Applied Mathematics & Information Sciences
An International Journal

A new class of Conjugate Gradient Methods with extended Nonmonotone Line Search

Hailin Liu[1] *and Xiaoyong Li*[2]

[1] School of Computer Science,Guangdong Polytechnic Normal University,Guangzhou, Guangdong 510665, P. R. China
[2] Laboratoire Collisions Agrgats Ractivit, Universit Paul Sabatier, 31062 Toulouse Cedex 09, France

Received: Apr 17, 2011; Revised Jul 21, 2011; Accepted Aug 4, 2011
Published online: 1 January 2012

Abstract: In this paper, we propose a new nonlinear conjugate gradient method for large-scale unconstrain optimization which possesses the following properties:(i)the sufficient descent condition $-g_k^T d_k \geq \frac{7}{8}\|g_k\|^2$ holds without any line searches;(ii)With exact line search, this method reduces to a nonlinear version of the Liu-Storey conjugate gradient scheme.(iii)Under some assumption, global convergence of this method is proved with a new nonmonotone line search.Preliminary numerical results show that this method is very efficient.

Keywords: Conjugate gradient, Sufficient descent, Hybrid method, Unconstrained optimization.

1. Introduction

Consider

$$\min_{x\in R^n} f(x) \tag{1}$$

where $f : R^n \longrightarrow R$ is a smooth nonlinear function and its gradient g is avaiable.Nonlinear conjugate gradient method is well suited for solving large scale problems,its iterative formula is given by

$$x_{k+1} = x_k + \alpha_k d_k \tag{2}$$

$$d_k = \begin{cases} -g_k, & for\ k = 1; \\ -g_k + \beta_k d_{k-1}, & for\ k \geq 2, \end{cases} \tag{3}$$

where $g_k = \nabla f(x_k), d_k$ is the search direction,α_k is a step-size obtained by a one-dimensional line search and β_k is a scalar.There are many formulas have been proposed to compute the scalar β_k for α_k is not the exact one-dimensional minimizer in practice and f is not a quadratic function.Among them,four well-known formulas for β_k are called the Hestense-Stiefel (HS)([1]), Flether-Reeves (FR) ([2]), Polak-Ribi. e olyak (PRP)([3]), Conjugate-Descent (CD)([4]), Liu-Storey(LS) ([5]), Dai-Yuan(DY)([6]) and Hager-Zhang (HZ)([7]) formulas are given by

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} \tag{4}$$

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \tag{5}$$

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} \tag{6}$$

$$\beta_k^{CD} = \frac{\|g_k\|^2}{-g_{k-1}^T d_{k-1}} \tag{7}$$

$$\beta_k^{LS} = \frac{g_k^T y_{k-1}}{-g_{k-1}^T d_{k-1}} \tag{8}$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} \tag{9}$$

$$\beta_k^{HZ} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - 2 g_k^T d_{k-1} \frac{||y_{k-1}||^2}{(d_{k-1}^T y_{k-1})^2} \tag{10}$$

respectively ,where $y_{k-1} = g_k - g_{k-1}$ and $\|.\|$ means the Euclidean norm.

In the already-existing convergence analysis and implementations of the conjugate gradient method,the extended strong Wofle line search,namely

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k \tag{11}$$

* Corresponding author: e-mail: lklhl@sina.com

$$\sigma_1 g_k^T d_k \leq g(x_k + \alpha_k d_k)^T d_k \leq -\sigma_2 g_k^T d_k \quad (12)$$

the extended nonmonotone line search,namely (1.12) and

$$f(x_k + \alpha_k d_k) - \max_{0 \leq j \leq m(k)} f(x_{k-j}) \leq \delta \alpha_k g_k^T d_k \quad (13)$$

where $m(k) = \min\{m(k-1)+1, M_0\}, m(0) = 1, 0 < \delta \leq \sigma_1 < 1, 0 < \sigma_2 < 1$ and $M_0 \in N$.

In addition,the sufficient descent condition ,namely,

$$-g_k^T d_k \geq c\|g_k\|^2 \quad (14)$$

has often been used in the literature to analyze the global convergence of conjugate gradient methods with inexact line searches,where c is a positive constant.

The convergence behaviors of (1.4),(1.5),(1.6)and (1.7) with some line search conditions have been studied by many authors for many years.Al-Baali [8] proved the global convergence of FR method with the strong Wolfe line search. Liu et al. [9] and Dai and Yuan [10] extended this result to $\sigma = \frac{1}{2}$. Although, in practical computation, the PRP method is generally believed to be the most efficient conjugate gradient method. However, Powell[11] constructed a counter example and showed that the PRP method can circle infinitely without approaching the solution, which implies that this method is not globally convergent for general functions.But the PRP+($\beta_k{}^{PRP+} = max\{0, \beta_k^{PRP}\}$) method with the Wolfe line search is globally convergent when the sufficient descent condition (1.11) is given,the HS method is very familiar with the PRP method.The DY method with the Wolfe and the strong Wolfe line search is globally convergent without the descent condition,but the descent property of the DY method depends on line search or convexity of the objective function.

In[12], the authors propose a nonmonotone Newton method, and analyze its convergence. Lucidi and Roma [15] present a nonmonotone algorithm of FR method in 1995.G. H. Liu, L. L. Jing, L. X. Han,and D. Han [14] introduce a class of nonmonotone conjugate gradient methods, this class of nonmonotone conjugate gradient methods is proved to be globally convergent when it is applied to solve unconstrained optimization problems with convex objective functions.In our paper we presented a new class of nonmonotone conjugate gradient methods which is globally convergent when it is applied to solve unconstrained optimization problems with general objective functions.

Motivated by the observation of the above ideas, we design our new conjugate gradient method as follows:

$$x_{k+1} = x_k + \alpha_k d_k \quad (15)$$

$$d_k = \begin{cases} -g_k, & for\ k = 1; \\ -g_k + \beta_k d_{k-1}, & for\ k \geq 2, \end{cases} \quad (16)$$

where

$$\beta_k^N = \frac{g_k^T y_{k-1}}{-g_{k-1}^T d_{k-1}} - 2 g_k^T d_{k-1} \frac{||y_{k-1}||^2}{(-g_{k-1}^T d_{k-1})^2} \quad (17)$$

with exact line search, we can get that $g_k^T d_{k-1} = 0$,so the N method reduce to LS method.the stepsize α_k satisfies the following new nonmonotone line search conditions

$$\begin{aligned} &\delta \alpha_k g_k^T d_k \\ &\geq f(x_k + \alpha_k d_k) - \lambda \max_{0 \leq j \leq m(k)} f(x_{k-j}) \\ &\quad -(1-\lambda) \min_{0 \leq j \leq m(k)} f(x_{k-j}) \end{aligned} \quad (18)$$

$$\sigma_1 g_k^T d_k \leq g(x_k + \alpha_k d_k)^T d_k \leq -\sigma_2 g_k^T d_k \quad (19)$$

where $m(k) = \min\{m(k-1)+1, M_0\}, m(0) = 1, 0 < \delta \leq \sigma_1 < 1, 0 < \sigma_2 < 1, 0 \leq \lambda \leq 1$ and $M_0 \in N$.The new nonmonotone line search can be viewed as some kind of convex combination of the extended strong Wofle line search and the extended nonmonotone line search, when $\lambda = 0$ the new nonmonotone line search reduce to the extended strong Wofle line search, and when $\lambda = 1$ the new nonmonotone line search reduce to the extended nonmonotone line search.

This paper is organized as follows.We will present a new algorithm (Algorithm 2.3), and the sufficient descent property (1.14) of Algorithm 2.2 is also given in the next section.In section 3 the global convergence results of the N method are established.At last the preliminary numerical results are reported.

2. New Algorithm

Throughout our paper, we assume that $g_k \neq 0$ for all k, for otherwise a stationary point has been found.Furthermore, in order to establish the global convergence result for the new algorithm, we give the following basic assumption on the objective function.

Assumption 2.1(i)The level set $\Omega = \{x \in R^n : f(x) \leq f(x_1)\}$is bounded, where x_1 is the starting point.(ii)$f(x)$ is strongly convex and Lipschitz continuous on Ω.

If f satisfies Assumption 2.1(i) and (ii), we can get that

$$\|g(x)\| \leq \bar{\gamma},\ for\ all\ x \in \Omega \quad (1)$$

where $\bar{\gamma}$ is a positive constant.

If f satisfies Assumption 2.1(ii), we can get that in some neighborhood $\mathcal{N}$ of Ω, f is differentiable and its gradient g is Lipschitz continuous, namely, there exists constants $L > 0$and $\mu > 0$such that for any $x, x' \in \mathcal{N}$

$$\|g(x) - g(x')\| \leq L\|x - x'\| \quad (2)$$

and

$$\mu\|x - x'\|^2 \leq (g(x) - g(x'))^T (x - x') \quad (3)$$

Now we give the following theorem, which illustrates that the formula (1.10) possesses the sufficient descent condition without any line searches.

Theorem 2.2 Consider any method (1.2) and (1.3), where $\beta_k = \beta_k^N$ .Then for all $k \geq 1$

$$-g_k^T d_k \geq \frac{7}{8}\|g_k\|^2 \quad (4)$$

Proof.Since $d_0 = -g_0$,we have $-g_0^T d_0 = \|g_0\|^2$,which satisfies (2.4).Multiplying (1.3) by g_k^Twe have

$$-g_k^T d_k = \|g_k\|^2 - \beta_k g_k^T d_{k-1} \quad (5)$$

when $\beta_k = \beta_k^N$ we can get that

$$\begin{aligned}&-g_k^T d_k = \|g_k\|^2 - g_k^T d_{k-1}(\tfrac{g_k^T y_{k-1}}{-g_{k-1}^T d_{k-1}}\\ &-2g_k^T d_{k-1}\tfrac{||y_{k-1}||^2}{(-g_{k-1}^T d_{k-1})^2})\\ &= (\|g_k\|^2(-g_{k-1}^T d_{k-1})^2 - g_k^T d_{k-1} g_k^T y_{k-1}(- g_{k-1}^T d_{k-1})\\ &+2(g_k^T d_{k-1})^2||y_{k-1}||^2)/(-g_{k-1}^T d_{k-1})^2\end{aligned} \quad (6)$$

we apply the inequality

$$u^T v \leq \frac{1}{2}(||u||^2 + ||v||^2)$$

to the second term in (2.6) with

$$u = \frac{1}{2} g_k(- g_{k-1}^T d_{k-1}), v = 2(g_k^T d_{k-1})y_{k-1}$$

Therefore , we can get that (2.4) is true for all $k \in N$.

Now we can present a new descent conjugate gradient method as follows:

Algorithm 2.3
step1:Given $x_1 \in R^n, \varepsilon \geq 0, 0 < \delta \leq \sigma_1 < 1, 0 < \sigma_2 < 1, 0 \leq \lambda \leq 1, M_0 \in N$, set $d_1 = -g_1, k := 1$, if $\|g_1\| \leq \varepsilon$, then stop.
step2:Find a $\alpha_k \geq 0$ by (1.18) and (1.19).
step3:Let $x_{k+1} = x_k + \alpha_k d_k$ and $g_{k+1} = g(x_{k+1})$ If $\|g_{k+1}\| \leq \varepsilon$, then stop.
step4:Compute $\beta_k = \beta_k^N$ by the formula (1.17) and generate d_{k+1} by (1.16).
step5:Set k:=k+1, go to step2.

Lemma 2.4 Suppose that Assumption 2.1 holds and α_k is obtained by the new Nonmonotone line search (1.18) and (1.19).Then there exists $c_1 > 0$, such that

$$\|s_k\| \geq \frac{c_1(-g_k^T d_k)}{\|d_k\|} \quad (7)$$

where $s_k = x_{k+1} - x_k$.

Proof.From Assumption 2.1 we can get that g is Lipschitz continuous in some neighborhood $\mathcal{N}$ of Ω, namely, there exists a constant $L > 0$, such that for any $x, x' \in \mathcal{N}$

$$\|g(x) - g(x')\| \leq L\|x - x'\| \quad (8)$$

that is $\|y_k\| \leq L\|s_k\|$, so we have

$$y_k^T d_k \leq L\|s_k\|\|d_k\| \quad (9)$$

and from (1.19) we can get that

$$y_k^T d_k = d_k^T(g_{k+1} - g_k) \geq (1 - \sigma_1)(-g_k^T d_k) \quad (10)$$

Thus, from (2.9) and (2.10) we obtain

$$\|s_k\| \geq \frac{1 - \sigma_1}{L}\frac{-g_k^T d_k}{\|d_k\|} \quad (11)$$

let $c_1 = \frac{1-\sigma_1}{L}$ and from (2.11), we obtain (2.7).Therefore, our proof is complete.

Lemma 2.5 Suppose that assumption 2.1 holds, α_k is given by (1.18) and (1.19), and $\beta_k = \beta_k^N$. Then there exists a positive constant $M = \frac{L^2}{\mu}$such that

$$\frac{\|y_k\|^2}{y_k^T s_k} \leq M \quad (12)$$

Proof.By the convexity assumption, we have

$$y_k^T d_k = d_k^T(g_{k+1} - g_k) \geq \mu\alpha_k\|d_k\|^2 \quad (13)$$

and from the Lipschitz continuity (2.8), we can get that

$$\begin{aligned}&||y_k|| = ||g_{k+1} - g_k|| = \|g(x_{k+1}) - g(x_k)\|\\ &\leq L||x_{k+1} - x_k|| = L\alpha_k\|d_k\|\end{aligned} \quad (14)$$

Utilizing (2.13) and (2.14), we can get

$$\frac{\|y_k\|^2}{y_k^T s_k} \leq \frac{L^2\alpha_k^2\|d_k\|^2}{\alpha_k y_k^T d_k} \leq \frac{L^2\alpha_k^2\|d_k\|^2}{\mu\alpha_k^2\|d_k\|^2} = \frac{L^2}{\mu} = M \quad (15)$$

which completes the proof.

Lemma 2.6 Assume that the following inequality holds for all k

$$0 < m_1 \leq ||g_k|| \leq m_2 \quad (16)$$

and α_k is given by (1.18) and (1.19), then
(1)there exist a positive constant $b > 1$such that

$$|\beta_k| \leq b \quad (17)$$

(2))there exist a positive constant λ, when $||y_{k-1}|| \leq \lambda$ we have $|\beta_k| \leq \epsilon$ for any $\epsilon > 0$.

Proof.When $\beta_k = \beta_k^N$, from (2.4) and (2.16) we can get

$$\begin{aligned}&|\beta_k| \leq |\tfrac{g_k^T y_{k-1}}{-g_{k-1}^T d_{k-1}}| + |2g_k^T d_{k-1}\tfrac{||y_{k-1}||^2}{(-g_{k-1}^T d_{k-1})^2}|\\ &\leq |\tfrac{g_k^T y_{k-1}}{-g_{k-1}^T d_{k-1}}|\\ &+2|\sigma_2(-g_{k-1}^T d_{k-1})\tfrac{||y_{k-1}||^2}{(-g_{k-1}^T d_{k-1})^2}|\\ &\leq \tfrac{2m_2^2}{-g_{k-1}^T d_{k-1}} + \tfrac{8\sigma_2 m_2^2}{-g_{k-1}^T d_{k-1}}\\ &\leq \tfrac{(16+64\sigma_2)m_2^2}{7m_1^2}\end{aligned} \quad (18)$$

let $b=\frac{(16+64\sigma_2)m_2^2}{7m_1^2}$, obviously we have $b>1$ because of $m_2\geq m_1>0$.
let $\lambda=\frac{7m_1^2}{(8+32\sigma_2)m_2}\epsilon$, when $||y_{k-1}||\leq\lambda$, we can get

$$\begin{aligned}|\beta_k| &\leq |\frac{g_k^T y_{k-1}}{-g_{k-1}^T d_{k-1}} - 2g_k^T d_{k-1}\frac{||y_{k-1}||^2}{(-g_{k-1}^T d_{k-1})^2}| \\ &\leq (|\frac{||g_k||}{-g_{k-1}^T d_{k-1}}| \\ &+|2\sigma(-g_{k-1}^T d_{k-1})\frac{||y_{k-1}||}{(-g_{k-1}^T d_{k-1})^2}|)||y_{k-1}|| \\ &\leq \frac{(8+32\sigma_2)m_2}{7m_1^2}||y_{k-1}||\leq\epsilon\end{aligned} \tag{19}$$

This completes the proof.
Lemma 2.7 Suppose that Assumption 2.1 (i) holds and α_k is obtained by the new Nonmonotone line search (1.18) and (1.19), denote $\xi_k=\delta\alpha_k g_k^T d_k$, then $\{f(x_k)\}$ is nonincreasing and

$$\lim_{k\to\infty}\xi_{l(k+1)-1}=0 \tag{20}$$

where

$$l(k)=\max\{i|0\leq k-i\leq m(k), f(x_i)=\max_{0\leq j\leq m(k)} f(x_{k-j})\}. \tag{21}$$

Proof.From (1.18) and (2.21) we can get that

$$\begin{aligned}f(x_{k+1}) &\leq \lambda f(x_{l(k)})+(1-\lambda)\min_{0\leq j\leq m(k)} f(x_{k-j})+\xi_k \\ &\leq f(x_{l(k)})+\xi_k\end{aligned} \tag{22}$$

because $\xi_k<0$, we can obtain

$$f(x_{k+1})\leq f(x_{l(k)})$$

from (2.21) and $m(k)\leq m(k-1)+1$, for all k.we can get that

$$\begin{aligned}f(x_{l(k)}) &\leq \max_{0\leq j\leq m(k-1)+1} f(x_{k-j}) \\ &\leq \max\{\max_{0\leq j\leq m(k-1)} f(x_{k-1-j}), f(x_k)\} \\ &= \max\{f(x_{l(k-1)}), f(x_k)\} \\ &= f(x_{l(k-1)}), k=1,2...\end{aligned} \tag{23}$$

therefore $\{f(x_k)\}$ is nonincreasing . Because $l(k+1)-1k+1-m(k+1)-1k-M_0$,we have

$$f(x_{l(l(k+1))-1})\leq f(x_{l(k-M_0)}) \tag{24}$$

from the above ineaquation and (2.22) we can obtain

$$\begin{aligned}&f(x_{l(k+1)})\leq f(x_{l(l(k+1))-1})+\xi_{l(k+1)-1} \\ &\leq f(x_{l(k-M_0)})+\xi_{l(k+1)-1}\end{aligned}$$

hence, we have

$$0\leq-\xi_{l(k+1)-1}\leq f(x_{l(k-M_0)})-f(x_{l(k+1)}) \tag{25}$$

by (2.25) when the Assumption 2.1 (i) holds, we have

$$\lim_{k\to\infty}\xi_{l(k+1)-1}=0$$

we complete the proof.
Lemma 2.8 Suppose that α_k is given by (1.18) and (1.19), and $\beta_k=\beta_k^N$. Then $\{l(k)\}$ is increasing.
Proof.Assume that

$$l(k+1)<l(k) \tag{26}$$

then we have

$$k+1\geq l(k)>l(k+1)\geq k+1-m(k+1) \tag{27}$$

by the definition of $l(k+1)$ and (2.23), we can obtain

$$f(x_{l(k+1)})\geq f(x_{l(k)}) \tag{28}$$

but from the Lemma 2.7we have

$$f(x_{l(k+1)})\leq f(x_{l(k)}) \tag{29}$$

hence, by the definition of $l(k+1)$and (2.29), we have

$$l(k+1)\geq l(k) \tag{30}$$

which is contradictory to (2.29).Hence
$l(k)\leq l(k+1)$, namely $\{l(k)\}$is increasing.

3. Convergence analysis

Theorem 3.1 Suppose that Assumption 2.1 holds and α_k is obtained by the new Nonmonotone line search (1.18) and (1.19), Consider any iteration method of the form (1.15) and (1.16), where $\beta_k=\beta_k^N$. Then

$$\lim_{k\to\infty}\inf||g_k||=0 \tag{1}$$

Proof.If (2.1) is not true, then there exists a constant γ such that $0<\gamma\leq||g_k||\leq\overline{\gamma}$ for all k.From (1.16) and Lemma 2.6 we have

$$\begin{aligned}||d_{l(k+1)-1}|| &\leq ||g_{l(k+1)-1}||+|\beta_{l(k+1)-1}|||d_{l(k+1)-2}|| \\ &\leq \overline{\gamma}+b||d_{l(k+1)-2}||\leq ... \\ &\leq \overline{\gamma}\sum_{j=0}^{l(k+1)-l(k)-3} b^j \\ &\quad +b^{l(k+1)-l(k)-2}||d_{l(k)+1}|| \\ &\leq \overline{\gamma}\sum_{j=0}^{l(k+1)-l(k)-2} b^j \\ &\quad +b^{l(k+1)-l(k)-2}|\beta_{l(k)+1}|||d_{l(k)}||\end{aligned} \tag{2}$$

because $l(k+1)-l(k)\leq k+1-[k-m(k)]<M_0+2$,we can get

$$||d_{l(k+1)-1}|| \leq \overline{\gamma}\sum_{j=0}^{M_0} b^j + b^{M_0}|\beta_{l(k)+1}|||d_{l(k)}|| \\ = q_1 + q_2|\beta_{l(k)+1}|||d_{l(k)}|| \quad (3)$$

where

$$q_1 = \overline{\gamma}\sum_{j=0}^{M_0} b^j, q_2 = b^{M_0}.$$

from Lemma 2.5 and (1.19) we have

$$0 \leq ||y_k|| \leq (My_k^T s_k)^{\frac{1}{2}} \leq (M(1-\sigma_2)(-\alpha_k g_k^T d_k))^{\frac{1}{2}}$$

From Lemma 2.6 we note that $\epsilon = \frac{1}{q_2 b^2}$,there exists an innegative integer k_0 when $k \geq k_0$ we have

$$|\beta_{l(k)}| < \epsilon$$

thus we can get

$$||d_{l(k+1)-1}|| \leq q_1 + \frac{1}{b^2}||d_{l(k)}|| \\ \leq q_1 + \frac{||g_{l(k)}|| + |\beta_{l(k)}|||d_{l(k)-1}||}{b^2} \\ \leq q_1 + \frac{\overline{\gamma} + b||d_{l(k)-1}||}{b^2} \leq q_3 + \frac{||d_{l(k)-1}||}{b}$$

for all $k \geq k_0$, where $q_3 = q_1 + \frac{\overline{\gamma}}{b}$.Thus, we have a recursive equation which leads to

$$||d_{l(k+1)-1}|| \leq q_3 + \frac{||d_{l(k)-1}||}{b} \\ \leq q_3 + \frac{q_3 + \frac{||d_{l(k-1)-1}||}{b}}{b} \\ = q_3 + \frac{q_3}{b} + \frac{||d_{l(k-1)-1}||}{b^2} \leq ... \\ \leq q_3\sum_{j=0}^{k-k_0}(\frac{1}{b})^j + \frac{1}{b}^{k-k_0+1}||d_{l(k_0)-1}|| \\ \leq q_3\sum_{j=0}^{\infty}(\frac{1}{b})^j + ||d_{l(k_0)-1}|| \\ \leq q_3\frac{b}{b-1} + ||d_{l(k_0)-1}|| \quad (4)$$

Applying $l(k) \geq k - m(k) \geq k - M_0$ and Lemma 2.8 we can assume that $l(i) - 1 \leq j \leq l(i+1) - 1, i \geq k_0 + 2$, for all $j \geq l(k_0+2) - 1$, thus we have

$$||d_j|| \leq ||g_j|| + |\beta_j|||d_{j-1}|| \\ \leq \overline{\gamma} + b(||g_{j-1}|| + |\beta_{j-1}|||d_{j-2}||) \quad (5) \\ \leq ... \leq \overline{\gamma}\sum_{t=0}^{j-l(i)} b^t + b^{j-l(i)+1}||d_{l(i)-1}||$$

Therefore, from

$$j - l(i) + 1 \leq [l(i+1) - 1] - l(i) + 1 \\ \leq i + 1 - [i - m(i)] \leq M_0 + 1$$

and (3.5), we have

$$||d_j|| \leq \overline{\gamma}\sum_{t=0}^{j-l(i)} b^t + b^{M_0+1}||d_{l(i)-1}|| \quad (6)$$

From (3.4) and (3.6), we have

$$||d_j|| \leq \overline{\gamma}\sum_{t=0}^{j-l(i)} b^t + b^{M_0+1}[q_3\frac{b}{b-1} + ||d_{l(k_0)-1}||] \quad (7)$$

for all $i - 1 \geq k_0$.By using Lemma 2.4 and Lemma 2.5 we have

$$-\xi_{l(k+1)-1} = \frac{\rho||s_{l(k+1)-1}||(-g_{l(k+1)-1}^T d_{l(k+1)-1})}{||d_{l(k+1)-1}||} \\ \geq \rho c_1\frac{(-g_{l(k+1)-1}^T d_{l(k+1)-1})^2}{||d_{l(k+1)-1}||} \quad (8) \\ \geq \rho c_1\frac{\gamma^4}{||d_{l(k+1)-1}||}$$

but from (2.20) and (3.8) we can get

$$\lim_{k\to\infty}\frac{1}{||d_{l(k+1)-1}||} = 0$$

which is contradictory to (3.7).Hence the theorem is valid.

4.Numerical experiments

In this section, we will test PRP, HZ and N conjugate methods with the new nonmonotone line search.In table 4-1 when $\varepsilon = 10^{-6}, \delta = 0.01, \sigma_1 = \sigma_2 = 0.1, \lambda = 0, M_0 = 100$, for when$\varepsilon = 10^{-6}, \delta = 0.01, \sigma_1 = \sigma_2 = 0.1, \lambda = \frac{1}{2}, M_0 = 100$ each method which with capital letters, and each method which with small letters. The problems that we test are from [15].

Table 4-1 show the computation results, where the columns of the tables have the following meanings:

Problem: the name of the test problem;

Dim: the dimension of the problem;

NI: the total number of iterations;

NF: the number of the function evaluations;

NG: the number of the gradient evaluations.

Time:the CPU total time

and the star * denotes that this result is the best one among these three methods.

Tabel 4-1

Problem	Dim	Method	NI	NF	NG
Var. dim.	1000	N*	2	12	3
		HZ	2	12	3
		PRP	2	12	3
		n	2	12	3
		hz	2	12	3
		prp	2	12	3
	5000	N*	3	22	6
		HZ	3	22	6
		PRP	3	22	6
		n	3	22	6
		hz	3	22	6
		prp	3	22	6
	10000	N*	2	12	2
		HZ	2	12	2
		PRP	2	12	2
		n	2	12	2
		hz	2	12	2
		prp	2	12	2
Penalty I	1000	N*	8	50	30
		HZ	1003	3290	1825
		PRP	29	104	72
		n	8	50	30
		hz	45	102	58
		prp	29	104	72
	5000	N*	12	54	33
		HZ	579	1882	1036
		PRP	94	332	208
		n	12	104	72
		hz	88	338	107
		prp	56	221	64
	10000	N*	16	38	17
		HZ	255	797	430
		PRP	50	185	105
		n	16	38	17
		hz	85	197	130
		prp	50	185	105
Penalty II	20	N	3332	10002	3350
		HZ	3332	10002	3356
		PRP	3332	10002	3354
		n*	321	867	342
		hz	321	868	346
		prp	321	868	346
	50	N	3304	10001	3365
		HZ	3303	10000	3365
		PRP	3303	10001	3396
		n	146	344	221
		hz*	128	322	201
		prp	128	344	221
	100	N	578	1768	631
		HZ	594	1817	650
		PRP	592	1761	634
		n*	224	644	231
		hz	244	689	260
		prp	242	688	255
Trig. metric	1000	N*	62	104	84
		HZ	76	142	123
		PRP	66	128	113
		n	62	104	84
		hz	76	142	123
		prp	66	128	113
	5000	N*	59	108	83
		HZ	67	126	111
		PRP	63	109	104
		n	59	108	83
		hz	67	126	111
		prp	63	109	104
	10000	N*	65	116	101
		HZ	81	168	142
		PRP	77	141	123
		n	65	116	101
		hz	81	168	142
		prp	77	141	123
Ext. Rosenbrock	1000	N	3249	10001	3418
		HZ	3248	10000	3421
		PRP	3273	10001	3403
		n	486	1301	518
		hz*	477	1299	526
		prp	409	1301	521
	5000	N	3249	10001	3418
		HZ	3248	10000	3421
		PRP	3273	10001	3403
		n	486	1301	521
		hz*	477	1299	526
		prp	490	1301	518
	10000	N	3249	10001	3418
		HZ	3248	10000	3421
		PRP	3249	10001	3403
		n	486	1301	521
		hz*	477	1299	526
		prp	490	1301	518
Ext. Powell sing.	1000	N	3333	10002	3320
		HZ	3333	10002	3340
		PRP	3333	10002	3338
		n*	1211	3421	1402
		hz	1211	3421	1430
		prp	1211	3421	1422
	5000	N	3333	10002	3320
		HZ	3333	10002	3340
		PRP	3333	10002	3338

		n*	1211	3421	1402
		hz	1211	3421	1430
		prp	1211	3421	1422
	10000	N	3333	10002	3320
		HZ	3333	10002	3340
		PRP	3333	10002	3338
		n*	1211	3421	1402
		hz	1211	3421	1430
		prp	1211	3421	1422
Chebyquad	200	N	2498	10001	3104
		HZ	2492	10003	3412
		PRP	2495	10000	3150
		n*	446	1330	476
		hz	521	1640	544
		prp	524	1677	532
	500	N*	2	12	1
		HZ	2	12	1
		PRP	2	12	1
		n	2	12	1
		hz	2	12	1
		prp	2	12	1
	2000	N*	2	12	1
		HZ	2	12	1
		PRP	2	12	1
		n	2	12	1
		hz	2	12	1
		prp	2	12	1
Brown & Dennis	4	N*	336	1390	363
		HZ	340	1406	367
		PRP	343	1425	373
		n*	121	367	142
		hz	144	421	211
		prp	156	331	191
Gulf research	3	N	3335	10001	3384
		HZ	3334	10001	3385
		PRP	3334	10001	3428
		n*	1222	3444	1312
		hz	1443	3866	1628
		prp	1443	3866	1628
Beale	2	N	836	2501	903
		HZ	836	2501	903
		PRP	842	2518	907
		n*	323	1003	421
		hz	334	1008	432
		prp	386	1120	449

In order to rank the iterative numerical methods, one can compute the total number of function and gradient evaluations by the formula

$$N_{total} = NF + 5 * NG \quad (1)$$

Similarly, we compare PRP+ method, MPRP method with PRP method as follows:for each problem i, compute the total numbers of function evaluations and gradient evaluations required by the evaluated methods and PRP method by formula (4.1), and denote them by $N_{total,i}(EM)$ and $N_{total,i}(PRP)$; then calculate the ratio

$$r_i(EM(j)) = \frac{N_{total,i}(EM(j))}{N_{total,i}(PRP)} \quad (2)$$

If $EM(j_0)$ method does not work for example i_0, but PRP method can work, we replace the $r_{i_0}EM(j_0)$ by a positive constant τ_1 which define as follows:

$$\tau_1 = max\{r_i(EM(j_0)) : (i, j_0)\bar{\in}S_1\} \quad (3)$$

where

$$S_1 = \{(i, j_0) : method\ j_0\ does\ not\ work\ for\ example\ i\} \quad (4)$$

If PRP method does not work for example i_0, but $EM(j_0)$ method can work, we replace the $r_{i_0}EM(j_0)$ by a positive constant τ_2 which define as follows:

$$\tau_2 = min\{r_i(EM(j_0)) : (i, j_0)\bar{\in}S_1\} \quad (5)$$

Neither PRP method nor $EM(j_0)$ method works, we define $r_{i_0}EM(j_0) = 1$.The geometric mean of these ratios for $EM(j)$ method over all the test problems isdefined by

$$r(EM(j)) = (\prod_{i\in s} r_i(EM(j))^{(1/|S|)} \quad (6)$$

where S denotes the set of the test problems and $|S|$ the number of elements in S.

According to the above rule, it is clear that $r(PRP) = 1$.The values of $r(HZ), r(N), r(hz), r(n)$ and $r(prp)$ are listed in Table 4-2.

Tabel 4-2

HZ	PRP	N	hz	n	prp
1.194	1.0	0.912	0.877	0.782	0.854

from tabel 4-2 we can see that the new method is more efficient than HZ method and PRP method.

Acknowledgments: This research supported by Natural Science Fundation of Guangdong Province No. 915100800 2000012.

References

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

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

[3] B. T. Polyak, The conjugate gradient method in extreme problems, USSR Comp. Math. Math. Phys., 9 (1969), 94-112.

[4] R. Fletcher, Practical Methods of Optimization vol. 1: Unconstrained Optimization, John Wiley

[5] Y. Liu and C. Storey, Efficient generalized conjugate gradient algorithms, Part 1: Theory, J. Optim. Theory Appl., 69 (1991), 129-137.

[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] Hager W W, Zhang H.A new conjugate gradient method with guaranteed descent and an efficient line search. SIAM J Optim. 2005, 16:170-192

[8]Al-Baali M. Descent Property and global convergence of the Fletcher-Reeves method with inexact line searches, SIMA Journal of Numerical Analysis, 5(1) (1985), 121-124.
[9]G. Liu, J. Han, H. Yin, Global convergence of the Fletcher ¨ CReeves algorithm with inexact line search, Appl. Math. JCU 10B (1995) 75¨C82.
[10]H. Dai and Y. Yuan, Convergence of the Fletcher-Reeves method under a generalized Wolfe wearch, J. Comput. Math., 2 (1996), 142¨C148.
[11]Powell M J D. Nonconvex minimization calculations and the conjugate gradient method, Lecture notes in Mathematics vol. Springer-Veriag(Berlin), 1006 (1984), 122-141.
[12] Grippo L, Lampariello F, Lucidi S. A nonmonotone line search technique for Neton¡äs methods. SIAM Journal on Numerical Analysis, 123 (1986), 707-716.
[13] Lucidi S, Roma M. Nonmonoton conjugate gradient methods for optimization, System Modeling and Optimization, Lecture Notes on Control and Information Sciences, Belin, Springer Verlag, 1995.
[14] Liu G H, Jing L L, Han D. A class of nonmonotone conjugate gradient methods for unconstrained optimization, SIAM J Optim, 101 (1999) 127-140.
[15]Jorge J M, Burton S G, Kenneth E H. Testing Unconstrained Optimization Software [J]. ACM Transactions on Mathematical Software 7 (1981) 17-41.

Hailin Liu obtained his M.Sc. degree in Applied Mathematics from the central south university of technology, china, he is a professor of school of computer science at Guangdong Polytechnic Normal University in Guangzhou, china, Liu's current research is divided into two areas:(i) optimization theory and it's application, (ii) computer science and it's application.

Xiaoyong Li is the responding author of this paper , he obtained his M.Sc. degree in Applied Mathematics from the Guangdong Polytechnic Normal University in 2011, he is a software engineer at a Information security corporation in Zhuhai city, his research interest is optimization theory and it's application.