A Modified Conjugate Gradient Method with Global Convergence Property for Unconstrained Optimization

Dr. Huda Issam Ahmed* Dhaam Aoweid Matrood**

Abstract:

In this paper, a modified formula for \( \beta^{DL} \) (Dai-Liao) is proposed for conjugate gradient method of solving unconstrained optimization problem. The new method has sufficient descent and global convergence properties. Numerical results show that this new method is very efficient compared with other similar methods in the same field.

1- Introduction

The conjugate gradient method presents 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 local and global convergence properties. For general unconstrained optimization problems.

\[
\text{minimize } f(x) \tag{1}
\]

Where \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) is a continuously differentiable function, bounded from below, starting. From an initial guess, a nonlinear conjugate

\* Assist Prof / Dept. operational research & intelligent techniques/ Computer Sciences and Mathematics College / Mosul University

\** Researcher / Dept. operational research & intelligent techniques/ Computer Sciences and Mathematics College / Mosul University

Received Date 8/10/2013 Accept Date 21/11/2013
A Modified Conjugate Gradient Method with Global Convergence . . . .

gradient algorithm generates a sequence of points \( (x_k) \), according to the following recurrence formula:

\[
x_{k+1} = x_k + \alpha_k d_k
\]  

(2)

Where \( \alpha_k \) is the step length, usually obtained by the Wolfe line search:

\[
f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k
\]  

(3)

\[
g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k
\]  

(4)

where \( 0 < \delta < \sigma < 1 \), which is known as weak Wolfe condition (W.W.C.) and for strong Wolfe condition (S.W.C.) is defined by:

\[
f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k
\]  

(5)

\[
|g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k
\]  

(6)

See (Nocedal and Wright, 1999).

Dai and Yuan (Dai and Yuan, 1996) showed that the conjugate gradient method are globally convergent when they generalized, the absolute value in (6) is replaced by pair of inequalities.

\[
\sigma_1 g_k^T d_k \leq g(x_k + \alpha_k d_k)^T d_k \leq -\sigma_2 g_k^T d_k
\]  

(7)

where \( 0 < \sigma_1 < 1, 0 < \sigma_2 < 1, \sigma_1 + \sigma_2 \leq 1 \)

The special case \( \sigma_1 = \sigma_2 = \sigma \) corresponds to the S.W.C (Hager and Zhan, 2006) the direction \( d_{k+1} \) are commented as:

\[
d_{k+1} = \begin{cases} 
- g_{k+1} & \text{for } k = 0 \\
- g_{k+1} + \beta_k d_k & \text{for } k \geq 1 
\end{cases}
\]  

(8)

where \( \beta_k \) is a scalar and \( g_k = \nabla f(x_k) \). Since 1952, there have been many formulas for the scalar, for example:
\[ \beta^{{\text{FR}}}_k = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \]  
(Fletcher and Reeves, 1964),  \hfill (9)

\[ \beta^{{\text{PR}}}_k = \frac{g^T_{k+1}y_k}{\|g_k\|^2} \]  
(Polak and Ribirer, 1969),  \hfill (10)

\[ \beta^{{\text{HS}}}_k = \frac{g^T_{k+1}y_k}{d^T_k y_k} \]  
(Hestenes and Stiefel, 1952),  \hfill (11)

\[ \beta^{{\text{LS}}}_k = \frac{g^T_{k+1}y_k}{-d^T_k g_k} \]  
(Liu and Story, 1991),  \hfill (12)

where \( y_k = g_{k+1} - g_k \) and \(||.||\) stands for the Euclidean norm.

The method (2) and (8) is called the linear conjugate methods, within the framework of linear conjugate gradient methods, the conjugate condition is defined by:

\[ \sum_{i=1}^{k} p_i^2 = \text{symmetric positive definite matrix}. \]

On the other hand, the method (2) and (8) is called the nonlinear conjugate gradient method for several unconstrained optimization problem. The conjugate condition is replaced by:

\[ d^T_{k+1} y_k = 0 \]  \hfill (13)

holds for strictly convex quadratic objective function. The extension of the conjugacy condition was studied by Perry (Perry, 1978), he tried to accelerate the conjugate gradient method by incorporating the second-order information into it. Specifically, he used the quasi-Newton (QN) method the search direction \( d_k \) can be calculate in the form:

\[ d_{k+1} = -H_{k+1} g_{k+1} \]  \hfill (14)

where \( H_{k+1} \) is an approximation to inverse Hessian, with quasi-Newton condition which is defined by:

\[ H_{k+1} y_k = s_k \]  \hfill (15)
A Modified Conjugate Gradient Method with Global Convergence . . .

where \( s_k = x_{k+1} - x_k = \alpha_k d_k \), by (14) and (15) we have

\[
d^T_{k+1} y_k = -(H_{k+1} g_{k+1})^T y_k
\]

\[
= -g^T_{k+1} (H_{k+1} y_k) = -g^T_{k+1} s_k
\]

Eq (16) is called Perry's condition, which implies (13) hold if line search is exact, since in this case \( g^T_{k+1} s_k = 0 \).

2. New formula for Beta and Algorithm

An idea is multiplying of \( H_{k+1} \) by scaling \( \rho_k \) before the update taking place. i.e. for every \( k \geq 1 \) the scalar Newton direction, is defined by:

\[
d_{k+1} = -\rho_k H_{k+1} g_{k+1}
\] (17)

Where \( H_{k+1} \) is an approximation to inverse Hessian, and \( \rho_k \) is scalar, this scalar is added to make the sequence and efficiency as problem dimension increase. The poor-scaling is an imbalance between the values of the function and change in \( x \). the function value may be change very little even though \( x \) is changing by good scaling factor for the updating \( H \) and the favorable in some asses especially when the number variable are large (Scales, 1985).

In this paper we use the scalar by Al-Assady (Al-Assady,1997) which defined by:

\[
\rho_k = \frac{s^T_k y_k}{2s^T_k g_k - 6(f_{k+1} - f_k)}
\] (18)

Now to drive the new methods using (8)

\[
d^T_{k+1} y_k = -g^T_{k+1} y_k + \beta_k d^T_k y_k
\]

and from (17) we get

\[
d^T_{k+1} y_k = -\rho_k (H_{k+1} g_{k+1})^T y_k
\]
\[ = -\rho_k g_k^{T} (H_{k+1} y_k) \]

Since \((H_{k+1} y_k) = s_k\) (QN- condition), then we get:

\[ d_{k+1}^T y_k = -\rho_k g_k^{T} s_k \]  

(20)

using (20) in (19) we get:

\[-\rho_k g_k^{T} s_k = -g_k^{T} y_k + \beta_k d_k^T y_k \]

\[ \beta_k = \frac{g_{k+1}^T y_k - \rho_k g_k^{T} s_k}{d_k^T y_k} \]  

(21)

Where \(\rho_k\) is defined in (18), i.e.:

\[ \beta_{k}^{new} = \frac{g_{k+1}^T y_k - s_k^T y_k}{2s_k^T g_k - 6(f_{k+1} - f_k)} \frac{g_{k+1}^T s_k}{d_k^T y_k} \]  

(22)

Observing that this new formula contains not only gradient value information, but also function value information at the present and previous step. If the function is quadratic and the line search is exact the new formula is equal to \(\beta_{k}^{HS}\). However, we consider general nonlinear function and inexact line search.:

If we compare the new version \(\beta_{k}^{new}\) with Dai and Liao (Dai-Liao, 2001) computational scheme:

\[ \beta_{k}^{DL} = \frac{s_k^T y_k}{d_k^T y_k} - t \frac{g_k^{T} s_k}{d_k^T y_k} \]  

(23)

where \(t\) is constant and \([0 < t < 1]\) in this paper we replace this parameter by the scalar \(\rho_k\), which can be viewed as adaptive of Dai-Liao computational schemes, corresponding to \(t\).

2.1 Algorithm of New Methods:
A Modified Conjugate Gradient Method with Global Convergence . . .

Step (1): Choose an initial point $x_1 \in \mathbb{R}^n$, set $k = 1$, $d_k = -g_k$.

Step (2): Compute $\alpha_k$ satisfying (S.W.C) by (5) and (6).

Step (3): Let $x_{k+1} = x_k + \alpha_k d_k$ and if $\|g_{k+1}\| \leq 1 \times 10^{-5}$ then stop, otherwise continue.

Step (4): Compute $\beta_k$ by (22) and the direction $d_{k+1}$ by (8).

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

The following assumptions are often used in the studies of the conjugate gradient methods.

Assumption (1)

i) The level set $\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$ is bounded, and $f(x)$ is bounded below in $\Omega$.

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

$$\|g(x) - g(y)\| \leq L \|x - y\|, \forall x, y \in N$$  \hspace{1cm} (24)

It follows directly from Assumption (1) that there exists two positive constants $D$ and $\gamma$ such that

$$\|x\| \leq D, \|g(x)\| \leq \gamma, \forall x \in \Omega$$  \hspace{1cm} (25)

3. Convergence Analysis of the New Method:

Since the conjugate gradient methods belong to the descent methods for solving unconstrained optimization problems, the new $\beta_k$ should be chosen such that $g_k^T d_k < 0$ if the line search is used. Furthermore, due to the sufficient descent condition
\[ g_k^T d_k \leq -c \| g_k \|^2 \]  \hspace{1cm} (26)

3.1 Theorem:

Suppose that Assumption (1) holds and \( \alpha_k \) satisfies the strong Wolfe condition (5) and (6), then the search direction (8) where \( \beta_{k+1} \) is defined by (22) is satisfy the sufficient descent condition.

proof:

For initial direction \( (k=1) \), we have

\[ d_1 = -g_1 \rightarrow g_1^T d_1 = -\| g_1 \|^2 \leq 0, \] which satisfies (26).

Now we suppose that \( d_i^T g_i \leq 0, \forall i = 1, 2, \ldots, k \)

multiplying (8) by \( g_{k+1}^T \), we get:

\[ d_{k+1}^T g_{k+1} = \| g_{k+1} \|^2 + \beta_k d_k^T g_{k+1} \]

\[ = \| g_{k+1} \|^2 + \frac{g_{k+1}^T y_k}{d_{k}^T y_k} d_{k}^T g_{k+1} \]  \hspace{1cm} (27)

Since

\[ d_{k+1}^T y_k = d_{k}^T g_{k+1} - d_{k}^T g_{k} \geq d_{k}^T g_{k+1} \]

\[ \therefore d_{k}^T g_{k+1} \leq d_{k}^T y_k \]  \hspace{1cm} (28)

Also from (6)

\[ \sigma d_{k}^T g_{k} \leq d_{k}^T g_{k+1} \leq -\sigma d_{k}^T g_{k} \]

\[ (\sigma - 1) d_{k}^T g_{k} \leq d_{k}^T y_k \leq (-\sigma - 1) d_{k}^T g_{k} \]

\[ -(1 - \sigma) d_{k}^T g_{k} \leq d_{k}^T y_k \leq -(\sigma + 1) d_{k}^T g_{k} \]

\[ \Rightarrow d_{k}^T y_k \geq -(1 - \sigma) d_{k}^T g_{k} \]

\[ \therefore S_k = \alpha_k d_k \Rightarrow d_k = \frac{S_k}{\alpha_k} \]
\[ s_k^T y_k \geq -(1 - \sigma) s_k^T g_k \]

\[ -s_k^T y_k \leq (1 - \sigma) s_k^T g_k \quad (29) \]

Substitute (28) and (29) in (27) we get:

\[
d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + \frac{g_{k+1}^T g_{k+1}}{d_k^T g_k} d_k^T y_k + \frac{(1 - \sigma) s_k^T g_k}{2s_k^T g_k - 5(f_{k+1} - f_k)} d_k^T y_k
\]

\[
= -\|g_{k+1}\|^2 + g_{k+1}^T y_k + \frac{(1 - \sigma) s_k^T g_k}{2s_k^T g_k - 6(f_{k+1} - f_k)} g_{k+1}^T s_k
\]

from (5) we get:

\[
f_{k+1} - f_k \leq \delta \alpha_k g_k^T d_k
\]

\[
\rightarrow -(f_{k+1} - f_k) \geq -\delta \alpha_k g_k^T d_k
\]

\[
\rightarrow -\frac{1}{f_{k+1} - f_k} \leq -\frac{1}{\delta \alpha_k g_k^T d_k}
\]

\[
\rightarrow \frac{1}{f_{k+1} - f_k} \leq \frac{1}{\delta s_k^T g_k}
\]

\[
d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + g_{k+1}^T y_k + \frac{(1 - \sigma) s_k^T g_k}{2s_k^T g_k - 6\delta s_k^T g_k} g_{k+1}^T s_k
\]

\[
= -\|g_{k+1}\|^2 + g_{k+1}^T y_k + \frac{(1 - \sigma) s_k^T g_k}{(2 - 6\delta) s_k^T g_k} \alpha_k d_k^T g_{k+1}
\]

\[
\leq -\|g_{k+1}\|^2 + g_{k+1}^T y_k + \frac{(1 - \sigma) \alpha_k}{(2 - 6\delta)} d_k^T g_{k+1}
\]

\[
\leq -\|g_{k+1}\|^2 + g_{k+1}^T y_k - \frac{(1 - \sigma) \alpha_k^2 d_k^T g_{k+1}}{(2 - 6\delta)}
\]

Since \( 0 < \delta < 1 \) and \( 0 < \delta \leq 0.001 \) this means \( \frac{(1 - \sigma)}{(2 - 6\delta)} > 0 \)

Since \( g_{k+1}^T y_k \leq \|g_{k+1}\| \|y_k\| \)
3.2 Global Convergence Property for Convex Functions

Where

\[ \langle d_{k+1} T g_{k+1} \rangle^2 \leq \|g_{k+1}^q\| + \|g_{k+1}^p\| \cdot \|y_k\| \leq \frac{(1 - \sigma)T a_k \sigma}{(2 - 6\delta)} \|d_k\| \cdot \|g_k\| \]

\[ \langle d_{k+1} T g_{k+1} \rangle^2 \leq -\|g_{k+1}^q\|^2 + \|g_{k+1}^p\| \cdot \|y_k\| \]

Since

\[ \|y_k\| = \|g_{k+1} - g_k\| \leq \|g_{k+1}\| \Rightarrow 0 < \frac{\|y_k\|}{\|g_{k+1}\|} < 1 \]

\[ \langle d_{k+1} T g_{k+1} \rangle^2 \leq -\left(1 - \frac{\|y_k\|}{\|g_{k+1}\|}\right) \|g_{k+1}\|^2 \]

\[ \langle d_{k+1} T g_{k+1} \rangle^2 \leq c \|g_{k+1}\|^2 \]

Where \[ c = \left(1 - \frac{\|y_k\|}{\|g_{k+1}\|}\right) > 0. \]

3.2 Global Convergence Property for Convex Functions

If \( f \) is a uniformly convex function, there exists a constant \( \mu > 0 \) such that:

\[ (g(x) - g(y)) (x - y) \geq \mu \|x - y\|^2, \forall x, y \in \Omega \]  \( (30) \)

We can rewrite (30) in the following manner:

\[ y_k^T s_k \leq \mu \|s_k\|^2 \]  \( (31) \)

Eq(31) with (24) implies that:

\[ \mu \|s_k\|^2 \leq y_k^T s_k \leq L \|s_k\|^2 \]  \( (32) \)

i.e. \[ \mu \leq L. \text{ see (Yabe and Sataiwa, 2005)} \]

Dai et al (Dai et al, 1999) proved that for any conjugate gradient method with strong Wolfe condition the followings results holds.

3.3 Lemma:
Suppose that Assumption (1) hold and consider any CG-methods (2), where $d$ is a descent direction and $\alpha_k$ is obtained by the strong Wolfe condition if

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

Then

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

### 3.3 Theorem:

Suppose that Assumption (1) hold and that $f$ is a uniformly convex function. the new algorithm of the form (2) (8) and (22) where $d_k$ satisfies the descent condition and $\alpha_k$ is obtained by the strong Wolfe condition (5) and (6) satisfies the global convergence (i.e.

$$\liminf_{k \to \infty} \|g_{k+1}\| = 0$$

**Proof:**

$$\|d_{k+1}\| = \| - \hat{g}_{k+1} + \beta_{k+1} d_k \|$$

$$\|d_{k+1}\| \leq \|\hat{g}_{k+1}\| + \|\beta_{k+1}\| \cdot \|d_k\|$$

$$\|d_{k+1}\| \leq \|\hat{g}_{k+1}\| + \left| \frac{s_k^T y_k}{d_k^T y_k} - \frac{s_k^T g_{k+1}}{2s_k^T g_k - 6(f_{k+1} - f_k)\cdot s_k^T g_k - 6(f_{k+1} - f_k)} \cdot \|d_k\| \right|$$

Since:

$$d_k^T s_{k+1} \leq d_k^T y_k \leq s_k^T y_k$$

$$-s_k^T y_k \leq -(1 - \sigma)s_k^T g_k$$

$$s_k = \alpha_k d_k \Rightarrow d_k = \frac{s_k}{\alpha_k}$$
\\[ \| d_{k+1} \| \leq \| s_{k+1} \| + \left( \frac{\| g_{k+1} \| + \| y_k \|}{\alpha_k} \right) + \frac{|1 - \sigma|s_k^Tg_k}{(2 - 6\delta)|s_k^Tg_k|}, \left| \frac{1}{s_k^T \hat{y}_k} \right|, \| d_k \| \]

\[ \leq \gamma + \left( \frac{\alpha_k \gamma L \| s_k \|}{\mu s_k^2} + c_2 \alpha_k \right) \| s_k \|, \text{ where } c_2 = \frac{|1 - \sigma|}{(2 - 6\delta)} \]

\[ \leq \gamma + \left( \frac{\gamma L}{\mu D} + c_2 \right) D \]

Since:

\[ \| s_k \| = \| x - x_k \|, D = \text{Max} \{ \| x - x_k \|, \forall x, x_k \in s \} \]

\[ \therefore \| d_{k+1} \| \leq \gamma + \left( \frac{\gamma L}{\mu D} + c_2 \right) D = D_2 \]

let: \[ \gamma + \left( \frac{\gamma L}{\mu D} + c_2 \right) D = D_2 \]

\[ \therefore \| d_{k+1} \| \leq D_2 \]

\[ \sum \frac{1}{\| d_{k+1} \|^2} \geq \frac{1}{D_2} \sum l = \infty. \]

Therefore, from Lemma 3.3 we have \( \lim_{k \to \infty} \| g_{k+1} \| = 0 \) which for uniformly convex function equivalent to \( \lim_{k \to \infty} \| g_{k+1} \| = 0 \).

### 3.4 Global Convergence for General Nonlinear Functions

For general nonlinear functions, we need to prove that the gradient of the new method cannot be bounded away from zero, we establish a bounded for the change (\( W_{k+1} - W_k \)) in the normalized direction \( w_k = d_k/\| d_k \| \). (Nocedal and Gillbert, 1992)

Also, we assume that there exists a positive constant \( \gamma > 0 \) such

\[ \| g \| \geq \gamma, \forall k \geq 0 \]  \hfill (33)

### 3.5 Lemma

[99]
Suppose that assumption (1) hold, consider the method (2), (8) and (22) where the direction satisfies the sufficient condition and \( a_k \) is obtained by the strong Wolfe condition (5) and (6), if (33) holds, then \( d_{k+1} \neq 0 \) and

\[
\sum_{k=1}^{\infty} \| w_{k+1} - w_k \|^2 < \infty
\]

(34)

Where \( w_k = d_k / \| d_k \| \)

Proof:

\[
d_{k+1} = -g_{k+1} + \left( \frac{g_{k+1}^T y_k}{d_k^T y_k} - \rho_k \frac{s_k^T g_{k+1}}{d_k^T y_k} \right) d_k
\]

We can rewrite it by

\[
d_{k+1} = v_{k+1} \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k, \quad \text{where} \quad v_{k+1} = -g_{k+1} - \rho_k \frac{s_k^T g_{k+1}}{d_k^T y_k} d_k
\]

Let \( u_{k+1} = \frac{v_{k+1}}{\|d_{k+1}\|}, \quad v_{k+1} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \|d_{k+1}\| \)

(35)

Therefore we have \( W_{k+1} = u_{k+1} + \theta_{k+1} W_k \)

Since \( \| W_k \| = \| W_{k+1} \| = 1 \), then from (36) we obtain:

\[
\| w_{k+1} - w_k \| \leq 2 \| u_{k+1} \|
\]

(37)

\[
\| u_{k+1} \| = \left\| -g_{k+1} + \left( \frac{s_k^T y_k}{2 s_k^T g_k - 6 (f_{k+1} - f_k)^T \frac{s_k^T y_k}{d_k^T y_k}} \right) d_k \right\|
\]

\[
\leq \left\| g_{k+1} + \left( \frac{1 - \sigma}{2 - 6 \delta} \right) s_k \frac{s_k^T y_k}{s_k^T g_k} \right\| \| d_k \|
\]

\[
\leq \| g_{k+1} \| + \left| \frac{1 - \sigma}{2 - 6 \delta} \right| | \alpha_k | \| d_k \|
\]

\[
\leq \| g_{k+1} \| + A | \alpha_k | | s_k |, \quad \text{where} \quad A = \left| \frac{1 - \sigma}{2 - 6 \delta} \right|
\]

\[
\leq \gamma + AD.
\]
Then from (37) and (35) we get \( \|w_{k+1} - w_k\| \leq \frac{2}{\|d_{k+1}\|} (\gamma + AD) \)

By taking the summation of the both sides and square of (37), we obtain

\[
\sum_{k=1}^{\infty} \|w_{k+1} - w_k\|^2 = 4 \sum_{k=1}^{\infty} \|u_{k+1}\|^2 
\leq 4 \sum_{k=1}^{\infty} \|v_{k+1}\|^2 
\leq 4 \sum_{k=1}^{\infty} \frac{(\gamma + AD)^2}{\|d_{k+1}\|^2} 
\leq 4(\gamma + AD)^2 \sum_{k=1}^{\infty} \frac{1}{\|d_{k+1}\|^2} < \infty
\]

i.e. (34) hold.

3.6 Lemma

Suppose that the assumption (1) hold, and consider the method (2), (8) and (22) where the direction satisfies the sufficient condition and \( \alpha_k \) is obtained by the strong Wolfe condition (5) and (6), and \( \omega \leq \alpha_k \leq \omega \), if (33) holds, then there exists the constant \( b > 1 \) and \( \lambda > 0 \), s.t. for all \( k \geq 1 \).

\[ |\beta_k| \leq \frac{1}{b} \]

and if \( \|s_k\| \leq \lambda \rightarrow |\beta_k| \leq \frac{1}{b} \quad \forall k \geq 1 \) \quad (38)

Proof :

We have from S.W.C.

\[
y_k^T s_k \geq (\sigma - 1) s_k^T g_k = -(1 - \sigma) \alpha_k d_k^T g_k
\]

\[
\geq (1 - \sigma) \alpha_k c \|g_k\|^2 \geq (1 - \sigma) c \omega \|g_k\|^2
\]

\[
|\beta_k| \leq \frac{g_{k+1}^T y_k}{d_k^T y_k} + \left| \frac{s_k^T y_k}{2s_k^T g_k - 6(f_{k+1} - f_k)} \right| s_k^T g_{k+1}
\]

Since \( s_k^T g_{k+1} \leq s_k^T y_k \)

[101]
\[ |\beta_k| \leq \frac{g_{k+1}^T y_k}{d_k^T y_k} + \frac{\| s_k^T y_k \|}{(2 - 6\delta) s_k^T g_k} \cdot \frac{1}{|\alpha_k|} |s_k^T y_k| \]

\[ \leq \frac{\| g_{k+1} \|}{|\alpha_k|} \cdot \frac{|s_k^T y_k|}{s_k^T \gamma_k} + \frac{\| s_k \|}{|\alpha_k|} \cdot \frac{|y_k|}{(6\delta - 2) s_k^T g_k} \cdot |\alpha_k| \]

\[ \leq \gamma L \frac{\| s_k \|}{s_k^T \gamma_k} \cdot |\alpha_k| + \frac{L^2 \omega}{c \| g_k \|^2} \]

\[ \leq \frac{L \gamma \omega D}{(1 - \sigma) c \omega \bar{y}^2} + \frac{L^2 \omega}{c \bar{y}^2} \quad \text{(since } \omega \leq \alpha_k \leq \omega) \]

\[ = \left( \frac{L \gamma \omega + (1 - \sigma) LD \omega}{(1 - \sigma) c \omega \bar{y}^2} \right) D \equiv b, b > 1 \quad (40) \]

Without lose of generality we can define \( b \) such that \( b > 1 \), let us define

\[ \lambda = \left( \frac{(1 - \sigma) c \omega \bar{y}^2}{L \gamma \omega + (1 - \sigma) LD \omega} \right)^2 \frac{1}{D} \quad (41) \]

If \( \| s_k \| \leq \lambda \), from (40) we have

\[ |\beta_k| \leq \left( \frac{(1 - \sigma) c \omega \bar{y}^2}{L \gamma \omega + (1 - \sigma) LD \omega} \right) \lambda \]

\[ = \left( \frac{L \gamma \omega + (1 - \sigma) LD \omega}{(1 - \sigma) c \omega \bar{y}^2} \right) \left( \frac{(1 - \sigma) c \omega \bar{y}^2}{L \gamma \omega + (1 - \sigma) LD \omega} \right)^2 \frac{1}{D} \]

\[ = \left( \frac{(1 - \sigma) c \omega \bar{y}^2}{L \gamma \omega + (1 - \sigma) LD \omega} \right)^2 \frac{1}{b} \]

The following theorem is similar to theorem (3.6) in (Dai and Liao, 2001) or to theorem (3.2) in (Hanger and Zhange, 2005), and the proof is omitted here.

**3.7 Theorem**
Suppose that Assumption (1) hold, consider the CG method (1),(8) and (22) where the direction $d_{k+1}$ satisfies the sufficient descent condition and $\alpha_k$ is obtained by strong Wolfe condition, then we have

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

4. Numerical results

We tested the HS method, DL method and our new conjugate gradient method (22). All results are obtained using Pentium 4 workstation and all programs are written in Fortran language. Our line search subroutine computes $\alpha_k$ such that the strong Wolfe condition (5)-(6) hold with $\delta = 0.001$ and $\sigma = 0.1$. The initial value of $\alpha_k$ is always compute by a cubic fitting procedure which was described in details by Bunday (Bunday, 1982) used as a line search procedure. Although our line search cannot always ensure the descent property of $d_k$ for all three methods, uphill search directions seldom occur in our numerical experiments. In the case when an uphill search direction does occur, we restart the algorithm by setting $d_k = -g_k$. For the DL method $t = 0.1$ is selected see (Dai and Liao, 2001).

We have test ten function with different dimension $n= 100, 100,10000$. The numerical results are given in the form of NOF and NOI where NOF denote the numbers of function evaluations, and NOI denote the numbers of iterations. The stopping condition is $\|g_{k+1}\| \leq 10^{-5}$.

Comparing the new method with HS method, DL method we could say that the new method is better than all especially for Powell function, Wood function, Helical function, Powell3 function, Helical function, edeger function and Resip function from the ten function test in this section as we see from the Table (4.1), (4.2), (4.3).

**Table (4.1)**
A Modified Conjugate Gradient Method with Global Convergence . . . 

Numerical comparisons of the new CG method with $n=100$

<table>
<thead>
<tr>
<th>function</th>
<th>HS method</th>
<th>DL method</th>
<th>New method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NOF</td>
<td>NOI</td>
<td>NOF</td>
</tr>
<tr>
<td>Powell</td>
<td>180</td>
<td>60</td>
<td>143</td>
</tr>
<tr>
<td>Wood</td>
<td>103</td>
<td>49</td>
<td>103</td>
</tr>
<tr>
<td>Rosen</td>
<td>68</td>
<td>26</td>
<td>66</td>
</tr>
<tr>
<td>Cubic</td>
<td>47</td>
<td>17</td>
<td>47</td>
</tr>
<tr>
<td>Powell3</td>
<td>43</td>
<td>20</td>
<td>48</td>
</tr>
<tr>
<td>Helical</td>
<td>250</td>
<td>123</td>
<td>250</td>
</tr>
<tr>
<td>Edger</td>
<td>16</td>
<td>6</td>
<td>16</td>
</tr>
<tr>
<td>Recip</td>
<td>31</td>
<td>11</td>
<td>31</td>
</tr>
<tr>
<td>Shallow</td>
<td>17</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>Beal</td>
<td>18</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>773</td>
<td>326</td>
<td>739</td>
</tr>
</tbody>
</table>

**Table (4.2)**

Numerical comparisons of the new CG method with $n=1000$

<table>
<thead>
<tr>
<th>Function</th>
<th>HS method</th>
<th>DL method</th>
<th>New method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NOF</td>
<td>NOI</td>
<td>NOF</td>
</tr>
<tr>
<td>Powell</td>
<td>219</td>
<td>66</td>
<td>143</td>
</tr>
<tr>
<td>Wood</td>
<td>103</td>
<td>49</td>
<td>103</td>
</tr>
<tr>
<td>Rosen</td>
<td>68</td>
<td>26</td>
<td>69</td>
</tr>
</tbody>
</table>
The 6th Scientific Conference of the College of Computer Sciences & Mathematics

<table>
<thead>
<tr>
<th>Function</th>
<th>NOF</th>
<th>NOI</th>
<th>NOF</th>
<th>NOI</th>
<th>NOF</th>
<th>NOI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cubic</td>
<td>47</td>
<td>17</td>
<td>47</td>
<td>17</td>
<td>59</td>
<td>19</td>
</tr>
<tr>
<td>Powell3</td>
<td>49</td>
<td>23</td>
<td>52</td>
<td>25</td>
<td>35</td>
<td>14</td>
</tr>
<tr>
<td>Helical</td>
<td>270</td>
<td>133</td>
<td>272</td>
<td>134</td>
<td>82</td>
<td>33</td>
</tr>
<tr>
<td>Edger</td>
<td>18</td>
<td>7</td>
<td>18</td>
<td>7</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>Recip</td>
<td>33</td>
<td>12</td>
<td>33</td>
<td>12</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>Shallow</td>
<td>17</td>
<td>6</td>
<td>17</td>
<td>6</td>
<td>26</td>
<td>10</td>
</tr>
<tr>
<td>Beal</td>
<td>18</td>
<td>8</td>
<td>18</td>
<td>8</td>
<td>28</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>842</td>
<td>347</td>
<td>772</td>
<td>333</td>
<td>557</td>
<td>195</td>
</tr>
</tbody>
</table>

Table (4.3)

Numerical comparisons of the new CG method with n=10000

<table>
<thead>
<tr>
<th>Function</th>
<th>HS method</th>
<th>DL method</th>
<th>New method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NOF</td>
<td>NOI</td>
<td>NOF</td>
</tr>
<tr>
<td>Powell</td>
<td>253</td>
<td>72</td>
<td>178</td>
</tr>
<tr>
<td>Wood</td>
<td>105</td>
<td>50</td>
<td>105</td>
</tr>
<tr>
<td>Rosen</td>
<td>68</td>
<td>26</td>
<td>69</td>
</tr>
<tr>
<td>Cubic</td>
<td>47</td>
<td>17</td>
<td>47</td>
</tr>
<tr>
<td>Powell3</td>
<td>51</td>
<td>24</td>
<td>52</td>
</tr>
<tr>
<td>Helical</td>
<td>249</td>
<td>145</td>
<td>294</td>
</tr>
<tr>
<td>Edger</td>
<td>18</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>Recip</td>
<td>33</td>
<td>12</td>
<td>33</td>
</tr>
</tbody>
</table>
A Modified Conjugate Gradient Method with Global Convergence . . .

<table>
<thead>
<tr>
<th></th>
<th>20</th>
<th>7</th>
<th>17</th>
<th>6</th>
<th>26</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shallow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beal</td>
<td>18</td>
<td>8</td>
<td>18</td>
<td>8</td>
<td>28</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>862</td>
<td>368</td>
<td>831</td>
<td>353</td>
<td>603</td>
<td>201</td>
</tr>
</tbody>
</table>

Reference


