A Conjugate Gradient Method with Strong Wolfe-Powell Line Search for Unconstrained Optimization

Mohamed Hamoda 1, Mustafa Mamat 2, Mohd Rivaie 3 and Zabidin Salleh 1

1 School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu (UMT), 21030 Kuala Terengganu, Malaysia

2 Department of Computer Science and Mathematics, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, 22200 Terengganu, Malaysia

3 Department of Computer Science and Mathematics, Universiti Teknologi MARA (UITM), 23000 Terengganu, Malaysia

Abstract

In this paper, a modified conjugate gradient method is presented for solving large-scale unconstrained optimization problems, which possesses the sufficient descent property with Strong Wolfe-Powell line search. A global convergence result was proved when the (SWP) line search was used under some conditions. Computational results for a set consisting of 138 unconstrained optimization test problems showed that this new conjugate gradient algorithm seems to converge more stable and is superior to other similar methods in many situations.

Keywords: conjugate gradient coefficient, inexact line search, strong Wolfe-Powell line search, global convergence, large scale, unconstrained optimization

1. Introduction

Nonlinear conjugate gradient methods are well suited for large-scale problems due to the simplicity of their iteration and their very low memory requirements,
that is, they are designed to solve the following unconstrained optimization problem:

$$\min f(x), x \in \mathbb{R}^n$$  \hspace{1cm} (1)

where \( f : \mathbb{R}^n \to \mathbb{R} \) is a smooth, nonlinear function, and its gradient is denoted by \( g(x) = \nabla f(x) \). The iterative formula of the conjugate gradient methods is given by

$$x_{k+1} = x_k + \alpha_k d_k, \hspace{1cm} k = 0, 1, 2, \ldots$$  \hspace{1cm} (2)

where \( x_k \) is the current iterate point and \( \alpha_k \) is the step length, which is computed by carrying out a line search, and \( d_k \) is the search direction defined by

$$d_k = \begin{cases} -g_k & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1} & \text{if } k \geq 1, \end{cases}$$  \hspace{1cm} (3)

where \( \beta_k \) is a scalar, and \( g_k = g(x_k) \).

Various conjugate gradient methods have been proposed, and they mainly differ in the choice of the parameter \( \beta_k \). Some well-known formulas for \( \beta_k \) are given below:

$$\begin{align*}
\beta_k^{HS} &= \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}}, \\
\beta_k^{FR} &= \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}, \\
\beta_k^{PRP} &= \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T g_{k-1}}, \\
\beta_k^{CD} &= -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}, \\
\beta_k^{LS} &= -\frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T g_{k-1}}, \\
\beta_k^{DY} &= \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}}.
\end{align*}$$

where \( \| \cdot \| \) denotes the \( l_2 \)-norm. The corresponding method is respectively called, \( HS \) (Hestenes-Stiefel [1]), \( FR \) (Fletcher-Reeves [2]), \( PRP \) (Polak-Ribiére-Polyak [3, 4]), \( CD \) (Conjugate Descent [5]), \( LS \) (Liu-Storey [6]), and \( DY \) (Dai-Yuan [7]) conjugate gradient method. The convergence behavior of the above formulas with some line search conditions has been studied by many authors for many years [5-17].

In the already-existing convergence analysis and implementations of the conjugate gradient method, the weak Wolfe–Powell (WWP) line search conditions are as follows:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k$$  \hspace{1cm} (4)

$$g_k^T d_k \geq \sigma g_k^T g_k$$  \hspace{1cm} (5)

where \( 0 < \delta < \sigma < 1 \) and \( d_k \) is a descent direction. The strong Wolfe–Powell conditions consist of (4) and,

$$\| g(x_k + \alpha_k d_k) \| \leq \sigma \| g_k \|$$  \hspace{1cm} (6)

Furthermore, the sufficient descent property, namely,

$$g_k^T d_k \leq -c \| g_k \|^2$$  \hspace{1cm} (7)

Where \( c \) is a positive constant, is crucial to ensure the global convergence of the nonlinear conjugate gradient method with the inexact line search techniques [12, 13].
2. New formula for $\beta_k$ and its properties

During the last decade, much effort has been devoted to developing new modifications of conjugate gradient methods which do not only possess strong convergence properties, but they are also computationally superior to the classical methods. Such methods can be found in [18-30].

Recently, Wei et al. [31] gave a variant of the PRP method which is called the WYL method. Zhang studied and improved based on WYL a new conjugate gradient method, NPRP, and he proved that the NPRP method satisfied descent condition under strong Wolfe line search. Moreover, Zhang et al. proposed another modified method known as the MPRP method, where Dai and Wen [32] proposed a modified NPRP method known as the DPRP method. In this paper, enlightened by the above ideas, a modified PRP conjugate gradient method was proposed as follows:

$$
\beta_k^{HRM} = \frac{g_k^T (g_k - \|g_k\|g_{k-1})}{u \|g_{k-1}\|^2 + (1-u)\|d_{k-1}\|^2}
$$

(8)

where, HRM denotes Hamoda, Rivaie, and Mamat. According to the results obtained by [30], the value of the parameter $u$ can be set to $0 < u < 1$, but in this paper, we will test our new method with an arbitrary value $u = 0.4$

Therefore, we first provide the following algorithm:

**Algorithm (2.1)**
Step 1: Given $x_0 \in \mathbb{R}^n, \varepsilon \geq 0$. Set $d_0 = -g_0$ if $\|g_0\| \leq \varepsilon$ then stop.
Step 2: Compute $\alpha_k$ by (SWP) line search.
Step 3: Let $x_{k+1} = x_k + \alpha_k d_k, g_{k+1} = g(x_{k+1})$ if $\|g_{k+1}\| < \varepsilon$ then stop.
Step 4: Compute $\beta_k$ by formula (8) and generate $d_{k+1}$ by (3).
Step 5: Set $k = k + 1$ go to Step 2.

The following assumptions are often used in previous studies of the conjugate gradient methods:

**Assumption A**
$f(x)$ 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 B**
In some neighborhood $N$ of $\Omega$, the objective function is continuously differentiable, and its gradient is Lipschitz continuous, that is, there exists a constant $L > 0$ such that

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

(9)
In 1992, Gilbert and Nocedal introduced the property (*) which plays an important role in the studies of CG methods. This property means that the next research direction approaches the steepest direction automatically when a small step-size is generated, and the step-sizes are not produced successively [33].

**Property (*)**
Consider a CG method of the form (2) and (3). Suppose that, for all \( k \geq 0 \),
\[
0 < \gamma \leq \|g_k\| \leq \bar{\gamma}
\]  
(10)
where \( \gamma \) and \( \bar{\gamma} \) are two positive constants. We say that the method has property (*), if there exist constants \( 1 > b \geq 0 \) such that for all \( k \geq 0 \),
\[
|\beta_k| \leq \frac{1}{2b}, \quad \text{where} \quad S_k = \alpha_k d_k.
\]
The following lemma shows that the new method \( \beta_k^{HRM} \) has the property (*).

**Lemma 2.1**
Consider the method of form (2) and (3), Suppose that Assumptions A and B hold, then, the method \( \beta_k^{HRM} \) has property (*).

**Proof**
Set \( b = \frac{5\bar{\gamma}^2(\gamma + \bar{\gamma})}{2\gamma^3} > 1, \quad \lambda = \frac{\gamma^2}{10L\bar{\gamma}b} \). By (8) and (10) we have
\[
|\beta_k^{HRM}| \leq \frac{\|s_k\|}{u\|g_{k-1}\|^2 + (1 - u)\|d_{k-1}\|^2} \leq \frac{\|s_k\|}{0.4\|g_{k-1}\|^2} \leq \frac{5\bar{\gamma}(\gamma + \bar{\gamma})}{2\gamma^2} = \frac{5\bar{\gamma}^2(\gamma + \bar{\gamma})}{2\gamma^3} = b
\]
From Assumption B, (9) holds. If \( \|s_k\| \leq \lambda \) then,
\[
|\beta_k^{HRM}| \leq \frac{\|g_k - g_{k-1}\| + \|g_{k-1} - g_{k-1}\|}{u\|g_{k-1}\|^2 + (1 - u)\|d_{k-1}\|^2} \leq \frac{(L\lambda + \|g_{k-1}\| - \|g_k\|)\|g_k\|}{u\|g_{k-1}\|^2} \leq \frac{2L\lambda\|g_k\|}{0.4\|g_{k-1}\|^2} = \frac{5L\lambda\bar{\gamma}}{\gamma^2} = \frac{1}{2b}
\]
The proof is finished.

**3. The global convergence properties**

The following theorem shows that the formula \( HRM \) with SWP line search possess the sufficient descent condition

**Theorem 3.1**
Suppose that the sequences \( \{g_k\} \) and \( \{d_k\} \) are generated by the method of form (2), (3) and (8), and the step length \( \alpha_k \) is determined by the (SWP) line search (4) and
(6), if \( g_k \neq 0 \), then the sequence \( \{d_k\} \) possesses the sufficient descent condition (7).

**Proof**

By the formulae (8), we have the following:

\[
\beta_{k_h}^{HRM} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \geq \frac{\|g_{k-1}\|^2}{\|g_{k-1}\|^2} = 0
\]

Thus we get, \( \beta_{k_h}^{HRM} \geq 0 \)

Also

\[
\beta_{k_h}^{HRM} = \frac{\|g_k\|^2 - \|g_{k-1}\|^2}{\|g_{k-1}\|^2} \leq \frac{\|g_{k-1}\|^2}{\|g_{k-1}\|^2} \leq \frac{2\|g_k\|^2}{\|g_{k-1}\|^2}
\]

Hence, we obtain

\[
0 \leq \beta_{k_h}^{HRM} \leq \frac{2\|g_k\|^2}{\|g_{k-1}\|^2}
\]

Using (6) and (11), we get

\[
\beta_{k_h+1}^{HRM} \leq \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \sigma_{g_k} g_k d_k
\]

By (3), we have \( d_{k+1} = -g_{k+1} + \beta_{k+1} d_k \)

\[
\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} = -1 + \beta_{k+1} \frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2}
\]

We prove the descent property of \( \{d_k\} \) by induction. Since \( g_0^T d_0 = -\|g_0\|^2 < 0 \), if \( g_0 \neq 0 \), now suppose that

d_{i}, i = 1, 2, ..., k, are all descent directions, that is \( g_i^T d_i < 0 \)

By (12), we get

\[
\beta_{k+1}^{HRM} \leq \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \sigma_{g_k} g_k d_k
\]

That is,

\[
\frac{\|g_{k+1}\|^2}{\|g_k\|^2} 5\sigma_{g_k} g_k d_k \leq \beta_{k+1}^{HRM} \leq \frac{\|g_{k+1}\|^2}{\|g_k\|^2} 5\sigma_{g_k} g_k d_k
\]

(13) and (15) deduce,
\[-1 + \frac{5\sigma}{\|s_k\|^2} \leq g_{k+1}^T d_k \leq \frac{5\sigma}{\|s_k\|^2} \leq -1 - \frac{5\sigma}{\|s_k\|^2}\]

By repeating this process and the fact \( g_0^T d_0 = -\|g_0\|^2 \), we have,

\[-\sum_{j=0}^{k} (5\sigma)^j \leq \frac{g_{k+1}^T d_{k+1}}{\|s_k\|^2} \leq -2 \sum_{j=0}^{k} (5\sigma)^j \tag{16}\]

Since

\[\sum_{j=0}^{k} (5\sigma)^j < \sum_{j=0}^{\infty} (5\sigma)^j = \frac{1}{1-5\sigma}\]

(16) can be written as

\[-\frac{1}{1-5\sigma} \leq g_{k+1}^T d_{k+1} \leq -2 \frac{1}{1-5\sigma}\]

(17)

By making the restriction \( \sigma \in (0,1) \), we have \( g_{k+1}^T d_{k+1} < 0 \). So by induction, \( g_k^T d_k < 0 \) holds for all \( k \geq 0 \).

Denote \( c = 2 - \frac{1}{1-5\sigma} \) then, \( 0 < c < 1 \), and (17) turns out to be

\[(c-2)\|g_k\|^2 \leq g_k^T d_k \leq -c\|g_k\|^2\]

(18)

This implies that (7) holds. The proof is complete.

The following condition known as Zoutendijk condition was used to prove the global convergence of nonlinear CG methods [15, 34].

**Lemma 3.1**

Suppose that Assumptions A and B hold. Consider a CG method of the form (2) and (3), where \( d_k \) satisfies \( g_k^T d_k < 0 \), for all \( k \), and \( \alpha_k \) is obtained by (SWP) line search (4) and (6), Then,

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

(19)

The proof had been given in [35, 36]. In [10], Gilbert and Nocedal introduced the following important theorem:

**Theorem 3.2**

Consider any CG method of form (2) and (3), that satisfies the following conditions:

1. \( \beta_k \geq 0 \)
2. The search directions satisfy the sufficient descent condition.
3. The Zoutendijk condition holds.
4. Property(*) holds.
A conjugate gradient method

If the Lipschitz and boundedness Assumptions hold, then the iterates are globally convergent.

From (7), (9), (17) and Lemma 2.1, we found that the HRM method with the parameter \(0 < \delta < \sigma < 1/10\) satisfies all four conditions in theorem 3.2 under the strong Wolfe-Powell line search, so the method is globally convergent.

4. Numerical Experiments

In the present numerical experiments, we selected thirty-two different functions which had been earlier considered in [37-39] for both small-scale and large-scale optimization problems. Each of these functions was tested with different variables which lie in the range from 2 to 10,000. We tested a set of 138 problems with strong Wolfe-Powell line search. The algorithm was implemented using MATLAB R2011b (7.13.0.564), applying the strong Wolfe-Powell line search. All of the numerical experiments were run on the same PC with an Intel (R) CoreTM i3-M350 (2.27GHz) CPU, 4GB of RAM, and the Windows 7 operating system. In order to assess the reliability of the new proposed method, HRM, we tested this method against the well-known classical and modified methods of the FR, PRP, MPRP, and DPRP methods using the same problems, and assumed that the best method should require fewer iterations and less CPU time. All of these algorithms terminated when \(\|g_k\| < 10^{-6}\). The step size \(\alpha_k\) satisfies the strong Wolfe-Powell conditions, with \(\delta = 10^{-4}\), and \(\sigma = 0.001\). For the MPRP method, we chose \(\mu_1 = \mu_2 = 1\) and \(\mu_2 = 3\), where \(\mu = 3\) in DPRP method. A list of test functions, Dimension, and the initial points used are shown in Table 1. In some cases, the computation stopped due to the failure of the line search to find the positive step size, and thus it was considered as a failure. In addition, we considered the search to have failed if the number of iterations exceeded 1,000 or CPU execution time exceeded 500 seconds. Numerical results were relatively compared with the CPU time and number of iterations. The performance results are shown in Figures 1 and 2 respectively, using a performance profile introduced by Dolan and More [40].

<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>Booth</td>
<td>2</td>
<td>10, 25, 50, 100</td>
</tr>
<tr>
<td>3</td>
<td>Treccani</td>
<td>2</td>
<td>5, 10, 20, 50</td>
</tr>
<tr>
<td>4</td>
<td>Zettl</td>
<td>2</td>
<td>5, 10, 20, 30</td>
</tr>
<tr>
<td>5</td>
<td>Leon</td>
<td>2</td>
<td>2, 5, 8, 10</td>
</tr>
<tr>
<td>6</td>
<td>Three Hump</td>
<td>2</td>
<td>20, 50, 60, 150</td>
</tr>
<tr>
<td>7</td>
<td>Extended Wood</td>
<td>4</td>
<td>3, 5, 20, 30</td>
</tr>
</tbody>
</table>
Under a strong Wolfe-Powell line search, the performance profile of all methods measured by the number of iterations required is shown in Figure 1, and the performance profile based on the CPU time used is in Figure 2. The shapes of the profile plots in both Figures 1 and 2 are almost alike. A thorough inspection of the left side of both figures indicates that the lowest curve represents the FR method. Therefore, this method possesses the lowest performance. The top left side curve indicates that the PRP method is the best performer. The curves for methods DPRP, MPRP, and HRM fall in between the two extreme curves. Thus, the performance of this set of methods is in the middle of the sets based the number of iterations and CPU time.
From results shown in Figures 1 and 2, it is evident that the FR method achieved a success rate of only 0.65, while the PRP method scored 0.79, and the DPRP method recorded 0.88. Furthermore, the MPRP method achieved 0.94, and our new method, HRM achieved 1, that is, our new method scored a 100% success rate. Such result indicates that HRM method is the best among the four methods in the perspectives of the number of iterations and the CPU time. Hence, our new method successfully solved all the test problems, and it is competitive with the well-known conjugate gradient methods for unconstrained optimization.

Figure 1: Performance profile relative to the number of iterations

Figure 2: Performance profile relative to the CPU time
5. Conclusion

In this paper, we proposed a new conjugate gradient method for unconstrained optimization. Results showed that it could satisfy the sufficient descent condition and converge globally if the strong Wolfe-Powell line search was used. Numerical results show that the HRM method is efficient for the addressed problems.

Acknowledgements. The authors would like to thank the University of Malaysia Terengganu (Grant no FRGS Vot 59256)

References


A conjugate gradient method


A conjugate gradient method


[40] E. D. Dolan and J. J. Mor, Benchmarking optimization software with
http://dx.doi.org/10.1007/s101070100263

Received: July 4, 2015; Published: March 8, 2016