Research Article
The Hybrid BFGS-CG Method in Solving Unconstrained Optimization Problems

Mohd Asrul Hery Ibrahim, 1 Mustafa Mamat, 2,3 and Wah June Leong 4

1 School of Applied Sciences and Foundation, Infrastructure University Kuala Lumpur, 43000 Kajang, Malaysia
2 Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Tembila Campus, 22200 Besut, Malaysia
3 Department of Mathematics, Faculty of Science and Technology, Universiti Malaysia Terengganu (UMT), 21030 Kuala Terengganu, Malaysia
4 Department of Mathematics, Faculty of Science, Universiti Putra Malaysia (UPM), 43400 Serdang, Malaysia

Correspondence should be addressed to Mohd Asrul Hery Ibrahim; asrulhery@iukl.edu.my

Received 22 April 2013; Accepted 23 January 2014; Published 4 March 2014

Academic Editor: Lucas Jodar

Copyright © 2014 Mohd Asrul Hery Ibrahim et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In solving large scale problems, the quasi-Newton method is known as the most efficient method in solving unconstrained optimization problems. Hence, a new hybrid method, known as the BFGS-CG method, has been created based on these properties, combining the search direction between conjugate gradient methods and quasi-Newton methods. In comparison to standard BFGS methods and conjugate gradient methods, the BFGS-CG method shows significant improvement in the total number of iterations and CPU time required to solve large scale unconstrained optimization problems. We also prove that the hybrid method is globally convergent.

1. Introduction

The unconstrained optimization problem only requires the objective function as

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

where \( \mathbb{R}^n \) is an \( n \)-dimensional Euclidean space and \( f : \mathbb{R}^n \to \mathbb{R} \) is continuously differentiable. The iterative methods are used to solve (1). On the \( i \)th iteration, an approximation point \( x_i \) and the \((i + 1)\)th iteration are given by

\[
x_{i+1} = x_i + \alpha_i d_i,
\]

where \( d_i \) denotes the search direction and \( \alpha_i \) denotes the step size. The search direction must satisfy the relation \( g_i^T d_i < 0 \), which guarantees that \( d_i \) is a descent direction of \( f(x) \) at \( x_i \). The different choices of \( d_i \) and \( \alpha_i \) yield the different convergence properties. Generally the first order condition \( \nabla f(x_*) = 0 \) is used to check for local convergence to stationary point \( x_* \). There are many ways to calculate the search direction depending on the method used, such as the steepest descent method, conjugate gradient (CG) method, Newton-Raphson method, and quasi-Newton method.

The different choices of the step size ensure that the sequence of iterates \( x_i \) defined by (2) is globally convergent with some rates of convergence. There are two ways to determine the values of the step size, the exact line search, and the inexact line search. For the exact line search, \( \alpha_i \) is calculated by using the formula \( \alpha_i = \arg\min_{\alpha > 0} f(x_i + \alpha d_i) \). However, it is difficult and often impossible to find the value of step size in practical computation using the exact line search. Hence, the inexact line search is proposed by previous researchers like Armijo [1], Wolfe [2, 3], and Goldstein [4] to overcome the problem. Recently Shi proposed a new inexact line search rule similar to the Armijo line search and analysed the global convergence [5]. Shi also claimed that among several well-known inexact line search procedures published by previous researchers, the Armijo line search rule
is one of the most useful and the easiest to be implemented in computational calculations. The Armijo line search rule can be described as follows:

\[
\alpha_t = \max \{ s, s\beta, s\beta^2, \ldots \} \quad \text{such that} \quad f(x_t) - f(x_t + \alpha_t d_t) \geq -\sigma \alpha_t g_t^T d_t,
\]

\(i = 0, 1, 2, \ldots\) Then, the sequence of \(\{x_i\}_{i=0}^{\infty}\) is converged to the optimal point, \(x^*\), which minimises \(f\) [6]. Hence, we will use the Armijo line search in this research associated with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method and the new hybrid method.

This paper is organised as follows. In Section 2, we elaborate the step size and search direction that are used in this research. Here, the BFGS method and CG method also will be presented. Then, the new hybrid method and convergence analysis will be discussed in Section 3. An explanation about the numerical results is provided in Section 4 and the paper ends with a short conclusion in Section 5.

2. The Search Direction

The different methods in solving unconstrained optimization problems depend on the calculation of search direction, \(d_i\) in (2). In this paper, we will focus on the CG method and quasi-Newton methods. The CG method is useful for finding the minimum value of functions or unconstrained optimization problems, which are introduced by [7]. The search direction of the CG method is

\[
d_i = \begin{cases} 
eg g_i, & i = 0, \neg g_i + \beta_i d_{i-1}, & i \geq 2, \end{cases}
\]

where \(\beta_i\) is known as the CG coefficient. There are many ways to calculate \(\beta_i\) and some well-known formulas are

\[
\begin{align*}
\beta_i^{\text{FR}} &= \frac{g_i^T g_i}{\|g_i\|^2}, \\
\beta_i^{\text{PR}} &= \frac{g_i^T (g_i - g_{i-1})}{\|g_{i-1}\|}, \\
\beta_i^{\text{HS}} &= \frac{g_i^T (g_i - g_{i-1})}{(g_i - g_{i-1})^T d_{i-1}},
\end{align*}
\]

where \(g_i\) and \(g_{i-1}\) are gradients of \(f(x)\) at points \(x_i\) and \(x_{i-1}\), respectively, while \(\| \cdot \|\) is a norm of vectors and \(d_{i-1}\) is a search direction for the previous iteration. The above corresponding coefficients are known as Fletcher-Reeves (CG-FR) [7], Polak-Ribière (CG-PR) [8–11], and Hestenes-Stiefel (CG-HS) [12].

In quasi-Newton methods, the search direction is the solution of linear system

\[
d_i = -H_i g_i,
\]

where \(H_i\) is an approximation of Hessian. Initial matrix \(H_0\) is chosen by the identity matrix, which subsequently updates by an update formula. There are a few update formulas that are widely used like Davidon-Fletcher-Powell (DFP), BFGS, and Broyden family formula. This research uses a BFGS formula in a classical algorithm and the new hybrid method. The update formula for BFGS is

\[
H_{i+1} = H_i - \frac{H_i s_i s_i^T H_i}{s_i^T H_i s_i} + \frac{y_i y_i^T}{s_i^T y_i}.
\]

with \(s_i = x_i - x_{i-1}\) and \(y_i = g_i - g_{i-1}\). The approximation that the Hessian must fulfil is

\[
H_{i+1} s_i = y_i.
\]

This condition is required to hold for the updated matrix \(H_{i+1}\). Note that it is only possible to fulfil the secant equation if

\[
s_i^T y_i > 0,
\]

which is known as the curvature condition.

3. The New Hybrid Method

The modification of the quasi-Newton method based on a hybrid method has already been introduced by previous researchers. One of the studies is a hybridization of the quasi-Newton and Gauss-Seidel methods, aimed at solving the system of linear equations in [13]. Luo et al. [14] suggest the new hybrid method, which can solve the system of nonlinear equations by combining the quasi-Newton method with chaos optimization. Han and Neumann [6] combine the quasi-Newton methods and Cauchy descent method to solve unconstrained optimization problems, which is recognised as the quasi-Newton-SD method.

Hence, the modification of the quasi-Newton method by previous researchers spawned the new idea of hybridizing the classical method to yield the new hybrid method. Hence, this study proposes a new hybrid search direction that combines the concept of search direction of the quasi-Newton and CG methods. It yields a new search direction of the hybrid method which is known as the BFGS-CG method. The search direction for the BFGS-CG method is

\[
d_i = \begin{cases} 
eg H_i g_i, & i = 0, 
eg H_i g_i + \eta (-g_i + \beta_i d_{i-1}), & i \geq 1, \end{cases}
\]

where \(\eta > 0\) and \(\beta_i = (g_i^T g_{i-1} / g_{i-1}^T d_{i-1})\).

Hence, the complete algorithms for the BFGS method, CG-HS, CG-PR, and CG-FR methods, and the BFGS-CG method will be arranged in Algorithms 1, 2, and 3, respectively.
Algorithm 1 (BFGS method). States the following.
Step 0. Given a starting point $x_0$ and $H_0 = I_n$, choose values for $s$, $\beta$, and $\sigma$ and set $i = 1$.

Step 1. Terminate if $\|g(x_{i+1})\| < 10^{-6}$ or $i \geq 10000$.
Step 2. Calculate the search direction by (6).
Step 3. Calculate the step size $\alpha_i$ by (3).
Step 4. Compute the difference between $s_i = x_i - x_{i-1}$ and $y_i = g_i - g_{i-1}$.
Step 5. Update $H_{i+1}$ by (7) to obtain $H_i$.
Step 6. Set $i = i + 1$ and go to Step 1.

Algorithm 2 (CG-HS, CG-PR, and CG-FR). States the following.
Step 0. Given a starting point $x_0$, choose values for $s$, $\beta$, and $\sigma$ and set $i = 1$.
Step 1. Terminate if $\|g(x_{i+1})\| < 10^{-6}$ or $i \geq 10000$.
Step 2. Calculate the search direction by (4) with respect to the coefficient of CG.
Step 3. Calculate the step size $\alpha_i$ by (3).
Step 4. Compute the difference between $s_i = x_i - x_{i-1}$ and $y_i = g_i - g_{i-1}$.
Step 5. Set $i = i + 1$ and go to Step 1.

Algorithm 3 (BFGS-CG method). States the following.
Step 0. Given a starting point $x_0$ and $H_0 = I_n$, choose values for $s$, $\beta$, and $\sigma$ and set $i = 1$.
Step 1. Terminate if $\|g(x_{i+1})\| < 10^{-6}$ or $i \geq 10000$.
Step 2. Calculate the search direction by (10).
Step 3. Calculate the step size $\alpha_i$ by (3).
Step 4. Compute the difference between $s_i = x_i - x_{i-1}$ and $y_i = g_i - g_{i-1}$.
Step 5. Update $H_{i+1}$ by (7) to obtain $H_i$.
Step 6. Set $i = i + 1$ and go to Step 1.

Based on Algorithms 1, 2, and 3 we assume that every search direction $d_i$ satisfies the descent condition
$$
\begin{align*}
g_i^T d_i < 0,
\end{align*}
$$
for all $i \geq 0$. If there exists a constant $c_1 > 0$ such that
$$
\begin{align*}
g_i^T d_i \leq c_1 \|g_i\|^2
\end{align*}
$$
for all $i \geq 0$, then the search directions satisfy the sufficient descent condition which can be proved in Theorem 6. Hence, we need to make a few assumptions based on the objective function.

Assumption 4. Consider the following.

H1: the objective function $f$ is twice continuously differentiable.
H2: the level set $L$ is convex. Moreover, positive constants $c_1$ and $c_2$ exist, satisfying
$$
\begin{align*}
c_1 \|z\|^3 \leq z^T F(x) z \leq c_2 \|z\|^2,
\end{align*}
$$
for all $z \in \mathbb{R}^n$ and $x \in L$, where $F(x)$ is the Hessian matrix for $f$.
H3: the Hessian matrix is Lipschitz continuous at the point $x^*$; that is, there exists the positive constant $c_3$ satisfying
$$
\begin{align*}
\|g(x) - g(x^*)\| \leq c_3 \|x - x^*\|
\end{align*}
$$
for all $x \in$ a neighbourhood of $x^*$.

Theorem 5 (see [15, 16]). Let $\{B_i\}$ be generated by the BFGS formula (8), where $B_i$ is symmetric and positive definite, and $y_i^T s_i > 0$ for all $i$. Furthermore, assume that $\{s_i\}$ and $\{y_i\}$ are such that
$$
\begin{align*}
\frac{\|y_i - G_s s_i\|}{\|s_i\|} \leq \epsilon_i
\end{align*}
$$
for some symmetric and positive definite matrix $G(x^*)$ and for some sequence $\{\epsilon_i\}$ with the property $\sum_{i=0}^{\infty} \epsilon_i < \infty$. Then
$$
\begin{align*}
\lim_{i \to \infty} \frac{\|B_i - G_s\|_2}{\|d_i\|} = 0
\end{align*}
$$
and the sequence $\{\|B_i\|\}$, $\{\|B_i^{-1}\|\}$ are bound.

Theorem 6. Suppose that Assumption 4 and Theorem 5 hold. Then condition (12) holds for all $i \geq 0$.

Proof. From (12), we see that
$$
\begin{align*}
g_i^T d_i &= -g_i^T B_i^{-1} g_i + \eta g_i^T \left( -g_i + \frac{g_i^T g_{i-1}}{g_i^T d_{i-1}} d_{i-1} \right) \\
&= -g_i^T B_i^{-1} g_i + \eta \left( -g_i^T g_i + \frac{g_i^T g_{i-1}}{g_i^T d_{i-1}} g_i^T d_{i-1} \right) \\
&= -g_i^T B_i^{-1} g_i + \eta \left( -g_i^T g_i + g_i^T d_{i-1} \right). \\

\end{align*}
$$
Based on Powell [17], $g_i^T g_{i-1} \geq \epsilon \|g_i\|^2$ with $\epsilon = (0, 1]$, and
$$
\begin{align*}
g_i^T d_i &= -g_i^T B_i^{-1} g_i + \eta \left( -\|g_i\|^2 + \epsilon \|g_i\|^2 \right) \\
&\leq -\lambda \|g_i\|^2 + (\eta - \eta \epsilon) \|g_i\|^2 \\
&\leq c_1 \|g_i\|^2,
\end{align*}
$$
where $c_1 = -(\lambda + \eta - \eta \epsilon)$ which is bound away from zero. Hence, $g_i^T d_i \leq c_1 \|g_i\|^2$ holds. The proof is completed. □
Lemma 7. Under Assumption 4, positive constants $\omega_1$ and $\omega_2$ exist, such that for any $x_i$ and any $d_i$, with $g_i^T d_i < 0$, the step size $\alpha_i$ produced by Algorithm 2 will satisfy either

$$f(x_i + \alpha_i d_i) - f(x_i) \leq -\omega_1 \frac{(g_i^T d_i)^2}{\|d_i\|^2}$$

(19)

or

$$f(x_i + \alpha_i d_i) - f(x_i) \leq \omega_1 g_i^T d_i.$$  

(20)

Proof. Suppose that $\alpha_i < 1$, which means that (3) failed for a step size $\alpha^T \leq \alpha_i / \tau$:

$$f(x_i + \alpha_i^T d_i) - f(x_i) \leq \omega \alpha_i^T g_i^T d_i.$$  

(21)

Then, by using the mean value theorem, we obtain

$$f(x_{i+1}) - f(x_i) = \overline{g}^T(x_{i+1} - x_i),$$

(22)

where $\overline{g} = \nabla f(\overline{x})$, for some $\overline{x} \in (x_i, x_{i+1})$. Now, by the Cauchy-Schwartz inequality, we get

$$\overline{g}^T(x_{i+1} - x_i) = g_i^T(x_{i+1} - x_i) + (\overline{g} - g_i)^T(x_{i+1} - x_i)$$

$$= g_i^T(x_{i+1} - x_i) + \|\overline{g} - g_i\|\|x_{i+1} - x_i\|$$

$$\leq g_i^T(x_{i+1} - x_i) + \mu \|x_{i+1} - x_i\|^2$$

$$\leq g_i^T(\alpha_i^T d_i) + \mu \|\alpha_i^T d_i\|^2$$

$$\leq g_i^T(\alpha_i^T d_i) + \mu (\|\alpha_i^T d_i\|^2).$$

(23)

Thus, from H3

$$(\omega - 1) \alpha_i^T g_i^T d_i < \alpha_i^T (\overline{g} - g_i)^T d_i \leq M \|\alpha_i^T d_i\|^2,$$  

(24)

which implies that

$$\alpha_i \geq \tau \alpha_i > \tau (1 - \omega) \frac{-g_i^T d_i}{M(\|\alpha_i^T d_i\|^2)}.$$  

(25)

Substituting this into (21), we have

$$f(x_i + \alpha_i^T d_i) - f(x_i) \leq \alpha_i \frac{-g_i^T d_i}{(\alpha_i^T d_i)}.$$  

(26)

where $\alpha_i = \tau(1 - \omega)/M$, which gives (19).

Theorem 8 (global convergence). Suppose that Assumption 4 and Theorem 5 hold. Then

$$\lim_{i \to \infty} g_i^T = 0.$$  

(27)

Proof. Combining descent property (12) and Lemma 7 gives

$$\sum_{i=0}^{\infty} \|g_i\|^2 < \infty.$$  

(28)

Hence, from Theorem 6, we can define that $\|d_i\| \leq -c\|g_i\|$. Then, (28) will be simplified as $\sum_{i=0}^{\infty} \|g_i\|^2 < \infty$. Therefore, the proof is completed.

4. Numerical Result

In this section, we use the test problem considered by Andre [18], Michalewicz [19], and Moré et al. [20] in Table 1 to analyse the improvement of the BFGS-CG method compared with the BFGS method and CG method. Each of the test problems is tested with dimensions varying from 2 to 1,000 variables. This represents a total of 159 test problems. As suggested by [20], for each of the test problems, the initial point $x_0$ will further subtract from the minimum point. In doing so, this leads us to test the global convergence properties and the robustness of our method. For the Armijo line search, we use $s = 1$, $\beta = 0.5$, and $\sigma = 0.1$. The stopping criteria we use are $\|g_i\| < 10^{-6}$ and the number of iterations exceeds its limit, which is set to be 10,000. In our implementation, the numerical tests were performed on an Acer Aspire with a Windows 7 operating system and using Matlab 2012 languages.

The performance results will be shown in Figures 1 and 2, respectively, using the performance profile introduced by Dolan and Moré [21]. The performance profile seeks to find how well the solvers perform relative to the other solvers on
Table 1: A list of problem functions.

<table>
<thead>
<tr>
<th>Test problem</th>
<th>Dimensional</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Powell badly scaled</td>
<td>2</td>
<td>Moré et al. [20]</td>
</tr>
<tr>
<td>Beale</td>
<td>2</td>
<td>Moré et al. [20]</td>
</tr>
<tr>
<td>Biggs Exp 6</td>
<td>6</td>
<td>Moré et al. [20]</td>
</tr>
<tr>
<td>Chebyquad</td>
<td>4, 6</td>
<td>Moré et al. [20]</td>
</tr>
<tr>
<td>Colville polynomial</td>
<td>4</td>
<td>Michalewicz [19]</td>
</tr>
<tr>
<td>Variably dimensioned</td>
<td>4, 8</td>
<td>Moré et al. [20]</td>
</tr>
<tr>
<td>Freudenstein and Roth</td>
<td>2</td>
<td>Moré et al. [20]</td>
</tr>
<tr>
<td>Goldstein price polynomial</td>
<td>2</td>
<td>Michalewicz [19]</td>
</tr>
<tr>
<td>Himmelblau</td>
<td>2</td>
<td>Andrei [18]</td>
</tr>
<tr>
<td>Penalty 1</td>
<td>2, 4</td>
<td>Moré et al. [20]</td>
</tr>
<tr>
<td>Extended Powell singular</td>
<td>4, 8</td>
<td>Moré et al. [20]</td>
</tr>
<tr>
<td>Extended Rosenbrock</td>
<td>2, 10, 100, 200, 500, 1000</td>
<td>Andrei [18]</td>
</tr>
<tr>
<td>Trigonometric</td>
<td>6</td>
<td>Andrei [18]</td>
</tr>
<tr>
<td>Watson</td>
<td>4, 8</td>
<td>Moré et al. [20]</td>
</tr>
<tr>
<td>Six-hump camel back polynomial</td>
<td>2</td>
<td>Michalewicz [19]</td>
</tr>
<tr>
<td>Extended shallow</td>
<td>2, 4, 10, 100, 200, 500, 1000</td>
<td>Andrei [18]</td>
</tr>
<tr>
<td>Extended strait</td>
<td>2, 4, 10, 100, 200, 500, 1000</td>
<td>Andrei [18]</td>
</tr>
<tr>
<td>Scale</td>
<td>2</td>
<td>Michalewicz [19]</td>
</tr>
<tr>
<td>Raydan 1</td>
<td>2, 4</td>
<td>Andrei [18]</td>
</tr>
<tr>
<td>Raydan 2</td>
<td>2, 4</td>
<td>Andrei [18]</td>
</tr>
<tr>
<td>Diagonal 3</td>
<td>2</td>
<td>Andrei [18]</td>
</tr>
<tr>
<td>Cube</td>
<td>2, 10, 100, 200</td>
<td>Moré et al. [20]</td>
</tr>
</tbody>
</table>

A set of problems. In general, $P(\tau)$ is the fraction of problems with performance ratio $\tau$; thus, a solver with high values of $P(\tau)$ or one that is located at the top right of the figure is preferable.

Figures 1 and 2 show that the BFGS-CG method has the best performance since it can solve 99% of the test problems compared with the BFGS (84%), CG-HS (65%), CG-PR (80%), and CG-FR (75%) methods. Moreover, we can also say that the BFGS-CG is the fastest solver on approximately 68% of the test problems for iteration and 52% of CPU time.

5. Conclusion

We have presented a new hybrid method for solving unconstrained optimization problems. The numerical results for a broad class of test problems show that the BFGS-CG method is efficient and robust in solving the unconstrained optimization problem. We also note that, as the size and complexity of the problem increase, greater improvements could be realised by our BFGS-CG method. Our future research will be to try the BFGS-CG method with coefficients of CG like Fletcher-Reeves, Hestenes-Stiefel, and Polak-Ribiére.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research was supported by Fundamental Research Grant Scheme (FRGS Vote no. 59256).

References


Submit your manuscripts at http://www.hindawi.com