Iraqi Journal of Statistical Sciences (25) 2013
The 6th Scientific Conference of the College of Computer Sciences & Mathematics
pp [89-107]

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 β^{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 filed.

تحسين طريقة التدرج المترافق ذات خاصية التقارب الشامل في الامثلية غير المقيدة

الخلاصة:

تم في هذا البحث تحسين طريقة التدرج المترافق بالاعتماد على المعلمة β^{DL} للعالمين (Dai-Liao) لحل مسائل الامثلية غير المقيدة. الطريقة الجديدة تمتلك خاصتي الانحدار الشديد والتقارب الشامل. النتائج العملية اثبتت بان الطريقة الجديدة اكثر كفاءةً عند مقارنتها مع الطرق المشابهة لها في هذا الحقل.

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.

$$\text{minimize } f(x) \qquad (1)$$

Where f: $R^n \rightarrow R$ 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

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

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

Where α_k is the step length, usually obtained by the Wolfe line search:

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta\, \alpha_k g_k^T d_k$$

(3) $g(x_k + \alpha_k d_k)^T d_k \ge \sigma\; g_k^T d_k$

(4)

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

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k g_k^T d_k \quad (5)$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \le -\sigma\; g_k^T d_k \quad (6)$$

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.

$$\sigma_1\; g_k^T d_k \le g(x_k + \alpha_k d_k)^T d_k \le -\sigma_2\; g_k^T d_k \quad (7)$$

where $0 < \sigma_1 < 1,\; 0 < \sigma_2 < 1,\; \sigma_1 + \sigma_2 \le 1$

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:

$$d_{k+1} = \begin{cases} -g_{k+1} & \text{for } k = 0 \\ -g_{k+1} + \beta_k d_k & \text{for } k \ge 1 \end{cases} \quad (8)$$

where β_k is a scalar and $g_k = \nabla f(x_k)$, since 1952, there have been many formulas for the scalar, for example:

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \quad \text{(Fletcher and Reeves, 1964)}, \tag{9}$$

$$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{\|g_k\|^2} \quad \text{(Polak and Ribirer, 1969)}, \tag{10}$$

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \quad \text{(Hestenes and Stiefel, 1952)}, \tag{11}$$

$$\beta_k^{LS} = \frac{g_{k+1}^T y_k}{-d_k^T g_k} \quad \text{(Liu and Story, 1991)}, \tag{12}$$

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: $d_i^T G d_j = 0 \;\; , i \neq j$, where $G \in R^{n\times n}$ is 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:

$$d_{k+1}^T y_k = 0 \tag{13}$$

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 calceolate in the form:

$$d_{k+1} = -H_{k+1} g_{k+1} \tag{14}$$

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

$$H_{k+1} y_k = s_k \tag{15}$$

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

$$d_{k+1}^T y_k = -(H_{k+1} g_{k+1})^T y_k$$

$$= -g_{k+1}^T (H_{k+1} y_k) \qquad = -g_{k+1}^T s_k \qquad (16)$$

Eq (16) is called Perry's condition, which implies (13) hold if line search is exact, since in this case $g_{k+1}^T s_k = 0$.

2. New formula for Beta and Algorithm

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

$$d_{k+1} = -\rho_k H_{k+1} g_{k+1} \qquad (17)$$

Where H_{k+1} is an approximation to inverse Hessian, and ρ_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:

$$\rho_k = \frac{s_k^T y_k}{2 s_k^T g_k - 6(f_{k+1} - f_k)} \qquad (18)$$

Now to drive the new methods using (8)

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

and from (17) we get

$$d_{k+1}^T y_k = -\rho_k (H_{k+1} g_{k+1})^T y_k$$

$$= -\rho_k g_{k+1}^T (H_{k+1}\, y_k)$$

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

$$d_{k+1}^T y_k = -\rho_k g_{k+1}^T s_k \qquad (20)$$

using (20) in (19) we get:

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

$$\beta_k = \frac{g_{k+1}^T y_k - \rho_k g_{k+1}^T s_k}{d_k^T y_k} \qquad (21)$$

Where ρ_k is defined in (18), i.e.:

$$\beta_k^{new} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - \frac{s_k^T y_k}{2 s_k^T g_k - 6(f_{k+1} - f_k)} \frac{g_{k+1}^T s_k}{d_k^T y_k} \qquad (22)$$

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 β_k^{HS}. However, we consider general nonlinear function and inexact line search.:

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

$$\beta_k^{DL} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - t \frac{g_{k+1}^T s_k}{d_k^T y_k} \qquad (23)$$

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

2.1 Algorithm of New Methods:

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

Step (2): Compute α_k satisfying (S.W.C) by (5) and (6).

Step (3): Let $x_{k+1} = x_k + \alpha_k d_k$ and if $\|g_{k+1}\| \le 1x10^{-5}$ then stop, otherwise continue.

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

Step (5): if k = n or $|g_{k+1}^T g_k| \ge 0.2 \, \|g_{k+1}\|^2$ 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 $\Omega = \{x \in R^n, f(x) \le f(x_1)\}$ is bounded, and f(x) is bounded below in Ω.

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

$$\|g(x) - g(y)\| \le L \, \|x - y\| \, , \forall \, x,y \in N \tag{24}$$

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

$$\|x\| \le D, \; \|g(x)\| \le \gamma, \forall \, x \in \Omega \tag{25}$$

3. Convergence Analysis of the New Method:

Since the conjugate gradient methods belong to the descent methods for solving unconstrained optimization problems, the new β_k should be chosen such that $g_k^T d_k < 0$ if the line search is used. Furthermore, due to the sufficient descant condition

$$g_k^T d_k \leq -c\,\|g_k\|^2 \tag{26}$$

3.1 Theorem:

Suppose that Assumption (1) holds and α_k satisfies the strong Wolfe condition (5) and (6),then the search direction (8) where β_{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:

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

$$= -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k^T g_{k+1} - \frac{s_k^T y_k}{2s_k^T g_k - 6(f_{k+1} - f_k)} \cdot \frac{g_{k+1}^T s_k}{d_k^T y_k} d_k^T g_{k+1} \tag{27}$$

Since

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

$$\therefore\ d_k^T g_{k+1} \leq d_k^T y_k \tag{28}$$

Also from (6)

$$\sigma d_k^T g_k \leq d_k^T g_{k+1} \leq -\sigma d_k^T g_k$$

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

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

$$\Rightarrow d_k^T y_k \geq -(1 - \sigma) d_k^T g_k$$

$$\because\ s_k = \alpha_k d_k \Rightarrow d_k = \frac{s_k}{\alpha_k}$$

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

$$-s_k^T y_k \leq (1-\sigma) s_k^T g_k \tag{29}$$

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

$$d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k^T y_k + \frac{(1-\sigma) s_k^T g_k}{2 s_k^T g_k - 6(f_{k+1} - f_k)} \frac{g_{k+1}^T s_k}{d_k^T y_k} d_k^T y_k$$

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

from (5) we get:

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

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

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

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

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

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

$$= -\|g_{k+1}\|^2 + g_{k+1}^T y_k + \frac{(1-\sigma)}{(2-6\delta)} . \alpha_k d_k^T g_{k+1}$$

$$\therefore d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + g_{k+1}^T y_k - \frac{(1-\sigma) \alpha_k \sigma\, d_k^T g_k}{(2-6\delta)}$$

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

Since $g_{k+1}^T y_k \leq \|g_{k+1}\| . \|y_k\|$

$$\therefore d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + \|g_{k+1}\|.\|y_k\| - \frac{(1-\sigma)\alpha_k\sigma}{(2-6\delta)}\|d_k\|.\|g_k\|$$

$$d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + \|g_{k+1}\|.\|y_k\|$$

Since

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

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

$$\therefore d_{k+1}^T g_{k+1} \leq -c\|g_{k+1}\|^2$$

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 μ>0 such that:

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

We can rewrite (30) in the following manner:

$$y_k^T s_k \leq \mu\,\|s_k\|^2 \qquad (31)$$

Eq(31) with (24)implies that:

$$\mu\,\|s_k\|^2 \leq y_k^T s_k \leq L\,\|s_k\|^2 \qquad (32)$$

i.e. $\mu \leq L$ 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 α_k is obtained by the strong Wolfe condition if

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

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 α_k is obtained by the strong Wolfe condition (5) and (6) satisfies the global convergence (i.e.

$$\operatorname{Lim}_{k\to\infty} \inf \|g_{k+1}\| = 0$$

Proof :

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

$$\|d_{k+1}\| \leq \|g_{k+1}\| + |\beta_{k+1}| . \|d_k\|$$

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

Since:-

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

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

$$\rightarrow s_k = \alpha_k d_k \rightarrow d_k = \frac{s_k}{\alpha_k}$$

$$\|d_{k+1}\| \le \|g_{k+1}\| + \left(\frac{\|g_{k+1}\|.\|y_k\|}{\frac{1}{\alpha_k}|s_k^T y_k|} + \frac{|1-\sigma||s_k^T g_k|}{|(2-6\delta)||s_k^T g_k|} \cdot \frac{|s_k^T y_k|}{\left|\frac{1}{\alpha_k} s_k^T y_k\right|}\right).\|d_k\|$$

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

$$\le \gamma + \left(\frac{\gamma L}{\mu\|S_k\|} + c_2\right)\|s_k\|$$

Since:- $\|s_k\| = \|x - x_k\|, D = \text{Max}\{\|x - x_k\|, \forall\, x, x_k \in s\}$

$$\therefore \|d_{k+1}\| \le \gamma + \left(\frac{\gamma L}{\mu D} + c_2\right) D$$

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

$$\therefore \|d_{k+1}\| \le D_2$$

$$\sum \frac{1}{\|d_{k+1}\|^2} \ge \frac{1}{D_2}\sum_{k\ge1} 1 = \infty \,.$$

Therefore, from Lemma 3.3 we have $\lim\inf_{k\to\infty} \|g_{k+1}\| = 0$ 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 $\bar{\gamma} > 0$ such

$$\|g\| \ge \bar{\gamma}, \forall_k \ge 0 \qquad (33)$$

3.5 Lemma

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

$$\sum_{k\geq 1} \|w_{k+1} - w_k\|^2 < \infty \tag{34}$$

Where $w_k = d_k/\|d_k\|$

Proof:

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

We can rewrite it by

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

Let $u_{k+1} = \frac{v_{k+1}}{\|d_{k+1}\|}, \quad \vartheta_{k+1} = \frac{g_{k+1}y_k}{d_k^T y_k} . \frac{\|d_k\|}{\|d_{k+1}\|}$ (35)

Therefore we have $w_{k+1} = u_{k+1} + \vartheta_{k+1} w_k$ (36)

Since $\|w_k\| = \|w_{k+1}\| = \mathbf{1}$, then from (36) we obtain:

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

$$\|v_{k+1}\| = \left\| -g_{k+1} - \left(\frac{s_k^T y_k}{2S_k^T g_k - 6(f_{k+1} - f_k)} . \frac{s_k^T g_{k+1}}{d_k^T y_k}\right) d_k \right\|$$

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

$$\leq \|g_{k+1}\| + \left|\frac{(1-\sigma)}{(2-6\delta)}\right| . |\alpha_k| \|d_k\|$$

$$\leq \|g_{k+1}\| + A|\alpha_k| . \frac{\|s_k\|}{|\alpha_k|}, \text{ where } A = \left|\frac{(1-\sigma)}{(2-6\delta)}\right|$$

$$\leq \gamma + AD.$$

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

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

$$\sum_{k\ge1}\|w_{k+1}-w_k\|^2 = \sum 4\|u_{k+1}\|^2 = 4\sum_{k\ge1}\frac{\|v_{k+1}\|^2}{\|d_{k+1}\|^2} \le 4\sum\frac{(\gamma+AD)^2}{\|d_{k+1}\|^2}$$
$$\le 4(\gamma+AD)^2\sum_{k\ge1}\frac{1}{\|d_{k+1}\|^2} < \infty$$

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 α_k is obtained by the strong Wolfe condition (5) and (6), and $\omega \le \alpha_k \le \omega$, if (33) holds, then there exists the constant $b > 1$ and $\lambda > 0$, s.t. for all $k \ge 1$.

$|\beta_k| \le b$

and if $\|s_k\| \le \lambda \to |\beta_k| \le \frac{1}{b}, \quad \forall k \ge 1$ (38)

Proof :

We have from S.W.C.

$$y_k^T s_k \ge (\sigma-1)s_k^T g_k = -(1-\sigma)\alpha_k d_k^T g_k$$
$$\ge (1-\sigma)\alpha_k c\|g_k\|^2 \ge (1-\sigma)c\omega\|g_k\|^2 \quad (39)$$

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

Since $s_k^T g_{k+1} \le s_k^T y_k$

$$\therefore |\beta_k| \le \left|\frac{g_{k+1}y_k}{d_k^T y_k}\right| + \left|\frac{s_k^T y_k}{(2-6\delta)s_k^T g_k}\right| . \frac{|s_k^T y_k|}{\frac{1}{|\alpha_k|}|s_k^T y_k|}$$

$$\le \frac{\|g_{k+1}\|.\|y_k\|}{\frac{1}{|\alpha_k|}|s_k^T y_k|} + \frac{\|s_k\|.\|y_k\|}{|-(6\delta-2)s_k^T g_k|}.|\alpha_k|$$

$$\le \frac{\gamma L\|s_k\||\alpha_k|}{|s_k^T y_k|} + \frac{L\|s_k\|^2|\alpha_k|}{c\|g_k\|^2}$$

$$\le \frac{L\gamma\omega\, D}{(1-\sigma)c\omega\bar{\gamma}^2} + \frac{LD^2\omega}{c\bar{\gamma}^2} \quad \text{(since } \omega \le \alpha_k \le \omega)$$

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

Without lose of generality we can define b such that $b > 1$, let us define

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

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

$$|\beta_k| \le \left(\frac{(1-\sigma)c\omega\bar{\gamma}^2}{L\gamma\omega + (1-\sigma)LD\omega}\right)\lambda$$

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

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

.

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 hero.

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 α_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 α_k such that the strong Wolfe condition (5)-(6) hold with $\delta = 0.001$ and $\sigma = 0.1$. The initial value of α_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 selectedsee (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 $\|g_{k+1}\| \le 10^{-5}$

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)

Numerical comparisons of the new CG method with n=100

	HS method		DL method		New method	
function	NOF	NOI	NOF	NOI	NOF	NOI
Powell	180	60	143	49	123	40
Wood	103	49	103	49	71	25
Rosen	68	26	66	25	79	29
Cubic	47	17	47	17	59	19
Powell3	43	20	48	23	35	14
Helical	250	123	250	123	82	33
Edger	16	6	16	6	15	6
Recip	31	11	31	11	16	5
Shallow	17	6	17	6	26	10
Beal	18	8	18	8	28	11
Tolal	773	326	739	317	534	192

Table (4.2)

Numerical comparisons of the new CG method with n=1000

	HS method		DL method		New method	
Function	NOF	NOI	NOF	NOI	NOF	NOI
Powell	219	66	143	49	140	41
Wood	103	49	103	49	77	27
Rosen	68	26	69	26	79	29

Cubic	47	17	47	17	59	19
Powell3	49	23	52	25	35	14
Helical	270	133	272	134	82	33
Edger	18	7	18	7	15	6
Recip	33	12	33	12	16	5
Shallow	17	6	17	6	26	10
Beal	18	8	18	8	28	11
Tolal	842	347	772	333	557	195

Table (4.3)

Numerical comparisons of the new CG method with n=10000

	HS method		**DL method**		**New method**	
Function	NOF	NOI	NOF	NOI	NOF	NOI
Powell	253	72	178	57	186	47
Wood	105	50	105	50	77	27
Rosen	68	26	69	26	79	29
Cubic	47	17	47	17	59	19
Powell3	51	24	52	25	35	14
Helical	249	145	294	145	82	33
Edger	18	7	18	7	15	6
Recip	33	12	33	12	16	5

Shallow	20	7	17	6	26	10
Beal	18	8	18	8	28	11
Tolal	862	368	831	353	603	201

Reference

1-Al-Assady, N. H. and huda, K. M., (1997),"A Rational logarithmic Science Model for Unconstrained Non-linear Optimization:, Rafideen Science J. ,Vol.8,pp.107-117.

2-Bunday, B. (1984), "Basic Optimization Methods", Edward Arnold Bedford Square, London, U.K.

3-Dai, Y. and Liao, L. Z. (2001), "Now conjugacy conditions and related nonlinear conjugate gradient methods", Applied Mathematics and Optimization, Vol. 43, PP. 87-101.

4-Dai, Y. and Yuan, Y.(1996),"convergence properties of the conjugate descent method", Adv. Math., Vol. 25, pp.552-562

5-Dai, H. Y., Han. Y. J., Lin, D. F., Sun, X. Yaan, Y., (1999), "Convergence properties of nonlinear conjugate gradient methods", SIAM J. on Optimization, PP. 177-182.

6-Fletcher R. and Reeves, C. M., (1964), "Function minimization by conjugate gradient", Computer Journal, Vol. 7, PP. 149-154.

7-Hager, W. and Zhang, H., (2005),"A new conjugate gradient method with guaranteed descent and an efficient line search", SIAM. J. Optimization,Vol. 16, pp. 170-192.

8-Hager, W. and Zhang, H., (2006),"A survey of nonlinear conjugate gradient methods", Pacific J. Optimization,Vol. 2,pp. 35-58.

9-Hestenes, M. R. and Stiefd, F., (1952), "Methods of conjugate gradients for solving linear systems", Journal of Research of the National Burean of standards, Vol. 5., No. 49, PP. 409-436.

10-Liu, D. and Story, C., (1991), "Efficient generalized conjugate gradient algorithms", Part 1: Theory, J. Optimization theory and applications, Vol. 69, PP. 129-137.

11-Nocedal, J. and Glibart, J., (1992), "Global convergence properties of conjugate gradient methods for optimizations, SIAM J. Optimization, Vol. 2, PP. 21-42.

12-Nocedal, J. and Wright, J. S. (1999), " Numerical Optimization", Springer Series in Operations Research, Springer-Verlag, New Yourk.

13-Perry, A. (1978), "A modified conjugate gradient algorithm", Operations Research, Vol. 26, pp. 1073-1078.

14-Polyak E. and Ribiere, G.,(1969),"Not sur La convergence de directions conjugue'e.", Rev. Franaise Informant Researche Operationnelle, 3e Anne'e., Vol.16, pp. 35-43.

15-Scales, L. E. (1985), "Introduction to Non-linear optimization", Macmillan, London.

16-Yabe, H. and Sataiwa, N., (2005), "A new nonlinear conjugate gradient method for unconstrained optimization", J. of Operation research society of Japan, Vol. 48, PP. 284-296.