Research Article

A Conjugate Gradient Method for Unconstrained Optimization Problems

Gonglin Yuan

College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi 530004, China

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

A hybrid method combining the FR conjugate gradient method and the WYL conjugate gradient method is proposed for unconstrained optimization problems. The presented method possesses the sufficient descent property under the strong Wolfe-Powell (SWP) line search rule relaxing the parameter $\sigma < 1$. Under the suitable conditions, the global convergence with the SWP line search rule and the weak Wolfe-Powell (WWP) line search rule is established for nonconvex function. Numerical results show that this method is better than the FR method and the WYL method.

1. Introduction

Consider the following n variables unconstrained optimization problem:

$$\min_{x \in \mathfrak{R}^n} f(x), \tag{1.1}$$

where $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$ is smooth and its gradient $g(x)$ is avaible. The nonlinear conjugate gradient (CG) method for (1.1) is designed by the iterative form

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \ldots, \tag{1.2}$$

where x_k is the kth iterative point, $\alpha_k > 0$ is a steplength, and d_k is the search direction defined by

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

where $\beta_k \in \Re$ is a scalar which determines the different conjugate gradient methods [1, 2], and g_k is the gradient of $f(x)$ at the point x_k. There are many well-known formulas for β_k, such as the Fletcher-Reeves (FR) [3], Polak-Ribière-Polyak (PRP) [4], Hestenses-Stiefel (HS) [5], Conjugate-Descent (CD) [6], Liu-Storrey (LS) [7], and Dai-Yuan (DY) [8]. The CG method is a powerful line search method for solving optimization problems, and it remains very popular for engineers and mathematicians who are interested in solving large-scale problems [9–11]. This method can avoid, like steepest descent method, the computation and storage of some matrices associated with the Hessian of objective functions. Then there are many new formulas that have been studied by many authors (see [12–20] etc.).

The following formula for β_k is the famous FR method:

$$\beta_k^{\mathrm{FR}} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \tag{1.4}$$

where g_k and g_{k+1} are the gradients $\nabla f(x_k)$ and $\nabla f(x_{k+1})$ of $f(x)$ at the point x_k and x_{k+1}, respectively, $\|\cdot\|$ denotes the Euclidian norm of vectors. Throughout this paper, we also denote $f(x_k)$ by f_k. Under the exact line search, Powell [21] analyzed the small-stepsize property of the FR conjugate gradient method and its global convergence, Zoutendijk [22] proved its global convergence for nonconvex function. Al-Baali [23] proved the sufficient descent condition and the global convergence of the FR conjugate gradient method with the SWP line search by restricting the parameter $\sigma < 1/2$. Liu et al. [24] extended the result to the the parameter $\sigma = 1/2$. Wei et al. (WYL) [17] proposed a new conjugate gradient formula:

$$\beta_k^{\mathrm{WYL}} = \frac{g_{k+1}^T \left(g_{k+1} - \left(\|g_{k+1}\| / \|g_k\|\right) g_k\right)}{\|g_k\|^2}. \tag{1.5}$$

The numerical results show that this method is competitive to the PRP method, the global convergence of this method with the exact line search and Grippo-Lucidi line search conditions is proved. Huang et al. [25] proved that by restricting the parameter $\sigma < 1/4$, under the SWP line search rule, this method has the sufficient descent property. Then it is an interesting task to extend the bound of the parameter σ and get the sufficient descent condition.

The sufficient descent condition

$$g_k^T d_k \leq -c\|g_k\|^2, \quad \forall k \geq 0, \tag{1.6}$$

where $c > 0$ is a constant, is crucial to insure the global convergence of the nonlinear conjugate gradient method [23, 26–28]. In order to get some better results of the conjugate gradient methods, Andrei [29, 30] proposed the hybrid conjugate gradient algorithms as convex combination of some other conjugate gradient algorithms. Motivated by the ideas of Andrei [29, 30] and the above observations, we will give a hybrid method combining the FR method and the WYL method. The proposed method, relaxing the parameter $\sigma < 1$, under the SWP line search technique, possesses the sufficient descent condition (1.6). The global convergence with the SWP line search and the WWP line search of our method is established for the nonconvex functions. Numerical results show that the presented method is competitive to the FR and the WYL method.

In the following section, the algorithm is stated. The properties and the global convergence of the new method are proved in Section 3. Numerical results are reported in Section 4 and one conclusion is given in Section 5.

2. Algorithm

Now we describe our algorithm as follows.

Algorithm 2.1 (the hybrid method).

Step 1. Choose an initial point $x_0 \in \mathfrak{R}^n$, $\varepsilon \in (0,1)$, $\lambda_1 \geq 0$, $\lambda_2 \geq 0$. Set $d_0 = -g_0 = -\nabla f(x_0)$, $k := 0$.

Step 2. If $\|g_k\| \leq \varepsilon$, then stop; otherwise go to the next step.

Step 3. Compute step size α_k by some line search rules.

Step 4. Let $x_{k+1} = x_k + \alpha_k d_k$. If $\|g_{k+1}\| \leq \varepsilon$, then stop.

Step 5. Calculate the search direction

$$d_{k+1} = -g_{k+1} + \beta_k^H d_k, \tag{2.1}$$

where $\beta_k^H = \lambda_1 \beta_k^{\text{WYL}} + \lambda_2 \beta_k^{\text{FR}}$.

Step 6. Set $k := k + 1$, and go to Step 3.

3. The Properties and the Global Convergence

In the following, we assume that $g_k \neq 0$ for all k, otherwise a stationary point has been found. The following assumptions are often used to prove the convergence of the nonlinear conjugate gradient methods (see [3, 8, 16, 17, 27]).

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

(ii) In an open convex, set Ω_0 that contains Ω, f is differentiable, and its gradient g is Lipschitz continuous, namely, there exists a constants $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in \Omega_0. \tag{3.1}$$

3.1. The Properties with the Strong Wolfe-Powell Line Search

The strong Wolfe-Powell (SWP) search rule is to find a step length α_k such that

$$f(x_k + \alpha_k d_k) \leq f_k + \delta \alpha_k g_k^T d_k, \tag{3.2}$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq \sigma \left| g_k^T d_k \right|, \tag{3.3}$$

where $\delta \in (0, 1/2)$, $\sigma \in (0, 1)$.

The following theorem shows that the hybrid algorithm with the SWP line search possesses the sufficient condition (1.6) only under the parameter $\sigma \in (\delta, 1)$.

Theorem 3.2. *Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm 2.1, and let the stepsize α_k be determined by the SWP line search (3.2) and (3.3), if $\sigma \in (0, 1)$, $2\lambda_1 + \lambda_2 \in (0, 1/2\sigma)$, then the sufficient descent condition (1.6) holds.*

Proof. By the definition λ_1, λ_2 and the formulae (1.4) and (1.5), we have

$$
\begin{aligned}
\beta_k^H &= \lambda_1 \beta_k^{\mathrm{WYL}} + \lambda_2 \beta_k^{\mathrm{FR}} \\
&\geq \frac{\lambda_1 \left(\|g_{k+1}\|^2 - (\|g_{k+1}\| / \|g_k\|) \|g_{k+1}\| \|g_k\| \right)}{\|g_k\|^2} + \frac{\lambda_2 \|g_{k+1}\|^2}{\|g_k\|^2} \\
&= \lambda_2 \frac{\|g_{k+1}\|^2}{\|g_k\|^2},
\end{aligned}
\tag{3.4}
$$

$$
\begin{aligned}
\left| \beta_k^H \right| &= \left| \lambda_1 \beta_k^{\mathrm{WYL}} + \lambda_2 \beta_k^{\mathrm{FR}} \right| \\
&\leq \frac{\lambda_1 \left(\|g_{k+1}\|^2 + (\|g_{k+1}\| / \|g_k\|) \|g_{k+1}\| \|g_k\| \right)}{\|g_k\|^2} + \lambda_2 \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \\
&= \frac{(2\lambda_1 + \lambda_2) \|g_{k+1}\|^2}{\|g_k\|^2}.
\end{aligned}
\tag{3.5}
$$

Using (3.3) and the above inequality, we get

$$
\left| \beta_k^H g_{k+1}^T d_k \right| \leq \frac{(2\lambda_1 + \lambda_2) \|g_{k+1}\|^2}{\|g_k\|^2} \sigma \left| g_k^T d_k \right|.
\tag{3.6}
$$

By (2.1), we have

$$
\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} = -1 + \beta_k^H \frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2}.
\tag{3.7}
$$

We prove the descent property of $\{d_k\}$ by induction. Since $g_0^T d_0 = -\|g_0\|^2 < 0$, if $g_0 \neq 0$, now we suppose that d_i, $i = 1, 2, \ldots, k$, are all descent directions, for example, $d_i^T g_i < 0$.

By (3.6), we get

$$
\left| \beta_k^H g_{k+1}^T d_k \right| \leq \frac{\sigma(2\lambda_1 + \lambda_2) \|g_{k+1}\|^2}{\|g_k\|^2} \left(-g_k^T d_k \right),
\tag{3.8}
$$

that is,

$$\frac{\|g_{k+1}\|^2}{\|g_k\|^2}(2\lambda_1 + \lambda_2)\sigma g_k^T d_k \le \beta_k^H g_{k+1}^T d_k \le -\frac{\|g_{k+1}\|^2}{\|g_k\|^2}(2\lambda_1 + \lambda_2)\sigma g_k^T d_k. \tag{3.9}$$

However, from (3.7) together with (3.9), we deduce

$$-1 + (2\lambda_1 + \lambda_2)\sigma \frac{g_k^T d_k}{\|g_k\|^2} \le \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \le -1 - (2\lambda_1 + \lambda_2)\sigma \frac{g_k^T d_k}{\|g_k\|^2}. \tag{3.10}$$

Repeating this process and using the fact $d_0^T g_0 = -\|g_0\|^2$ imply

$$-\sum_{i=0}^{k}[(2\lambda_1 + \lambda_2)\sigma]^i \le \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \le -2 + \sum_{i=0}^{k}[(2\lambda_1 + \lambda_2)\sigma]^i. \tag{3.11}$$

By the restriction $\sigma \in (0,1)$ and $2\lambda_1 + \lambda_2 \in (0, 1/2\sigma)$, we have $(2\lambda_1 + \lambda_2)\sigma \in (0, 1/2)$. So

$$\sum_{i=0}^{k}[(2\lambda_1 + \lambda_2)\sigma]^i < \sum_{i=0}^{\infty}[(2\lambda_1 + \lambda_2)\sigma]^i = \frac{1}{1 - 4\lambda_1\sigma}. \tag{3.12}$$

Then (3.11) can be rewritten as

$$-\frac{1}{1 - (2\lambda_1 + \lambda_2)\sigma} \le \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \le -2 + \frac{1}{1 - (2\lambda_1 + \lambda_2)\sigma}. \tag{3.13}$$

Thus, by induction, $g_k^T d_k < 0$ holds for all $k \ge 0$.

Denote $c = 2 - 1/(1 - (2\lambda_1 + \lambda_2)\sigma)$, then $c \in (0,1)$ and (3.13) turns out to be

$$c - 2 \le \frac{g_k^T d_k}{\|g_k\|^2} \le -c, \tag{3.14}$$

this implies that (1.6) holds. The proof is complete. □

Lemma 3.3. *Suppose that Assumption 3.1 holds. Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm 2.1, let the stepsize α_k be determined by the SWP line search (3.2) and (3.3), and let the conditions in Theorem 3.2 hold. Then the Zoutendijk condition [22]*

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty \tag{3.15}$$

holds.

By the same way, if Assumption 3.1 and the condition $g_k^T d_k < 0$ (for all k) hold, (3.15) also holds for the exact line search, the Armijo-Goldstein line search, and the weak Wolfe-Powell line search. The proofs can be seen in [31, 32]. Now, we prove the global convergence theorem of Algorithm 2.1 with the SWP line search.

Theorem 3.4. *Suppose that Assumption 3.1 holds. Let the sequence $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm 2.1, let the stepsize α_k be determined by the SWP line search (3.2) and (3.3), let the conditions in Theorem 3.2 hold, and let the parameter $2\lambda_1 + \lambda_2 \leq 1$. Then*

$$\lim_{k \to \infty} \inf \|g_k\| = 0. \tag{3.16}$$

Proof. By (1.6), (3.3), and the Zoutendijk condition (3.15), we get

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty. \tag{3.17}$$

Denote

$$t_k = \frac{\|d_k\|^2}{\|g_k\|^4}, \tag{3.18}$$

so (3.17) can be rewritten as

$$\sum_{k=0}^{\infty} \frac{1}{t_k} < +\infty. \tag{3.19}$$

We prove the result of this theorem by contradiction. Assume that this theorem is not true, then there exists a positive constant $\gamma > 0$ such that

$$\|g_k\| \geq \gamma, \quad \forall k \geq 0. \tag{3.20}$$

Squaring both sides of (2.1), we obtain

$$\|d_k\|^2 = \|g_k\|^2 - 2\beta_{k-1}^H g_k^T d_{k-1} + \left(\beta_{k-1}^H \|d_{k-1}\|\right)^2. \tag{3.21}$$

Dividing both sides by $\|g_k\|^4$, applying (3.4), (3.18), and the parameter $2\lambda_1 + \lambda_2 \leq 1$, we get

$$t_k \leq t_{k-1} + \frac{1}{\|g_k\|^2}\left(1 + \frac{2|g_k^T d_{k-1}|}{\|g_{k-1}\|^2}\right). \tag{3.22}$$

Using (3.3) and (1.6), we have

$$t_k \le t_{k-1} + \frac{1}{\|g_k\|^2}\left(1 + \frac{2\sigma|g_{k-1}^T d_{k-1}|}{\|g_{k-1}\|^2}\right) \le t_{k-1} + [1 + 2\sigma(2-c)]\frac{1}{\|g_k\|^2}. \tag{3.23}$$

Repeating this process and using the fact $t_0 = 1/\|g_0\|^2$, we get

$$t_k \le [1 + 2\sigma(2-c)]\sum_{i=1}^{\infty}\frac{1}{\|g_i\|^2}. \tag{3.24}$$

Now, combining (3.24) and (3.20), we get another formula

$$t_k \le [1 + 2\sigma(2-c)]\frac{k+1}{\gamma^2}. \tag{3.25}$$

Thus

$$\sum_{k=0}^{\infty}\frac{1}{t_k} = +\infty, \tag{3.26}$$

this contradicts the condition (3.19). However, the conclusion of this theorem is correct. The following theorem will show that Algorithm 2.1 with the SWP line search, only under the descent condition

$$g_k^T d_k < 0, \quad \forall k \ge 0, \tag{3.27}$$

is global convergence too. □

Theorem 3.5. *Suppose that Assumption 3.1 holds. Let the sequence $\{g_k, d_k\}$ be generated by Algorithm 2.1, let the stepsize α_k be determined by the SWP line search (3.2) and (3.3), let the parameter $2\lambda_1 + \lambda_2 \le 1$, and let (3.27) hold. Then, for all $k \ge 0$, the following inequalities holds:*

$$\min\{r_k, r_{k+1}\} \ge \frac{1}{2}, \tag{3.28}$$

where

$$r_k = -\frac{g_k^T d_k}{\|g_k\|^2}, \tag{3.29}$$

furthermore, (3.16) holds

Proof. By (2.1), (3.3), (3.27), and $2\lambda_1 + \lambda_2 \le 1$, we can deduce that (3.10) holds for all $\sigma < 1$ and $\sigma_* = (2\lambda_1 + \lambda_2)\sigma < 1$. According to the second inequality of (3.10), we have

$$\sigma_* r_k + r_{k+1} \ge 1. \tag{3.30}$$

Similar to the way of [31], it is not difficult to get (3.28).

Secondly, using (3.30) and the Cauchy-Schwarz inequality implies that (or see [31])

$$r_k^2 + r_{k+1}^2 \ge \left(1 + \sigma_*^2\right)^{-1}. \tag{3.31}$$

Moreover, repeating the process of the first inequality of (3.10) and using $r_0 = 1$, we get

$$r_{k+1} < (1 - \sigma_*)^{-1}. \tag{3.32}$$

By (3.3), (3.22), and the above inequality, we have

$$t_{k+1} \le \frac{1 + \sigma_*}{1 - \sigma_*} \sum_{i=0}^{k+1} \frac{1}{\|g_i\|^2}. \tag{3.33}$$

Thus, using (3.20), we have

$$t_{k+1} \le b_1(k + 2), \tag{3.34}$$

where $b_1 = 1 + \sigma_* / (1 - \sigma_*)\gamma^2$. By Zoutendijk condition (3.15), (3.31), and (3.34), we obtain

$$\begin{aligned}
+\infty > \sum_{k=0}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} &= \sum_{k=0}^{\infty} \frac{r_k^2}{t_k} = \sum_{k=0}^{\infty} \left(\frac{r_{2k-1}^2}{t_{2k-1}} + \frac{r_{2k}^2}{t_{2k}} \right) \\
&\ge \sum_{k=0}^{\infty} \frac{r_{2k-1}^2 + r_{2k}^2}{2b_1(k + 1)} \ge \sum_{k=0}^{\infty} \frac{1}{2b_1\left(1 + \sigma_*^2\right)(k + 1)} \\
&= +\infty.
\end{aligned} \tag{3.35}$$

This contradiction shows that (3.16) is true.   □

3.2. *The Properties with the Weak Wolfe-Powell (WWP) Line Search*

The weak Wolfe-Powell line search is to find a step length α_k satisfying (3.2) and

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k, \tag{3.36}$$

where $\delta \in (0, 1/2)$, $\sigma \in (\delta, 1)$.

From the computation point of view, one of the well-known formulas for β_k is the PRP method. The global convergence with the exact line search had been proved by Polak and Ribière [4] when the objective function is convex. Powell [33] gave a counter example to show that there exist nonconvex functions on which the PRP method does not converge globally even if the exact line search is used. He suggested that β_k should not be less than zero. Considering this suggestion, under the assumption of the sufficient descent condition, Gilbert and Nocedal [27] proved that the modified PRP method $\beta_k^+ = \max\{0, \beta_k^{\mathrm{PRP}}\}$ is globally convergent with the WWP line search technique. For the new formula β_k^H, we know that it is always larger than zero. Then we can also get the global convergence of the hybrid method with the WWP line search.

Lemma 3.6 (see [31, Lemma 3.3.1]). *Let Assumption 3.1 hold and let the sequences $\{g_k, d_k\}$ be generated by Algorithm 2.1. The stepsize α_k is determined by (3.2) and (3.36). Suppose that (3.20) is true, and the sufficient descent condition (1.6) holds. Then we have $d_k \neq 0$ and*

$$\sum_{k=0}^{\infty} \|u_{k+1} - u_k\|^2 < \infty, \tag{3.37}$$

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

The following **Property 1** was introduced by Gilbert and Nocedal [27], which pertains to the PRP method under the sufficient descent condition. The WYL also has this property. Now we will prove that this **Property 1** pertains to the new method.

Property 1. Suppose that

$$0 < \gamma_1 \le \|g_k\| \le \gamma_2. \tag{3.38}$$

We say that the method has **Property 1** if for all k, there exists constants $b > 1$ and $\lambda > 0$ such that $|\beta_k| \le b$ and

$$\|s_k\| \le \lambda \Longrightarrow |\beta_k| \le \frac{1}{2b}. \tag{3.39}$$

Lemma 3.7. *Let Assumption 3.1 hold, let the sufficient descent condition (1.6) hold, and let the sequences $\{g_k, d_k\}$ be generated by Algorithm 2.1. Suppose that there exists a constant $M > 0$ such that $\|d_k\| \le M$ for all k. Then this method possesses Property 1.*

Proof. By (3.36), (1.6), (3.1), and (3.38), we have

$$\begin{aligned}
\gamma_1 c(1 - \sigma) \|g_k\| \\
\le c(1 - \sigma)\|g_k\|^2 \le -(1 - \sigma)g_k^T d_k \le \left(g_{k+1} - g_k\right)^T d_k \le L\|s_k\|\|d_k\| \le LM\|s_k\|.
\end{aligned} \tag{3.40}$$

Then we get

$$\|g_k\| \le \frac{LM}{\gamma_1 c(1 - \sigma)} \|s_k\|. \tag{3.41}$$

By (3.1) again, we obtain

$$\|g_{k+1}\| - \|g_k\| \leq \|g_{k+1} - g_k\| \leq L\|s_k\|. \tag{3.42}$$

Combing the above inequality and (3.41) implies that

$$\|g_{k+1}\| \leq \|g_k\| + L\|s_k\| \leq \left(\frac{LM}{\gamma_1 c(1-\sigma)} + L \right) \|s_k\|. \tag{3.43}$$

By (3.5) and (3.38), we have

$$\left| \beta_k^H \right| = \frac{(2\lambda_1 + \lambda_2)\|g_{k+1}\|^2}{\|g_k\|^2} \leq \frac{(2\lambda_1 + \lambda_2)\gamma_2^2}{\gamma_1^2}, \tag{3.44}$$

let $b = \max\{2, (2\lambda_1 + \lambda_2)\gamma_2^2/\gamma_1^2\} > 1$, $\lambda = \gamma_1^2/2b\gamma_2 L(2\lambda_1 + M/\gamma_1 c(1-\sigma) + 1)$. If $\|s_k\| \leq \lambda$, using (3.1), (3.38), (3.43), and the above equation, we obtain

$$
\begin{aligned}
\left| \beta_k^H \right| &\leq \|g_{k+1}\| \frac{\lambda_1 \|g_{k+1} - (\|g_{k+1}\|/\|g_k\|)g_k\| + \lambda_2 \|g_{k+1}\|}{\|g_k\|^2} \\
&= \gamma_2 \frac{\lambda_1 \|g_{k+1} - g_k + g_k - (\|g_{k+1}\|/\|g_k\|)g_k\| + \lambda_2 \|g_{k+1}\|}{\|g_k\|^2} \\
&\leq \gamma_2 \frac{\lambda_1 \|g_{k+1} - g_k\| + \lambda_1 |\|g_k\| - \|g_{k+1}\|| + \lambda_2 \|g_{k+1}\|}{\|g_k\|^2} \\
&\leq \gamma_2 \frac{2\lambda_1 L\|s_k\| + (LM/\gamma_1 c(1-\sigma) + L)\|s_k\|}{\|g_k\|^2} \\
&\leq \frac{\gamma_2 L(2\lambda_1 + M/\gamma_1 c(1-\sigma) + 1)}{\gamma_1^2} \lambda = \frac{1}{2b}.
\end{aligned}
\tag{3.45}
$$

Therefore, the conclusion of this lemma holds. □

Lemma 3.8 (see [31, Lemma 3.3.2]). *Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm 2.1 and the conditions in Lemma 3.7 hold. If $\beta_k^H \geq 0$ and has Property 1, then there exists $\lambda > 0$ such that for any $\Delta \in N$ and any index k_0, there is an index $k > k_0$ satisfying*

$$\left| \kappa_{k,\Delta}^\lambda \right| > \frac{\lambda}{2}, \tag{3.46}$$

where $\kappa_{k,\Delta}^\lambda = \{i \in N : k \leq i \leq k + \Delta - 1, \|s_i\| > \lambda\}$, N denotes the set of positive integers, $|\kappa_{k,\Delta}^\lambda|$ denotes the numbers of elements in $\kappa_{k,\Delta}^\lambda$.

Finally, by Lemmas 3.6 and 3.8, we present the global convergence theorem of Algorithm 2.1 with the WWP line search. Similar to [31, Theorem 3.3.3], it is not difficult to prove the result, so we omit it.

Theorem 3.9. *Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm 2.1 with the weak Wolfe-Powell line search and the conditions in Lemma 3.7 hold. Then* (3.16) *holds.*

4. Numerical Results

In this section, we report some results of the numerical experiments. It is well known that there exist many new conjugate gradient methods (see [1, 13–16, 18, 19, 29, 30]) which have good properties and good numerical performances. Since the given formula is the hybrid of the FR formula and the WYL formula, we only test Algorithm 2.1 under the WWP line search on problems in [34] with the given initial points and dimensions, and compare its performance with those of the FR [3] and the WYL [17] methods. The parameters are chosen as follows: $\delta = 0.1$, $\sigma = 0.2$, $\lambda_1 = \lambda_2 = 0.5$. The following *Himmeblau* stop rule is used as follows.

If $|f(x_k)| > e_1$, let $stop1 = |f(x_k) - f(x_{k+1})|/|f(x_k)|$; otherwise, let $stop1 = |f(x_k) - f(x_{k+1})|$.

If $\|g(x)\| < \varepsilon$ or $stop1 < e_2$ was satisfied, we will stop the program, where $e_1 = e_2 = 10^{-6}$. We also stop the program if the iteration number is more than one thousand. All codes were written in MATLAB and run on PC with 2.60 GHz CPU processor and 256 MB memory and Windows XP operation system. The detail numerical results are listed at http://210.36.18.9:8018/publication.asp?id=36990.

Dolan and Moré [35] 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:

$t_{p,s}$ = computing time (the number of function evaluations or others) 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

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s} : s \in S\}}. \tag{4.1}$$

Suppose that a parameter $r_M \geq r_{p,s}$ for all p, s is chosen, and $r_{p,s} = r_M$ if and only if solver s does not solve problem p.

The performance of solver s on any given problem might be of interest, but we would like to obtain an overall assessment of the performance of the solver, then they defined

$$\rho_s(t) = \frac{1}{n_p} \text{size}\{p \in P : r_{p,s} \leq t\}, \tag{4.2}$$

thus $\rho_s(t)$ was the probability for solver $s \in S$ that a performance ratio $r_{p,s}$ was within a factor $t \in \Re$ of the best possible ratio. Then function ρ_s was the (cumulative) distribution
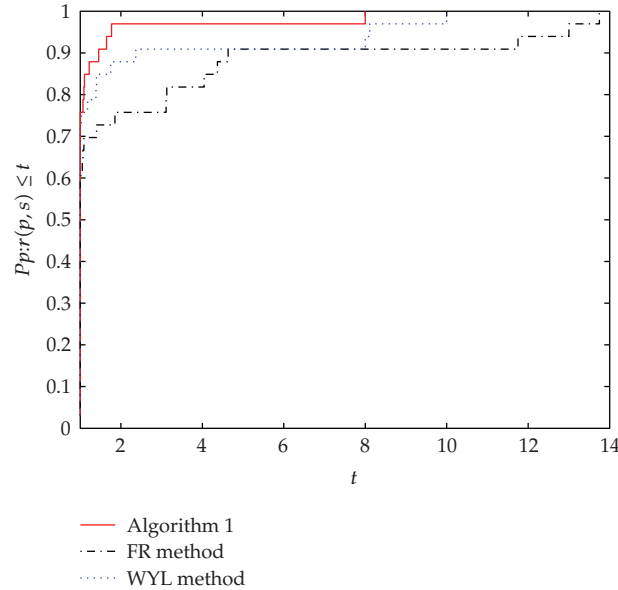
Figure 1: Performance profiles of these three methods (NI).

function for the performance ratio. The performance profile $\rho_s : \mathfrak{R} \mapsto [0,1]$ for a solver was a nondecreasing, piecewise constant function, continuous from the right at each breakpoint. The value of $\rho_s(1)$ was the probability that the solver would win over the rest of the solvers.

According to the above rules, we know that one solver whose performance profile plot is on top right will win over the rest of the solvers.

Figures 1-2 show that the performances of these methods are relative to the iteration number (NI) and the number of the function and gradient (NFN), where the "FR" denotes the FR formula with WWP rule, the "WYL" denotes the WYL formula with WWP rule, and Algorithm 2.1 denotes the new method with WWP rule, respectively.

From Figures 1-2, it is easy to see that Algorithm 2.1 is the best among the three methods, and the WYL method is much better than FR methods. Notice that the global convergence of the FR method with the WWP line search has not been established yet. In other words, the given method is competitive to the other two normal methods and the hybrid formula is notable.

5. Conclusions

This paper gives a hybrid conjugate gradient method for solving unconstrained optimization problems. Under the SWP line search, this method possesses the sufficient descent condition only with the parameter $\sigma < 1$. The global convergence with the SWP line search and the WWP line search is established for the nonconvex functions. Numerical results show that the given method is competitive to other two conjugate gradient methods.

For further research, we should study the new method with the nonmonotone line search technique. Moreover, more numerical experiments for large practical problems (such as the problems [36]) should be done, and the given method should be compared with other famous formulas in the future. How to choose the parameters λ_1 and λ_2 in the algorithm is another aspect of future investigation.
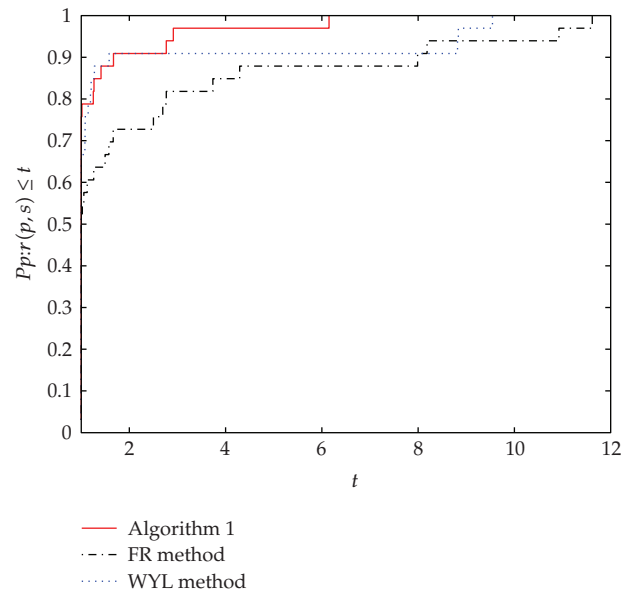
Figure 2: Performance profiles of these three methods (NFN).

Acknowledgments

References

[1] G. H. Yu, *Nonlinear self-scaling conjugate gradient methods for large-scale optimization problems*, Doctorial thesis, Sun Yat-Sen University, Guangzhou, China, 2007.

[2] G. Yuan and Z. Wei, "New line search methods for unconstrained optimization," *Journal of the Korean Statistical Society*, vol. 38, no. 1, pp. 29–39, 2009.

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

[4] E. Polak and G. Ribière, "Note sur la convergence de méthodes de directions conjuguées," *Revue Francaised Informatiquet de Recherche Operiionelle*, vol. 3, no. 16, pp. 35–43, 1969.

[5] 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.

[6] R. Fletcher, *Practical Methods of Optimization. Vol. 1: Unconstrained Optimization*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1997.

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

[8] 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.

[9] J. Nocedal, "Conjugate gradient methods and nonlinear optimization," in *Linear and Nonlinear Conjugate Gradient-Related Methods (Seattle, WA, 1995)*, L. Adams and J. L. Nazareth, Eds., pp. 9–23, SIAM, Philadelphia, Pa, USA, 1996.

[10] E. Polak, *Optimization: Algorithms and Consistent Approximations*, vol. 124 of *Applied Mathematical Sciences*, Springer, New York, NY, USA, 1997.

[11] Y. Yuan and W. Sun, *Theory and Methods of Optimization*, Science Press of China, Beijing, China, 1999.

[12] Y. Dai, "A nonmonotone conjugate gradient algorithm for unconstrained optimization," *Journal of Systems Science and Complexity*, vol. 15, no. 2, pp. 139–145, 2002.

[13] 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.

[14] 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.

[15] W. W. Hager and H. Zhang, "A survey of nonlinear conjugate gradient methods," *Pacific Journal of Optimization*, vol. 2, pp. 35–58, 2006.

[16] Z. Wei, G. Li, and L. Qi, "New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems," *Applied Mathematics and Computation*, vol. 179, no. 2, pp. 407–430, 2006.

[17] 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.

[18] 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.

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

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

[21] M. J. D. Powell, "Restart procedures for the conjugate gradient method," *Mathematical Programming*, vol. 12, no. 2, pp. 241–254, 1977.

[22] G. Zoutendijk, "Nonlinear programming, computational methods," in *Integer and Nonlinear Programming*, J. Abadie, Ed., pp. 37–86, North-Holland, Amsterdam, The Netherlands, 1970.

[23] 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.

[24] G. H. Liu, J. Y. Han, and H. X. Yin, "Global convergence of the Fletcher-Reeves algorithm with inexact linesearch," *Applied Mathematics: A Journal of Chinese Universities*, vol. 10, no. 1, pp. 75–82, 1995.

[25] 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.

[26] D. Touati-Ahmed and C. Storey, "Efficient hybrid conjugate gradient techniques," *Journal of Optimization Theory and Applications*, vol. 64, no. 2, pp. 379–397, 1990.

[27] 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.

[28] Y. F. Hu and C. Storey, "Global convergence result for conjugate gradient methods," *Journal of Optimization Theory and Applications*, vol. 71, no. 2, pp. 399–405, 1991.

[29] N. Andrei, "Another hybrid conjugate gradient algorithm for unconstrained optimization," *Numerical Algorithms*, vol. 47, no. 2, pp. 143–156, 2008.

[30] N. Andrei, "Hybrid conjugate gradient algorithm for unconstrained optimization," *Journal of Optimization Theory and Applications*, vol. 141, no. 2, pp. 249–264, 2009.

[31] Y. H. Dai and Y. Yuan, *Nonlinear Conjugate Gradient Methods*, Shanghai Scientific and Technical, Shanghai, China, 1998.

[32] Z. F. Li, J. Chen, and N. Y. Deng, "Convergence properties of conjugate gradient methods with Goldstein line search," *Journal of China Agricultural University*, vol. 1, no. 4, pp. 15–18, 1996.

[33] M. J. D. Powell, "Nonconvex minimization calculations and the conjugate gradient method," in *Numerical Analysis (Dundee, 1983)*, vol. 1066 of *Lecture Notes in Mathematics*, pp. 122–141, Springer, Berlin, Germany, 1984.

[34] J. J. Moré, B. S. Garbow, and K. E. Hillstrom, "Testing unconstrained optimization software," *ACM Transactions on Mathematical Software*, vol. 7, no. 1, pp. 17–41, 1981.

[35] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles," *Mathematical Programming*, vol. 91, no. 2, pp. 201–213, 2002.

[36] K. E. Bongartz, A. R. Conn, N. I. M. Gould, and P. L. Toint, "CUTE: constrained and unconstrained testing environments," *ACM Transactions on Mathematical Software*, vol. 21, pp. 123–160, 1995.