A Modified Hestenes-Stiefel Conjugate Gradient Method and its Global convergence for unconstrained optimization

Ghada M. Al-Naemi*
Department of Mathematics, College of Computers Science and Mathematics, University of Mosul, Mosul, Iraq

Abstract
In this paper, we proposed a modified Hestenes-Stiefel (HS) conjugate gradient method. This achieves a high order accuracy in approximating the second order curvature information of the objective function by utilizing the modified secant condition which is proposed by Babaie-Kafaki [1], also we derive a non-quadratic conjugate gradient model. The important property of the suggestion method that is satisfy the descent property and global convergence independent of the accuracy of the line search. In addition, we prove the global convergence under some suitable conditions, and we reported the numerical results under these conditions.

Keywords: Conjugate gradient, modified secant condition, decent property, global convergence, non-quadratic model.

تطوير طريقة Hestenes-Stiefel للتدرج المترافق وتقاربها الشمولي في الأمثلية غير المقيدة

غادة مؤيد ألنعيمي*
قسم الرياضيات ، كلية علوم الحاسوب والرياضيات ، جامعة الموصل ، الموصل، العراق.

الخلاصة:
في هذا البحث، تم اقتراح تطوير لطريقة (HS) Hestenes-Stiefel للتدرج المترافق. وللحصول على تقريب عالي ودقيق لدالة الهدف، ثم استخدام تطوير شرط القاطع الذي اقترح من قبل مجموعة من الباحثين في بحث قدم من قبل ([1] Babaie-Kafaki)، وكذلك في هذا البحث تم اشتقاق نموذج غير تربيعي للتدرج المترافق. الخاصية المهمة للطريقة المقترحة أنها تحقق الانحدار والتقارب المطلق بغض النظر عن أي نوع من أنواع خط البحث. بالإضافة إلى ذلك تم برهان تقاربها المطلق تحت بعض الشروط المناسبة، لقد تم الحصول على النتائج العددية باستخدام هذه الشروط.

Introduction

Consider the following n-dimensional unconstrained minimization problem

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

where $f : R^n \to R$ is continuously differentiable and its gradient $g(x_k) = \nabla f(x_k)$ is available. The conjugate gradient method is a powerful line search method for solving unconstrained optimization problems, and it remains very popular for engineers and mathematicians who are interested in solving large scale problems.

Conjugate gradient method is an iterative method, this method can be described by the following:

$$x_k = x_{k-1} + \alpha_{k-1} d_{k-1}, \quad (2)$$

*Email: ghalnaemi@yahoo.fr

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

where $x_k \in R^n$ is the k-th approximation to the solution of (1), α_k is a positive step size which is determined by some line search direction, d_k is a search direction, and $\beta_k \in R$ is a scalar 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 [2, 3, 4, 5, 6, 7] formulas are well known and they are given by:

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \qquad \beta_k^{PR} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2},$$

$$\beta_k^{LS} = -\frac{g_k^T y_{k-1}}{d_{k-1}^T g_{k-1}}, \qquad \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^{CD} = -\frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}}$$

Respectively, where $y_{k-1} = g_k - g_{k-1}$, $s_{k-1} = x_k - x_{k-1} = \alpha_{k-1} d_{k-1}$ and $\|.\|$ denote the Euclidean norm. All these methods are equivalent, if f(x) is a strictly convex quadratic function, and the performed line search is exact, but for general function different choices of β_k give rise to distinct conjugate gradient methods with quite different computational efficiency and converge properties [8].

The reminder of this paper is organized as follows: In section 2, we present our motivation and our proposed new conjugate gradient method and we derive a non-quadratic model. While in Section 3, we present the global convergence analysis for general functions and generalize our technique to the rest of the conjugate gradient method. Finally in section 4, the numerical; experiments are reported.

Modified Hestenes-Stiefel (HS) conjugate gradient method

For any unconstrained optimization method, including quasi-Newton method the search direction can be written in the form

$$d_k = -B_k g_k,$$

where B_k is $n \times n$ symmetric and positive definite matrix satisfying the quasi-Newton equation:

$$B_k s_{k-1} = y_{k-1}. \qquad (4)$$

The researcher in [9, 10] modified condition (4) with vector parameter, by the following formula:

$$B_k s_{k-1} = z_{k-1}, \text{ and } z_{k-1} = y_{k-1} + \eta_{k-1} \frac{\theta_{k-1}}{s_{k-1}^T u} u,$$

and θ_{k-1} is defined by

$$\theta_{k-1} = 6(f_{k=1} - f_k) + 3(g_k + g_{k-1})^T s_{k-1}. \qquad (6)$$

where $u \in R^n$ is a vector parameter satisfying $s_k^T u \neq 0$. Observing that this modified quasi-Newton equation contains not only gradient value information but also function value information at the previous and the present step.

After that [11] proposed the following direction:

$$d_k = -\left[1 + \max\left(\frac{g_k^T y_{k-1}^*}{d_{k-1}^T y_{k-1}^*}, 0\right) - t\frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}^*}\right] g_k + \bar{\beta}_k^* d_{k-1},$$

and,

$$y_{k-1}^{*} = y_{k-1} + \eta_{k-1}\frac{\max\{\theta_{k-1},0\}}{s_{k-1}^{T}u}u, \qquad (7)$$

where θ_{k-1} is defined in (6), u is any vector satisfying $s_{k-1}^{T}u > 0$, the parameter η_{k-1} is restricted to values of $\{0,1\}$.

Moreover, [12], proved that if $\|s_{k-1}\|$ is sufficiently small, then

$$s_{k-1}^{T}(G_k s_{k-1} - y_{k-1}) = O\left(\|s_{k-1}\|^{3}\right),$$

$$s_{k-1}^{T}(G_k s_{k-1} - z_{k-1}) = O\left(\|s_{k-1}\|^{4}\right).$$

Clearly, the above equations imply that the modified secant equation (5) is superior than the classical one in equation (4), because that z_k better approximates $\nabla^2 f_k s_{k-1}$ than y_{k-1}. Recently, [1] noticed that if $\|s_{k-1}\| > 1$, the standard secant equation (4) is expected to be more accurate than the modified secant equation (5). In order to overcome this difficulty, he considered an extension of the modified secant equation (5) by the following formula:

$$B_k s_{k-1} = \bar{z}_{k-1}, \qquad (8a)$$

and

$$\bar{z}_{k-1} = y_{k-1} + \eta_{k-1}\frac{\max\{\theta_{k-1},0\}}{s_{k-1}^{T}u}u, \qquad (8b)$$

where the parameter $\eta_{k-1} \in \{0,1\}$ and adaptively switch between standard secant equation (4), by setting $\eta_{k-1} = 1$ if $\|s_{k-1}\| \le 1$ and setting $\eta_{k-1} = 0$, otherwise.

By taking into consideration the theoretical advantages of the modified secant equation (8), and the strong converge properties of the conjugate gradient method which have the property $\beta_k \ge 0$ [6, 13]. So, we are going to propose a new conjugate gradient method as follows, let the search direction be defined by:

$$d_k = -\left(1 + \beta_k^{MHS}\frac{g_k^{T}d_{k-1}}{\|g_k\|^{2}}\right)g_k + \beta_k^{MHS}d_{k-1}, \qquad (9)$$

where β_k^{MHS} is the propose modification of formula β_k^{HS} as follows:

$$\beta_k^{MHS} = \beta_k^{*} = \frac{g_k^{T}\bar{z}_{k-1}}{\bar{z}_{k-1}^{T}d_{k-1}} \qquad (10)$$

where θ_{k-1} and $\bar{z}_{k-1}$ are defined in (6) and (8b) respectively.

In this paper we will going to evaluate η_{k-1} in (8b) by using a more general model than quadratic one is suggested as a basis for a conjugate gradient algorithm instead of using the parameter $\eta_{k-1} \in \{0,1\}$.

If q(x) is a quadratic function, then the function F(q(x)) is defined as a non-linear scaling of q(x) if the following condition holds:

$$f(x) = F(q(x)),\ \frac{df}{dq} = f' > 0 \text{ and } q(x) > 0.$$

The following scaling properties are immediately derived from the above conditions:

- Every counter line of q(x) is a counter line of f(x).
- If x^* is a minimizer of q(x), then it is a minimizer of f(x).
- If x^* is a local minimizer of q(x), then it is a local minimizer of f(x).

Boland [14] was the first who observed that q(x) and F(q(x)) have determined the same search directions so that the finite termination property for their algorithm was satisfied. A conjugate gradient algorithm that is minimizes the function

$f(x)=q(q(x))^p$, p>0 and $x \in R^n$,

in at most n-step which has been described by [15].

We summarized some various published works which is a special case of the above as following:

- A polynomial model which has been investigated by [14] which is defined by:

$F(q(x)) = \varepsilon_1 q(x) + \frac{1}{2}\varepsilon_2 q^2(x)$, where ε_1 *and* ε_2 are positive scalar.

- Two rational models which have been investigated by [16], the first one was

$F(q(x)) = \frac{(\varepsilon_1 q(x)+1)}{\varepsilon_2 q(x)}$, $\varepsilon_1 > 0,\ \varepsilon_2 < 0$, where ε_1 *and* ε_2 are scalar.

The second one was

$F(q(x)) = \frac{\varepsilon q(x)}{(q(x)+1)}$, $\varepsilon > 0$.

- Another rational model which has been investigated by [17], which defined by:

$F(q(x)) = \frac{\varepsilon_1 q(x)}{(1-\varepsilon_2 q(x))}$, $\varepsilon_1 > 0, \varepsilon_2 < 0$.

In this study , we consider a new rational model, define by:

$$F(q(x)) = \frac{1}{e^{q^2(x)} - 1},\ \ q(x) > 0 \ \ and \ \ \frac{df}{dq} > 0. \tag{11}$$

The unknown quantities η_{k-1} were expressed in term of available quantities of algorithm (i.e. function gradient value of the objective function) using the expression for η_{k-1}

$$\eta_{k-1} = \frac{f'_{k-1}}{f'_k}, \tag{12}$$

from the relations

$$g_k = f'_k G\left(x_k - x^*\right) \tag{13}$$

$$g_{k-1} = f'_{k-1} G\left(x_{k-1} - x^*\right). \tag{14}$$

Where G is the Hessian matrix and x^* is the minimum point, so from (12) we get:

$$\eta_{k-1} = \frac{g_{k-1}^T\left(x_k - x^*\right)}{g_k^T\left(x_{k-1} - x^*\right)},$$

but, $g_{k-1}^T\left(x_k - x^*\right) = g_{k-1}^T\left(x_{k-1} + \alpha_{k-1}d_{k-1} - x^*\right) = g_{k-1}^T\left(x_{k-1} - x^*\right) + \alpha_{k-1}g_{k-1}^T d_{k-1}$,

and, $g_k^T\left(x_{k-1} - x^*\right) = g_k^T\left(x_k - \alpha_{k-1}d_{k-1} - x^*\right) = g_k^T\left(x_k - x^*\right)$.

Since $g_k^T d_{k-1} = 0$, therefore, we can express η_{k-1}

$$\eta_{k-1} = \frac{g_{k-1}^T\left(x_{k-1} - x^*\right) + \alpha_{k-1}g_{k-1}^T d_{k-1}}{g_k^T\left(x_k - x^*\right)}. \tag{15}$$

From (13), (14) and (15), we get:

$$\eta_{k-1} = \frac{f'_{k-1}\left(x_{k-1} - x^*\right)G\left(x_{k-1} - x^*\right) + \alpha_{k-1}g_{k-1}^T d_{k-1}}{f'_k\left(x_k - x^*\right)G\left(x_k - x^*\right)},$$

therefore

$$\eta_{k-1} = \frac{2f'_{k-1}q_{k-1} + \alpha_{k-1}g_{k-1}^T d_{k-1}}{2f'_k q_k}. \tag{16}$$

If we calculate f'_{k-1} *and* f'_k from (11), we get:

$$f'_{k-1} = 2q_{k-1} f_{k-1}(1+f_{k-1}), \; and \; f'_k = 2q_k f_k(1+f_k). \tag{17}$$

Now, to find q(x) we can solving (11), its yield:

$$q(x) = \left[\ln\left(\frac{1+f}{f}\right)\right]^{\frac{1}{2}}, \tag{18}$$

By substituting (17) and (18) in (16), we get:

$$\eta_{k-1} = \frac{M_1 + w}{M_2}, \tag{19}$$

where $M_1 = \left[\ln\left(\frac{1+f_{k-1}}{f_{k-1}}\right)\right]^{\frac{1}{2}} * f_{k-1} * (1+f_{k-1})$,

$$M_2 = \left[\ln\left(\frac{1+f_k}{f_k}\right)\right]^{\frac{1}{2}} * f_k * (1+f_k),$$

and $w = \frac{\alpha_{k-1} g_{k-1}^T d_{k-1}}{4}$.

Note that, it easy to prove that the direction in (9) satisfy the descent condition with any line search is used, i.e.

$$g_k^T d_k = -\|g_k\|^2 < 0. \tag{20}$$

Moreover, if the objective function is a convex quadratic function and the step size is exact line search, then we have $\theta_{k-1} = 0$, $\bar{z}_{k-1} = y_{k-1}$ and $g_k^T s_{k-1} = 0$. In this case, the new parameter β_k^* reduces to the β_k^{HS} formula within the framework of linear conjugate gradient methods.

After this, we need to find that β_k^* satisfied a descent direction. This requires that

$$d_k^T g_k = -\|g_k\|^2 + \beta_k^* d_{k-1}^T g_k < 0. \tag{21}$$

The right hand side of equation (21) can be written as

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

$$= -\|g_k\|^2 + \beta_k^* d_{k-1}^T y_{k-1} + \beta_k^* d_{k-1}^T g_{k-1}.$$

By using descent direction (20) in the above inequality, yields

$$-\|g_k\|^2 + \beta_k^* d_{k-1}^T g_{k-1} \le -\|g_k\|^2 + \beta_k^* d_{k-1}^T y_{k-1}, \tag{22}$$

equation (21) can be rewritten by

$$d_k^T g_k \le -\|g_k\|^2 + \beta_k^* d_{k-1}^T y_{k-1} \tag{23}$$

Therefore, the non-positively of descent direction is sufficient to show that condition (21) is satisfied. This reduces to the condition

$$\|g_k\|^2 \ge \beta_k^* d_{k-1}^T y_{k-1}. \tag{24}$$

Algorithm (I):-

1. Choose an initial point $x_0 \in R^n, \varepsilon > 0, d_k = -g_k$, set k=0.
2. If $\|g_k\| \le \varepsilon$, stop; otherwise, go to the next.
3. Determine a stepsize α_k by some line search rule.
4. Let $x_{k+1} = x_k + \alpha_k d_k$, and compute f_{k+1}, g_{k+1}.
5. Compute the descent direction d_k by (9), (10) and (19).

6. If $\|g_{k+1}\| \le \varepsilon$, stop; otherwise, go to the next
7. Set k=k+1 and go to 3.
8.

Convergence Analysis

To establish convergence properties of this method, it is usually required that the step size $\alpha_k > 0$ should satisfy some conditions, one of them is the weak Wolfe condition (WWC) [18] which is defined by:

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k g_k^T d_k , \tag{25}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k . \tag{26}$$

The other one is the strong Wolfe conditions (SWC), satisfying (25) and

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \le -\sigma g_k^T d_k , \tag{27}$$

where $0 < \delta < \sigma < 1$. [19] showed that the conjugate gradient method are globally convergent when they generalized , the absolute value in (27) is replaced by pair of inequalities:

$$\sigma_1 g_k^T d_k \le g(x_k + \alpha_k d_k)^T d_k \le -\sigma_2 g_k^T d_k , \tag{28}$$

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 [20].

Another one is the Wolfe type line search which is proposed by [21]

$$f(x_k + \alpha_k d_k) - f(x_k) \le -\delta \alpha_k^2 \|d_k\|^2 , \tag{29}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge -2\sigma \alpha_k \|d_k\|^2 , \tag{30}$$

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

Also, we need the following assumptions for the objective function f.

Assumption (1):

(i) The level set $\Psi = \{x \in R^n \,/\, f(x) \le f(x_0)\}$ is bounded, and f(x) is bounded in Ψ.

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

$$\|g(x) - g(y)\| < L\|x - y\| \quad , \ \forall x, y \in N. \tag{31}$$

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

In order to guarantee the global convergence of Algorithm (I), we will impose that steplength α_k satisfies the SWC in (21) and the special case of (24), also we will used Wolfe type line search in (25) and (26) in other one.

At first, we give the following useful theorem which essentially proved by [22] and by [18, 23], which are useful in showing the global convergence of any conjugate gradient method with inexact line search.

Theorem (1)

Suppose that Assumption (I) holds. Any conjugate gradient method of the form (2), where d_k is a descent search direction and α_k satisfies the line search. Then the following holds

$$\sum_{k=1}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} < +\infty , \tag{32}$$

or equivalently

$$\sum_{k=1}^{\infty} \|g_k\|^2 \cos^2 \phi_k < +\infty$$

where ϕ_k is the angle between the search direction d_k and $-g_k$.

In following theorem we are going to prove that Algorithm (I) are satisfy Zoutendijk condition, i.e. it satisfy (32).

Theorem (2)

Suppose that Assumption (1) holds. Let the sequence $\{x_k\}$ be generated by Algorithm (I), and α_k is determined by any line searches. Then, our algorithm either terminates at a stationary point or converges in sense that

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

Then $\lim\limits_{k\to\infty} \inf \|g_k\| = 0$ holds.

Proof:

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

$$\|g_k\| > \bar{\gamma}, \ \forall k \geq 0. \tag{33}$$

We get from (9) that

$$\|d_k\|^2 = d_k^T d_k$$
$$= \left(-M_k g_k + \beta_k^* d_{k-1}\right)^T \left(-M_k g_k + \beta_k^* d_{k-1}\right),$$

where $M_k = 1 + \beta_k^* \dfrac{g_k^T d_{k-1}}{\|g_k\|^2}$

$$\|d_k\|^2 = M_k^2 \|g_k\|^2 - 2M_k \beta_k^* g_k^T d_{k-1} + \left(\beta_k^*\right)^2 \|d_{k-1}\|^2$$
$$= \left(\beta_k^*\right)^2 \|d_{k-1}\|^2 - 2M_k g_k^T d_k - M_k^2 \|g_k\|^2.$$

Dividing both sides by $\left(g_k^T d_k\right)$, we get

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

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

By (3)

$$\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} \leq \frac{\left(\beta_k^*\right)^2 \left(g_{k-1}^T d_{k-1}\right)^2}{\left(-M_k \|g_k\| + \beta_k^* g_k^T d_{k-1}\right)^2} \frac{\|d_{k-1}\|^2}{\left(g_{k-1}^T d_{k-1}\right)^2} + \frac{1}{\|g_k\|^2}. \tag{34}$$

Using (21), $g_k^T d_k < 0$ and (24), we have

$$\left(-M_k \|g_k\|^2 + \beta_k^* g_k^T d_{k-1} + \beta_k^* g_{k-1}^T d_{k-1}\right)\left(-M_k \|g_k\|^2 + \beta_k^* g_k^T d_{k-1} - \beta_k^* g_{k-1}^T d_{k-1}\right) \geq 0$$

Thus,

$$\frac{\left(\beta_k^*\right)^2\left(g_{k-1}^T d_{k-1}\right)^2}{\left(-M_k\|g_k\|^2+\beta_k^* g_k^T d_{k-1}\right)^2}\leq 1. \tag{35}$$

From (34) and (35), we get the following inequality:

$$\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2}\leq\frac{\|d_{k-1}\|^2}{\left(g_{k-1}^T d_{k-1}\right)^2}+\frac{1}{\|g_k\|^2}. \tag{36}$$

By noting $\frac{\|d_0\|^2}{\left(g_0^T d_0\right)^2}=\frac{1}{\|g_0\|^2}$, equation (36) yields that

$$\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2}\leq\sum_{i=0}^{k}\frac{1}{\|g_i\|^2} \text{ for all k.} \tag{37}$$

Therefore, it follows from (33) and (37) that

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

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

This contradiction to the Zoutendijk condition (32), so the proof is complete.

The above theorem show that the new proposed method is independent to any line search is descent and global convergent.

We are going to prove the global convergence of Algorithm (I) by using SWC (21) and (24).

Theorem (3)

Suppose that Assumption (1) holds. Consider any conjugate gradient method in the form (2), (9) and (10), where d_k is descent and α_k is determined by SWC (21) and (24). If

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

then $\underset{k\to\infty}{Lim}\inf\|g_k\|=0$ holds.

Proof:

Note that from (19) we have $0<\eta_{k-1}\leq\tau$, since

$$M_1\leq\left[\ln\left(\frac{1+f_{k-1}}{f_{k-1}}\right)\right]^{\frac{1}{2}}\|f_{k-1}\|\|1+f_{k-1}\|=\tau_1,$$

and $M_2\leq\left[\ln\left(\frac{1+f_k}{f_k}\right)\right]^{\frac{1}{2}}\|f_k\|\|1+f_k\|=\tau_2$.

We have $\alpha_{k-1}>0$, we can fined ϖ *and* ω which are positive numbers such that $\varpi\leq\alpha_{k-1}\leq\omega$, and we know that

$$\|d_{k-1}\|=\|s_{k-1}\|/\alpha_{k-1}\leq B/\varpi=E,$$

and

$$w \leq \frac{\alpha_{k-1}\|g_{k-1}\|\|d_{k-1}\|}{4} \leq \frac{\omega\gamma E}{4} = \tau_3,$$

Therefore, $0 < M_1 \leq \tau_1$, $0 < M_2 \leq \tau_2$ and $0 < w \leq \tau_3$, let $\tau = \frac{\tau_1 + \tau_3}{\tau_2}$

From (6) and (20), Babaie-Kafaki [1] proved that

$$|\theta_{k-1}| \leq 3L\|s_{k-1}\|^2 \text{ , and } \|\bar{z}_{k-1}\| \leq (1+3\eta)L\|s_{k-1}\|$$

The above inequality with $\eta \leq \tau$, its yield

$$|\theta_{k-1}| \leq 3L\|s_{k-1}\|^2 \text{ , and } \|\bar{z}_{k-1}\| \leq (1+3\tau)L\|s_{k-1}\| = \kappa\|s_{k-1}\|, \quad (38)$$

where $\kappa = (1+3\tau)L$.

We have, $\bar{z}_{k-1}^T d_{k-1} \geq y_{k-1}^T d_{k-1} \geq (\sigma - 1) g_{k-1}^T d_{k-1}$. (39)

Put (20) in the above inequality, we get

$$\bar{z}_{k-1}^T d_{k-1} \geq (1-\sigma)\|g_{k-1}\|^2. \quad (40)$$

From (10), (38) and (39) , yields

$$\left|\beta_k^*\right| \leq \frac{\kappa\|g_k\|\|s_{k-1}\|}{(1-\sigma)\|g_{k-1}\|^2}. \quad (41)$$

Take the norm of the both sides of equation (9), and used (41), we get

$$\|d_k\| \leq \|g_k\| + 2\frac{\kappa\|g_k\|\|s_{k-1}\|}{(1-\sigma)\|g_{k-1}\|^2}\|d_{k-1}\|$$

$$\leq \left(1 + 2\frac{\kappa\alpha_{k-1}\|d_{k-1}\|^2}{(1-\sigma)\|g_{k-1}\|^2}\right)\|g_k\|$$

$$\leq \left(1 + 2\frac{\kappa\omega E^2}{(1-\sigma)\bar{\gamma}^2}\right)\gamma = \mu\gamma.$$

Where $\mu = 1 + 2\frac{\kappa\omega E^2}{(1-\sigma)\bar{\gamma}^2}$.

So we get, $\sum_{k\geq 1}\frac{1}{\|d_k\|^2} \geq \frac{1}{\mu\gamma}\sum_{k\geq 1} 1 = \infty$.

Therefore, $\lim_{k\to\infty}\inf\|g_k\| = 0$. The proof is complete.

Next, we present a Lemma which shows that, asymptotically, the search directions change slowly.

Lemma (4)

Suppose that Assumption (1) hold. Let $\{x_k\}$ *and* $\{d_k\}$ be generated by Algorithm (I) and α_k is obtained by SWC (21) and (24), and

$$\sum_{k\geq 1}\|w_k - w_{k-1}\|^2 < \infty, \quad (42)$$

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

Proof

Firstly, note that $d_k \neq 0$, for otherwise (20), would imply $g_k = 0$. Therefore, w_k is well defined. Now, let us define

$$u_k = \frac{v_k}{\|d_k\|}, \quad r_k = \beta_k^* \frac{\|d_{k-1}\|}{\|d_k\|} = \frac{g_k^T \bar{z}_{k-1}}{\bar{z}_{k-1}^T d_{k-1}} \frac{\|d_{k-1}\|}{\|d_k\|}. \tag{43}$$

Where $v_k = -\left(1 + \frac{g_k^T \bar{z}_{k-1}}{\bar{z}_{k-1}^T d_{k-1}} \frac{g_k^T d_{k-1}}{\|g_k\|^2}\right) g_k$. (44)

Therefore, from (9), for $k \geq 1$, we obtain

$$w_k = u_k + r_k w_{k-1}.$$

Using this relation with the identity $\|w_k\| = \|w_{k-1}\| = 1$, we have

$$\|u_k\| = \|w_k - r_k w_{k-1}\| = \|r_k w_k - w_{k-1}\|.$$

In addition, using this with the condition $t_k \geq 0$ and the triangular inequality, we have

$$\|w_k - w_{k-1}\| \leq \|w_k - r_k w_{k-1}\| + \|r_k w_k - w_{k-1}\|$$

so

$$\|w_k - w_{k-1}\| = 2\|u_k\|. \tag{45}$$

Now, taking the norm of both sides of equation (44), and used (41)

$$\|v_k\| = \left\| -\left(1 + \frac{g_k^T \bar{z}_{k-1}}{\bar{z}_{k-1}^T d_{k-1}} \frac{g_k^T d_{k-1}}{\|g_k\|^2}\right) g_k \right\|,$$

$$\leq \left(1 + \frac{\delta\|g_k\| \|s_{k-1}\|}{(1-\sigma)\|g_k\|^2} \frac{\|g_k\| \|d_{k-1}\|}{\|g_k\|^2}\right) \|g_k\|,$$

$$\leq \left(1 + \frac{\delta \omega E^2}{(1-\sigma)\bar{\gamma}^2}\right) \gamma \equiv D.$$

From (45)

$$\|w_k - w_{k-1}\| \leq 2 \frac{D}{\|d_k\|},$$

Now squaring the both sides of the above inequality, and taking the summation we get

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

Which complete the proof.

Next, we present a lemma which shows that β_k^* will be small when the step s_{k-1} is small which implies that the Algorithm (I) prevents the inefficient behavior of the jamming phenomenon from occurring [24].

Lemma (5)

Suppose that Assumption (1) hold. Let $\{x_k\}$ *and* $\{d_k\}$ be generated by Algorithm (I), if there exists a positive constant $\bar{\gamma} > 0$ such that equation (33) holds; then there exists constants b>1 and $\zeta > 0$ such that for all k:

$$|\beta_k^*| \leq b, \tag{46}$$

and

$$\|s_{k-1}\| \leq \zeta \Rightarrow |\beta_k^*| \leq \frac{1}{b}. \tag{47}$$

Proof

Using equations (20), (33) and (41), we obtain the following:

$$\left|\beta_k^*\right| \leq \left(\frac{\delta\|g_k\|}{(1-\sigma)\|g_k\|^2}\right)\|s_{k-1}\|$$

$$\left|\beta_k^*\right| \leq \left(\frac{\delta\gamma}{(1-\sigma)\bar{\gamma}^2}\right)B \equiv b \quad . \tag{48}$$

Note that b can be defined such that b>1 without loss generality. Now, we define

$$\zeta = \left(\frac{(1-\sigma)\bar{\gamma}^2}{\delta\gamma}\right)^2 \frac{1}{b}. \tag{49}$$

If $\|s_{k-1}\| \leq \zeta$, we have from inequality (48) that

$$\left|\beta_k^*\right| \leq \left(\frac{\delta\gamma}{(1-\sigma)\bar{\gamma}^2}\right)\zeta \equiv \frac{1}{b}.$$

Therefore, for the b and ζ in (48) and (49), relations (46) and (47) holds.

Now, we are going to prove the global convergence of Algorithm (I) by using Wolfe type line search (29) and (30).

Theorem (6)

Suppose that Assumption (1) holds, Consider any conjugate gradient method in the form (2), (9) and (10), where d_k is descent and α_k is determined by Wolfe type line search (29) and (30), if

$$\sum_{k=1}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} < +\infty, \tag{50}$$

then the following holds

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

Proof

From (29), (30), and Assumption (1), we obtain

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

Then, we know that

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

Squaring both sides of the above formula, we get

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

Taking the summations of the both sides for the above inequalities, we get

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

By (29), we have

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

Therefore, $\lim_{k\to\infty}\inf\|g_k\| = 0$ .The proof is complete.

Remark: Motivated by the above modified HS conjugate gradient method, we use similarly technique to classical algorithm, we get other modified formulas as follows:

$$d_k^{MPR} = -\left(1+\frac{g_k^T \bar{z}_{k-1}}{\|g_{k-1}\|^2}\frac{g_k^T d_{k-1}}{\|g_k\|^2}\right)g_k + \frac{g_k^T \bar{z}_{k-1}}{\|g_{k-1}\|^2}d_{k-1},$$

$$d_k^{MLS} = -\left(1-\frac{g_k^T \bar{z}_{k-1}}{d_{k-1}^T g_{k-1}}\frac{g_k^T d_{k-1}}{\|g_k\|^2}\right)g_k - \frac{g_k^T \bar{z}_{k-1}}{d_{k-1}^T g_{k-1}}d_{k-1},$$

$$d_k^{MDY} = -\left(1+\frac{\|g_k\|^2}{d_{k-1}^T \bar{z}_{k-1}}\frac{g_k^T d_{k-1}}{\|g_k\|^2}\right)g_k - \frac{\|g_k\|^2}{d_{k-1}^T \bar{z}_{k-1}}d_{k-1},$$

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

where $\beta_k^{MHZ} = \frac{g_k^T \bar{z}_{k-1}}{d_{k-1}^T \bar{z}_{k-1}} - 2\frac{\|\bar{z}_{k-1}\|^2}{d_{k-1}^T \bar{z}_{k-1}}\frac{g_k^T d_{k-1}}{d_{k-1}^T \bar{z}_{k-1}}$.

Note, we give the above equations ($d_k^{MPR}, d_k^{MLS}, d_k^{MDY}\, and\, d_k^{MHZ}$) the number (51) and $\bar{z}$ is defined in (8b).

Theorem (7)

Let scalar β_k of equation (9) be replaced by other conjugate gradient equation, respectively. Moreover, if d_k of equation (9) is replaced by the above formulas of equation (51), then we have the descent property, i.e.

$$g_k^T d_k = -\|g_k\|^2 < 0.$$

If the conditions of Theorems (2)- (6) holds. Then the corresponding algorithm satisfy global convergence, i.e.

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

too.

4. Numerical results

In this section, we compare the performance of new modified HS conjugate gradient method to other conjugate gradient methods. Especially, we will compare this method with well-known routine HS, and modified Dai-Liao which is introduced by [11] using the same test functions conjugate gradient methods. The cods were written in FORTRAN and in double precision arithmetic. We selected some of large-scale unconstrained optimization test problems. For each test function we considered three experiments with the number of variables 1000, 10000 and 100000, respectively. The test problems are the unconstrained problems in CUTE [25] library, along with other large-scale optimization test problems in [26].

In table 1, we give the comparison between the proposed method, the Livieris [11] and the standard Hestenes-Stiefel conjugate gradient method [2]. We used in this table the SWC (21) and (24) with parameters $\delta = 0.001\, and\, \sigma = 0.9$. While in able 2, we give the comparison between the proposed method, the Livieris [11] and the standard Hestenes-Stiefel conjugate gradient method [2]. We used in this table the Wolfe type line searches (29) and (30) with parameters $\delta = 1*E-5$ *and* $\sigma = 0.01$.for each test problem, the termination criterion is:

$$\|g_{k+1}\| < 1\times 10^{-6} \quad . \tag{52}$$

We also force these routines stooped if the number of iterations exceed (600) without achieving convergence.

We record the number of iterations calls (NOI), and the number of function and gradient evaluations calls (NOFG), for purpose of our comparisons. We can see all these results reported in Table (1) with SWC. While in Table (2) we reported our comparisons with Wolfe Type Line Search.

Note that,

1- The symbol F means that the algorithm is fail to converges.
2- The symbol F^* means that the number of iteration exceed (600) without achieving convergence.
3- We denote the new proposed method by MHS, the method which proposed method in [11] by MDL, and Hestenes-Stiefel by HS conjugate gradient methods.
4- We shaded the better results of our proposed method in below tables.

Table 1- the comparison between MHS, MDL and HS with SWC

Test Functions	Dimensions	MHS		MDL		HS	
		NOI	NOFG	NOI	NOFG	NOI	NOFG
Ex. Powell	1000	41	119	45	139	46	145
	10000	41	119	45	139	49	164
	100000	44	140	60	201	55	184
DENSCHNF	1000	22	47	23	49	F	
	10000	21	44	23	49	F	
	100000	22	47	F		F	
Ex. Wood	1000	30	68	31	69	33	73
	10000	30	69	35	78	34	75
	100000	32	72	34	77	34	75
Dixmaana (Cute)	1000	5	13	6	15	5	14
	10000	5	13	6	15	5	14
	100000	5	13	6	15	6	14
Dixmaanc (Cute)	1000	6	15	7	17	6	16
	10000	6	15	7	17	6	16
	100000	6	15	7	17	6	16
NOND	1000	27	65	26	63	26	63
	10000	27	67	27	67	26	67
	100000	27	67	27	67	27	67
Ex. Block Diagonal - 1	1000	18	39	33	68	23	48
	10000	22	46	22	46	26	54
	100000	23	49	24	51	28	58
ENGVAL- 1 (Cute)	1000	21	46	21	46	23	83
	10000	25	88	25	88	26	230
	100000	F		F		F	
Strait	1000	7	18	8	19	8	19
	10000	7	18	8	19	8	19
	100000	7	18	8	19	8	19
Wolfe	1000	68	137	79	159	79	159
	10000	121	243	159	276	105	214

	100000	112	228	114	232	114	232
Sum	1000	18	88	19	97	19	99
	10000	35	138	31	125	31	125
	100000	76	340	71	321	72	377
Ex. Freudenstein and Roth	1000	9	22	10	24	9	23
	10000	9	22	10	24	F	
	100000	9	21	F		F	
Shallow	1000	9	24	10	25	10	25
	10000	9	24	10	25	11	28
	100000	10	23	11	27	11	27
NONDIA (Shanno-78)	1000	7	18	8	22	F	
	10000	6	20	8	21	9	25
	100000	5	18	6	20	9	25
Miele	1000	69	290	75	301	54	209
	10000	60	247	89	383	61	249
	100000	62	263	103	467	65	271
Gcentrel	1000	48	214	50	220	48	214
	10000	50	219	60	244	52	223
	100000	53	225	93	610	54	227

Table 2- the comparison between MHS, MDL and HS with Wolfe Type Line Search.

Test Functions	Dimensions	MHS		MDL		HS	
		NOI	NOFG	NOI	NOFG	NOI	NOFG
Ex. Powell	1000	41	119	45	139	46	145
	10000	41	119	60	200	49	164
	100000	44	140	60	200	55	184
DENSCHNF	1000	22	47	23	49	23	51
	10000	21	44	23	49	25	55
	100000	22	47	25	55	26	57
Ex. Wood	1000	30	68	31	69	33	73
	10000	30	69	35	78	34	75
	100000	32	72	34	77	34	75
Dixmaana (Cute)	1000	5	14	6	16	6	16
	10000	5	14	6	16	5	14
	100000	5	14	7	17	6	16
Dixmaanc (Cute)	1000	6	15	7	17	6	16
	10000	6	15	7	17	6	16
	100000	6	15	7	17	6	16
NOND	1000	24	57	27	63	28	66
	10000	24	57	27	67	28	66
	100000	24	57	27	67	28	69
Ex. Block Diagonal - 1	1000	20	42	26	53	33	67
	10000	18	38	25	51	33	67
	100000	21	44	32	65	35	71
ENGVAL- 1 (Cute)	1000	24	51	22	48	37	77
	10000	51	104	99	308	F*	
	100000	F*		F*		F*	
Strait	1000	7	18	8	19	8	19
	10000	7	18	8	19	8	19
	100000	7	18	8	19	8	19
Wolfe	1000	68	137	79	159	79	159
	10000	121	243	159	276	105	214

	100000	112	228	114	232	114	232
Sum	1000	28	158	19	97	19	99
	10000	38	203	45	275	37	203
	100000	65	334	89	393	68	312
Ex. Freudenstein and Roth	1000	8	20	9	22	9	22
	10000	8	20	9	22	11	24
	100000	10	23	11	25	11	25
Shallow	1000	9	21	10	25	10	25
	10000	10	23	11	26	12	28
	100000	10	23	11	27	12	27
NONDIA (Shanno-78)	1000	7	18	8	22	7	20
	10000	6	20	8	21	7	20
	100000	5	18	8	22	9	25
Miele	1000	69	290	75	301	54	209
	10000	60	247	89	383	61	249
	100000	62	263	103	467	65	271
Gcentrel	1000	46	209	55	238	51	228
	10000	47	212	45	210	52	231
	100000	50	223	67	263	56	241

References:

1. Babaie-Kafaki, S., Ghanbari, R., and Mahdavi-Amiri, N. **2010**. Two new conjugate gradient methods based on modified secant equations, *Journal of Computational and Applied Mathematics*, 234, pp:1374-1386.
2. Hestenes, M.R. and Stiefel**,** E. **1952**. Methods of conjugate gradients for solving linear systems, *Journal of Research of the National Bureau of Standards* , 5, pp:409-436.
3. Fletcher, R. and Reeves, C.M. **1964.** Function minimization by conjugate gradients, *Computer Journal*., 7, pp:149-154.
4. Polak, E., and Ribière, G. **1969**. Not sur la convergence de directions conjugue'e. *Rev. Franaise Informant Researche operationelle,* 3e Anne'e., 16, pp:35-43.
5. Liu, D. and Story, C**.** **1991**. Efficient generalized conjugate gradient algorithms., part l : *Theory, Journal of Optimization Theory and Applications*, 69, pp;129-137.
6. Dai, Y. H. and Yuan, Y. **1999**. A nonlinear conjugate gradient method with a strong global convergence property, *SIAM Journal of Optimization*., 10, pp:177-182**.**
7. Fletcher, R., **1987**, *Practical Methods of Optimization*, Second Edition, A Wiley-Interscience Publication, John Wiley & Sons, Inc., NY, USA.
8. Nocedal, J. and Wright, S.J. **1999**, *Numerical Optimization*, *Springer Series in Operations Research,* Second Edition, Springer-Verlag, New York, USA.
9. Zhang, J.Z., Deng, N.Y., Chen, L.H. **1999**. New quasi-Newton equation and related methods for unconstrained optimization., *Journal of Optimization Theory and Appl*ications, 102, pp:147-167.
10. Zhang, J.Z. and Xu, C.X., **2001**, Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations, *Journal of Computational and Applied Mathematics*, 137, pp:269-278 .
11. Livieris, I.E., and Pintelas, P.**2012**, A descent Dai-Liao conjugate gradient method based on a modified secant equation and its global *convergence. ISRN Comp. Math.*, 2012, Article ID 43495, 1-8. DOI:10.5402/2012/435495.
12. Zhang, L., Zhou, W. and Li, D. H. **2011**, Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line-search, *Numerical Mathmatics*,104, pp:561-572. DOI:10.1007/s0021-006-0028-z..
13. Gilbert, J. C. and Nocedal, J**.** **1992**. Global convergence Properties of conjugate gradient methods for optimization, *SIAM Journal of Optimization*, 2, pp:21-42.
14. Boland, W., Kamgnia, E., and Kowallik, J**.**, **1979**. A conjugate gradient optimization method invariant to non-linear scaling. *Journal of Optimization Theory and Appl*ications**,** 27, pp:221-230.

15. Fried, I. **1971**. N-step conjugate gradient minimization scheme for non-quadratic functions. *AIAA Journal*,7, pp:2286-2287. DOI:10.2514/3.6507.

16. Tassopoulos, A. and Story, C. **1984**. A conjugate gradient direction method based on non-quadratic model., *Journal of Optimization Theory and Appli*cations., 43, pp:371-381. DOI:10.10007/BF00934461.

17. Al-Bayati**, A.Y.**, **1993**. A new non-quadratic model for unconstrained non-linear optimization. *Mu'tah J. Stud. Nat. Applied Science. Series*, 8, pp:131-155.

18. Wolfe, P.**, 1969**. Convergence conditions for ascent methods II: some corrections. *SIAM Review*, 13, pp:185-188**.**

19. Dai, Y. H. and Yuan, Y**. 1996.** Convergence properties of the conjugate descent method. *Advance Mathematics Journal*, 25 , pp:552-562.

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

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

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

23. Wolfe, P. **1969**. Convergence conditions for ascent methods. *SIAM Review*, 11**,** pp:226-235**.**

24. Powell, M.J.D**. 1977**. Restart procedure for the conjugate gradient methods, *Mathematics programming*, 12, pp: 241-254.

25. Bongratz, I., 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.

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