DEVELOPMENT OF THE HESTENES AND STEIFEL ALGORITHM TO SOLVE UNCONSTRAINED OPTIMIZATION

Basim A. Hassan¹ and Ali J. AL-Issa *²

¹Department of Mathematics, College of Computers Sciences and Mathematics University of Mosul, IRAQ
²Department of Mathematics, College of Computers Sciences and Mathematics University of Aleppo, Syria

*alieisa558@gmail.com.

Abstract

In this paper we present the development of the algorithm the Hestenes and Steifel (HS), that will be through suggest conjugate coefficient, then analyzing, studying the convergence of the suggested algorithm and proving the sufficient descent condition and global convergence. The numerical results have shown the effectiveness of the suggested algorithm after applying it on a group of standard tests problems

Keywords: Unconstrained Optimization, Conjugate Gradient, Conjugate Coefficient, Global Convergence.
1. INTRODUCTION

Given unconstrained optimization problem as follows[1]:

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

To solve the problem (1), we start with the following iterative relationship:

\[
x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \ldots
\]

\(\alpha_k > 0\), is a step size. Calculates strong Wolfe–Powell[2] conditions:

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

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

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

d_k is the search direction computed as follow[3]:

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

where \(\beta_k \in \mathbb{R}\) is known CG coefficient that characterizes different CG methods. Some classical methods such as the Hestenes and Stiefel (HS) [4], and Fletcher Reeves (FR) [5], Polak_Ribiére_Polyak (PRP) [6, 7].

\[
\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}} , \quad \beta_k^{FR} = \frac{g_k^T g_k}{\|g_k\|^2} , \quad \beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{\|g_k - g_{k-1}\|^2}
\]

In this Paper, we suggest anew modified nonlinear conjugate gradient method based on Hestenes and Steifel [4], our new modified formula algorithm is presented in Section 2. The global convergence of the method, in Section 3. Some Numerical applications is presented in Section 4, a short conclusion in section 5.

1. NEW CONJUGATE GRADIENT COEFFICIENT

We recall that for quasi-Newton methods, an approximation matrix \(B_k\) to the Hessian \(\nabla^2 f(x)\) is updated so that a new matrix \(B_k\) satisfies the following secant condition:

\[
B_{k+1} v_k = y_k
\]
By expanding condition (5), Marko et al. [8] proposed the following modified secant condition:

\[ v_k = \gamma_{k+1}^{-1} y_k, \quad G = \gamma_{k+1}^{-1} I_{n \times n} = \left( \frac{\| y_k \|}{\| v_k \|} \right) \times I_{n \times n} \]  \hspace{1cm} (6)

This means that the ideal case would be for the vector \( v_k \) to be a scalar multiple of \( y_k \).

\[ B_k v_k = y_k \Rightarrow \gamma_{k+1} v_k = y_k \]

In this section, we will give a new conjugate gradient method by using the modified quasi-Newton condition (6) instead of the ordinary one (5):

\[ v_k^{-1} G v_k = v_k^{-1}(\gamma_{k+1}^{-1} I_{n \times n}) v_k = \gamma_{k+1}^{-1} v_k^{-1}(I_{n \times n}) v_k = \gamma_{k+1}^{-1} v_k^{-1} v_k \]  \hspace{1cm} (7)

On the other hand since \( y_k = g_{k+1} - g_k \) then:

\[ v_k^T G v_k = v_k^T y_k = v_k^T g_{k+1} - v_k^T g_k \]  \hspace{1cm} (8)

\[ v_k^T g_{k+1} - v_k^T g_k = \gamma_{k+1}^{-1} v_k^T v_k \]  \hspace{1cm} (9)

\[ v_k^T g_{k+1} = \gamma_{k+1}^{-1} v_k^T v_k + v_k^T g_k \]

It follows from Perry's conjugacy conditions and (9) we have:

\[ d_k^T y_k = -\gamma_{k+1}^{-1} v_k^T v_k - v_k^T g_k \]  \hspace{1cm} (10)

Additionally, \( d_{k+1} = -g_{k+1} + \beta_k d_k \) and (10) imply that:

\[ -g_{k+1}^T y_k + \beta_k d_k^T y_k = -\gamma_{k+1}^{-1} v_k^T v_k - v_k^T g_k \]  \hspace{1cm} (11)

\[ \beta_k d_k^T y_k = -\gamma_{k+1}^{-1} v_k^T v_k - v_k^T g_k + g_{k+1}^T y_k \]

which implies that:

\[ \beta_{RS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - \frac{v_k^T g_k}{d_k^T y_k} - \gamma_{k+1}^{-1} \frac{v_k^T v_k}{d_k^T y_k} \]  \hspace{1cm} (12)

whereas: \( v_k = x_{k+1} - x_k, y_k = g_{k+1} - g_k \).
New Algorithm

Step1. Initialization. Select \( x_0 \in \mathbb{R}^n \) and the parameters \( 0 < \rho < \sigma < 1 \). Compute \( f(x_0) \) and \( g_0 \). Consider \( d_0 = -g_0 \) and \( \alpha_0 = 1/\|g_0\| \). Set \( k = 0 \).

Step2. Test for continuation of iterations. If \( \|g_k\| \leq 10^{-6} \), then stop, else set \( k = k + 1 \).

Step3. Line search. Compute \( \alpha_k \) satisfying the Wolfe line search conditions (4) and (5) and update the variables \( x_{k+1} = x_k + \alpha_k d_k \).

Step 4. Compute \( \beta_{kRS} \) as in (12).

Step 5. Compute the search direction \( d_{k+1} \) as in (4). If the restart criterion of Powell \( |g_{k+1}^T g_k| \geq 0.2\|g_{k+1}\|^2 \), is satisfied, then set \( d_{k+1} = -g_{k+1} \) otherwise set \( k = k + 1 \) and go to Step 2.

2. CONVERGENCE ANALYSIS

In this section, we present the convergence analysis of RS method under inexact line search. A convergent algorithm has to satisfy the sufficient descent and global convergence properties.

Assumption[9]

(i) The level set \( S = \{ x \in \mathbb{R}^n | f(x) \leq f(x_0) \} \) is bounded.
(ii) In a neighborhood \( N \) of \( S \) the function \( f \) is continuously differentiable and its gradient is Lipschitz continuous, i.e. there exists \( L > 0 \) such that \( \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \) for all \( x \) and \( y \) from \( N \).
(iii) there exists a constant \( \Gamma \geq 0 \) such \( \|g_k\| \leq \Gamma \) that for all \( x \)

3.1 SUFFICIENT DESCENT CONDITION:

The descent property ([10], Al-Baali, 1985) and the sufficient descent property are important conditions for the global convergence.

Theorem 1.

Consider a CG method with the search direction \( d_{k+1} \) (4) and \( \beta_k \) given as (12), if \( 0 < L < 1 \) and \( d_k^T y_k \neq 0 \) \( \forall k \geq 1 \) then Sufficient descent Condition will also hold true.
\[d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T y_k}{d_k^T y_k} \frac{v_k^T g_k}{d_k^T y_k} - \frac{v_k^T v_k}{d_k^T y_k} d_k\]

Then we have:
\[g_{k+1}^T d_{k+1} \leq -c\|g_{k+1}\|^2\]  \hspace{1cm} (14)

**Proof:**
Since \(d_0 = -g_0\), we have \(g_0^T d_0 = -\|g_0\|^2 \leq 0\). Suppose that \(g_k^T d_k \leq -c\|g_k\|^2\) for all \(k \in n\). To complete the proof, we have to show that the theorem is true for all \(k + 1\). Multiplying (13) by \(g_{k+1}^T\), we have:
\[d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + \left[ \frac{g_{k+1}^T y_k}{d_k^T y_k} \frac{v_k^T g_k}{d_k^T y_k} - \frac{v_k^T v_k}{d_k^T y_k} \right] d_k^T g_{k+1}\]

Since \(\gamma_k = \frac{\|y_k\|}{\|v_k\|} \leq L\|v_k\| \leq L\) and \(v_k^T y_k \leq L v_k^T v_k \implies -\frac{v_k^T v_k}{L} \geq -v_k^T y_k\) then
\[-\gamma_k v_k \leq L \left( -\frac{v_k^T y_k}{L} \right) = -v_k^T y_k\]

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

But, \(d_k^T y_k > 0\), therefore:
\[d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k (v_k^T g_{k+1} + v_k^T y_k)}{(v_k^T y_k)^2} - \frac{(v_k^T g_{k+1})^2}{(v_k^T y_k)^2}\]

From the following inequality:
\[u_k^T v_k \leq \frac{1}{2} \left[ \|u_k\|^2 + \|v_k\|^2 \right], u_k, v_k \in \mathbb{R}^n\]  \hspace{1cm} (17)

it can be derived that:
\[ u_k = g_{k+1}(v_k^T y_k), \quad s_k = (v_k^T g_{k+1}) y_k \]
\[ y_k^T g_{k+1}(v_k^T g_{k+1})(v_k^T y_k) \leq \frac{1}{2} \left[ ||g_{k+1}||^2 (v_k^T y_k)^2 + ||y_k||^2 (v_k^T g_{k+1})^2 \right] \]
So, it follows from (5) and (7) that:
\[ g_{k+1}^T d_{k+1} = - g_{k+1}^T y_k + \frac{1}{2} \left[ ||g_{k+1}||^2 (v_k^T y_k)^2 + ||y_k||^2 (v_k^T g_{k+1})^2 \right] \]
\[ \leq - \left( 1 - \frac{1}{2} \right) ||g_{k+1}||^2 + \left[ \frac{1}{2} ||y_k||^2 - (v_k^T y_k) \right] \frac{(v_k^T g_{k+1})^2}{(v_k^T y_k)^2} \]
\[ \leq - \frac{1}{2} ||g_{k+1}||^2 + \left[ \frac{1}{2} ||y_k||^2 - (v_k^T y_k) \right] \frac{(v_k^T g_{k+1})^2}{(v_k^T y_k)^2} \]
Since \( y_k^T y_k \leq L v_k^T y_k \), then:
\[ g_{k+1}^T d_{k+1} \leq - \frac{1}{2} ||g_{k+1}||^2 \leq - c ||g_{k+1}||^2 \] (18)
the sufficient descent property (14) holds.

### 3.2 GLOBAL CONVERGENCE PROPERTIES:

Next, we will show that CG methods with \( \beta_k \) converge globally. However, we first need to assume that the step size \( \alpha_k \) is obtained by the strong Wolfe line search (3) so that our convergence proof will be markedly easier.

Dai et al. [11] proved that for any conjugate gradient method with strong Wolfe line search the following general result holds:

**Lemma 1.**

Suppose that the assumptions (i) and (ii) hold and consider any conjugate gradient method (2), where \( d_{k+1} \) is a descent direction and \( \alpha_k \) is obtained by the strong Wolfe line search. If:
\[ \frac{1}{||d_{k+1}||^2} = \infty \] (19)
then:
\[ \lim_{k \to \infty} (\inf ||g_{k+1}||) = 0 \] (20)
Therefore, the following theorem can be proved.
Theorem 2.

Suppose that the assumptions (i) and (ii) hold and consider the conjugate gradient method \((2)\), where \(d_{k+1}\) is a descent direction. Assume that \(m < \nabla^2 f(x_{k+1}) < M\), where \(m\) and \(M\) are positive constants, then:

\[
\lim_{k \to \infty} (\inf \|g_{k+1}\|) = 0
\]

Proof:

Since \(m < \nabla^2 f(x_{k+1}) < M\) it follows that:

\[
|\beta_k| = \left| \frac{g_{k+1}^T y_k}{v_k^T y_k} - \frac{v_k^T g_k}{v_k^T y_k} \right| \leq \left| \frac{g_{k+1}^T y_k}{v_k^T y_k} \right| + \left| \frac{v_k^T g_{k+1}}{v_k^T y_k} \right|
\]

But, by using:

\[
|\beta_k| \leq \frac{M}{m} \frac{|v_k^T g_{k+1}|}{\|v_k\|^2} + \frac{|v_k^T g_{k+1}|}{m \|v_k\|^2} \leq \frac{(M+1)}{m} \frac{|v_k^T g_{k+1}|}{\|v_k\|^2} \leq \frac{(M+1)}{m} \frac{\|g_{k+1}\|}{\|v_k\|}
\]

Therefore,

\[
\|d_{k+1}\| \leq \|g_{k+1}\| + |\beta_k| \|v_k\| \leq \|g_{k+1}\| + \frac{(M+1)}{m} \frac{\|g_{k+1}\|}{\|v_k\|} \|v_k\| \leq \frac{\Gamma(m+M+1)}{m}
\]

Hence,

\[
\sum_{k \geq 1} \frac{1}{\|d_k\|^2} \geq \left( \frac{m}{\Gamma(m+M+1)} \right)^2 \sum_{k \geq 1} 1 = \infty
\]

By Lemma 1 we have:

\[
\lim_{k \to \infty} (\inf \|g_{k+1}\|) = 0
\]

3. NUMERICAL APPLICATIONS

We use our new method (RS) to solve a standard problem [12], and then compare it with a method in terms of number of iterations (NI), number of evaluations of the function (NF)
The step size was selected according to Strong Wolfe-Powell conditions with $\rho = 10^{-4}$, $\sigma = 0.9$. We stop the iterations if the inequality $\|g_k\| \leq 10^{-6}$.

**Table 1**: Comparison of different CG-algorithms with different test functions and different dimensions.

<table>
<thead>
<tr>
<th>Test functions</th>
<th>Algorithm HS</th>
<th>Algorithm RS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>NI / NF</td>
</tr>
<tr>
<td>Freudenstein and Roth</td>
<td>100</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>23 / 324</td>
</tr>
<tr>
<td>Extended Trigonometric</td>
<td>100</td>
<td>19 / 35</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>39 / 67</td>
</tr>
<tr>
<td>Extended Rosenbrock</td>
<td>100</td>
<td>34 / 72</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>35 / 77</td>
</tr>
<tr>
<td>Generalized Tridiagonal-1</td>
<td>100</td>
<td>22 / 44</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>40 / 484</td>
</tr>
<tr>
<td>Extended Three ExponentialTerms</td>
<td>100</td>
<td>13 / 20</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>26 / 260</td>
</tr>
<tr>
<td>Generalized Tridiagonal-2</td>
<td>100</td>
<td>42 / 62</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>67 / 26</td>
</tr>
<tr>
<td>Extended PSC1</td>
<td>100</td>
<td>8 / 17</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>26 / 505</td>
</tr>
<tr>
<td>Extended Maratos</td>
<td>100</td>
<td>66 / 155</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>71 / 161</td>
</tr>
<tr>
<td>Extended Hiebert</td>
<td>100</td>
<td>79 / 174</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>79 / 171</td>
</tr>
<tr>
<td>Quadratic QF1</td>
<td>100</td>
<td>100 / 152</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>371 / 586</td>
</tr>
<tr>
<td>Extended Quadratic Penalty QP1</td>
<td>100</td>
<td>10 / 24</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>F</td>
</tr>
<tr>
<td>ENGVAL1</td>
<td>100</td>
<td>29 / 50</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>F</td>
</tr>
<tr>
<td>DENSCHNA</td>
<td>100</td>
<td>21 / 33</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>18 / 30</td>
</tr>
<tr>
<td>SINCOS</td>
<td>100</td>
<td>8 / 17</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>26 / 505</td>
</tr>
</tbody>
</table>
Development of the Hestenes and Steifel Algorithm to Solve Unconstrained Optimization

Fig. 1. Number of iteration performance profile.

Fig. 1. Number of function evaluation performance profile.
From figures 1-2, it is easy to see that the method RS is the best among the HS method in the perspectives of the number of iterations and number of evaluations of the function (NF).

4. CONCLUSION

A new method (RS) has been developed to solve convex functions of high dimensions and study the convergence of the proposed method. From the results recorded in the table we find the efficiency of the method.

References

Development of the Hestenes and Steifel Algorithm to Solve Unconstrained Optimization

