Three New Hybrid Conjugate Gradient Methods for Optimization*

Anwa Zhou, Zhibin Zhu, Hao Fan, Qian Qing
School of Mathematics and Computational Science, Guilin University of Electronic Technology, Guilin, China
E-mail: zhouanwa@126.com
Received November 10, 2010; revised January 9, 2011; accepted January 14, 2011

Abstract

In this paper, three new hybrid nonlinear conjugate gradient methods are presented, which produce sufficient descent search direction at every iteration. This property is independent of any line search or the convexity of the objective function used. Under suitable conditions, we prove that the proposed methods converge globally for general nonconvex functions. The numerical results show that all these three new hybrid methods are efficient for the given test problems.

Keywords: Conjugate Gradient Method, Descent Direction, Global Convergence

1. Introduction

In this paper, we consider the unconstrained optimization problem:

$$\min f(x), x \in \mathbb{R}^n$$ (1.1)

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and the gradient of $f$ at $x$ is denoted by $g(x)$.

Due to its simplicity and its very low memory requirement, the conjugate gradient (CG) method plays a very important role for solving (1.1). Especially, when the scale is large, the CG method is very efficient. Let $x_0 \in \mathbb{R}^n$ be the initial guess of the solution of problem (1.1). A nonlinear conjugate gradient method is usually designed by the iterative form

$$x_{k+1} = x_k + \alpha_k d_k, k = 0, 1, \cdots, (1.2)$$

where $x_k$ is the current iterate point, $\alpha_k > 0$ is a step-length which is determined by some line search, and $d_k$ is the search direction defined by

$$d_k = \begin{cases} 
- g_k, & \text{if } k = 0, \\
- g_k + \beta_k d_{k-1}, & \text{if } k > 0,
\end{cases}$$ (1.3)

where $g_k$ denotes $g(x_k)$, and $\beta_k$ is a scalar. There are some well-known formulas for $\beta_k$, which are given as follows:

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_k\|^2},$$ [1];
$$\beta_k^{CD} = \frac{\|g_k\|^2}{d_k^T g_{k+1}},$$ [2];
$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_k^T d_{k+1}},$$ [3];
$$\beta_k^{HS} = \frac{g_k^T y_k}{d_k^T y_k},$$ [4];
$$\beta_k^{PRP} = \frac{g_k^T y_k}{\|g_k\|^2},$$ [5, 6];
$$\beta_k^{LS} = \frac{g_k^T y_k}{d_k^T g_{k+1}},$$ [7];

where $y_{k+1} = g_k - g_{k+1}$ and $\|\|$ stands for the Euclidean norm of vectors.

Although the methods above are equivalent [8, 9] when $f$ is a strictly convex quadratic function and $\alpha_k$ is calculated by the exact line search, their behaviors for general objective functions may be far different. Generally, in the convergence analysis of conjugate gradient methods, one hopes the inexact line search such as the Wolfe conditions, the strong Wolfe conditions or the strong *Wolfe conditions, which are showed respectively as follows:

1) The Wolfe line search is to find $\alpha_k$ such that
As we all know, the FR, DY and CD methods are descent methods, but their descent properties depend on the line search such as the strong Wolfe line search (1.5). Similar to the descent three terms PRP method in [17], Zhang et al. [18,19] proposed a modified FR method which we call the MFR method, that is,

\[
MFR: d_k = -g_k + \beta_k^{FR} d_{k-1} - \frac{\alpha_k}{\|g_k\|} g_k \left(1 + \frac{\beta_k^{FR}}{\|g_k\|} \right) g_k + \beta_k^{FR} d_{k-1}. 
\]

And [18] also gave an equivalent form to the MFR method. Similarly, Zhang [19] also proposed a modified DY method called the MDY method, that is,

\[
MDY: d_k = -g_k + \beta_k^{DY} d_{k-1} - \frac{\alpha_k}{\|g_k\|} g_k \left(1 + \frac{\beta_k^{DY}}{\|g_k\|} \right) g_k + \beta_k^{DY} d_{k-1}. 
\]

It is easy to see that the MFR and MDY methods have an important property that the search directions satisfy

\[
g_k^T d_k = -\|g_k\|^2, 
\]

which depends neither on the line search used nor on the convexity of the objective function; moreover these two methods reduce to the FR method and the DY method respectively with exact line search.

[18] has explored the convergence and efficiency of the MFR method for nonconvex functions with the Wolfe line search or Armijo line search. Based on the idea of the H1 and the H2 methods, recently, Zhang-Zhou [20] replaced \(\beta_k^{FR}\) in (1.9) and \(\beta_k^{DY}\) in (1.10) with \(\beta_k^{H1}\) in (1.7) and \(\beta_k^{H2}\) in (1.8), respectively, and proposed two new hybrid PRP–FR and HS–DY methods called the NH1 method and the NH2 method, respectively, that is,

\[
NH1: d_k = -g_k + \beta_k^{H1} \frac{g_k^T d_{k-1}}{\|g_k\|} g_k + \beta_k^{H1} d_{k-1} 
\]

\[
NH2: d_k = -g_k + \beta_k^{H2} \frac{g_k^T d_{k-1}}{\|g_k\|} g_k + \beta_k^{H2} d_{k-1} 
\]

Obviously, these two new hybrid methods still satisfy

\[
g_k^T d_k = -\|g_k\|^2, 
\]

which shows that they are descent and independent of any line search used. [20] proved the global convergence of these two methods and also showed their efficiency in real computations.

Similarly, based on the idea of the methods all above, we consider the CD method, and propose three new hy-
brid conjugate gradient(CG) methods which we call the H3 method, the MCD method and the NH3 method, respectively, that is

\[ H3: \beta_k^{H3} = \max \left\{ 0, \min \left\{ \beta_k^{LS}, \beta_k^{CD} \right\} \right\} \]  

(1.14)

\[ MCD: d_k = -g_k + \beta_k^{CD} d_{k-1} + \frac{g^T_k d_{k-1}}{d^T_k g_{k-1}} g_k \]  

(1.15)

\[ NH3: d_k = -\left( 1 + \beta_k^{H3} \frac{g^T_k d_{k-1}}{g^T_k g_k} \right) g_k + \beta_k^{H3} d_{k-1} \]  

(1.16)

From these three methods above, it is not difficult to see that the MCD and the NH3 methods also satisfy \( g^T_k d_k = -\|g_k\|^2 \), which shows that they are sufficient descent methods. In the next section, the new algorithms are given. The global convergence of the proposed methods are proved in Section 3. We give the numerical experiments in Section 4, and in Section 5, the conclusion is presented.

2. Algorithm

2.1. Algorithm 1 (The H3 Algorithm)

Step 0: Choose an initial point \( x_0 \in \mathbb{R}^n \), \( 0 < \varepsilon \ll 1 \), \( 0 < \delta < \frac{1}{2} \), and \( \delta < \sigma < 1 \). Set \( d_0 = -g_0, k := 0 \).

Step 1: If \( \|g_k\| < \varepsilon \), then stop; Otherwise go to the next step.

Step 2: Compute step size \( \alpha_k \) by strong *Wolfe line search rule (1.6).

Step 3: Let \( x_{k+1} = x_k + \alpha_k d_k \). If \( \|g_{k+1}\| < \varepsilon \), then stop.

Step 4: Calculate the search direction \( d_{k+1} = -g_{k+1} + \beta_k^{H3} d_k \).

Step 5: Set \( k := k + 1 \), and go to Step 2.

2.2. Algorithm 2 (The MCD (Or the NH3) Algorithm)

Step 0: Choose an initial point \( x_0 \in \mathbb{R}^n \), \( 0 < \varepsilon \ll 1 \), \( 0 < \delta < \frac{1}{2} \), and \( \delta < \sigma < 1 \). Set \( d_0 = -g_0, k := 0 \).

Step 1: If \( \|g_k\| < \varepsilon \), then stop; otherwise go to the next step.

Step 2: Compute step size \( \alpha_k \) by Wolfe line search rule (1.4).

Step 3: Let \( x_{k+1} = x_k + \alpha_k d_k \). If \( \|g_{k+1}\| < \varepsilon \), then stop.

Step 4: Calculate the search direction \( d_{k+1} = -g_{k+1} + \beta_k^{CD} d_k \).

Step 5: Set \( k := k + 1 \), and go to Step 2.

3. The Global Convergence

Assumption A

1) The level set \( \Omega = \{ x \in \mathbb{R}^n \mid f(x) \leq f(x_0) \} \) is bounded, where \( x_0 \in \mathbb{R}^n \), is a given point.

2) In an open convex set \( N \) that contains \( \Omega \), \( f \) is continuously differentiable and its gradient \( g \) is Lipschitz continuous, namely, there exists a constant \( L > 0 \) such that

\[ \|g(x) - g(y)\| \leq L\|x - y\|, \forall x, y \in N. \]  

(3.1)

Since \( \{f(x_k)\} \) is decreasing, it is clear that the sequence \( \{x_k\} \) generated by Algorithm 1 and Algorithm 2 is contained in \( \Omega \). In addition, we can get from Assumption A that there exists a constant \( B \) and \( \gamma \), such that

\[ \|x\| \leq B, \|g(x)\| \leq \gamma, \forall x \in \Omega. \]  

(3.2)

In the latter part of the paper, we always suppose that the conditions in Assumption A hold. Then there is an useful lemma, which was originally given in [10,21].

**Lemma 3.1** Let \( \{x_k\} \) be generated by (1.2) and \( d_k \) is a descent direction. If \( \alpha_k \) is determined by the Wolfe line search (1.4), then we have

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

(3.3)

From Lemma 3.1 and (1.13) that for the MCD and NH3 methods with the Wolfe line search, we can easily obtain the following condition

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

(3.4)

We now establish the global convergence theorem for Algorithm 1 and Algorithm 2.

3.1. The Global Convergence of the H3 Method

For simplicity, here, we list the Theorem 2.3 in [14] as the following Lemma 3.2 without proof.

**Lemma 3.2** Suppose that \( x_0 \) is an initial point. Consider the method (1.2) and (1.3), where \( \alpha_k \) is computed by the Wolfe line search (1.4), and \( \beta_k \) is such that

\[ r_k \in [-1,1], \text{ where } r_k = \frac{\beta_k}{\beta_k^{LS}}, 0 = \frac{1-\sigma}{1+\sigma} > 0. \]

Then if \( g_k \neq 0 \) for all \( k \geq 0 \), we have that...
Further, the method converges in the sense that
\[ \lim_{k \to \infty} \|g_k\| = 0. \]
From the Lemma 3.2 above, similar to Corollary 2.4 in [14], we give the global convergence of the H3 method (Algorithm 1).

**Theorem 3.3** Suppose that \( x_0 \) is an initial point, Consider the Algorithm 1, then we have either \( \|g_k\| = 0 \) for some \( k \geq 0 \), or \( \lim \inf_{k \to \infty} \|g_k\| = 0 \).

**Proof** From the second inequality of (1.6) and the definitions of \( H \), \( CD \), and \( D \), it follows that
\[
306 \leq \beta_{k}^{CD} \leq \beta_{k}^{DY}.
\]
Therefore the statement follows lemma 3.2.

### 3.2. The Global Convergence of the MCD Method

Now, we establish the global convergence theorem for the MCD method.

**Theorem 3.4** Let \( \{x_k\} \) be generated by the MCD method (Algorithm 2, where \( d_k \) satisfies (1.15)), then we have \( \lim \inf_{k \to \infty} \|g_k\| = 0. \)

**Proof** Suppose by contradiction that the desired conclusion is not true, that is to say, there exists a constant \( \epsilon > 0 \) such that
\[
\|g_k\| \geq \epsilon, \forall k \geq 0. \quad (3.6)
\]
Set \( h_k = 1 + \beta_{k}^{CD} \frac{1}{2} \|d_{k-1}\| - h_k \|g_k\|^2 \). From (1.15) and (1.13), it follows that
\[
\|d_k\|^2 = \beta_{k}^{CD} \|d_{k-1}\|^2 - h_k \|g_k\|^2.
\]
(3.7)
Dividing both sides of (3.7) by \( \|g_k\|^2 \), we get, from (1.13) and the definition of \( \beta_{k}^{CD} \), that
\[
\frac{\|d_k\|^2}{\|g_k\|^2} = \left( \beta_{k}^{CD} \right)^2 \frac{\|d_{k-1}\|^2}{\|g_k\|^2} - 2 h_k \frac{1}{\|g_k\|^2} \|g_k\| \|g_k\|.
\]
(3.8)

### 3.3. The Global Convergence of the NH3 Method

The same as the Theorem 3.4, we can establish the following global convergence theorem for the NH3 method.

**Theorem 3.5** Let \( \{x_k\} \) be generated by the NH3 method (Algorithm 2, where \( d_k \) satisfies (1.16), then we have
\[
\lim \inf_{k \to \infty} \|g_k\| = 0. \quad (3.8)
\]

**Proof** Suppose by contradiction that the desired conclusion is false, that is to say, there exists a constant \( \epsilon > 0 \) such that
\[
\|g_k\| \geq \epsilon, \forall k \geq 0. \quad (3.9)
\]
Similar to (3.7), we get from (1.16) that
\[
\|d_k\|^2 = \left( \beta_{k}^{NH3} \right)^2 \|d_{k-1}\|^2 - 2 h_k \|d_k\|^2 - h_k \|g_k\|^2 \|g_k\|^2.
\]
(3.10)
where
\[
h_k = 1 + \beta_{k}^{NH3} \frac{1}{\|g_k\|^2}.
\]
Notice that
\[
\beta_{k}^{NH3} \leq \frac{1}{\|g_k\|^2}, \forall k \geq 0. \quad (3.11)
\]
Dividing both sides of (3.10) by \( \left( \beta_{k}^{CD} \right)^2 \), we get, from (3.11), (1.13) and the definition of \( \beta_{k}^{NH3} \), that
\[
\frac{\|d_k\|^2}{\|g_k\|^2} = \left( \beta_{k}^{CD} \right)^2 \frac{\|d_{k-1}\|^2}{\|g_k\|^2} - 2 h_k \frac{1}{\|g_k\|^2} \|g_k\| \|g_k\|.
\]
(3.12)
\[
\frac{\|d_k\|^2}{\|g_k\|^2} = \frac{\|d_k\|^2}{(g_k^T d_k)^2} = \left(\beta_k^H\right)^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{2h_k}{g_k^T d_k} \frac{\|g_k\|^2}{(g_k^T d_k)^2} \\
\leq \left(\beta_k^C D\right)^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{2h_k}{g_k^T d_k} \frac{\|g_k\|^2}{(g_k^T d_k)^2} \\
= \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^2} \frac{\|d_k\|^2}{(g_k^T d_k)^2} - \frac{2h_k}{g_k^T d_k} \frac{\|g_k\|^2}{(g_k^T d_k)^2} \leq \frac{1}{e^2} \sum_{i=0}^{k-1} \frac{1}{\|g_i\|^2} \leq k \frac{1}{e^2},
\]

which contradicts (3.4). This finishes the proof. 

4. Numerical Experiments

In this section, we carry out some numerical experiments. These three algorithms have been tested on some problems from [22]. The results are summarized in the following three tables: Table 1-3. For each test problem, No. is the number of the test problem in [22], \(x_0\) is the initial point, \(x_f\) the final point, \(k\) the number of times of iteration for each problem.

These three tables show the performance of these three methods relative to the iterations. It is easy to see that, for each algorithm, are all very efficient, especially for the problems such as s201, s207, s240, s311. The results for each problem are accurate, and with less number of times of iteration.

5. Conclusions

We have proposed three new hybrid conjugate gradient (CG) methods, that are, the H3 method, the MCD method and the NH3 method, where the last two methods produce sufficient descent search direction at every iteration. This property depends neither on the line search used nor on the convexity of the objective function. Under suitable conditions, we proposed the global convergence of these three new methods even for nonconvex minimization. And numerical experiments in section 4 showed that the new three algorithms are all efficient for the given test problems.

6. References

A. W. ZHOU ET AL.

No. 1, 1969, pp. 35-43.


Copyright © 2011 SciRes.