Australian Journal of Basic and Applied Sciences, 7(14) December 2013, Pages: 622-632

AENSI Journals

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

New Conjugate gradient method with Wolfe type line searches for Nonlinear Programming

[1]Ghada M. Al-Naemi and [2]Eman T. Hamed

[1]*University of Mosul, College of Computers Science and Mathematics, Department of Mathematics, Mosul-Iraq*
[2]*University of Mosul, College of Computers Science and Mathematics, Department of Operation Res. And Int. Tech., Mosul-Iraq*

ARTICLE INFO

Article history:
Received in revised form 20 December
Accepted 23 December 2013
Available online 1 February 2014

Keywords:
Unconstrained Optimization,
Conjugate Gradient,
Memoryless Self –Scale BFGS,
Global Convergence.

ABSTRACT

In this paper, based on the memoryless self-scale BFGS quasi-Newton method, we propose a new formula β_k of the conjugate gradient method. The proposed method is descent methods even with inexact line searches by using Wolfe type line searches. When exact line search is used, the proposed method reduce to the standard HS method. Convergence properties of the proposed method is discussed. Numerical results are reported.

To Cite This Article: Ghada M. Al-Naemi, Eman T. Hamed., New Conjugate gradient method with Wolfe type line searches for Nonlinear Programming. ***Aust. J. Basic & Appl. Sci.,*** *7(14): 622-632, 2013*

INTRODUCTION

In this paper, we consider the following large scale unconstrained minimization problem

Minimize f(x), $x \in R^n$ (1)

where $f : R^n \to R$ is continuously differentiable and its gradient $g(x) = \nabla f(x)$ is available. There are several kinds of numerical methods for solving (1), which include the steepest descent, the Newton method and quasi-Newton method, for example. A among them, the conjugate gradient method is one choice for solving large scale problems, because it does not need any matrices [Al-Baali, 1985].

Iterative methods are widely used for solving (1) and the iterative formula is given by:

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

where $x_k \in R^n$ is the k-th approximation to the solution of (1), α_k is a positive step size obtained by carrying out a line search and d_k is a search direction.

Due to the simplicity of its iteration and low memory requirements, the nonlinear conjugate gradient method are one of the most famous methods for solving the above unconstrained optimization problem (1), especially in case of the dimension n of f(x) is large [Chen, 2012]. The search direction are usually defined by:

$$d_{k+1} = \begin{cases} -g_{k+1} & if\ k = 0 \\ -g_{k+1} + \beta_k d_{k,} & if\ k \ge 1 \end{cases}, \tag{3}$$

where $g_{k+1} = \nabla f(x_{k+1})$ and $\beta_k \in R$ is scalar parameter which characterizes conjugate gradient methods. Usually the parameter β_k is chosen so that (2) and (3) reduces the linear conjugate gradient method if f(x) is a strictly convex quadratic function and if α_k is calculated by the exact line search. Several kinds of formulas for β_k has been proposed. For example, the Fletcher-Reeves (FR), Polak-Ribie're-Polyak (PRP), Hestens-Stiefel (HS) and Dai-Yuan (DY) formulas are well known and they are given by:

Corresponding Author: Dr. Ghada M. Al-Naemi, University of Mosul, College of Computers Science and Mathematics, Department of Mathematics, Mosul-Iraq
E-mail: ghalnaemi@yahoo.fr.

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \quad \text{[Fletcher and Reeves, 1964]} \tag{4}$$

$$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{\|g_k\|^2} \quad \text{[Polak, Ribiere and Polyak, 1969a-b]} \tag{5}$$

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \quad \text{[Hestenes and Stiefel, 1952]} \tag{6}$$

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k} \quad \text{[Dai and Yuan, 1996]} \tag{7}$$

where $y_k = g_{k+1} - g_k$ and $\|.\|$ denote the Euclidean norm.

The global convergence properties of the FR, PRP and HS methods without regular restarts have been studied by many researchers, including [Zoutendijk, 1970] and [Gilbert and Nocedal,1992]. The CG-method with regular restart was also found in [Nocedal and Wright, 1999].

To establish convergence properties of these methods, it is usually required that the step size $\alpha_k > 0$ should satisfy some conditions, one of them is the week Wolfe condition (WWC) [Wolfe, 1971], which defined by:

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k g_k^T d_k \tag{8}$$

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

Or strong Wolfe conditions (SWC), satisfying (8) and

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \le -\sigma g_k^T d_k \tag{10}$$

Where $0 < \delta < \sigma < 1$.

[Dai and Yuan, 1999] showed that the conjugate gradient method are globally convergent when they generalized , the absolute value in (10) is replaced by pair of inequalities:

$$\sigma_1 g_k^T d_k \le g(x_k + \alpha_k d_k)^T d_k \le -\sigma_2 g_k^T d_k, \tag{11}$$

where $\sigma_1 \ge 0,\ 0 \le \delta < \sigma_1 < 1,\ 0 < \sigma_2 < 1,\ \sigma_1 + \sigma_2 \le 1$.

The special case $\sigma_1 = \sigma_2 = \sigma$ corresponds to the SWC [Hager and Zhang, 2006].
Another one is the Wolfe type line search which is proposed by [More' and et al, 1981]

$$f(x_k + \alpha_k d_k) - f(x_k) \le -\delta \alpha_k^2 \|d_k\|^2 \tag{12}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge -2\sigma \alpha_k \|d_k\|^2 \tag{13}$$

where $0 < \rho < \sigma < 1$.
Line search strategies require the descent conditions

$$g_k^T d_k < 0, \quad \forall k \tag{14}$$

However, most of CG-methods don't always generate a descent search direction, so condition (14) is usually assumed in the analysis and implementation. Some strategies have been studied which produce a descent search direction within the framework of CG-methods [Gilbert and Nocedal,1992].

The structure of the paper is as follows: In section 2 we derived the new formula β_k and its algorithm. In section 3, the descent property of new algorithm is proved in this section, and also the global convergence of the new algorithm was proven. The preliminary numerical results are contained in section 4.

2. New formulas for β 's and the algorithm:

The BFGS formula can be written in the following manner

$$H_{k+1} = H_k - \left[\frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k}\right] + \frac{s_k s_k^T}{s_k^T y_k} + \frac{y_k^T H_k y_k}{s_k^T y_k}\frac{s_k s_k^T}{s_k^T y_k}. \tag{15}$$

We are going to used the self-scaling quasi-Newton method is to scale the Hessian matrix H_k. Self-scaling variable metric algorithms was introduced by Oren [Oren, 1974a-b], Oren scaled some terms of BFGS. The general strategy of our proposed method, is to scaled all the terms of BFGS, i.e. updated the matrix H_k by self-scaling BFGS of the form

$$H_{k+1} = \rho_k \{H_k - \left[\frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k}\right] + \frac{s_k s_k^T}{s_k^T y_k} + \frac{y_k^T H_k y_k}{s_k^T y_k}\frac{s_k s_k^T}{s_k^T y_k}\}, \tag{16a}$$

$$\rho_k = \frac{s_k^T y_k}{y_k^T H_k y_k}, \tag{16b}$$

where ρ_k is the self-scaling factor [Oren, 1977a-b].

When H_k is replaced by I (i.e. if $H_k \equiv I$, where I is the identity matrix) then the above self-scaling BFGS method becomes the memoryless self-scaling BFGS. So the memoryless BFGS can be written by

$$H_{k+1} = \frac{s_k^T y_k}{y_k^T y_k}\{I - \left[\frac{y_k s_k^T + s_k y_k^T}{s_k^T y_k}\right] + \frac{s_k s_k^T}{s_k^T y_k} + \frac{y_k^T y_k}{s_k^T y_k}\frac{s_k s_k^T}{s_k^T y_k}\}. \tag{17}$$

To derive the new formula multiply both sides of equation (17) by $(-g_{k+1})$, since, $d_{k+1} = -H_{k+1} g_{k+1}$ and in the conjugate gradient

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

hence

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

multiply both sides of the above equation by y_k, we will get

$$\beta_k^{new} = \frac{y_k^T g_{k+1}}{d_k^T y_k} - \alpha_k^2 \frac{d_k^T g_{k+1}}{y_k^T y_k}. \tag{18}$$

If exact line search is used the new formula of β 's reduced to the formula of HS.

Now, we are going to give the general algorithm of the conjugate gradient method with the new formula of β_k, using Wolfe type line search.

Algorithm (I):

Given $x_1 \in R^n$ and $\varepsilon > 0$, set k=1

<u>*Step (1):*</u> set $d_k = -g_k = -\nabla f(x_k)$, if $\|g_k\| < \varepsilon$ then stop.

<u>*Step (2):*</u> Find $\alpha_k > 0$ satisfying the Wolfe type line search in (12) and (13).

<u>*Step(3):*</u> Let $x_{k+1} = x_k + \alpha_k d_k$ and $g_{k+1} = g(x_{k+1})$. If $\|g_{k+1}\| < \varepsilon$, then stop; Otherwise continue.

<u>*Step(4):*</u> Compute β_k^{new} by the formula (18) and generated d_{k+1} by (3) .

<u>*Step(5):*</u> If k=n or $|g_{k+1}^T g_k| \geq 0.2\|g_{k+1}\|^2$ is satisfy, go to step (1), else k=k+1 and go to step (2).

Australian Journal of Basic and Applied Sciences, 7(14) December 2013, Pages: 622-632

3. The sufficient descent property and global convergence:

In order to establish the global convergence result for the new algorithm, we make the following basic assumptions on the objective function.

Assumption (1):

(i) The level set $\Omega = \{x \in R^n / f(x) \le f(x_1)\}$ is bounded, and f(x) is bounded in Ω.

(ii) In some neighborhood N of Ω, f(x) is continuously differentiable and its gradient is Lipschtiz continuous, namely, there exists a constant L>0 such that

$$\|g(x) - g(y)\| < L\|x - y\| \quad , \ \forall x, y \in N. \tag{19}$$

It follows directly from Assumption (1) that there exists a positive constants γ *and* A such that $\|g_k\| \le \gamma$, $\forall x \in \Omega$, and $\|x_k\| \le A$, respectively.

And g(x) also satisfies the following condition for uniformly convex function

$$(g(x) - g(y))^T (x - y) \ge m\|x - y\|^2 \ , \quad for \ \ m > 0 \ \ and \ \ \forall x, y \in N. \tag{20}$$

In order to establish the global convergence of Algorithm (I). Firstly, we are going to study the descent property of the new proposed method, but in the begging we have the following theorem:

Theorem (4.1):

Suppose that Assumption (1) holds and α_k satisfies the Wolfe type line search in (12) and (13) with β_k in (18). Then the new formula $\beta_k^{new} \ge 0$.

Proof: by second condition of Wolfe type line search (13), we have

$$\beta_k \ge \frac{g_{k+1}^T y_k}{d_k^T y_k} + 2\sigma\alpha_k^3 \frac{\|d_k\|^2}{\|y_k\|^2},$$

since $y_k = g_{k+1} - g_k$, then

$$\beta_k \ge \frac{\|g_{k+1}\|^2 - g_{k+1}^T g_k}{d_k^T y_k} + 2\sigma\alpha_k^3 \frac{\|d_k\|^2}{\|y_k\|^2}$$

but , $-g_{k+1}^T g_k \ge -\|g_{k+1}\|\|g_k\|$, therefore the above inequality become

$$\beta_k \ge \frac{\|g_{k+1}\|^2 - \|g_{k+1}\|\|g_k\|}{\|d_k\| \ \|y_k\|} + 2\sigma\alpha_k^3 \frac{\|d_k\|^2}{\|y_k\|^2}$$

$$\ge \frac{\|g_{k+1}\| \left(\|g_{k+1}\| - \|g_k\|\right)}{\|d_k\| \ \|y_k\|} + 2\sigma\alpha_k^3 \frac{\|d_k\|^2}{\|y_k\|^2},$$

we know that $\|y_k\| = \|g_{k+1} - g_k\| \ge \|g_{k+1}\| - \|g_k\|$, and $0 < \sigma < 1$ therefore

$$\beta_k \ge \frac{\|g_{k+1}\|}{\|d_k\|} + 2\sigma\alpha_k^3 \frac{\|d_k\|^2}{\|y_k\|^2},$$

$$\therefore \beta_k \ge 0, \tag{21}$$

the proof of theorem is complete.

After this, suppose that the current search direction is descent direction d_k i.e. $g_k^T d_k < 0$. Now we need to find that β_k that produces a descent direction d_{k+1}. This requires that

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \beta_k d_k^T g_{k+1} < 0. \tag{22}$$

The right hand side of equation (22) can be written as

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

$$= -\|g_{k+1}\|^2 + \beta_k d_k^T y_k + \beta_k d_k^T g_k \ .$$

By using (14) in the above inequality, yields

$$-\|g_{k+1}\|^2 + \beta_k d_k^T g_{k+1} \le -\|g_{k+1}\|^2 + \beta_k d_k^T y_k , \tag{23}$$

equation (22) can be rewritten by

$$d_{k+1}^T g_{k+1} \le -\|g_{k+1}\|^2 + \beta_k d_k^T y_k \tag{24}$$

Therefore, the non-positively of (23) is sufficient to show that condition (22) is satisfied. This reduces to the condition

$$\|g_{k+1}\|^2 \ge \beta_k d_k^T y_k . \tag{25}$$

By summarizing these relations, we have the following lemma.

Lemma (4.2)*:*

We have $\beta_k \ge 0$. If β_k satisfies inequality (24), then d_{k+1} is a descent direction for the objective function.

Theorem (4.3):

Suppose that Assumption (1) holds and α_k satisfies the Wolfe type line search in (12) and (13) with β_k in (18). Then the search direction (3) satisfies the descent condition (14).

Proof: For initial direction (k=1), we have

$d_1 = -g_1 \Rightarrow d_1^T g_1 = -\|g_1\|^2 \le 0$, which satisfies (14).

Now, we need to show that for all $k \ge 1$, condition (14) holds.

From (3), multiply the both sides by g_{k+1}, then

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \beta_k d_k^T g_{k+1} . \tag{26}$$

If exact line search are used i.e. $d_k^T g_{k+1} = 0,$ and thus (26) become

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 \le 0 .$$

If inexact line search are used, from (24) and (25), inequality (26) yields

$$d_{k+1}^T g_{k+1} \le -\|g_{k+1}\|^2 + \beta_k d_k^T y_k < 0, \ \forall k \ge 1.$$

Therefore, (14) is holds for all $k \ge 1$.

Additionally, we can prove it by multiply both sides (3) by g_{k+1}, we get

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

Therefore,

$$\begin{aligned} d_{k+1}^T g_{k+1} &\le -\|g_{k+1}\|^2 + \frac{\|y_k\|\|d_k\|}{d_k^T y_k}\|g_{k+1}\|^2 - \alpha_k^2 \frac{\|d_k\|^2}{\|y_k\|^2}\|g_{k+1}\|^2 \\ &\le -\|g_{k+1}\|^2 + \frac{\|y_k\|\|d_k\|}{d_k^T y_k}\|g_{k+1}\|^2 . \end{aligned}$$

From (19) and (20), the above inequality become

$$\begin{aligned} d_{k+1}^T g_{k+1} &\le -(1-\frac{L}{m})\|g_{k+1}\|^2 \\ &\le -(1-\frac{L}{m})\|g_{k+1}\|^2 , \end{aligned}$$

let $c = 1 - \frac{L}{m},\ with\, L < m$, so

$$d_{k+1}^T g_{k+1} \le -c\|g_{k+1}\|^2 < 0 .$$

That is (14) is holds also for all $k \ge 1$.

The following important results was obtained by [Zoutendijk, 1970] and [Dai and Liao, 2001] which are useful in showing the global convergence of our method.

***Theorem (4.4)*:**

Suppose that Assumption (1) holds, α_k is determined by Wolfe type line search (12) and (13), we have

$$\sum_{k=1}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} < +\infty \tag{27}$$

<u>Proof:</u> From (12), (13), Theorem (4.3) and Assumption (1), we obtain

$$(2\sigma + L)\alpha_k \|d_k\|^2 \ge -d_k^T g_k .$$

Then, we know that

$$(2\sigma + L)\alpha_k \|d_k\| \ge \left(-\frac{d_k^T g_k}{\|d_k\|}\right).$$

Squaring and taking the summations of the both sides for the above inequalities, we get

$$\sum_{k=1}^{\infty} \frac{\left(d_k^T g_k\right)^2}{\|d_k\|^2} \le (2\sigma + L)^2 \sum_{k=1}^{\infty} \alpha_k^2 \|d_k\|^2$$

By (12), we have

$$\sum_{k=1}^{\infty} \frac{\left(d_k^T g_k\right)^2}{\|d_k\|^2} \le \frac{(2\sigma + L)^2}{\rho} \sum_{k=1}^{\infty} \{f(x_k) - f(x_{k+1})\} < +\infty .$$

So, we get (27), and this complete the proof.

***Theorem (4.5)*:**

Consider Algorithm (I), suppose that Assumption (1) holds. Then we have

$$\underset{k\to\infty}{Lim} \inf \|g_k\| = 0 . \tag{28}$$

<u>Proof:</u> For the sake of contradiction, we suppose the conclusion is not true, then there exists a positive constant $\bar{\gamma} > 0$ such that:

$$\|g_k\| > \bar{\gamma} , \quad \forall k \ge 0 . \tag{29}$$

We get from (3) that

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

Squaring the both sides of the above equation, and rearrange it yields

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

Dividing both sides by $\left(g_{k+1}^T d_{k+1}\right)^2$, we get

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

$$= \frac{(\beta_k)^2 \|d_k\|^2}{\left(g_{k+1}^T d_{k+1}\right)^2} - \left[\frac{\|g_{k+1}\|}{g_{k+1}^T d_{k+1}} + \frac{1}{\|g_{k+1}\|} \right]^2 + \frac{1}{\|g_{k+1}\|^2}$$

$$\frac{\|d_{k+1}\|^2}{\left(g_{k+1}^T d_{k+1}\right)^2} \le \frac{(\beta_k)^2 \left(g_k^T d_k\right)^2}{\left(g_{k+1}^T d_{k+1}\right)^2} \frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} + \frac{1}{\|g_{k+1}\|^2} .$$

By (3)

$$\frac{\|d_{k+1}\|^2}{\left(g_{k+1}^T d_{k+1}\right)^2} \le \frac{(\beta_k)^2 \left(g_k^T d_k\right)^2}{\left(-\|g_{k+1}\| + \beta_k g_{k+1}^T d_k\right)^2} \frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} + \frac{1}{\|g_{k+1}\|^2} . \tag{30}$$

Since, $g_k^T d_k < 0$, therefore

$$\left(-\|g_{k+1}\|^2 + \beta_k g_{k+1}^T d_k + \beta_k g_k^T d_k\right)\left(-\|g_{k+1}\|^2 + \beta_k g_{k+1}^T d_k - \beta_k g_k^T d_k\right) \ge 0$$

Thus,

$$\frac{(\beta_k)^2 \left(g_k^T d_k\right)^2}{\left(-\|g_{k+1}\|^2 + \beta_k g_{k+1}^T d_k\right)^2} \le 1.$$

From (30) and the above inequalities, we get the following:

$$\frac{\|d_{k+1}\|^2}{\left(g_{k+1}^T d_{k+1}\right)^2} \le \frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} + \frac{1}{\|g_{k+1}\|^2} . \tag{31}$$

By noting $\frac{\|d_0\|^2}{\left(g_0^T d_0\right)^2} = \frac{1}{\|g_0\|^2}$, equation (31) yields that

$$\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} \le \sum_{i=0}^{k} \frac{1}{\|g_i\|^2} \quad \text{for all k.}$$

Therefore, put (29)in the above inequalities, we get

$$\frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} \ge \frac{\bar{\gamma}^2}{k+1},$$

$$\Rightarrow \sum_{k \ge 0} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} = +\infty .$$

This contradiction to the Zoutendijk condition (27), so the proof is complete.

To a void that the gradients cannot be bounded away from zero, then we will use the condition $\left(w_{k+1} - w_k\right)$ in the normalized direction $w_k = d_k / \|d_k\|$, [Gilbert and Nocedal, 1992].

Lemma (4.6):

Suppose that Assumption (1) hold. Let $\{x_k\}$ *and* $\{d_k\}$ be generated by Algorithm (I) and α_k is obtained by Wolfe type line search (12) and (13), and

$$\sum_{k\geq 1}\left\|w_{k+1}-w_k\right\|^2 < \infty , \tag{32}$$

where $w_k = d_k / \|d_k\|$.

Proof: Firstly, note that $d_{k+1} \neq 0$, for otherwise (14) would imply $g_k = 0$. Therefore, w_k well be defined. We have

$$d_{k+1} = -g_{k+1} + \left(\frac{y_k^T g_{k+1}}{d_k^T y_k} - \alpha_k^2 \frac{d_k^T g_{k+1}}{y_k^T y_k}\right) d_k , \tag{33}$$

we can rewrite the above equation in the following form

$$d_{k+1} = v_{k+1} + \frac{y_k^T g_{k+1}}{d_k^T y_k} d_k ,$$

where

$$v_{k+1} = -g_{k+1} - \alpha_k^2 \frac{d_k^T g_{k+1}}{y_k^T y_k} d_k , \tag{34}$$

now, we define

$$u_{k+1} = \frac{v_{k+1}}{\|d_{k+1}\|} \quad and \quad r_{k+1} = \frac{y_k^T g_{k+1}}{d_k^T y_k} \frac{\|d_k\|}{\|d_{k+1}\|} .$$

From (33), we have for $\forall k \geq 1$

$$w_{k+1} = u_{k+1} + r_{k+1} w_k . \tag{35}$$

Now, since $\|w_{k+1}\| = \|w_k\| = 1$ and from (35), we have

$$\|u_{k+1}\| = \|w_{k+1} - r_{k+1} w_k\| = \|r_{k+1} w_{k+1} - w_k\| \tag{36}$$

(the last equality can be verified by squaring both sides). Using the condition $r_{k+1} \geq 0$, the triangle inequality, and (36), we obtain

$$\begin{aligned}\|w_{k+1} - w_k\| &\leq \|(1 + r_{k+1})(w_{k+1} - w_k)\| \\ &\leq \|w_{k+1} - r_{k+1} w_k\| + \|r_{k+1} w_{k+1} - w_k\|\end{aligned}$$

so

$$\|w_{k+1} - w_k\| = 2\|u_{k+1}\| . \tag{37}$$

Now, taking the norm of both sides of equation (34), we get

$$\begin{aligned}\|v_{k+1}\| &= \left\| - g_{k+1} - \alpha_k^2 \frac{d_k^T g_{k+1}}{y_k^T y_k} d_k \right\| \\ &\leq \|g_{k+1}\| + \alpha_k^2 \frac{\|d_k\|\|g_{k+1}\|}{\|y_k\|^2} \|d_k\| \\ &\leq \left(1 + \omega^2 \frac{B^2}{\bar{\gamma}^2}\right) \gamma = M .\end{aligned}$$

From (37)

$$\|w_{k+1} - w_k\| \leq 2 \frac{M}{\|d_{k+1}\|} , \tag{38}$$

Now squaring the both sides of (38), and taking the summation we get

$$\sum_{k\geq 1}\|w_{k+1} - w_k\|^2 \leq 4M^2 \sum_{k\geq 1}\frac{1}{\|d_{k+1}\|^2} < \infty ,$$

we complete the proof.

Australian Journal of Basic and Applied Sciences, 7(14) December 2013, Pages: 622-632

5. Numerical results:

In this section, the main work is report the performance of the Algorithm (I) on a set of test functions. The cods were written in Fortran2010 and in double precision arithmetic. We selected (20) large-scale unconstrained optimization test problems. For each test function we considered two experiments with the number of variables (n=1000, and 10000). The test problems are the unconstrained problems in CUTE [Bongratz and et al, 1995] library, along with other large-scale optimization test problems in [Andrei, 2008].

In order to asses the reliability of our algorithms, we also tested these methods against the well-known routine HS, DY and PR using the same test functions. All these algorithms are implemented with Wolfe type line searches (12) and (13) with $\delta = 0.001$ *and* $\sigma = 0.9$. In the all these methods terminate when the following stopping criterion is met:

$$\|g_{k+1}\| < 1\times 10^{-5} \quad . \tag{39}$$

We also force these routines stooped if the number of iterations exceed (750) without achieving convergence.

We record the number of iterations calls (NOI), and the number of function evaluations calls (NOF), for purpose of our comparisons. We can see all these results reported in Table (1) and Table (3) with n=1000 and n=10000 respectively. While in Table (2) and Table (4) we reported the percentage with respect to NOI and NOF.

Note that,

1- The symbol F means that the algorithm is fail to converges.
2- To calculate the total NOI and NOF for Table (1), we put instead of F the greatest amount of each column

Table 1: Comparison between HS, FR, PR, and β_k^{new} w. r. t. n=1000.

Test Functions		HS Method		FR Method		PR Method		β_k^{new} Method	
		NOI	NOF	NOI	NOF	NOI	NOF	NOI	NOF
Powell		41	109	38	106	45	123	39	107
Ex. Block Diagonal		23	48	22	46	27	87	22	46
Miele		47	172	52	169	53	180	45	161
Wood		26	60	21	50	25	58	22	52
Wolfe		70	141	52	105	64	129	50	101
Sum		18	82	21	106	21	110	17	79
Powell-3		14	31	20	43	21	46	14	31
OSP		163	467	165	459	160	459	162	465
NOND		43	97	46	107	42	96	40	91
Quadratic QF1		174	352	174	352	174	352	174	352
Ex. Himmelbau		21	220	22	232	21	219	20	217
Ex. Trigeometric 2		26	53	26	53	26	53	24	49
Ex. Wood		30	67	27	61	29	67	25	53
Ex. Rosen		30	67	29	66	29	66	29	66
Ex. Beal U63		10	27	10	27	10	27	10	27
Cubic		17	46	16	44	19	51	14	39
Dixmaana		5	13	5	13	5	13	5	13
Dixmaang		44	145	45	147	46	150	43	141
Cosine		11	29	10	26	12	32	9	23
Diagonal -2		146	555	143	552	153	588	142	550
TOTAL	NOI	959		944		982		906	
	NOF	2781		2764		2993		2663	

Table 2: Percentage between HS, FR, PR, and β_k^{new} w. r. t. n=1000.

Measurement	HS Method	FR Method	PR Method	β_k^{new} Method
NOI	97.66%	96.13%	100%	92.26%
NOF	92.92%	92.34%	100%	88.97%

Table 3: Comparison between HS, FR, PR, and β_k^{new} w. r. t. n=10000.

Test Functions		HS Method	FR Method	PR Method	β_k^{new} Method
		NOI NOF	NOI NOF	NOI NOF	NOI NOF
Powell		43 114	40 115	52 147	41 113
Ex. Block Diagonal		26 54	25 52	30 94	24 49
Miele		54 192	59 208	53 180	52 177
Wood		26 60	F	25 58	24 55
Wolfe		98 200	114 232	118 238	95 194
Sum		30 107	23 102	32 161	25 103
Powell-3		F	20 43	F	15 33
OSP		557 1706	530 1892	648 2063	529 1646
NOND		61 137	78 169	87 188	59 129
Quadratic QF1		565 1133	565 1133	565 1133	565 1133
Ex. Himmelbau		8 391	8 390	8 386	8 381
Ex. Trigeometric 2		29 71	31 73	30 73	28 62
Ex. Wood		33 73	29 66	29 67	28 64
Ex. Rosen		30 67	29 66	29 66	29 66
Ex. Beal U63		11 29	11 29	11 29	11 29
Cubic		18 48	18 47	21 57	15 42
Dixmaana		5 14	5 14	5 14	5 14
Dixmaang		363 1090	356 1068	362 1088	351 1057
Cosine		11 25	9 24	11 26	9 23
Diagonal -2		F	389 1559	408 1636	381 1539
TOTAL	NOI	3098	2904	3172	2294
	NOF	8923	9174	13767	6909

Table 4: Percentage between HS, FR, PR, and β_k^{new} w. r. t. n=10000.

Measurement	HS Method	FR Method	PR Method	β_k^{new} Method
NOI	97.67%	91.55%	100%	72.32%
NOF	64.81%	66.64%	100%	50.19%

REFERENCES

Al-Baali, M., 1985. Descent property and global convergence of the Fletcher-Reeves method with inexact line search, SIAM J. Numerical Analysis, 5: 121-124.

Andrei, N., 2008. An unconstrained optimization test functions collection. Journal of Advance Modeling and Optimization, 10: 147-161.

Bongratz, I., A.R. Conn, N.I.M. Gould and P.L. Toint, 1995. CUTE: constrained and unconstrained testing environments, ACM Trans. Math. Softw., 21: 123-160.

Chen, Y., 2012. Global convergence of a new conjugate gradient method with Wolfe type line search, J. Information and Computing Science, 7: 67-71.

Dai, Y. and L.Z. Liao, 2001. New conjugacy conditions and related nonlinear conjugate gradient methods, Applied Mathematics and Optimization, 43: 87-101.

Dai, Y.H. and Y. Yuan, 1999. A nonlinear conjugate gradient method with a strong global convergence property, SIAM Journal on Optimization, 10: 177-182.

Dai, Y.H. and Y. Yuan, 1996. Convergence properties of the conjugate descent method. Adv. Math., 25: 552-562.

Fletcher, R. and C.M. Reeves, 1964. Function minimization by conjugate gradients, Computer Journal, 7: 149-154.

Hestenes, M.R. and E. Stiefel, 1952. Methods of conjugate gradients for solving linear systems, Journal of Research of the National Bureau of Standards, 5: 409-436.

Hager, W. and H. Zhang, 2006. A survey of non-linear conjugate gradient methods, Pacific J. Optimization, 2: 35-58.

Gilbert, J.C. and J. Nocedal, 1992. Global convergence Properties of conjugate gradient methods for optimization, SIAM J. Optimization, 2: 21-42.

More', J.J., B.S. Garbow and K.E. Hillstrom, 1981. "Testing unconstrained optimization software:, ACM Trans. Math. Software, 7: 17-41.

Oren, S.S., 1974a. Self-scaling variable metric (SSVM) algorithms .II. Implementation and experiments. Management Sci., Mathematical Programming, 20: 864-874.

Oren, S.S. and D.G. Luenberger, 1974b. Self-scaling variable metric (SSVM) algorithms. I. Criteria and sufficient conditions for scaling a class of algorithms. Management Sci., Mathematical Programming, 20: 845-862.

Polak, E. and G. Ribiere, 1969a. Not sur la convergence de directions conjugue'e. Rev. Franaise Informant Researche operationelle, 3e Anne'e., 16: 35-43.

Polyak, B.T., 1969b. The conjugate gradient method in extreme problems. URSS Comp. Math. Phys., 9: 94-112.

Wolfe, P., 1971. Convergence conditions for a scent methods II: some corrections. SIAM Review, 13: 185-188.

Zoutendijk, G., 1970. Nonlinear programming computational methods, In Integer and Nonlinear programming, J. Abadie (Ed.), North-Holland, Amsterdam.