New Hybrid Conjugate Gradient Method with Global Convergence Properties under Exact Line Search

Yasir Salih¹, Mustafa Mamat¹∗, Mohd Rivaie², Abdelrhaman Abashar³, Mohamad Afendee Mohamed¹

¹Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UNISZA), Kuala Terengganu, Terengganu, Malaysia
²Department of Computer Science and Mathematics, Universiti Teknologi MARA (UiTM), Terengganu, Malaysia
³Faculty of Engineering, Red Sea University, Sudan
∗Corresponding author E-mail: mast@unisza.edu.my

Abstract

Conjugate Gradient (CG) method is a very useful technique for solving large-scale nonlinear optimization problems. In this paper, we propose a new formula for β_c, which is a hybrid of PRP and WYL methods. This method possesses sufficient descent and global convergence properties when used with exact line search. Numerical results indicate that the new formula has higher efficiency compared with other classical CG methods.

Keywords: Nonlinear optimization; conjugate gradient coefficient; exact line search; global convergence; large scale.

1. Introduction

In this paper, we focus on the unconstrained optimization problem defined by

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

where \(f : \mathbb{R}^n \to \mathbb{R}\) is a continuously differentiable function. The iterates of CG method for solving (1) are obtained by

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

where \(\alpha_k\) is the step-size and computed by a line search procedure, that is

\[
\alpha_k = \arg \min_{\alpha \geq 0} f(x_k + \alpha d_k).
\]

The search direction \(d_k\) is given as

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

where \(\beta_k\) is the CG coefficient. There are many formulas for determining \(\beta_k\), each one defines different CG method. Examples of the widely known formulas for \(\beta_k\) are

\[
\beta^H_k = \frac{\|g_k\|^2}{\|g_k - g_{k-1}\|^2}, \quad \beta^{FR}_k = \frac{\|g_k - g_{k-1}\|^2}{\|g_k - 2g_{k-1} + g_{k-2}\|^2}, \quad \beta^{PRP}_k = \frac{\|g_k\|^2}{\|g_k - g_{k-1}\|^2}, \quad \beta^{LS}_k = \frac{\|g_k - g_{k-1}\|^2}{\|g_k - 2g_{k-1} + g_{k-2}\|^2}, \quad \beta^{DY}_k = \frac{\|g_k - g_{k-1}\|^2}{\|g_k - g_{k-2}\|^2},
\]

The above formulas are called Hestenes-Stiefel (HS) [1], Fletcher-Reeves (FR) [2], Polak-Ribiere-Polyak (PRP) [3,4], Conjugate Descent (CD) [5], Liu-Storey (LS) [6] and Dai-Yuan (DY) [7] respectively. All of these methods have finite convergence properties under the exact line search, provided that \(f(x)\) is strictly a convex quadratic function. For further information, readers can refer to [8-19].

2. New Parameter and Algorithm

In this section, we present a new formula for CG coefficient, referred here as \(\beta^{YHM}_k\) where YHM denotes Yasir, Hamoda and Mamat. The \(\beta^{YHM}_k\) is defined by

\[
\beta^{YHM}_k = \begin{cases} 
g_k^T (g_k - g_{k-1}) & \text{if } 0 \leq g_k^T g_{k-1} \leq \|g_k\|^2 \\
g_k^T g_{k-1} & \text{if otherwise}
\end{cases}
\]

The following algorithm is implemented using the proposed hybrid CG method.
Algorithm 2.1: Hybrid CG Method
Step 1: Given $x_0 \in \mathbb{R}^n$, $\varepsilon \geq 0$, set $d_0 = -g_0$, $k = 0$, if $\|g_0\| \leq \varepsilon$, then stop.
Step 2: Compute $\alpha_k$ by exact line search.
Step 3: Let $x_{k+1} = x_k + \alpha_k d_k$, if $\|g_{k+1}\| \leq \varepsilon$, then stop.
Step 4: Compute $\beta_k$ by (5) and generate $d_{k+1}$ by (4).
Step 5: Set $k = k + 1$ then go to Step 2.

3. Global Convergence Properties
In this section, we study the global convergence properties of $\beta_k^{YHM}$. Firstly, we need to simplify $\beta_k^{YHM}$ to ease the proving process. From (5), we have

$$
\beta_k^{YHM} = \begin{cases} 
\frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} & \text{if } 0 \leq g_k^T g_{k-1} \leq \|g_k\|^2 \\
\frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} & \text{otherwise}
\end{cases}
$$

If $0 \leq g_k^T g_{k-1} \leq \|g_k\|^2$, then

$$
\beta_k^{YHM} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} = \frac{\|g_k\|^2 - g_k^T g_{k-1}}{\|g_{k-1}\|^2} \geq 0.
$$

Otherwise,

$$
\beta_k^{YHM} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} \geq \frac{\|g_k\|^2 - \|g_k\| \|g_{k-1}\|}{\|g_{k-1}\|^2}.
$$

Using the Cauchy-Schwartz inequalities implies that

$$
\beta_k^{YHM} \geq \frac{\|g_k\|^2 - \|g_k\|^2}{\|g_{k-1}\|^2} = 0.
$$

Hence, we get $\beta_k^{YHM} \geq 0$.

3.1. Sufficient descent condition
The sufficient descent condition is defined by:

$$
g_k^T d_k \leq -c \|g_k\|^2 \quad \text{for } k > 0, c > 0.
$$

The following theorem shows that our new CG coefficient with exact line search possesses sufficient descent property.

**Theorem 1.** Suppose that $x_k$ and $d_k$ are generated by the iterative CG method of the form (2), (4) and (5) and the step-size $\alpha_k$ is determined by exact line search, then condition (6) holds for all $k \geq 0$.

**Proof.** The proof is by induction. If $k = 0$, then

$$
g_k^T d_0 = g_0^T (-g_0) = -\|g_0\|^2 < 0.
$$

Hence, condition (6) holds true for $k \geq 1$.

Multiply equation (4) by $g_k^T$, then

$$
g_k^T d_k = g_k^T (-g_k + \beta_k^{YHM} d_{k-1}) = -\|g_k\|^2 + \beta_k^{YHM} g_k^T d_{k-1}
$$

Note that for exact line search $g_k^T d_{k-1} = 0$. Thus, we get

$$
g_k^T d_k \leq -\|g_k\|^2.
$$

Hence, condition (6) also holds true for all $k \geq 1$.

3.2. Global Convergence Properties
The following assumptions are necessary for proving the convergence of CG method.

**Assumption 1.**
$f(x)$ is bounded from below on the level set $\Omega = \{x \in \mathbb{R}^n, f(x) \leq f(x_0)\}$ where $x_0$ is the starting point.

**Assumption 2.**
The gradient $g(x)$ is Lipschitz continuous in $N$ and there exists a constant $L > 0$ such that $\|g(x) - g(y)\| \leq L \|x - y\|$, $\forall x, y \in N$.

After this assumption, we have the following lemma which was proved by [20].

**Lemma 1.** Suppose that Assumptions 1 and 2 hold, let $x_k$ be generated by Algorithm 2.1 and $d_k$ satisfies $g_k^T d_k < 0$ for all $k$ and $\alpha_k$ is obtained by (3), then

$$
\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty.
$$

The following theorem is based on Lemma 1.

**Theorem 2.** Suppose that Assumption 1 and 2 hold, the sequence $\{x_k\}$ is generated by Algorithm 2.1. If $\|x_k\| = \|x_k d_k\| \to 0$ while $k \to \infty$, then

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

**Proof.** Case 1: If $0 \leq g_k^T g_{k-1} \leq \|g_k\|^2$ then $\beta_k^{YHM} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}$ to prove this theorem, we use contradiction. If Theorem 1 is not true, then the constant $c > 0$ exists such that

$$
\|g_k\| \geq c.
$$

From equation (4), we have $d_k = -g_k + \beta_k^{YHM} d_{k-1}$. Now, by squaring both sides of the equation, we obtain

$$
\|d_k\|^2 = \|g_k\|^2 - 2\beta_k^{YHM} g_k d_{k-1} + (\beta_k^{YHM})^2 \|d_{k-1}\|^2.
$$

For exact line search, $g_k^T d_{k-1} = 0$. Therefore,

$$
\|d_k\|^2 = (\beta_k^{YHM})^2 \|d_{k-1}\|^2 + \|g_k\|^2.
$$
Then, $\|d_k\|^2 \leq \|g_k\|^2 + \frac{\|g_k\|^4}{\|g_{k-1}\|^2} \|d_{k-1}\|^2$

We divide the above equation by $\|g_k\|^4$ to get

$$\|d_k\|^2 \leq \frac{1}{\|g_k\|^2} \|d_{k-1}\|^2 + \frac{1}{\|g_{k-1}\|^2} \|d_{k-1}\|^2$$

(12)

Utilizing (12) recursively and noting that

$$\|d_0\|^2 = -\frac{\|d_0\|}{\|g_0\|^2} \|d_0\|^2 = \|g_0\|^2,$$

we get

$$\|g_k\|^4 \leq \sum_{k=0}^{\infty} \|g_k\|^2$$

By (9), we know that $\|d_k\|^2 \leq \frac{k}{c^2}$. Therefore,

$$\|d_k\|^2 \geq \frac{c^2}{k} \cdot \|g_k\|^2 \geq \frac{c^2}{k}$$

This contradicts the Zoutendijk condition $\sum_{k=0}^{\infty} \frac{\|g_k\|^2}{|d_k|^2} < \infty$.

The proof is completed.

Case 2: If $\beta_k^{YHM} = \frac{\|g_k\|^2}{\|g_k\|^2 + \|g_{k-1}\|^2}$, the proof of this theorem can be seen in [21].

4. Results and Discussion

In this section, we present the computational performance of the new hybrid CG method under exact line search. Thirty-one unconstrained optimization test problems are taken from [22] and fully listed in Table 1. We analyze the efficiency of our new formula by comparing its numerical performance with the FR [2], PRP [3, 4], WYL [21] and HRM3 [23] methods. All of the tested CG algorithms use $\|g_k\| < \varepsilon$ as the stopping criterion with $\varepsilon = 10^{-5}$ as suggested by [24]. Four initial points are selected for the numerical test, starting from a point close to the solution and then moving on to other points farther from it. The numerical testing process uses MATLAB version 8.3.0.532 (R2014a) software on a computer with Intel(R) Core(TM) i5-MS20 (2.4GHz) CPU processor and 4GB RAM. In some cases, failure occurs when the line search is unable to find a positive step length.

<table>
<thead>
<tr>
<th>No.</th>
<th>Function</th>
<th>Dimension</th>
<th>Initial Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Six Hump Camel</td>
<td>2</td>
<td>10, -8, 8, 10</td>
</tr>
<tr>
<td>2</td>
<td>Three Hump Function</td>
<td>2</td>
<td>-10, 10, 20, 30</td>
</tr>
<tr>
<td>3</td>
<td>Booth</td>
<td>2</td>
<td>10, 25, 50, 100</td>
</tr>
<tr>
<td>4</td>
<td>Treccani</td>
<td>2</td>
<td>5, 10, 20, 50</td>
</tr>
<tr>
<td>5</td>
<td>Zettl</td>
<td>2</td>
<td>5, 10, 20, 30</td>
</tr>
<tr>
<td>6</td>
<td>Ex – Rosenbrock</td>
<td>2, 4, 10, 100, 500, 100, 10000</td>
<td>13, 25, 30, 50</td>
</tr>
<tr>
<td>7</td>
<td>Diagonal 4</td>
<td>2, 4, 10, 100, 500, 1000</td>
<td>1, 3, 6, 12</td>
</tr>
<tr>
<td>8</td>
<td>Shallow</td>
<td>2, 4, 10, 100, 500, 1000, 10000</td>
<td>10, 25</td>
</tr>
</tbody>
</table>

Table 1: A list of problem functions

We use the performance profile introduced by [25] to assess the efficiency of each tested method. Figure 1 and 2 show the numerical performances of those methods relative to the number of iterations and CPU time respectively. From the results, we can see that YHM solves 100% of the test problems, followed by HRM3 which solves 99.5%. PRP solves 99.4%, WYL solves 98% and FR solves 72% of the test problems. Since only YHM manage to solve all of the test functions, we can say that it has the best overall performance.
5. Conclusion

In this paper, we propose a new formula for $\beta_k$ under exact line search. The proposed method is shown to possess sufficient descent and global convergence properties. Based on the numerical results, our new method has the best performance compared to other tested CG methods.

Acknowledgement

The authors would like to thank the Universiti Sultan Zainal Abidin (Unisza), and Red Sea University of Sudan for financial support.

References