



# A new family of conjugate gradient methods<sup>☆</sup>

Zhen-Jun Shi<sup>a,b,\*</sup>, Jinhua Guo<sup>b</sup>

<sup>a</sup> College of Operations Research and Management, Qufu Normal University, Rizhao, Shandong 276826, PR China

<sup>b</sup> Department of Computer and Information Science, University of Michigan, Dearborn, MI 48128-1491, USA

## ARTICLE INFO

### Article history:

Received 28 December 2007

### MSC:

90C30

65K05

49M37

### Keywords:

Unconstrained optimization

Conjugate gradient method

Global convergence

## ABSTRACT

In this paper we develop a new class of conjugate gradient methods for unconstrained optimization problems. A new nonmonotone line search technique is proposed to guarantee the global convergence of these conjugate gradient methods under some mild conditions. In particular, Polak–Ribière–Polyak and Liu–Storey conjugate gradient methods are special cases of the new class of conjugate gradient methods. By estimating the local Lipschitz constant of the derivative of objective functions, we can find an adequate step size and substantially decrease the function evaluations at each iteration. Numerical results show that these new conjugate gradient methods are effective in minimizing large-scale non-convex non-quadratic functions.

© 2008 Elsevier B.V. All rights reserved.

## 1. Introduction

Consider an unconstrained minimization problem

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

where  $R^n$  denotes an  $n$ -dimensional Euclidean space and  $f : R^n \rightarrow R^1$  is a continuously differentiable function.

Many approaches to solving (1) are iterative methods, such as line search and trust region methods. For convenience, if  $x_k$  is the current iterate and  $x^*$  is a minimal solution or stationary point of (1) then we denote  $f(x_k)$  by  $f_k$ ,  $\nabla f(x_k)$  by  $g_k$ ,  $\nabla^2 f(x_k)$  by  $G_k$  and  $f(x^*)$  by  $f^*$ , respectively. If  $G_k$  is available and invertible then  $d_k = -G_k^{-1}g_k$  leads to the Newton method and  $d_k = -g_k$  results in the steepest descent method (e.g. [13,23,26]).

Line search methods have the form

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \dots, \quad (2)$$

where  $d_k$  is a search direction of  $f(x)$  at the current iterate  $x_k$  and  $\alpha_k$  is a step size. The search direction  $d_k$  is generally required to satisfy

$$g_k^T d_k < 0, \quad (3)$$

which guarantees that  $d_k$  is a descent direction of  $f(x)$  at  $x_k$  [13,23,43]. In order to guarantee the global convergence, we sometimes require  $d_k$  to satisfy the sufficient descent condition

$$g_k^T d_k \leq -c \|g_k\|^2, \quad (4)$$

<sup>☆</sup> The work was supported in part by NSF CNS-0521142, USA.

\* Corresponding author at: Department of Computer and Information Science, University of Michigan, Dearborn, MI 48128-1491, USA.

E-mail addresses: [zjshi@umd.umich.edu](mailto:zjshi@umd.umich.edu) (Z.-J. Shi), [jinhua@umd.umich.edu](mailto:jinhua@umd.umich.edu) (J. Guo).

where  $c > 0$  is a constant. Moreover, the angle property is often used in proving the global convergence of related line search methods, that is

$$\cos\langle -g_k, d_k \rangle = -\frac{g_k^T d_k}{\|g_k\| \cdot \|d_k\|} \geq \tau,$$

where  $1 \geq \tau > 0$ .

After the descent direction  $d_k$  is determined at the  $k$ th iteration, we should seek a step size along the descent direction and complete one iteration.

There are many approaches to finding an available step size. Among them, the exact line search is too difficult or too time-consuming to carry out and some inexact line searches are sometimes useful and powerful in practical computation, such as Armijo line search [1], Goldstein and Wolfe line searches [12,26,43]. The Armijo line search is commonly used and easy to implement in practical computation. The step size can also be defined by nonmonotone line search procedure (e.g. [6,17–19,29,33,40–42,45]). We describe the Armijo line search [1] and nonmonotone Armijo line search [33] as follows.

*Armijo line search.* Let  $s > 0$  be a constant,  $\rho \in (0, 1)$  and  $\mu \in (0, 1)$ . Choose  $\alpha_k$  to be the largest  $\alpha$  in  $\{s, s\rho, s\rho^2, \dots\}$  such that

$$f_k - f(x_k + \alpha d_k) \geq -\alpha \mu g_k^T d_k.$$

*Nonmonotone Armijo line search.* Given a nonnegative integer  $m$ , the index  $m(k)$  is defined by

$$m(0) = 0, \quad 0 \leq m(k+1) \leq \min(m(k) + 1, m). \quad (5)$$

Let  $s > 0$  be a constant,  $\rho \in (0, 1)$  and  $\mu \in (0, 1)$ . Choose  $\alpha_k$  to be the largest  $\alpha$  in  $\{s, s\rho, s\rho^2, \dots\}$  such that

$$\max_{0 \leq j \leq m(k)} [f_{k-j}] - f(x_k + \alpha d_k) \geq -\alpha \mu g_k^T d_k.$$

The shortcoming of the above Armijo-type line searches is how to choose the initial step size  $s$ . If  $s$  is too large then the procedure needs to call several more function evaluations. If  $s$  is too small then the efficiency of the related algorithm will be decreased. Therefore, we should choose an adequate initial step size at each iteration so as to find a suitable step size  $\alpha_k$  easily.

The conjugate gradient method is an important line search method that has the form (2) with

$$d_k = \begin{cases} -g_k, & \text{if } k = 0; \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geq 1, \end{cases} \quad (6)$$

where  $\beta_k$  can be defined by

$$\begin{aligned} \beta_k^{FR} &= \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, & \beta_k^{PRP} &= \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}, \\ \beta_k^{HS} &= \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1})}, & \beta_k^{LS} &= -\frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T g_{k-1}}, \\ \beta_k^{CD} &= -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}, & \beta_k^{DY} &= \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}}. \end{aligned}$$

or by other formulae (e.g. see [12,21,32]). The corresponding method is respectively called FR (Fletcher–Reeves [13]), PRP (Polak–Ribière–Polyak [27,28]), HS (Hestenes–Stiefel [23]), LS (Liu–Storey [25]), CD (Conjugate Descent [12]) and DY (Dai–Yuan [9]) conjugate gradient method.

The conjugate gradient method is an approach for solving large scale minimization problems due to its decreased storage requirements and simple computation (e.g. [4,7,12,13,30,31]). This method was motivated by Hestenes and Stiefel in solving symmetric positive definite linear equations [23] and developed by Fletcher and Reeves (e.g. [8,9,13]) in solving unconstrained minimization problems. The conjugate gradient methods have wide applications in many fields, such as control science, engineering, management science and operations research, etc., [12,43].

Although the above mentioned conjugate gradient algorithms are equivalent to each other for minimizing strong convex quadratic functions under exact line search, they have different performance when they are used to minimize non-quadratic functions or when some inexact line searches are used in these algorithms. For non-quadratic objective functions, the global convergence of FR method was proved when the exact line search or strong Wolfe line search [2,8] was used. The PRP method has no global convergence under some traditional line searches. Some convergent versions were proposed by using some new complicated line searches, or through restricting the parameter  $\beta_k$  to a nonnegative number [16,20,34,36,35]. The CD method and DY method were proved to have global convergence under strong Wolfe line search [9,12]. We can find some splendid literature on conjugate gradient method, e.g. [4,5,7,10,11,14,15,24,25,30,31,44]. However, to the best of our knowledge, the global convergence of PRP, LS and HS methods has not been established under all mentioned line searches. The main reason is that many conjugate gradient methods cannot guarantee the descent of objective function values at each iteration [21]. Thereby, we should seek some new line search approaches to overcome this drawback.

In this paper we propose a new nonmonotone line search in which an appropriate initial step size  $s$  is defined and varies at each iteration. The new nonmonotone line search enables us to find a suitable step size  $\alpha_k$  easily at each iteration and guarantees the global convergence of many conjugate gradient methods under some mild conditions. The global convergence and linear convergence rates are analyzed and numerical results show that the new class of conjugate gradient methods with the new nonmonotone line search are more effective than other similar methods in solving large scale minimization problems. Numerical results also show that the new nonmonotone line search can substantially decrease the function evaluations at each iteration.

The rest of this paper is organized as follows. In the next section we introduce a new nonmonotone Armijo line search and establish a new class of conjugate gradient methods. In Sections 3 and 4 the global convergence and linear convergence rate are analyzed respectively. Some numerical results are reported in Section 5 and conclusion remarks are given in Section 6.

## 2. New nonmonotone line search

We first assume that

(H1). The objective function  $f(x)$  is continuously differentiable and has a lower bound on  $R^n$ .

(H2). The gradient  $g(x) = \nabla f(x)$  of  $f(x)$  is Lipschitz continuous on an open convex set  $B$  that contains the level set  $L(x_0) = \{x \in R^n | f(x) \leq f(x_0)\}$  with  $x_0$  given, i.e., there exists an  $L > 0$  such that

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

Since  $L$  is usually not known a priori in practical computation but it plays an important rule in algorithm design, we need to estimate it for the new nonmonotone line search. Later on, we should discuss the problem and present some approaches for estimating  $L$ . Recently, some approaches for estimating  $L$  were proposed [37,38]. If  $k \geq 1$  then we set  $\delta_{k-1} = x_k - x_{k-1}$  and  $y_{k-1} = g_k - g_{k-1}$  and obtained the following three estimation formulae

$$L \simeq \frac{\|y_{k-1}\|}{\|\delta_{k-1}\|}, \quad (7)$$

$$L \simeq \frac{\|y_{k-1}\|^2}{|\delta_{k-1}^T y_{k-1}|}, \quad (8)$$

$$L \simeq \frac{|\delta_{k-1}^T y_{k-1}|}{\|\delta_{k-1}\|^2}. \quad (9)$$

In fact, if  $L$  is a Lipschitz constant then any  $L'$  greater than  $L$  is also a Lipschitz constant, which allows us to find a large Lipschitz constant. However, very large Lipschitz constant can lead to a very small step size and makes conjugate gradient methods with the new nonmonotone line search converge very slowly. Therefore, we should seek Lipschitz constants that are as small as possible in practical computation.

In the  $k$ th iteration we take respectively the approximated Lipschitz constant as

$$L_k = \max \left( L_0, \frac{\|y_{k-1}\|}{\|\delta_{k-1}\|} \right), \quad (10)$$

$$L_k = \max \left( L_0, \min \left( \frac{\|y_{k-1}\|^2}{|\delta_{k-1}^T y_{k-1}|}, M'_0 \right) \right), \quad (11)$$

$$L_k = \max \left( L_0, \frac{|\delta_{k-1}^T y_{k-1}|}{\|\delta_{k-1}\|^2} \right) \quad (12)$$

with  $L_0 > 0$  and  $M'_0$  being a large positive number. The corresponding conjugate gradient methods are denoted by CG1, CG2 and CG3, respectively, in which the following new nonmonotone line search is used in practical computation. Their global convergence and convergence rate will be given in the subsequent section.

**New nonmonotone line search.** Given  $\mu \in (0, \frac{1}{2})$ ,  $\rho \in (0, 1)$ ,  $c \in (\frac{1}{2}, 1)$  and  $u \in [0, 1]$ . Set  $s_k = \frac{1-c}{L_k} \cdot \frac{(1-u)\|g_k\|^2 - u g_k^T d_k}{\|d_k\|^2}$  and  $\alpha_k$  is the largest  $\alpha$  in  $\{s_k, s_k \rho, s_k \rho^2, \dots\}$  such that

$$\max_{0 \leq j \leq m(k)} [f_{k-j}] - f(x_k + \alpha d_k) \geq -\alpha \mu \left[ g_k^T d_k + \frac{1}{2} \alpha L_k \|d_k\|^2 \right],$$

and

$$g(x_k + \alpha d_k)^T d(x_k + \alpha d_k) \leq -c \|g(x_k + \alpha d_k)\|^2,$$

where

$$d(x_k + \alpha d_k) = -g(x_k + \alpha d_k) + \frac{g(x_k + \alpha d_k)^T (g(x_k + \alpha d_k) - g_k)}{(1-u)\|g_k\|^2 - ug_k^T d_k} d_k,$$

$m(k)$  is defined by (5) and  $L_k$  is estimated by (10), (11) or (12), respectively.

**Algorithm (A).**

**Step 0.** Choose  $x_0 \in R^n$ ,  $u \in [0, 1]$  and set  $d_0 = -g_0$ ,  $k := 0$ .

**Step 1.** If  $\|g_k\| = 0$  then stop else go to Step 2;

**Step 2.** Set  $x_{k+1} = x_k + \alpha_k d_k$  where  $d_k$  is defined by (6) with

$$\beta_k = \frac{g_k^T (g_k - g_{k-1})}{(1-u)\|g_{k-1}\|^2 - ug_{k-1}^T d_{k-1}}$$

and  $\alpha_k$  is defined by the new nonmonotone line search.

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

It is interesting that if  $u = 0$  then Algorithm (A) reduces to PRP method and if  $u = 1$  then Algorithm (A) reduces to LS method. This new nonmonotone line search is motivated based on [39].

Some simple properties of the above algorithm are given as follows.

**Lemma 2.1.** Assume that (H1) and (H2) hold, and Algorithm (A) with the new nonmonotone line search generates an infinite sequence  $\{x_k\}$ . Then, there exist  $m_0 > 0$  and  $M_0 > 0$  such that

$$m_0 \leq L_k \leq M_0. \quad (13)$$

**Proof.** Obviously,  $L_k \geq L_0$ , and we can take  $m_0 = L_0$ . For (10), by (H2) we have

$$L_k = \max \left( L_0, \frac{\|y_{k-1}\|}{\|\delta_{k-1}\|} \right) \leq \max(L_0, L).$$

For (11), we have

$$L_k = \max \left( L_0, \min \left( \frac{\|y_{k-1}\|^2}{\delta_{k-1}^T y_{k-1}}, M'_0 \right) \right) \leq \max(L_0, M'_0).$$

For (12), by using Cauchy–Schwartz inequality, we have

$$L_k = \max \left( L_0, \frac{\delta_{k-1}^T y_{k-1}}{\|\delta_{k-1}\|^2} \right) \leq \max(L_0, L).$$

By letting  $M_0 = \max(L_0, L, M'_0)$ , we complete the proof.  $\square$

**Lemma 2.2.** Assume that (H1) and (H2) hold, and Algorithm (A) with the new nonmonotone line search generates an infinite sequence  $\{x_k\}$ . If  $g_k^T d_k < 0$  and

$$\alpha_k \leq \frac{1-c}{L} \cdot \frac{(1-u)\|g_k\|^2 - ug_k^T d_k}{\|d_k\|^2},$$

then

$$g_{k+1}^T d_{k+1} \leq -c\|g_{k+1}\|^2.$$

**Proof.** By the two inequalities and Cauchy–Schwartz inequality, we have

$$\begin{aligned} -(1-c)[(1-u)\|g_k\|^2 - ug_k^T d_k] &\geq \alpha_k L \|d_k\|^2 \\ &= \frac{\alpha_k L \|g_{k+1}\| \cdot \|d_k\|}{\|g_{k+1}\|^2} \cdot \|g_{k+1}\| \cdot \|d_k\| \\ &\geq \frac{\|g_{k+1}\| \cdot \|g_{k+1} - g_k\|}{\|g_{k+1}\|^2} |g_{k+1}^T d_k| \\ &\geq \frac{|g_{k+1}^T (g_{k+1} - g_k)|}{\|g_{k+1}\|^2} \cdot |g_{k+1}^T d_k| \end{aligned}$$

$$\begin{aligned}
&\geq \frac{g_{k+1}^T(g_{k+1} - g_k)}{(1-u)\|g_k\|^2 - ug_k^T d_k} \frac{(1-u)\|g_k\|^2 - ug_k^T d_k}{\|g_{k+1}\|^2} g_{k+1}^T d_k \\
&= \beta_{k+1} \frac{(1-u)\|g_k\|^2 - ug_k^T d_k}{\|g_{k+1}\|^2} g_{k+1}^T d_k.
\end{aligned}$$

Therefore

$$(1-c)\|g_{k+1}\|^2 \geq \beta_{k+1} g_{k+1}^T d_k,$$

and thus

$$-c\|g_{k+1}\|^2 \geq -\|g_{k+1}\|^2 + \beta_{k+1} g_{k+1}^T d_k = g_{k+1}^T d_{k+1}.$$

The proof is finished.  $\square$

**Corollary 2.1.** *If the conditions in Lemma 2.2 hold then for  $k \geq 0$  we have*

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

**Lemma 2.3.** *Assume that (H1) and (H2) hold. Then the new nonmonotone line search is well-defined.*

**Proof.** On the one hand, since

$$\begin{aligned}
&\lim_{\alpha \rightarrow 0} \left[ \frac{\max_{0 \leq j \leq m(k)} [f_{k-j}] - f(x_k + \alpha d_k) + \mu \alpha (g_k^T d_k + \frac{1}{2} L_k \alpha \|d_k\|^2)}{\alpha} \right] \\
&\geq \lim_{\alpha \rightarrow 0} \left[ \frac{f_k - f(x_k + \alpha d_k) + \mu \alpha (g_k^T d_k + \frac{1}{2} \alpha L_k \|d_k\|^2)}{\alpha} \right] \\
&= -(1-\mu)g_k^T d_k \\
&> 0,
\end{aligned}$$

there is an  $\alpha'_k > 0$  such that

$$\max_{0 \leq j \leq m(k)} [f_{k-j}] - f(x_k + \alpha d_k) \geq -\mu \alpha \left[ g_k^T d_k + \frac{1}{2} \alpha L_k \|d_k\|^2 \right], \quad \forall \alpha \in [0, \alpha'_k].$$

Thus, letting  $\alpha''_k = \min(s_k, \alpha'_k)$  yields

$$\max_{0 \leq j \leq m(k)} [f_{k-j}] - f(x_k + \alpha d_k) \geq -\alpha \mu \left[ g_k^T d_k + \frac{1}{2} \alpha L_k \|d_k\|^2 \right], \quad (14)$$

for  $\alpha \in [0, \alpha''_k]$ . On the other hand, by Lemma 2.2, we can obtain

$$g(x_k + \alpha d_k)^T d(x_k + \alpha d_k) \leq -c\|g(x_k + \alpha d_k)\|^2, \quad (15)$$

if  $\alpha < \frac{1-c}{L} \frac{(1-u)\|g_k\|^2 - ug_k^T d_k}{\|d_k\|^2}$ . Letting

$$\bar{\alpha}_k = \min \left( \alpha''_k, \frac{1-c}{L} \frac{(1-u)\|g_k\|^2 - ug_k^T d_k}{\|d_k\|^2} \right),$$

we can see that both (14) and (15) hold whenever  $\alpha \in [0, \bar{\alpha}_k]$ . This shows that the new nonmonotone line search is well-defined. The proof is completed.  $\square$

### 3. Global convergence

**Lemma 3.1.** *Assume that (H1) and (H2) hold, and Algorithm (A) with the new nonmonotone line search generates an infinite sequence  $\{x_k\}$ . Then,*

$$\|d_k\| \leq \left( 1 + \frac{L(1-c)}{m_0} \right) \|g_k\|, \quad \forall k, \quad (16)$$

where  $m_0$  is defined in Lemma 2.1.

**Proof.** For  $k = 0$ , we have

$$\|d_k\| = \|g_k\| \leq \left(1 + \frac{L(1-c)}{m_0}\right) \|g_k\|.$$

For  $k > 0$ , by the new nonmonotone line search and Lemma 2.1, we have

$$\alpha_k \leq \frac{1-c}{L_k} \cdot \frac{(1-u)\|g_k\|^2 - ug_k^T d_k}{\|d_k\|^2} \leq \frac{1-c}{m_0} \cdot \frac{(1-u)\|g_k\|^2 - ug_k^T d_k}{\|d_k\|^2}.$$

By Cauchy–Schwartz inequality and the above inequality, noting the new conjugate gradient formula, we have

$$\begin{aligned} \|d_{k+1}\| &= \|-g_{k+1} + \beta_{k+1}d_k\| \\ &\leq \|g_{k+1}\| + \frac{|g_{k+1}(g_{k+1} - g_k)|}{(1-u)\|g_k\|^2 - ug_k^T d_k} \|d_k\| \\ &\leq \|g_{k+1}\| (1 + \alpha_k L \|d_k\|^2 / [(1-u)\|g_k\|^2 - ug_k^T d_k]) \\ &\leq \left(1 + \frac{L(1-c)}{m_0}\right) \|g_{k+1}\|. \end{aligned}$$

The proof is completed.  $\square$

**Theorem 3.1.** Assume that (H1) and (H2) hold, and Algorithm (A) with the new nonmonotone Armijo line search generates an infinite sequence  $\{x_k\}$ . Then there exists  $\eta > 0$  such that

$$\max_{0 \leq j \leq m(k)} [f_{k-j}] - f_{k+1} \geq \eta \|g_k\|^2. \quad (17)$$

**Proof.** Let  $\eta_0 = \inf_{k \in K} \{\alpha_k\}$ . If  $\eta_0 > 0$  then we have

$$\begin{aligned} \max_{0 \leq j \leq m(k)} [f_{k-j}] - f_{k+1} &\geq -\mu \alpha_k \left[ g_k^T d_k + \frac{1}{2} \alpha_k L_k \|d_k\|^2 \right] \\ &\geq -\mu \alpha_k \left[ g_k^T d_k + \frac{1}{2} s_k L_k \|d_k\|^2 \right] \\ &= -\mu \alpha_k \left[ g_k^T d_k - \frac{1}{2} (1-c) g_k^T d_k \right] \\ &= -\frac{\mu(1+c)}{2} \alpha_k g_k^T d_k \\ &\geq \frac{1}{2} \mu c (1+c) \eta_0 \|g_k\|^2. \end{aligned}$$

By letting  $\eta = \frac{1}{2} \mu c \eta_0 (1+c)$  we can prove the truth of (17).

In the following, we will prove that  $\eta_0 > 0$ . For the contrary, assume that  $\eta_0 = 0$ . Then, there exists an infinite subset  $K \subseteq \{0, 1, 2, \dots\}$  such that

$$\lim_{k \in K, k \rightarrow \infty} \alpha_k = 0. \quad (18)$$

By Lemmas 2.1, 2.2 and 3.1, we obtain

$$\begin{aligned} s_k &= \frac{1-c}{L_k} \cdot \frac{(1-u)\|g_k\|^2 - ug_k^T d_k}{\|d_k\|^2} \\ &\geq \frac{(1-c)}{L_k} \cdot \frac{(1-u+uc)\|g_k\|^2}{\|d_k\|^2} \\ &\geq \frac{(1-c)(1-u+uc)}{M_0} \left(1 + \frac{L(1-c)}{m_0}\right)^{-2} \\ &> 0. \end{aligned}$$

Therefore, (18) implies that there is a  $k'$  such that

$$\alpha_k / \rho \leq s_k, \quad \forall k \geq k' \text{ and } k \in K.$$

Letting  $\alpha = \alpha_k / \rho$ , at least one of the following two inequalities

$$\max_{0 \leq j \leq m(k)} [f_{k-j}] - f(x_k + \alpha d_k) \geq -\alpha \mu \left[ g_k^T d_k + \frac{1}{2} \alpha L_k \|d_k\|^2 \right], \quad (19)$$

and

$$g(x_k + \alpha d_k)^T d(x_k + \alpha d_k) \leq -c \|g(x_k + \alpha d_k)\|^2, \quad (20)$$

doesn't hold. If (19) doesn't hold, then we can deduce

$$f_k - f(x_k + \alpha d_k) < -\alpha \mu g_k^T d_k.$$

Using the mean value theorem on the left-hand side of the above inequality, there exists  $\theta_k \in [0, 1]$  such that

$$-\alpha g(x_k + \theta_k \alpha d_k)^T d_k < -\alpha \mu g_k^T d_k.$$

Thus

$$g(x_k + \theta_k \alpha d_k)^T d_k > \mu g_k^T d_k. \quad (21)$$

By (H2), Cauchy–Schwartz inequality, (21) and Lemma 2.2, we have

$$\begin{aligned} L\alpha \|d_k\|^2 &\geq \|g(x_k + \alpha \theta_k d_k) - g_k\| \cdot \|d_k\| \\ &\geq (g(x_k + \alpha \theta_k d_k) - g_k)^T d_k \\ &\geq -(1 - \mu) g_k^T d_k \\ &\geq c(1 - \mu) \|g_k\|^2. \end{aligned}$$

We can obtain from Lemma 3.1 that

$$\alpha_k \geq \frac{c\rho(1-\mu)}{L} \frac{\|g_k\|^2}{\|d_k\|^2} \geq \frac{c\rho(1-\mu)}{L(1+\frac{L(1-c)}{m_0})^2} > 0, \quad k \geq k', \quad k \in K,$$

which contradicts (18).

If (20) doesn't hold, then we have

$$g(x_k + \alpha d_k)^T d(x_k + \alpha d_k) > -c \|g(x_k + \alpha d_k)\|^2,$$

and thus,

$$\frac{g(x_k + \alpha d_k)(g(x_k + \alpha d_k) - g_k)}{(1-u)\|g_k\|^2 - u g_k^T d_k} g(x_k + \alpha d_k)^T d_k > (1-c) \|g(x_k + \alpha d_k)\|^2.$$

By using Cauchy–Schwartz inequality on the left-hand side of the above inequality we have

$$\alpha L \frac{\|d_k\|^2}{(1-u)\|g_k\|^2 - u g_k^T d_k} > 1 - c.$$

Combining Lemma 3.1 we have

$$\begin{aligned} \alpha_k &> \frac{\rho(1-c)}{L} \frac{(1-u)\|g_k\|^2 - u g_k^T d_k}{\|d_k\|^2} \\ &\geq \frac{\rho(1-c)(1-u+uc)}{L(1+\frac{L(1-c)}{m_0})^2} \\ &> 0, \quad k \geq k', \quad k \in K, \end{aligned}$$

which also contradicts (18). This shows that  $\eta_0 > 0$ . The whole proof is completed.  $\square$

**Lemma 3.2.** If the conditions of Theorem 3.1 hold, then,

$$\max_{1 \leq j \leq m} [f(x_{m+j})] \leq \max_{1 \leq i \leq m} [f(x_{m(l-1)+i})] - \eta \min_{1 \leq j \leq m} \|g_{m+j-1}\|^2, \quad (22)$$

and

$$\sum_{l=1}^{\infty} \min_{1 \leq j \leq m} \|g_{m+j-1}\|^2 < +\infty. \quad (23)$$

**Proof.** By (H1) and Lemma 3.1, it suffices to show that the following inequality holds for  $j = 1, 2, \dots, m$ ,

$$f(x_{ml+j}) \leq \max_{1 \leq i \leq m} [f(x_{m(l-1)+i})] - \eta \|g_{ml+j-1}\|^2. \quad (24)$$

Since the new nonmonotone Armijo line search and Theorem 3.1 imply

$$f(x_{ml+1}) \leq \max_{0 \leq i \leq m(ml)} [f(x_{ml-i})] - \eta \|g_{ml}\|^2, \quad (25)$$

it follows from this and

$$0 \leq m(ml) \leq m$$

that (24) holds for  $j = 1$ . Suppose that (24) holds for any  $j : 1 \leq j \leq m - 1$ . With the descent property of  $d_k$ , this implies

$$\max_{1 \leq i \leq j} [f(x_{ml+i})] \leq \max_{1 \leq i \leq m} [f(x_{m(l-1)+i})]. \quad (26)$$

By the new nonmonotone line search, the induction hypothesis,

$$m(ml+j) \leq m,$$

Theorem 3.1 and (25), we obtain

$$\begin{aligned} f(x_{ml+j+1}) &\leq \max_{0 \leq i \leq m(ml+j)} [f(x_{ml+j-i})] - \eta \|g_{ml+j}\|^2 \\ &\leq \max\{\max_{1 \leq i \leq m} f(x_{m(l-1)+i}), \max_{1 \leq i \leq j} f(x_{ml+i})\} - \eta \|g_{ml+j}\|^2 \\ &\leq \max_{1 \leq i \leq m} [f(x_{m(l-1)+i})] - \eta \|g_{ml+j}\|^2. \end{aligned}$$

Thus, (24) is also true for  $j + 1$ . By induction, (24) holds for  $1 \leq j \leq m$ . This shows that (22) holds.

Since  $f(x)$  is bounded from below by (H1), it follows that

$$\max_{1 \leq i \leq m} [f(x_{ml+i})] > -\infty.$$

By summing (22) over  $l$ , we can get

$$\sum_{l=1}^{\infty} \min_{1 \leq j \leq m} \|g_{ml+j-1}\|^2 < +\infty.$$

Therefore (23) holds. The proof is end.  $\square$

**Theorem 3.2.** If the conditions of Theorem 3.1 hold, then

$$\lim_{k \rightarrow \infty} \|g_k\| = 0. \quad (27)$$

**Proof.** Cauchy–Schwartz inequality and Lemma 2.2 imply that

$$\|d_k\| \geq c \|g_k\|. \quad (28)$$

By the new nonmonotone line search, Cauchy–Schwartz inequality, Lemmas 2.1, 2.2 and 3.1, and (28), we have

$$\begin{aligned} \alpha_k \leq s_k &= \frac{1-c}{L_k} \cdot \frac{(1-u)\|g_k\|^2 - u g_k^T d_k}{\|d_k\|^2} \\ &\leq \frac{1-c}{m_0} \cdot \frac{[(1-u)/c + u](-g_k^T d_k)}{\|d_k\|^2} \\ &\leq \frac{(1-c)(1-u+uc)}{cm_0} \frac{\|g_k\|}{\|d_k\|} \leq \frac{(1-c)(1-u+uc)}{m_0 c^2}. \end{aligned}$$

Noting the above inequality, (H2) and Lemma 3.1, it holds that

$$\begin{aligned} \|g_{k+1}\| &= \|g_{k+1} - g_k + g_k\| \\ &\leq \|g_{k+1} - g_k\| + \|g_k\| \\ &\leq L\alpha_k \|d_k\| + \|g_k\| \\ &\leq \frac{L(1-c)}{m_0 c} \|d_k\| + \|g_k\| \\ &\leq \left[ 1 + \frac{L(1-c)(1-u+uc)}{m_0 c^2} \left( 1 + \frac{L(1-c)}{m_0} \right) \right] \|g_k\|. \end{aligned}$$



Let

$$c_3 = 1 + \frac{L(1-c)(1-u+uc)}{m_0 c^2} \left( 1 + \frac{L(1-c)}{m_0} \right),$$

it follows that

$$\|g_{k+1}\| \leq c_3 \|g_k\|. \quad (29)$$

Let

$$\|g_{m+l+\phi(l)}\| = \min_{0 \leq i \leq m-1} \|g_{m+l+i}\|.$$

By Lemma 3.2 we obtain

$$\lim_{l \rightarrow \infty} \|g_{m+l+\phi(l)}\| = 0, \quad (30)$$

where

$$0 \leq \phi(l) \leq m-1.$$

By (29), we have

$$\|g_{m(l+1)+i}\| \leq c_3^{2m} \|g_{m+l+\phi(l)}\|, \quad i = 0, \dots, m-1. \quad (31)$$

Therefore, it follows from (30) and (31) that (27) holds.  $\square$

#### 4. Linear convergence rate

We further assume that

(H3). The sequence  $\{x_k\}$  generated by Algorithm (A) with the new nonmonotone Armijo line search converges to  $x^*$ ,  $\nabla^2 f(x^*)$  is a positive definite matrix and  $f(x)$  is twice continuously differentiable on  $N(x^*, \epsilon_0) = \{x \mid \|x - x^*\| < \epsilon_0\}$ .

**Lemma 4.1.** Assume that (H3) holds. Then there exist  $m', M'$  and  $\epsilon_0$  with  $0 < m' \leq M'$  and  $\epsilon \leq \epsilon_0$  such that

$$m' \|y\|^2 \leq y^T \nabla^2 f(x) y \leq M' \|y\|^2, \quad \forall x, y \in N(x^*, \epsilon); \quad (32)$$

$$\frac{1}{2} m' \|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{2} M' \|x - x^*\|^2, \quad \forall x \in N(x^*, \epsilon); \quad (33)$$

$$M' \|x - y\|^2 \geq (g(x) - g(y))^T (x - y) \geq m' \|x - y\|^2, \quad \forall x, y \in N(x^*, \epsilon); \quad (34)$$

and thus

$$M' \|x - x^*\|^2 \geq g(x)^T (x - x^*) \geq m' \|x - x^*\|^2, \quad \forall x \in N(x^*, \epsilon). \quad (35)$$

By (35) and (34) we can also obtain, from Cauchy–Schwartz inequality, that

$$M' \|x - x^*\| \geq \|g(x)\| \geq m' \|x - x^*\|, \quad \forall x \in N(x^*, \epsilon), \quad (36)$$

and

$$\|g(x) - g(y)\| \leq M' \|x - y\|, \quad \forall x, y \in N(x^*, \epsilon). \quad (37)$$

Its proof can be seen from the literature (e.g. [43]).

**Theorem 4.1.** Assume that (H3) holds, Algorithm (A) with the new nonmonotone line search generates an infinite sequence  $\{x_k\}$ . Then  $\{x_k\}$  converges to  $x^*$  at least  $R$ -linearly.

**Proof.** If (H3) holds then there exists  $k'$  such that  $x_k \in N(x^*, \epsilon_0)$ ,  $\forall k \geq k'$ . Without loss of generality, we assume that  $x_0 \in N(x^*, \epsilon_0)$ , and thus, (H1) and (H2) hold. By the proof of Theorem 3.1 we have

$$\max_{0 \leq j \leq m(k)} [f_{k-j} - f_{k+1}] \geq \eta \|g_k\|^2, \quad (38)$$

where  $\eta = \frac{1}{2} \mu c \eta_0 (1 + c)$ .

By Lemma 2.2 and Cauchy–Schwartz inequality we have

$$\|g_k\| \cdot \|d_k\| \geq -g_k^T d_k \geq c \|g_k\|^2,$$

which yields

$$\|d_k\| \geq c\|g_k\|. \quad (39)$$

According to the procedure of Algorithm (A) with the new nonmonotone line search and Lemma 4.1, we have

$$m' \leq L_k. \quad (40)$$

By the definition of  $\eta_0$  in the proof of Theorem 3.1, Cauchy–Schwartz inequality, (39) and (40), we have

$$\begin{aligned} \eta_0 &\leq \frac{1-c}{L_k} \cdot \frac{(1-u)\|g_k\|^2 - ug_k^T d_k}{\|d_k\|^2} \\ &\leq \frac{(1-c)(1-u+uc)}{cL_k} \cdot \frac{-g_k^T d_k}{\|d_k\|^2} \\ &\leq \frac{(1-c)(1-u+uc)}{cL_k} \cdot \frac{\|g_k\|}{\|d_k\|} \\ &\leq \frac{(1-c)(1-u+uc)}{c^2 m'}. \end{aligned}$$

Therefore

$$\eta_0 \leq \frac{(1-c)(1-u+uc)}{c^2 m'}. \quad (41)$$

By Lemma 4.1 and (38) we obtain

$$\begin{aligned} \max_{0 \leq j \leq m(k)} [f_{k-j}] - f_{k+1} &\geq \eta \|g_k\|^2 \\ &\geq \eta m'^2 \|x_k - x^*\|^2 \\ &\geq \frac{2\eta m'^2}{M'} (f_k - f^*). \end{aligned}$$

By setting

$$\theta = m' \sqrt{\frac{2\eta}{M'}},$$

we have

$$\max_{0 \leq j \leq m(k)} [f_{k-j}] - f_{k+1} \geq \theta^2 (f_k - f^*), \quad (42)$$

and we can prove that  $\theta < 1$ . In fact, by the definition of  $\eta$ , (41) and noting that  $m' < M' \leq L$ , we obtain

$$\begin{aligned} \theta^2 &= \frac{2m'^2 \eta}{M'} \leq \frac{m'^2 \eta_0 c \mu (1+c)}{M'} \\ &\leq \frac{m'^2 c (1+c) (1-c) (1-u+uc) \mu}{m' M' c^2} \\ &\leq \frac{\mu (1-c^2) (1-u+uc)}{c} \\ &< \frac{\mu (1-c^2)}{c} < 1. \end{aligned}$$

Noting (42), the following proof is similar to that of [6,40].  $\square$

## 5. Numerical reports

In this section, we will conduct some numerical experiments to show the efficiency of the new class of conjugate gradient methods with the new nonmonotone line search. CG1, CG2, and CG3 denote Algorithm (A) with the new nonmonotone line search corresponding to the estimation (10)–(12) respectively. LS denotes the original LS method with strong Wolfe line search. LS+ denotes the LS method with

$$\beta_k = \max(0, \beta^{LS})$$

and strong Wolfe line search [7,20,26].

In the same way, PRP denotes the original PRP method with strong Wolfe line search. PRP+ denotes the PRP method with

$$\beta_k = \max(0, \beta^{PRP})$$

and strong Wolfe line search

*Strong Wolfe line search.* Choose  $\alpha_k$  to satisfy

$$f_k - f(x_k + \alpha d_k) \geq -\mu \alpha g_k^T d_k, \quad (43)$$

and

$$|g(x_k + \alpha d_k)^T d_k| \leq -\sigma g_k^T d_k, \quad (44)$$

with  $\mu = 0.25$  and  $\sigma = 0.75$ .

Strong Wolfe line search has been proved to be an efficient line search technique in some literature. Several numerical results also have shown that strong Wolfe line search has a good performance in practical computation. However, it needs to solve a system of two inequalities in one variable, which increased the cost of computation. In fact, we should estimate some adequate step size at each iteration. In so doing, we can save computational time and memory cost. How do we estimate the step size in terms of iterative information? In this paper, we analyzed three approaches to estimating the Lipschitz constant and step sizes. Numerical results showed that the estimation is useful in practical computation.

We first chose 15 test problems (Problems 21–35) with the dimension  $n = 10000$  and initial points from literature [31] to implement the new class of conjugate gradient methods with the new nonmonotone line search on a portable computer with Pentium IV/1.2 MHz CPU. The double precision and C++ language were used to program the procedure. We set the parameters  $m = 3$ ,  $\mu = 0.25$ ,  $\rho = 0.75$ ,  $c = 0.75$  and  $L_0 = 1$  in the numerical experiment.

In practical computation, if  $m$  is too large then we can find a large step size. In the beginning of iteration, taking large  $m$  is reasonable because we hope the iterates move to an optimal point quickly. But, if the iterates are close neighbors of the optimal point, taking large  $m$  is not reasonable because we need not large step size in this case. Practical computation can remind us to take large or small  $m$  in solving different problems. In common cases, taking  $m = 3$  is reasonable for most optimization problems.

The stop criterion is

$$\|g_k\| \leq 10^{-8} \|g_0\|,$$

and the numerical results are given in Table 1.

In Table 1, CPU denotes the total CPU time (seconds) for solving all the 15 test problems. It can be seen from Table 1 that the LS method (Algorithm (A) corresponding to  $u = 0$ ) and PRP method (Algorithm (A) corresponding to  $u = 1$ ) with the new nonmonotone line search are effective for solving some large scale problems. In particular, LS1 and PRP1 methods seem to be the best ones among the eight algorithms because they use the least iterative number and function evaluations to the same precision. This shows that the estimation formula (10) may be more reasonable than other formulae. In fact, we have

$$\frac{|\delta_{k-1}^T y_{k-1}|}{\|\delta_{k-1}\|^2} \leq \frac{\|y_{k-1}\|}{\|\delta_{k-1}\|} \leq \frac{\|y_{k-1}\|^2}{|\delta_{k-1}^T y_{k-1}|}.$$

This motivates us to guess that the available Lipschitz constant should be chosen in the interval

$$\left[ \frac{\delta_{k-1}^T y_{k-1}}{\|\delta_{k-1}\|^2}, \frac{\|y_{k-1}\|^2}{\delta_{k-1}^T y_{k-1}} \right].$$

It can be seen from Tables 1 and 2 that LS and PRP methods with the new nonmonotone line search are superior to LS, LS+, PRP and PRP+ conjugate gradient methods, respectively. Moreover, LS and PRP methods may fail in some cases if we chose inadequate parameters. Although LS+ and PRP+ conjugate gradient methods have global convergence, their numerical performance is not better than LS and PRP methods in many situations. Numerical experiments show that the new nonmonotone line search proposed in this paper is effective for LS and PRP methods in practical computation. The reason is, we used local Lipschitz constant estimation in the new line search and could define an adequate initial step size  $s_k$  so as to seek a suitable step size  $\alpha_k$  for LS and PRP methods, which reduced the function evaluations at each iteration and improved the efficiency of LS and PRP methods.

However, since the 15 test problems from [31] are almost quadratic, the above comparisons seem to be inadequate in some sense. We chose another 15 test problems from CUTer (see [3,21,22]) and implemented the new conjugate gradient methods with the new nonmonotone line search, LS1, LS2, LS3, PRP1, PRP2, PRP3, LS, LS+, PRP and PRP+ conjugate gradient methods. The stop criteria is

$$\|g_k\|_\infty \leq 10^{-5} (1 + |f_0|).$$

Numerical results are summarized in Tables 3 and 4.

From Tables 3 and 4, we can see that the new conjugate gradient methods PRP1, PRP2, PRP2 are superior to PRP and PRP+ methods in practical computation. In the same way, LS1, LS2 and LS3 methods are superior to LS and LS+ methods.

All the numerical results show that the step size estimation at each iteration is very important in practical computation. It also shows that, if we find an adequate step size estimation at each iteration, Armijo line search can be superior to Wolfe line search in practical computation.

**Table 1**

Iterations and function evaluations for LS methods

$P$	$n$	LS1	LS2	LS3	LS	LS+
21	$10^4$	32/198	33/93	52/118	52/187	65/142
22	$10^4$	30/89	47/189	54/223	61/321	63/387
23	$10^4$	25/78	30/98	32/78	44/136	51/221
24	$10^4$	31/114	46/214	52/281	77/256	81/268
25	$10^4$	43/142	42/178	57/211	43/118	52/176
26	$10^4$	32/138	35/126	33/113	52/174	55/192
27	$10^4$	36/167	42/181	33/135	49/227	42/189
28	$10^4$	52/273	62/226	72/223	72/186	78/346
29	$10^4$	28/96	26/106	31/128	42/191	52/189
30	$10^4$	46/178	38/167	33/126	41/167	42/788
31	$10^4$	62/145	72/126	67/178	72/168	72/189
32	$10^4$	52/167	52/194	75/186	78/197	72/216
33	$10^4$	32/186	45/173	42/184	36/127	48/175
34	$10^4$	38/183	36/263	36/268	59/250	54/271
35	$10^4$	56/273	62/226	51/331	68/239	61/278
CPU	–	89 s	118 s	169 s	248 s	295 s

**Table 2**

Iterations and function evaluations for PRP methods

$P$	$n$	PRP1	PRP2	PRP3	PRP	PRP+
21	$10^4$	31/165	32/73	46/97	48/138	59/133
22	$10^4$	30/93	38/167	58/186	52/273	66/291
23	$10^4$	24/66	32/112	31/93	46/138	54/184
24	$10^4$	35/134	42/189	55/178	66/183	78/224
25	$10^4$	45/157	42/162	55/216	43/164	58/196
26	$10^4$	47/164	36/95	38/115	56/123	54/160
27	$10^4$	32/137	48/141	36/193	44/186	46/173
28	$10^4$	46/189	61/185	62/164	78/216	72/281
29	$10^4$	33/78	29/151	35/142	46/167	56/173
30	$10^4$	46/163	38/184	33/152	46/169	43/346
31	$10^4$	52/121	73/166	64/171	76/166	73/164
32	$10^4$	49/137	52/156	78/191	71/187	75/222
33	$10^4$	33/146	48/175	44/165	36/106	42/157
34	$10^4$	32/164	38/180	35/221	54/282	57/287
35	$10^4$	51/232	55/186	56/257	69/213	53/242
CPU	–	82 s	121 s	173 s	256 s	318 s

**Table 3**

Iterations and function evaluations for LS methods

$P$	$n$	LS1	LS2	LS3	LS+	LS
ARWHEAD	$10^4$	12/17	13/18	12/17	22/56	25/55
DQDRTIC	$10^4$	13/29	17/28	14/23	16/28	15/42
ENGVAL1	$10^4$	15/18	14/19	12/19	13/25	14/38
VAREIGVL	$10^4$	18/23	18/27	19/28	18/28	19/54
WOODS	$10^4$	15/18	17/24	15/24	18/34	19/56
LIARWHD	$10^4$	27/32	25/38	23/33	28/42	38/49
MOREBV	$10^4$	76/82	73/81	73/85	79/83	74/97
NONDIA	$10^4$	22/26	22/28	22/26	22/36	28/52
TQUARTIC	$10^4$	23/29	23/28	23/27	22/29	24/46
POWELLSG	$10^4$	56/79	58/64	53/69	48/67	53/92
QUARTC	$10^4$	23/45	22/36	27/31	22/64	28/48
SCHMVETT	$10^4$	22/29	24/29	25/32	24/36	25/38
SPARSQUR	$10^4$	34/38	45/86	42/53	36/68	49/69
SROSENBR	$10^4$	18/23	19/26	21/42	19/36	27/73
TOINTGSS	$10^4$	16/27	18/26	16/25	18/35	19/49
CPU	–	86 s	113 s	82 s	132	183 s

## 6. Conclusion

In this paper, we proposed a new class of conjugate gradient methods for minimizing functions that have Lipschitz continuous partial derivatives. In particular, the new class of conjugate gradient methods contains PRP and LS methods as its special cases. A new nonmonotone line search was proposed for guaranteeing the global convergence of these new conjugate gradient methods. It needs to estimate the local Lipschitz constant of the derivative of objective functions in the

**Table 4**

Iterations and function evaluations for PRP methods

P	n	PRP1	PRP2	PRP3	PRP	PRP+
ARWHEAD	10 <sup>4</sup>	11/19	14/18	12/19	23/53	24/77
DQDRTIC	10 <sup>4</sup>	13/32	18/36	15/28	19/39	18/43
ENGVAL1	10 <sup>4</sup>	16/19	15/23	14/27	17/53	19/58
VAREIGVL	10 <sup>4</sup>	18/28	18/25	19/26	23/48	21/57
WOODS	10 <sup>4</sup>	16/23	15/19	17/28	19/38	19/46
LIARWHD	10 <sup>4</sup>	23/36	23/38	23/34	27/51	27/69
MOREBV	10 <sup>4</sup>	71/87	72/89	72/89	79/98	75/94
NONDIA	10 <sup>4</sup>	21/28	21/32	21/26	26/42	28/59
TQUARTIC	10 <sup>4</sup>	22/27	26/48	26/49	32/48	33/65
POWELLSG	10 <sup>4</sup>	48/74	53/68	51/73	54/87	58/96
QUARTC	10 <sup>4</sup>	24/45	22/38	23/39	27/78	28/81
SCHMVETT	10 <sup>4</sup>	23/31	21/27	23/29	31/39	42/68
SPARSQR	10 <sup>4</sup>	32/48	42/58	42/55	32/73	33/85
SROSENBR	10 <sup>4</sup>	21/28	21/36	23/47	31/67	29/73
TOINTGSS	10 <sup>4</sup>	18/23	16/32	15/39	22/58	21/64
CPU	–	89 s	107 s	91 s	152 s	197 s

new nonmonotone line search. The global convergence and linear convergence rate of the new class of conjugate gradient methods with the new nonmonotone line search were analyzed under some mild conditions. Numerical results showed that the corresponding LS and PRP methods with the new nonmonotone line search are effective and superior to LS, LS+, PRP and PRP+ conjugate gradient methods with strong Wolfe line search. The main reason is that, using local Lipschitz constant estimation we can choose an adequate initial step size  $s_k$  and a suitable step size  $\alpha_k$  at each iteration, which decreases the function evaluations and improves the efficiency of conjugate gradient methods. In fact, we need only to estimate the local Lipschitz constant of the derivative of objective functions, although we used the global Lipschitz constant in analyzing the global convergence of these conjugate gradient methods.

For future research, we should investigate the estimation for the Lipschitz constant of the derivative of objective functions and give a wider scope for step sizes. Furthermore, we can investigate these conjugate gradient methods when  $u \in (0, 1)$ .

## References

- [1] L. Armijo, Minimization of functions having Lipschitz continuous partial derivatives, *Pacific J. Math.* 16 (1966) 1–3.
- [2] M. Al-Baali, New property and global convergence of the Fletcher–Reeves method with inexact line searches, *IMA J. Numer. Anal.* 5 (1985) 122–124.
- [3] I. Bongartz, A.R. Conn, N.I.M. Gould, P.L. Toint, CUTE: Constrained and unconstrained testing environments, *ACM Trans. Math. Softw.* 21 (1995) 123–160.
- [4] X.D. Chen, J. Sun, Global convergence of a two-parameter family of conjugate gradient methods without line search, *J. Comput. Appl. Math.* 146 (2002) 37–45.
- [5] Y.H. Dai, New properties of a nonlinear conjugate gradient method, *Numer. Math.* 89 (2001) 83–98.
- [6] Y.H. Dai, On the nonmonotone line search, *J. Optim. Theory Appl.* 112 (2) (2002) 315–330.
- [7] Y.H. Dai, Y. Yuan, Convergence properties of the conjugate descent method, *Adv. Math.* 25 (1996) 552–562.
- [8] Y.H. Dai, Y. Yuan, Convergence properties of the Fletcher–Reeves method, *IMA J. Numer. Anal.* 16 (1996) 155–164.
- [9] Y.H. Dai, Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, *SIAM J. Optim.* 10 (1999) 177–182.
- [10] Y.H. Dai, J.Y. Han, G.H. Liu, D.F. Sun, H.X. Yin, Y.X. Yuan, Convergence properties of nonlinear conjugate gradient methods, *SIAM J. Optim.* 10 (1999) 345–358.
- [11] Y.H. Dai, A family of hybrid conjugate gradient methods for unconstrained optimization, *Math. Comp.* 72 (2003) 1317–1328.
- [12] R. Fletcher, *Practical Method of Optimization*, 2nd ed., in: *Unconstrained Optimization*, vol. I, Wiley, New York, 1987.
- [13] R. Fletcher, C. Reeves, Function minimization by conjugate gradients, *Comput. J.* 7 (1964) 149–154.
- [14] G. Fasano, Planar conjugate gradient algorithm for large-scale unconstrained optimization, Part 1: Theory, *J. Optim. Theory Appl.* 125 (2005) 523–541.
- [15] G. Fasano, Planar conjugate gradient algorithm for large-scale unconstrained optimization, Part 2: Applications, *J. Optim. Theory Appl.* 125 (2005) 543–558.
- [16] L. Grippo, S. Lucidi, A globally convergent version of the Polak–Ribière conjugate gradient method, *Math. Program* 78 (1997) 375–391.
- [17] L. Grippo, F. Lampariello, S. Lucidi, A nonmonotone line search technique for Newton's method, *SIAM J. Numer. Anal.* 23 (1986) 707–716.
- [18] L. Grippo, F. Lampariello, S. Lucidi, A truncated Newton method with nonmonotone line search for unconstrained optimization, *J. Optim. Theory Appl.* 60 (1989) 401–419.
- [19] L. Grippo, M. Sciandrone, Nonmonotone globalization techniques for the Barzilai–Borwein gradient method, *Comput. Optim. Appl.* 23 (2002) 143–169.
- [20] J.C. Gilbert, J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM J. Optim.* 2 (1) (1992) 21–42.
- [21] W.W. Hager, H.C. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, *SIAM J. Optim.* 16 (2005) 170–192.
- [22] W.W. Hager, H.C. Zhang, Algorithm 851: CG\_DESCENT, A conjugate gradient method with guaranteed descent, *ACM Trans. Math. Softw.* 32 (2006) 113–137.
- [23] M.R. Hestenes, E. Stiefel, Method of conjugate gradient for solving linear systems, *J. Res. Natl. Bur. Stand.* 49 (1952) 409–436.
- [24] Y.F. Hu, C. Storey, Global convergence result for conjugate gradient methods, *J. Optim. Theory Appl.* 71 (1991) 399–405.
- [25] Y. Liu, C. Storey, Efficient generalized conjugate gradient algorithms, part 1: theory, *J. Optim. Theory Appl.* 69 (1991) 129–137.
- [26] J. Nocedal, J.S. Wright, *Numerical Optimization*, Springer-Verlag New York, Inc., 1999.
- [27] E. Polak, G. Ribière, Note sur la convergence de directions conjuguées, *Rev. Française Informat Recherche Opérationnelle*, 3e Année 16 (1969) 35–43.
- [28] B.T. Polyak, The conjugate gradient method in extreme problems, *USSR Comp. Math. Math. Phys.* 9 (1969) 94–112.
- [29] M. Raydan, The Barzilai and Borwein gradient method for the large-scale unconstrained minimization problem, *SIAM J. Optim.* 7 (1997) 26–33.
- [30] J. Sun, J.P. Zhang, Convergence of conjugate gradient methods without line search, *Ann. Oper. Res.* 103 (2001) 161–173.
- [31] J. Sun, X.Q. Yang, X.D. Chen, Quadratic cost flow and the conjugate gradient method, *European J. Oper. Res.* 164 (2005) 104–114.
- [32] D.F. Shanno, On the convergence of a new conjugate gradient algorithm, *SIAM J. Numer. Anal.* 15 (1978) 1247–1257.

- [33] W.Y. Sun, J.Y. Han, J. Sun, Global convergence of nonmonotone descent methods for unconstrained optimization problems, *J. Comput. Appl. Math.* 146 (2002) 89–98.
- [34] Z.J. Shi, Restricted PR conjugate gradient method and its global convergence, *Adv. in Math.* 31 (1) (2002) 47–55. (in Chinese).
- [35] Z.J. Shi, Modified HS conjugate gradient method and its global convergence, *Math. Numer. Sin.* 23 (2001) 383–406. (in Chinese).
- [36] Z.J. Shi, J. Shen, Convergence of Polak–Rebière–Polyak conjugate gradient method, *Nonlinear Anal.* 66 (2007) 1428–1441.
- [37] Z.J. Shi, J. Shen, Convergence of descent method without line search, *Appl. Math. Comput.* 167 (2005) 94–107.
- [38] Z.J. Shi, J. Shen, Step-size estimation for unconstrained optimization methods, *Comput. Appl. Math.* 24 (3) (2005) 399–416.
- [39] Z.J. Shi, J. Shen, New inexact line search method for unconstrained optimization, *J. Optim. Theory Appl.* 127 (2) (2005) 425–446.
- [40] Z.J. Shi, J. Shen, Convergence of nonmonotone line search methods, *J. Comput. Appl. Math.* 193 (2006) 397–412.
- [41] P.L. Toint, A nonmonotone trust-region algorithm for nonlinear optimization subject to convex constraints, *Math. Prog.* 77 (1997) 69–94.
- [42] P.L. Toint, An assessment of nonmonotone line search techniques for unconstrained optimization, *SIAM J. Sci. Comput.* 17 (1996) 725–739.
- [43] Y. Yuan, *Numerical Methods for Nonlinear Programming*, Shanghai Science and Technology Publisher, 1993.
- [44] Y. Yuan, Analysis on the conjugate gradient method, *Optim. Methods Softw.* 2 (1993) 19–29.
- [45] H.C. Zhang, W.W. Hager, A nonmonotone line search technique and its application to unconstrained optimization, *SIAM J. Optim.* 14 (2004) 1043–1056.

**Zhen-Jun Shi** completed his Master Degree in computational mathematics from Nanjing University in 1991; Ph.D. in operations research from Dalian University of Technology in 2002; obtained postdoctoral training in Chinese Academy of Science from 2003 to 2005; and he is currently a research fellow in Department of Computer & Information Science, University of Michigan–Dearborn, USA. His research area includes numerical optimization, nonlinear programming, numerical linear algebra, and computer application with operations research, etc.

**Jinhua Guo** completed his Master Degree in computer science from Dalian University of Technology in 1995 and his Ph.D. in computer science from University of Georgia in 2002, USA. He is currently an assistant professor in the Department of Computer & Information Science, University of Michigan, Dearborn, USA. He has published over 20 technical papers. His research areas include distributed computing, grid computing, computational steering and optimization, etc.