COMPUTATIONAL & APPLIED MATHEMATICS

Volume 27, N. 1, pp. 93–106, 2008

ISSN 0101-8205
www.scielo.br/cam

A new algorithm of nonlinear conjugate gradient method with strong convergence*

ZHEN-JUN SHI[1,2] and JINHUA GUO[2]

[1]College of Operations Research and Management, Qufu Normal University
Rizhao, Sahndong 276826, P.R. China

[2]Department of Computer and Information Science, University of Michigan
Dearborn, Michigan 48128-1491, USA

E-mails: zjshi@qrnu.edu.cn; zjshi@umd.umich.edu / jinhua@umd.umich.edu

Abstract. The nonlinear conjugate gradient method is a very useful technique for solving large scale minimization problems and has wide applications in many fields. In this paper, we present a new algorithm of nonlinear conjugate gradient method with strong convergence for unconstrained minimization problems. The new algorithm can generate an adequate trust region radius automatically at each iteration and has global convergence and linear convergence rate under some mild conditions. Numerical results show that the new algorithm is efficient in practical computation and superior to other similar methods in many situations.

Mathematical subject classification: 90C30, 65K05, 49M37.

Key words: unconstrained optimization, nonlinear conjugate gradient method, global convergence, linear convergence rate.

1 Introduction

Consider an unconstrained minimization problem

$$\min \ f(x), \qquad x \in R^n, \tag{1}$$

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

#724/07. Received: 10/V/07. Accepted: 24/IX/07.

*The work was supported in part by NSF CNS-0521142, USA.

When n is very large (for example, $n > 10^6$) the related problem is called large scale minimization problem. In order to solve large scale minimization problems, we need to design special algorithms that avoid the high storage and computation cost of some matrices.

The conjugate gradient method is a suitable approach to solving large scale minimization problems. For strictly convex quadratic objective functions, the conjugate gradient method with exact line searches has the finite convergence property. If the objective function is not a quadratic or the inexact line searches are used, the conjugate gradient method has no finite convergence property or even no global convergence property [6, 20].

When the conjugate gradient method is used to minimize non-quadratic objective functions, the related algorithm is called the nonlinear conjugate gradient method [17, 18]. There has been much literature to study the nonlinear conjugate gradient methods [3, 4, 5]. Meanwhile, some new nonlinear conjugate gradient methods have appeared [8, 11].

The conjugate gradient method has the form

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

where x_0 is an initial point, α_k is a step size, and d_k can be taken as

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

in which $g_k = \nabla f(x_k)$. Different β_k will determine different conjugate gradient methods. Some famous formulae for β_k are as follows.

$$\beta^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \qquad \text{(Fletcher-Reeves [10])} \tag{4}$$

$$\beta^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2}, \qquad \text{(Polak-Ribiére-Polyak [15, 16])} \tag{5}$$

$$\beta^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})}, \qquad \text{(Hestenes-Stiefiel [12])} \tag{6}$$

$$\beta^{CD} = -\frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}}, \qquad \text{(Conjugate-Descent [11])} \tag{7}$$

$$\beta^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T(g_k - g_{k-1})}, \qquad \text{(Dai-Yuan [16])} \tag{8}$$

$$\beta^{LS} = -\frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T g_{k-1}}. \qquad \text{(Liu-Storey [13])} \tag{9}$$

Although some conjugate gradient methods have good numerical performance in solving large scale minimization problems, they have no global convergence in some situations [6]. We often have two questions. Whether can we construct a conjugate gradient method that has both global convergence and good numerical performance in practical computation? Whether can we design a conjugate gradient method that is suitable to solve ill-conditioned minimization problems (the Hessian of objective functions at a stationary point is ill-conditioned)?

Yuan and Stoer [19] studied the conjugate gradient method on a subspace and obtained a new conjugate gradient method. In their algorithm, the search direction was taken from the subspace $span\{g_k, d_{k-1}\}$ at the kth iteration ($k \geq 1$), i.e.,

$$d_k = \gamma_k g_k + \beta_k d_{k-1}, \tag{10}$$

where γ_k and β_k are parameters.

Motivated by [19], we can apply the trust region technique to the conjugate gradient method and propose a new algorithm of nonlinear conjugate gradient methods. This new algorithm has both global convergence and good numerical performance in practical computation. Theoretical analysis and numerical results show that the proposed algorithm is promising and can solve some ill-conditioned minimization problems.

The paper is organized as follows. Section 1 is the introduction. In Section 2, we introduce the new conjugate gradient method. In Sections 3 and 4, we analyze the global convergence and convergence rate of the new method. Numerical results are reported in Section 5.

2 New Algorithm

We first assume that

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

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

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in B. \tag{11}$$

Lemma 2.1. *Assume that* (H2) *holds and* $x_k, x_k + d_k \in B$*, then*

$$f(x_k + d_k) - f_k \leq g_k^T d_k + \frac{1}{2}L\|d_k\|^2. \tag{12}$$

Proof. The proof is easy to obtain from mean value theorem and here is omitted. □

Algorithm (A)

Step 0. Choose parameters $\mu \in (0, 1)$, $\rho \in (0, 1)$ and $M_0 \gg L_0 > 0$; given initial point $x_0 \in R^n$, set $k := 0$.

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

Step 2. $x_{k+1} = x_k + d_k(\alpha_k)$, where α_k is the largest one in $\{1, \rho, \rho^2, \ldots,\}$ such that

$$\frac{f_k - f(x_k + d_k(\alpha))}{q_k(0) - q_k(d_k(\alpha))} \geq \mu, \tag{13}$$

in which

$$d_k(\alpha) = \begin{cases} -\gamma g_k, & k = 0; \\ -\gamma g_k + \beta d_{k-1}, & k \geq 1, \end{cases} \tag{14}$$

and $(\gamma, \beta)^T \in R^2$ is a solution to

$$\begin{aligned} \min q_k(d_k(\alpha)) \;&=\; f_k + g_k^T d_k(\alpha) + \frac{1}{2}L_k\|d_k(\alpha)\|^2, \; s.t. \; \|d_k(\alpha)\| \\ &\leq\; \frac{\alpha\|g_k\|}{L_k}. \end{aligned} \tag{15}$$

Step 3.

$$L_{k+1} = \max\left(L_0, \min\left(\frac{|(g_{k+1} - g_k)^T(x_{k+1} - x_k)|}{\|x_{k+1} - x_k\|^2}, M_0\right)\right); \tag{16}$$

or

$$L_{k+1} = \max\left(L_0, \min\left(\frac{\|g_{k+1} - g_k\|}{\|x_{k+1} - x_k\|}, M_0\right)\right); \tag{17}$$

or

$$L_{k+1} = \max\left(L_0, \min\left(\frac{\|g_{k+1} - g_k\|^2}{|(g_{k+1} - g_k)^T (x_{k+1} - x_k)|}, M_0\right)\right); \tag{18}$$

Step 4. Set $k := k + 1$ and goto **Step 1**.

Remark 2.1. In Algorithm (A), the main task is to solve (15). In fact, if $k = 0$ then the problem (15) has a solution $\gamma = \alpha / L_k$. If $k \geq 1$ then the problem (5) has the solution

$$y = (\gamma, \beta)^T = \begin{cases} -\dfrac{y_k}{L_k}, & \|y_k\| \leq \alpha \|g_k\|; \\ -\dfrac{\alpha \|g_k\|}{L_k \|y_k\|} y_k, & \|y_k\| > \alpha \|g_k\|, \end{cases} \tag{19}$$

where $y_k = (\gamma', \beta')^T$ is a solution of the equations in two variables

$$\begin{cases} \|g_k\|^2 \gamma - (g_k^T d_{k-1}) \beta = \|g_k\|^2, \\ -(g_k^T d_{k-1}) \gamma + \|d_{k-1}\|^2 \beta = -g_k^T d_{k-1}. \end{cases} \tag{20}$$

Moreover L_k is an approximation to the Lipschitz constant L of the gradient of the objective function. If we set $\beta \equiv 0$ then Algorithm (A) is very similar to BB method [1, 7]. However, Algorithm (A) has global convergence.

Lemma 2.2. *If* (H2) *holds then*

$$L_0 \leq L_k \leq \max(L, M_0). \tag{21}$$

In fact, by the Cauchy-Schwartz inequality, we have

$$\frac{|(g_{k+1} - g_k)^T (x_{k+1} - x_k)|}{\|x_{k+1} - x_k\|^2} \leq \frac{\|g_{k+1} - g_k\|}{\|x_{k+1} - x_k\|} \leq \frac{\|g_{k+1} - g_k\|^2}{|(g_{k+1} - g_k)^T (x_{k+1} - x_k)|},$$

and thus, L_{k+1} should be in the interval

$$\left[\frac{|(g_{k+1}-g_k)^T(x_{k+1}-x_k)|}{\|x_{k+1}-x_k\|^2}, \frac{\|g_{k+1}-g_k\|^2}{|(g_{k+1}-g_k)^T(x_{k+1}-x_k)}\right].$$

Generally, we take

$$L_{k+1} = \frac{\|g_{k+1}-g_k\|}{\|x_{k+1}-x_k\|}$$

in practical computation.

3 Global convergence

Lemma 3.2. *Assume that* (H1) *and* (H2) *hold, then*

$$q_k(0) - q_k(d_k(\alpha)) \geq \frac{1}{2}\alpha\|g_k\|^2/L_k. \tag{22}$$

Proof. Set $\overline{d}_k(\alpha) = -\gamma g_k$ such that $\|\overline{d}_k(\alpha)\| = \alpha\|g_k\|/L_k$, then $\overline{d}_k(\alpha)$ is a feasible solution to (15). By noting $\alpha \in (0,1]$ and $d_k(\alpha)$ being an optimal solution to (15), we have

$$\begin{aligned} q_k(0) - q_k(d_k(\alpha)) &\geq q_k(0) - q_k(\overline{d}_k(\alpha)) \\ &= -g_k^T\overline{d}_k(\alpha) - \frac{1}{2}L_k\|\overline{d}_k(\alpha)\|^2 \\ &= \alpha\|g_k\|^2/L_k - \frac{1}{2}\alpha^2\|g_k\|^2/L_k \\ &\geq \frac{1}{2}\alpha\|g_k\|^2/L_k. \end{aligned} \qquad \square$$

Theorem 3.1. *Assume that* (H1) *and* (H2) *hold. Algorithm* (A) *generates an infinite sequence* $\{x_k\}$. *Then*

$$\lim_{k\to\infty}\|g_k\| = 0. \tag{23}$$

Proof. It is easy to obtain from (H1), (H2) and Lemmas 2.1, 2.2 and 3.1, that

$$\left|\frac{f_k - f(x_k + d_k(\alpha)}{q_k(0) - q_k(d_k(\alpha))} - 1\right| = \left|\frac{f_k - f(x_k + d_k(\alpha)) - q_k(0) + q_k(d_k(\alpha))}{q_k(0) - q_k(d_k(\alpha)}\right|$$

$$
\begin{aligned}
&\le \frac{\frac{1}{2}L\|d_k(\alpha)\| + \frac{1}{2}L_k\|d_k(\alpha)\|^2}{\frac{1}{2}\alpha\|g_k\|^2/L_k} \\
&\le \frac{2\max(L, M_0)\|d_k(\alpha)\|^2}{\alpha\|g_k\|^2/L_k} \\
&\le \frac{2\max(L, M_0)\alpha}{L_k} \\
&\le \frac{2\max(L, M_0)}{L_0}\alpha \\
&\to 0(\alpha \to 0).
\end{aligned}
$$

This shows that if $\alpha \le \dfrac{(1-\mu)L_0}{2\max(L, M_0)}$ then we have $\dfrac{f_k - f(x_k + d_k(\alpha)}{q_k(0) - q_k(d_k(\alpha))} \ge \mu$. Therefore, there exists $\eta_0 > 0$ such that $\alpha_k \ge \eta_0$. By Lemma 3.2 and the procedure of Algorithm (A), we have

$$
\begin{aligned}
f_k - f(x_k + d_k(\alpha_k)) &\ge \mu[q_k(0) - q_k(d_k(\alpha_k))] \\
&\ge \frac{1}{2}\mu\alpha\|g_k\|^2/L_k \\
&\ge \frac{1}{2}\mu\eta_0\|g_k\|^2/\max(L, M_0).
\end{aligned}
$$

By (H1) and the above inequality, we assert that $\{f_k\}$ is a monotone decreasing number sequence and has a lower bound. Therefore, $\{f_k\}$ has a limit and thus,

$$
\frac{1}{2}\mu\eta_0/\max(L, M_0)\sum_{k=0}^{+\infty}\|g_k\|^2 \le f_0 - \lim_{k\to\infty} f_k < +\infty,
$$

which implies that (23) holds. □

4 Linear convergence rate

We further assume that

(H3) The sequence $\{x_k\}$ generated by Algorithm (A) 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 ϵ with $0 < m' \leq M'$ and $\epsilon \leq \epsilon_0$ such that*

$$m'\|y\|^2 \leq y^T \nabla^2 f(x) y \leq M'\|y\|^2, \quad \forall x, y \in N(x^*, \epsilon); \tag{24}$$

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

$$M'\|x - y\|^2 \geq (g(x) - g(y))^T(x - y) \geq m'\|x - y\|^2, \quad \forall x, y \in N(x^*, \epsilon); \tag{26}$$

and thus

$$M'\|x - x^*\|^2 \geq g(x)^T(x - x^*) \geq m'\|x - x^*\|^2, \quad \forall x \in N(x^*, \epsilon). \tag{27}$$

By (27) *and* (26) *we can also obtain, from Cauchy-Schwartz inequality, that*

$$M'\|x - x^*\| \geq \|g(x)\| \geq m'\|x - x^*\|, \quad \forall x \in N(x^*, \epsilon), \tag{28}$$

and

$$\|g(x) - g(y)\| \leq M'\|x - y\|, \quad \forall x, y \in N(x^*, \epsilon). \tag{29}$$

Its proof can be seen from the literature (e.g. [11]).

Lemma 4.2. *Assume that* (H3) *holds and Algorithm* (A) *generates an infinite sequence $\{x_k\}$. Then*

$$\eta_0 = \inf_{\forall k} \{\alpha_k\} > 0.$$

Proof. Without loss of generality, suppose that $x_0 \in N(x^*, \epsilon)$. By Lemma 4.1 it follows that (H1) and (H2) hold. By the proof of Theorem 3.1, as long as

$$\alpha \leq \frac{(1-\mu)L_0}{2\max(L, M_0)},$$

we have

$$\frac{f_k - f(x_k + p_k(\alpha))}{q_k(0) - q_k(y_k(\alpha))} - 1 \geq -\frac{2\max(L, M_0)}{L_0}\alpha.$$

Therefore,

$$\frac{f_k - f(x_k + p_k(\alpha))}{q_k(0) - q_k(y_k(\alpha))} \geq \mu,$$

which shows that there exists η_0:

$$0 < \eta_0 \leq \frac{(1-\mu)L_0}{2\max(L, M_0)}$$

such that $\alpha_k \geq \eta_0$. The proof is finished. □

Theorem 4.1. *If the conditions of Lemma* 4.2 *hold, then* $\{x_k\}$ *converges to* x^* *at least* R*-linearly.*

Proof. By the proof of Theorem 3.1 and Lemma 4.2, and noting Lemmas 2.2 and 4.1, we have

$$\begin{aligned} f_k - f_{k+1} &\geq \mu[q_k(0) - q_k(p_k)] \geq \frac{\mu\alpha_k}{2L_k}\|g_k\|^2 \\ &\geq \frac{\mu\eta_0}{2\max(L, M_0)}\|g_k\|^2 = \eta\|g_k\|^2 \geq \eta m'^2\|x_k - x^*\|^2 \\ &\geq \frac{2\eta m'^2}{M'}(f_k - f^*), \end{aligned}$$

where

$$\eta = \frac{\mu\eta_0}{2\max(L, M_0)}.$$

By setting

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

we can prove that $\theta < 1$. In fact, since $m' \leq L \leq \max(L, M_0)$ and $\eta_0 \leq 1$, by the definition of η, we obtain

$$\theta^2 = \frac{2m'^2\eta}{M'} \leq \frac{2m'^2\mu\eta_0}{2\max(L, M_0)M'} \leq \mu < 1.$$

By setting

$$\omega = \sqrt{1 - \theta^2},$$

(obviously $\omega < 1$), we obtain that

$$\begin{aligned} f_{k+1} - f^* &\leq \big(1 - \theta^2\big)\big(f_k - f^*\big) \\ &= \omega^2\big(f_k - f^*\big) \leq \dots \\ &\leq \omega^{2(k-k')}\big(f_{k'+1} - f^*\big). \end{aligned}$$

By Lemma 4.1 we have

$$\begin{aligned} \|x_{k+1} - x^*\|^2 &\leq \frac{2}{m'}\big(f_{k+1} - f^*\big) \\ &\leq \omega^{2(k-k')}\frac{2\big(f_{k'+1} - f^*\big)}{m'}, \end{aligned}$$

and thus,

$$\|x_{k+1}-x^*\| \le \omega^{k-k'}\sqrt{\frac{2(f_{k'+1}-f^*)}{m'}}, \quad \text{i.e.,} \quad \|x_k-x^*\| \le \omega^k\sqrt{\frac{2(f_{k'+1}-f^*)}{m'\omega^{2(k'+1)}}}.$$

We finally have

$$\lim_{k\to\infty} \|x_k - x^*\|^{1/k} \le \omega < 1,$$

which shows that $\{x_k\}$ converges to x^* at least R-linearly. □

5 Numerical results

We choose the following numerical examples from [2, 9, 14] to test the new conjugate gradient method.

Problem 1. Penalty function I (problem (23) in [14])

$$f(x) = \sum_{i=1}^{n} 10^{-5}(x_i-1)^2 + \left[\left(\sum_{i=1}^{n} x_i^2\right) - \frac{1}{4}\right]^2, \qquad [x_0]_i = i.$$

Problem 2. Variable dimensioned function (problem (25) in [14])

$$f(x) = \sum_{i=1}^{n}(x_i-1)^2 + \left[\sum_{i=1}^{n} i(x_i-1)\right]^2 + \left[\sum_{i=1}^{n} i(x_i-1)\right]^4,$$

$$[x_0]_i = 1 - i/n.$$

Problem 3. Trigonometric function (problem (26) in [14])

$$f(x) = \sum_{i=1}^{n}\left[n - \sum_{j=1}^{n}\cos(x_j) + i(1-\cos(x_i)) - \sin(x_i)\right]^2,$$

$$[x_0]_i = 1/n.$$

Problem 4. A penalty function (problem (18) in [2])

$$f(x) = 1 + \sum_{i=1}^{n} x_i + 10^3\left(1 - \sum_{i=1}^{n} 1/x_i\right)^2 + 10^3\left(1 - \sum_{i=1}^{n} i/x_i\right)^2,$$

$$[x_0]_i = 1.$$

Problem 5. Extended Rosenbrock function (problem (21) in [14])

$$f(x) = \sum_{i=1}^{n} \left[100(x_{2i} - x_{2i-1})^2 + (1 - x_{2i-1})^2\right], \quad [x_0]_{2i-1} = -1.2,\ [x_0]_{2i} = 1.$$

Problem 6. Penalty function II (modification of problem (24) in [14])

$$\begin{aligned} f(x) &= (x_1 - 0.2)^2 + 10^{-5} \sum_{i=2}^{n} \left[\exp\left(\frac{x_i}{m}\right) + \exp\left(\frac{x_{i-1}}{m}\right) - y_i\right]^2 \\ &\quad + 10^{-5} \sum_{i=n+1}^{2n-1} \left[\exp\left(\frac{x_{i-n+1}}{m}\right) - \exp\left(\frac{-1}{m}\right)\right]^2 + \left(\left[\sum_{i=1}^{n} (n - i + 1) x_i^2\right] - 1\right)^2, \\ y_i &= \exp\left(\frac{i}{m}\right) + \exp\left(\frac{i-1}{m}\right), \qquad [x_0]_i = 0.5,\ m = \frac{n}{10}. \end{aligned}$$

Problem 7. Brown almost linear function (problem (27) in [14])

$$f(x) = \sum_{i=1}^{n-1} \left[x_i + \sum_{j=1}^{n} -(n+1)\right]^2 + \left[\left(\prod_{i=1}^{n}\right) - 1\right]^2, \qquad [x_0]_i = 0.5.$$

Problem 8. Linear function-rank 1 (problem (33) in [14], with modified initial values)

$$f(x) = \sum_{i=1}^{m} \left[i\left(\sum_{j=1}^{n} j x_j\right) - 1\right]^2 \quad (m \geq n), \qquad [x_0]_i = 1/i.$$

In the numerical experiment, we set the parameters $\mu = 0.013$, $\rho = 0.5$, $L_0 = 0.00001$ and $M_0 = 10^{30}$. We use Matlab 6.1 to program the procedure and stop criterion is

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

The numerical results are summarized in Table 1. Strong Wolfe line search is used in the traditional conjugate gradient methods such as FR, PRP, CD, DY, HS and LS.

Strong Wolfe line serach. α_k is defined by

$$f(x_k + \alpha d_k) - f_k \leq \mu \alpha g_k^T d_k, \tag{30}$$

T	n	NM	PRP	HS	FR	CD	DY	LS
1	10^4	29/63	37/78	17/58	37/78	37/78	37/78	37/78
	5000	29/55	37/75	40/78	37/75	37/75	37/75	37/75
2	10^4	33/72	fail	fail	37/128	37/128	37/128	fail
	5000	28/64	fail	fail	37/121	37/121	37/121	fail
3	1000	19/38	23/42	35/74	19/40	29/50	25/46	24/45
	500	24/40	29/49	fail	30/49	36/56	27/46	31/51
4	10^4	32/76	39/96	fail	38/80	38/80	38/80	fail
	5000	23/55	40/85	37/76	41/83	39/79	40/80	fail
5	10^4	28/52	fail	36/87	37/69	32/58	33/62	35/72
	5000	21/48	fail	27/67	32/72	30/69	26/57	27/71
6	10^4	25/53	29/66	28/61	28/53	26/64	28/53	22/78
	5000	18/45	26/74	27/72	24/69	29/68	27/83	22/72
7	10^4	26/51	fail	fail	28/60	28/72	28/72	fail
	5000	21/51	fail	fail	25/63	fail	24/58	fail
8	10^4	28/69	35/76	fail	34/72	33/73	32/69	fail
	5000	18/48	26/57	27/67	23/62	29/35	28/72	27/46
CPU	–	178s	–	–	336s	>275s	234s	–

Table 1 – Number of iterations and functional evaluations.

and

$$|g(x_k + \alpha d_k)^T d_k| \leq -\sigma g_k^T d_k, \tag{31}$$

in which

$$\mu = 0.25 \quad \text{and} \quad \sigma = 0.75.$$

In Table 1, a pair of numbers denote the number of iterations and functional evaluations. The symbol "fail" means that the corresponding conjugate gradient method fails in solving the problem. "CPU" denotes the total CPU time of the corresponding algorithm for solving all the problems. It can be seen from Table 1 that the new nonlinear conjugate gradient method (NM) is effective in practical computation and superior (total CPU time (seconds)) to other similar methods in many situations. Moreover, PRP, HS and LS may fail to converge in solving some problems, while NM always converges in a stable manner when solving the mentioned problems. The new method has the strong convergence property and is more stable than FR, CD and DY conjugate gradient methods.

Numerical results also show that the proposed new method has the best numerical performance in practical computation. Meanwhile, the Lischitz constant estimation of the derivative of objective functions plays an important role in the new method.

6 Conclusion

In this paper, we presented a new nonlinear conjugate gradient method with strong convergence for unconstrained minimization problems. The new method can generate an adequate trust region radius automatically at each iteration and have global convergence and linear convergence rate under some mild conditions. Numerical results showed that the new conjugate gradient method is effective in practical computation and superior to other similar conjugate gradient methods in many situations.

Acknowledgements. The authors are very grateful to the referees and the editor for their valuable comments and suggestions that greatly improved the paper.

REFERENCES

[1] J. Barzilai and J.M. Borwein, *Two point step size gradient methods*. IMA J. Numer. Anal., **8** (1988), 141–148.

[2] A.R. Conn, N.I.M. Gould and P.L. Toint, *Testing a class of methods for solving minimization problems with simple bounds on the variables*. Math. Comput., **50** (1998), 399–430.

[3] Y.H. Dai, *Conjugate gradient methods with rmijo-type line seaches*. Acta Mathematicae Applicatae Sinica, English Series, **18** (2002), 123–130.

[4] Y.H. Dai, J.Y. Han, G.H. Liu, D.F. Sun, H.X. Yin and Y. Yuan, *Convergence properties of nonlinear conjugate gradient methods*. SIAM J. Optim., **10** (2000), 345–358.

[5] Y.H. Dai and Y. Yuan, *Convergence properties of the conjugate descent method*. Adv. Math., **25**(6) (1996), 552–562.

[6] Y.H. Dai and Y. Yuan, *Nonlinear Conjugate Gradient Methods*. Shanghai Science and Technology Press, Shanghai, 2000.

[7] Y.H. Dai and L.Z. Liao, *R-linear convergence of the Barzilai and Borwein gradient method*. IMA J. Numer. Anal., **22** (2002), 1–10.

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

[9] E.D. Dolan and J.J. Moré, *Benchmarking optimization software with performance profiles*. Math. Program., **91**(2) (2002), Ser. A, 201–213.

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

[11] R. Fletcher, *Practical Methods of Optimization, Vol. 1: Unconstrained Optimization [M]*. John Wiley & Sons, New York, 1987.

[12] M.R. Hestenes and E.L. Stiefel, *Methods of conjugate gradients for solving linear systems*. J. Res. Nat. Bur. Standards, **49** (1952), 409–436.

[13] Y. Liu and C. Storey, *Efficient generalized conjugate gradient algorithms: I. Theory*. J. Optim. Theory Appl., **69** (1991), 129–137.

[14] J. Moré, B. Garbow and K. Hillstrom, *Testing unconstrained optimization software*. ACM Transactions on Mathematical Software, **7** (1981), 17–41.

[15] E. Polak and G. Ribiére, *Note sur la convergence de directions conjugées*. Rev. Francaise Informat Recherche Opertionelle, 3e Année, **16** (1969), 35–43.

[16] B.T. Polyak, *The conjugate gradient method in extremem problems*. USSR Comp. Math. and Math. Phys., **9** (1969), 94–112.

[17] S. Sanmtias and E. Vercher E, *A generalized conjugate gradient algrithm*. J. Optim Theory Appl., **98** (1998), 489–502.

[18] Z.J. Shi, *Nonlinear conjugate gradient method with exact line search (in Chinese)*. Acta Math. Sci. Ser. A Chin. Ed., **24**(6) (2004), 675–682.

[19] Y. Yuan and J. Stoer, *A subspace study on conjugate gradient algorithms*. Z. Angew Math. Mech., **75** (1995), 69–77.

[20] Y. Yuan and W.Y. Sun, *Optimization Theory and Methods*. Science Press, Beijing, 1997.