Global Convergence of Conjugate Gradient Method in Unconstrained Optimization Problems

Huda Y. Najm, Eman T. Hamed and Huda I. Ahmed

ABSTRACT: In this study, we propose a new parameter in conjugate gradient method. It is shown that the new method fulfills the sufficient descent condition with the strong Wolfe condition when inexact line search has been used. The numerical results of this suggested method also shown that this method outperforms to other standard conjugate gradient method.

Key Words: Unconstrained optimization, Conjugate Gradient method, Inexact line search, Global convergence.

Contents

1 Introduction 227
2 New formula for $\beta_k$ and the algorithm 228
3 Sufficient Descent Property and Global Convergence Analysis 228
4 Numerical Results 229
5 Conclusion 230

1. Introduction

Regarding the following problem:

$$\min f(x), \quad x \in \mathbb{R}^n \quad (1.1)$$

where $f$ is a continuously differentiable function. The non linear conjugate gradient methods are efficient to solve this problem by iterative method at the $(k+1)$ iteration by the following iteration form:

$$x_{k+1} = x_k + \alpha_k d_k \quad (1.2)$$

Where the step length $\alpha_k > 0$ and $d_k$ denoted by the search direction:

$$d_k = \begin{cases} -g_k & \text{for } k = 1 \\ -g_k + \beta_k d_{k-1} & \text{for } k > 1 \end{cases} \quad (1.3)$$

where $\beta_k$ is a scalar conjugacy coefficient, there are some well-known formulas of his scalar such as: Hestenes-Stiefel (HS) [1], Fletcher-Reeves (FR) [2], Polak-Ribière

2010 Mathematics Subject Classification: 65k10.
Submitted February 04, 2019. Published September 01, 2019
The convergence behavior of conjugate gradient methods are different. There are many convergence results with some line search conditions has been widely studied, there method can guarantee the descent property of each direction which provided the step length computed by carrying out a line search and its satisfies the strong Wolfe conditions.

2. New formula for $\beta_k$ and the algorithm

In this section, a new coefficient of conjugate gradient using Hestenes-Steifel formula in the original numerator, and the denominator is introduced in the following formula:

$$
\beta_{k}^{new} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1} + \mu \| g_k^T d_{k-1} \|},
$$

where $g_{k-1}(g_k - g_{k-1})$ and $\mu$ is positive constant.

Algorithm (2.1)

Step 1: For the initial point $x_1 \in \mathbb{R}^n$, Set $d_1 = -g_1$, $k = 1$, if $\|g_1\| \leq \varepsilon$ then stop.

Step 2: Set $d_k = -g_k$.

Step 3: Find $\alpha_k > 0$ satisfying the Strong Wolfe Conditions.

Step 4: Let $x_{k+1} = x_k + \alpha_k d_k$ and If $|g_1| \leq \varepsilon$ then stop.

Step 5: Compute $\beta_k$ by the new formula (6) where $y_{k-1} = (g_k - g_{k-1})$ and $\mu > 1$, then generate $d_k$ by (3).

Step 6: If $k = n$ or $\frac{|g_k^T g_1|}{\|g_1\|^2} \geq 0.2$ Powell restart [12], then go to step 2.

Step 7: Set $k := k + 1$, go to Step 4.

3. Sufficient Descent Property and Global Convergence Analysis

We make the following basic assumptions on the objective function in order to establish the global convergence results for the new algorithm:

Assumption (3.1) (see [14])

(i) $f$ is bounded below on the level set $\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$

(ii) In some neighborhood $\Omega_0$ of $\Omega$, $f$ is differentiable and it is gradient $g(x)$ is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$
\|g(x) - g(y)\| \leq \|x - y\|, \forall x, y \in \Omega
$$

Under these assumptions, there exists a constant $\epsilon > 0$ such that

$$
\|g_k\| \leq \epsilon, \forall k
$$

Lemma (3.2) Suppose that Assumption (3.1) holds, let the sequence $\{x_k\}$ generated by the algorithm (2.1) and the step length $\alpha_k$ satisfies Wolfe conditions, then

$$
g_k^T d_k \leq -c\|g_k\|^2
$$
Global Convergence of Conjugate Gradient Method

Numerical Results by compare between the new-CG Method and Standard HS language. These test problems are contributed in CUTE, we can find the details of a set of test problems of (35) nonlinear unconstrained problems by using Fortran.

Proof: For the initial direction \( k = 1 \), since \( d_0 = -g_0 \), then \( g_0^T \leq -\|g_0\|^2 \) which satisfied (8). For some \( k > 1 \) and by using (3) and (5), we get

\[
g_k^T d_k = -\|g_k\|^2 + \beta_{k}^{new} g_k^T d_{k-1} = -\|g_k\|^2 + \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1} + \mu |g_k^T d_{k-1}|} g_k^T d_{k-1}
\]

\[
= -\|g_k\|^2 + \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T d_{k-1} + \mu |g_k^T d_{k-1}|} g_k^T d_{k-1}.
\]

We have

\[
d_{k-1}^T y_{k-1} = g_k^T d_{k-1} - g_{k-1}^T d_{k-1} \geq \sigma g_k^T d_{k-1} - g_{k-1}^T d_{k-1} = (\sigma - 1) g_k^T d_{k-1} > 0,
\]

i.e \( d_{k-1}^T y_{k-1} > 0 \), and from the strong Wolfe conditions \( \sigma g_k^T d_k \leq g((x_k + \alpha d_k)^T d_k \leq -\sigma g_k^T d_k \), we get

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

\[
\leq -\|g_k\|^2 + \frac{\|g_k\|^2}{\mu} + \frac{g_k^T g_{k-1}}{\mu}
\]

(3.5)

From Powell restart condition \( g_k^T g_{k-1} > 0, 0.2\|g_k\|^2 \), [7] we have:

\[
g_k^T d_k \leq -\|g_k\|^2 - \frac{\|g_k\|^2}{\mu} - \frac{(0.2)\|g_k\|^2}{\mu}
\]

\[
\leq - \left(1 + \frac{1}{\mu} + \frac{(0.2)}{\mu}\right)\|g_k\|^2
\]

\[
\leq - c\|g_k\|^2
\]

where \( c = 1 + \frac{1}{\mu} + \frac{(0.2)}{\mu} \), and \( \mu > 1 \).

Theorem (3.3) Consider the iteration method \( x_{k+1} = x_k + \alpha_d d_k \) where \( d_k \) defined by (3), (5) and suppose that Assumption (3.1) holds. Then the new algorithm either stops at stationary point

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

(3.7)

4. Numerical Results

In this section, the main idea to report the performance of the new method on a set of test problems of (35) nonlinear unconstrained problems by using Fortran language. These test problems are contributed in CUTE, we can find the details of the test functions, in [8] and [9]. For each test function, the number of variables \( n = 100, 200, ..., 500 \). In order to evaluate the reliability of the new proposed method. Numerical Results by compare between the new-CG Method and Standard HS.
Method by depend the following tools, $n$: dimension of the problem, $iter$: number of iterations, $irs$: number of restart, $fgcnt$: number of function and gradient evaluations., $time$: total time required to complete the evaluation process. $fxnew$: the value of function. And $gnorm$: the minimum gradient values, but we do not give the results of all test function due to page limit, see Table (1).

Table 1: Compare Numerical Results for NEW-CG Method and Standard HS-CG Method

<table>
<thead>
<tr>
<th>Function test</th>
<th>Standard HS-CG Method</th>
<th>NEW-CG Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>$fgcnt$</td>
</tr>
<tr>
<td>DIXMAANA (CUTE)</td>
<td>100</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>6</td>
</tr>
<tr>
<td>LIARWHD (CUTE)</td>
<td>100</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>19</td>
</tr>
<tr>
<td>Extended Freudenstein &amp; Roth</td>
<td>100</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>12</td>
</tr>
<tr>
<td>Extended Trigonometric Function</td>
<td>100</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function test</th>
<th>Standard HS-CG Method</th>
<th>NEW-CG Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$</td>
<td>$fgcnt$</td>
</tr>
<tr>
<td>NONDIA (Shanno-78) (CUTE)</td>
<td>100</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>10</td>
</tr>
<tr>
<td>DIXMAANB (CUTE)</td>
<td>100</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>11</td>
</tr>
<tr>
<td>Extended Powell</td>
<td>100</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>58</td>
</tr>
</tbody>
</table>

5. Conclusion

We have proposed a new $\beta_k$, also we have provided proof the global convergence. The effectiveness of the new proposed method $\beta_k^{new}$ has the good performance.
compared with other standard conjugate gradient method depended on the selected list of test functions problems. Comparison in Total for all Function Test of new-CG Method against HS Method for 35 Test Problems:

References