Performance of New Line Search Methods with Three-Term Hybrid Descent Approach for Unconstrained Optimization Problems

Okundalaye Oluwaseun Olumide \(^a\)

\(^a\)Department of Mathematical Sciences, Faculty of Science, Adekunle Ajasin University, Akungba-Akoko (AAUA), P.M.Box 01, Akungba-Akoko, Ondo State, Nigeria.

Author’s contribution

The sole author designed, analysed, interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/ARJOM/2022/v18i230358

Open Peer Review History:

This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/74877

Received: 20 December 2021
Accepted: 24 January 2022
Published: 26 February 2022

Original Research Article

Abstract

In this paper, I demonstrates the performance and efficiency of new line search methods with three-term hybrid descent method for the solution of unconstrained optimization problems (UOPs). The techniques advanced the sustainable range of step-length to a broader level than the previous studies and gave a suitable initial step-length at each step of the iterations. The global convergence rate of the new line with three-term hybrid descent method search is carried studied. Some numerical results through performance profile shows that among the new search method modified Wolfe line search method in CPU time and iterations is best in practical computation.

Keywords: Quasi-Newton method; search direction; step-length; global convergence; performance profile.
1 Introduction

Considering an unconstrained optimization problem of the form

$$\min_{x \in \mathbb{R}^n} f(x).$$

where $\mathbb{R}^n$ is the n-dimensional Euclidean space and $f : \mathbb{R}^n \to \mathbb{R}$ is continuous differentiable ($C^2$).

The solution of (1.1) require using an iterative methods with a starting initial point $x_0$ to obtain a point of sequence $\{x_k\}$, $k = 1, 2, 3...n$, and show progressive approximations to the required solution using the iterative technique below

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, ...n.$$  \hspace{1cm} (1.2)

Where $x_k$ is the current iterate, $d_k$ is the search direction, and $\alpha_k$ the step-length. Let $(x^*)$ be a minimizer of (1.1) and thus be a stationary point that satisfies $g(x^*) = 0$. We denote $f(x_k)$ by $f_k$, $f_{x_k}$ by $f_x$, and $\nabla f(x_k)$ by $g_k$ respectively. The line search method required two step at each iterations; the first step is to find a search direction $d_k$ and the second step is to select the step-length $\alpha_k$ along the search direction. On the other hand, the $d_k$ is typically needed to satisfy the descent condition $g^T_k d_k < 0$ that guarantees $d_k$ is a descent direction of $f(x)$ at $x_k$. From the past studies and literature, search direction perform an essential role in line search techniques and that the step-length methods mainly guarantee global convergence.

The following condition holds in order to obtain the global convergence of the line search methods.

$$-\frac{g_k^T d_k}{\|g_k\|\|d_k\|} \geq c.$$  \hspace{1cm} (1.3)

where $c \in (0,1]$ is a constant. The condition (1.3) is called angle property. Several methods of selecting $d_k$ and $\alpha_k$ yield different convergence properties. Also, for the step-length, the sequence of iterates $x_k$ given by (1.2) converges globally with additional rate of convergence.

Exact line search and inexact line search are the notable two ways of determining the step-length. For the exact line search, $\alpha_k$ is obtained by using the following approach

$$\alpha_k = \arg \min_{\alpha > 0} \{f(x_k + \alpha d_k)\}.$$  \hspace{1cm} (1.4)


This work focused on new inexact line search rule called modified line search rules that advanced the scope of appropriate step-length and give a good initial step-length at each iteration.

Modified Armijo rule.

Set scalar $l_k > 0, \beta \in (1,0), \sigma \in (0, \frac{1}{2})$, and set $S_k = -\frac{g_k^T d_k}{l_k \|d_k\|^2}$. Let $\alpha_k$ be the largest $\alpha$ in $\{s_k, \beta s_k, \beta^2 s_k, ...\}$ such that

$$f_k - f(x_k + \alpha d_k) \geq \sigma \alpha \|d_k\|w_k(\alpha),$$  \hspace{1cm} (1.5)
Modified Goldstein rule.
A fixed scalar $\sigma = (0, \frac{1}{2})$ is selected and $\alpha_k$ is chosen to satisfy

$$-(1-\sigma)\alpha_k g_k^T d_k \geq f_k - f(x_k + \alpha_k d_k) \geq \sigma \alpha_k ||d_k||w_k(\alpha_k),$$

(1.6)

Modified Wolfe rule.
The step $\alpha_k$ is chosen to satisfy

$$f_k - f(x_k + \alpha_k d_k) \geq \sigma \alpha_k ||d_k||w_k(\alpha_k),$$

and

$$g(x_k + \alpha_k d_k)^T d_k \geq y g_k^T d_k,$$

(1.7)

where $\sigma$ and $y$ are some scalars with $\sigma \in (0, \frac{1}{2})$ and $y \in (0, 1)$ for $k = 0, 1, 2, \ldots n$. Then, the sequence of $\{x_k\}_{k=0}^\infty$ converges to the optimal point $x^*$ which minimizes $f(x)$. Hence, modified Armijo, Goldstein, and Wolfe line search methods are used in this research associated with three-term hybrid descent search direction.

This work is prepared as follows; In section (2), I illustrate and discussed extensively the importance of search direction in iterative method. The new three-term hybrid method and its convergence analysis are discussed in section (3). Numerical results and discussion are given in section (4) and short conclusion in section (5).

## 2 The Use of Search Direction

In an iterative method of solving an unconstrained optimization problem, search direction is most important and essential which includes; conjugate gradient method (CGM), Newton method (NM), and quasi-Newton method (QNM). Many of these techniques used in solving UOPs relies solely on the results of search direction $d_k$. The conjugate gradient (CG) approach contain a class of UOPs algorithms with a properties of low memory, easy computation and global strong convergence, making them efficient for solving large-scale problems in the form of $\min_{x \in \mathbb{R}^n} f(x)$ with the differentiable non-linear function $f: \mathbb{R}^n \to \mathbb{R}$. The CG method is also uses to obtain the minimum value of functions of an UOPs. \cite{10} introduced a CG method of providing solution to a linear system of equations, after which \cite{11} used the CG method to solve UOPs. Recently, CG methods is becoming more popular iterative methods to solve large-scale UOPs, since they do not required the storage of matrices \cite{12, 13, 14, 15, 16, 17, 18}. The conjugate gradient search direction approach is defined by

$$d_k = \begin{cases} -g_k & k = 0, \\ -g_k + \beta_k d_{k-1} & k \geq 1. \end{cases}$$

(2.1)

where $g_k = \nabla f(x_k)$ and $\beta_k$ is known as the CG coefficient. We have several ways of calculating $\beta_k$ and some well-known formulae are;

$$\beta_k^{FR} = \frac{g_k^T g_k}{||g_{k-1}||^2},$$

$$\beta_k^{PR} = \frac{g_k^T (g_k - g_{k-1})}{||g_{k-1}||^2},$$

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}},$$

$$\beta_k^{BAN} = \frac{g_k^T (g_k - g_{k-1})}{g_k^T g_k - (g_{k-1})^T (g_{k-1})},$$

$$\beta_k^{HZ} = \left( g_k - \frac{2d_k (||g_k||^2)}{(d_k g_k)^2} \right)^T \left( g_{k+1} + \frac{2d_k (||g_k||^2)}{(d_k g_k)^2} \right).$$
where \( g_k \) and \( g_{k-1} \) are gradients of \( f(x) \) at the points \( x_k \) and \( x_{k-1} \) and \( y_k = g_k - g_{k-1} \) respectively. While \( ||.|| \) is a norm of vectors and \( d_{k-1} \) is a direction for the previous iteration. The above corresponding coefficients are known as \( \{\text{Fletcher and Reeves, 1964}, \text{Polak and Ribiere, 1969}\} \) and \( \{\text{Hestenes and Stiefel, 1952}\} \) and \([\text{19, 20, 21, 22, 23, 24}]\). Recently, \([\text{25, 26, 27, 28, 29}]\) proposed three-term CG methods which satisfy the sufficient descent condition.

\[
g_T^k d_k \leq -\bar{c}||g_k||^2 \quad \text{for all} \quad k = 0, 1, 2...n. \tag{2.2}\]

and a positive constant \( \bar{c} \), independently of line searches. They proposed the modified FR method defined by

\[
d_k = -\theta_k g_k + \beta^{FR} d_{k-1} \]

where \( \theta_k = \frac{d_k^T y_{k-1}}{||y_{k-1}||^2} \). Since this search direction satisfies \( g_T^k d_k < -||g_k||^2 \) for all \( k \), it can be written by the three-term form:

\[
d_k = -g_k + \beta^{FR} d_{k-1} - \theta_k^1 g_k, \tag{2.3}\]

where \( \theta_k^1 = \frac{g_k^T d_{k-1}}{||y_{k-1}||^2} \). They also proposed the modified PR methods and the modified HS method, which are respectively given by

\[
d_k = -g_k + \beta^{PR} d_{k-1} - \theta_k^{(2)} y_{k-1}, \tag{2.4}\]

\[
d_k = -g_k + \beta^{HS} d_{k-1} - \theta_k^{(3)} y_{k-1}. \tag{2.5}\]

Where \( \theta_k^{(2)} = \frac{g_k^T d_{k-1}}{||y_{k-1}||^2} \) and \( \theta_k^{(3)} = \frac{g_k^T d_{k-1}}{y_k^T y_{k-1}} \), Cheng gave another modification of PR method:

\[
d_k = -g_k + \beta_k^{PR} (I - \frac{g_k y_k^T}{g_k^T y_k}) d_{k-1} = -g_k + \beta^{PR} d_{k-1} - \beta^{PR}_k \frac{g_k^T d_{k-1}}{g_k^T g_k} y_k. \tag{2.6}\]

They obtained their global convergence properties under appropriate line searches. The recent modification can be seen \([\text{8, 30, 31, 32, 27}]\). We observe that these approach always satisfy \( g_T^k d_k = -||g_k||^2 < 0 \) for all \( k \), which indicate the sufficient descent condition with \( \bar{c} = 1 \).

In quasi-Newton (QN) family, the search direction given below is the solution of it’s linear system

\[
d_k = -H_k g_k. \tag{2.7}\]

where \( H_k \) is the Hessian approximation. The initial matrix give as \( H_0 \) is selected by the identity matrix \( I \), which can be obtained by an update formula. There are a few update formulae that are widely used like Davidon-Fletcher-Powell\( (DFP) \), BFGS, and Broyden family formula. This study employs a BFGS formula in a classical algorithm and the new hybrid method. The update formula for BFGS is

\[
H_{k+1} = H_k - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}, \tag{2.8}\]

with \( s_k = x_k - x_{k+1} \) and \( y_k = g_k - g_{k-1} \). The approximation that the Hessian must fulfil is

\[
H_{k+1} s_k = y_k, \tag{2.9}\]

The above condition is essential to hold for the updated matrix \( H_{k+1} \). Note that it is only feasible to fulfil the equations if

\[
s_k^T y_k > 0. \tag{2.10}\]

which is called the curvature condition.
3 The Proposed Three-Term Method

The modification on three-term approach have been proposed by many researchers; One of the research is by Ludwig [33] which is a hybrid between quasi-Newton methods with Gauss-Siedel method to solve the system of linear equation. Then, Luo et al. [34] suggested the new hybrid method which can provide solution to system of non-linear equations using quasi-Newton method with chaos optimization. Besides, Han and Newman [35] used Quasi-Newton methods and Cauchy descent method to solve UOPs. Also Ibrahim et al. [36] proposed BFGS-CG method which is between quasi-Newton and conjugate gradient method and come out with this search direction.

\[ d_k = \begin{cases} -H_k g_k & k = 0, \\ -H_k g_k + \eta (-g_k + \beta_k d_{k-1}) & k \geq 1. \end{cases} \]

where \( \eta > 0 \) \( \beta_k = \frac{g_k^T g_k - 1}{g_k^T d_k - 1} \).

Hence, the advancement on Quasi-Newton by previous authors provide new idea on hybrid. Therefore, a new hybrid search direction which hybridize the search direction of quasi-Newton and conjugate gradient method is establish. This gave a new search direction technique known as Three-term BFGS-CG method.

Search direction for Three-term BFGS-CG method

\[ d_k = \begin{cases} -H_k g_k & k = 0, \\ -H_k g_k + \eta (-g_k + \beta_k d_{k-1}) - \beta_k \frac{g_k^T g_k - 1}{g_k^T d_k - 1} d_k & k \geq 1. \end{cases} \] (3.1)

Hence, the complete algorithms for BFGS, (CG-HS, CG-PR, CG-FR), and Three-Term BFGS-CG method will be arranged in Algorithm(16), Algorithm(17) and (18) respectively.

Algorithm(1) Modified Armijo line search (MALS) for Three-term BFGS-CG.

Step 1. From an initial starting point \( x_0 \) and \( H_0 = I_n \), select values for \( s, \beta, \sigma \), and set \( k = 1 \).

Step 2. End if \( \|g(x_{k+1})\| < 10^{-6} \) or \( k \leq 1000 \).

Step 3. Compute the search direction using (3.1).

Step 4. Compute the step-length \( \alpha_k \) using (1.5).

Step 5. Calculate the difference \( s_k = x_k - x_{k-1} \) and \( y_k = g_k - g_{k-1} \).

Step 6. Substitute \( H_{k+1} \) by (12) to obtain \( H_k \).

Step 7. Set \( k = k + 1 \) and go to step 2.

Algorithm(2) Modified Goldstein line search (MGLS) for Three-term BFGS-CG method.

Step 1. From an initial starting point \( x_0 \) and select values for \( s, \beta, \sigma \), and set \( k = 1 \).

Step 2. End if \( \|g(x_{k+1})\| < 10^{-6} \) or \( k \leq 1000 \).

Step 3. Compute the search direction using (3.1).

Step 4. Compute the step size \( \alpha_k \) using (1.6).

Step 5. Calculate the difference \( s_k = x_k - x_{k-1} \) and \( y_k = g_k - g_{k-1} \).

Step 6. Substitute \( H_{k+1} \) by (2.8) to obtain \( H_k \).

Step 7. Set \( k = k + 1 \) and go to step 2.

Algorithm(3) Modified Wolfe line search (MWLS) for Three-term BFGS-CG method.

Step 1. From an initial starting point \( x_0 \) and \( H_0 = I_n \), select values for \( s, \beta, \sigma \), and set \( k = 1 \).

Step 2. End if \( \|g(x_{k+1})\| < 10^{-6} \) or \( k \leq 1000 \).

Step 3. Compute the search direction using (3.1).
Step 4. Compute the step length \( \alpha_k \) using (1.7).
Step 5. Calculate the difference between \( s_k = x_k - x_{k-1} \) and \( y_k = g_k - g_{k-1} \).
Step 6. Substitute \( H_{k+1} \) by (2.8) to obtain \( H_k \).
Step 7. Set \( k = k + 1 \) and go to step 2.

From the three algorithms we assume that each \( d_k \) satisfied the descent condition \( g_k^T d_k < 0 \).

The following few assumption are make concerning objective function

Assumption 3.1
H1: The objective function \( f \) is twice \( C^2 \).
H2: The level set \( L \) is a convex. Also, positive constants \( c_1 \) and \( c_2 \) exist, satisfying

\[
|g(x) - g(x^*)| \leq c_3|x - x^*|.
\]

for all \( x \in \mathbb{R}^n \) and \( x \in L \) where \( f(x) \) is the Hessian matrix (H) of \( f \).
H3: The \( H \) is Lipschitz continuous at the point \( x^* \) that is, there exist the positive constant \( c_3 \)
satisfying

\[
|g(x) - g(x^*)| \leq \epsilon_{c_k} c_3|x - x^*|.
\]

for all \( x \) in a neighbourhood of \( x^* \)

Theorem (3.2). Let \( \{B_k\} \) be generated by BFGS (2.8), where \( B_k \) is symmetric and positive definite, and \( y_k^T s_k > 0 \) for \( k \). Also, assume that \( \{s_k\} \) and \( \{y_k\} \) are such that

\[
\frac{||y_k - G_k s_k||}{||s_k||} \leq \epsilon_k.
\]

for some symmetric and definite matrix \( G(x^*) \) and for some sequence \( \epsilon_k \) with the property. \( \sum_{k=1}^{\infty} \epsilon_k < \infty \). Then

\[
\lim_{k \to \infty} \frac{||(B_k - G_k d_k)||}{||d_k||} = 0.
\]

and the sequence \( ||\{B_k\}||, ||\{B_k^{-1}\}|| \) are bounded.

Theorem (3.3). Global convergence.

Given that Assumption (3.1) and Theorem(3.2) hold. Then

\[
\lim_{k \to \infty} ||g_k||^2 = 0.
\]

Proof.
from the condition \( g_k^T d_k < 0 \), we see that

\[
g_k^T d_k = -g_k^T B_k^{-1} g_k + \eta g_k^T (-g_k + \beta_k d_{k-1} - \beta_k g_k^T d_{k-1} g_k).
\]

(3.7)

\[
= -g_k^T B_k^{-1} g_k + \eta (-g_k^T g_k + g_k^T g_k + g_k^T d_{k-1} g_k - g_k^T g_k + g_k^T d_{k-1} g_k + g_k^T g_k + g_k^T d_{k-1} g_k + g_k^T g_k)
\]

(3.8)

\[
= -g_k^T B_k^{-1} g_k + \eta (-g_k^T g_k + g_k^T g_k + g_k^T g_k - g_k^T g_k).
\]

(3.9)

then

\[
g_k^T d_k = -g_k^T B_k^{-1} g_k + \eta (-||g_k||^2),
\]

(3.10)

\[
\leq -\lambda_k ||g_k||^2 + \eta (-||g_k||^2),
\]

(3.11)

\[
g_k^T d_k \leq c_1 ||g_k||^2,
\]

(3.12)
where \( c_1 = -(\lambda_k + \eta) \) which is bounded away from zero. Hence, from the Armijo line search condition, we have that

\[
f_k - f_{k+1} \leq \sigma \alpha_k g_k^T d_k,
\]

\[
\leq \sigma \alpha_k c_1 \| g_k \|^2,
\]

holds for all \( k \). Since \( f_k \) is decreasing and the sequence \( \{ f_k \} \) is bounded below by \( H_2 \), we have that

\[
\lim_{k \to \infty} (f_k - f_{k+1}) = 0,
\]

Hence, this (3.14) and (3.15) imply

\[
\lim_{k \to \infty} \| g_k \|^2 = 0.
\]

4 Numerical Results and Discussion

Some selected unconstrained optimization problems from the CUTEr suite were chosen as given on Table 1 to analyse the efficiency and performance of the new line search methods with three-term hybrid descent method. Each of the selected UOPs were tested with dimensions varying from 2 to 1000. For each of the test problems, the initial starting point \( x_0 \) take further away from the minimum point which leads to test the global convergence properties and the robustness of the new line search methods. For the MALS, MGLS, and MWLS using \( \sigma = (\alpha, \frac{1}{2}) \), stopping criteria \( \| g_k \| \leq 10^{-6} \), and the number of iterations exceeds a limit of 10,000. In general \( p(\tau) \) is the fraction of problems with performance ratio \( \tau \); thus, a solver with high values of \( p(\tau) \) is desirable. Performance profile were drawn to see how well the new line search method perform. The performance profile seeks to find how well the solvers perform relative to the other solvers on a set of problems. The implementations and numerical tests was performed on Matlab 2021a software. Table 1 show the unconstrained optimization test problems used to test the efficiency of the proposed line search methods. Performance profiles of methods are illustrated in Figs. 1 and 2. From Figs. 1 and 2, three-term hybrid modified Wolfe line search approach has the best performance since it can solve (97%) of the test problems compared with the three-term hybrid modified Armijo line of search(77%), and three-term hybrid modified Goldstein line of search(62%).

<table>
<thead>
<tr>
<th>Test Problems</th>
<th>n-dimension</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Powell badly scaled</td>
<td>2</td>
<td>More et al.</td>
</tr>
<tr>
<td>Beale</td>
<td>2</td>
<td>More et al.</td>
</tr>
<tr>
<td>Higgs Exp</td>
<td>6, 6</td>
<td>More et al.</td>
</tr>
<tr>
<td>Chebyshev polynomial</td>
<td>4</td>
<td>More et al.</td>
</tr>
<tr>
<td>Colville polynomial</td>
<td>4</td>
<td>Michalewicz</td>
</tr>
<tr>
<td>Variably dimensioned</td>
<td>4, 8</td>
<td>More et al.</td>
</tr>
<tr>
<td>Freudenstein and Roth</td>
<td>2</td>
<td>More et al.</td>
</tr>
<tr>
<td>Goldstein price polynomial</td>
<td>4</td>
<td>Michalewicz</td>
</tr>
<tr>
<td>Himemstein</td>
<td>2</td>
<td>Andrei</td>
</tr>
<tr>
<td>Penalty</td>
<td>1, 2, 4</td>
<td>More et al.</td>
</tr>
<tr>
<td>Extended Powell singular</td>
<td>4, 8</td>
<td>More et al.</td>
</tr>
<tr>
<td>Extended Rosenbrock</td>
<td>2, 10, 100, 200, 500, 1000</td>
<td>Andrei</td>
</tr>
<tr>
<td>Airhood</td>
<td>10,50,100,500,1000</td>
<td>Andrei</td>
</tr>
<tr>
<td>FSC 1</td>
<td>2</td>
<td>More et al.</td>
</tr>
<tr>
<td>Six-hump camel back polynomial</td>
<td>2</td>
<td>Michalewicz</td>
</tr>
<tr>
<td>Extended Cliff</td>
<td>2, 4, 10, 100, 200, 500, 1000</td>
<td>Andrei</td>
</tr>
<tr>
<td>Extended Hobart</td>
<td>2, 4, 10, 100, 200, 500, 1000</td>
<td>Andrei</td>
</tr>
<tr>
<td>Extended EP1</td>
<td>2,4,10</td>
<td>Michalewicz</td>
</tr>
<tr>
<td>Raylan</td>
<td>1, 2, 4</td>
<td>Andrei</td>
</tr>
<tr>
<td>Raylan</td>
<td>2, 2, 4</td>
<td>Andrei</td>
</tr>
<tr>
<td>Dingunet</td>
<td>4</td>
<td>Andrei</td>
</tr>
<tr>
<td>Cube</td>
<td>2, 10, 100, 200</td>
<td>More et al.</td>
</tr>
</tbody>
</table>
5 Conclusion

In summary, I presented performance and efficiency of a several line search methods with three-term BFGS-CG descent search for solving unconstrained optimization problems that guaranteed sufficient descent condition. I deduced that modified Wolfe line search technique perform best on performance profile as shown on the Figures. I also concluded that forming an hybrid method out of the existing methods is more efficient especially when the strength of the component are the target of the hybridization. Researchers and authors can implement modified Wolfe line search for large scale unconstrained optimization problem.
Competing Interests

Author has declared that no competing interests exist.

References


© 2022 Olumide; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar) https://www.sdiarticle.com/review-history/74877