.................}(77)$$

this relation shows that

$$\sum_{k\ge1}\frac{1}{\|d_k\|^2} \ge \left(\frac{\mu}{\mu + (1-2c)L}\right)^2 \frac{1}{\bar{\gamma}^2}\sum_{k\ge1} 1 = \infty \qquad \text{.................}(78)$$

Therefore, from Lemma 1, we have $\lim_{k\to\infty}(\inf\|g_k\|) = 0$, which for uniformly convex function is equivalent to $\lim_{k\to\infty}\|g_k\| = 0$.

Theorem(4.2):

Suppose that the assumption 1 holds and consider conjugate gradient algorithm in the forms $(2)-(3)$ with β_k^{BH2} as (43), where d_k is descent direction and α_k is obtained by Wolf line search. If

$$\sum_{k>1}\frac{1}{\|d_{k+1}\|^2} \le \infty \qquad \text{.................}(79)$$

then

$$\lim_{k\to\infty}(\inf\|g_k\|) = 0. \qquad \text{.................}(80)$$

Proof :

Substiting (49) into (43), we obtain :

$$\left|\beta_k^{BH2}\right| = \left|\frac{g_{k+1}^T y_k}{d_k^T y_k} - 2c\frac{g_{k+1}^T y_k}{d_k^T y_k} + \frac{g_{k+1}^T s_k}{d_k^T y_k}\right| \quad \text{……………….(81)}$$

then

$$\left|\beta_k^{BH2}\right| \le \|g_{k+1}\|\left[\frac{L\alpha_k\|d_k\|}{\left|d_k^T y_k\right|} - 2c\frac{L\alpha_k\|d_k\|}{\left|d_k^T y_k\right|} + \frac{\alpha_k\|d_k\|}{\left|d_k^T y_k\right|}\right] \quad \text{……………….(82)}$$

From (71) and (82) we have

$$\|d_{k+1}\| \le \|g_{k+1}\| + \left|\beta_k^{BH2}\right|\|d_k\|$$

$$\le \|g_{k+1}\| + \|g_{k+1}\|\left[\frac{L\alpha_k\|d_k\|}{\left|d_k^T y_k\right|} - 2c\frac{L\alpha_k\|d_k\|}{\left|d_k^T y_k\right|} + \frac{\alpha_k\|d_k\|}{\left|d_k^T y_k\right|}\right]\|d_k\| \quad \text{……………….(83)}$$

$$= \|g_{k+1}\|\left[1 + \frac{L\alpha_k\|d\|^2}{\left|d_k^T y_k\right|} - 2c\frac{L\alpha_k\|d_k\|^2}{\left|d_k^T y_k\right|} + \frac{\alpha_k\|d_k\|^2}{\left|d_k^T y_k\right|}\right]$$

From the strong Wolfe conditions (18), (19) and sufficient descent condition, we have

$$d_k^T y_k \ge (\sigma - 1)g_k^T d_k = (1-\sigma)\|g_k\|^2 > 0 \quad \text{……..………..…(84)}$$

$$\|d_{k+1}\| \le \|g_{k+1}\|\left[1 + \frac{L\alpha_k\|d_k\|^2}{(1-\sigma)\|g_k\|^2} - 2c\frac{L\alpha_k\|d_k\|^2}{(1-\sigma)\|g_k\|^2} + \frac{\alpha_k\|d_k\|^2}{(1-\sigma)\|g_k\|^2}\right]$$

$$\le \|g_{k+1}\|\left[1 + \frac{L\alpha_k B^2}{(1-\sigma)\bar{\gamma}^2} - 2c\frac{L\alpha_k B^2}{(1-\sigma)\bar{\gamma}^2} + \frac{\alpha_k B^2}{(1-\sigma)\bar{\gamma}^2}\right] \quad \text{……………….(85)}$$

$$= \|g_{k+1}\|M \le \bar{\gamma}M \,. \quad \text{………………. (86)}$$

where, $M = 1 + \frac{L\alpha_k B^2}{(1-\sigma)\bar{\gamma}^2} - 2c\frac{L\alpha_k B^2}{(1-\sigma)\bar{\gamma}^2} + \frac{\alpha_k B^2}{(1-\sigma)\bar{\gamma}^2}$. This relation implies

$$\sum_{k\ge 1}\frac{1}{\|d_{k+1}\|^2} \ge \frac{1}{M^2\bar{\gamma}^2}\sum_{k\ge 1} 1 = \infty \quad \text{………………. (87)}$$

therefore, we have $\lim_{k\to\infty}\inf\|g_k\| = 0$.

5. Numerical Results:

In this section, we have reported some numerical results obtained with the implementation of new formals β_k^{BH1} and β_k^{BH2} on a set of unconstrained optimization test problems taken from [1]. We have selected (15) large scale unconstrained optimization problems in extended or generalized form, for each test function , we have considered numerical experiments with the number of variable $100 \le n \le 1000$. These two new algorithms are compared with Polak-Ribe`re (PR) algorithm, by using the standard Wolfe line search conditions (21) and (22) with $\delta = 0.001$ and $\sigma = 0.9$. In all these cases, the stopping criteria is the $\|g_k\| = 10^{-6}$. The programs are written in Fortran 90. The test functions are commonly used for unconstrained test problems with standard starting points and a summary of the results of these test functions were given in Table (5.1). We tabulate for comparison of these algorithms, the Total number of iterations (TNOI) and the Total number of restart (TIRS).

Table(5.1)

Test problems	β_k^{PR}		β_k^{BH1}		β_k^{BH2}	
	TNOI	TIRS	TNOI	TIRS	TNOI	TIRS
Extended Rosenbrock	745	436	379	185	392	201
Extended While & Holst	2296	2012	396	184	410	189
Extended PSC 1	176	136	94	65	93	64
Extended Maratos	3282	2631	2695	2442	2664	2419
Quadratic QF2	4480	2743	2873	865	2714	813
Arwhead	158	95	128	85	137	93
Nondia	1371	1308	128	76	132	76
Partial Perturbed Quad.	2407	815	2279	631	2193	566
Liarwhd	315	174	213	119	210	117
Denschnc	2487	2247	2135	2084	2138	2084
Denschnf	223	196	200	172	204	174
Extended Block Diagonal	1280	650	154	94	155	95
Generalized Quad. GQ1	110	65	94	57	96	59
Sincos	176	136	94	65	93	64
Generalized Quad. GQ2	730	329	412	73	413	170
Total	**20236**	**13973**	**12274**	**7197**	**12044**	**7184**

6. Conclusions and Discussions:

In this paper, we have proposed a modified class of non-linear CG- algorithms based on the Taylor expansion to second order terms defined by (38) and (43) respectively. Under some assumptions, the two new algorithms have been shown to be globlly convergent for uniformly convex, functions and satisfied the sufficient descent property. The computational experiments show that the new two kinds given in this paper are successful .

Table (5.1) gives a comparison between the new-algorithm and the Polak-Ribiere (PR) algorithm for convex optimization, this table indicates, (see Table (6.1)), that the new algorithm saves $(59.50 - 60.65)\%$ NOI and $(51.41 - 51.50)\%$ IRS, overall against the standard Polak-Ribiere (PR) algorithm, especially for our selected group of test problems.

Table(6.1): Relative Efficiency of the New Algorithm

Tools	NOI	IRS
PR Algorithm	100 %	100 %
New Algorithm with β_k^{BH1}	39.35 %	48.59 %
New Algorithm with β_k^{BH2}	40.50 %	48.50 %

REFERENCES

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

[2] M. R. Hestenes and E. Stiefel, (1952), Methods of conjugate gradients for solving linear systems, Journal of Research of National Burean of Standard,49, pp. 409-436.

[3] R. Fletcher and C. M. Reeves,(1964), Funtion minimization by conjagate gradients, Computer Journal 7, pp. 149-154.

[4] A. Perry, (1978) . A modified conjugate gradient algorithm , Operations Research 26 , pp. 1073-1078.

[5] R. Fletcher, (1989), Practical Method of Optimization (2nd Edition), John Wiley & Sons, New York .

[6] Y. H. Dai and L. Z. Liao, (2001), New conjugacy condition and related nonlinear conjugate gradient methods, Applied Mathematics and Optimization 43, pp. 87-101.

[7] G. Zoutendijk, 1970. Nonlinear Porgramming, computational methods, in: Integr and Nonlinear Programming, North-Holland, Amsterdam, pp. 37-86.

[8] M. Al-Baali, (1985), Descent property and global convergence of the Flecher-Reeves method with inexact line search, IMA Journal of Numerical Analysis 5, pp. 121-124.

[9] J. C. Gilbert and J. Nocedal, (1992), Global convergence properties of conjugate gradient methods for optimization, SIAM Journal on Optimization 2, pp. 21-42.

[10] J. Nocedal and S. J. Wright, (1999), Numerical Optimization, Springer Series in Operations Research, Springer Verlag, New York.

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

[12] E. Polak and G. Ribiere, (1969), Note sur la Convergence de Directions Conjugate, Revue Francaise Informant, Reserche. Opertionelle16, pp. 35-43.

[13] Y. H. Dai and Y. Yuan, (1999), A nonlinear conjugate gradient method with a strong global convergence property, SIAM J. optimization, pp. 177-182.