Stud. Univ. Babeş-Bolyai Math. 67(2022), No. 4, 855–869
DOI: 10.24193/subbmath.2022.4.14

Hybrid conjugate gradient-BFGS methods based on Wolfe line search

Khelladi Samia and Benterki Djamel

Abstract. In this paper, we present some hybrid methods for solving unconstrained optimization problems. These methods are defined using proper combinations of the search directions and included parameters in conjugate gradient and quasi-Newton method of Broyden–Fletcher–Goldfarb–Shanno (CG-BFGS). Their global convergence under the Wolfe line search is analyzed for general objective functions. Numerical experiments show the superiority of the modified hybrid (CG-BFGS) method with respect to some existing methods.

Mathematics Subject Classification (2010): 65K05, 90C26, 90C30.

Keywords: Unconstrained optimization, global convergence, conjugate gradient methods, quasi-Newton methods, Wolfe line search.

1. Introduction

Conjugate gradient methods are very important ones for solving unconstrained optimization problems, especially for large scale problems. It is well known that Fletcher-Reeves (FR) [7], Conjugate Descent (CD) [6] and Dai-Yuan (DY) [4] conjugate gradient methods have strong convergence properties, but they may not perform well in practice. On the other hand, Hestnes-Stiefel (HS) [9], Polak-Ribiere-Polyak (PRP) [13, 14] and Liu-Storey (LS) [12] conjugate gradient methods may not converge in general, but they often perform better than FR, CD and DY. To combine the best numerical performances of the LS method and the global convergence properties of the CD method, Yang et al. [17] proposed a hybrid LS-CD method. Dai and Liao [3] proposed an efficient conjugate gradient method (Dai-Liao type method). Later, some more efficient Dai-Liao type conjugate gradient method, known as DHSDL and DLSDL were proposed in [21].

The rest of this paper is organized as follows. In Section 2, we give various possibilities to determine the step size and the search direction. A hybridization of

Received 23 December 2019; Accepted 08 February 2020.

the conjugate gradient method (CG) and the BFGS method will also be presented. In Section 3, we consider the modification of LSCD method, termed as MLSCD and the modification of (DHSDL and DLSDL) termed as MMDL [15] and we prove the global convergence using the Wolfe line search instead of backtracking line search used by the authors in [15]. In Section 4, we consider the hybrid method BFGS-CG termed as H-BFGS-CG1 in [15] and we prove the global convergence with the Wolfe line search termed WH-BFGS-CG. In section 5, we report some numerical results and compare the performance of the different considered methods. Finally, we give some conclusions to end this paper.

2. Preliminaries

Consider the following unconstrained optimization problem

$$\min f(x), \qquad x \in \mathbb{R}^n, \tag{2.1}$$

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a continuously differentiable function. Let g_k be the gradient of $f(x)$ at the current iterative point x_k, then the classical conjugate gradient method for (2.1) is given by

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2.2}$$

in which $\alpha_k > 0$ is the step size found by one of the line search methods, and d_k is the search direction defined by

$$d_k = \begin{cases} -g_0, & k = 0, \\ -g_k + \beta_k d_{k-1}, & k \geq 1, \end{cases} \tag{2.3}$$

where β_k is an appropriately defined real scalar, known as the conjugate gradient parameter.

Since Fletcher and Reeves introduced the nonlinear conjugate gradient method in 1964, many formulae have been proposed using various modifications of the conjugate gradient direction d_k and the parameter β_k. The most popular parameters β_k are:

$$\begin{aligned}
\beta_k^{FR} &= \frac{\| g_k \|^2}{\| g_{k-1} \|^2}, \quad \beta_k^{CD} = -\frac{\| g_k \|^2}{g_{k-1}^T d_{k-1}}, \quad \beta_k^{DY} = \frac{\| g_k \|^2}{y_{k-1}^T d_{k-1}}, \\
\beta_k^{HS} &= \frac{g_k^T y_{k-1}}{y_{k-1}^T d_{k-1}}, \quad \beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\| g_{k-1} \|^2}, \quad \beta_k^{LS} = -\frac{g_k^T y_{k-1}}{g_{k-1}^T d_{k-1}}, \\
\beta_k^{DHSDL} &= \frac{\| g_k \|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} \mid g_k^T g_{k-1} \mid}{\mu \mid g_k^T d_{k-1} \mid + d_{k-1}^T y_{k-1}} - t\frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}, \quad \mu > 1, \ t > 0, \\
\beta_k^{DLSDL} &= \frac{\| g_k \|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} \mid g_k^T g_{k-1} \mid}{\mu \mid g_k^T d_{k-1} \mid - d_{k-1}^T g_{k-1}} - t\frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}, \quad \mu > 1, \ t > 0,
\end{aligned}$$

where

$$y_{k-1} = g_k - g_{k-1}, \ s_{k-1} = x_k - x_{k-1}$$

and $\| \cdot \|$ denotes the Euclidean vector norm.

In this paper, the step size α_k is determined using the following Wolfe line search conditions

$$\begin{aligned} f(x_k+\alpha_k d_k) &\le f(x_k)+\rho\alpha_k g_k^T d_k,\\ g_{k+1}^T d_k &\ge \sigma g_k^T d_k,\ \ 0<\rho<\sigma<1. \end{aligned} \tag{2.4}$$

To combine the best numerical performances of the PRP method and the global convergence properties of the FR method, Touati-Ahmed and Storey [16] proposed a hybrid PRP-FR method which is called the H1 method in [19], with the gradient parameter is defined as

$$\beta_k^{H1}=\max\{0,\min\{\beta_k^{PRP},\beta_k^{FR}\}\}. \tag{2.5}$$

Gilbert and Nocedal in [8] modified (2.5) to

$$\beta_k=\max\{-\beta_k^{FR},\min\{\beta_k^{PRP},\beta_k^{FR}\}\}.$$

A hybrid HS-DY conjugate gradient method was proposed by Dai and Yuan in [5], termed as the H2 method in [19] where the gradient parameter is defined as

$$\beta_k^{H2}=\max\{0,\min\{\beta_k^{HS},\beta_k^{DY}\}\}. \tag{2.6}$$

We consider hybrid CG methods where the search direction d_k, $k\ge 1$, from (2.3) is modified using one of the following tow rules [15]

$$d_k=\mathcal{D}(\beta_k,g_k,d_{k-1})=-\left(1+\beta_k\frac{g_k^T d_{k-1}}{\|g_k\|^2}\right)g_k+\beta_k d_{k-1} \tag{2.7}$$

$$d_k=\mathcal{D}_1(\beta_k,g_k,d_{k-1})=-B_k g_k+\mathcal{D}(\beta_k,g_k,d_{k-1}) \tag{2.8}$$

and the conjugate gradient parameter β_k is defined using some proper combinations of the parameters β_k given above and already defined hybridizations of these parameters.

Zhang et al. in [20, 18] proposed a modification to the FR method, termed as the MFR method, using the search direction

$$d_k=\mathcal{D}(\beta_k^{FR},g_k,d_{k-1}) \tag{2.9}$$

Zhang in [18] also proposed a modified DY method, which is known as the MDY method, using the search direction

$$d_k=\mathcal{D}(\beta_k^{DY},g_k,d_{k-1}) \tag{2.10}$$

The MFR and MDY methods posses very useful property

$$g_k^T d_k=-\|g_k\|^2 \tag{2.11}$$

If the exact line search is used, then MFR and the MDY methods reduce to the FR and the DY methods, respectively.

The MFR method has proven to be globally convergent for non convex functions with the Wolfe line search or the Armijo line search, and it is very efficient in real computations [20].

However, it is not known whether the MDY method converges globally. So, in [19], the authors replaced β_k^{FR} in (2.9) and β_k^{DY} in (2.10) by β_k^{H1} and β_k^{H2}, respectively. Then, they defined new hybrid PRP-FR and HS-DY methods, which they call the NH1 method and the NH2 method, respectively. These methods are based on the search directions

$$\text{NH1}: d_k = \mathcal{D}(\beta_k^{H1}, g_k, d_{k-1}) \tag{2.12}$$

$$\text{NH2}: d_k = \mathcal{D}(\beta_k^{H2}, g_k, d_{k-1}). \tag{2.13}$$

It is clear that NH1 and NH2 are descent methods, they satisfy (2.11).
On the other hand, the search direction d_k in quasi-Newton methods is obtained as a solution of the linear algebraic system

$$B_k d_k = -g_k, \tag{2.14}$$

where B_k is an approximation of the Hessian. The initial approximation is the identity matrix ($B_0 = I$) and the subsequent updates B_k are defined by an appropriate formula.
Here, we are interested in the BFGS update formula, defined by

$$B_{k+1} = B_k + \frac{y_k y_k^T}{s_k^T y_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}, \tag{2.15}$$

where $s_k = x_{k+1} - x_k, \quad y_k = g_{k+1} - g_k$. The next secant equation must hold

$$B_{k+1} s_k = y_k, \tag{2.16}$$

which is possible only if the curvature condition

$$y_k^T s_k > 0 \tag{2.17}$$

is satisfied.

The three-term hybrid BFGS conjugate gradient method was proposed in [10]. That method uses best properties of both BFGS and CG methods and defines a hybrid BFGS-CG method for solving some selected unconstrained optimization problems, resulting in improvement in the total number of iterations and the CPU time.

3. Modification of LSCD, DHSDL and DLSDL methods

3.1. A modified LSCD conjugate gradient method

We consider the modification of LSCD method, defined in [17] by

$$\beta_k^{LSCD} = \max\left\{0, \min\left\{\beta_k^{LS}, \beta_k^{CD}\right\}\right\}, \tag{3.1}$$

$$d_k = \begin{cases} -g_0 & k = 0 \\ d_k = -g_k + \beta_k^{LSCD} d_{k-1} & k \geq 1, \end{cases}$$

and define the MLSCD method [15] with the search direction

$$d_k = \mathcal{D}(\beta_k^{LSCD}, g_k, d_{k-1}). \tag{3.2}$$

Now, we give the algorithm of this method using the Wolfe line search.

3.1.1. Algorithm WMLSCD.

- Step0: Given a starting point x_0 and a parameter $0 < \varepsilon < 1$.
- Step1: Set $k = 0$ and compute $d_0 = -g_0$.
- Step2: If $\|g_k\| \leq \varepsilon$, STOP; else go to Step3.
- Step3: Find the step size $\alpha_k \in]0, 1]$ using the Wolfe line search.
- Step4: Compute $x_{k+1} = x_k + \alpha_k d_k$.
- Step5: Compute $y_k = g_{k+1} - g_k$ and go to Step6.
- Step6: Compute

$$\begin{aligned} \beta_{k+1}^{LS} &= -\frac{y_k^T g_{k+1}}{g_k^T d_k}, \ \beta_{k+1}^{CD} = -\frac{\|g_{k+1}\|^2}{g_k^T d_k}, \\ \beta_{k+1}^{LSCD} &= \max\left\{0, \min\left\{\beta_{k+1}^{LS}, \beta_{k+1}^{CD}\right\}\right\}. \end{aligned}$$

- Step7: Compute the search direction $d_{k+1} = \mathcal{D}(\beta_{k+1}^{LSCD}, g_{k+1}, d_k)$.
- Step8: Let $k := k + 1$ and go to Step2.

3.1.2. Convergence of the WMLSCD conjugate gradient method. It is easy to prove the next theorem.

Theorem 3.1. *Let β_k be any CG parameter. Then, the search direction*

$$d_k = \mathcal{D}(\beta_k, g_k, d_{k-1})$$

satisfies

$$g_k^T d_k = -\|g_k\|^2. \tag{3.3}$$

To prove the global convergence of the WMLSCD method, we need the following assumptions.

Assumption 3.1 The level set $\mathcal{L} = \{x \in \mathbb{R}^n / f(x) \leq f(x_0)\}$ is bounded.

Assumption 3.2 The function f is continuously differentiable in some neighbourhood $\mathcal{N}$ of $\mathcal{L}$ and its gradient is Lipschitz continuous. Namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \text{ for all } x, y \in \mathcal{N}. \tag{3.4}$$

It is well known that if Assumption 3.2 holds, then there exists a positive constant γ, such that

$$\|g_k\| \leq \gamma, \forall k \tag{3.5}$$

The next lemma, often called the Zoutendijk condition [22], is used to prove the global convergence of nonlinear CG method.

Lemma 3.2. [15] *Let the Assumption 3.1 and Assumption 3.2 be satisfied. Let the sequence $\{x_k\}$ be generated by the MLSCD method with the Wolfe line search. Then it holds that*

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty \tag{3.6}$$

Theorem 3.3. *Let the Assumption 3.1 and Assumption 3.2 hold. Then, the sequence $\{x_k\}$ generated by the WMLSCD method with the Wolfe line search satisfies*

$$\liminf_{k \to \infty} \|g_k\| = 0 \tag{3.7}$$

Proof. In order to gain the contradiction, let us suppose that (3.7) does not hold. Then, there exists a constant $c > 0$ such that

$$\|g_k\| \geq c, \text{ for all } k \tag{3.8}$$

Clearly, (3.2) can be rewritten into the form

$$d_k = -l_k g_k + \beta_k^{LSCD} d_{k-1}, \; l_k = 1 + \beta_k^{LSCD} \frac{g_k^T d_{k-1}}{\|g_k\|^2}. \tag{3.9}$$

Now from (3.9), it follows that

$$d_k + l_k g_k = \beta_k^{LSCD} d_{k-1}$$

which further implies

$$(d_k + l_k g_k)^2 = \left(\beta_k^{LSCD} d_{k-1}\right)^2$$

$$\Longleftrightarrow \|d_k\|^2 + 2 l_k d_k^T g_k + l_k^2 \|g_k\|^2 = \left(\beta_k^{LSCD}\right)^2 \|d_{k-1}\|^2,$$

and subsequently

$$\|d_k\|^2 = \left(\beta_k^{LSCD}\right)^2 \|d_{k-1}\|^2 - 2 l_k d_k^T g_k - l_k^2 \|g_k\|^2. \tag{3.10}$$

Notice that

$$\beta_k^{LSCD} = \max\left\{0, \min\left\{\beta_k^{LS}, \beta_k^{CD}\right\}\right\} \leq \left|\beta_k^{CD}\right| \tag{3.11}$$

Dividing both sides of (3.10) by $(g_k^T d_k)^2$, we get from (3.11), (3.3), (3.8) and the definition of β_k^{CD} that

$$\begin{aligned}
\frac{\|d_k\|^2}{\|g_k\|^4} &= \frac{\|d_k\|^2}{(g_k^T d_k)^2} = \left(\beta_k^{LSCD}\right)^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{2 l_k d_k^T g_k}{(g_k^T d_k)^2} - l_k^2 \frac{\|g_k\|^2}{(g_k^T d_k)^2} \\
&\leq \left(\beta_k^{CD}\right)^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{2 l_k}{g_k^T d_k} - l_k^2 \frac{\|g_k\|^2}{(g_k^T d_k)^2} \\
&= \left(\frac{\| g_k \|^2}{-g_{k-1}^T d_{k-1}}\right)^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{2 l_k}{g_k^T d_k} - l_k^2 \frac{\|g_k\|^2}{(g_k^T d_k)^2}
\end{aligned}$$

Finally

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leq \left(\frac{\| g_k \|^2}{-g_{k-1}^T d_{k-1}}\right)^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{2 l_k}{g_k^T d_k} - l_k^2 \frac{\|g_k\|^2}{(g_k^T d_k)^2} \tag{3.12}$$

Now, applying (3.3), (3.12) becomes

$$\begin{aligned}
\frac{\|d_k\|^2}{\|g_k\|^4} &\leq \frac{\| g_k \|^4}{\| g_{k-1} \|^4}^2 \frac{\|d_{k-1}\|^2}{\| g_k \|^4} - \frac{2l_k}{\| g_k \|^2} - l_k^2 \frac{\|g_k\|^2}{\| g_k \|^4} \\
&= \frac{\|d_{k-1}\|^2}{\| g_{k-1} \|^4} + \frac{2l_k}{\| g_k \|^2} - l_k^2 \frac{1}{\| g_k \|^2} \\
&= \frac{\|d_{k-1}\|^2}{\| g_{k-1} \|^4} - \frac{(l_k - 1)^2}{\| g_k \|^2} + \frac{1}{\| g_k \|^2} \\
&\leq \frac{\|d_{k-1}\|^2}{\| g_{k-1} \|^4} + \frac{1}{\| g_k \|^2} \\
&\leq \sum_{j=0}^{k} \frac{1}{\|g_j\|^2} \\
&\leq \frac{k+1}{c^2}.
\end{aligned}$$

The last inequalities imply

$$\sum_{k\geq 1} \frac{\|g_k\|^4}{\|d_k\|^2} \geq c^2 \sum_{k\geq 1} \frac{1}{k+1} = \infty$$

which contradicts to (3.6). This completes the proof. □

3.2. A modified DHSDL and DLSDL conjugate gradient method

In this part, we have the hybrid MMDL method, proposed in [15], which is defined by the search direction d_k as follows

$$\beta_k^{MMDL} = \max\left\{0, \min\left\{\beta_k^{DHSDL}, \beta_k^{DLSDL}\right\}\right\}$$
$$d_k = \mathcal{D}(\beta_k^{MMDL}, g_k, d_{k-1}).$$

We give the algorithm of this method where we have changed the backtracking line search by the Wolfe line search.

3.2.1. Algorithm WMMDL.

- Step0: Given a starting point x_0, a parameter $0 < \varepsilon < 1$ and $\mu > 1$.
- Step1: Set $k = 0$ and compute $d_0 = -g_0$.
- Step2: If $\|g_k\| \leq \varepsilon$, STOP; else go to Step3.
- Step3: Find the step size $\alpha_k \in]0, 1]$ using the Wolfe line search.
- Step4: Compute $x_{k+1} = x_k + \alpha_k d_k$.
- Step5: Compute $y_k = g_{k+1} - g_k$, $s_k = x_{k+1} - x_k$ and go to Step6.
- Step6: Compute

$$\beta_{k+1}^{DHSDL} = \frac{\| g_{k+1} \|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} \mid g_{k+1}^T g_k \mid}{\mu \mid g_{k+1}^T d_k \mid + d_k^T y_k} - \alpha_k \frac{g_{k+1}^T s_k}{d_k^T y_k}$$

$$\beta_{k+1}^{DLSDL} = \frac{\| g_{k+1} \|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} \mid g_{k+1}^T g_k \mid}{\mu \mid g_{k+1}^T d_k \mid - d_k^T g_k} - \alpha_k \frac{g_{k+1}^T s_k}{d_k^T y_k}$$

$$\beta_{k+1}^{MMDL} = \max\left\{0, \min\left\{\beta_{k+1}^{DHSDL}, \beta_{k+1}^{DLSDL}\right\}\right\}.$$

- Step7: Compute the search direction $d_{k+1} = \mathcal{D}(\beta_{k+1}^{MMDL}, g_{k+1}, d_k)$.
- Step8: Let $k := k+1$ and go to Step2.

3.2.2. Convergence of the WMMDL conjugate gradient method. The following theorem prove the global convergence of the WMMDL method.

Theorem 3.4. *Let the Assumption 3.1 and Assumption 3.2 be satisfied. Then the sequence $\{x_k\}$ generated by the WMMDL method with the Wolfe line search satisfies*

$$\liminf_{k\to\infty} \|g_k\| = 0 \tag{3.13}$$

Proof. Assume, on the contrary, that (3.13) does not hold. Then, there exists a constant $c > 0$ such that

$$\|g_k\| \geq c, \text{ for all } k \tag{3.14}$$

Denote

$$l_k = 1 + \beta_k^{MMDL} \frac{g_k^T d_{k-1}}{\|g_k\|^2}$$

Then we can write

$$d_k + l_k g_k = \beta_k^{MMDL} d_{k-1}$$

and further

$$(d_k + l_k g_k)^2 = \left(\beta_k^{MMDL} d_{k-1}\right)^2$$

$$\Longleftrightarrow \|d_k\|^2 + 2 l_k d_k^T g_k + l_k^2 \|g_k\|^2 = \left(\beta_k^{MMDL}\right)^2 \|d_{k-1}\|^2.$$

Thus,

$$\|d_k\|^2 = \left(\beta_k^{MMDL}\right)^2 \|d_{k-1}\|^2 - 2 l_k d_k^T g_k - l_k^2 \|g_k\|^2. \tag{3.15}$$

Having in view, $\mu > 1$ as well as $d_k^T g_k < 0$ and applying the extended conjugacy condition $d_k^T y_{k-1} = -\alpha g_k^T s_{k-1}$, $\alpha > 0$, which was exploited in [3, 21], we get

$$\begin{aligned}
\beta_{k+1}^{DHSDL} &= \frac{\| g_{k+1} \|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} \mid g_{k+1}^T g_k \mid}{\mu \mid g_{k+1}^T d_k \mid + d_k^T y_k} - \alpha_k \frac{g_{k+1}^T s_k}{d_k^T y_k} \\
&\leq \frac{\| g_{k+1} \|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} \mid g_{k+1}^T g_k \mid}{\mu \mid g_{k+1}^T d_k \mid + d_k^T y_k} \\
&= \frac{\| g_{k+1} \|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} \mid g_{k+1}^T g_k \mid}{\mu \mid g_{k+1}^T d_k \mid + d_k^T (g_{k+1} - g_k)} \\
&= \frac{\| g_{k+1} \|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} \mid g_{k+1}^T g_k \mid}{\mu \mid g_{k+1}^T d_k \mid + d_k^T g_{k+1} - d_k^T g_k} \\
&\leq \frac{\| g_{k+1} \|^2}{\mu \mid g_{k+1}^T d_k \mid + d_k^T g_{k+1} - d_k^T g_k} \\
&\leq \frac{\| g_{k+1} \|^2}{-d_k^T g_k}.
\end{aligned}$$

Further

$$\begin{aligned} \beta_{k+1}^{DLSDL} &= \frac{\| g_{k+1} \|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} \mid g_{k+1}^T g_k \mid}{\mu \mid g_{k+1}^T d_k \mid - d_k^T g_k} - \alpha_k \frac{g_{k+1}^T s_k}{d_k^T y_k} \\ &\leq \frac{\| g_{k+1} \|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} \mid g_{k+1}^T g_k \mid}{\mu \mid g_{k+1}^T d_k \mid - d_k^T g_k} \\ &\leq \frac{\| g_{k+1} \|^2}{\mu \mid g_{k+1}^T d_k \mid - d_k^T g_k} \\ &\leq \frac{\| g_{k+1} \|^2}{-d_k^T g_k}. \end{aligned}$$

Now, we conclude

$$\beta_k^{MMDL} = \max\left\{0, \min\left\{\beta_k^{DHSDL}, \beta_k^{DLSDL}\right\}\right\} \leq \frac{\| g_k \|^2}{-d_{k-1}^T g_{k-1}} \tag{3.16}$$

Next, dividing both sides of (3.15) by $(g_k^T d_k)^2$, we get from (3.3), (3.16) and (3.14) that

$$\begin{aligned} \frac{\|d_k\|^2}{\|g_k\|^4} &= \frac{\|d_k\|^2}{(g_k^T d_k)^2} = \left(\beta_k^{MMDL}\right)^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{2 l_k d_k^T g_k}{(g_k^T d_k)^2} - l_k^2 \frac{\|g_k\|^2}{(g_k^T d_k)^2} \\ &= \left(\beta_k^{MMDL}\right)^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{2 l_k}{g_k^T d_k} - l_k^2 \frac{\|g_k\|^2}{(g_k^T d_k)^2} \\ &\leq \left(\frac{\| g_k \|^2}{-g_{k-1}^T d_{k-1}}\right)^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{2 l_k}{g_k^T d_k} - l_k^2 \frac{\|g_k\|^2}{(g_k^T d_k)^2} \\ &= \frac{\| g_k \|^4}{\| g_{k-1} \|^4}^2 \frac{\|d_{k-1}\|^2}{\| g_k \|^4} - \frac{2 l_k}{\| g_k \|^2} - l_k^2 \frac{\|g_k\|^2}{\| g_k \|^4} \\ &= \frac{\|d_{k-1}\|^2}{\| g_{k-1} \|^4} + \frac{2 l_k}{\| g_k \|^2} - l_k^2 \frac{1}{\| g_k \|^2} \\ &= \frac{\|d_{k-1}\|^2}{\| g_{k-1} \|^4} - \frac{(l_k - 1)^2}{\| g_k \|^2} + \frac{1}{\| g_k \|^2} \\ &\leq \frac{\|d_{k-1}\|^2}{\| g_{k-1} \|^4} + \frac{1}{\| g_k \|^2} \\ &\leq \sum_{j=0}^{k} \frac{1}{\|g_j\|^2} \\ &\leq \frac{k+1}{c^2}. \end{aligned}$$

These inequalities imply

$$\sum_{k \geq 1} \frac{\|g_k\|^4}{\|d_k\|^2} \geq c^2 \sum_{k \geq 1} \frac{1}{k+1} = \infty$$

Therefore, $\|g_k\| \geq c$ causes a contradiction to (3.6). Consequently, (3.13) is verified. This completes the proof. □

4. Hybrid BFGS-CG methods

It is known that conjugate gradient method are better compared to the quasi-Newton method in terms of the CPU time. In addition, BFGS is more costly in terms of the memory storage requirements than CG. On the other hand, the quasi-Newton methods are better in terms of the number of iterations and the number of function evaluations. For this purpose, various hybridizations of quasi-Newton methods and CG methods have been proposed by various researchers.

In [10], the authors proposed a hybrid search direction that combines the quasi-Newton and CG methods, where d_k is defined by

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

$$\text{where } \eta > 0 \text{ and } \beta_k = \frac{g_k^T g_{k-1}}{g_k^T d_{k-1}}.$$

A hybrid direction search between BFGS update of the Hessian matrix and the conjugate parameter β_k was proposed in [1, 11].

4.1. WH-BFGS-CG method

P. S. Stanimirovic et al. proposed in [15] a three-term hybrid BFGS-CG method, called H-BFGS-CG, defined by the search direction

$$d_k = \begin{cases} -B_k g_k, & k = 0 \\ \mathcal{D}_1(\beta_{k+1}^{LSCD}, g_k, d_{k-1}), & k \geq 1 \end{cases} \tag{4.1}$$

The following algorithm correspond to this method, where we have changed the backtracking line search by the Wolfe line search.

4.1.1. Algorithm WH-BFGS-CG.

- Step0: Given a starting point x_0 and a parameter $0 < \varepsilon < 1$.
- Step1: Set $k = 0$ and compute g_0, $B_0 = I$, $d_0 = -B_0 g_0$.
- Step2: If $\|g_k\| \leq \varepsilon$, STOP; else go to Step3.
- Step3: Find the step size $\alpha_k \in]0, 1]$ using the Wolfe line search.
- Step4: Compute $x_{k+1} = x_k + \alpha_k d_k$.
- Step5: Compute $y_k = g_{k+1} - g_k$, $s_k = x_{k+1} - x_k$ and go to Step6.
- Step6: Compute

$$\beta_{k+1}^{LS} = -\frac{y_k^T g_{k+1}}{g_k^T d_k}, \ \beta_{k+1}^{CD} = -\frac{\|g_{k+1}\|^2}{g_k^T d_k},$$

$$\beta_{k+1}^{LSCD} = \max\left\{0, \min\left\{\beta_{k+1}^{LS}, \beta_{k+1}^{CD}\right\}\right\}.$$

- Step7: Compute B_{k+1} using (2.15).
- Step8: Compute the search direction $d_{k+1} = \mathcal{D}_1(\beta_{k+1}^{LSCD}, g_{k+1}, d_k)$.
- Step9: Let $k := k + 1$ and go to Step2.

4.2. Convergence analysis of WH-BFGS-CG method

Assumption 4.1:

$H1$: The objective function f is twice continuously differentiable.

$H2$: The level set $\mathcal{L}$ is convex. Moreover, there exist positive constants c_1 and c_2 such that

$$c_1 \|z\|^2 \leq z^T H(x) z \leq c_2 \|z\|^2, \text{ for all } z \in \mathbb{R}^n \text{ and } x \in \mathcal{L},$$

where $H(x)$ is the Hessian of f.

$H3$: The gradient g is Lipschitz continuous at the point x^*, that is, there exists a positive constant c_3 satisfying

$$\|g(x) - g(x^*)\| \leq c_3 \|x - x^*\|,$$

for all x in a neighbourhood of x^*.

Theorem 4.1. [2] *Let $\{B_k\}$ be generated by the BFGS update formula* (2.15)*, where $s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$. Assume that the matrix B_k is symmetric positive definite and satisfies* (2.16) *and* (2.17) *for all k. Furthermore, assume that $\{s_k\}$ and $\{y_k\}$ satisfy the inequality*

$$\frac{\|y_k - G_* s_k\|}{\|s_k\|} \leq \epsilon_k,$$

for some symmetric positive definite matrix G_* and for some sequence $\{\epsilon_k\}$ possessing the property

$$\sum_{k=1}^{\infty} \epsilon_k < \infty,$$

then

$$\lim_{k \longrightarrow \infty} \frac{\|(B_k - G_*) s_k\|}{\|s_k\|} = 0,$$

and the sequences $\{\|B_k\|\}$, $\left\{\left\|B_k^{-1}\right\|\right\}$ are bounded.

Theorem 4.2. *(Sufficient descent and global convergence) Consider Algorithm WH-BFGS-CG. Assume that the conditions H1, H2 and H3 in Assumption 4.1 are satisfied as well as conditions of Theorem* 4.1*. Then*

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

Proof. From (4.1), we have

$$\begin{aligned} g_k^T d_k &= -g_k^T B_k g_k - g_k^T g_k - \beta_k^{LSCD} g_k^T d_{k-1} + \beta_k^{LSCD} g_k^T d_{k-1} \\ &\leq -c_1 \|g_k\|^2 - \|g_k\|^2 = -(c_1 + 1) \|g_k\|^2 \\ &\leq -\|g_k\|^2, \quad 0 < c_1 + 1 \leq 1, \end{aligned}$$

then

$$g_k^T d_k \leq -\|g_k\|^2. \tag{4.2}$$

We conclude that the sufficient descent holds.

Further, from Wolfe line search conditions and (4.2), it holds

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\rho \alpha_k g_k^T d_k \geq \rho \alpha_k \|g_k\|^2. \tag{4.3}$$

Since $f(x_k)$ is decreasing and the sequence $f(x_k)$ is bounded below and by the condition $H2$, we have

$$\lim_{k\to\infty} f(x_k) - f(x_k + \alpha_k d_k) = 0. \tag{4.4}$$

Hence (4.3) and (4.4) imply

$$\lim_{k\to\infty} \rho\alpha_k \|g_k\|^2 = 0.$$

Now, since $\rho > 0$ and $\alpha_k > 0$, we have

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

This completes the proof. □

5. Numerical results

In this section, some numerical results are reported to illustrate the behaviours of WMLSCD, WMMDL and WH-BFGS-CG methods. The step size α_k is determined using the Wolfe line search.

We use the Matlab Langage with a precision $\varepsilon = 10^{-6}$.

We designate by:

- k: The number of iterations required to obtain the solution.
- Time: The execution time in second.

Example 5.1. We take the function

$$f(x) = \sum_{i=1}^{n} (\exp(x_i) - x_i).$$

We take as starting point $x_0 = (1, 1, \ldots, 1)^T$.
The minimum of this function is reached at the point

$$x^* = (0, 0, \ldots, 0)^T \text{ and } f(x^*) = n.$$

The results obtained are summarised in the following tables:
For $n = 3$, we have

Methods	k	Time	$\| g_k \|$
WMLSCD	19	0.149532	$8.0732e-07$
WMMDL	19	0.161138	$8.0732e-07$
WH-BFGS-GC	5	0.073673	$1.4372e-08$

For $n = 100$, we have

Methods	k	Time	$\| g_k \|$
WMLSCD	22	3.883876	$5.8263e-07$
WMMDL	22	3.803220	$5.8263e-07$
WH-BFGS-GC	5	1.622640	$8.2976e-08$

For $n = 500$, we have

Methods	k	Time	$\parallel g_k \parallel$
WMLSCD	24	74.325460	$6.4631e - 07$
WMMDL	24	70.101070	$6.4631e - 07$
WH-BFGS-GC	5	21.087659	$1.8554e - 07$

Example 5.2. We take the function

$$f(x) = \sum_{i=1}^{n} \ln(\exp(x_i) + \exp(-x_i)).$$

We take as starting point $x_0 = (1.1, 1.1, \ldots, 1.1)^T$
The minimum of this function is reached at the point

$$x^* = (0, 0, \ldots, 0)^T \text{ and } f(x^*) = n \ln(2).$$

The results obtained are summarised in the following tables:
For $n = 3$, we have

Methods	k	Time	$\parallel g_k \parallel$
WMLSCD	96	0.348543	$9.6801e - 07$
WMMDL	95	0.443647	$9.5309e - 07$
WH-BFGS-GC	47	0.375461	$8.4400e - 08$

For $n = 100$, we have

Methods	k	Time	$\parallel g_k \parallel$
WMLSCD	104	40.083872	$9.9132e - 07$
WMMDL	104	83.918822	$9.9369e - 07$
WH-BFGS-GC	66	20.465962	$8.4827e - 07$

For $n = 200$, we have

Methods	k	Time	$\parallel g_k \parallel$
WMLSCD	107	83.209667	$9.1391e - 07$
WMMDL	108	80.273199	$9.2334e - 07$
WH-BFGS-GC	69	52.410529	$8.2027e - 07$

For $n = 300$, we have

Methods	k	Time	$\parallel g_k \parallel$
WMLSCD	109	171.535865	$9.8675e - 07$
WMMDL	111	205.430203	$9.5399e - 07$
WH-BFGS-GC	70	110.807414	$7.9846e - 07$

Commentaries: The numerical tests show clearly that the proposed hybrid algorithm WH-BFGS-GC Wolfe based on line search is more efficient in terms of number of iterations and computation time than WMLSCD and WMMDL methods.

6. Conclusion

We have considered the hybrid conjugate gradient methods, MLSCD, MMDL and H-BFGS-CG, for solving unconstrained optimization problems where we have changed the backtracking line search given in [15] by the Wolfe line search. Firstly, we have shown that the obtained WMLSCD, WMMDL and WH-BFGS-CG algorithms are globally convergent for general functions.

Secondly, the numerical simulations confirm the effectiveness of the approach WH-BFGS-CG. In fact, the WH-BFGS-CG method is the most efficient in terms of number of iterations and computation time compared to WMLSCD and WMMDL methods which was not the case with backtracking line search, where the computation time of H-BFGS-GC was greater than MLSCD and MMDL [15].

Acknowledgment. . This work was supported by the Directorate-General for Scientific Research and Technological Development, MESRS Algeria.

References

[1] Baluch, B., Salleh, Z., Alhawarat, A., Roslan, U.A.M., *A new modified three-term conjugate gradient method with sufficient descent property and its global convergence*, J. Math. (2017), Article ID 2715854, 12 pp.

[2] Byrd, R.H., Nocedal, J., *A tool for the analysis of quasi-Newton methods with application to unconstrained minimization*, SIAM J. Numer. Anal., **26**(1989), no. 3, 727-739.

[3] Dai, Y.H., Liao, L.Z., *New conjugacy conditions and related nonlinear conjugate gradient methods*, Appl. Math. Optim., **43**(2001), no. 1, 87-101.

[4] Dai, Y.H., Yuan, Y., *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM J. Optim., **10**(1999), no. 1, 177-182.

[5] Dai, Y.H., Yuan, Y., *An efficient hybrid conjugate gradient method for unconstrained optimization*, Ann. Oper. Res., **103**(2001), 33-47.

[6] Fletcher, R., *Practical Methods of Optimization. Unconstrained Optimization*, vol. 1, Wiley, New York, 1987.

[7] Fletcher, R., Reeves, C.M., *Function minimization by conjugate gradients*, Comput. J., **7**(1964), no. 2, 149-154.

[8] Gilbert, J.C., Nocedal, J., *Global convergence properties of conjugate gradient methods for optimization*, SIAM J. Optim., **2**(1992), no. 1, 21-42.

[9] Hestenes, M.R., Stiefel, E., *Methods of conjugate gradients for solving linear systems*, J. Res. Natl. Bur. Stand., **49**(1952), no. 6, 409-436.

[10] Ibrahim, M.A.H., Mamat, M., Leong, W.J., *The hybrid BFGS-CG method in solving unconstrained optimization problems*, Abstr. Appl. Anal. Hindawi Publishing Corporation. Article ID 507102, 6 pp.

[11] Khanaiah, Z., Hmod, G., *Novel hybrid algorithm in solving unconstrained optimizations problems*, Int. J. Novel Res. Phys. Chem. Math., **4**(2017), no. 3, 36-42.

[12] Liu, Y., Storey, C., *Efficient generalized conjugate gradient algorithms, part 1: theory*, J. Optim. Theory Appl., **69**(1991), no. 1, 129-137.

[13] Polak, E., Ribiere, G., *Note sur la convergence des méthodes de directions conjuguées*, Rev. Française d'Informatique et de Recherche Opérationnelle **3**(1969), no. R1, 35-43.

[14] Polyak, B.T., *The conjugate gradient method in extreme problems*, Comput. Math. Phys., **9**(1969), 94-112.

[15] Stanimirović, P.S., Ivanov, B., Djordjević, S., Brajević, I., *New hybrid conjugate gradient and Broyden–Fletcher–Goldfarb–Shanno conjugate gradient methods*, J. Optim. Theory Appl., **178**(2018), no. 3, 860-884.

[16] Touati-Ahmed, D., Storey, C., *Efficient hybrid conjugate gradient techniques*, J. Optim. Theory Appl., **64**(1990), no. 2, 379-397.

[17] Yang, X., Luo, Z., Dai, X., *A global convergence of LS-CD hybrid conjugate gradient method*, Adv. Numerical Anal, (2013), Hindawi Publishing Corporation, Article ID 517452, 5 pp.

[18] Zhang, L., *Nonlinear Conjugate Gradient Methods for Optimization Problems*, Ph.D. Thesis, College of Mathematics and Econometrics, Hunan University, Changsha China, 2006.

[19] Zhang, L., Zhou, W., *Two descent hybrid conjugate gradient methods for optimization*, J. Comput. Appl. Math., **216**(2008), 251-264.

[20] Zhang, L., Zhou, W.J., Li, D.H., *Global convergence of a modified Fletcher–Reeves conjugate method with Armijo-type line search*, J. Numer. Math., **104**(2006), 561-572.

[21] Zheng, Y., Zheng, B., *Two new Dai-Liao-type conjugate gradient methods for unconstrained optimization problems*, J. Optim. Theory Appl., **175**(2017), 502-509.

[22] Zoutendijk, G., *Nonlinear Programming, Computational Methods*, Abadie, J. (ed.), Integer and Nonlinear Programming, (1970), 37-86.

Khelladi Samia
Laboratory of Fundamental and Numerical Mathematics LMFN,
Faculty of Sciences,
Setif-1 Ferhat Abbas University,
19000 Setif, Algeria
e-mail: samia.boukaroura@univ-setif.dz

Benterki Djamel
Laboratory of Fundamental and Numerical Mathematics LMFN,
Faculty of Sciences,
Setif-1 Ferhat Abbas University,
19000 Setif, Algeria
e-mail: djbenterki@univ-setif.dz