Hindawi
Mathematical Problems in Engineering
Volume 2018, Article ID 4813030, 15 pages
https://doi.org/10.1155/2018/4813030

Research Article

A Three-Term Conjugate Gradient Algorithm with Quadratic Convergence for Unconstrained Optimization Problems

Gaoyi Wu,[1] Yong Li,[2] and Gonglin Yuan[3]

[1] *College of Business Administration, Guangxi University of Finance and Economics, Nanning, Guangxi 530003, China*
[2] *Department of Mathematics, Baise University, Baise, Guangxi 533000, China*
[3] *College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi 530004, China*

Correspondence should be addressed to Gonglin Yuan; glyuan@gxu.edu.cn

Received 4 December 2017; Accepted 30 May 2018; Published 27 June 2018

Academic Editor: Nibaldo Rodríguez

This paper further studies the WYL conjugate gradient (CG) formula with $\beta_k^{WYL} \geq 0$ and presents a three-term WYL CG algorithm, which has the sufficiently descent property without any conditions. The global convergence and the linear convergence are proved; moreover the n-step quadratic convergence with a restart strategy is established if the initial step length is appropriately chosen. Numerical experiments for large-scale problems including the normal unconstrained optimization problems and **the engineer problems (Benchmark Problems)** show that the new algorithm is competitive with the other similar CG algorithms.

1. Introduction

Consider the following minimization optimizations modelling:

$$\min_{x\in\Re^n} \quad f(x), \tag{1}$$

where $f(x) : \Re^n \to \Re$ is a continuously differentiable function. The CG algorithms for (1) have the following iterative processes:

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \cdots, \tag{2}$$

where x_k is the kth iterate, α_k is the step length, and d_k is the search direction defined by

$$d_k = \begin{cases} -g_k + \beta_k d_{k-1}, & \text{if } k \geq 1 \\ -g_k, & \text{if } k = 0, \end{cases} \tag{3}$$

where $g_k = \nabla f(x_k)$ is the gradient and the parameter β_k is a scalar determining different formulas (see [1–8], etc.). The PRP algorithm [6, 7] is one of the most effective CG algorithms and its convergence can be found (see [7, 9, 10], etc.). Powell [9] suggested that β_k should not be less than zero; then many new CG formulas are proposed (see [11–17], etc.) to ensure the scalar $\beta_k \geq 0$. At present, there are many results obtained in CG algorithms (see [11, 18–26], etc.) and a modified weak Wolfe-Powell line search technique is presented to study open unconstrained optimization (see [27, 28]). If a restart strategy is used, the PRP algorithm is n-step quadratic convergence (see [29–31]). Li and Tian [32] proved that a three-term CG algorithm has quadratic convergence with a restart strategy under some inexact line searches and the suitable assumptions.

Recently, Wei et al. [21] proposed a new CG formula defined by $\beta_k^{WYL} = g_k^T y_{k-1}^{WYL}/\|g_{k-1}\|^2$, where $y_{k-1}^{WYL} = g_k - (\|g_k\|/\|g_{k-1}\|)g_{k-1}$, g_k and g_{k+1} are the gradient of f at x_k and x_{k+1}, respectively, and $\|\cdot\|$ denotes the Euclidean norm of vectors. It is easy to deduce that $\beta_k^{WYL} = g_k^T(g_k - (\|g_k\|/\|g_{k-1}\|)g_{k-1})/\|g_{k-1}\|^2 \geq g_k^T(\|g_k\| - (\|g_k\|/\|g_{k-1}\|)\|g_{k-1}\|)/\|g_{k-1}\|^2 = 0$. The global convergences of the WYL algorithm with exact linear search, the Grippo-Lucidi Armijo line search, and the Wolfe-Powell line search have been established by [21, 33–35]. By restricting the parameter $\varsigma_2 < 1/4$ under the strong Wolfe-Powell linear search, the WYL algorithm can meet the sufficiently descent property.

However the quadratic convergence is still open based on this algorithm. In this paper, we mainly further research the WYL algorithm. On the base of paper [32] and the paper [21], we propose a new WYL three-term CG formula. We show that the new CG algorithm has global convergence for general functions and has the n-step quadratic convergence for uniformly convex functions with r-step restart and standard Armijo line search under appropriate conditions. The numerical results show that the new algorithm performs quite well. The main attributes of this algorithm are listed as follows.

(i) A new WYL three-term CG algorithm is introduced, which has sufficiently descent property automatically.

(ii) The global convergence, linear convergent rate, and the n-step quadratic convergence are established.

(iii) Numerical results show that this algorithm is competitive with the normal algorithm for the given problems.

This paper is arranged as follows. In Section 2, we mainly review the motivation and introduce the modified WYL algorithm. We show that the global convergence and r-step linear convergence of the new algorithm with the standard Armijo line search in Section 3. In Section 4, the n-step quadratic convergence of the given algorithm is proved. In Section 5, some numerical experiments are done.

2. Motivation and Algorithm

In this section, we will give motivations based on the WYL formulas. Consider the WYL search direction

$$d_k = \begin{cases} -g_k + \beta_k^{WYL} d_{k-1}, & \text{if } k \geq 1 \\ -g_k, & \text{if } k = 0, \end{cases} \tag{4}$$

and we all know that d_k is sufficiently descent by restricting the parameter $\varsigma_2 < 1/4$ under the strong Wolfe-Powell linear search [33]. By the definition of d_k in (4), for $k \geq 1$, we get

$$d_k^T g_k = -\|g_k\|^2 + \frac{g_k^T y_{k-1}^{WYL}}{\|g_{k-1}\|^2} d_{k-1}^T g_k. \tag{5}$$

In order to ensure that the sufficiently descent property holds, then the first term of the above equality should be maintained. So the directive idea is to add another term to eliminate the second term of the above equality; at the same time, the conjugacy should be guaranteed. Therefore the new conjugate gradient formula called MWYL algorithm is defined by

$$d_k^{WYL} = \begin{cases} -g_k + \beta_k^{WYL} d_{k-1}^{WYL} - \theta_k^{WYL} y_{k-1}^{WYL}, & \text{if } k \geq 1 \\ -g_k, & \text{if } k = 0, \end{cases} \tag{6}$$

where $\theta_k^{WYL} = g_k^T d_{k-1}^{WYL} / \|g_{k-1}\|^2$ and $\beta_k^{WYL} = g_k^T y_{k-1}^{WYL} / \|g_{k-1}\|^2$. It is easy to see that the above search direction is the normal WYL algorithm if the exact linear search is used. It is not difficult to see from (6) that d_k^{WYL} is a descent direction of f at x_k; namely, we have

$$\left(d_k^{WYL}\right)^T g_k = (-g_0)^T g_0 = -\|g_0\|^2, \quad (k = 0) \tag{7}$$

and

$$\begin{aligned} \left(d_k^{WYL}\right)^T g_k &= \left(-g_k + \beta_k^{WYL} d_{k-1}^{WYL} - \theta_k^{WYL} y_{k-1}^{WYL}\right)^T g_k \\ &= -\|g_k\|^2 + \frac{g_k^T \cdot y_{k-1}^{WYL}}{\|g_{k-1}\|^2} \cdot \left(d_{k-1}^{WYL}\right)^T g_k \\ &\quad - \frac{g_k^T d_{k-1}^{WYL}}{\|g_{k-1}\|^2} \cdot \left(y_{k-1}^{WYL}\right)^T g_k = -\|g_k\|^2, \\ &\qquad (k \geq 1). \end{aligned} \tag{8}$$

Moreover we obtain

$$\left(d_k^{WYL}\right)^T g_k = -\|g_k\|^2, \quad k \geq 0 \tag{9}$$

and

$$\|g_k\| \leq \left\|d_k^{WYL}\right\|. \tag{10}$$

Now we list some linear search techniques that will be used in the following sections.

(i) The exact line search is to find α_k such that the function is minimized along the direction d_k, that is, α_k, satisfying

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} \quad f(x_k + \alpha d_k) \tag{11}$$

(ii) The Armijo line search is to find a step length $\alpha_k = \max\{\rho^i \mid i = 0, 1, 2, \cdots\}$ which satisfies

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \varsigma_1 \alpha_k g_k^T d_k, \tag{12}$$

where $\rho \in (0, 1)$ and $\varsigma_1 \in (0, 1/2)$.

(iii) The Wolfe line search conditions are

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + \varsigma_1 \alpha_k g_k^T d_k \\ g(x_k + \alpha_k d_k)^T d_k &\geq \varsigma_2 g_k^T d_k, \end{aligned} \tag{13}$$

where $0 < \varsigma_1 < 1/2$ and $\varsigma_1 < \varsigma_2 < 1$.

In the following we will give the MWYL algorithm.

Algorithm 1 (a modified WYL three-terms CG algorithm, called MWYL).

Step 0: Choose an initial point $x_0 \in \mathfrak{R}^n$, $\varsigma_1 \in (0, 1/2)$, $\rho \in [0, 1)$, $\epsilon \in [0, 1)$; let $d_0 = -g_0$, $k := 0$.

Step 1: If $\|g_k\| \leq \epsilon$, stop. Otherwise go to the next step.

Step 2: Compute step size α_k by the Armijo line search or Wolfe line search.

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

Step 4: Compute the search direction using (6).

Step 5: Let $k := k + 1$. Go to step 2.

3. Convergence of Algorithm 1

In this part, we will prove the global convergence and the r-linear convergence of the *MWYL* algorithm with the Armijo line search and Wolfe line search. The following assumptions are required.

Assumption i. The level set $\Omega = \{x \mid f(x) \leq f(x_1)\}$ is bounded and, in some neighborhood N of $\overline{co}\Omega$, f is continuously differentiable and bounded below, and its gradient is globally Lipschitz continuous; namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in N, \tag{14}$$

where $\overline{co}\Omega$ is the closed convex hull of Ω.

Now we establish the global convergence of Algorithm 1.

Theorem 2. *Let Assumption i hold and the sequence $\{x_k\}$ be generated by Algorithm 1. Then the relation*

$$\lim_{k\to\infty} \inf \|g_k\| = 0 \tag{15}$$

holds.

Proof. We will prove this theorem by contradiction. Suppose that (15) does not hold, then, for all k, there exists a constant $\varepsilon_1 > 0$ satisfying

$$\|g_k\| \geq \varepsilon_1. \tag{16}$$

Using (9) and (12), if f is bounded from below, it is not difficult to get

$$\sum_{k=0}^{\infty} \alpha_k^2 \left\|d_k^{WYL}\right\|^2 < \infty. \tag{17}$$

In particular, we have

$$\lim_{k\to\infty} \alpha_k \left\|d_k^{WYL}\right\| = 0. \tag{18}$$

If $\lim_{k\to\infty}\alpha_k > 0$, by (9) and (18), we obtain $\lim_{k\to\infty}\inf\|g_k\| = 0$. This contradicts (16); then (15) holds.

Otherwise if $\lim_{k\to\infty}\alpha_k = 0$. Then there exists an infinite index set N satisfying

$$\lim_{k\in N, k\to\infty} \alpha_k = 0. \tag{19}$$

By Step 2 of Algorithm 1, when k is sufficiently large, $\alpha_k\rho^{-1}$ does not satisfy (12), which implies that

$$f\left(x_k + \alpha_k\rho^{-1}d_k^{WYL}\right) - f(x_k) > -\varsigma_1\rho^{-2}\alpha_k^2\left\|d_k^{WYL}\right\|^2. \tag{20}$$

By (16), similar to the proof of Lemma 2.1 in [36], it is easy to deduce that there exists a constant $\varrho > 0$ such that

$$\left\|d_k^{WYL}\right\| \leq \varrho, \quad \forall k. \tag{21}$$

Using (21), (9), and the mean-value theorem, we have

$$\begin{aligned} f&\left(x_k + \alpha_k\rho^{-1}d_k^{WYL}\right) - f(x_k) \\ &= \rho^{-1}\alpha_k g\left(x_k + \xi_0\rho^{-1}\alpha_k d_k^{WYL}\right)^T d_k^{WYL} \\ &= \rho^{-1}\alpha_k g_k^T d_k^{WYL} \\ &\quad + \rho^{-1}\alpha_k\left(g\left(x_k + \xi_0\rho^{-1}\alpha_k d_k^{WYL}\right) - g_k\right)^T d_k^{WYL} \\ &\leq \rho^{-1}\alpha_k g_k^T d_k^{WYL} + M\rho^{-2}\alpha_k^2\left\|d_k^{WYL}\right\|^2, \end{aligned} \tag{22}$$

where $\xi_0 \in (0,1)$ and the last inequality follows (26). Combining with (20), for all $k \in$ N sufficiently large, we obtain

$$\|g_k\|^2 \leq \rho^{-1}(M + \varsigma_1)\alpha_k\left\|d_k^{WYL}\right\|^2. \tag{23}$$

By (21) and $\lim_{k\to\infty}\alpha_k = 0$, then the above inequality implies that $\lim_{k\in N,k\to\infty}\|g_k\| = 0$. This is a contradiction too. The proof is complete. □

In the next, we will prove the linear convergence of the sequence $\{x_k\}$ by the MWYL algorithm with the Armijo or Wolfe line search. The following assumption is further needed.

Assumption ii. f is twice continuously differentiable and the uniformly convex function. In other words, there are positive constants $M \geq m > 0$ such that

$$m\|d\|^2 \leq d^T\nabla^2 f(x)\, d \leq M\|d\|^2, \quad \forall x, d \in \Re^n, \tag{24}$$

where $\nabla^2 f(x)$ denotes the Hessian matrix of f at x.

It is not difficult to see that, under the Assumption ii, $\nabla^2 f(x)$ is continuous and g is Lipschitz continuous and problem (1) has a unique solution x^* which satisfies

$$\frac{1}{2}m\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{2}M\|x - x^*\|^2, \quad \forall x \in \Re^n \tag{25}$$

and

$$m\|x - x^*\| \leq \|g_k\| \leq M\|x_k - x^*\|. \tag{26}$$

Lemma 3. *Let Assumption ii hold and the sequence $\{x_k\}$ be generated by the MWYL algorithm with the Armijo or Wolfe line search, one has*

$$\begin{gathered} -\sum_{k=0}^{\infty} g_k^T s_k^{WYL} < \infty, \\ \sum_{k=0}^{\infty} \left\|s_k^{WYL}\right\| < \infty, \\ c_1\alpha_k\left\|d_k^{WYL}\right\|^2 \leq -g_k^T d_k^{WYL}, \end{gathered} \tag{27}$$

where $s_k^{WYL} = x_{k+1} - x_k$ and $c_1 = (1/2)(1-\sigma_1)^{-1}m$. In addition, if the Wolfe line search is used, the following holds:

$$-g_k^T d_k^{WYL} \leq (1 - \varsigma_2)^{-1} M\alpha_k\left\|d_k^{WYL}\right\|^2. \tag{28}$$

Proof. We have from line search (12) that

$$f(x_0) - f(\widehat{x}_k) = \sum_{k=0}^{\widehat{k}-1} (f(x_k) - f(x_{k+1})) \geq -\sum_{k=0}^{\widehat{k}-1} \sigma_1 g_k^T s_k^{WYL} > 0 \tag{29}$$

hold for any $\widehat{k} > 1$, because the objective function f is uniformly convex of Assumption ii, f is bounded below, so the inequality $-\sum_{k=0}^{\widehat{k}-1} g_k^T s_k^{WYL} < +\infty$ holds. Combining the Taylor theorem and Assumption ii, we obtain

$$f(x_{k+1}) = f(x_k) + g_k^T s_k^{WYL} + \frac{1}{2} \left(s_k^{WYL}\right)^T \nabla^2 f(\xi_k) s_k^{WYL}, \tag{30}$$

where ξ_k belong to the segment $[x_k, x_{k+1}]$. Therefore, we get

$$\sum_{k=0}^{\widehat{k}-1} (f(x_{k+1}) - f(x_k)) - \sum_{k=0}^{\widehat{k}-1} g_k^T s_k^{WYL} = \sum_{k=0}^{\widehat{k}-1} \frac{1}{2} \left(s_k^{WYL}\right)^T \nabla^2 f(\xi_k) s_k^{WYL} \geq \sum_{k=0}^{\widehat{k}-1} \frac{1}{2} m \left\| s_k^{WYL} \right\|^2. \tag{31}$$

By the inequalities $\sum_{k=0}^{\widehat{k}-1} (f(x_{k+1}) - f(x_k)) < 0$ and $-\sum_{k=0}^{\widehat{k}-1} g_k^T s_k^{WYL} < +\infty$, we get $\sum_{k=0}^{\infty} \| s_k^{WYL} \|^2 < +\infty$. Using (12), (30), and Assumption ii, we obtain

$$\begin{aligned} \varsigma_1 g_k^T s_k^{WYL} &\geq f(x_{k+1}) - f(x_k) \\ &= g_k^T s_k^{WYL} + \frac{1}{2} \left(s_k^{WYL}\right)^T \nabla^2 f(\xi_k) s_k^{WYL} \\ &\geq g_k^T s_k^{WYL} + \frac{1}{2} m \left\| s_k^{WYL} \right\|^2, \end{aligned} \tag{32}$$

which includes $-g_k^T s_k^{WYL} \geq (m/2(1-\varsigma_1)) \| s_k^{WYL} \|^2$.

It is not difficult to get that

$$-g_k^T d_k^{WYL} \geq \frac{m}{2(1-\varsigma_1)} \cdot \alpha_k \left\| d_k^{WYL} \right\|^2 = c_1 \alpha_k \left\| d_k^{WYL} \right\|^2. \tag{33}$$

By the second inequality of (13), we get

$$g\left(x_k + \alpha_k d_k^{WYL}\right) = \left[g_k + \int_0^1 \nabla^2 f\left(x_k + \tau \alpha_k d_k^{WYL}\right) d\tau \cdot \alpha_k d_k^{WYL} \right]^T \cdot d_k^{WYL} \geq \varsigma_2 g_k^T d_k^{WYL} \tag{34}$$

and

$$\alpha_k \left(d_k^{WYL}\right)^T \int_0^1 \nabla^2 f\left(x_k + \tau \alpha_k d_k^{WYL}\right) d\tau \cdot d_k^{WYL} \geq (\varsigma_2 - 1) g_k^T d_k^{WYL}. \tag{35}$$

By Assumption ii, we obtain $-g_k^T d_k^{WYL} \leq (1-\varsigma_2)^{-1} M \alpha_k \| d_k^{WYL} \|^2$. This completes the proof. □

Lemma 4. *Let the sequence $\{x_k\}$ be generated by the MWYL algorithm with the Armijo or Wolfe line search and Assumption ii hold; then there is a constant $c > 0$ such that*

$$\alpha_k > c, \quad \forall k \geq 0. \tag{36}$$

Proof. Set

$$G_{k-1} = \int_0^1 \nabla^2 f(x_{k-1} + \tau s_{k-1}) d\tau, \tag{37}$$

where $s_{k-1}^{WYL} = x_k - x_{k-1} = \alpha_{k-1} d_{k-1}^{WYL}$. By the mean-value theorem, we have

$$y_k = g_k - g_{k-1} = G_{k-1} s_{k-1} = \alpha_{k-1} G_{k-1} d_{k-1}^{WYL} \tag{38}$$

and

$$\begin{aligned} \left| y_{k-1}^{WYL} \right| &= \left\| g_k - \frac{\|g_k\|}{\|g_{k-1}\|} \cdot g_{k-1} \right\| \\ &= \left\| g_k - g_{k-1} + g_{k-1} - \frac{g_k}{g_{k-1}} \cdot g_{k-1} \right\| \\ &\leq \| g_k - g_{k-1} \| + \left| \|g_{k-1}\| - \|g_k\| \right| \\ &\leq 2 \| g_k - g_{k-1} \| \leq 2\alpha_{k-1} \cdot \|G_{k-1}\| \cdot \left\| d_{k-1}^{WYL} \right\| \end{aligned} \tag{39}$$

Therefore, by (9), (39), Lemma 3, and the Assumption ii, we get

$$\begin{aligned} \left| \beta_k^{WYL} \right| &= \frac{\|g_k\| \cdot \left\| y_{k-1}^{WYL} \right\|}{\|g_{k-1}\|^2} \leq \frac{2\alpha_k \|g_k\| \cdot \|G_{k-1}\| \left\| d_{k-1}^{WYL} \right\|}{c_1 \alpha_k \left\| d_{k-1}^{WYL} \right\|^2} \\ &\leq \frac{2c_1^{-1} M \|g_k\|}{\left\| d_{k-1}^{WYL} \right\|} \end{aligned} \tag{40}$$

and

$$\left| \theta_k^{WYL} \right| = \frac{\left| g_k^T \cdot d_{k-1}^{WYL} \right|}{\|g_{k-1}\|^2} = \frac{\left| g_k^T \cdot d_{k-1}^{WYL} \right|}{-g_{k-1}^T \cdot d_{k-1}^{WYL}} \leq \frac{2c_1^{-1} M \|g_k\|}{\left\| d_{k-1}^{WYL} \right\|} \tag{41}$$

By the above conclusion, (6), and the Lipschitz continuity of g, we get

$$\begin{aligned} \left\| d_k^{WYL} \right\| &\leq \|g_k\| + \left| \beta_k^{WYL} \right| \cdot \left\| d_{k-1}^{WYL} \right\| + \left| \theta_k^{WYL} \right| \cdot \left\| y_{k-1}^{WYL} \right\| \\ &\leq \|g_k\| + 2Mc_1^{-1} \|g_k\| + 2Mc_1^{-1} \|g_k\| \\ &\leq \left(1 + 4Mc_1^{-1}\right) \cdot \|g_k\| \end{aligned} \tag{42}$$

If the Armijo line search is used, using the line search rule, if $\alpha_k \neq 1$, then $\alpha_k' = \alpha_k \rho^{-1}$ will not satisfy line search condition (12). Namely,

$$f\left(x_k + \alpha_k' d_k^{WYL}\right) - f(x_k) > \varsigma_1 \alpha_k' g_k^T d_k^{WYL}. \tag{43}$$

Using the mean-value theorem and the above inequality, there exists $\mu_k \in (0,1)$ satisfying

$$\begin{aligned}\varsigma_1 \alpha_k' g_k^T d_k^{WYL} &< f\left(x_k + \alpha_k' d_k^{WYL}\right) - f\left(x_k\right) \\ &= \alpha_k' g\left(x_k + \mu_k \alpha_k' d_k^{WYL}\right)^T d_k^{WYL} \\ &= \alpha_k' \left(g\left(x_k + \mu_k \alpha_k' d_k^{WYL}\right) - g\left(x_k\right)\right)^T d_k^{WYL} \\ &\quad + \alpha_k' g_k^T d_k^{WYL} \\ &= \mu_k \left(\alpha_k'\right)^2 \left(d_k^{WYL}\right)^T \int_0^1 \nabla^2 f\left(x_k + \tau \mu_k \alpha_k' d_k^*\right) d\tau \\ &\quad \cdot d_k^{WYL} + \alpha_k' g_k^T d_k^{WYL} \\ &\leq \left(\alpha_k'\right)^2 M \cdot \left\|d_k^{WYL}\right\|^2 + \alpha_k' g_k^T d_k^{WYL}.\end{aligned} \tag{44}$$

Thus, by the above conclusion and (42), we get

$$\begin{aligned}\alpha_k = \rho \alpha_k' &\geq -\frac{\left(1-\varsigma_1\right)\rho}{M} \cdot \frac{g_k^T d_k^{WYL}}{\left\|d_k^{WYL}\right\|^2} \\ &= \frac{\left(1-\varsigma_1\right)\rho}{M} \cdot \frac{\left\|g_k\right\|^2}{\left\|d_k^{WYL}\right\|^2} \\ &\geq M^{-1}\left(1-\varsigma_1\right)\rho\left(1+4Mc_1^{-1}\right)^{-2} \equiv c^*,\end{aligned} \tag{45}$$

and letting $c = \min\{1, c^*\}$, we have (36). If the Wolfe line search is used, from the second inequality of (13), we obtain

$$\begin{aligned}M\alpha_k \left\|d_k^{WYL}\right\|^2 &\geq \left(g\left(x_k + \alpha_k d_k^{WYL}\right) - g_k\right)^T d_k^{WYL} \\ &\geq -\left(1-\varsigma_2\right) g_k^T d_k^{WYL}.\end{aligned} \tag{46}$$

By similar way to that for the Armijo line search, we can find a lower positive bound of α_k; the proof is completed. □

Similar to [32], It is easy to get the r-linear convergence theorem of Algorithm 1. So we only state it as follows but omit the proof.

Theorem 5. *Let Assumption ii hold, x^* be the unique solution of (1), and the sequence $\{x_k\}$ be generated by the MWYL algorithm with the Armijo or Wolfe line search. Then there are constants $a > 0$ and $r \in (0,1)$ satisfying*

$$\left\|x_k - x^*\right\| \leq a r^k. \tag{47}$$

Proof. By (12) or the first relation of (13), we get

$$\begin{aligned}f\left(x_{k+1}\right) - f\left(x^*\right) &\leq f\left(x_k\right) - f\left(x^*\right) + \varsigma_1 \alpha_k d_k^T g_k \\ &= f\left(x_k\right) - f\left(x^*\right) - \varsigma_1 \alpha_k \left\|g_k\right\|^2 \\ &\leq f\left(x_k\right) - f\left(x^*\right) - c\varsigma_1 m^2 \alpha_k \left\|x_k - x^*\right\|^2 \\ &\leq f\left(x_k\right) - f\left(x^*\right) - \frac{2c\varsigma_1 m^2}{M}\left[f\left(x_k\right) - f\left(x^*\right)\right] \\ &= \left[1 - \frac{2c\varsigma_1 m^2}{M}\right]\left[f\left(x_k\right) - f\left(x^*\right)\right],\end{aligned} \tag{48}$$

where the first equality follows (9), the second inequality follows (26), and the last inequality (25). Setting $r = [1 - 2c\varsigma_1 m^2/M]^{1/2} \in (0,1)$ generates

$$\begin{aligned}f\left(x_{k+1}\right) - f\left(x^*\right) &\leq r^2\left[f\left(x_k\right) - f\left(x^*\right)\right] \leq \cdots \\ &\leq r^{2k}\left[f\left(x_0\right) - f\left(x^*\right)\right].\end{aligned} \tag{49}$$

By (25) again, we have

$$\begin{aligned}\left\|x_k - x^*\right\| &\leq \frac{2}{m}\left[f\left(x_k\right) - f\left(x^*\right)\right] \\ &\leq \frac{2}{m}\left[f\left(x_0\right) - f\left(x^*\right)\right] r^{2k},\end{aligned} \tag{50}$$

and this relation shows that the proof is complete. □

4. The Restart MWYL Algorithm's N-Step Quadratic Convergence

Setting $\overline{\alpha_k}$ as exact line search step length, then

$$g\left(x_k + \overline{\alpha_k} d_k^{WYL}\right)^T d_k^{WYL} = 0 \tag{51}$$

holds. Thus,

$$\begin{aligned}\left\|g_k\right\|^2 &= -g_k^T d_k^{WYL} \\ &= \left(g\left(x_k + \overline{\alpha_k} d_k^{WYL}\right) - g\left(x_k\right)\right)^T d_k^{WYL} \\ &= \overline{\alpha_k}\left(d_k^{WYL}\right)^T \overline{G_k} d_k^{WYL},\end{aligned} \tag{52}$$

where $\overline{G_k} = \int_0^1 \nabla^2 f(x_k + \tau \overline{\alpha_k} d_k^{WYL}) d\tau$. It is feasible to use the initial step length of the Armijo or Wolfe line search as an approximation of $\overline{\alpha_k}$, where $\overline{\alpha_k}$ is defined by

$$\begin{aligned}\overline{\alpha_k} &\equiv \frac{\left\|g_k\right\|^2}{\left(d_k^{WYL}\right)^T \overline{G_k} d_k^{WYL}} \\ &\approx \frac{\epsilon_k \left\|g_k\right\|^2}{\left(d_k^{WYL}\right)^T\left(g\left(x_k + \epsilon_k d_k^{WYL}\right) - g\left(x_k\right)\right)} \equiv \gamma_k,\end{aligned} \tag{53}$$

where the integer sequence $\{\epsilon_k\} \to 0$ as $k \to \infty$. If f is a quadratic function, then γ_k and $\overline{\alpha_k}$ are consistent; namely,

$$\begin{aligned}|\overline{\alpha_k}-\gamma_k| &= \left|\frac{\|g_k\|^2}{(d_k^{WYL})^T\overline{G_k}d_k^{WYL}}-\frac{\epsilon_k\|g_k\|^2}{(d_k^{WYL})^T(g(x_k+\epsilon_k d_k^{WYL})-g(x_k))}\right| \\ &= \frac{\left|(d_k^{WYL})^T(g(x_k+\epsilon_k d_k^{WYL})-g_k)\|g_k\|^2-(d_k^{WYL})^T\overline{G}_k d_k^{WYL}\epsilon_k\|g_k\|^2\right|}{\left((d_k^{WYL})^T\overline{G}_k d_k^{WYL}\right)(d_k^{WYL})^T(g(x_k+\epsilon_k d_k^{WYL})-g(x_k))} \le \frac{o\left(\epsilon_k\|d_k^{WYL}\|^2\right)\cdot\|g_k\|^2}{m^2\epsilon_k\|d_k^{WYL}\|^4}\longrightarrow 0.\end{aligned} \tag{54}$$

The above discussions can also be found in [32]; in fact, our ideas are motivated by this paper partly. The following Theorem 6 will show that, for sufficiently large k, the inexact line search step γ_k which is defined by (53) satisfies the Armijo and Wolfe conditions.

Theorem 6. *Let sequence $\{x_k\}$ be generated by the MWYL algorithm and Assumption ii hold. Then, when k is sufficiently large, γ_k satisfies the Armijo and Wolfe conditions.*

Proof. Let $A_k=\int_0^1\nabla^2 f(x_k+\tau\epsilon_k d_k^{WYL})d\tau$, using Assumption ii and (10), we have

$$\gamma_k=\frac{\|g_k\|^2}{(d_k^{WYL})^T A_k d_k^{WYL}}\ge\frac{\|g_k\|^2}{M\|d_k^{WYL}\|^2} \tag{55}$$

and

$$\gamma_k=\frac{\|g_k\|^2}{(d_k^{WYL})^T A_k d_k^{WYL}}\le\frac{\|g_k\|^2}{m\|d_k^{WYL}\|^2}\le\frac{1}{m}. \tag{56}$$

Using $\{d_k^{WYL}\}\to 0$, Assumption ii, (47), and (55), we get

$$\begin{aligned}&f\left(x_k+\gamma_k d_k^{WYL}\right)\\&\quad=f(x_k)+\gamma_k g_k^T d_k^{WYL}+\frac{1}{2}\gamma_k^2\left(d_k^{WYL}\right)^T A_k d_k^{WYL}\\&\qquad+\gamma_k^2 o\left(\|d_k^{WYL}\|^2\right)\\&\quad=f(x_k)+\frac{1}{2}\gamma_k g_k^T d_k^{WYL}+\gamma_k^2 o\left(\|d_k^{WYL}\|^2\right)\\&\quad=f(x_k)+\varsigma_1\gamma_k g_k^T d_k^{WYL}-\left(\frac{1}{2}-\varsigma_1\right)\gamma_k\|g_k\|^2\\&\qquad+\gamma_k^2 o\left(\|d_k^{WYL}\|^2\right)\\&\quad\le f(x_k)+\varsigma_1\gamma_k g_k^T d_k^{WYL}\\&\qquad-\left(\frac{1}{2}-\varsigma_1\right)\frac{\left(1+2Mc_1^{-1}\right)^{-4}}{M}\|d_k^{WYL}\|^2\\&\qquad+\gamma_k^2 o\left(\|d_k^{WYL}\|^2\right)\end{aligned} \tag{57}$$

For k is sufficiently large, we have

$$f\left(x_k+\gamma_k d_k^{WYL}\right)\le f(x_k)+\varsigma_1\gamma_k g_k^T d_k^{WYL}. \tag{58}$$

When k is sufficiently large, $\alpha_k=\gamma_k$ satisfies the Armijo condition. Setting $\overline{A_k}=\int_0^1\nabla^2 f(x_k+\tau\gamma_k d_k^{WYL})d\tau$, we get

$$\begin{aligned}&g(x_k+\gamma_k d_k)^T d_k^{WYL}-\varsigma_2 g_k^T d_k^{WYL}\\&\quad=\left(g\left(x_k+\gamma_k d_k^{WYL}\right)-g_k\right)^T d_k^{WYL}\\&\qquad+(1-\varsigma_2)\,g_k^T d_k^{WYL}\\&\quad=\gamma_k\left(d_k^{WYL}\right)^T\overline{A_k}d_k^{WYL}-(1-\varsigma_2)\|g_k\|^2\\&\quad=\left(\frac{\left(d_k^{WYL}\right)^T\overline{A_k}d_k^{WYL}}{\left(d_k^{WYL}\right)^T A_k d_k^{WYL}}-1\right)\|g_k\|^2+\varsigma_2\|g_k\|^2\\&\quad\le f(x_k)+\varsigma_1\gamma_k g_k^T d_k^{WYL}\\&\qquad-\left(\frac{1}{2}-\varsigma_1\right)\frac{\left(1+2Mc_1^{-1}\right)^{-4}}{M}\|d_k^{WYL}\|^2\\&\qquad+\gamma_k^2 o\left(\|d_k^{WYL}\|^2\right)\\&\quad=\left(\frac{\left(d_k^{WYL}\right)^T\left(\overline{A_k}-A_k\right)d_k^{WYL}}{\left(d_k^{WYL}\right)^T A_k d_k^{WYL}}\right)\|g_k\|^2+\varsigma_2\|g_k\|^2\\&\quad=\varsigma_2\|g_k\|^2+o\left(\|g_k\|^2\right).\end{aligned} \tag{59}$$

So, for sufficiently large k, we have

$$g\left(x_k+\gamma_k d_k^{WYL}\right)^T d_k^{WYL}\ge\varsigma_2 g_k^T d_k^{WYL}. \tag{60}$$

This implies that $\alpha_k=\gamma_k$ satisfy the Wolfe line search. The proof is complete. □

If we use the restart *MWYL* algorithm, the n-step quadratic convergence is desirable. In the next, we use the $|\gamma_k|$ as the initial step-length and give the algorithm steps of the restart MWYL algorithm.

Algorithm 7 (called RWYL).

Step 0: Given $x_0\in\mathfrak{R}^n$, $r>0$, $\varsigma_1\in(0,1/2)$, $\rho\in[0,1)$, $\epsilon\in[0,1)$, let $k:=0$.

Step 1: If $\|g_k\|\le\epsilon$, stop.

Step 2: If the inequality $f(x_k+|\gamma_k|d_k^{WYL})\le f(x_k)+\varsigma_1|\gamma_k|g_k^T d_k^{WYL}$ holds, we set $\alpha_k=|\gamma_k|$. Otherwise, we determine $\alpha_k = \max\{|\gamma_k|\rho^j \mid j = 0, 1, 2, \cdots\}$ satisfying

$$f\left(x_k + \alpha_k d_k^{WYL}\right) \leq f\left(x_k\right) + \varsigma_1 \alpha_k g_k^T d_k^{WYL}. \tag{61}$$

Step 3: Let $x_{k+1} = x_k + \alpha_k d_k^{WYL}$, and $k := k + 1$.

Step 4: If $\|g_k\| \leq \epsilon$, stop.

Step 5: If $k = r$, we let $x_0 := x_k$. Go to step 1.

Step 6: Compute d_k^{WYL} by (6). Go to step 2.

Lemma 8. *Let Assumption ii hold and $\{x_k\}$ be generated by the RWYL algorithm. Then there exist positive numbers c_i^*, $i = 1, 2, 3, 4$, such that*

$$\begin{aligned} \|g_{k+1}\| &\leq c_1^* \left\|d_k^{WYL}\right\|, \\ \left|\beta_{k+1}^{WYL}\right| &\leq c_2^*, \\ \left|\theta_{k+1}^{WYL}\right| &\leq c_3^*, \\ \left\|d_{k+1}^{WYL}\right\| &\leq c_4^* \left\|d_k^{WYL}\right\|. \end{aligned} \tag{62}$$

Proof. Considering the first inequality of (62), we get

$$\begin{aligned} \|g_{k+1}\| &= \|g_k + (g_{k+1} - g_k)\| \\ &\leq \|g_k\| + |\gamma_k| \cdot \left\|\widehat{A_k} d_k^{WYL}\right\| \\ &\leq \left\|d_k^{WYL}\right\| + \frac{M}{m} \cdot \left\|d_k^{WYL}\right\| = \left(1 + \frac{M}{m}\right) \left\|d_k^{WYL}\right\| \\ &\equiv c_1^* \left\|d_k^{WYL}\right\|, \end{aligned} \tag{63}$$

where $\widehat{A_k} = \int_0^1 \nabla^2 f(x_k + \tau|\gamma_k| d_k^{WYL}) d\tau$. By the definition of β_{k+1}^{WYL} we discuss the other three inequalities of (62), respectively. Starting from β_{k+1}^{WYL}, by the (39) and (62), we get

$$\begin{aligned} \left|\beta_{k+1}^{WYL}\right| &= \frac{\left\|g_{k+1}^T y_k^{WYL}\right\|}{\|g_k\|^2} \leq \frac{\left\|g_{k+1}^T\right\| \cdot \left\|y_k^{WYL}\right\|}{\|g_k\|^2} \\ &\leq \frac{2c_1^{-1} M \|g_{k+1}\|}{\left\|d_k^{WYL}\right\|} \leq 2c_1^{-1} M c_1^* \equiv c_2^*. \end{aligned} \tag{64}$$

By (40), (62), and the definition of θ_{k+1}^{WYL}, we obtain

$$\begin{aligned} \left|\theta_{k+1}^{WYL}\right| &= \left|\frac{g_{k+1}^T d_k^{WYL}}{\|g_k\|^2}\right| \leq \|g_{k+1}\| \cdot \frac{\left\|d_k^{WYL}\right\|}{\|g_k\|^2} \\ &\leq \frac{c_1^* \left\|d_k^{WYL}\right\|^2}{\|g_k\|^2} \leq c_1^* \left(1 + 4Mc_1^{-1}\right)^2 \equiv c_3^*. \end{aligned} \tag{65}$$

By the above conclusion and the definition of d_{k+1}^{WYL}, we have

$$\begin{aligned} \left\|d_{k+1}^{WYL}\right\| &= \left\|-g_{k+1} + \beta_{k+1}^{WYL} d_k^{WYL} - \theta_{k+1}^{WYL} y_k^{WYL}\right\| \\ &\leq \|g_{k+1}\| + \left|\beta_{k+1}^{WYL}\right| \cdot \left\|d_k^{WYL}\right\| + \left|\theta_{k+1}^{WYL}\right| \\ &\quad \cdot \left\|y_k^{WYL}\right\| \\ &\leq c_1^* \left\|d_k^{WYL}\right\| + c_2^* \left\|d_k^{WYL}\right\| + 2c_3^* M\alpha_k \left\|d_k^{WYL}\right\| \\ &\leq \left(c_1^* + c_2^* + 2Mc_3^*\right) \left\|d_k^{WYL}\right\| \equiv c_4^* \left\|d_k^{WYL}\right\|. \end{aligned} \tag{66}$$

The proof is complete. □

In the following, we will prove the n-order quadratic convergence of the RWYL algorithm. We always let Assumption ii hold and $\{x_k\}$ be generated by the RWYL algorithm. Using x^* as the unique solution of problem (1), by Theorem 2, we have $\{x_k\} \to x^*$. The equation $\alpha_k = \gamma_k$ always holds if only k is large enough by the Theorem 6. In order to establish this convergence of the RWYL algorithm, we further need the following assumption.

Assumption iii. In some neighborhood N of x^*, $\nabla^2 f$ is Lipschitz continuous.

Based on Assumption iii and the above lemma, we have the following remarks. Let $\widehat{f_{kr}}(x)$ be the second-order approximate function of f in the neighborhood of the initial point x_{kr}, then we have

$$\begin{aligned} \widehat{f_{kr}}(x) &= f\left(x_{kr}\right) + g\left(x_{kr}\right)^T \left(x - x_{kr}\right) \\ &\quad + \frac{1}{2}\left(x - x_{kr}\right)^T \nabla^2 f\left(x_{kr}\right)\left(x - x_{kr}\right). \end{aligned} \tag{67}$$

Let $\{x_{kr}^i\}$ and $\{d_{kr}^{WYL(i)}\}$ be the iterations and directions generated by the RWYL algorithm to minimize the quadratic function $\widehat{f_{kr}}(x)$ with initial point $x_{kr}^0 = x_{kr}$. Specifically, the sequence $\{x_{kr}^i\}$ is generated by using the following process:

$$\begin{aligned} x_{kr}^0 &= x_{kr}, \\ x_{kr}^{i+1} &= x_{kr}^0 + \alpha_{kr}^i d_{kr}^{WYL(i)}, \\ &\qquad i = 0, 1, \cdots, \end{aligned} \tag{68}$$

and

$$d_{kr}^{WYL(i)} = \begin{cases} -g_{kr}^i + \beta_{kr}^{WYL(i)} d_{kr}^{WYL(i-1)} - \theta_{kr}^i y_{kr}^{WYL(i-1)}, & \text{if } k \geq 1 \\ -g_{kr}^0, & \text{if } k = 0, \end{cases} \tag{69}$$

where $g_{kr}^i = g(x_{kr}^i)$ for $i = 1, 2, \cdots$:

$$\begin{aligned} \beta_{kr}^{WYL(i)} &= \frac{\left(g_{kr}^i\right)^T y_{kr}^{WYL(i-1)}}{\left\|g_{kr}^{i-1}\right\|^2}, \\ y_{kr}^{WYL(i-1)} &= g_{kr}^i - g_{kr}^{i-1}, \end{aligned}$$

$$\theta_{kr}^{WYL(i)} = \frac{\left(g_{kr}^{i}\right)^{T} d_{kr}^{WYL(i-1)}}{\left\|g_{kr}^{i-1}\right\|^{2}} \tag{70}$$

From the proof process of Theorem 6, it is not difficult to see that when k is sufficiently large, step length γ_k can always be found. Because $\widehat{f_{kr}}(x)$ is a quadratic function, γ_k is the same as the step length obtained by the exact line search. Consequently, we have $\theta_{kr}^{WYL(i)} = 0$; moreover there is an index $j(kr) \leq n$ such that $x_{kr}^{j(kr)}$ is the exact minimizer of $\widehat{f_{kr}}$.

Similar to Lemmas (A.8)-(A-10) in the paper [30], it is not difficult to get the following relations:

$$\begin{aligned} &\left\|\beta_{kr+i}^{WYL} d_{kr+i-1}^{WYL} - \beta_{kr}^{WYL(i)} d_{kr}^{WYL(i-1)}\right\| \\ &\quad = O\left(\left\|d_{kr+i-1}^{WYL} - d_{kr}^{WYL(i-1)}\right\|\right) + O\left(\left\|d_{kr}^{WYL}\right\|^{2}\right), \end{aligned} \tag{71}$$

$$\begin{aligned} &\left\|g_{kr+i} - g_{kr}^{i}\right\| \\ &\quad \leq \left\|g_{kr+i-1} - g_{kr}^{i-1}\right\| + O\left(\left\|d_{kr}^{WYL}\right\|^{2}\right) \\ &\qquad + M\left\|\alpha_{kr+i-1} d_{kr+i-1}^{WYL} - \alpha_{kr}^{WYL(i-1)} d_{kr}^{WYL(i-1)}\right\|, \end{aligned} \tag{72}$$

and

$$\begin{aligned} &\left\|\alpha_{kr+i} d_{kr+i}^{WYL} - \alpha_{kr}^{i} d_{kr}^{WYL(i)}\right\| \\ &\quad = O\left(\left\|g_{kr+i} - g_{kr}^{i}\right\|\right) + O\left(\left\|d_{kr+i}^{WYL} - d_{kr}^{WYL(i)}\right\|\right) \\ &\qquad + O\left(\left\|d_{kr}^{WYL}\right\|^{2}\right). \end{aligned} \tag{73}$$

The following lemma shows that the parameter θ_{k}^{WYL} will converge to 0.

Lemma 9. *For the parameter θ_{k}^{WYL}, one has*

$$\left|\theta_{kr+i+1}^{WYL}\right| = O\left(\left\|g_{kr+i}\right\|\right) = O\left(\left\|g_{kr}\right\|\right). \tag{74}$$

Proof. Let $\widehat{A}_{kr+i}$ and A_{kr+i} be defined by Lemma 8; we get

$$\begin{aligned} &\left|\frac{\left(d_{kr+i}^{WYL}\right)^{T} \widehat{A}_{kr+i} d_{kr+i}^{WYL}}{\left(d_{kr+i}^{WYL}\right)^{T} A_{kr+i} d_{kr+i}^{WYL}} - 1\right| \\ &\quad \leq \frac{\left|\left(d_{kr+i}^{WYL}\right)^{T} \widehat{A}_{kr+i} d_{kr+i}^{WYL} - \left(d_{kr+i}^{WYL}\right)^{T} A_{kr+i} d_{kr+i}^{WYL}\right|}{m\left\|d_{kr+i}^{WYL}\right\|^{2}} \\ &\quad \leq \frac{1}{m\left\|d_{kr+i}^{WYL}\right\|^{2}}\left\{\left\|\int_{0}^{1} \nabla^{2} f\left(x_{kr+i} + \tau\alpha_{kr+i} d_{kr+i}^{WYL}\right) d\tau\right.\right. \\ &\quad \left. - \int_{0}^{1} \nabla^{2} f\left(x_{kr+i} + \epsilon_{kr+i}\tau\alpha_{kr+i} d_{kr+i}^{WYL}\right) d\tau\right\| \\ &\quad \left.\cdot\left\|d_{kr+i}^{WYL}\right\|^{2}\right\} \leq \frac{L}{m}\int_{0}^{1}\left(1-\epsilon_{kr+i}\right)\tau d\tau \cdot \left\|\alpha_{kr+i}\right| \\ &\quad \cdot d_{kr+i}^{WYL}\| \leq \frac{L}{2m}\left(1-\epsilon_{kr+i}\right)\cdot\left\|d_{kr+i}^{WYL}\right\| \\ &\quad \leq \frac{L\left(1-\epsilon_{kr+i}\right)\left(1+4Mc_{1}^{-1}\right)}{2m}\left\|g_{kr+i}\right\| = O\left(\left\|g_{kr+i}\right\|\right) \end{aligned} \tag{75}$$

where L is Lipschitz constant for $\nabla^2 f$ on set N. Then we get

$$\left|\frac{\left(d_{kr+i}^{WYL}\right)^{T} \widehat{A}_{kr+i} d_{kr+i}^{WYL}}{\left(d_{kr+i}^{WYL}\right)^{T} A_{kr+i} d_{kr+i}^{WYL}} - 1\right| = O\left(\left\|g_{kr+i}\right\|\right). \tag{76}$$

For k sufficiently large, we get $\alpha_k = \gamma_k$. By the mean value theorem, we have

$$\begin{aligned} &g\left(x_{kr+i}\right)^{T} d_{kr+i}^{WYL} = g\left(x_{kr+i} + \gamma_{kr+i} d_{kr+i}^{WYL}\right)^{T} d_{kr+i}^{WYL} \\ &\quad = g\left(x_{kr+i}\right)^{T} d_{kr+i}^{WYL} \\ &\qquad + \left(g\left(x_{kr+i} + \gamma_{kr+i} d_{kr+i}^{WYL}\right) - g\left(x_{kr+i}\right)\right)^{T} d_{kr+i}^{WYL} \\ &\quad \leq g_{kr+i}^{T} d_{kr+i}^{WYL} + \gamma_{kr+i}\left(d_{kr+i}^{WYL}\right)^{T} \widehat{A}_{kr+i} d_{kr+i}^{WYL} \\ &\quad = \left(\frac{\left(d_{kr+i}^{WYL}\right)^{T} \widehat{A}_{kr+i} d_{kr+i}^{WYL}}{\left(d_{kr+i}^{WYL}\right)^{T} A_{kr+i} d_{kr+i}^{WYL}} - 1\right)\left\|g_{kr+i}\right\|^{2} \\ &\quad = O\left(\left\|g_{kr+i}\right\|^{3}\right). \end{aligned} \tag{77}$$

Therefore, by the definition of θ_{k}^{WYL}, we have

$$\begin{aligned} \left|\theta_{kr+i+1}^{WYL}\right| &= \frac{\left|g\left(x_{kr+i+1}\right)^{T} d_{kr+i}^{WYL}\right|}{\left\|g_{kr+i}\right\|^{2}} = O\left(\left\|g_{kr+i}\right\|\right) \\ &= O\left(\left\|g_{kr}\right\|\right). \end{aligned} \tag{78}$$

This completes the proof. □

Theorem 10. *Let Assumptions ii and iii hold; then, for all $i = 0, 1, \cdots$, one gets*

$$\left\|\alpha_{kr+i} d_{kr+i}^{WYL} - \alpha_{kr}^{i} d_{kr}^{WYL(i)}\right\| = O\left(\left\|d_{kr}^{WYL}\right\|^{2}\right) \tag{79}$$

and

$$\left\|\alpha_{kr+i} d_{kr+i}^{WYL} - \alpha_{kr}^{i} d_{kr}^{WYL(i)}\right\| = O\left(\left\|x_{kr} - x^{*}\right\|^{2}\right). \tag{80}$$

Proof. First, we will prove the following relationship. For all $i = 0, 1, \cdots, j(kr)$, we have

$$\left\|g_{kr+i} - g_{kr}^{i}\right\| = O\left(\left\|d_{kr}^{WYL}\right\|^{2}\right), \tag{81}$$

$$\left\|d_{kr+i}^{WYL} - d_{kr}^{WYL(i)}\right\| = O\left(\left\|d_{kr}^{WYL}\right\|^{2}\right), \tag{82}$$

and

$$\left\|\alpha_{kr+i} d_{kr+i}^{WYL} - \alpha_{kr}^{i} d_{kr}^{WYL(i)}\right\| = O\left(\left\|d_{kr}^{WYL}\right\|^{2}\right). \tag{83}$$

For the RWYL algorithm, by $x_{kr}^0 = x_{kr}$ and $\alpha_{kr} = \gamma_{kr}$. The equalities (81)-(83) obviously hold for $i = 0$. Suppose that (81)-(83) hold for $i \geq 0$, we prove that the equalities (81)-(83) hold for $i+1$. By inequality (72), we get

$$\begin{aligned} \left\|g_{kr+i+1} - g_{kr}^{i+1}\right\| &\leq \left\|g_{kr+i} - g_{kr}^{i}\right\| + O\left(\left\|d_{kr}^{WYL}\right\|^2\right) \\ &\quad + M\left\|\alpha_{kr+i}d_{kr}^{WYL} - \alpha_{kr}^{i}d_{kr}^{WYL(i)}\right\| \\ &= O\left(\left\|d_{kr}^{WYL}\right\|^2\right) + O\left(\left\|d_{kr}^{WYL}\right\|^2\right) \\ &\quad + MO\left(\left\|d_{kr}^{WYL}\right\|^2\right) = O\left(\left\|d_{kr}^{WYL}\right\|^2\right) \end{aligned} \tag{84}$$

Using the equality $d_{kr} = -g_{kr}$ and the mean value theorem, we get

$$\begin{aligned} &\left\|d_{kr+i+1}^{WYL} - d_{kr}^{WYL(i+1)}\right\| = \left\|-g_{kr+i+1} + \beta_{kr+i+1}^{WYL}d_{kr+i}^{WYL}\right. \\ &\quad \left. - \theta_{kr+i+1}^{WYL}y_{kr+i}^{WYL} + g_{kr}^{i+1} - \beta_{kr}^{WYL(i+1)}d_{kr}^{WYL(i)}\right\| \\ &\quad \leq \left\|g_{kr+i+1} - g_{kr}^{i+1}\right\| + \left|\theta_{kr+i+1}^{WYL}\right| \cdot \left\|y_{kr+i}^{WYL}\right\| \\ &\quad + \left\|\beta_{kr+i+1}^{WYL}d_{kr+i}^{WYL} - \beta_{kr}^{WYL(i+1)}d_{kr}^{WYL(i)}\right\| \\ &\quad \leq O\left(\left\|d_{kr}^{WYL}\right\|^2\right) + O\left(\left\|d_{kr}^{WYL}\right\|\right) \cdot O\left(\left\|g_{kr}\right\|\right) \\ &\quad + O\left(\left\|d_{kr}^{WYL}\right\|^2\right) = O\left(\left\|d_{kr}^{WYL}\right\|^2\right), \end{aligned} \tag{85}$$

and

$$\begin{aligned} &\left\|\alpha_{kr+i+1}d_{kr+i+1}^{WYL} - \alpha_{kr}^{i+1}d_{kr}^{WYL(i+1)}\right\| \\ &\quad = O\left(\left\|g_{kr+i+1} - g_{kr}^{i+1}\right\|\right) + O\left(\| d_{kr+i+1}^{WYL} - d_{kr}^{WYL(i+1)}\right) \\ &\quad + O\left(\left\|d_{kr}^{WYL}\right\|^2\right) \\ &\quad \leq O\left(\left\|d_{kr}^{WYL}\right\|^2\right) + O\left(\left\|d_{kr}^{WYL}\right\|^2\right) + O\left(\left\|d_{kr}^{WYL}\right\|^2\right) \\ &\quad = O\left(\left\|d_{kr}^{WYL}\right\|^2\right), \end{aligned} \tag{86}$$

where the above inequalities follow (71), (72), (73), (81), and (82). Thus, equalities (81)-(83) hold for all $i = 1, 2, \cdots j(kr)$. Moreover equality (79) holds too. Now we prove that (80) holds. Considering

$$\begin{aligned} \left\|d_{kr}^{WYL}\right\| &= \left\|g_{kr}\right\| = \left\|g\left(x_{kr}\right) - g\left(x^*\right)\right\| \\ &\leq M \cdot \left\|x_{kr} - x^*\right\|, \end{aligned} \tag{87}$$

we get

$$\left\|d_{kr}^{WYL}\right\|^2 \leq M^2 \left\|x_{kr} - x^*\right\|^2. \tag{88}$$

Therefore equality (80) holds. This completes the proof. □

Based on the above lemmas, similar to Theorem 4.2 of [32], we can get the n-step quadratic convergence of the RWYL algorithm. Here we only state it as follows but omit the proof too.

Theorem 11. *Let Assumptions ii and iii hold; then there exists a constant $c' > 0$ satisfying*

$$\lim_{k\to\infty} \sup \frac{\left\|x_{kr+n} - x^*\right\|}{\left\|x_{kr} - x^*\right\|^2} \leq c' < \infty. \tag{89}$$

Namely, the RWYL algorithm is quadratically convergent.

5. Numerical Results

This section reports some numerical experiments with Algorithm 7 (RWYL). In order to show the effectiveness of the given algorithm, we will test the algorithm in [15, 19] (Hager-Zhang), the algorithm in [6, 7] (PRP), and the algorithm in [21] (WYL). In these four algorithms, the Wolfe-Powell line search technique is used as well as the parameters $\varsigma_1 = 0.1$ and $\varsigma_2 = 0.9$. The restart constant $r = 10$. The program will be stopped if $\|g(x_k)\|_\infty \leq \max\{10^{-6}, 10^{-12}\|g(x_0)\|_\infty\}$ or $\|g(x_k)\|_\infty \leq 10^{-6}(1 + |f(x_k)|)$ holds.

5.1. Normal Unstrained Optimization Problems. The unconstrained optimization problems with the given initial points can be found at

http://camo.ici.ro/neculai/THREECG/funcname.txt,

which were collected by Neculai Andrei. The programs are written by Fortran and the codes are downloaded from

http://users.clas.ufl.edu/hager/papers/Software/,

which are written by Hager and Zhang.

All codes run on PC Core 2 Duo CPU at 3.2 GHz, 2.00GB of RAM, and Windows 7 operation system. The dimension of the test problems is 10000, 50000, and 10,0000 variables. The dimension (dim) of the variable, the CPU time in seconds (CPU) and the number of iterations (NI), function evaluations and gradient evaluations (NFG), the final function values, and the norm value of the gradient when the program is stopped for these four algorithms are computed. The profiles of Dolan and Moré [37] are used to analyze the performance data of these four algorithms. The fraction P of problems is plotted where any given algorithm is within a factor t of the best time. In a performance profile plot, the top curve shows that the algorithm solves the most problems in a time, which was within a factor t of the best time.

Figures 1–3 show the performance of the RWYL, Hager-Zhang, PRP, and WYL algorithms with the dimension 10000 about NI, NFG, and CPU time, respectively. It is not difficult to see that these four algorithms can successfully solve the given problems. The RWYL is the best profile among these four algorithms and the normal WYL algorithm has the worst performance.

In Figures 4–6, we use CPU time, NI, and NFG to compare the performance of the conjugate gradient codes RWYL, Hager-Zhang, PRP, and WYL algorithms on the dimension 50000. These four figures indicate that, relative to the CPU time, NI, and NFG, RWYL is fastest, then PRP,

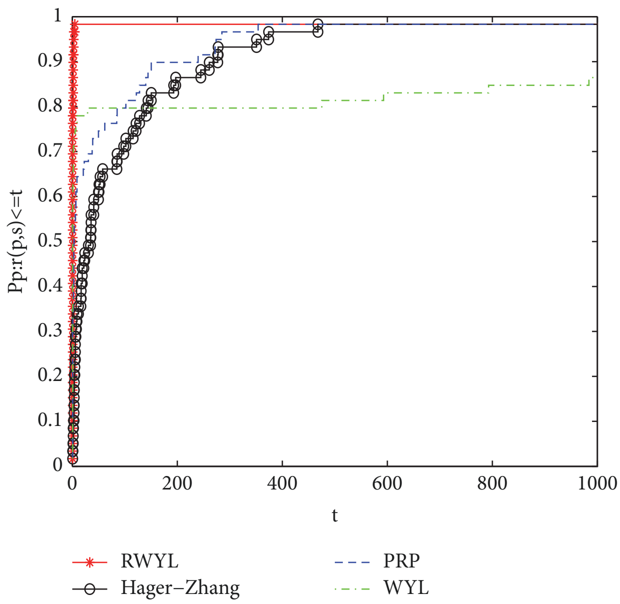

Figure 1: Performance profiles of these algorithms on NI (dim=10000).

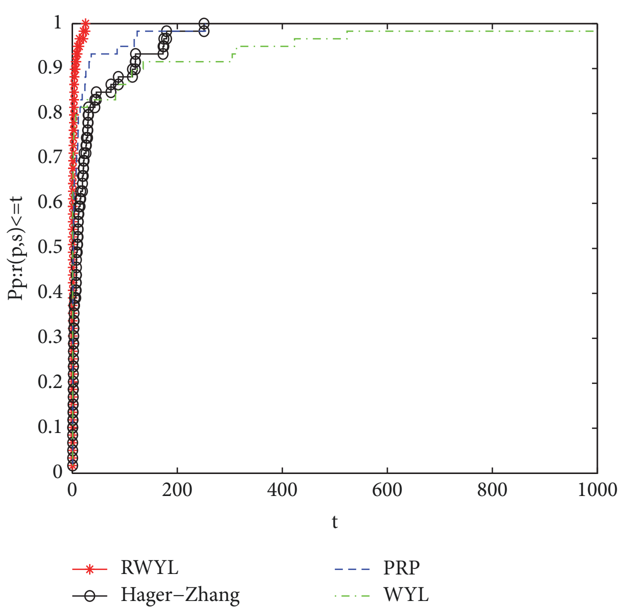

Figure 2: Performance profiles of these algorithms on NFG (dim=10000).

then Hager-Zhang algorithm, and then WYL. These codes only differ in their choice of the search direction; then we can conclude that the RWYL generates the best search directions for these test problems, on average.

In Figures 7–9, we use CPU time, NI, and NFG to compare the performance of the conjugate gradient codes RWYL, Hager-Zhang algorithm, PRP, and WYL on the dimension 100000. These four figures indicate that, relative to the CPU time, NI, and NFG, RWYL is fastest, then Hager-Zhang algorithm, then PRP, and then WYL.

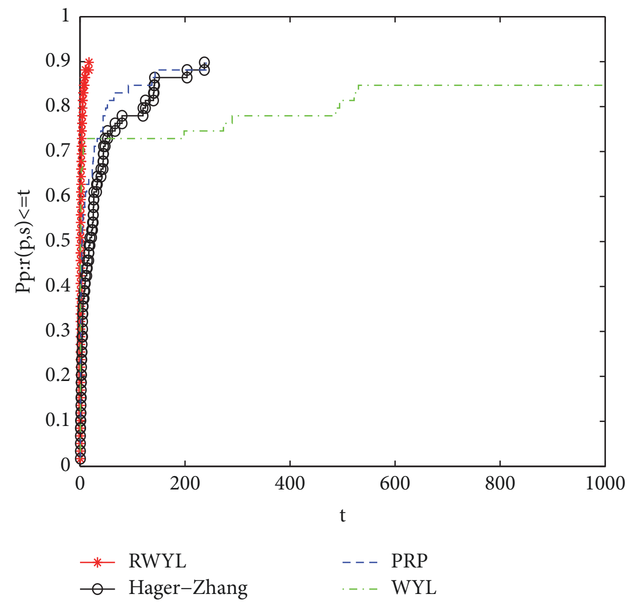

Figure 3: Performance profiles of these algorithms on CPU (dim=10000).

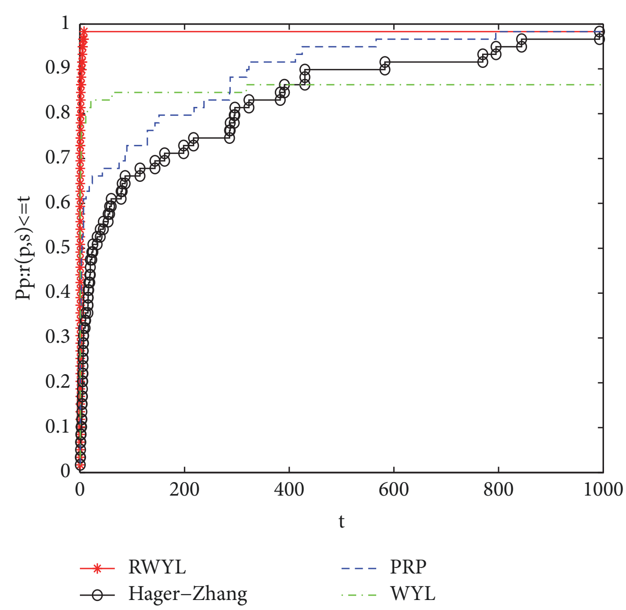

Figure 4: Performance profiles of these algorithms on NI (dim=50000).

According to these nine figures, it is easy to see that the RWYL has the best performance for the dimensions 10000, 50000, and 100000. The Hager-Zhang algorithm becomes competitive with the dimension become large, which shows that the Hager-Zhang algorithm is very effective for large-scale problems. The PRP algorithm has the stable numerical performance for any dimensions. The normal WYL can also successfully solve the optimization problems and its efficiency is limited. To directly show the CPU time, NI, and NFG of these four algorithms, each algorithm number is listed in Table 1.

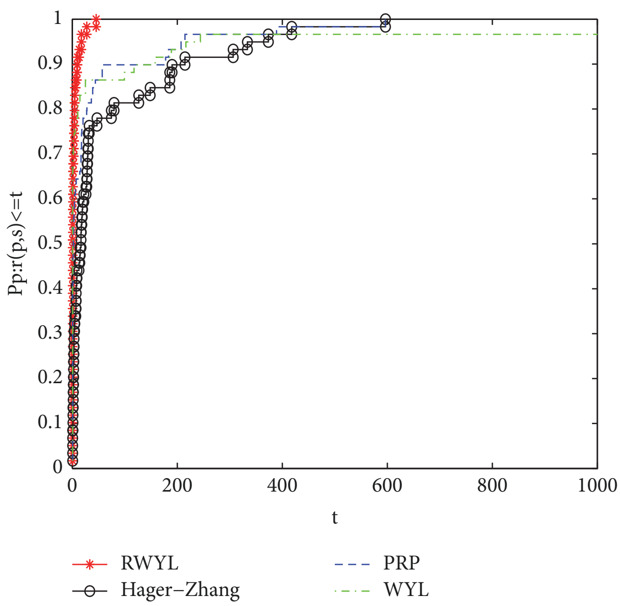

Figure 5: Performance profiles of these algorithms on NFG (dim=50000).

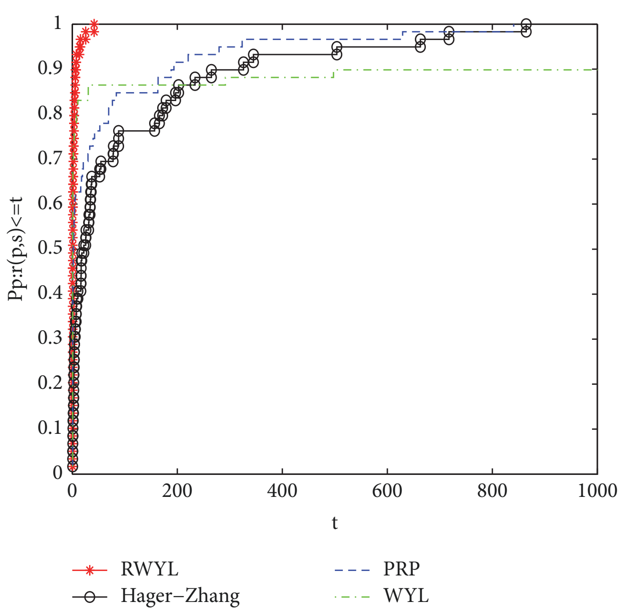

Figure 6: Performance profiles of these algorithms on CPU (dim=50000).

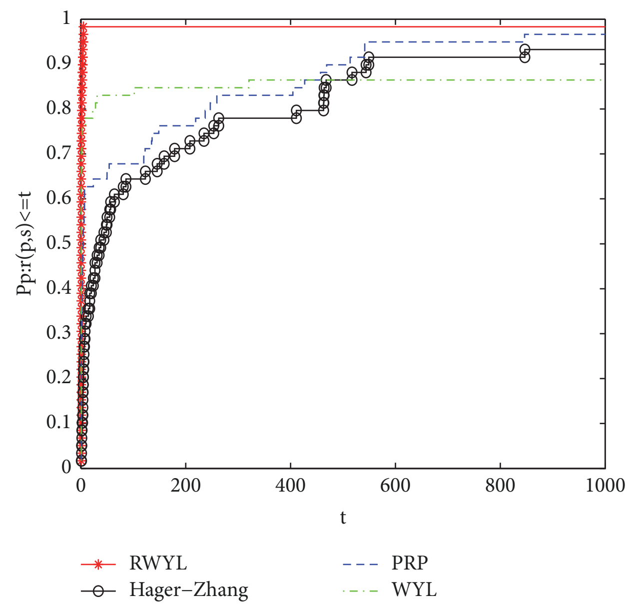

Figure 7: Performance profiles of these algorithms on NI (dim=100000).

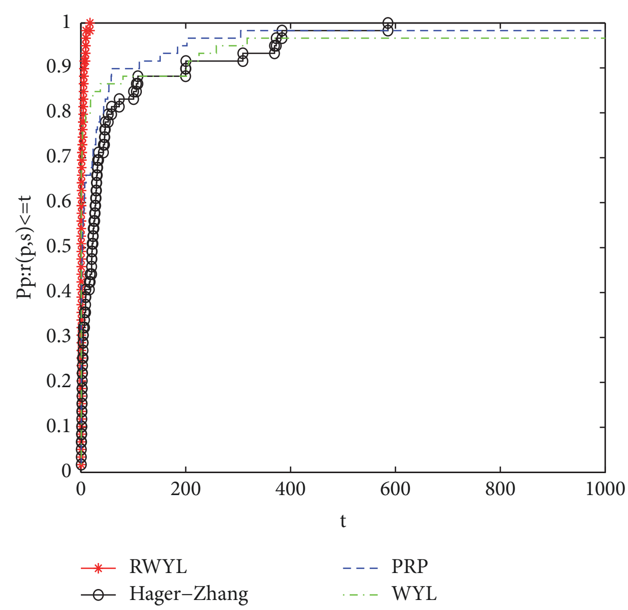

Figure 8: Performance profiles of these algorithms on NFG (dim=100000).

5.2. Benchmark Problems (Engineer Problems). The following Benchmark Problems can be found at

http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume24/ortizboyer05a-html/node6.html.

(1) Sphere function:

$$f_{sph}(x) = \sum_{i=1}^{n} x_i^2, \quad x_i \in [-5.12, 5.12]$$
$$x^* = (0, 0, \cdots, 0), \quad f_{sph}(x^*) = 0. \tag{90}$$

(2) Schwefel function:

$$f_{sch}(x) = 418.9828n + \sum_{i=1}^{n} x_i \sin\sqrt{|x_i|},$$
$$x_i \in [-512.03, 511.97] \tag{91}$$
$$x^* = (-420.9678, -420.9678, \cdots, -420.9678),$$
$$f_{sch}(x^*) = 0.$$

Table 1: Results of total.

Algorithm	dim=10000			dim=50000			dim=100000		
	NI	NFG	CPU	NI	NFG	CPU	NI	NFG	CPU
RWYL	144	4035	8.78	222	4546	76.88	151	3653	141.79
Hager-Zhang	8655	33062	116.14	16329	58406	950.75	20661	72516	2251.47
PRP	6769	25186	144.00	12563	43475	699.07	116090	354062	10985.57
WYL	59433	181937	415.10	89221	271059	3117.06	167044	503781	11868.37

Table 2: Results of Benchmark Problems for RWYL.

No.	1	2	3	4	5
x_0	$(-0.001,\cdots)$	$(-0.001,\cdots)$	$(-0.001,\cdots)$	$(-0.001,\cdots)$	$(-0.001,\cdots)$
	NI/NFG/CPU	NI/NFG/CPU	NI/NFG/CPU	NI/NFG/CPU	NI/NFG/CPU
n=300	2/6/4.680e-2	1/3/1.560e-2	150/450/1.926e+1	31/93/1.092e+0	3/9/9.360e-2
n=1000	2/6/1.560e-2	1/3/0.000e+0	452/1356/2.033e+3	51/153/1.070e+1	801/6394/4.570e+0
x_0	$(0.001,\cdots)$	$(0.001,\cdots)$	$(0.001,\cdots)$	$(0.001,\cdots)$	$(0.001,\cdots)$
	NI/NFG/CPU	NI/NFG/CPU	NI/NFG/CPU	NI/NFG/CPU	NI/NFG/CPU
n=300	2/6/1.560e-2	1/3/1.560e-02	150/450/1.931e+1	31/93/7.020e-1	3/9/6.240e-2
n=1000	2/6/1.560e-2	1/3/0.000e+0	452/1356/2.033e+3	51/153/1.065e+1	801/6394/4.243e+0

(3) Schwefel's function:

$$f_{schds}(x) = \sum_{i=1}^{n}\left(\sum_{j=1}^{i} x_j\right)^2, \quad x_i \in [-65.536, 65.536]$$
$$x^* = (0,0,\cdots,0), \quad f_{schds}(x^*) = 0. \tag{92}$$

(4) Griewank function:

$$f_{gri}(x) = 1 + \sum_{i=1}^{n}\frac{x_i^2}{4000} - \prod_{i=1}^{n}\cos\frac{x_i}{i}, \quad x_i \in [-600, 600]$$
$$x^* = (0,0,\cdots,0), \quad f_{gri}(x^*) = 0. \tag{93}$$

(5) Rastrigin function:

$$f_{ras}(x) = 10n + \sum_{i=1}^{n}\left(x_i^2 - 10\cos(2\pi x_i)\right), \quad x_i \in [-5.12, 5.12]$$
$$x^* = (0,0,\cdots,0), \quad f_{ras}(x^*) = 0. \tag{94}$$

Benchmark Problems are from the engineer fields and there are many scholars focusing on the studies of these problems. The given algorithm of this paper can also successfully solve them. We do the experiments about the RWYL and the normal WYL for comparing and omit the other two methods Hager-Zhang and PRP. The parameters and the stopping rule are the same as the above subsection. The codes are written by Matlab 2017 and run on PC Core 2 Duo CPU at 2.26 GHz, 6.00GB of RAM, and Windows 7 operation system. The dimension is 300 and 1000 variables. The detailed numerical results are listed in Tables 2 and 3.

To directly see the results of Tables 2 and 3, we compute the total NI and NFG and set them in Table 4.

The results of Table 4 show that the restart algorithm is more competitive with the normal algorithm for the Benchmark Problems.

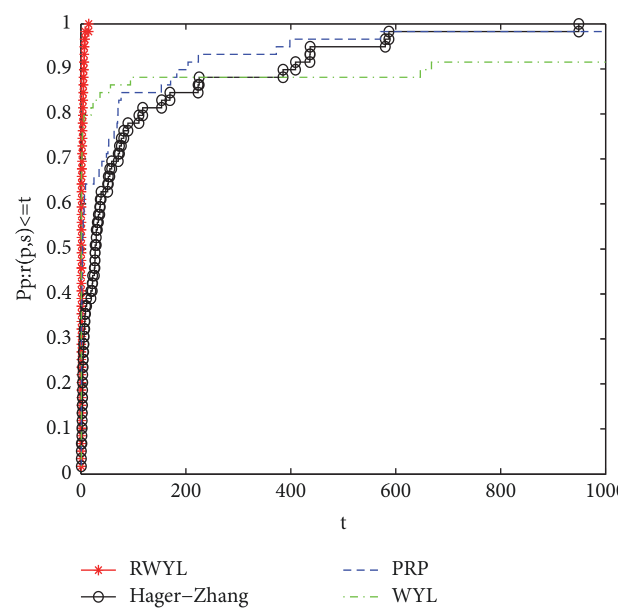

Figure 9: Performance profiles of these algorithms on CPU (dim=100000).

Table 3: Results of Benchmark Problems for WYL.

No.	1	2	3	4	5
x_0	$(-0.001,\cdots)$	$(-0.001,\cdots)$	$(-0.001,\cdots)$	$(-0.001,\cdots)$	$(-0.001,\cdots)$
	NI/NFG/CPU	NI/NFG/CPU	NI/NFG/CPU	NI/NFG/CPU	NI/NFG/CPU
n=300	519/2075/1.684e+0	1/3/1.560e-2	801/4002/5.396e+1	141/389/1.716e+0	801/4003/1.981e+0
n=1000	549/2195/1.809e+0	1/3/0.000e+0	801/4803/1.824e+3	152/419/1.622e+1	801/4003/2.199e+0
x_0	$(0.001,\cdots)$	$(0.001,\cdots)$	$(0.001,\cdots)$	$(0.001,\cdots)$	$(0.001,\cdots)$
	NI/NFG/CPU	NI/NFG/CPU	NI/NFG/CPU	NI/NFG/CPU	NI/NFG/CPU
n=300	519/2075/1.606e+0	1/3/0.000e+0	801/4002/5.343e+1	141/389/1.638e+0	801/4003/1.996e+0
n=1000	549/2195/1.450e+0	1/3/0.000e+0	801/4803/1.822e+3	152/419/1.597e+1	801/4003/2.230e+0

Table 4: Results of Tables 2 and 3.

	RWYL	WYL
NI	2988	9134
NFG	16946	43790

5.3. Parameters Estimation of Nonlinear Muskingum Models. The basic Muskingum model, the continuity, and storage equations are defined by

$$\frac{dS_t}{dt} = I_t - Q_t, \quad S_t = k\left[xI_t + (1-x)Q_t\right], \tag{95}$$

where, at time t, S_t is channel storage, I_t is rate of inflow and Q_t = denotes outflow, k is storage-time constant, and x is weighting factor for the river reach. The generalized trapezoidal formula [38] is

$$\begin{aligned} \min \quad & f(k,x,m) \\ & = \sum_{i=1}^{n-1} \min\left(1-\frac{\Delta t}{2}\theta\right) k\left[xI_{i+1} + (1-x)Q_{i+1}\right]^m \\ & \quad - \left(1-\frac{\Delta t}{2}\theta\right) k\left[xI_i + (1-x)Q_i\right]^m \\ & \quad - \frac{\Delta t}{2}\left[I_i - Q_i\right] + \frac{\Delta t}{2}\left(1-\Delta_t\theta\right)\left[I_{i+1} - Q_{i+1}\right]. \end{aligned} \tag{96}$$

To conveniently estimate the parameters k, x, and m in the nonlinear Muskingum model, the objective function can rewritten as

$$\begin{aligned} \min \quad & SSQ1 \\ & = \sum_{i=1}^{n-1}\left(\left(1-\frac{\Delta t}{2}\right)k\left[xI_{i+1}+(1-x)Q_{i+1}\right]^m - \left(1-\frac{\Delta t\theta}{2}\right)k\left[xI_i+(1-x)Q_i\right]^m - \frac{\Delta t}{2}\left(I_i-Q_i\right) - \frac{\Delta t}{2}\left(1-\Delta t\theta\right)\left(I_{i+1}-Q_{i+1}\right)\right)^2, \end{aligned} \tag{97}$$

$$\begin{aligned} \min \quad & SSQ2 \\ & = \sum_{i=1}^{n-1}\left(\left(1-\frac{\Delta t}{2}\right)k\left[xI_{i+1}^m+(1-x)Q_{i+1}^m\right] - \left(1-\frac{\Delta t\theta}{2}\right)k\left[xI_i^m+(1-x)Q_i^m\right] - \frac{\Delta t}{2}\left(I_i-Q_i\right) - \frac{\Delta t}{2}\left(1-\Delta t\theta\right)\left(I_{i+1}-Q_{i+1}\right)\right)^2. \end{aligned} \tag{98}$$

This subsection will use our RWYL to estimate the parameters of the above two Muskingum models, named Model 1 (97) and Model 2 (98).

All in all, we can conclude that that the restart algorithm is competitive with the norm algorithm without restart technique and other similar algorithms.

6. Conclusions

Nonlinear conjugate gradient algorithm is one of the most effective algorithms in optimization algorithms, especially for large-scale optimization problems. Many scholars have obtained many interesting results in this field. This paper focuses on a modified WYL CG algorithm with restart technique for large-scale optimization. In our opinion, there are at least seven issues that warrant further research and improvement: (i) The first issue that should be considered is the choice of the restart parameter r in the RWYL algorithm; the value ($r = 10$) used here is not the only choice. (ii) The second important issue is the termination condition; better termination conditions may exist for the CG algorithms, which may improve the numerical performance and convergence. (iii) Under the restart strategy, other similar CG algorithms with quadratic convergence are worth studying. (iv) It would be interesting to test the performance of the given algorithm when applied to other optimization problems that arise in the image processing field. (v) We all know that the nonmonotone line search techniques are very effective. In the future, we will study the possibility of combining CG algorithms with nonmonotonic techniques for large-scale optimization problems and will attempt to obtain good results. (vi) In the experiment, there are 59 optimization problems with the dimension 10000, 50000, and 100000 variables that are tested. We also do the test for the Benchmark Problems which has wild applications in engineer fields. In the future, more problems and numerical experiments should be done to turn out the performance of the CG algorithms. (vii) The last issue is the use of the CG algorithm for nonsmooth optimization and nonlinear equations, which we consider to be important for future research. All these topics will be the focus of our future work.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant no. 11661009), the Guangxi Science Fund for Distinguished Young Scholars (no. 2015GXNSFGA139001), the Guangxi Natural Science Key Fund (no. 2017GXNSFDA198046), and the Guangxi Science Fund of Young and Middle-Aged Teachers for the Basic Ability Promotion (no. 2017KY0019).

References

[1] Y. H. Dai and Y. Yuan, "A nonlinear conjugate gradient method with a strong global convergence property," *SIAM Journal on Optimization*, vol. 10, no. 1, pp. 177–182, 1999.

[2] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1987.

[3] R. Fletcher and C. M. Reeves, "Function minimization by conjugate gradients," *The Computer Journal*, vol. 7, pp. 149–154, 1964.

[4] 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 (1953), 1952.

[5] Y. Liu and C. Storey, "Efficient generalized conjugate gradient algorithms, Part 1," *Journal of Optimization Theory and Applications*, vol. 69, no. 1, pp. 129–137, 1991.

[6] B. Polak, "The conjugate gradient method in extreme problems," *Computational Mathematics and Mathematical Physics*, vol. 9, pp. 94–112, 1969.

[7] 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.

[8] G. Yuan and Z. Sheng, *Press of Science*, Press of Science, Beijing, China, 2017.

[9] 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.

[10] Y. Yuan, "Analysis on the conjugate gradient method," *Optimization Methods and Software*, vol. 2, no. 1, pp. 19–29, 2010.

[11] Y. H. Dai and Y. Yuan, *Nonlinear conjugate gradient Methods*, vol. 14, Shanghai Scientific and Technical Publishers, 1998.

[12] X. Dong, H. Liu, and Y. He, "A self-adjusting conjugate gradient method with sufficient descent condition and conjugacy condition," *Journal of Optimization Theory and Applications*, vol. 165, no. 1, pp. 225–241, 2015.

[13] X. L. Dong, H. W. Liu, Y. B. He, and X. . Yang, "A modified Hestenes-Stiefel conjugate gradient method with sufficient descent condition and conjugacy condition," *Journal of Computational and Applied Mathematics*, vol. 281, pp. 239–249, 2015.

[14] 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.

[15] W. W. Hager and H. Zhang, "Algorithm 851: CG-DESCENT, a conjugate gradient method with guaranteed descent," *ACM Transactions on Mathematical Software*, vol. 32, no. 1, pp. 113–137, 2006.

[16] M. J. Powell, "Convergence properties of algorithms for nonlinear optimization," *SIAM Review*, vol. 28, no. 4, pp. 487–500, 1986.

[17] G. Yuan, "Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems," *Optimization Letters*, vol. 3, no. 1, pp. 11–21, 2009.

[18] Y.-h. Dai, *Analysis of conjugate gradient methods [Ph.D. Thesis]*, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1997.

[19] W. W. Hager and H. 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.

[20] X. Li, S. Wang, Z. Jin, and H. Pham, "A conjugate gradient algorithm under Yuan-Wei-Lu line search technique for large-scale minimization optimization models," *Mathematical Problems in Engineering*, vol. 2018, Article ID 4729318, 11 pages, 2018.

[21] 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.

[22] G. Yuan and X. Lu, "A modified PRP conjugate gradient method," *Annals of Operations Research*, vol. 166, pp. 73–90, 2009.

[23] G. Yuan, X. Lu, and Z. Wei, "A conjugate gradient method with descent direction for unconstrained optimization," *Journal of Computational and Applied Mathematics*, vol. 233, no. 2, pp. 519–530, 2009.

[24] G. Yuan, Z. Meng, and Y. Li, "A modified Hestenes and Stiefel conjugate gradient algorithm for large-scale nonsmooth minimizations and nonlinear equations," *Journal of Optimization Theory and Applications*, vol. 168, no. 1, pp. 129–152, 2016.

[25] G. Yuan, Z. Wei, and G. Li, "A modified Polak-Ribière-Polyak conjugate gradient algorithm for nonsmooth convex programs," *Journal of Computational and Applied Mathematics*, vol. 255, pp. 86–96, 2014.

[26] G. Yuan and M. Zhang, "A three-terms Polak-Ribière-Polyak conjugate gradient algorithm for large-scale nonlinear equations," *Journal of Computational and Applied Mathematics*, vol. 286, pp. 186–195, 2015.

[27] G. Yuan, Z. Sheng, B. Wang, W. Hu, and C. Li, "The global convergence of a modified BFGS method for nonconvex functions," *Journal of Computational and Applied Mathematics*, vol. 327, pp. 274–294, 2018.

[28] G. Yuan, Z. Wei, and X. Lu, "Global convergence of BFGS and PRP methods under a modified weak Wolfe-Powell line search," *Applied Mathematical Modelling: Simulation and Computation for Engineering and Environmental Systems*, vol. 47, pp. 811–825, 2017.

[29] W. Burmeister, "Die Konvergenzordnung des Fletcher-Powell-Algorithmus," *Zeitschrift für Angewandte Mathematik und Mechanik. Ingenieurwissenschaftliche Forschungsarbeiten*, vol. 53, pp. 693–699, 1973.

[30] A. I. Cohen, "Rate of convergence of several conjugate gradient algorithms," *SIAM Journal on Numerical Analysis*, vol. 9, pp. 248–259, 1972.

[31] K. Ritter, "On the rate of superlinear convergence of a class of variable metric methods," *Numerische Mathematik*, vol. 35, no. 3, pp. 293–313, 1980.

[32] D.-H. Li and B.-S. Tian, "n-step quadratic convergence of the MPRP method with a restart strategy," *Journal of Computational and Applied Mathematics*, vol. 235, no. 17, pp. 4978–4990, 2011.

[33] H. Huang, Z. Wei, and Y. Shengwei, "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.

[34] 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.

[35] 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.

[36] L. Zhang, W. Zhou, and D. H. Li, "A descent modified Polak-Ribiere-Polyak conjugate gradient method and its global convergence," *IMA Journal of Numerical Analysis (IMAJNA)*, vol. 26, no. 4, pp. 629–640, 2006.

[37] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles," *Mathematical Programming*, vol. 91, pp. 201–213, 2002.

[38] M. M. Chawla, M. A. Al-Zanaidi, and D. J. Evans, "Generalized trapezoidal formulas for parabolic equations," *International Journal of Computer Mathematics*, vol. 70, no. 3, pp. 429–443, 1999.

Advances in Operations Research
Advances in Decision Sciences
Journal of Applied Mathematics
The Scientific World Journal
Journal of Probability and Statistics
International Journal of Mathematics and Mathematical Sciences
Hindawi
Submit your manuscripts at
www.hindawi.com
Journal of Optimization
International Journal of Engineering Mathematics
International Journal of Analysis
Journal of Complex Analysis
Advances in Numerical Analysis
Mathematical Problems in Engineering
International Journal of Differential Equations
Discrete Dynamics in Nature and Society
International Journal of Stochastic Analysis
Journal of Mathematics
Journal of Function Spaces
Abstract and Applied Analysis
Advances in Mathematical Physics