

Research Article

A Hybrid of DL and WYL Nonlinear Conjugate Gradient Methods

Shengwei Yao^{1,2} and Bin Qin¹

¹ School of Information and Statistics, Guangxi University of Finance and Economics, Nanning 530003, China

² School of Science, East China University of Science and Technology, Shanghai 200237, China

Correspondence should be addressed to Shengwei Yao; idhot@163.com

Received 29 November 2013; Accepted 27 January 2014; Published 25 March 2014

Academic Editor: Abdon Atangana

Copyright © 2014 S. Yao and B. Qin. 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.

The conjugate gradient method is an efficient method for solving large-scale nonlinear optimization problems. In this paper, we propose a nonlinear conjugate gradient method which can be considered as a hybrid of DL and WYL conjugate gradient methods. The given method possesses the sufficient descent condition under the Wolfe-Powell line search and is globally convergent for general functions. Our numerical results show that the proposed method is very robust and efficient for the test problems.

1. Introduction

The nonlinear conjugate gradient (CG) method has played a special role in solving large-scale nonlinear optimization due to the simplicity of their iterations and their very low memory requirements. In fact, the CG method is not among the fastest or most robust optimization algorithms for nonlinear problems available today, but it remains very popular for engineers and mathematicians who are interested in solving large-scale problems. As we know, the nonlinear conjugate gradient method is the extended based on linear conjugate gradient method. The first linear conjugate gradient method is proposed by Hestenes and Stiefel 1952 [1] for solving linear equations. In 1964, Fletcher and Reeves [2] extended it to nonlinear problems and get the first nonlinear conjugate gradient method (FR method).

In this paper, we focus on solving the following nonlinear unconstrained optimization problem by conjugate gradient method:

$$\min \{f(x) : x \in R^n\}, \quad (1)$$

where $f(x) : R^n \rightarrow R$ is a smooth, nonlinear function whose gradient will be denoted by $g(x)$. The iterative formula of the conjugate gradient method is given by

$$x_k = x_{k-1} + \alpha_{k-1} d_{k-1}, \quad (2)$$

where, d_{k-1} is the search direction at x_{k-1} and α_{k-1} is the step-length. For nonlinear conjugate gradient method, d_k is computed by

$$d_k = -g_k + \beta_k d_{k-1}, \quad d_1 = -g_1, \quad (3)$$

where β_k is a scalar and $g_k = \nabla f(x_k)$ denotes the gradient. Different conjugate gradient methods correspond to the different ways to compute β_k . Some well-known formulae for β_k are the Fletcher-Reeves (FR), Polak-Ribière (PR), Hestense-Stiefel (HS), Dai-Yuan (DY), and CG-DESCENT, which are given, respectively, by

$$\beta_k^{\text{FR}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \quad (4)$$

(see [2]),

$$\beta_k^{\text{PR}} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} \quad (5)$$

(see [3]),

$$\beta_k^{\text{HS}} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} \quad (6)$$

(see [1]),

$$\beta_k^{\text{DY}} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} \quad (7)$$

(see [4]),

$$\beta_k^{\text{N}} = \left(y_{k-1} - 2d_{k-1} \frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}} \right)^T \frac{g_k}{d_{k-1}^T y_{k-1}} \quad (8)$$

(see [5]), where $y_{k-1} = g_k - g_{k-1}$ denotes the gradient change.

Although all these methods are equivalent in the linear case, namely, when f is a strictly convex quadratic function and α_k are determined by exact line search, their behaviors for general objective functions may be far different. For general functions, Zoutendijk [6] proved the global convergence of FR method with exact line search. (Here and throughout this paper, for global convergence, we mean the sequence generated by the corresponding methods will either terminate after finite steps or contain a subsequence such that it converges to a stationary point of the objective function from a given initial point.) Although one would be satisfied with its global convergence properties, the FR method performs much worse than the PR (HS) method in real computations. Powell [7] analyzed a major numerical drawback of the FR method, namely, if a small step is generated away from the solution point, the subsequent steps may be also very short. On the other hand, in practical computation, the HS method resembles the PR method, and both methods are generally believed to be the most efficient conjugate gradient methods since these two methods essentially perform a restart if a bad direction occurs. However, Powell [8] constructed a counterexample and showed that the PR method and HS method can cycle infinitely without approaching the solution. This example suggests that these two methods have a drawback that they are not globally convergent for general functions. Therefore, over the past few years, much effort has been put to find out new formulas for conjugate methods such that they are not only globally convergent for general functions but also have good numerical performance. The similar counterexamples are also constructed by Dai and Yuan [9].

From the structure of the above formulae β_k , we know that β_k^{FR} and β_k^{DY} have the common numerator $\|g_k\|^2$. They are globally convergent if the objective function Lipschitz continuous and the level set $\{x \in R^n : f(x) < f(x_0)\}$ is bounded. For inexact line search, Al-Baali [10] proved the global convergence of FR method under the strong Wolfe-Powell line search with the restriction $\sigma < 1/2$. Based on Al-Baali's result, Liu et al. [11] extended the global convergence of FR method to the case $\sigma = 1/2$. Dai and Yuan [12] proved that the sufficient descent condition must hold for one of the directions d_{k+1} and d_k , and proposed the global convergence of FR method with general Wolfe line searches.

β_k^{PR} and β_k^{HS} share the common numerator $g_k^T y_{k-1}$, they possess a built-in restart feature to avoid the jamming problem as follows: when the step $x_k - x_{k-1}$ is small, the factor y_{k-1} in the numerator of β_k tends to be zero. Hence, the next search

direction d_k is essentially the steepest descent direction $-g_k$. So, the numerical performance of these methods is better than the performance of the methods with $\|g_k\|^2$ in numerator of β_k . In [3] Polak and Ribière proved that if the objective function is strongly convex and line search is exact, the PR method is globally convergent. For general functions, Powell [7, 8] analyzed the convergence properties of PR method and constructed an example which shows that the PR method may cycle infinitely between nonstationary points. To get the global convergence, Gilbert and Nocedal [13] made the following nonnegative restriction on β_k :

$$\beta_k^{\text{PR+}} = \max \{ \beta_k^{\text{PR}}, 0 \}. \quad (9)$$

Generally speaking, methods with numerator $g_k^T y_{k-1}$ possess better convergence than the methods with numerator $\|g_k\|^2$. But from numerical performance point of view, methods with numerator $\|g_k\|^2$ outperform the methods with the numerator $g_k^T y_{k-1}$. So, a lot of effort has been made to find the method which has nice convergence properties and efficient numerical performance in the past decades. In [14], authors proposed a new conjugacy condition which made use of not only gradient values but also function values. Based on the given conjugacy condition, a class of nonlinear conjugate gradient methods is proposed. The PR method outperforms a lot of methods in numerical experiments, but it does not possess the sufficient descent condition. So, some modified forms of PR method have been studied in [15, 16], the given methods possess the sufficient descent condition and are globally convergent for general functions.

2. Motivations and the New Formula

Since the PR method is considered as one of the most efficient nonlinear conjugate gradient methods, a lot of effort has been made on its convergence properties and its modifications. In [13], with the sufficient descent assumption, Gilbert and Nocedal proved the global convergence of PR+ method under the Wolfe line search. Grippo and Lucidi [17] constructed an Armijo-type line search and proved that under this line search, directions d_k generated by PR method satisfy the sufficient descent condition.

Recently, Wei et al. [18] and Huang et al. [19] gave a modified formula for computing β_k as follows:

$$\beta_k^{\text{WYL}} = \frac{g_k^T \widetilde{y_{k-1}}}{\|g_k\|^2}, \quad (10)$$

where $\widetilde{y_{k-1}} = g_k - (\|g_k\|/\|g_{k-1}\|)g_{k-1}$.

The method with formula β_k^{WYL} not only has nice numerical results but also possess the sufficient descent condition and global convergence properties under the strong Wolfe-Powell line search. From the structure of β_k^{WYL} , we know that, the method with β_k^{WYL} can also avoid jamming such that when the step $x_k - x_{k-1}$ is small, $\|g_k\|/\|g_{k-1}\|$ tends to be 1 and the next search direction tends to be the steepest descent direction which is similar to PR method. But WYL

method has some advantages, such as under strong Wolfe-Powell line search, $\beta_k^{\text{WYL}} \geq 0$, and if the parameter $\sigma \leq 1/4$ in SWP, WYL method possesses the sufficient descent condition which deduces the global convergence of the WYL method.

In [20, 21], Shengwei et al. and Huang et al. extended such modification to HS method as follows:

$$\beta_k^{\text{MHS}} = \frac{g_k^T y_{k-1}^*}{d_{k-1}^T y_{k-1}}, \quad y_{k-1}^* = g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}. \quad (11)$$

The above formulae β_k^{WYL} and β_k^{MHS} can be considered as the modification form of β_k^{PR} and β_k^{HS} by using y_{k-1}^* to replace y_{k-1} , respectively. In [20, 21], the corresponding methods are proved to be globally convergent for general functions under the strong Wolfe-Powell line search and Grippo-Lucidi line search. Based on the same approach, some authors extended other discussions and modifications in [22–24]. In fact, y_{k-1}^* is not our point at the beginning, our purpose is involving the information of the angle between g_k and g_{k-1} . From this point of view, β_k^{WYL} has the following form:

$$\beta_k^{\text{WYL}} = \beta_k^{\text{FR}} (1 - \cos(\overline{\theta}_k)), \quad (12)$$

where $\overline{\theta}_k$ is the angle between g_k and g_{k-1} . By multiplying β_k^{FR} with $1 - \cos \overline{\theta}_k$, the method not only has similar convergence properties with FR method, but also avoid jamming which is similar to PR method.

Recently, Dai and Liao [25] proposed a new conjugacy condition which is based on the Quasi-Newton techniques. According to the new conjugacy condition, the following formula β_k^{DL} is given:

$$\beta_k^{\text{DL1}} = \frac{g_k^T (y_{k-1} - t s_{k-1})}{d_{k-1}^T y_{k-1}}, \quad (13)$$

where $t \geq 0$, for simplicity, we call the method with (13) as DL1 method. It is obviously that

$$\beta_k^{\text{DL1}} = \beta_k^{\text{HS}} - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}. \quad (14)$$

In [25], for the method with β_k^{DL1} , if the line search is exact, DL1 method has the same convergence properties with PR method, which indicates that DL1 method does not converge for general functions. To get the global convergence, Dai and Liao replace (13) by

$$\beta_k^{\text{DL}} = \max\{\beta_k^{\text{HS}}, 0\} - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}. \quad (15)$$

The formula (15) can be considered as a modified form of β_k^{HS} , by adding the part $t(g_k^T s_{k-1}/d_{k-1}^T y_{k-1})$ which may contain some information of Hessian $\nabla^2 f(x)$ [25]. From the convergence analysis in [25], the nonnegative restriction $\max\{\beta_k^{\text{HS}}, 0\}$ and the sufficient descent condition are significant for the global convergence results.

Motivated by the above discussion, in this paper, we give the following formula to compute the parameter β_k :

$$\beta_k^* = \beta_k^{\text{WYL}} - t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}. \quad (16)$$

The formula β_k^* can be considered a modification of β_k^{WYL} , namely, by adding $t(g_k^T s_{k-1}/d_{k-1}^T y_{k-1})$, the β_k^* may contain some Hessian information [25]. It also can be considered as a modified form of β_k^{DL} by substituting $\max\{\beta_k^{\text{HS}}, 0\}$ with β_k^{WYL} . We call the method with (2), (3), and (16) as WYLDL method and give the corresponding algorithm as follows.

Algorithm 1 (WYLDL).

Step 1. Given $x_1 \in R^n$, $\varepsilon \geq 0$, set $d_1 = -g_1$, $k = 1$, if $\|g_1\| \leq \varepsilon$, then stop;

Step 2. Compute α_k by the Strong Wolfe-Powell line search;

Step 3. Let $x_k = x_{k-1} + \alpha_{k-1} d_{k-1}$, $g_k = g(x_k)$, if $\|g_k\| \leq \varepsilon$, then stop;

Step 4. Compute β_k by (16) and generate d_k by (3);

Step 5. Set $k := k + 1$, go to Step 2.

The convergence properties of Algorithm 1 will be discussed in Section 3.

3. Convergence Analysis

For conjugate gradient methods, during the iteration process, the gradient of the objective function is required. We make the following basic assumptions on the objective functions.

Assumption 2. (i) The level set $\Gamma = \{x \in R^n : f(x) \leq f(x_1)\}$ is bounded, namely, there exists a constant $B > 0$ such that

$$\|x\| \leq B, \quad \forall x \in \Gamma. \quad (17)$$

(ii) In some neighborhood N of Γ , f is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L \|x - y\|, \quad \forall x, y \in N. \quad (18)$$

Under the above assumptions of f , there exists a constant $\bar{\gamma} \geq 0$ such that

$$\|\nabla f(x)\| \leq \bar{\gamma}, \quad \forall x \in \Gamma. \quad (19)$$

Exact Line Search. Suppose that d_k is a descent direction and step length α_k is the solution of

$$\min_{\alpha_k > 0} f(x_k + \alpha d_k). \quad (20)$$

For exact line search,

$$g_k^T d_{k-1} = 0. \quad (21)$$

Form (16) and (21), we can get that $\beta_k^* = \beta_k^{\text{WYL}}$. In [18], author proved that if the line search is exact, the method with β_k^{WYL} is globally convergent for the uniformly convex functions.

For conjugate gradient methods, the sufficient descent condition is significant to the global convergence. We say the sufficient descent condition holds if there exists a constant $c > 0$ such that

$$g_k^T d_k \leq -c \|g_k\|^2, \quad \forall k. \quad (22)$$

In nonlinear optimization algorithm, the Strong Wolfe-Powell conditions, namely,

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \quad (23)$$

$$|g(x_k + \alpha_k d_k)^T d_k| \leq \sigma |g_k^T d_k|, \quad (24)$$

where $0 < \delta < \sigma < 1$, are often imposed on the line search.

In [25], Dai and Liao proved that, if directions d_k satisfy the sufficient descent condition (22), the DL method is globally convergent under the strong Wolfe-Powell line search for general functions. In this section, we will prove that the directions d_k generated by Algorithm 1 satisfy the sufficient descent condition (22). Based on this result, the global convergence of Algorithm 1 will be established.

Lemma 3. *Suppose that the sequence $\{x_k\}$ is generated by Algorithm 1, the step-length α_k satisfy the strong Wolfe-Powell conditions (23) and (24), if $0 < \sigma < 1/4$; then, the generated directions d_k satisfy the sufficient descent condition (22).*

Proof. We prove this result by induction. By using (3), we have $g_k^T d_k = -\|g_k\|^2 + \beta_k^{\text{WYLDL}} g_k^T d_{k-1}$, and combining this equation with (β_k^{WYLDL}) , we can deduce that

$$\begin{aligned} \frac{g_k^T d_k}{\|g_k\|^2} &= -1 + \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} (1 - \cos(\bar{\theta}_k)) \\ &\quad - t \frac{\alpha_k (g_k^T d_{k-1})^2}{\|g_k\|^2 d_{k-1}^T y_{k-1}}. \end{aligned} \quad (25)$$

By strong Wolfe-Powell condition (24), it follows that $d_{k-1}^T y_{k-1} \geq (\sigma - 1) g_{k-1}^T d_{k-1} > 0$. Which means that

$$t \frac{\alpha_k (g_k^T d_{k-1})^2}{\|g_k\|^2 d_{k-1}^T y_{k-1}} > 0. \quad (26)$$

From (25), the following inequality holds:

$$\frac{g_k^T d_k}{\|g_k\|^2} \leq -1 + \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} (1 - \cos(\bar{\theta}_k)). \quad (27)$$

(25) and Wolfe-Powell condition (24) deduce that

$$\frac{g_k^T d_k}{\|g_k\|^2} \leq -1 + \frac{-\sigma g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} (1 - \cos(\bar{\theta}_k)), \quad (28)$$

namely,

$$\frac{g_k^T d_k}{\|g_k\|^2} \leq -1 + \frac{-2\sigma g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2}. \quad (29)$$

The repeating of (29) can deduce that

$$\frac{g_k^T d_k}{\|g_k\|^2} \leq -2 + \sum_{j=0}^{k-1} (2\sigma)^j. \quad (30)$$

Since

$$\sum_{j=0}^{k-1} (2\sigma)^j < \sum_{j=0}^{\infty} (2\sigma)^j = \frac{1}{1 - 2\sigma}, \quad (31)$$

(30) can be expressed as

$$\frac{g_k^T d_k}{\|g_k\|^2} \leq -2 + \frac{1}{1 - 2\sigma}. \quad (32)$$

With the restrictions $\sigma \in (0, 1/4)$ and $g_1^T d_1 = -\|g_1\|^2$, for $c = (1 - 4\sigma)/(1 - 2\sigma)$, the inequality (32) means that

$$g_k^T d_k \leq -c \|g_k\|^2. \quad (33)$$

□

For PR method, when a small step-length occurs, β_k^{PR} will tend to be zero, and the next search direction d_k automatically approaches to $-g_k$. By such way, the PR method automatically avoids jamming. This property was first studied by Gilbert and Nocedal [13], which is called Property (*). We are going to show that the method with β_k^{WYLDL} possesses such property (*).

Property 1. (*) Consider a method of form (2) and (3). Suppose that

$$0 < \gamma \leq \|g_k\| \leq \bar{\gamma} \quad \forall k \geq 1. \quad (34)$$

We say that the method has Property (*), if for all k , there exist constants $b > 1$, $\lambda > 0$ such that $|\beta_k| \leq b$, and if $\|s_{k-1}\| \leq \lambda$, we have $|\beta_k| \leq 1/2b$.

Lemma 4. *Consider a method of form (2) and (3). If β_k is determined by β_k^{WYLDL} , α_k satisfies the Wolfe-Powell condition (24), then the method possesses Property 1(*).*

Proof. By Lemma 3, we know that the sufficient descent condition (22) holds. Combining with Wolfe-Powell condition (24), we have

$$d_{k-1}^T y_{k-1} \geq (\sigma - 1) g_{k-1}^T d_{k-1} \geq (1 - \sigma) c \gamma^2. \quad (35)$$

It follows from (17), (35), and β_k^{WYLDL} that

$$\begin{aligned}
 & \left| \beta_k^{\text{WYLDL}} \right| \\
 & \leq \frac{\|g_k\| (\|g_k - g_{k-1}\| + \|g_{k-1} - (\|g_k\| / \|g_{k-1}\|) g_{k-1}\|)}{\|g_{k-1}\|^2} \\
 & \quad + t \frac{\|g_k\| \|s_{k-1}\|}{\left| d_{k-1}^T y_{k-1} \right|} \\
 & \leq \frac{\|g_k\| (2\|g_k - g_{k-1}\|)}{\|g_{k-1}\|^2} + t \frac{\|g_k\| \|s_{k-1}\|}{c(1-\sigma)\|g_{k-1}\|^2} \\
 & \leq \frac{\|g_k\| (2L\|s_{k-1}\|)}{\|g_{k-1}\|^2} + t \frac{\|g_k\| \|s_{k-1}\|}{c(1-\sigma)\|g_{k-1}\|^2} \\
 & \leq \frac{2B\bar{\gamma}(c(1-\sigma)2L+t)}{\gamma^2 c(1-\sigma)}.
 \end{aligned} \tag{36}$$

Set

$$b := \frac{2B\bar{\gamma}(c(1-\sigma)2L+t)}{\gamma^2 c(1-\sigma)}, \tag{37}$$

which means that $|\beta_k^{\text{WYLDL}}| < b$. By setting

$$\lambda := \frac{\gamma^2 c(1-\sigma)}{2b\bar{\gamma}(c(1-\sigma)2L+t)}, \tag{38}$$

we have

$$\begin{aligned}
 \left| \beta_k^{\text{WYLDL}} \right| & < \frac{\|s_{k-1}\| \bar{\gamma}(c(1-\sigma)2L+t)}{\gamma^2 c(1-\sigma)}, \\
 & < \frac{1}{2b}.
 \end{aligned} \tag{39}$$

□

For nonlinear conjugate gradient methods, Dai et al. [4] proposed the following general conclusion.

Lemma 5. *Suppose that Assumption 2 holds. Consider any conjugate gradient method, where d_k is a descent direction and α_k is obtained by the strong Wolfe-Powell line search. if*

$$\sum \frac{1}{\|d_k\|^2} = \infty, \tag{40}$$

we have

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \tag{41}$$

By Lemma 3, we know that the method with β_k^{WYLDL} possesses the sufficient descent condition under the Wolfe-Powell line search. Combining with Lemma 5, we can have the following theorem.

Theorem 6. *Suppose that Assumption 2 holds. Consider WYLDL method, where α_k is obtained by strong Wolfe-Powell*

lien search with $\sigma < 1/4$. If there exists a constant $\gamma > 0$ such that

$$\|g_k\| \geq \gamma, \quad \forall k \geq 1, \tag{42}$$

then $d_k \neq 0$ and

$$\sum_{k \geq 2} \|u_k - u_{k-1}\|^2 < \infty, \tag{43}$$

where $u_k = d_k / \|d_k\|$.

Proof. Firstly, note that $d_k \neq 0$; otherwise, the sufficient descent condition (22) fails. Therefore, u_k is well defined. In addition, by relation (42) and Lemma 5 we have

$$\sum_{k \geq 1} \frac{1}{\|d_k\|^2} < \infty. \tag{44}$$

Now, we divide formula β_k^{WYLDL} into two parts as follows:

$$\beta_k^1 = \beta_k^{\text{WYL}}, \quad \beta_k^2 = -t \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}, \tag{45}$$

and define

$$r_k := \frac{\vartheta_k}{\|d_k\|}, \quad \delta_k := \beta_k^1 \frac{\|d_{k-1}\|}{\|d_k\|}, \tag{46}$$

where $\vartheta_k = -g_k + \beta_k^2 d_{k-1}$.

Then by (3), we have, for all $k \geq 2$,

$$u_k = r_k + \delta_k u_{k-1}. \tag{47}$$

Using the identity $\|u_k\| = \|u_{k-1}\| = 1$ and (47), we can obtain

$$\|r_k\| = \|u_k - \delta_k u_{k-1}\| = \|\delta_k u_k - u_{k-1}\|. \tag{48}$$

Using the condition $\delta_k = \beta_k^{\text{WYL}} (\|d_{k-1}\| / \|d_k\|) = (\|g_k\|^2 / \|g_{k-1}\|^2) (1 - \cos(\theta_k)) (\|d_{k-1}\| / \|d_k\|) \geq 0$, the triangle inequality, and (48), we obtain

$$\begin{aligned}
 \|u_k - u_{k-1}\| & \leq \|(1 + \delta) u_k - (1 + \delta) u_{k-1}\| \\
 & \leq \|u_k - \delta_k u_{k-1}\| + \|\delta_k u_k - u_{k-1}\| \\
 & = 2 \|r_k\|.
 \end{aligned} \tag{49}$$

On the other hand, line search condition (24) gives

$$y_{k-1}^T d_{k-1} \geq (\sigma - 1) g_{k-1}^T d_{k-1}. \tag{50}$$

Equations (22), (24), and (50) imply that

$$\left| \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} \right| \leq \frac{\sigma}{1-\sigma}. \tag{51}$$

It follows from the definition of ϑ_k , (17), (36), and (51) that

$$\begin{aligned}
 \|\vartheta_k\| & \leq \|g_k\| + t \left| \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}} \right| \|d_{k-1}\| \\
 & = \|g_k\| + t \left| \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} \right| \|s_{k-1}\| \\
 & \leq \bar{\gamma} + t \frac{\sigma}{1-\sigma} 2B.
 \end{aligned} \tag{52}$$

So we have

$$\begin{aligned} \sum \|u_k - u_{k-1}\|^2 &\leq 4 \sum \|r_k\|^2 \\ &\leq 4 \sum \frac{\vartheta_k^2}{\|d_k\|^2} \\ &\leq 4 \left(\bar{\gamma} + t \frac{\sigma}{1-\sigma} 2B \right)^2 \sum \frac{1}{\|d_k\|^2} \\ &< \infty. \end{aligned} \quad (53)$$

Let N^* denote the set of positive integers. For $\lambda > 0$ and a positive integer Δ , denote

$$K_{k,\Delta}^\lambda := \{i \in N^* : k \leq i \leq k + \Delta - 1, \|s_{k-1}\| > \lambda\}. \quad (54)$$

Let $|K_{k,\Delta}^\lambda|$ denote the number of elements in $K_{k,\Delta}^\lambda$. Dai and Liao [25] pointed out that for conjugate gradient method which satisfies

- (i) Property 1(*);
- (ii) the sufficient descent condition;
- (iii) Theorem 6;

if (42) holds, then the small step-sizes should not be too many. This property is described as follows. \square

Lemma 7. *Suppose that Assumption 2 holds. Consider WYLDL method, where α_k is obtained by the strong Wolfe-Powell line search in which $\sigma < 1/4$. Then if (42) holds, there exists $\lambda > 0$ such that, for any $\Delta \in N^*$ and any index k_0 , there is an index $k \geq k_0$ such that*

$$|K_{k,\Delta}^\lambda| > \frac{\Delta}{2}. \quad (55)$$

Proof. It follows from Lemmas 3 and 4 and Theorem 6 that WYLDL method possesses the above three conditions in [25]. So, according to Lemma 3.5 in [25], the Lemma 7 holds. We omit the detailed proof of this Lemma 7. \square

According to the above lemmas and theorems, we can prove the following convergence theorem for WYLDL method.

Theorem 8. *Suppose that Assumption 2 holds. Consider WYLDL method, if α_k is obtained by strong Wolfe-Powell line search with $\sigma < 1/4$, then we have*

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (56)$$

Proof. We prove this theorem by contradiction. If $\liminf_{k \rightarrow \infty} \|g_k\| > 0$, then (42) must hold. Then the conditions of Theorem 6 and Lemma 7 hold. Defining $u_i = d_i/\|d_i\|$, we have, for any indices l, k , with $l \geq k$,

$$\begin{aligned} x_l - x_{k-1} &= \sum_{i=k}^l x_i - x_{i-1} \\ &= \sum_{i=k}^l \alpha_{i-1} d_{i-1} = \sum_{i=k}^l u_{i-1} \|s_{i-1}\| \\ &= \sum_{i=k}^l \|s_{i-1}\| u_{k-1} + \sum_{i=k}^l \|s_{i-1}\| (u_{i-1} - u_{k-1}). \end{aligned} \quad (57)$$

(57), $\|u_i\| = 1$, and (17) give the following:

$$\begin{aligned} \sum_{i=k}^l \|s_{i-1}\| &\leq \|x_l - x_{k-1}\| \\ &+ \sum_{i=k}^l \|s_{i-1}\| \|u_{i-1} - u_{k-1}\| \\ &\leq 2B + \sum_{i=k}^l \|s_{i-1}\| \|u_{i-1} - u_{k-1}\|. \end{aligned} \quad (58)$$

Let $\lambda > 0$ be given by Lemma 7 and define $\Delta := \lceil 8B/\lambda \rceil$ to be the smallest integer not less than $8B/\lambda$. By Theorem 6, we can find an index $k_0 \geq 1$ such that

$$\sum_{i \geq k_0} \|u_{i-1} - u_{k-1}\|^2 \leq \frac{1}{4\Delta}. \quad (59)$$

With this Δ and k_0 , Lemma 7 gives an index $k \geq k_0$ such that

$$|K_{k,\Delta}^\lambda| > \frac{\Delta}{2}. \quad (60)$$

For any index $i \in [k, k + \Delta - 1]$, by Cauchy-Schwartz inequality and (59),

$$\begin{aligned} \|u_i - u_{k-1}\| &\leq \sum_{j=k}^i \|u_j - u_{j-1}\| \\ &\leq (i - k + 1)^{1/2} \left(\sum_{j=k}^i \|u_j - u_{j-1}\|^2 \right)^{1/2} \\ &\leq \Delta^{1/2} \left(\frac{1}{4\Delta} \right)^{1/2} = \frac{1}{2}. \end{aligned} \quad (61)$$

From these relations (61) and (60) and taking $l = k + \Delta - 1$ in (58), we get

$$2B \geq \frac{1}{2} \sum_{i=k}^{k+\Delta-1} \|s_{i-1}\| > \frac{\lambda}{2} |K_{k,\Delta}^\lambda| > \frac{\lambda\Delta}{4}. \quad (62)$$

Thus $\Delta < 8B/\lambda$, which contradicts the definition of Δ . The proof is completed. \square

4. Numerical Experiments

In this section, we report the performance of the Algorithm 1 (WYLDL) on a set of test problems. The codes were written in Fortran 77 and in double precision arithmetic. All the tests were performed on the same PC. The experiments were performed on a set of 73 nonlinear unconstrained problems in [26]. Some of the problems are from CUTE [27] library. For each test problem, we have performed 10 numerical experiments with number of variables $n = 1000, 2000, \dots, 10000$.

In order to assess the reliability of the WYLDL algorithm, we also tested this method against the DL method and WYL method using the same problems. All these algorithms are terminated when $\|g_k\| \leq 10^{-5}$. We also force stopped the routines if the iterations exceeded 1000 or the number of function evaluations reached 2000. In the Wolfe-Powell line search conditions (23) and (24), the parameters are $\delta = 10^{-4}$, $\sigma = 10^{-1}$. For DL method, $t = 0.1$, which is the same with [25]. We also test WYLDL algorithm with $t = 0.1$ which is the best choice.

The comparing data contain the iterations, function and gradient evaluations, and CPU time. To approximately assess the performance of WYLDL, WYL and DL methods, we use the profile of Dolan and Moré [28] as an evaluated tool.

Dolan and Moré [28] gave a new tool to analyze the efficiency of algorithms. They introduced the notion of a performance profile as a means to evaluate and compare the performance of the set of solvers S on a test set P . Assuming that there exist n_s solvers and n_p problems, for each problem p and solver s , they defined that

$t_{p,s}$ = computing cost (iterations or function and gradient evaluations or CPU time) is required to solve problem p by solver s .

Requiring a baseline for comparisons, they compared the performance on problem p by solver s with the best performance by any solver on this problem; that is, using the performance ratio as follows:

$$r_{p,s} = \frac{t_{p,s}}{\min \{t_{p,s} : s \in S\}}. \tag{63}$$

Suppose that a parameter $M \geq r_{p,s}$ for all p, s . Set $r_{p,s} = M$ if and only if solver s does not solve problem p . Then they defined

$$\rho_s(\tau) = \frac{1}{n_p} \text{size} \{p \in P : r_{p,s} \leq \tau\}, \tag{64}$$

thus $\rho_s(\tau)$ is the probability for solver s that a performance ratio $r_{p,s}$ is within a factor $\tau \geq 1$ of the best possible ratio. Then function ρ_s is the distribution function for the performance ratio. The performance profile ρ_s is a nondecreasing, piecewise constant function. That is, for subset of the methods being analyzed, we plot the fraction P of the problems for which any given method is within a factor τ of the best.

For the testing problems, if all three methods cannot terminate successfully, then we got rid of them. In case one method fails, but there are other methods that terminate

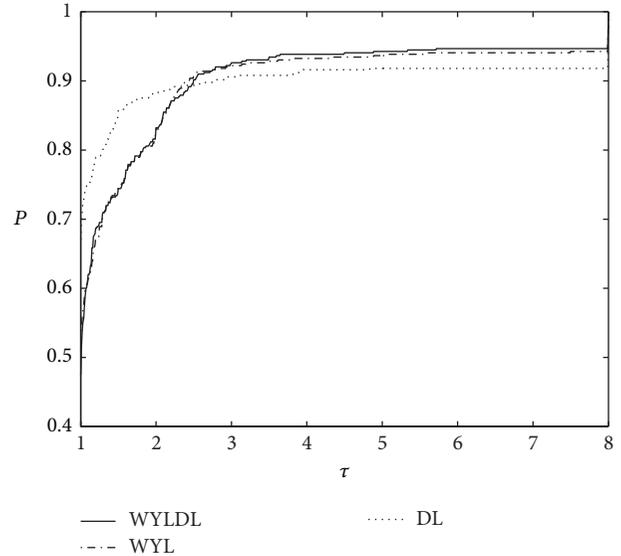


FIGURE 1: Performance profiles based on iterations.

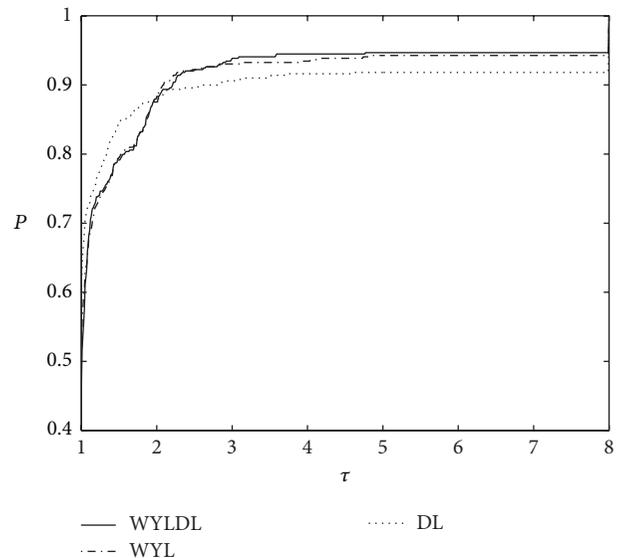


FIGURE 2: Performance profiles based on function and gradient evaluations.

successfully, then the performance ratio of the failed method is set to be M (M are the maxima of the performance ratios). The performance profiles based on iterations, function and gradient evaluations, and CPU-time of the three methods are plotted in Figures 1, 2, and 3, respectively.

From Figure 1, which plots the performance profile based on iterations, when $\tau = 1$, the DL method performs better than WYL and WYLDL methods. With the increasing of τ , when $\tau \geq 2.2$, the profiles of WYLDL and WYL methods outperform DL method. This means that, from the iteration point of view, for a subset of problems, DL method is better than WYL and WYLDL methods. But, for all the testing problems, WYLDL method is more robust than DL method.

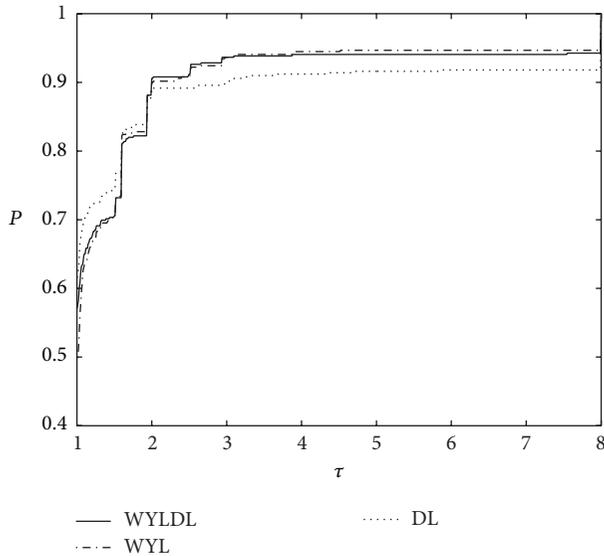


FIGURE 3: Performance profiles based on CPU time.

From Figure 2, which plots the performance profile based on function and gradient evaluations, it can be seen that for $\tau < 2$, DL method performs better than WYL and WYLDL methods. Comparing with Figure 1, the difference of these methods is much less than the iterations' profile. One of the possible reason is as follows: for WYLDL and WYL methods, the average times of function and gradient evaluations required during the iterations are less than DL method. From this point of view, the CPU time consumed by WYLDL or WYL methods should be less than DL method, since the CPU time is mainly dependent on function and gradient evaluations. Figure 3 validates this phenomenon. From Figures 1 to 3, it is easy to see that the performances of WYL method and WYLDL method are quite similar. The possible reason I think is that the second part of β_k^{WYLDL} : $-t(g_k^T s_{k-1}/d_{k-1}^T y_{k-1})$, $t = 0.1$ is very small compared with β_k^{WYL} . One of the reasons may be relevant to the Wolfe line search. Since the line search used in this paper is based on Lemarechal [29], Fletcher [30], or More and Thuente's [31] strategy, this may make the directional derivative $|g_k^T d_{k-1}|$ very small.

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 the Guangxi Universities Foundation Grant no. 2013BYB210.

References

[1] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *Journal of Research of the National Bureau of Standards*, vol. 49, pp. 409–436, 1952.

- [2] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *The Computer Journal*, vol. 7, pp. 149–154, 1964.
- [3] E. Polak and G. Ribière, "Note sur la convergence de méthodes de directions conjuguées," *Revue Française d'Informatique et de Recherche Opérationnelle*, vol. 3, no. 16, pp. 35–43, 1969.
- [4] Y. Dai, J. Han, G. Liu, D. Sun, H. Yin, and Y.-X. Yuan, "Convergence properties of nonlinear conjugate gradient methods," *SIAM Journal on Optimization*, vol. 10, no. 2, pp. 345–358, 2000.
- [5] W. W. Hager and H. C. Zhang, "A new conjugate gradient method with guaranteed descent and an efficient line search," *SIAM Journal on Optimization*, vol. 16, no. 1, pp. 170–192, 2005.
- [6] G. Zoutendijk, "Nonlinear programming, computational methods," in *Integer and Nonlinear Programming*, J. Hadji, Ed., pp. 37–86, North-Holland, Amsterdam, The Netherlands, 1970.
- [7] M. J. D. Powell, "Restart procedures for the conjugate gradient method," *Mathematical Programming*, vol. 12, no. 2, pp. 241–254, 1977.
- [8] M. J. D. Powell, "Nonconvex minimization calculations and the conjugate gradient method," in *Numerical analysis*, vol. 1066 of *Lecture Notes in Mathematics*, pp. 122–141, Springer, Berlin, Germany, 1984.
- [9] Y. H. Dai and Y. Yuan, *Nonlinear Conjugate Gradient Method*, Shanghai science and Technology Press, 2000.
- [10] M. Al-Baali, "Descent property and global convergence of the Fletcher-Reeves method with inexact line search," *IMA Journal of Numerical Analysis*, vol. 5, no. 1, pp. 121–124, 1985.
- [11] G. H. Liu, J. Y. Han, and H. X. Yin, "Global convergence of the Fletcher-Reeves algorithm with inexact linesearch," *Applied Mathematics*, vol. 10, no. 1, pp. 75–82, 1995.
- [12] Y. H. Dai and Y. Yuan, "Convergence properties of the Fletcher-Reeves method," *IMA Journal of Numerical Analysis*, vol. 16, no. 2, pp. 155–164, 1996.
- [13] J. C. Gilbert and J. Nocedal, "Global convergence properties of conjugate gradient methods for optimization," *SIAM Journal on Optimization*, vol. 2, no. 1, pp. 21–42, 1992.
- [14] G. Li, C. Tang, and Z. Wei, "New conjugacy condition and related new conjugate gradient methods for unconstrained optimization," *Journal of Computational and Applied Mathematics*, vol. 202, no. 2, pp. 523–539, 2007.
- [15] Z. X. Wei, G. Y. Li, and L. Q. Qi, "Global convergence of the Polak-Ribière-Polyak conjugate gradient method with an Armijo-type inexact line search for nonconvex unconstrained optimization problems," *Mathematics of Computation*, vol. 77, no. 264, pp. 2173–2193, 2008.
- [16] G. Yu, L. Guan, and G. Li, "Global convergence of modified Polak-Ribière-Polyak conjugate gradient methods with sufficient descent property," *Journal of Industrial and Management Optimization*, vol. 4, no. 3, pp. 565–579, 2008.
- [17] L. Grippo and S. Lucidi, "A globally convergent version of the Polak-Ribière conjugate gradient method," *Mathematical Programming*, vol. 78, no. 3, pp. 375–391, 1997.
- [18] Z. Wei, S. Yao, and L. Liu, "The convergence properties of some new conjugate gradient methods," *Applied Mathematics and Computation*, vol. 183, no. 2, pp. 1341–1350, 2006.
- [19] H. Huang, Z. Wei, and S. Yao, "The proof of the sufficient descent condition of the Wei-Yao-Liu conjugate gradient method under the strong Wolfe-Powell line search," *Applied Mathematics and Computation*, vol. 189, no. 2, pp. 1241–1245, 2007.

- [20] Y. Shengwei, Z. Wei, and H. Huang, "A note about WYL's conjugate gradient method and its applications," *Applied Mathematics and Computation*, vol. 191, no. 2, pp. 381–388, 2007.
- [21] H. Huang, S. Yao, and H. Lin, "A new conjugate gradient method based on HS-DY methods," *Journal of Guangxi University of Technology*, vol. 4, pp. 63–66, 2008.
- [22] L. Zhang, "An improved Wei-Yao-Liu nonlinear conjugate gradient method for optimization computation," *Applied Mathematics and Computation*, vol. 215, no. 6, pp. 2269–2274, 2009.
- [23] L. Zhang and S. Jian, "Further studies on the Wei-Yao-Liu nonlinear conjugate gradient method," *Applied Mathematics and Computation*, vol. 219, no. 14, pp. 7616–7621, 2013.
- [24] Z. Dai and F. Wen, "Another improved Wei-Yao-Liu nonlinear conjugate gradient method with sufficient descent property," *Applied Mathematics and Computation*, vol. 218, no. 14, pp. 7421–7430, 2012.
- [25] Y.-H. Dai and L.-Z. Liao, "New conjugacy conditions and related nonlinear conjugate gradient methods," *Applied Mathematics and Optimization*, vol. 43, no. 1, pp. 87–101, 2001.
- [26] J. J. Moré, B. S. Garbow, and K. E. Hillstom, "Testing unconstrained optimization software," *ACM Transactions on Mathematical Software*, vol. 7, no. 1, pp. 17–41, 1981.
- [27] I. Bongartz, A. R. Conn, N. Gould, and P. L. Toint, "CUTE: constrained and unconstrained testing environment," *ACM Transactions on Mathematical Software*, vol. 21, no. 1, pp. 123–160, 1995.
- [28] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles," *Mathematical Programming*, vol. 91, no. 2, pp. 201–213, 2002.
- [29] C. Lemarechal, "A view of line-searches," in *Optimization and Optimal Control*, A. Auslander, W. Oettli, and J. Stoer, Eds., vol. 30 of *Lecture Notes in Control and Information Science*, pp. 59–78, Springer, Berlin, Germany, 1981.
- [30] R. Fletcher, *Practical Methods of Optimization. Vol. 1: Unconstrained Optimization*, John Wiley and Sons, New York, NY, USA, 1989.
- [31] J. Moré and D. J. Thuente, "line search algorithms with guaranteed sufficient decrease," *ACM Transactions on Mathematical Software*, vol. 20, pp. 286–307, 1994.