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.

Introduction
Consider the following n-dimensional unconstrained minimization problem

\[
\text{Minimize } f(x), \quad x \in \mathbb{R}^n
\]

where \( f : \mathbb{R}^n \to \mathbb{R} \) 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:

\[
x_k = x_{k-1} + \alpha_{k-1}d_{k-1},
\]

where

\[
\alpha_k = \frac{g_k^T s_k}{d_k^T s_k}, \quad \text{with } s_k = r_k + \alpha_{k-1}d_{k-1}, \quad d_k = -g_k + \beta_k d_{k-1},
\]

\[
\beta_k = \frac{\frac{\|r_k\|^2}{\|r_{k-1}\|^2}}{\frac{\|r_k\|^2}{\|r_{k-1}\|^2}}.
\]

*Email: ghalnaemi@yahoo.fr
d_k = \begin{cases} -g_k & \text{if } k = 0 \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geq 1 \end{cases} \quad (3)

where \( x_k \in \mathbb{R}^n \) is the k-th approximation to the solution of (1), \( \alpha_k \) is a positive step size which is determined by some line search direction, \( d_k \) is a search direction, and \( \beta_k \in \mathbb{R} \) 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 \( \alpha_k \) is calculated by the exact line search. Several kinds of formulas for \( \beta_k \) has been proposed. For example [2, 3, 4, 5, 6, 7] formulas are well known and they are given by:

\[
\beta_{k}^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \quad \beta_{k}^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \beta_{k}^{PR} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \quad \beta_{k}^{LS} = -\frac{g_k^T y_{k-1}}{d_{k-1}^T g_{k-1}}, \quad \beta_{k}^{DV} = \frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}}, \quad \beta_{k}^{CD} = -\frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}}.
\]

Respectively, where \( y_{k-1} = g_k - g_{k-1}, \quad s_{k-1} = x_k - x_{k-1} = \alpha_{k-1} d_{k-1} \) and \( \| \cdot \| \) 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 \( \beta_k \) 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:

\[
B_k s_{k-1} = y_{k-1}. \quad (4)
\]

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

\[
B_k s_{k-1} = z_{k-1}, \quad z_{k-1} = y_{k-1} + \eta_{k-1} \frac{\theta_{k-1}}{s_{k-1}^T u} u,
\]

and \( \theta_{k-1} \) is defined by

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

where \( u \in \mathbb{R}^n \) is a vector parameter satisfying \( s_{k-1}^T u \neq 0 \). 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:

\[
d_k = \left[ 1 + \max \left\{ \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, 0 \right\} \right]^{-1} \left[ \frac{g_k^T s_{k-1}}{d_{k-1}^T s_{k-1}} \right] g_k + \beta_k^* d_{k-1},
\]

and,

\[ y_{k-1}^* = y_{k-1} + \eta_{k-1} \left( \frac{\max[\theta_{k-1}, 0]}{s_{k-1}^T u} \right) u, \]  

(7)

where \( \theta_{k-1} \) is defined in (6), \( u \) is any vector satisfying \( s_{k-1}^T u > 0 \), the parameter \( \eta_{k-1} \) 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:

\[ B_k s_{k-1} = \tilde{z}_{k-1}, \]  

(8a)

and

\[ \tilde{z}_{k-1} = y_{k-1} + \eta_{k-1} \left( \frac{\max[\theta_{k-1}, 0]}{s_{k-1}^T u} \right) u, \]  

(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 \( \|s_{k-1}\| \leq 1 \) 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 \( \beta_k \geq 0 \) [6, 13]. So, we are going to propose a new conjugate gradient method as follows, let the search direction be defined by:

\[ d_k = -\left(1 + \beta_k^{MHS} g_k^T d_{k-1} / \|g_k\|^2\right) g_k + \beta_k^{MHS} d_{k-1}, \]  

(9)

where \( \beta_k^{MHS} \) is the propose modification of formula \( \beta_k^{HS} \) as follows:

\[ \beta_k^{MHS} = \beta_k^{*} = \frac{g_k^T \tilde{z}_{k-1}}{\tilde{z}_{k-1} d_{k-1}} \]  

(10)

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

In this paper we will going to evaluate \( \eta_{k-1} \) 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 
\[
f(x)=q(q(x))^3, \quad p>0 \text{ and } x \in \mathbb{R}^n,
\]
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:
  \[
  F(q(x)) = \epsilon_1 q(x) + \frac{1}{2} \epsilon_2 q^2(x), \quad \text{where } \epsilon_1 \text{ and } \epsilon_2 \text{ are positive scalar.}
  \]

- Two rational models which have been investigated by [16], the first one was
  \[
  F(q(x)) = \frac{(\epsilon_1 q(x) + 1)}{\epsilon_2 q(x)}, \quad \epsilon_1 > 0, \quad \epsilon_2 < 0, \quad \text{where } \epsilon_1 \text{ and } \epsilon_2 \text{ are scalar.}
  \]

The second one was
\[
F(q(x)) = \frac{\epsilon_1 q(x)}{(q(x) + 1)}, \quad \epsilon > 0.
\]

- Another rational model which has been investigated by [17], which defined by:
  \[
  F(q(x)) = \frac{\epsilon_1 q(x)}{(1 - \epsilon_2 q(x))}, \quad \epsilon_1 > 0, \epsilon_2 < 0.
  \]

In this study, we consider a new rational model, define by:
\[
F(q(x)) = \frac{1}{\epsilon q^*(x) - 1}, \quad q(x) > 0 \text{ and } \frac{df}{dq} > 0.
\]

The unknown quantities \( \eta_{k-1} \) were expressed in term of available quantities of algorithm (i.e. function gradient value of the objective function) using the expression for \( \eta_{k-1} \)
\[
\eta_{k-1} = \frac{f_{k-1}}{f_k},
\]
from the relations
\[
g_k = f_k G(x_k - x^*) \quad (13)
\]
\[
g_{k-1} = f_{k-1} G(x_{k-1} - x^*) \quad (14)
\]
Where \( G \) is the Hessian matrix and \( x^* \) is the minimum point, so from (12) we get:
\[
\eta_{k-1} = \frac{g_k^T(x_k - x^*)}{g_k^T(x_k - x^*)},
\]
but, 
\[
g_k^T(x_k - x^*) = g_{k-1}^T(x_{k-1} + \alpha_k d_{k-1} - x^*) = g_{k-1}^T(x_{k-1} - x^*) + \alpha_k g_{k-1}^T d_{k-1},
\]
and,
\[
g_k^T(x_k - x^*) = g_k^T(x_k - \alpha_k d_{k-1} - x^*) = g_k^T(x_k - x^*).
\]
Since \( g_k^T d_{k-1} = 0 \), therefore, we can express \( \eta_{k-1} \)
\[
\eta_{k-1} = \frac{g_k^T(x_{k-1} - x^*) + \alpha_{k-1} g_{k-1}^T d_{k-1}}{g_k^T(x_k - x^*)}.
\]
From (13), (14) and (15), we get:
\[
\eta_{k-1} = \frac{f_{k-1}^T(x_{k-1} - x^*) G(x_{k-1} - x^*) + \alpha_{k-1} g_{k-1}^T d_{k-1}}{f_k^T(x_k - x^*) G(x_k - x^*)},
\]
therefore
\[
\eta_{k-1} = \frac{2 f_{k-1}^T q_{k-1} + \alpha_{k-1} g_{k-1}^T d_{k-1}}{2 f_k^T q_k}.
\]
If we calculate $f_{k-1}'$ and $f_k'$ from (11), we get:

$$f_{k-1}' = 2q_{k-1}f_{k-1}(1 + f_{k-1}), \text{ and } f_k' = 2q_kf_k(1 + f_k).$$

(17)

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

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

(18)

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

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

(19)

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

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

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

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

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

(20)

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

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

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

(21)

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

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

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

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

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

(22)

equation (21) can be rewritten by

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

(23)

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

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

(24)

**Algorithm (1):**

1. Choose an initial point $x_0 \in R^n, \varepsilon > 0, d_k = -g_k$, set $k = 0$.
2. If $\|g_k\| \leq \varepsilon$, stop; otherwise, go to the next.
3. Determine a stepsize $\alpha_k$ 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 \( \| g_{k+1} \| \leq \varepsilon \), 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:

\[
\begin{align*}
\mathbf{f}(x_k + \alpha_k d_k) - \mathbf{f}(x_k) & \leq \delta \alpha_k \mathbf{g}_k^T d_k , \\
g(x_k + \alpha_k d_k)^T d_k & \geq \sigma \mathbf{g}_k^T d_k .
\end{align*}
\]

(25)
(26)

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

\[
\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq -\sigma \mathbf{g}_k^T d_k ,
\]

(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:

\[
\begin{align*}
\sigma_1 \mathbf{g}_k^T d_k & \leq g(x_k + \alpha_k d_k)^T d_k \leq -\sigma_2 \mathbf{g}_k^T d_k ,
\end{align*}
\]

(28)

where \( \sigma_1 \geq 0, 0 \leq \delta < \sigma_1 < 1, 0 < \sigma_2 < 1, \sigma_1 + \sigma_2 \leq 1 \). 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]

\[
\begin{align*}
f(x_k + \alpha_k d_k) - f(x_k) & \leq -\delta \alpha_k^2 \| d_k \|^2 , \\
g(x_k + \alpha_k d_k)^T d_k & \geq -2\sigma \alpha_k \| d_k \|^2 ,
\end{align*}
\]

(29)
(30)

where \( 0 < \delta < \sigma < 1 \).

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

Assumption (1):

(i) The level set \( \Psi = \{ x \in \mathbb{R}^n \mid f(x) \leq f(x_0) \} \) is bounded, and \( f(x) \) is bounded in \( \Psi \).

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

\[
\| g(x) - g(y) \| \leq L \| x - y \| , \quad \forall x, y \in N .
\]

(31)

It follows directly from Assumption (1) that there exists a positive constants \( \gamma \) and \( A \) such that

\[
\| g_k \| \leq \gamma , \quad \forall x \in \Psi , \quad \text{and} \quad \| v_k \| \leq A , \quad \forall x \in \Psi , \quad \text{respectively} .
\]

In order to guarantee the global convergence of Algorithm (I), we will impose that steplength \( \alpha_k \) 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 (1) holds. Any conjugate gradient method of the form (2), where \( d_k \) is a descent search direction and \( \alpha_k \) satisfies the line search. Then the following holds

\[
\sum_{k=1}^{\infty} \frac{\| g_k^T d_k \|^2}{\| d_k \|^2} < +\infty ,
\]

(32)

or equivalently

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

where \( \phi_k \) 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 \( \alpha_k \) is determined by any line searches. Then, our algorithm either terminates at a stationary point or converges in sense that

\[
\sum_{k=1}^{\infty} \left( g_k^T d_k \right)^2 \|d_k\|^2 < +\infty.
\]

Then \( \lim_{k \to \infty} g_k = 0 \) holds.

**Proof:**

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

\[
\|g_k\| > \tilde{\gamma}, \quad \forall k \geq 0. \tag{33}
\]

We get from (9) that

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

where \( M_k = 1 + \beta_k^* \frac{g_k^T d_{k-1}}{\|g_k\|^2} \)

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

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

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

By (3)

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

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

\[
\left( -M_k g_k \right)^2 + \beta_k^* g_k^T d_{k-1} + \beta_k^* s_k^T d_{k-1} \left( -M_k g_k \right)^2 + \beta_k^* g_k^T d_{k-1} - \beta_k^* s_k^T d_{k-1} \right) \geq 0
\]

Thus,
\[
\frac{(\beta_k^*)^2(g_k^T d_k)}{(-M_k s_k^2 + \beta_k^* g_k^T d_k)^2} \leq 1. \tag{35}
\]

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

\[
\frac{\|d_k\|^2}{(s_k^T d_k)^2} \leq \frac{\|d_{k-1}\|^2}{(s_{k-1}^T d_{k-1})^2} + \frac{1}{\|g_k\|^2}. \tag{36}
\]

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

\[
\frac{\|d_k\|^2}{(s_k^T d_k)^2} \leq \sum_{i=1}^{k} \frac{1}{\|s_i\|^2} \text{ for all } k. \tag{37}
\]

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

\[
\frac{(s_k^T d_k)^2}{\|d_k\|^2} \geq \frac{\gamma^2}{k + 1},
\]

\[
\Rightarrow \sum_{k \geq 1} \frac{(s_k^T d_k)^2}{\|d_k\|^2} = +\infty.
\]

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 \(\alpha_k\) is determined by SWC (21) and (24). If

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

then \(\lim \inf_{k \to \infty} \|g_k\| = 0\) holds.

**Proof:**

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

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

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

We have \(\alpha_{k-1} > 0\), we can fined \(\sigma\) and \(\omega\) which are positive numbers such that \(\sigma \leq \alpha_{k-1} \leq \omega\), and we know that

\[
\|d_{k-1}\| = \|s_{k-1}\|/\alpha_{k-1} \leq B/\sigma = E,
\]

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

\[ |\theta_{k-1}| \leq 3L \|s_{k-1}\|^2, \text{ and } \|x_{k-1}\| \leq (1 + 3\eta)L \|s_{k-1}\| \]

The above inequality with \( \eta \leq \tau \), its yield

\[ |\theta_{k-1}| \leq 3L \|s_{k-1}\|^2, \text{ and } \|x_{k-1}\| \leq (1 + 3\tau)L \|s_{k-1}\| = \kappa \|s_{k-1}\|. \] (38)

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

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

Put (20) in the above inequality, we get

\[ \bar{z}_{k-1}^Td_{k-1} \geq (1 - \sigma)\|g_{k-1}\|^2. \] (40)

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

\[ |\beta_k^*| \leq \frac{\kappa \|g_k\| \|s_{k-1}\|}{(1 - \sigma)\|g_{k-1}\|^2}. \] (41)

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

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

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

\[
\leq \left(1 + 2 \frac{\kappa \alpha \|s_{k-1}\|^2}{(1 - \sigma)\|g_{k-1}\|^2}\right)\gamma = \mu \gamma .
\]

Where \( \mu = 1 + 2 \frac{\kappa \alpha \|g_k\|^2}{(1 - \sigma)\gamma^2} \).

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

Therefore, \( \lim_{k \to \infty} \|g_k\| = 0\). 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 \( \alpha_k \) is obtained by SWC (21) and (24), and

\[ \sum_{k \geq 1} \|w_{k} - w_{k-1}\|^2 < \infty , \] (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

$$u_k = \frac{v_k}{d_k}, \quad r_k = \beta_k^* \frac{d_{k-1}}{d_k}, \quad k \leq \frac{g_k^T z_{k-1}}{z_{k-1}^T d_k}.$$  \hspace{1cm} (43)

Where $v_k = \left(1 + \frac{g_k^T z_{k-1} - T d_{k-1}}{g_k^T d_{k-1}} \right) g_k.$  \hspace{1cm} (44)

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

$$w_k = u_k + r_k w_{k-1}.$$  \hspace{1cm} (45)

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

$$\|u_k\| = \|w_k - r_k w_{k-1}\| = \|r_k w_k - w_{k-1}\|.$$  \hspace{1cm} (46)

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

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

which complete the proof.

Next, we present a lemma which shows that $\beta_k^*$ will be small when the step $s_k$ 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 $\gamma > 0$ such that equation (33) holds; then there exists constants $b > 1$ and $\zeta > 0$ such that for all $k$:

$$|\beta_k^*| \leq b,$$  \hspace{1cm} (46)

and

$$|s_{k-1}| \leq \zeta \Rightarrow |\beta_k^*| \leq \frac{1}{b}.$$  \hspace{1cm} (47)
Proof

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

\[
|\beta^*_k| \leq \frac{\delta \|g_k\|}{(1 - \sigma \|g_k\|^2)} \|s_{k-1}\| \\
|\beta^*_k| \leq \frac{\delta \gamma}{(1 - \sigma \gamma^2)} B \equiv b
\]  \hspace{1cm} (48)

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

\[
\zeta = \frac{(1 - \sigma \gamma^2)^2}{\delta \gamma} \frac{1}{b}.
\]  \hspace{1cm} (49)

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

\[
|\beta^*_k| \leq \frac{\delta \gamma}{(1 - \sigma \gamma^2)} \zeta = \frac{1}{b}.
\]

Therefore, for the \(b\) and \(\zeta\) 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 \(\alpha_k\) is determined by Wolfe type line search (29) and (30), if

\[
\sum_{k=1}^{\infty} \left( \frac{g_k^T d_k}{\|d_k\|^2} \right)^2 < + \infty,
\]  \hspace{1cm} (50)

then the following holds

\[
\liminf_{k \to \infty} \|g_k\| = 0.
\]

**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

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

By (29), we have
\[
\sum_{k=1}^{\infty} \left( d_k^T g_k \right)^2 \leq \frac{(2\sigma + L)^2}{\delta} \sum_{k=1}^{\infty} \{ f(x_k) - f(x_{k+1}) \} < +\infty.
\]

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:

\[
d_k^{MPR} = \left( 1 + \frac{g_k^T z_{k-1} g_k^T d_{k-1}}{\| g_k \|^2} \right) g_k + \frac{g_k^T z_{k-1} d_{k-1}}{\| g_k \|^2} d_{k-1}.
\]

\[
d_k^{MLS} = \left( 1 - \frac{g_k^T z_{k-1} d_{k-1}}{d_{k-1}^T g_k - \| g_k \|^2} \right) g_k - \frac{g_k^T z_{k-1} d_{k-1}}{d_{k-1}^T g_k - \| g_k \|^2} d_{k-1}.
\]

\[
d_k^{MDY} = \left( 1 + \frac{\| g_k \|^2}{d_{k-1}^T z_{k-1} g_k} \right) g_k - \frac{\| g_k \|^2}{d_{k-1}^T z_{k-1} g_k} d_{k-1},
\]

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

where \( \beta_k^{MHZ} = \frac{g_k^T z_{k-1} d_{k-1}}{d_{k-1}^T z_{k-1} - 2 \frac{\| z_{k-1} \|^2}{d_{k-1}^T z_{k-1}} \frac{g_k^T d_{k-1}}{\| g_k \|^2} d_{k-1}} \).

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

**Theorem (7)**

Let scalar \( \beta_k \) 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.

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

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 codes 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 $\delta = 1 \times 10^{-5}$ and $\sigma = 0.01$. For each test problem, the termination criterion is:

$$\left| g_{k+1} \right| < 1 \times 10^{-6}$$

(52)

We also force these routines to stop 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

<table>
<thead>
<tr>
<th>Test Functions</th>
<th>Dimensions</th>
<th>MHS</th>
<th>MDL</th>
<th>HS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NOI</td>
<td>NOFG</td>
<td>NOI</td>
</tr>
<tr>
<td>Ex. Powell</td>
<td>1000</td>
<td>41</td>
<td>119</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>41</td>
<td>119</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>44</td>
<td>140</td>
<td>60</td>
</tr>
<tr>
<td>DENSHCHNF</td>
<td>1000</td>
<td>22</td>
<td>47</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>21</td>
<td>44</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>22</td>
<td>47</td>
<td>F</td>
</tr>
<tr>
<td>Ex. Wood</td>
<td>1000</td>
<td>30</td>
<td>68</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>30</td>
<td>69</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>32</td>
<td>72</td>
<td>34</td>
</tr>
<tr>
<td>Dixmaana (Cute)</td>
<td>1000</td>
<td>5</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>5</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>5</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>Dixmaanc (Cute)</td>
<td>1000</td>
<td>6</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>6</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>6</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>NOND</td>
<td>1000</td>
<td>27</td>
<td>65</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>27</td>
<td>67</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>27</td>
<td>67</td>
<td>27</td>
</tr>
<tr>
<td>Ex. Block Diagonal - 1</td>
<td>1000</td>
<td>18</td>
<td>39</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>22</td>
<td>46</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>23</td>
<td>49</td>
<td>24</td>
</tr>
<tr>
<td>ENGVAL- 1 (Cute)</td>
<td>1000</td>
<td>21</td>
<td>46</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>25</td>
<td>88</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>Strait</td>
<td>1000</td>
<td>7</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>7</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>7</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>Wolfe</td>
<td>1000</td>
<td>68</td>
<td>137</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>121</td>
<td>243</td>
<td>159</td>
</tr>
</tbody>
</table>
Table 2- the comparison between MHS, MDL and HS with Wolfe Type Line Search.

<table>
<thead>
<tr>
<th>Test Functions</th>
<th>Dimensions</th>
<th>MHS</th>
<th>MDL</th>
<th>HS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NOI</td>
<td>NOFG</td>
<td>NOI</td>
</tr>
<tr>
<td>Ex. Powell</td>
<td>1000</td>
<td>41</td>
<td>119</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>60</td>
<td>200</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>60</td>
<td>200</td>
<td>55</td>
</tr>
<tr>
<td>DENSCHNF</td>
<td>1000</td>
<td>22</td>
<td>47</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>22</td>
<td>47</td>
<td>23</td>
</tr>
<tr>
<td>Ex. Wood</td>
<td>1000</td>
<td>30</td>
<td>68</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>30</td>
<td>69</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>32</td>
<td>72</td>
<td>34</td>
</tr>
<tr>
<td>Dixmaana (Cute)</td>
<td>1000</td>
<td>5</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>Dixmaanc (Cute)</td>
<td>1000</td>
<td>6</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>NOND</td>
<td>1000</td>
<td>24</td>
<td>57</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>24</td>
<td>57</td>
<td>27</td>
</tr>
<tr>
<td>Ex. Block Diagonal - 1</td>
<td>1000</td>
<td>20</td>
<td>42</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>18</td>
<td>38</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>21</td>
<td>44</td>
<td>32</td>
</tr>
<tr>
<td>ENGVAl-1 (Cute)</td>
<td>1000</td>
<td>24</td>
<td>51</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>51</td>
<td>104</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>F*</td>
<td>F*</td>
<td>F*</td>
</tr>
<tr>
<td>Strait</td>
<td>1000</td>
<td>7</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>7</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>100000</td>
<td>7</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>Wolfe</td>
<td>1000</td>
<td>68</td>
<td>137</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>10000</td>
<td>121</td>
<td>243</td>
<td>159</td>
</tr>
</tbody>
</table>
References:


