ISSN: 1017-060X (Print)

ISSN: 1735-8515 (Online)

Bulletin of the Iranian Mathematical Society

Vol. 43 (2017), No. 6, pp. 2067–2084

Title:

A new hybrid conjugate gradient algorithm for unconstrained optimization

Author(s):

X. Han, J. Zhang and J. Chen

Published by the Iranian Mathematical Society
http://bims.ims.ir

Bull. Iranian Math. Soc.
Vol. 43 (2017), No. 6, pp. 2067–2084
Online ISSN: 1735-8515

A NEW HYBRID CONJUGATE GRADIENT ALGORITHM FOR UNCONSTRAINED OPTIMIZATION

X. HAN, J. ZHANG AND J. CHEN*

(Communicated by Maziar Salahi)

Abstract. In this paper, a new hybrid conjugate gradient algorithm is proposed for solving unconstrained optimization problems. The new method possesses the sufficient descent property for any line search. Moreover, the global convergence of the proposed method is proved under the Wolfe line search. Numerical experiments are also presented to show the efficiency of the proposed algorithm, especially for solving highly dimensional problems.
Keywords: Unconstrained optimization problem, hybrid conjugate gradient algorithm, sufficient descent directions, global convergence.
MSC(2010): Primary: 90C46; Secondary: 90C06, 65K05.

1. Introduction

In this paper, we consider the following unconstrained optimization problem:

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

where $f:\mathbb{R}^n\to\mathbb{R}$ is a continuously differentiable function, $\mathbb{R}^n$ denotes an n-dimensional Euclidean space, and the gradient $\mathrm{g}(x) := \nabla f(x)$ is available. Conjugate gradient (CG) methods are efficient for solving unconstrained optimization problem (1.1). Generally, the CG method for (1.1) is of the form:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.2}$$

where $x_k\in\mathbb{R}^n$ is the current iterate and $\alpha_k>0$ is a step length determined by some suitable line search. $d_k\in\mathbb{R}^n$ is the search direction generated by the following:

$$d_k = \begin{cases} -\mathrm{g}_k, & \text{if } k=1, \\ -\mathrm{g}_k + \beta_k d_{k-1}, & \text{if } k\ge 2, \end{cases} \tag{1.3}$$

Article electronically published on 30 November, 2017.
Received: 1 July 2016, Accepted: 24 December 2016.
*Corresponding author.

where g_k denotes the gradient $g(x_k)$ and $\beta_k \in \mathbb{R}$ is a scalar parameter. Some well-known CG methods are the Hestenes-Stiefel (HS) method [14], the Dai-Yuan (DY) method [6], the Polak-Ribière-Polyak (PRP) method [20, 21], the Fletcher-Reeves (FR) method [10]. The parameters β_k in these methods are given respectively by

$$\begin{aligned} \beta_k^{HS} &= \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, \\ \beta_k^{PRP} &= \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \qquad \beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}. \end{aligned} \tag{1.4}$$

In (1.4), $y_{k-1} := g_k - g_{k-1}$ and $\|\cdot\|$ stands for the Euclidean norm.

In practice, the positive step length α_k of the CG method usually is computed by the standard Wolfe line search conditions:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{1.5a}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{1.5b}$$

or the strong Wolfe line search conditions, i.e., (1.5a) and $|g(x_k + \alpha_k d_k)^T d_k| \leq \sigma |g_k^T d_k|$, where $0 < \delta < \sigma < 1$.

As is well-known, the FR and DY algorithms possess good convergence properties, but their practical performances are not so good in general. Under the strong Wolfe line search, Al-Baali [1] proved the global convergence of the FR algorithm with the parameter $0 < \sigma < \frac{1}{2}$. Dai and Yuan [7] showed that the DY algorithm is global convergent under the standard Wolfe line search conditions (1.5a) and (1.5b). In contrary, the PRP and HS algorithms are generally regarded to be two of the most efficient CG algorithms in computational point of view, but their convergence properties are not outstanding. Hager and Zhang [13] conducted a detailed analysis to show that the PRP and HS algorithms can automatically adjust β_k to avoid jamming and their practical performance is outbalance the algorithms with $\|g_k\|^2$ in the numerator of β_k. In [23], Powell's example shows that if the function is not strongly convex, the PRP algorithm may not converge, even with an exact line search. By Powell's example, the HS algorithm may not converge for a general nonlinear function, with an exact line search. There has been much research on convergence properties and computational performance of these CG algorithms (see [2, 5, 13, 19, 23] etc).

In recent years, based on the original CG algorithms and their variants, several hybrid algorithms have been studied. Wei et al. [26] gave a variation of the FR method (shortly, VFR method), that is

$$\beta_k^{VFR} = \frac{\mu_1 \|g_k\|^2}{\mu_2 |g_k^T d_{k-1}| + \mu_3 \|g_{k-1}\|^2},$$

where $\mu_1 > 0$, $\mu_3 > 0$, $\mu_2 > \mu_1 + \epsilon_1$ and $\epsilon_1 > 0$ is an any given constant. The VFR algorithm always satisfies the sufficient condition $g_k^T d_k \leq -(\mu_2 -$

$\mu_1)/\mu_2\|g_k\|^2$, independently of choices of the parameter β_k and the line search. Simultaneously, Wei et al. [27] gave a variant of the PRP method (shortly, $VPRP$ method), that is

$$\beta_k^{VPRP} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{\|g_{k-1}\|^2}.$$

The $VPRP$ method inherits the excellent properties of the PRP method, such as good numerical performance and the [11, Property (∗)]. Furthermore, under the strong Wolfe line search with the parameter $0 < \sigma < \frac{1}{4}$, Huang et al. [16] showed that the $VPRP$ method possesses the sufficient descent property and the global convergence. Shortly afterwards, Yao et al. [25] extended this result to the HS method and introduced a new method which we call the YWH method with

$$\beta_k^{YWH} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{d_{k-1}^T y_{k-1}}.$$

Yao et al. [25] established that the YWH method can also generate sufficient descent directions under the strong Wolfe line search with greater parameter $\sigma \in (0, \frac{1}{3})$. Zhang [28] made a little modification to the $VPRP$ method and suggested a new method, that is

$$\beta_k^{NPRP} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} |g_k^T g_{k-1}|}{\|g_{k-1}\|^2},$$

which we call the $NPRP$ method. Zhang [28] showed that the $NPRP$ method possesses the sufficient descent property if $0 < \sigma < \frac{1}{2}$ in the strong Wolfe line search and converges globally for nonconvex optimization with the strong Wolfe line search. According to the idea of [26], Dai and Wen [8] proposed a new method, denoted it by $DPRP$ method, that is

$$\beta_k^{DPRP} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} |g_k^T g_{k-1}|}{\mu |g_k^T d_{k-1}| + \|g_{k-1}\|^2}, \tag{1.6}$$

where $\mu > 1$. Dai and Wen [8] proved the sufficient descent property for any line search and the global convergence of the $DPRP$ method with the standard Wolfe line search conditions (1.5a) and (1.5b). Recently, according to works [6, 25], Jiang et al. [18] introduced a hybrid method which we call the JHJ method with

$$\beta_k^{JHJ} = \frac{\|g_k\|^2 - \max\{0, \frac{\|g_k\|}{\|d_{k-1}\|} g_k^T d_{k-1}, \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}\}}{d_{k-1}^T y_{k-1}}. \tag{1.7}$$

Jiang et al. [18] proved the global convergence of the JHJ method with the standard Wolfe line search conditions (1.5a) and (1.5b). More recently, Jian et al. [17] took a little modification to the JHJ method, and proposed the N method as follows,

$$\beta_k^N = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|}\max\{0, g_k^T g_{k-1}\}}{\max\{\|g_{k-1}\|^2, d_{k-1}^T y_{k-1}\}}. \tag{1.8}$$

Jian et al. [17] proved the descent property for any line search and the global convergence of the N method with the standard Wolfe line search conditions (1.5a) and (1.5b).

Motivated by the works of Wei et al. [27], Dai and Wen [8] and Jian et al. [17], we introduce a new hybrid method (referred to as NHC method):

$$\beta_k^{NHC} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|}\max\{0, g_k^T g_{k-1}\}}{\max\{\max\{0, u g_k^T d_{k-1}\} + \|g_{k-1}\|^2, d_{k-1}^T y_{k-1}\}}, \tag{1.9}$$

where $u > 1$. It is easy to see that β_k^{NHC} is one of hybrids of β_k^N, β_k^{FR}, β_k^{DY}, β_k^{VFR}, β_k^{YWH}, β_k^{VPRP} and β_k^{DPRP}.

The remainder of this paper is organized as follows. In the next section, a new hybrid CG algorithm is introduced, and then we show the sufficient descent property of the new algorithm for any line search. In Section 3, some properties of the proposed algorithm are presented, and the global convergence of the new algorithm is also proved under the standard Wolfe line search. In Section 4, some numerical results are reported to show the efficiency and feasibility of the introduced algorithm. Finally, we give the conclusion.

2. A new hybrid conjugate gradient algorithm

In this section, based on (1.9), a new hybrid CG algorithm is proposed and the sufficient descent property of the introduced algorithm is also proved.

The framework of the proposed algorithm is given as follows:

Algorithm NHC

Step 1: Select a starting point $x_1 \in \mathbb{R}^n$, and accuracy tolerance $\epsilon > 0$. Compute $d_1 = -g_1$ and the initial guess $\alpha_1 = 1/\|g_1\|$. Let $k = 1$.

Step 2: If $\|g_k\| \le \epsilon$, then stop. Otherwise, skip to Step 3.

Step 3: Line search. Compute step length $\alpha_k > 0$ by the standard Wolfe line search conditions (1.5a) and (1.5b).

Step 4: Let $x_{k+1} = x_k + \alpha_k d_k$ and $g_{k+1} = g(x_{k+1})$. Compute β_{k+1}^{NHC}.

Step 5: Generate $d_{new} = -g_{k+1} + \beta_{k+1}^{NHC} d_k$. If the Powell [22] restart criterion

$$|g_{k+1}^T g_k| \ge 0.2\|g_{k+1}\|^2$$

holds, then let $d_{k+1} = -g_{k+1}$, otherwise define $d_{k+1} = d_{new}$. Compute the initial guess $\alpha_k = \alpha_{k-1}\|d_{k-1}\|/\|d_k\|$.

Step 6: Let $k := k + 1$ and return to Step 3.

It is well-known that the initial selection of the step length α_k weightily affects the practical behavior of the standard Wolfe line search. If f is bounded below along the search direction d_k, then there exists a step length α_k satisfying the standard Wolfe line search conditions (1.5a) and (1.5b). At the first iteration, the initial guess of the step length is given as $\alpha_1 = 1/\|g_1\|$. At the following iteration, the starting guess for the step length α_k in the standard Wolfe line search is computed as $\alpha_{k-1}\|d_{k-1}\|/\|d_k\|$. This selection is firstly introduced by Shanno and Phua in CONMIN [24].

The following theorem shows that the search direction d_k in Algorithm NHC satisfies the sufficient descent property without any line search. The characterization of d_k plays an extremely significant role in analyzing the global convergence property.

Theorem 2.1. *Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm NHC. Then, for all $k \geq 1$,*

$$g_k^T d_k \leq -c\|g_k\|^2,$$

where $c = 1 - \frac{1}{u}$ and $u > 1$.

Proof. For $k = 1$, it is easy to know that $g_1^T d_1 = -\|g_1\|^2$.

If $\beta_k^{NHC} = 0$ for $k > 1$, then from (1.3), $g_k^T d_k = -\|g_k\|^2$. Since $0 < c < 1$, the conclusion holds.

We next assume that $\beta_k^{NHC} \neq 0$. To prove $g_k^T d_k \leq -c\|g_k\|^2$ for $k > 1$, we divide the proof into eight cases:

(i) $g_k^T d_{k-1} \leq 0$, $d_{k-1}^T y_{k-1} \leq \|g_{k-1}\|^2$ and $g_k^T g_{k-1} > 0$. In this case, the denominator of (1.9) is equal to $\|g_{k-1}\|^2$, and the numerator of (1.8) is equal to $\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}$. Then, from (1.8), we have

$$\beta_k^{NHC} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{\|g_{k-1}\|^2} = \beta_k^{VPRP}.$$

It follows from (1.3) that

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{\|g_{k-1}\|^2} g_k^T d_{k-1} \\ &= -\|g_k\|^2 + \frac{\|g_k\|^2 (1 - \cos\langle g_k, g_{k-1}\rangle)}{\|g_{k-1}\|^2} g_k^T d_{k-1} \\ &\leq -\|g_k\|^2 \\ &< -(1 - \frac{1}{u})\|g_k\|^2. \end{aligned} \tag{2.1}$$

(ii) $g_k^T d_{k-1} \leq 0$, $d_{k-1}^T y_{k-1} \geq \|g_{k-1}\|^2$ and $g_k^T g_{k-1} \leq 0$. In this case, the denominator of (1.9) is equal to $d_{k-1}^T y_{k-1}$, and the numerator of (1.9) is equal to $\|g_k\|^2$. Then, from (1.9), we have

$$\beta_k^{NHC} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} = \beta_k^{DY}.$$

It follows from (1.3) that

$$g_k^T d_k = -\|g_k\|^2 + \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} g_k^T d_{k-1} \le -\|g_k\|^2 < -(1-\frac{1}{u})\|g_k\|^2. \tag{2.2}$$

(iii) $g_k^T d_{k-1} \le 0$, $d_{k-1}^T y_{k-1} \ge \|g_{k-1}\|^2$ and $g_k^T g_{k-1} > 0$. In this case, the denominator of (1.9) is equal to $d_{k-1}^T y_{k-1}$, and the numerator of (1.9) is equal to $\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}$. From (1.9), we have

$$\beta_k^{NHC} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{d_{k-1}^T y_{k-1}} = \beta_k^{YWH}.$$

Again, from (1.3), we deduce that

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{d_{k-1}^T y_{k-1}} g_k^T d_{k-1} \\ &= -\|g_k\|^2 + \frac{\|g_k\|^2(1-\cos\langle g_k, g_{k-1}\rangle)}{d_{k-1}^T y_{k-1}} g_k^T d_{k-1} \\ &\le -\|g_k\|^2 \\ &< -(1-\frac{1}{u})\|g_k\|^2. \end{aligned} \tag{2.3}$$

(iv) $g_k^T d_{k-1} \le 0$, $d_{k-1}^T y_{k-1} \le \|g_{k-1}\|^2$ and $g_k^T g_{k-1} \le 0$. In this case, the denominator of (1.9) is equal to $\|g_{k-1}\|^2$, and the numerator of (1.9) is equal to $\|g_k\|^2$. Then,

$$\beta_k^{NHC} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} = \beta_k^{FR}.$$

Hence, by (1.3), we obtain

$$g_k^T d_k = -\|g_k\|^2 + \frac{\|g_k\|^2}{\|g_{k-1}\|^2} g_k^T d_{k-1} \le -\|g_k\|^2 < -(1-\frac{1}{u})\|g_k\|^2. \tag{2.4}$$

(v) $g_k^T d_{k-1} > 0$, $d_{k-1}^T y_{k-1} \le u g_k^T d_{k-1} + \|g_{k-1}\|^2$ and $g_k^T g_{k-1} \le 0$. In this case, the denominator of (1.9) is equal to $u g_k^T d_{k-1} + \|g_{k-1}\|^2$, and the numerator of (1.9) is equal to $\|g_k\|^2$. Note that $\mu_1 = 1$, $\mu_2 = u > 1$ and $\mu_3 = 1$. Moreover, we have

$$\beta_k^{NHC} = \frac{\|g_k\|^2}{u g_k^T d_{k-1} + \|g_{k-1}\|^2} = \beta_k^{VFR}.$$

So, from (1.3), one has

$$
\begin{aligned}
g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2}{u g_k^T d_{k-1} + \|g_{k-1}\|^2} g_k^T d_{k-1} \\
&< -\|g_k\|^2 + \frac{\|g_k\|^2}{u g_k^T d_{k-1}} g_k^T d_{k-1} \\
&= -(1 - \frac{1}{u})\|g_k\|^2. \qquad (2.5)
\end{aligned}
$$

(vi) $g_k^T d_{k-1} > 0$, $d_{k-1}^T y_{k-1} \leq u g_k^T d_{k-1} + \|g_{k-1}\|^2$ and $g_k^T g_{k-1} > 0$. In this case, the denominator of (1.9) is equal to $u g_k^T d_{k-1} + \|g_{k-1}\|^2$, and the numerator of (1.9) is equal to $\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}$. Then, one has

$$
\beta_k^{NHC} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{u g_k^T d_{k-1} + \|g_{k-1}\|^2} = \beta_k^{DPRP}.
$$

Consequently, from (1.3), we obtain

$$
\begin{aligned}
g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{u g_k^T d_{k-1} + \|g_{k-1}\|^2} g_k^T d_{k-1} \\
&< -\|g_k\|^2 + \frac{\|g_k\|^2}{u g_k^T d_{k-1}} g_k^T d_{k-1} \\
&= -(1 - \frac{1}{u})\|g_k\|^2. \qquad (2.6)
\end{aligned}
$$

(vii) $g_k^T d_{k-1} > 0$, $d_{k-1}^T y_{k-1} \geq u g_k^T d_{k-1} + \|g_{k-1}\|^2$ and $g_k^T g_{k-1} > 0$. In this case, the denominator of (1.9) is equal to $d_{k-1}^T y_{k-1}$, and the numerator of (1.9) is equal to $\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}$. Again, from (1.9), we get

$$
\beta_k^{NHC} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{d_{k-1}^T y_{k-1}} = \beta_k^{YWH}.
$$

It follows from (1.3) that

$$
\begin{aligned}
g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{d_{k-1}^T y_{k-1}} g_k^T d_{k-1} \\
&< -\|g_k\|^2 + \frac{\|g_k\|^2}{u g_k^T d_{k-1}} g_k^T d_{k-1} = -(1 - \frac{1}{u})\|g_k\|^2. \qquad (2.7)
\end{aligned}
$$

(viii) $g_k^T d_{k-1} > 0$, $d_{k-1}^T y_{k-1} \geq u g_k^T d_{k-1} + \|g_{k-1}\|^2$ and $g_k^T g_{k-1} \leq 0$. In this case, the denominator of (1.9) is equal to $d_{k-1}^T y_{k-1}$, and the numerator of (1.9) is equal to $\|g_k\|^2$. Then,

$$\beta_k^{NHC} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} = \beta_k^{DY}.$$

It follows from (1.3) that

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} g_k^T d_{k-1} \\ &\leq -\|g_k\|^2 + \frac{\|g_k\|^2}{u g_k^T d_{k-1}} g_k^T d_{k-1} = -(1-\frac{1}{u})\|g_k\|^2. \end{aligned} \tag{2.8}$$

Hence, from (2.1)-(2.8), we obtain the desired result. □

Remark 2.2. It is easy to find that $0 \leq \beta_k^{NHC} \leq \beta_k^N$ by expressions of β_k^N and β_k^{NHC}. Furthermore, it is easy to gain descent of our algorithm by means of idea of the [17, Lemma 1]. But this idea is invalid to show the sufficient descent property of the proposed algorithm. From the proof process of Theorem 2.1, it clearly shows that the range of choice of scalar β_k is widen further and the sufficient descent property of our algorithm is also proved.

3. Convergence analysis

In this section, we analyze the global convergence property of Algorithm NHC. To do this, the following basic assumptions for the objective function are given.

Assumption 3.1

- **(a):** The level set $\mathcal{L} = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$ is bounded.
- **(b):** In some neighborhood $\mathcal{N}$ of $\mathcal{L}$, the gradient $g(x) = \nabla f(x)$ is Lipschitz, i.e., there exists a constant $L > 0$ such that $\|g(x) - g(y)\| \leq L\|x - y\|$, for all $x, y \in \mathcal{N}$.

The following well-known lemma is usually called the Zoutendijk condition [29].

Lemma 3.1. *Assume that Assumption 3 is satisfied. Consider any CG method of the form (1.2) and (1.3), where d_k is a descent direction and α_k satisfies the standard Wolfe line search conditions (1.5a) and (1.5b). Then,*

$$\sum_{k=1}^{+\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty. \tag{3.1}$$

From Theorem 2.1, it is easy to derive the following interesting property about the parameter β_k in (1.9).

Theorem 3.2. *Let β_k^{NHC} be defined by (1.9). Then*

$$0 \le \beta_k^{NHC} \le \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}, \ \forall\ k > 1. \tag{3.2}$$

Proof. We first show that the left inequality of (3.2) holds. By the definition of β_k^{NHC} in (1.9), we obtain

$$\begin{aligned}\beta_k^{NHC} &= \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|}\max\{0, g_k^T g_{k-1}\}}{\max\{\max\{0, u g_k^T d_{k-1}\} + \|g_{k-1}\|^2, d_{k-1}^T y_{k-1}\}} \\ &\ge \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|}\|g_k\|\|g_{k-1}\|}{\max\{u|g_k^T d_{k-1}| + \|g_{k-1}\|^2, d_{k-1}^T y_{k-1}\}} = 0.\end{aligned}$$

Let us now show that the right inequality of (3.2) holds. If $\beta_k^{NHC} = 0$ and $g_k \ne 0$, from Theorem 2.1, we have

$$\beta_k^{NHC} = 0 < \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}, \ \forall\ k > 1.$$

Assume that $\beta_k^{NHC} > 0$. To prove $\beta_k^{NHC} \le \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}$ for all $k > 1$, we consider eight cases: (i) $g_k^T d_{k-1} \le 0$, $d_{k-1}^T y_{k-1} \le \|g_{k-1}\|^2$ and $g_k^T g_{k-1} > 0$. In this case, $\beta_k^{NHC} = \beta_k^{VPRP}$. Then, we have

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{\|g_{k-1}\|^2} g_k^T d_{k-1} \\ &\le -\|g_k\|^2 + \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{\|g_{k-1}\|^2} (\|g_{k-1}\|^2 + g_{k-1}^T d_{k-1}) \\ &= -\frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}^T g_{k-1} + \beta_k^{NHC} g_{k-1}^T d_{k-1} \\ &< \beta_k^{NHC} g_{k-1}^T d_{k-1}. \end{aligned}$$

By Theorem 2.1, we obtain $\beta_k^{NHC} < \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}$.

(ii) $g_k^T d_{k-1} \le 0$, $d_{k-1}^T y_{k-1} \ge \|g_{k-1}\|^2$ and $g_k^T g_{k-1} \le 0$. In this case, $\beta_k^{NHC} = \beta_k^{DY}$. Then, we have

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} g_k^T d_{k-1} \\ &= -\|g_k\|^2 + \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} (d_{k-1}^T y_{k-1} + g_{k-1}^T d_{k-1}) \\ &= \beta_k^{NHC} g_{k-1}^T d_{k-1}. \end{aligned}$$

Therefore, we obtain $\beta_k^{NHC} = \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}$.

(iii) $g_k^T d_{k-1} \leq 0$, $d_{k-1}^T y_{k-1} \geq \|g_{k-1}\|^2$ and $g_k^T g_{k-1} > 0$. In this case, $\beta_k^{NHC} = \beta_k^{YWH}$. Then, we have

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{d_{k-1}^T y_{k-1}} g_k^T d_{k-1} \\ &= -\|g_k\|^2 + \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{d_{k-1}^T y_{k-1}} (d_{k-1}^T y_{k-1} + g_{k-1}^T d_{k-1}) \\ &= -\frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1} + \beta_k^{NHC} g_{k-1}^T d_{k-1} \\ &< \beta_k^{NHC} g_{k-1}^T d_{k-1}. \end{aligned}$$

It follows from Theorem 2.1 that $\beta_k^{NHC} < \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}$.

(iv) $g_k^T d_{k-1} \leq 0$, $d_{k-1}^T y_{k-1} \leq \|g_{k-1}\|^2$ and $g_k^T g_{k-1} \leq 0$. In this case, $\beta_k^{NHC} = \beta_k^{FR}$. Then, we have

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2}{\|g_{k-1}\|^2} g_k^T d_{k-1} \\ &\leq -\|g_k\|^2 + \frac{\|g_k\|^2}{\|g_{k-1}\|^2} (\|g_{k-1}\|^2 + g_{k-1}^T d_{k-1}) \\ &= \beta_k^{NHC} g_{k-1}^T d_{k-1}. \end{aligned}$$

Again, from Theorem 2.1, we have $\beta_k^{NHC} \leq \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}$.

(v) $g_k^T d_{k-1} > 0$, $d_{k-1}^T y_{k-1} \leq u g_k^T d_{k-1} + \|g_{k-1}\|^2$ and $g_k^T g_{k-1} \leq 0$. In this case, $\beta_k^{NHC} = \beta_k^{VFR}$. Then, we have

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2}{u g_k^T d_{k-1} + \|g_{k-1}\|^2} g_k^T d_{k-1} \\ &\leq -\|g_k\|^2 + \frac{\|g_k\|^2}{u g_k^T d_{k-1} + \|g_{k-1}\|^2} (u g_k^T d_{k-1} + \|g_{k-1}\|^2 + g_{k-1}^T d_{k-1}) \\ &= \beta_k^{NHC} g_{k-1}^T d_{k-1}. \end{aligned}$$

By Theorem 2.1, we obtain $\beta_k^{NHC} \leq \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}$.

(vi) $g_k^T d_{k-1} > 0$, $d_{k-1}^T y_{k-1} \leq u g_k^T d_{k-1} + \|g_{k-1}\|^2$ and $g_k^T g_{k-1} > 0$. In this case, $\beta_k^{NHC} = \beta_k^{DPRP}$. Then, we have

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{u g_k^T d_{k-1} + \|g_{k-1}\|^2} g_k^T d_{k-1} \\ &\leq -\|g_k\|^2 + \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{u g_k^T d_{k-1} + \|g_{k-1}\|^2} (u g_k^T d_{k-1} + \|g_{k-1}\|^2 + g_{k-1}^T d_{k-1}) \\ &= -\frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1} + \beta_k^{NHC} g_{k-1}^T d_{k-1} \\ &< \beta_k^{NHC} g_{k-1}^T d_{k-1}. \end{aligned}$$

From Theorem 2.1, we obtain $\beta_k^{NHC} < \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}$.

(vii)$g_k^T d_{k-1} > 0$, $d_{k-1}^T y_{k-1} \geq u g_k^T d_{k-1} + \|g_{k-1}\|^2$ and $g_k^T g_{k-1} > 0$. In this case, $\beta_k^{NHC} = \beta_k^{YWH}$. Then, we have

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{d_{k-1}^T y_{k-1}} g_k^T d_{k-1} \\ &= -\|g_k\|^2 + \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1}}{d_{k-1}^T y_{k-1}} (d_{k-1}^T y_{k-1} + g_{k-1}^T d_{k-1}) \\ &= -\frac{\|g_k\|}{\|g_{k-1}\|} g_k^T g_{k-1} + \beta_k^{NHC} g_{k-1}^T d_{k-1} \\ &< \beta_k^{NHC} g_{k-1}^T d_{k-1}. \end{aligned}$$

Consequently, one has $\beta_k^{NHC} < \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}$.

(viii) $g_k^T d_{k-1} > 0$, $d_{k-1}^T y_{k-1} \geq u g_k^T d_{k-1} + \|g_{k-1}\|^2$ and $g_k^T g_{k-1} \leq 0$. In this case, $\beta_k^{NHC} = \beta_k^{DY}$. Then, we have

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} g_k^T d_{k-1} \\ &= -\|g_k\|^2 + \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} (d_{k-1}^T y_{k-1} + g_{k-1}^T d_{k-1}) \\ &= \beta_k^{NHC} g_{k-1}^T d_{k-1}. \end{aligned}$$

Therefore, we get $\beta_k^{NHC} = \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}$.

Hence, based on the discussion above, the desired result holds. □

The following result shows the global convergence of Algorithm NHC.

Theorem 3.3. *Assume that Assumption 3 holds. Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm NHC. Then,*

$$\liminf_{k\to+\infty} \|g_k\| = 0. \tag{3.3}$$

Proof. Suppose that (3.3) fails. Then, there exists a constant $r > 0$ such that

$$\|g_k\| \geq r, \tag{3.4}$$

for all k. It follows from (1.3) that $d_k + g_k = \beta_k^{NHC} d_{k-1}$. This combining with (3.2) implies that

$$\begin{aligned} \|d_k\|^2 &= (\beta_k^{NHC})^2 \|d_{k-1}\|^2 - \|g_k\|^2 - 2g_k^T d_k \\ &\leq (\frac{g_k^T d_k}{g_{k-1}^T d_{k-1}})^2 \|d_{k-1}\|^2 - \|g_k\|^2 - 2g_k^T d_k. \end{aligned} \tag{3.5}$$

Dividing both sides of (3.5) by $(g_k^T d_k)^2$, we have

$$\begin{aligned} (\frac{\|d_k\|}{g_k^T d_k})^2 &\leq (\frac{\|d_{k-1}\|}{g_{k-1}^T d_{k-1}})^2 - (\frac{\|g_k\|}{g_k^T d_k})^2 - \frac{2}{g_k^T d_k} \\ &= (\frac{\|d_{k-1}\|}{g_{k-1}^T d_{k-1}})^2 - [(\frac{\|g_k\|}{g_k^T d_k})^2 + (\frac{1}{\|g_k\|})^2 + 2\frac{\|g_k\|}{g_k^T d_k}\frac{1}{\|g_k\|}] + \frac{1}{\|g_k\|^2} \\ &= (\frac{\|d_{k-1}\|}{g_{k-1}^T d_{k-1}})^2 + \frac{1}{\|g_k\|^2} - (\frac{\|g_k\|}{g_k^T d_k} + \frac{1}{\|g_k\|})^2 \\ &\leq (\frac{\|d_{k-1}\|}{g_{k-1}^T d_{k-1}})^2 + \frac{1}{\|g_k\|^2}. \end{aligned} \tag{3.6}$$

Since $\|g_k\| \geq r$ and $(\frac{\|d_1\|}{g_1^T d_1})^2 = \frac{1}{\|g_1\|^2}$, by a recurrence of (3.6), we obtain

$$\begin{aligned} (\frac{\|d_k\|}{g_k^T d_k})^2 &\leq (\frac{\|d_{k-1}\|}{g_{k-1}^T d_{k-1}})^2 + \frac{1}{\|g_k\|^2} \\ &\leq (\frac{\|d_{k-2}\|}{g_{k-2}^T d_{k-2}})^2 + \frac{1}{\|g_k\|^2} + \frac{1}{\|g_{k-1}\|^2} \\ &\leq \cdots \leq \sum_{j=1}^{k} \frac{1}{\|g_j\|^2} \leq \frac{k}{r^2}. \end{aligned}$$

Hence, $(\frac{g_k^T d_k}{\|d_k\|})^2 \geq \frac{r^2}{k}$, this evidently implies that $\sum_{k=1}^{+\infty}(\frac{g_k^T d_k}{\|d_k\|})^2 = +\infty$, which contradicts (3.1). Therefore, (3.4) does not hold and the conclusion (3.3) is proved. □

4. Numerical experiments

In this section, we present some numerical results to show the efficiency and feasibility of the proposed method. All codes were written in Matlab 7.10 and run on a DELL with 1.60 GHz CPU processor, 4.00 GB RAM memory, and Windows 10 operation system.

In order to examine the numerical performance of our algorithm in practice, the following five CG algorithms are tested and compared:

NHC: with $\beta_k = \beta_k^{NHC}$ defined by this paper, i.e., formula (1.9),

N: with $\beta_k = \beta_k^{N}$ defined by Jian et al. [17], i.e., formula (1.8),
JHJ: with $\beta_k = \beta_k^{JHJ}$ defined by Jiang et al. [18], i.e., formula (1.7),
DPRP: with $\beta_k = \beta_k^{DPRP}$ defined by Dai and Wen [8], i.e., formula (1.6),
HuS : with $\beta_k = \max\{0, \min\{\beta_k^{FR}, \beta_k^{PRP}\}\}$ defined by Hu and Storey [15].

The above methods are tested by taking advantage of 71 unconstrained optimization test problems from [3]. In all CG algorithms, the standard Wolfe line search conditions (1.5a) and (1.5b) are implemented with $\sigma = 0.9$, $\delta = 10^{-4}$. In $DPRP$ and NHC we set $\mu = u = 1.1$. The initial guess of the step length at the first iteration is $\alpha_1 = 1/\|g_1\|$. At the following iteration, the starting guess for the step length α_k is generated as $\alpha_{k-1}\|d_{k-1}\|/\|d_k\|$. This selection is firstly introduced by Shanno and Phua in CONMIN [24] and is proved to be one of the best selections of the initial guess of the step length. Stop the program if criterion $\|g_k\|_\infty \leq \epsilon = 10^{-7}$ is satisfied, where $\|\cdot\|_\infty$ is the maximum absolute component of a vector. In addition, we declare failure if the criterion does not hold after 1200 seconds.

In order to make comprehensive comparisons according to Time, Iter, Fn and Gnx [1], respectively, we use the performance profile introduced by Dolan and Moré [9] to evaluate and compare the performance. Performance profile gives, for every $t \geq 1$, the proportion $p_s(t)$ of the best problems that each considered algorithmic variant has a performance within a factor of t of the best. The left axis of the figure gives the percentage of the test problems for which an algorithm is the fastest; the right axis of the figure gives the percentage of the test problems that are successfully solved by each of the algorithms, which is a measure of an algorithm's robustness. The top curve is the algorithm that solved the most problems in a time that is within a factor t of the best time.

From Figures 1 to 4, we can see that our algorithm possesses evident advantages based on Time and Iter. The main reason is that the standard Wolfe line

[1]For convenience of our discussion, some illustrations are given as follows:

Time : the computing time of CPU for computing a problem (Units: second);
Iter : the number of iterations;
Fn : the number of function evaluations;
Gn : the number of gradient evaluations.

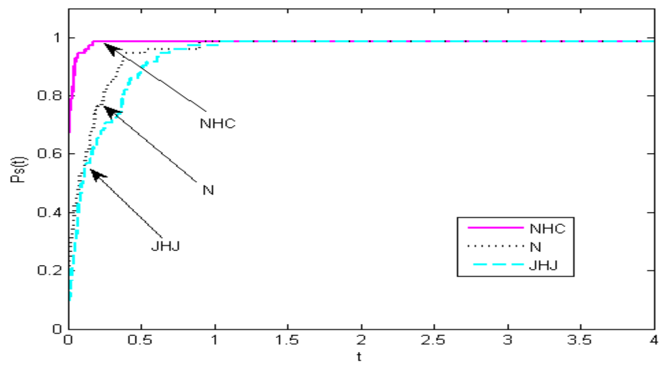

FIGURE 1. Performance profiles on Time (NHC versus N and JHJ)

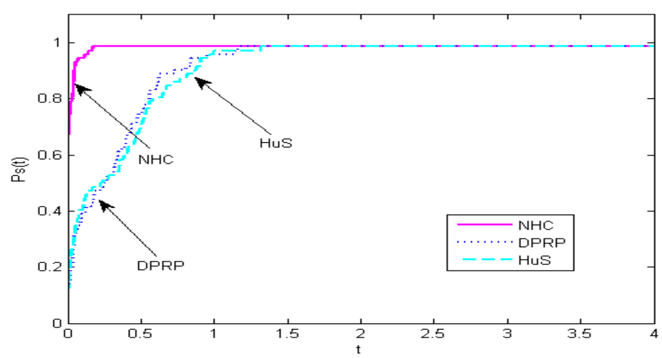

FIGURE 2. Performance profiles on Time (NHC versus $DPRP$ and HuS)

search at each iteration can be more easily satisfied with the sufficient descent property. In addition, Figures 5 and 6 show that the introduced algorithm is also competitive according to $Fn + 3Gn$ by Hager and Zhang [12]. The reason may be that the parameters β_k in these algorithms possess the different structure. From the above analysis, the NHC algorithm is preferable and promising.

5. **Conclusion**

In this paper, a new hybrid CG algorithm has been proposed for solving unconstrained optimization problems. The proposed algorithm always generates a sufficient descent direction at each iteration unrelated to the line search

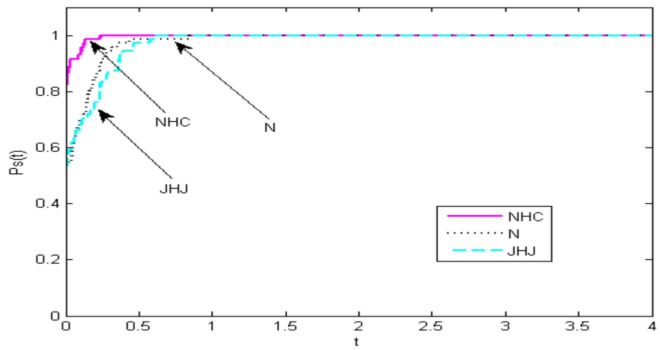

FIGURE 3. Performance profiles on Iter (NHC versus N and JHJ)

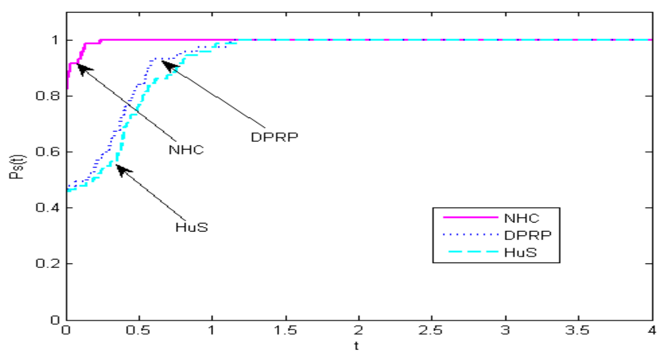

FIGURE 4. Performance profiles on Iter (NHC versus $DPRP$ and HuS)

strategy. Furthermore, the global convergence of our algorithm is proved under the standard Wolfe line search. Finally, a mass of numerical experiments are done, which show that the proposed algorithm is effective and feasible.

Acknowledgements

The authors would like to express their deeply gratitude to the editor and the anonymous referees for their valuable comments and suggestions, which enhance this paper greatly. Additionally, the first author would like to appreciate assistant professor Chaofeng Zhang from Sichuan University of Arts and Science for his valuable suggestions. This research was partially supported by the National Natural Science Foundation of China (Nos:11401487, 11401058), the Basic and Advanced Research Project of Chongqing (Nos: cstc2016jcyjA0239,

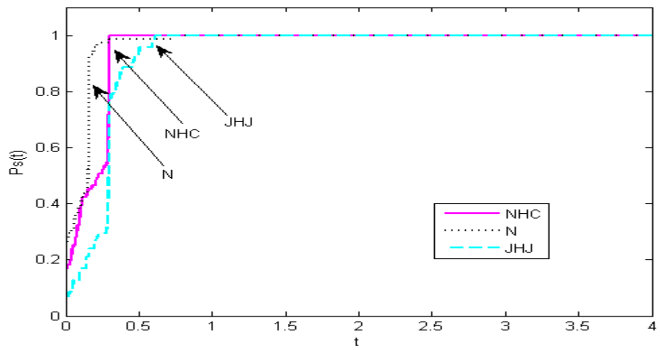

FIGURE 5. Performance profiles on Fn (NHC versus N and JHJ)

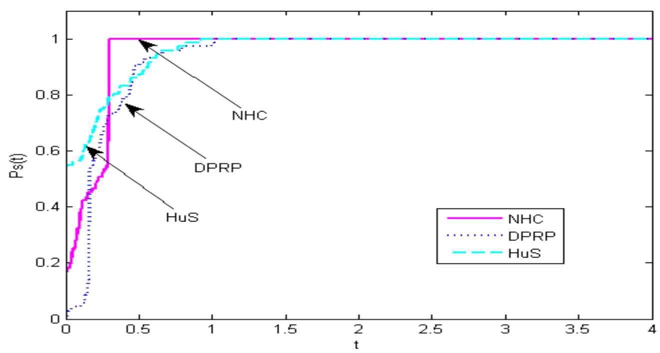

FIGURE 6. Performance profiles on Fn (NHC versus $DPRP$ and HuS)

cstc2015jcyjB0078) and the Fundamental Research Funds for the Central Universities.

REFERENCES

[1] M. Al-Baali, Descent property and global convergence of the Fletcher-Reeves method with inexact line search, *IMA J. Numer. Anal.* **5** (1985) 121–124.

[2] N. Andrei, Numerical comparison of conjugate gradient algorithms for unconstrained optimization, *Stud. Inform. Control.* **16** (2007) 333–352.

[3] N. Andrei, An unconstrained optimization test functions collection, *Adv. Model. Optim.* **10** (2008) 147–161.

[4] J. Chen, Q. Ansari, Y. Liu and J.C. Yao, A proximal point algorithm based on decomposition method for cone constrained multiobjective optimization problems, *Comput. Optim. Appl.* **65** (2016) 289–308.

[5] Y.H. Dai and Q. Ni, Testing different conjugate gradient methods for large-scale unconstrained optimization, *J. Comput. Math.* **22** (2003) 311–320.

[6] Y.H. Dai and Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, *SIAM J. Optim.* **10** (1999) 177–182.

[7] Y.H. Dai and Y. Yuan, An efficient hybrid conjugate gradient method for unconstrained optimization, *Ann. Oper. Res.* **103** (2001) 33–47.

[8] Z. Dai and F. Wen, Another improved Wei-Yao-Liu nonlinear conjugate gradient method with sufficient descent property, *Appl. Math. Comput.* **218** (2012) 7421–7430.

[9] E.D. Dolan and J.J. Moré, Benchmarking optimization software with performance profiles, *Math. Program.* **91** (2002) 201–213.

[10] R. Fletcher and C. Reeves, Function minimization by conjugate gradients, *Comput. J.* **7** (1964) 149–154.

[11] J.C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM J. Optim.* **2** (1992) 21–42.

[12] W.W. Hager and H. Zhang, Algorithm 851: CG_Descent, a conjugate gradient method with guaranteed descent, *ACM Trans. Math. Softw.* **32** (2006) 113–137.

[13] W.W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, *Pac. J. Optim.* **2** (2006) 35–58.

[14] M.R. Hestenes and E.L. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. Natl. Bur. Stand. Sec. B* **49** (1952) 409–432.

[15] Y.F. Hu and C. Storey, Global convergence result for conjugate gradient methods, *J. Optim. Theory Appl.* **71** (1991) 399–405.

[16] 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, *Appl. Math. Comput.* **189** (2007) 1241–1245.

[17] J. Jian, L. Han and X. Jiang, A hybrid conjugate gradient method with descent property for unconstrained optimization, *Appl. Math. Model.* **39** (2015) 1281–1290.

[18] X. Jiang, L. Han and J. Jian, A globally convergent mixed conjugate gradient method with Wolfe line search, *Math. Numer. Sin.* **34** (2012) 103–112.

[19] J. Nocedal and S.J. Wright, Numerical Optimization, Springer, New York, 2006.

[20] E. Polak and G. Ribière, Note sur la convergence de directions conjugées, *Rev. Fr. Inf. Rech. Oper.* **16** (1969) 35–43.

[21] B.T. Polyak, The conjugate gradient method in extremal problems, *USSR Comput. Math. Math. Phys.* **9** (1969) 94–112.

[22] M.J.D. Powell, Restart procedures for the conjugate gradient method, *Math. Program.* **12** (1977) 241–254.

[23] M.J.D. Powell, Convergence properties of algorithms for nonlinear optimization, *SIAM Rev.* **28** (1986) 487–500.

[24] D.F. Shanno and K.H. Phua, Algorithm 500. Minimization of unconstrained multivariate functions, *ACM Trans. Math. Softw.* **2** (1976) 87–94.

[25] S. Yao, Z. Wei and H. Huang, A note about WYL's conjugate gradient method and its application, *Appl. Math. Comput.* **191** (2007) 381–388.

[26] Z. Wei, G. Li and L. Qi, New nonliear conjugate gradient formulas for large-scale unconstrained optimization problems, *Appl. Math. Comput.* **179** (2006) 407–430.

[27] Z. Wei, S. Yao and L. Liu, The convergence properties of some new conjugate gradient methods, *Appl. Math. Comput.* **183** (2006) 1341–1350.

[28] L. Zhang, An improved Wei-Yao-Liu nonlinear conjugate gradient method for optimization computation, *Appl. Math. Comput.* **215** (2009) 2269–2274.

[29] G. Zoutendijk, Nonlinear programming computational methods, *Integer and Nonlinear Programming* in: J. Abadie (Ed.), Integer and Nonlinear Programming, North-Holland, Amsterdam, 1970, pp. 37–86.

(Xin Han) School of Mathematics and Statistics, Southwest University, Chongqing 400715, P.R. China.
E-mail address: hanmath@163.com

(Junrong Zhang) School of Mathematics and Statistics, Southwest University, Chongqing 400715, P.R. China.
E-mail address: zhangjr@swu.edu.cn

(Jiawei Chen) School of Mathematics and Statistics, Southwest University, Chongqing 400715, P.R. China.
E-mail address: J.W.Chen713@163.com