A hybrid conjugate gradient method for optimization problems

Xiangrong Li, Xupei Zhao

Department of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, China; xrl68@163.com

Received 14 October 2010; revised 18 November 2010; accepted 22 November 2010.

ABSTRACT

A hybrid method of the Polak-Ribière-Polyak (PRP) method and the Wei-Yao-Liu (WYL) method is proposed for unconstrained optimization problems, which possesses the following properties: i) This method inherits an important property of the well known PRP method: the tendency to turn towards the steepest descent direction if a small step is generated away from the solution, preventing a sequence of tiny steps from happening; ii) The scalar holds automatically; iii) The global convergence with some line search rule is established for nonconvex functions. Numerical results show that the method is effective for the test problems.

Keywords: Line Search; Unconstrained Optimization; Conjugate Gradient Method; Global Convergence

1. INTRODUCTION

We are interested to consider the unconstrained optimization problem

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

for (1.1), where \( x \) is the current iterate point, \( \alpha > 0 \) is a steplength, and \( d_k \) is the search direction designed by

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

where \( \beta_k \in \mathbb{R} \) is a scalar which determines the different conjugate gradient methods \([4,5,8,9,12,13,15,16,18,20,21,25,33]\) etc., and \( g_k \) is the gradient of \( f(x) \) at the point \( x \). The well-known formula for \( \beta_k \) from the computation point of view is the following PRP method

\[ \beta_k^{PRP} = \frac{g_k^T (g_{k+1} - g_k)}{\|g_k\|^2}, \]  \( (1.4) \)

where \( g_k \) and \( g_{k+1} \) are the gradients \( \nabla f(x_k) \) and \( \nabla f(x_{k+1}) \) of \( f(x) \) at the point \( x_k \) and \( x_{k+1} \), respectively, and \( \|g_k\| \) denotes the Euclidian norm of vectors. Throughout this paper, we also denote \( f(x_k) \) by \( f_k \).

Polak and Ribière [18] proved that this method with the exact line search is globally convergent when the objective function is convex. Powell [19] gave a counter example to show that there exist nonconvex functions on which the PRP method does not converge globally even the exact line search is used. He suggested that \( \beta_k \) should not be less than zero. Considering this suggestion, Gilbert and Nocedal [10] proved that the modified PRP method \( \beta_k = \max \{0, \beta_k^{PRP}\} \) is globally convergent with the weak Wolfe-Powell (WWP) line search technique and the assumption of sufficient descent condition. However, the global convergence of the PRP method is still open under the WWP line search rule.

Recently, Wei, Yao, and Liu (WYL) [21] propose a new conjugate gradient formula

\[ \beta_k^{WYL} = \frac{g_k^T (g_{k+1} - g_k)}{\|g_k\|^2} \]  \( (1.5) \)

It is not difficult to deduce that

\[ x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \ldots \]  \( (1.2) \)

This work is supported by China NSF grands 10761001 and Guangxi SF grands 0991028.
\[ \beta^W_{k+1} = \frac{g^T_{k+1} (g_{k+1} - g_k)}{g^T_k g_k} \geq \frac{\|g_{k+1}\|^2}{\|g_k\|^2} = 0 \]

is true. The numerical results show that this method is competitive to the PRP method for the test problems of [17]. Under the sufficient descent condition, this method is globally convergent with the WWP line search.

These observations make us know that the sufficient descent condition

\[ g^T_k d_k \leq -\epsilon \|g_k\|^2, \quad c > 0 \text{ is a constant holds for all } k \geq 0 \quad (1.6) \]

is very important to ensure the global convergence [1,2,10,14], and the scalar \( \beta \geq 0 \) also plays a very important role [10,19]. This motivates us to propose a hybrid method combining the PRP method and the WYL method. The hybrid method will possess some better properties of the PRP method and the WYL method: (i) the tendency to turn towards the steepest descent direction if a small step is generated away from the solution, preventing a sequence of tiny steps from happening; (ii) The scalar \( \beta \geq 0 \) holds automatically.

The global convergence with the WWP line search of the presented method is established for nonconvex objective function. Numerical results show that this given method is competitive to the PRP method and the WYL method.

This paper is organized as follows. In the next section, the algorithm is stated. The global convergence is proved in Section 3, and the numerical results are reported in Section 4. The last section gives one conclusion.

2. ALGORITHM

Now we describe the given algorithm as follows. Here we call it Algorithm 1.

**Algorithm 1** (The hybrid algorithm of the PRP method and the WYL method)

Step 0: Choose an initial point \( x_0 \in \mathbb{R}^n, \epsilon \in (0,1) \).

Set \( d_0 = -g_0 = -\nabla f(x_0), k = 0 \).

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

Step 2: Compute step size \( \alpha_k \) by some line search rules.

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

Step 4: Calculate the search direction

\[ d_{k+1} = -g_{k+1} + \beta^W_{k+1} d_k, \quad (2.1) \]

where \( \beta^W_{k} = \max \{ \beta^{PRP}_{k}, \beta^{WYL}_{k} \} \).

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

**Remark** i) If \( x_{k+1} = x_k \), we have \( g_{k+1} = g_k \) and \( \|g_{k+1}\| \geq \|g_k\| \) which imply that \( \beta^W_{k} \rightarrow 0 \), and \( \beta^{PRP}_{k} \rightarrow 0 \), which means that \( \beta^W_{k} \rightarrow 0 \) if a small step is generated for all \( k \geq 0 \). Thus the given method inherits the better property of the PRP method: the directions will turn out to be the steepest descent directions if the tiny steps from happening.

ii) By the definition of the new formula \( \beta^W_{k} \), we have

\[ \beta^W_{k} = \max \{ \beta^{PRP}_{k}, \beta^{WYL}_{k} \} \geq \beta^{WYL}_{k} \]

\[ \geq \frac{\|g_{k+1}\|^2}{\|g_k\|^2} = 0 \]

3. THE GLOBAL CONVERGENCE

The following assumptions are often needed to prove the convergence of the nonlinear conjugate gradient methods (see [5,9,10,20,21] etc.).

**Assumption 3.1** i) The function \( f(x) \) has a lower bound on the level set \( \Omega = \{ x \in \mathbb{R}^n | f(x) \leq f(x_0) \} \), where \( x_0 \) is a given point and \( \Omega \) is bounded.

ii) In an open convex set \( \Omega_0 \) that contains \( \Omega, J \) is differentiable and its gradient \( g \) is Lipschitz continuous, namely, there exists a constants \( L > 0 \) such that

\[ \|g(x) - g(y)\| \leq L \|x - y\|, \quad \forall x, y \in \Omega_0. \quad (3.1) \]

3.1. The global Convergence with the Weak Wolfe-Powell Line Search

The weak Wolfe-Powell (WWP) search rule is to find a step length \( \alpha_k \) such that

\[ f(x_k + \alpha_k d_k) \leq f_k + \delta \alpha_k g^T_k d_k \quad (3.2) \]

and

\[ g(x_k + \alpha_k d_k)^T d_k \geq \sigma g^T_k. \quad (3.3) \]

where \( \delta \in (0,1/2), \sigma \in (\delta, 1) \). This line search technique is often used to study the convergence of conjugate gradient algorithms [6,27,34]. At present, the global convergence of the PRP method with the WWP line search is still open.

**Lemma 3.1** Suppose that Assumption 3.1 holds. Let the sequence \( \{g_k\} \) and \( \{d_k\} \) be generated by Algorithm 1. \( g^T_k d_k \leq 0 \), and the stepsize \( \alpha_k \) be determined.
by the WWP line search (3.2) and (3.3) Then the
zoutendijk condition [34]
\[ \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty \] (3.4)
holds.

Proof. By (3.3) and Assumption 3.1 ii), we have
\[- (1 - \sigma) g_k^T d_k \leq (g_{k+1} - g_k)^T d_k \leq \alpha_k L \|d_k\|^2,\]
this means that \( \alpha_k \geq - (1 - \sigma) g_k^T d_k / L \|d_k\|^2 \),
which together with \( g_k^T d_k \leq 0 \), and (3.2) implies that
\[ \frac{(1 - \sigma)(g_k^T d_k)^2}{L \|d_k\|^2} \leq f_k - f_{k+1}, \]
summing up this inequality from \( k = 0 \) to \( \infty \), and using Assumption 3.1 i), we can obtain this lemma. This completes the proof.

We will prove the global convergence of Algorithm 1 by contradiction. Then we assume that there exists a positive constant \( \gamma > 0 \) such that
\[ \|g_k\| \geq \gamma, \quad \forall k \geq 0. \] (3.5)

Using (3.5) deduces a contradiction to obtain our conclusion.

Similar to Lemma 3.3.1 in [6], based on Assumption 3.1, Lemma 3.1, the fact \( \beta_k^{\text{PP}} \geq 0 \), and (3.5), we can get the following lemma.

Lemma 3.2 Let Assumption 3.1 hold and the sequences \( \{g_k\} \) and \( \{d_k\} \) be generated by Algorithm 1. The sufficient descent condition (1.6) holds, and the stepsize \( \alpha_k \) is determined by (3.2) and (3.3). Suppose that the inequalities (3.5) is true. Then we have \( d_k \neq 0 \) and
\[ \sum_{k=0}^{\infty} \|u_{k+1} - u_k\|^2 < \infty, \]
where \( u_k = \frac{d_k}{\|d_k\|} \).

Proof. These two inequalities (1.6) and (3.5) imply that \( d_k \neq 0 \) is true, for otherwise \( g_k = 0 \), then \( u_k = \frac{d_k}{\|d_k\|} \) is reasonable. Denote
\[ r_{k+1} = \frac{g_{k+1}}{d_{k+1}}, \quad \delta_k = \beta_k^{\text{PP}}, \quad \|d_k\| \]
By (2.1), for \( k \geq 0 \), we have
\[ u_{k+1} = r_{k+1} + \delta_k u_k, \]
this combining with \( \|u_{k+1}\| = \|u_k\| = 1 \) shows that
\[ \|r_{k+1}\| = \|u_{k+1} - \delta_k u_k\| = \|\delta_k u_{k+1} - u_k\| \] (3.6)
The inequality \( \beta_k^{\text{PP}} \geq 0 \) implies that \( \delta_k \geq 0 \) is true, then it follows that from (3.6) and triangular inequality
\[ \|u_{k+1} - u_k\| \]
\[ \leq \|1 + \delta_k\| u_{k+1} - (1 + \delta_k) u_k\|
\leq \|u_{k+1} - \delta_k u_k\| + \|\delta_k u_{k+1} - u_k\|
= 2 \|r_{k+1}\|. \] (3.7)
By (1.6) and (3.4), we get
\[ \sum_{k=0}^{\infty} \|d_{k+1}\|^2 = \sum_{k=0}^{\infty} \|r_{k+1}\|^2 \|g_{k+1}\|^2 < \infty \]
Which together with (3.5), we obtain
\[ \sum_{k=0}^{\infty} \|r_{k+1}\|^2 < \infty \]
By the above inequality and (3.7), we get this lemma. The proof is complete.

The following property (*) was introduced by Gilbert and Nocedal [10], which pertains to the WYL formula also has this property. Now we show that this property (*) pertains to our method.

Property (*). Suppose that
\[ 0 < \gamma \leq \|g_k\| \leq \gamma. \] (3.8)
We say that the method has Property (*), if for all \( k \), there exists constants \( b > 1 \) and \( \lambda > 0 \) such that
\[ \|\beta_k\| \leq b \]
and
\[ \|s_k\| \leq \lambda \Rightarrow \|\beta_k\| \leq \frac{1}{2b} \]

Lemma 3.3 Let Assumption 3.1 hold and the sequences \( \{g_k\} \) and \( \{d_k\} \) be generated by Algorithm 1. Then the new formula \( \beta_k^{\text{PP}} \) possesses property (*).

Proof. The result of this lemma is proved by the following two cases.

Case i: we consider \( \beta_k^{\text{PP}} \) By (3.1), we have
\[ \|\beta_k^{\text{PP}}\| = \frac{g_k^T (g_{k+1} - g_k)}{\|g_k\|^2} \leq \frac{L \|g_{k+1}\|^2}{\|g_k\|^2}. \] (3.9)
From Assumption 3.1 i), then there exists a constant \( M_1 > 0 \) such that
\[ \|s_k\| \leq M_1. \] (3.10)
Let \( b = \max \left\{ \frac{2}{L_2 (L_2^2 M_1)^{1/2}}, 1 \right\} \) and \( \lambda = \gamma_1^2 / 4b \), it follows that
\[ \|\beta_k^{\text{PP}}\| \leq b \]
and

Copyright © 2011 SciRes.
Let $\beta_k^{\text{PRP}}$ be generated by Algorithm 1 and the conditions in Lemma 3.3 hold. Then $\lim_{i \to \infty} \inf \|g_i\| = 0$.

**Proof.** We will get this theorem by contradiction. Suppose that (3.5) is true, then the conditions in Lemma 3.2 and 3.3 hold. By Assumption 3.1 i), then there exists a constant $\xi_0 > 0$ such that

$$\|x\| \leq \xi_0, \forall x \in \Omega \tag{3.13}$$

We also denote $u_i = d_i/\|d_i\|$, then for all integers $l,k (i \geq k)$, we have

$$x_i - x_{k+1} = \sum_{j=k+1}^{i} \|u_j\| u_i = \sum_{j=k+1}^{i} \|u_{j-1} + u_{k+1} - u_{k+1}\|.$$

Taking the norm in both sides of the above equality, and using (3.13) we get

$$\sum_{j=k+1}^{i} \|u_{j-1} - u_{k+1}\| \leq 2\xi_0 + \sum_{j=k+1}^{i} \|u_{j-1} - u_{k+1}\|$$

Let $\Delta = [8\xi_0/\lambda]$ be the smallest integer where $\Delta$ does not less than $8\xi_0/\lambda$. By Lemma 3.2, there exists an index $k_0$ such that

$$\sum_{j=k+1}^{i} \|u_{j-1} - u_{k+1}\| \leq \frac{1}{4\Delta} \tag{3.14}$$

On the other hand, by Lemma 3.3, there exists $k \geq k_0$ satisfying

$$k_{k,A}^d > \frac{\Delta}{2} \tag{3.15}$$

For all $i \in [k,k+\Delta-1]$, by Cauchy-Schwarz inequality and (3.14), we obtain

$$\|u_{i-1} - u_{k+1}\| \leq \sum_{j=k+1}^{i} \|u_j - u_{j-1}\| \leq (i-k)\left(\sum_{j=k+1}^{i} \|u_j - u_{j-1}\|^2\right)^{1/2} \leq \Delta^2 \left(\frac{1}{4\Delta}\right)^{1/2} = \frac{1}{2}.$$

By the above inequality, (3.15) and (3.13), we have

$$2\xi_0 > \frac{1}{2} \sum_{j=k+1}^{i} \|u_{j-1}\| > \frac{\lambda}{2} k_{k,A}^d > \frac{\lambda\Delta}{4},$$

Thus $\Delta < 8\xi_0/\lambda$, this contradicts with the definition of $\Delta$. Therefore, the conclusion of this theorem is right. This completes the proof.

**4. NUMERICAL RESULTS**

In this section, we report some numerical experiments.
The unconstrained optimization problems with the given initial points can be found at:


which were collected by Neculai Andrei. Since this new method is the hybrid method of the PRP method and the WYL method, we test Algorithm 1 with the WWP line search and compare its performance with those of the WYL [21] and the PRP [18] methods. The stop criteria are given below: we stop the program if the inequality \( \| g(x_k) \| \leq \varepsilon \) is satisfied or the inequality

\[
\| g(x_k) \| \leq \varepsilon (1 + |f(x_k)|)
\]

is satisfied, where \( \varepsilon = 1.0 \times 10^{-5} \). All the codes were written in Fortran and run on PC with 2.60 GHz CPU processor and 256 MB memory and Windows XP operation system. In the experiments, the parameters were chosen as \( \delta = 1.0 \times 10^{-2}, \sigma = 1.0 \times 10^{-1} \). The dimension of the test problems is from 500 to 5000. The detailed numerical results are listed on the web site


In Figure 1, “WYL”, “PRP”, and “MPRP-WYL” stand for the WYL method, the PRP method, and the new method, respectively.

Figure 1 shows the performance of these methods relative to the iterative number of the function and gradient(NFN), which were evaluated using the profiles of Dolan and Moré [7]. It is easy to see that the MPRP-WYL is predominant among these three methods and the new method can solve about 99% of the test problems successfully. The PRP method is better than the WYL method for \( 1 \leq t \leq 1.2 \) and the WYL method is better than the PRP method for \( 1.2 \leq t \leq 6 \). Moreover, the PRP method solves about 98% of the test problems and the WYL method solve about 99% of the test problems successfully, respectively. In a word, the given method is competitive to the other two methods and the hybrid formula is notable.

5. CONCLUSION

This paper gives a hybrid conjugate gradient method for solving unconstrained optimization. The global convergence for nonconvex functions with the WWP line search is established. The numerical results show that the given method is competitive to the other standard conjugate gradient methods for the test problems.

For further research, we should study the convergence of the new algorithm under other line search rules. Moreover, more numerical experiments and testing environments (such that [3]) for large practical problems should be done in the future.

REFERENCES


