Iraqi Journal of Statistical Science (20) 2011
The Fourth Scientific Conference of the College of Computer Science & Mathematics
pp [222-236]

A descent modified Hager-Zhang conjugate gradient method and its global convergence

Dr. Huda I. Ahmed* **Ghada M. Al-Naemi****

Abstract

In this paper, based on the memoryless BFGS quasi-Newton method, we propose a new modified Hager-Zhang (HZ) type method. An attractive property of the proposed method is that the direction generated by the method is always a descent direction for the objective function. Moreover, if the exact line search is used, the new method reduces to the ordinary HS method. Under appropriate conditions, we show that the modified HZ method is globally convergent for convex and general functions. Numerical results are also reported.

تحسين طريقة التدرج المترافق المعتمدة على (HZ) وتقاربها الشمولي

الملخص

اعتمد هذا البحث على خاصية تقليص الذاكرة لطريقة BFGS، اقترحت طريقة جديدة لتحسين طريقة (HZ) Hager–Zhang. الطريقة تحتفظ بخاصية المتجه السلبي بالنسبة إلى دالة الهدف فضلا عن استخدام خط البحث المضبوط فأن الطريقة الجديدة تعود إلى طريقة (HS) Hesttenes–Stiefel القياسية، وباستخدام بعض الشروط تم إثبات أن الطريقة الجديدة تحقق شرط التقارب الشمولي لكل الدوال المحدبة والعامة. النتائج العملية أثبتت كفاءة الطريقة المقترحة.

1- Introduction.

We consider the unconstrained nonlinear optimization problem:

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

where $f : R^n \to R$ is smooth, nonlinear function whose gradient will be denoted by g. nonlinear Conjugate Gradient (CG) method is one of the

*Assistant Prof. \ College of Computers Sciences and Math.\ University of Mosul.
**Lecture\ College of Computers Sciences and Math.\ University of Mosul.

Received:1/10 /2011 ____________________Accepted: 21 /12 / 2011

effective methods for solving unconstrained nonlinear optimization problem (1), its iterative formula is given by:

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

and

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

Where d_k is the search direction, α_k step-length which is computed by carrying out a line search.

The main step-length rules are as follows:

1. Armigo-Goldenstien rule. Find an α_k>0 such that

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k g_k^T d_k \qquad (4)$$

$$f(x_k + \alpha_k d_k) - f(x_k) \ge (1-\delta)\alpha_k g_k^T d_k , \delta \in (0,1/2) \qquad (5)$$

2. Weak Wolfe-Powell rule (WWP). Find an α_k>0 satisfying (4) and

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k , \sigma \in (\delta,1) \qquad (6)$$

3. Strong Wolfe-Powell rule (SWP). Find an α_k>0 satisfying (4) and

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \le -\sigma g_k^T d_k , \sigma \in (\delta,1) \qquad (7)$$

See (Hong & et al, 2009), where β_k is a scalar and g_k denotes $g(x_k)$. There are some famous formulas for β_k such as

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \quad \text{(Fletcher \& Reeves, 1964)} \qquad (8)$$

$$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}, \quad \text{(Polat \& Reber're, 1969)} \qquad (9)$$

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}, \quad \text{(Hesteness \& Stiefel, 1952)} \qquad (10)$$

$$\beta_k^{LS} = -\frac{g_{k+1}^T y_k}{d_k^T g_k}, \quad \text{(liu \& Story, 1991)} \qquad (11)$$

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

Hager-Zhang (Hager & Zhang, 2005) proposed a new formula as follows:

$$\beta_k^{HZ} = \left[y_k - \frac{2 d_k \|y_k\|^2}{d_k^T y_k} \right]^T \frac{g_{k+1}}{d_k^T y_k} \qquad (12)$$

Obviously, formula (12) can be rewrite as:

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

$$= \beta_k^{HS} - 2\frac{\|y_k\|^2 g_{k+1}^T d_k}{\left(d_k^T y_k\right)^2} \qquad (13)$$

With exact line search, the method reduces to a nonlinear version of the Hestenes- Stiefel CG- scheme. Moreover, d_k satisfies $g_k^T d_k \le -\frac{7}{8}\|g_k\|^2$.
This method can be regarded as a modified HS method.
Hager and Zhang proved that this method with the Wolfe line search converges globally (Hager & Zhang, 2006). Other interested results about the global convergence of CG methods can be found in (Dai & Yuan, 1996),(Dai, 1988), (Sun & Zhang, 2001) and (Liu & et al, 1995).

Our paper is organized as follows: In section 2 we defined the memoryless BFGS method. In section 3 we first derived the new modified of the HZ method and present a new algorithm, the sufficient descent property of new algorithm is also proved in this section. In section 4, we establish the global convergence of the new algorithm for the convex and general functions. The preliminary numerical results are contained in section 5.

2- The memoryless quasi-Newton method.

In order to introduce our method, let us simply recall the well-known BFGS quasi-Newton method. This type of method was suggested for the first time by Perry (Perry, 1978), he noted that in eq.(3) the scalar β_k was chosen to make d_k and d_{k+1} conjugate using an exact line search, since, in general, line search is not exact, Perry relaxed this requirement and he rewrote eq.(3) where β_k is defined by (10) in an equivalent form, but assuming inexact line search, thus he obtained:

$$d_{k+1} = -\left[I - \frac{d_k y_k^T}{d_k^T y_k}\right] g_{k+1} \tag{14}$$

but this matrix is not of full rank; hence he modified (14) as:

$$d_{k+1} = -\left[I - \frac{s_k y_k^T}{s_k^T y_k} + \frac{y_k s_k^T}{s_k^T y_k}\right] g_{k+1} \tag{15a}$$

$$= -Q_{k+1} g_{k+1} \tag{15b}$$

where $s_k = x_{k+1} - x_k$. Then (Shanno, 1978) addressed the issue that (15b) does not satisfy the actual QN- condition, so he modified it in order to make it do so, he then obtained

$$Q'_{k+1} = I - \left[\frac{s_k y_k^T + y_k s_k^T}{s_k^T y_k}\right] + \left[1 + \frac{y_k^T y_k}{s_k^T y_k}\right] \frac{s_k s_k^T}{s_k^T y_k} \tag{16}$$

This new form of the projection matrix Q'_{k+1} has a special relationship with the BFGS update formula which is defined by (Dennis & More', 1974, 1977) and (Al-Bayati,1996)

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

It is easily seen that (16) is equivalent to (17) with H_k replaced by I (i.e. if $H_k \equiv I$, where I is the identity matrix) then the above BFGS method becomes the memoryless BFGS.

The CG-method, which is referred to as a memoryless BFGS method, defined by

$$d_{k+1} = -Q'_{k+1} g_{k+1} \tag{18}$$

then

$$d_{k+1} = -g_{k+1} + \left[\frac{y_k^T g_{k+1}}{s_k^T y_k} - \left(1 + \frac{y_k^T y_k}{s_k^T y_k} \right) \frac{s_k^T g_{k+1}}{s_k^T y_k} \right] s_k + \frac{s_k^T g_{k+1}}{s_k^T y_k} y_k \tag{19}$$

The above equality can be rewritten in the following equivalent form:

$$d_{k+1} = -g_{k+1} + \left[\frac{g_{k+1}^T y_k}{s_k^T y_k} - \frac{y_k^T y_k}{s_k^T y_k} \frac{s_k^T g_{k+1}}{s_k^T y_k} \right] s_k + \frac{g_{k+1}^T s_k}{s_k^T y_k} (y_k - s_k) \tag{20}$$

for more details see (Zhang, 2006) and (Zhang, 2009),

since $s_k = x_{k+1} - x_k = \alpha_k d_k$, then (20) is equivalent:

$$d_{k+1} = -g_{k+1} + \left[\frac{g_{k+1}^T y_k}{d_k^T y_k} - \frac{y_k^T y_k}{d_k^T y_k} \frac{g_{k+1}^T d_k}{d_k^T y_k} \right] d_k + \frac{g_{k+1}^T d_k}{d_k^T y_k} (y_k - s_k) \tag{21}$$

equation (21) introduce a new CG- method:

$$d_{k+1} = -g_{k+1} + \beta_k d_k + \theta_k (y_k - s_k) \tag{22}$$

Where $$\beta_k = \left[\frac{g_{k+1}^T y_k}{d_k^T y_k} - \frac{\|y_k\|^2 (g_{k+1}^T d_k)}{(d_k^T y_k)^2} \right] \tag{23}$$

and $$\theta_k = \frac{g_{k+1}^T d_k}{d_k^T y_k} \tag{24}$$

If we use exact line search (21) reduces to the standard HS method.

(Zhang, 2009) replace term $d_k^T y_k$ in (21) by $(-d_k^T g_k)$ and define a new LS type CG-method defined by:

$$d_{k+1} = -g_{k+1} + \beta^{MLS} d_k + \theta_k (y_k - s_k) \tag{25}$$

Where $$\beta_k^{MLS} = \left[\frac{g_{k+1}^T y_k}{-d_k^T g_k} - \frac{\|y_k\|^2 (g_{k+1}^T d_k)}{(-d_k^T g_k)^2} \right] \tag{26}$$

and $$\theta_k = \frac{g_{k+1}^T d_k}{-d_k^T g_k} \tag{27}$$

he called the method (2) with (25) as a new LS (MLS) method.

It is clear that the MLS method reduces to the standard LS method if exact line search is used since, this line search ensure $g_{k+1}^T d_k = 0$.

3- A new Hager- Zhang conjugate gradient (MHZ) method.

(Al-Bayati, 1991) investigated another family of QN- method for which the updating matrix H_{k+1} was defined by:

$$H_{k+1} = H_k + \left[\frac{2y_k^T H_k y_k}{\left(s_k^T y_k\right)^2}\right] s_k s_k^T - \frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k} \qquad (28)$$

The above updating formula generates positive definite matrices. If we use memoryless (i.e. $H_k \equiv I$) then the above matrix is defined by:

$$H_{k+1} = I + \left[\frac{2y_k^T y_k}{\left(s_k^T y_k\right)^2}\right] s_k s_k^T - \frac{y_k s_k^T + s_k y_k^T}{s_k^T y_k} \qquad (29)$$

Since, $d_{k+1} = -H_{k+1} g_{k+1}$ (30)

hence

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

$$= -g_{k+1} - \left[\frac{2y_k^T y_k}{s_k^T y_k}\frac{s_k^T g_{k+1}}{s_k^T y_k}\right] s_k + \frac{s_k^T g_{k+1}}{s_k^T y_k} y_k + \frac{y_k^T g_{k+1}}{s_k^T y_k} s_k$$

$$= -g_{k+1} + \left[\frac{y_k^T g_{k+1}}{s_k^T y_k} - 2\frac{y_k^T y_k}{s_k^T y_k}\frac{s_k^T g_{k+1}}{s_k^T y_k}\right] s_k + \frac{s_k^T g_{k+1}}{s_k^T y_k} y_k \qquad (31)$$

Since $s_k = \alpha_k d_k$

$$d_{k+1} = -g_{k+1} + \left[\frac{y_k^T g_{k+1}}{d_k^T y_k} - 2\frac{\|y_k\|^2}{d_k^T y_k}\frac{d_k^T g_{k+1}}{d_k^T y_k}\right] d_k + \frac{d_k^T g_{k+1}}{d_k^T y_k} y_k \qquad (32)$$

equation (32) represent a new HZ type define by:

$$d_{k+1} = -g_{k+1} + \beta_k^{HZ} d_k + \theta_k y_k \qquad (33)$$

where β_k^{HZ} is defined in (12) and $\theta_k = \frac{g_{k+1}^T d_k}{d_k^T y_k}$. We call the method (2) and (33) as MHZ.

Now, we present concrete algorithm as follows:

3.1 Algorithm of the MHZ method.

Step 1: set k=1, $d_k = -g_k / \|g_k\|$.

Step 2: set $x_{k+1} = x_k + \alpha_k d_k$, where α is a scalar chosen in such a way that $f_{k+1} < f_k$.

Step 3: check for convergence, i.e. if $\|g_{k+1}\| < \in$, where $\in$ is a small positive tolerance, stop.

step 4: otherwise, if (k=n), or $\left|g_{k+1}^T g_k\right| \ge 0.2\left|g_{k+1}^T g_{k+1}\right|$

Compute the new search direction defined by:

$d_{k+1} = -g_{k+1}\left(\alpha_k d_k^T d_k / g_{k+1}^T g_{k+1}\right)$;

set k=1, and go to step 2. Else, set k=k+1.
Step 5: compute the new search direction defined in (33) and go to step 2.
Next, we will establish the search direction defined by (32) always yield descent direction.

3.2 Theorem

Let $\{x_{k+1}\} and \{d_{k+1}\}$ be generated by the new HZ (MHZ), then we have

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

Proof:

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

$$= -\|g_{k+1}\|^2 + \frac{(g_{k+1}^T y_k)(g_{k+1}^T d_k)}{d_k^T y_k} - 2\frac{\|y_k\|^2 (g_{k+1}^T d_k)^2}{(d_k^T y_k)^2} + \frac{(g_{k+1}^T d_k)(g_{k+1}^T y_k)}{d_k^T y_k}$$

$$= -\|g_{k+1}\|^2 + 2\frac{(g_{k+1}^T y_k)(g_{k+1}^T d_k)}{d_k^T y_k} - 2\frac{\|y_k\|^2 (g_{k+1}^T d_k)^2}{(d_k^T y_k)^2},$$

use the fact $u^T v \le \frac{1}{2}(u^2 + v^2)$ with, $u = \frac{1}{\sqrt{2}} g_{k+1}$ and $v = \frac{\sqrt{2} g_{k+1}^T d_k}{d_k^T y_k} y_k$ to get:

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

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

$$= -\|g_{k+1}\|^2 + \frac{1}{2}\|g_{k+1}\|^2$$

$$= -\frac{1}{2}\|g_{k+1}\|^2, \text{ let } (c=1/2>0) \tag{34}$$

$$\therefore d_{k+1}^T g_{k+1} \le -c\|g_{k+1}\|^2. \tag{35}$$

Which satisfy the sufficient descent condition.

4- The global convergence of MHZ method.

In order to establish the global convergence result for the MHZ, we will impose the following assumptions of f, which have been used often in the literature to analyze the global convergence of CG methods with inexact line search.

Assumption (A):

(i) the level set $\Omega = \{x / f(x) \le f(x_0), x \in R^n\}$ is bounded.

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

$$\|g(x)-g(y)\| \le L\|x-y\|, \forall x, y \in \Omega. \quad (36)$$

The above assumption implies that there exists $\gamma > 0$, such that

$$\|g_k\| \le \gamma, \forall x \in \Omega \quad (37)$$

4.1 Theorem

Suppose that the Assumption (A) holds and consider any CG methods (2) and (3), where the direction d_{k+1} given by (32) is a descent direction and α_k is obtained by the SWP if

$$\sum_{k\ge 1}\frac{1}{\|d_k\|^2} = \infty$$

then

$$\lim_{k\to\infty}\inf\|g_k\| = 0 \quad (38)$$

Proof:

From (36) we get ($y \le Ls$), then

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

$$\le \|g_{k+1}\|\left[\frac{L\alpha_k\|d_k\|}{|d_k^T y_k|} + 2\frac{L^2\alpha_k^2\|d_k\|^3}{|d_k^T y_k|^2}\right] \quad (39)$$

also,

$$|\theta_k| = \left|\frac{g_{k+1}^T d_k}{d_k^T y_k}\right| \le \|g_{k+1}\|\frac{\|d_k\|}{|d_k^T y_k|} \quad (40)$$

We have from (37), (39), (40) and the SWP that

$$\therefore \|d_{k+1}\| \le \|g_{k+1}\| + |\beta_k^{HZ}|\|d_k\| + |\theta_k|\|y_k\|$$

$$\le \|g_{k+1}\| + \|g_{k+1}\|\left[\frac{L\alpha_k\|d_k\|}{|d_k^T y_k|} + 2\frac{L^2\alpha_k^2\|d_k\|^3}{|d_k^T y_k|^2}\right]\|d_k\| + \|g_{k+1}\|\frac{\|d_k\|}{|d_k^T y_k|}L\alpha_k\|d_k\|$$

$$= \|g_{k+1}\|\left[1 + \frac{L\alpha_k\|d_k\|^2}{|d_k^T y_k|} + 2\frac{L^2\alpha_k^2\|d_k\|^4}{|d_k^T y_k|^2} + \frac{L\alpha_k\|d_k\|^2}{|d_k^T y_k|}\right]$$

$$= \|g_{k+1}\|\left[1 + 2\frac{L\alpha_k\|d_k\|^2}{|d_k^T y_k|} + 2\frac{L^2\alpha_k^2\|d_k\|^4}{|d_k^T y_k|^2}\right]$$

From the SWP, we set

$$d_k^T y_k \ge (\sigma-1)g_k^T d_k \ge (\sigma-1)c\|g_k\|^2$$

$$\|d_{k+1}\| \le \|g_{k+1}\|\left[1 + 2\frac{L\alpha_k\|d_k\|^2}{(\sigma-1)c\|g_k\|^2} + 2\frac{L^2\alpha_k^2\|d_k\|^4}{(\sigma-1)^2c^2\|g_k\|^2}\right]$$

$$\leq \|g_{k+1}\| \left[1 + 2\frac{L\alpha_k B^2}{(\sigma-1)c\bar{\gamma}^2} + 2\frac{L^2\alpha_k^2 B^4}{(\sigma-1)^2 c^2 \bar{\gamma}^4} \right], \text{(since } \|d_k\| \leq B\text{)}$$

$$= \|g_{k+1}\| M \leq \gamma M.$$

This relation implies

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

Therefore, we have $\underset{k\to\infty}{Lim} \inf \|g_k\| = 0$.

Which for uniformly convex function equivalent to $\underset{k\to\infty}{Lim} \|g_k\| = 0$.

Now to prove that the new algorithm is global convergence for general function, we establish a bounded for the change $(w_{k+1} - w_k)$ in the normalized direction $w_k = d_k / \|d_k\|$, which we will use to conclude, by contradiction, that the gradients cannot be bounded away from zero (Nocedal & Glibart, 1992).

4.2 Lemma

Suppose that Assumption (A) holds, consider the method (2)-(3), with $\beta_k \geq 0$, and the line search satisfying the Zoutendijk condition and the sufficient descent condition if (37) holds, then $d_k \neq 0$ and

$$\sum_{k=1}^{\infty} \|w_{k+1} - w_k\|^2 < \infty \qquad (41)$$

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

Proof:

We can rewrite (33) in the following form

$$d_{k+1} = -g_{k+1} + \theta_k y_k + \beta_k^{MHZ} d_k = v_k + \beta_k^{MHZ} d_k .$$

Now, we define

$$u_{k+1} = \frac{v_{k+1}}{\|d_{k+1}\|} \ \ and \ \ r_{k+1} = \frac{\beta_k^{MHZ}\|d_k\|}{\|d_{k+1}\|}. \qquad (42)$$

From (3), we have for $\forall k \geq 1$

$$w_{k+1} = u_{k+1} + r_{k+1} w_k . \qquad (43)$$

Using the identity $\|w_{k+1}\| = \|w_k\| = 1$ and (43), we have

$$\|u_{k+1}\| = \|w_{k+1} - r_{k+1}w_k\| = \|r_{k+1}w_{k+1} - w_k\| \qquad (44)$$

(the last equality can be verified by squaring both sides). Using the condition $r_{k+1} \geq 0$, the triangle inequality, and (44), we obtain

$$\|w_{k+1} - w_k\| \leq \|(1 + r_{k+1})(w_{k+1} - w_k)\|$$
$$\leq \|w_{k+1} - r_{k+1}w_k\| + \|r_{k+1}w_{k+1} - w_k\|$$
$$= 2\|u_{k+1}\|. \qquad (45)$$

Now, using SWP and (37), we get

$$\|v_{k+1}\| = \left\| -g_{k+1} + \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k \right\| \le \left\| -g_{k+1} + \frac{\sigma}{(\sigma-1)} \frac{g_k^T d_k}{g_k^T d_k} y_k \right\|$$

$$\le \|g_{k+1}\| + \left| \frac{\sigma}{(\sigma-1)} \right| \|y_k\|$$

since, $y_k = g_{k+1} - g_k$, therefore $\|y_k\| \le \|g_{k+1}\| + \|g_k\|$

$$\|v_{k+1}\| \le \gamma + \left| \frac{\sigma}{(\sigma-1)} \right| 2\gamma$$

$$= \gamma \left(1 + 2 \left| \frac{\sigma}{(\sigma-1)} \right| \right) = A.$$

Now, by taking the summation of the first part of equation (42),we obtain

$$\sum_{k=0}^{\infty} \|u_{k+1}\|^2 = \sum_{k=0}^{\infty} \frac{\|v_{k+1}\|^2}{\|d_{k+1}\|^2} \le \sum_{k=0}^{\infty} \frac{A^2}{\|d_k\|^2}$$

$$= \sum_{k=0}^{\infty} \frac{A^2}{\|g_k\|^4} \frac{\|g_k\|^4}{\|d_k\|^2} \le \frac{A^2}{\bar{\gamma}^4} \sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty .$$

The above inequality with (45) completes our proof.

Property (*):

Consider a method of the form (2)-(3), and suppose that

$$0 < \bar{\gamma} \le \|g_k\| \le \gamma , \qquad (46)$$

for all $k \ge 1$. Under this assumption we say that the method has property (*) if there exists a constant b>1 and λ>0 such that for all k:

$$|\beta_k| \le b, \qquad (47)$$

And

$$\|s_{k-1}\| \le \lambda \Rightarrow |\beta_k| \le \frac{1}{2b}. \qquad (48)$$

See (Nocedal& Glibart , 1992).

We use this property to show that asymptotically the search directions are generated by the Algorithm (3.1) where β_k is defined by (32) change slowly.

4.3 Lemma

Suppose that the Assumption (A) holds. Consider the CG-algorithm (3.1), and α_k is obtained by the SWP, and for all $k \ge 0$, there exists the positive constant $\bar{\gamma}$ such that $\|g_k\| \ge \bar{\gamma}$, then the new algorithm (3.1) has property (*).

Proof:

It follows from the definition of β_k^{MHZ}, the second SWP and (46)

$$\begin{aligned}
\left|\beta_k^{MHZ}\right| &= \left|\frac{g_{k+1}^T y_k}{d_k^T y_k} - 2\frac{\|y_k\|^2}{d_k^T y_k}\frac{g_{k+1}^T d_k}{d_k^T y_k}\right| \\
&\le \left|\frac{g_{k+1}^T y_k}{(1-\sigma)\|g_k\|^2} - 2\frac{\|y_k\|^2}{(1-\sigma)\|g_k\|^2}\frac{\sigma}{(1-\sigma)}\frac{g_k^T d_k}{d_k^T g_k}\right| \\
&\le \frac{2\gamma^2}{(1-\sigma)\bar{\gamma}^2} + \frac{2(4\gamma^2)}{(1-\sigma)\bar{\gamma}^2}\frac{\sigma}{(1-\sigma)} \\
&\le \frac{\left(2\gamma^2(1-\sigma)+8\gamma^2\sigma\right)}{(1-\sigma)^2\bar{\gamma}^2} \equiv b. \qquad (49)
\end{aligned}$$

Without loss of generality we can define b such that b>1. let us define

$$\lambda \equiv \frac{(1-\sigma)^2\bar{\gamma}^2}{2(\gamma(1-\sigma)+4\gamma\sigma)Lb}. \qquad (50)$$

Obviously, if $\|s_{k-1}\| \le \lambda$, then from (49) we have

$$\begin{aligned}
\left|\beta_k^{MHZ}\right| &\le \left[\frac{\|g_{k+1}\|}{\left|d_k^T y_k\right|} + 2\frac{\|y_k\|\left|g_{k+1}^T d_k\right|}{\left|d_k^T g_k\right|^2}\right]\|y_k\| \\
&\le \left[\frac{\gamma}{(1-\sigma)\bar{\gamma}^2} + 2\frac{2\gamma\sigma}{(1-\sigma)^2}\right]L\lambda \\
&= \frac{\gamma(1-\sigma)+4\gamma\sigma}{(1-\sigma)\bar{\gamma}^2} = \frac{1}{2b} \qquad (51)
\end{aligned}$$

Therefore, for b and λ are defined in (49) and (50), respectively, it follows that the relations (46) and (47) holds.

It is clear that many other choices of β_k give rise to algorithm with property (*). For example, if $|\beta_k|$ and $\beta_k^+ = \max\{\beta_k, 0\}$.

The next Lemma shows that if the gradient is bounded away from zero, and if the method has property (*), then a fraction of the steps cannot be too small. Therefore, the algorithm makes a rapid progress to the optimum. We let N^* denote the set of positive integers, and for λ >0, we define the set of index

$$K_{k,\Delta}^{\lambda} = \left\{i \in N^* : k \le i \le k+\Delta-1, \|s_{k-1}\| \ge \lambda\right\}, \qquad (52)$$

the following Lemma is similar to Lemma 3.5 in (Dai, 2001) and Lemma 4.2 in (Nocedal & Glibart, 1992).

4.4 Lemma

Suppose that all Assumptions of Lemma 4.4 are satisfied. Then there exists a λ >0 such that for any $\Delta \in N^*$ and any index k_0, there is a greater index $k \ge k_0$ such that $\left|K_{k,\Delta}^{\lambda}\right| > \Delta/2$.

The prove of Lemma is similar to Theorem 3.6 in (Dai, 2001) or to Theorem 3.2 in (Hager & Zhang, 2005) and the proof is omitted here.

4.5 Theorem

Suppose that Assumption (A) holds. Consider the method (2)-(3) with the following three properties:

(i) $\beta_k \geq 0$ for all k;

(ii) the line search satisfies the Zoutendijk condition (38), and the sufficient descent condition (35);

(iii) Property (*) holds.

Then $\lim_{k\to\infty} \inf \|g_k\| = 0$, for more details see (Nocedal & Glibart, 1992).

i.e. the new algorithm (3.1) with β_k^{MHZ} is satisfying the global convergence.

5- Numerical results

In this section, we compare the performance of the new method MHZ on a set of test functions. The cods were written in Fortran77 and in double precision arithmetic. Our experiments are performed on a set of (25) nonlinear unconstrained test functions. These test functions are contributed in CUTE (Bongratz & et al, 1995) and (Andrei, 2008).

All these algorithms are implemented with strong Wolfe Powell line search conditions (4) and (7), with $\delta = 0.001 \, and \, \sigma = 0.9$. In all these methods terminate when the following stopping criterion is met:

$$\|g_{k+1}\| < 1 \times 10^{-5} \qquad (53)$$

We record the Number Of Iterations calls (NOI), the Number Of Function evaluations calls (NOF), and the dimensions of test problems calls (N), for the purpose of our comparisons.

In Table (1), we compare some numerical results for Shanno and MHZ methods. In order to summarize our numerical results, we are concerned only with the total of N different dimensions, N=100, 500, 1000, 5000, 10000 and 50000.

In Table (2), we compare the percentage performance of the new MHZ against Shanno Methods for the total of our 25 test functions. While in Table (3) we compare between Shanno , Hestenes and Stiefel (HS), Hager-Zhang (HZ) and Modified Hager- Zhang (MHZ) methods for the total of N different dimensions [N=100, 1000 & 10000] for each test function.

From these three tables we see that for more test functions the MHZ method are really much better than others.

Table (1): comparison between Shanno and MHZ methods for the total of N different dimensions [N=100, 500, 1000, 5000,10000 & 50000] for each test function.

Test functions	A 3-Terms Method (Shanno) NOI/NOF	A 3-Terms New Method (MHZ) NOI/NOF	Test functions	A 3-Terms Method (Shanno) NOI/NOF	A 3-Terms New Method (MHZ) NOI/NOF
Extended Block diagonal	288/618	213/494	Diagonal (5)	17/71	17/71
Extended Cliff	46/212	46/212	Extended Trigonometric	66/195	67/197
Extended Beale	57/179	61/194	Extended PSC1	51/417	47/135
Generalized Tridiagonal (1)	114/266	120/277	Himmelblau (CUTE)	100/1445	99/1442
QUARTC (QF2) (CUTE)	34/121	34/121	Cubic	94/289	87/278
Strait	36/125	35/123	Fred	48/161	51/142
Extended Tridiagonal (1)	50/202	50/206	Extended White & Holst	94/289	87/278
Generalized Tridiagonal (2)	244/520	239/507	DQDRTIC (CUTE)	30/107	30/106
DIXMAANA (CUTE)	29/94	28/94	Diagonal (6)	12/63	12/63
DIXMAANB (CUTE)	30/110	30/110	Rosenbrock	179/467	173/464
Diagonal (4)	12/63	12/63	Wolfe	374/814	370/801
Cosine (CUTE)	50/139	50/139	Extended Rosenbrock	179/476	173/464
Extended Powell	281/802	278/801	TOTAL	4720/12653	4651/12275

Table (2) :percentage performance of the new MHZ against Shanno Methods for the total of our 25 test functions.

Tools	Shanno method	MHZ method
NOI	100 %	98.5 %
NOF	100 %	97.0 %

Table (3): between Shanno , Hestenes and Stiefel (HS), Hager-Zhang (HZ) and Modified Hager-Zhang (MHZ) methods for the total of N different dimensions [N=100, 1000 & 10000] for each test function.

Test functions	N	Shanno Method NOI/ NOF	HS Method NOI/ NOF	HZ Method NOI/ NOF	MHZ Method NOI/ NOF
Extended Block diagonal (1)	100 1000 10000	56/ 116 51/ 107 52/ 112	57/ 118 54/ 113 52/ 112	35/ 74 59/ 123 50/ 109	52/108 24/ 37 17/ 48
Extended Cliff	100 1000 10000	6/ 26 7/31 8/37	6/ 26 7/ 31 7/ 40	8/ 33 8/ 36 7/ 34	8/ 33 6/ 28 8/ 35
Extended Beale	100 1000 10000	10/ 32 10/ 31 8/ 28	10/ 32 10/ 31 8/ 28	10/ 31 10/ 31 8/ 28	9/ 27 9/27 8/ 28
Generalized Tridiagonal (1)	100 1000 10000	19/ 42 19/ 44 19/ 46	20/ 45 20/ 47 19/ 46	19/ 42 19/ 44 19/ 45	20/ 44 20/ 46 20/ 47
QUARTC (QF2)	100	87/ 179	81/ 168	78/ 161	83/ 171

(CUTE)	1000	315/ 636	334/ 675	287/580	282/570
	10000	1131/2270	1093/2194	1049/2106	982/1972
Strait	100	6/ 18	6/ 18	6/ 18	6/ 18
	1000	6/ 21	6/ 21	5/ 21	5/ 21
	10000	6/ 22	6/ 22	6/ 22	6/ 22
Extended Tridiagonal (1)	100	9/ 30	9/ 30	9/ 30	9/ 30
	1000	8/ 33	9/ 31	8/ 33	8/ 33
	10000	9/ 43	9/ 43	9/ 42	9/ 42
Generalized Tridiagonal (2)	100	35/ 73	36/ 75	42/ 87	36/ 85
	1000	38/ 82	38/ 82	40/ 84	40/ 84
	10000	42/ 90	41/ 89	42/ 90	39/ 83
DIXMAANA (CUTE)	100	4/ 12	4/ 12	4/ 12	4/ 12
	1000	4/ 15	4/ 15	4/ 15	4/ 15
	10000	5/ 17	5/ 17	5/ 17	5/ 17
DIXMAANB (CUTE)	100	5/ 15	5/ 15	5/ 15	5/ 15
	1000	5/ 23	5/ 23	5/ 23	5/ 23
	10000	5/ 18	5/ 18	5/ 18	5/ 18
Diagonal (4)	100	2/ 8	2/ 8	2/ 8	2/ 8
	1000	2/ 10	2/ 10	2/ 10	2/ 10
	1000	2/ 12	2/ 12	2/ 12	2/ 12
Cosine (CUTE)	100	8/20	8/20	8/20	8/20
	1000	8/22	8/22	8/22	8/22
	10000	9/ 25	9/ 25	9/ 25	9/ 25
Extended Powell	100	49/ 121	46/ 123	32/ 86	50/ 147
	1000	50/ 142	52/ 157	47/ 136	23/ 71
	10000	53/ 163	44/ 129	59/ 160	63/ 181
Diagonal (5)	100	3/ 10	3/ 10	3/ 10	3/ 10
	1000	2/ 10	2/ 10	2/ 10	2/ 10
	10000	3/ 13	3/ 13	3/ 13	3/ 13
Extended Trigonometric	100	11/30	11/30	11/30	11/30
	1000	11/ 32	11/ 32	11/ 32	11/ 32
	10000	11/ 34	11/ 34	12/ 36	11/ 34
Extended PSC1	100	8/ 21	8/ 21	8/ 21	8/ 21
	1000	8/ 22	8/ 22	8/ 22	8/ 22
	10000	8/ 24	8/ 24	8/ 24	8/ 24
Himmelblau (CUTE)	100	15/122	14/ 114	18/ 150	14/ 112
	1000	23/ 233	24/ 218	21/ 213	22/ 223
	10000	17/ 242	11/ 181	11/ 184	20/ 277
Cubic	100	16/ 45	16/ 45	16/ 45	16/ 45
	1000	16/ 46	16/ 46	16/ 46	16/ 46
	10000	16/56	16/56	13/49	13/50
Fred	100	8/ 24	8/ 24	9/ 25	6/ 17
	1000	8/ 28	8/ 25	9/ 25	10/ 26
	10000	8/ 27	8/ 27	8/ 27	8/ 22
Extended White & Holst	100	16/ 45	16/ 45	16/ 45	16/ 45
	1000	16/ 46	16/ 46	16/ 46	16/ 46
	10000	16/ 56	16/ 56	13/ 49	13/ 50
DQDRTIC (CUTE)	100	5/ 15	5/ 15	5/ 15	5/ 15
	1000	5/ 17	5/ 17	5/ 17	5/ 17
	10000	5/ 19	5/ 19	5/ 19	5/ 19
Diagonal (6)	100	2 /8	2 /8	2 /8	2 /8
	1000	2/ 10	2/ 10	2/ 10	2/ 10
	10000	2/ 12	2/ 12	2/ 12	2/ 12
Rosenbrock	100	30/ 78	30/ 78	33/ 83	30/ 76
	1000	29/ 77	29/ 77	32/ 85	27/ 73
	10000	30/ 81	30/ 81	32/ 88	30/ 82
Wolfe	100	65/ 136	71/ 149	70/ 146	69/ 144
	1000	62/ 132	64/ 137	62/131	57/121
	10000	60/ 133	69/ 150	58/ 125	60/ 132
Extended Rosenbrock	100	30/ 78	30/ 78	33/ 83	30/ 78
	1000	29/ 77	29/ 77	32/85	27/ 73
	10000	30/ 81	30/ 81	32/ 88	30/ 82

References

[1] Al-Bayati, A. Y., (1991), **"Anew family of self-scaling variable metric algorithms of unconstrained optimization"**. Journal of Education and Science, Mosul University, 12, pp.25-54.

[2] Al-Bayati, A. Y., Ahmed, H. I., (1996)," **Investigation of single-update memoryless CG-methods"**. Qutar University Science Joural, University of QATAR, 16 (2), pp.183-192.

[3] Andrei, N., **(2008), "An unconstrained optimization test functions collection"**. Journal of Advance Modeling and Optimization, 10, pp0147-161.

[5] Bongratz, K. E., Conn, A. R., Gould, N. I. M. and Toint, P. L., (1995), CUTE: **constrained and unconstrained testing environments**. ACM Trans. Math. Softw., 21, pp.123-160.

[6] Dai, Y. H., (1988), **"Some new properties of nonlinear conjugate gradient method"**. Research report ICM-98-010. Beijing: Institute of Computational Math. And Scientific/ Engineery Computing, Chinese Academy of Sciences.

[7] Dai, Y. H. and Yuan, Y., (1996), **"convergence properties of the conjugate descent method"**. Adv. Math., 25, pp.552-562.

[8]Dennis, J. E. and More', J. J., (1974), **"A characterization of superlinear convergence and its application to quasi-Newton methods"**. Math. Compute., 28, pp.549-560.

[9] Dennis, J. E. and More', J. J., (1977)**, "Quasi-Newton method, motivation and theory"**. SIAM Rev., 19, pp.46-89.

[10] Fletcher, R., and Reeves, C., (1964), Function minimization by conjugate gradient, *Computer J.*, 7, pp.149-154.

[11] Hager, W. and Zhang, H., (2005), **"A new conjugate gradient method with gruanteed descent and efficient line search"**. SIAM J. on Optimization, (16), pp.170-192.

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

[13] Hong, L., Mo, L. and Wei, Z. X.,(2009). **"a modified conjugate gradient method with global convergence property"**. Journal of Math. Research and Exposition, 29, pp.65-75. DOI:10.3770/j.issn:1000-341 X.2009.01.009.

[14] Hestenes, M. R., and Stiefel, E. L., (1952**), "Methods of conjugate gradient for solving linear systems"**. *J. Research Nat. Bur. Standards*, 49, pp.409-436.

[15] Liu, Y. and Story,(1991), **"Efficient generalized conjugate gradient algorithm"**, Part I: Theory J. Optim. Theory Appl., 69, pp.129-137.

[16] Liu, G., Han, J. and Yin, H., (1995), **"Global convergence of the Fletcher- Reeves algorithm with inexact line search"**. Appl. Math. J. chin. Univ. Ser. B, 10, pp.75-82.
[17] Nocedal, J. and Glibart, J., (1992), **" Global convergence properties of conjugate gradient methods for optimization"**, SIAM J., Optimization, 2, pp.21-42
[18] Perry, A., (1978), "**A modified conjugate gradient algorithm"**. Journal of Operations Research, 26, pp.1073-1078.
[19] Polat, E., and Ribe're, G. R., (1969), **"Note sur la convergence de direction conjuge'cs"**. Rev. Francaise Informat conjugate gradient methods Research Operational, 3e Anne's, pp.16: 35-43
[20] Polyak, B. T., (1969), "**The conjugate gradient method in extreme problems"**. USSR Comp. math. And Math. Phys., 9, pp.94-112**.**
[21] Shanno, F. D., (1978**)**, **"Conjugate gradient methods with inexact searches"**, Mathematics of Operations Research, 3, pp.244-256.
[22] Sun, J. and Zhang, J., (2001), **"Convergence of conjugate gradient methods without line search"**. Ann. Oper. Res., 103, pp.161-173.
[23] Wolfe, P., (1969), **"Converge condition for ascent methods"**. *SIAM Review*, 11, pp.226-235.
[24] Wolfe, P., (1971), **" Convergence conditions for ascent methods II: some corrections"**. SIAM Review, 13,pp.185-188.
[25] Zhang, L., zhou, W., and Li, D. H., (2006),**" A descent modified Polak-Ribire're-Polyak conjugate gradient mehod and its global convergence"**.IAM Journal of Numerical Analysis, 26, pp.629-640.
[26] Zhang, L., (2009), **"A new Liu- Story type nonlinear conjugate gradient method for unconstrained optimization problems"**. Journal of Compute. And Applied Math., 225, pp.146-157.
[27] Zoutendijk, G., (1970), **"Nonlinear programming, computational methods"**. Integer and Nonlinear Programming (J. bbadie ed.). North-Holland, pp.37-86.