Mathematica Aeterna, Vol. 5, 2015, no. 4, 537 - 549

A spectral conjugate gradient method under modified nonmonotone line search technique

Ke Su

College of Mathematics and information Science
Hebei University, Baoding, 071002 China

ZiXing Rong

College of Mathematics and information Science
Hebei University, Baoding, 071002 China

Abstract

In this paper,we introduce a spectral Conjugate Gradient direction, which have the properties of both HS conjugate gradient method and PRP conjugate gradient method via introducing parameter λ, Motivated by the ideas of introducing parameter, we give a modified nonmonotone line search by doing convex combination. besides we introduce a new nonmonotone line search, and propose the algorithm of nonmonontone spectral conjugate gradient method, then the convergence of new nonmonotone spectral Conjugate Gradient method is established under mild conditions. in the end, some some numerical experiments are given.

Mathematics Subject Classification: xxxxx

Keywords: spectral conjugate gradient, global convergence, unconstrained optimization, modified nonmonotone line search

1 Introduction

Consider the following unconstrained optimization problem

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

where $f : R^n \to R$ is a continuously differentiable function, R^n is an Euclidean space.

Unconstrained optimization problem is an important research topic in mathematical programming fields. There exist many methods for solving unconstrained optimization problem, these methods have research topic in many fields such as economy, management, control science, etc. It is well-known that conjugate gradient method [1] is a good method for solving the unconstrained optimization problem in management and engineering. It has the following formula for solving problem (1):

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

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

where $d_k$is a search direction, β_k is a scalar, $g_k = \nabla f(x_k)$, α_k is a positive step-size along the search direction.

In many literatures(refer to [2-4]), the main thing is to choose the scalarβ_k, which leads to many different conjugate gradient methods according to the differentβ_k. A well-known conjugate gradient method was proposed by Fletcher and Reeves[1], and some other choices as followed:

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2},$$
$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \qquad \beta_k^{CD} = -\frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}},$$
$$\beta_k^{LS} = -\frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}},$$

where $\|\cdot\|$ denotes Euclidean norm, $y_{k-1} = g_k - g_{k-1}$.

Raydan introduced the spectral gradient method (SGM) for potentially large-scale unconstrained optimization [5]. The main feature of this method is that only gradient directions are used at each line search whereas a non-monotone strategy guarantees global convergence. surprisingly, this algorithm outperforms sophisticated conjugate gradient algorithms in many problems. The numerical results in [5, 6, 7] and others suggested us that spectral gradient and conjugate gradient ideas could be combined in order to obtain even more efficient algorithms. Assume that $f : R^n \to R$ has continuous partial derivatives. The problem considered in this paper is (1) the iteration formulation for (2) is as follows:

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

where d_k is a search direction and α_k is a step length chosen to induce the value of $f(x)$ The direction is generated by

$$d_k = -\theta_k g_k + \beta_k d_k,$$

where g_k denotes $\nabla f(x_k)$and $d_0 = -\theta_0 g_0$ and θ_k is a arbitrary parameter

$$\beta_k = \frac{(\theta_k y_{k-1} + s_{k-1})^T g_k}{d_{k-1}^T y_{k-1}},$$

where $s_k = x_{k+1} - x_k = \alpha_k d_k$ and $y_k = g_{k+1} - g_k$.

In general, we denote θ_k as

$$\theta_k = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T g_{k-1}}([8])$$

Now we introduce a new spectral conjugate gradient direction[9]

$$d_k = \begin{cases} -g_0, & k = 0, \\ -\theta_k g_k + \beta_k d_{k-1}, & k \geq 1. \end{cases} \tag{2}$$

Where

$$\theta_k = 1 + \frac{\beta_k d_{k-1}^T g_k}{\|g_k\|^2} \tag{3}$$

$$\beta_k = \frac{g_k^T y_{k-1}}{(1-\lambda)\|g_{k-1}\|^2 + \lambda d_{k-1}^T y_{k-1}} \tag{4}$$

and $\lambda \in [0, 1]$

It is not difficult to find that there is difference between the one generated by formula (2)(3),(4) and previous ones. Firstly, the formula (4) combine HS conjugate gradient method with PRP conjugate gradient method together by introducing parameter λ. secondly, in the formula (3), the choice of spectral parameter ensure that the Conjugate Gradient direction d_k is descent sufficiently undepend on line search. we can give the proof as *Lemma* 3.1 in section 3.

Modified by the ideas of convex combination, in this paper, we introduced a modified monmonotone line search [10] via combining monotone Armijo line search with Armijo line search, which have both properties, the best convergence result can be obtained in case that α_k is chosen by nonmonotone when the iterates were far from the optimum, and α_k is chosen by monotone when the iterates were near an optimum, which is same as [11], and then we introduced anther new monmonotone line search we combined the new spectral conjugate gradient direction with different nonmonotone line search technique. which perform efficiently both in theory and numerical results. Furthermore, we prove that the new method with nonmonotone line search rule is globally convergent under mild conditions. Numerical experiments show that the proposed method is efficient.

The paper is organized as follows. in section 2, we give modified nonmonotone line search which is a convex combination of monotone line search and Armijo nonmonotone line search and a new nonmonotone line search proposed by Yu, besides, we give the algorithms. in section 3, we establish the global convergence of the seconds algorithm ,in section 4, all the essential parameters of its implementation is given, some numerical experiments are given.

2 A Modified Nonmonotone Line Search and Algorithms

The next task is to find a step size α_k along the search direction. The ideal line search rule is the exact one which satisfies:

$$f(x_k + \alpha_k d_k) = \min_{\alpha > 0} f(x_k + \alpha d_k)$$

in fact, the exact step size is difficult or even impossible to seek in practical computation, and thus many researchers constructed some inexact line search rule, such as Armijo rule, Goldtein rule, Wolfe rule and nonmontone line search [12]

In 1982, Chamberlain et al. in [13] proposed a watchdog technique for constrained optimization, in which some standard line search conditions were relaxed to overcome the Marotos effect. Motivated by this idea, Grippo, Lampariello and Lucidi in [12] presented a nonmonotone Armijo-type line search technique for the Newton method. The traditional line search rules require the function value descent monotonically at each iteration. It may considerably slow the rate of convergence in the intermediate stages of the minimization process, especially in the presence of the narrow curved valley. However the nonmontone line search rules are effective or even powerful at some iteration, especially when the iterates are trapped in a narrow curved valley of objective functions.

The earliest nonmonotone line search framework was developed by Grippo, Lampariello, and Lucidi in [12] for Newton′s methods. thanks to its excellent numerical exhibition, many nonmonotone techniques have been developed in recent years, To describe the nonmonotone technique, we describe the nonmonotone Armijo rule. α_k is a stepsize with $\alpha_k \geq 0$ and d_k is a search direction satisfied $g_k^T d_k \leq 0$, Let $a > 0$, $\gamma \in (0, 1)$, $\beta \in (0, 1)$and let M be a nonnegative integer. For each k , let $m(k)$ satisfies

$$m(0) = 0, 0 \leq m(k) \leq \min[m(k-1) + 1, M], for\ k \geq 1,$$

Let $\alpha_k = \beta^{p^k} a$ and p^k be the smallest nonnegative integer p such that

$$f(x_k + \alpha_k) \leq \max_{0 \leq j \leq m(k)} [f(x_{k-j})] + \gamma \beta^{p^k} a g_k^T d_k.$$

If the search is the nonmonotone Goldstein line search, so α_k should satisfied the following condition:

$$f(x_k + \alpha_k) \le \max_{0 \le j \le m(k)} [f(x_{k-j})] + \mu_1 \lambda_k g_k^T d_k,$$

$$f(x_k + \alpha_k) \le \max_{0 \ge j \le m(k)} [f(x_{k-j})] + \mu_2 \lambda_k g_k^T d_k,$$

where $0 < \mu_1 \le \mu_2 < 1$

If the search is the nonmonotone wolfe line search, so α_k should satisfied the following condition:

$$f(x_k + \alpha_k) \le \max_{0 \le j \le m(k)} [f(x_{k-j})] + \gamma_1 \alpha_k g_k^T d_k,$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \gamma_2 g_k^T d_k,$$

where $0 < \gamma_1 \le \gamma_2 < 1$

Motivated by the ideas of introducing parameter λ, we give a modified nonmonotone line search which is a convex combination of monotone line search and Armijo nonmonotone line search, when the value of μ closer to 1, the modified line search closely approximates the usual monotone line search, and as μ approachs 0, the scheme becomes more nonmonotone. the modified line search as following:

Let M be a nonnegative integer, $\mu \in [0, 1]$. For each k , let $m(k)$ satisfies

$$m(0) = 0, 0 \le m(k) \le \min[m(k-1) + 1, M], for\ k \ge 1,$$

$\alpha_k = \beta^{p_k}$ is bounded above and satisfy

$$f(x_k + \alpha_k d_k) \le \mu f(x_k) + (1 - \mu) \max_{0 \le j \le m(k)} [f(x_{k-j})] + \delta a_k g_k^T d_k.$$

Although these nonmonotone technique work well in many case, there are some drawbacks, First, a good function value generated in any iteration is essentially discard due to the max. Second, in some case, the numerical performance is very dependent on the choice of M [5.12.14] now we give a new nonmonotone line search proposed by Yu as follows: [15]

Let $\lambda \in (0, 1]$ $M \ge 1$ is a positive integer,defined $m(k) = \min[k + 1, M]$

$$\lambda_{kr} \ge \lambda,\ r = 0, 1, 2, \cdots, m(k) - 1 \quad \sum_{r=0}^{m(k)-1} \lambda_{kr} = 1,$$

Let $\alpha_k = \beta^{p_k}$ be bounded above and satisfy:

$$f(x_k + \alpha_k d_k) \le \max[f(x_k), \sum_{r=0}^{m(k)-1} \lambda_{kr} f(x_{k-r})] + \gamma \alpha_k \langle d_k, g(x_k) \rangle.$$

Algorithm 2.1

Step0. Give starting guess x_1, and some constants, $0 < \beta < 1$, $0 < \delta < 1$, $\lambda \in [0, 1]$, $\varepsilon > 0$ and $k = 1$.

Step1. Compute g_k, and $\|g_k\| \leq \varepsilon$, STOP.

Step2. Compute d_k by (2) (3),(4).

Step3 Set trial step $\alpha_k = 1$.

Step4 Set $x_{k+1} = x_k + \alpha_k d_k$.

Step5 Let M be a nonnegative integer, $\mu \in [0, 1]$. For each k , let $m(k)$ satisfies

$$m(0) = 0, 0 \leq m(k) \leq \min[m(k-1) + 1, M], for\ k \geq 1,$$

$\alpha_k \geq 0$ is bounded above and satisfy

$$f(x_k + \alpha_k d_k) \leq \mu f(x_k) + (1 - \mu) \max_{0 \leq j \leq m(k)} [f(x_{k-j})] + \delta a_k g_k^T d_k. \tag{5}$$

.

If (5) does not holds, define $\alpha_k = \alpha_k \beta$ and go to step 5

Step6. Set $k := k + 1$, and go to step 1 .

Note that the algorithm cannot cycle indefinitely in step 3.2 in fact, when $\alpha_k = 0$, $f(x_k + \alpha_k d_k) = f(x_k)$ there must exist a sufficient small α_k

$$f(x_k + \alpha_k d_k) \leq f_k + \delta \alpha_k \nabla f(x_k) d_k,$$

because of $\nabla f(x_k) d_k \leq 0$ and $0 < \delta < 1$. what is more

$$f(x_k) \leq \mu f(x_k) + (1 - \mu) \max_{0 \leq j \leq m(k)} [f(x_{k-j})],$$

So we have

$$f(x_k + \alpha_k d_k) \leq \mu f(x_k) + (1 - \mu) \max_{0 \leq j \leq m(k)} [f(x_{k-j})] + \delta \alpha_k \nabla f(x_k) d_k.$$

Algorithm 2.2

Step0. Give starting guess x_1, and some constants, $0 < \beta < 1$, $0 < \delta < 1$, $\lambda \in [0, 1]$, $\varepsilon > 0$ and $k = 1$.

Step1. Compute g_k, and $\|g_k\| \leq \varepsilon$, STOP.

Step2. Compute d_k by (2) (3),(4).

Step3. Set trial step $\alpha_k = 1$.

Step4. Set $x_{k+1} = x_k + \alpha_k d_k$.

Step5. Let $\lambda \in (0,1]$ $M \geq 1$ is a positive integer,defined $m(k) = \min[k+1, M]$

$$\lambda_{kr} \geq \lambda, \ r = 0,1,2,\cdots, m(k)-1 \quad \sum_{r=0}^{m(k)-1} \lambda_{kr} = 1,$$

Let $\alpha_k \geq 0$ be bounded above and satisfy:

$$f(x_k + \alpha_k d_k) \leq \max[f(x_k), \sum_{r=0}^{m(k)-1} \lambda_{kr} f(x_{k-r})] + \gamma \alpha_k \langle d_k, g(x_k) \rangle. \tag{6}$$

If (6) does not holds, define $\alpha_k = \alpha_k \beta$ and go to step 5

Step6. Set $k := k+1$, and go to step 1 .

3 The Global Convergence of the Second Algorithm

In this section, we discuss the global convergence property of algorithm with the new nonmonotone line search. In order to achieve the convergence of the second algorithm, we give some Assumptions and lemmas as follow:

Assumption 3.1

A: f(x) is bounded above on the level set $L = \{x | f(x) \leq f(x_0)\}$

B: In some neighborhood Ω of L ,f is continuously differentiable,and its gradient $\nabla f(x)$ is Lipschitz continuous, namely, there exists a constant L such that

$$\|\nabla f(x) - \nabla f(x_k)\| \leq L \|x - x_k\|$$

Lemma 3.1 Assume that the d_k is generated by (2)(3),(4) so for any $k \geq 0$ we have

$$d_k^T g_k = -\|g_k\|^2 \tag{7}$$

proof: when $k = 0$, $d_0^T g_0 = -\|g_0\|^2$ when $k \geq 1$,

$$d_k = -\theta_k g_k + \beta_k d_{k-1} = -g_k - \frac{\beta_k d_{k-1}^T g_k}{\|g_k\|^2} g_k + \beta_k d_{k-1}$$

so we have

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

which means that d_k is descent sufficiently undepend on line search.
Lemma 3.2 Assumption B is hold , α_k satisfy the formula(6)of algorithm, then there exist β satisfy :

$$\alpha_k \geq \min\{1, \frac{(1-\gamma)\beta}{L} \frac{|\langle g_k, d_k \rangle|}{\|d_k\|^2}\}. \tag{8}$$

Proof:

At the kth iterate , if $\alpha_k = 1$ satisfy the formula (6). then $\alpha_k = 1$,

otherwise, there exist β , which does not satisfy formula (6) for $\alpha_k/\beta > 0$, in other words :

$$\begin{aligned} f(x_k + \frac{\alpha_k}{\beta} d_k) &> \max\{f(x_k), \sum_{r=0}^{m_k-1} \lambda_{kr} f(x_{k-r})\} + \gamma \frac{\alpha_k}{\beta} \langle d_k, g(x_k) \rangle \\ &> f(x_k) + \gamma \frac{\alpha_k}{\beta} \langle d_k, g(x_k) \rangle. \end{aligned} \tag{9}$$

by mean value theorems, we have:

$$\begin{aligned} f(x_k + \alpha d_k) - f(x_k) &= \int_0^\beta \langle g(x_k + t d_k) - g(x_k) d_k \rangle d_t + \alpha \langle g(x_k), d_k \rangle. \\ &\leq \frac{1}{2} L \alpha^2 \|d_k\|^2 + \alpha \langle g(x_k), d_k \rangle. \end{aligned}$$

with(9), we have

$$\alpha_k \geq \min\{1, \frac{(1-\gamma)\beta}{L} \frac{|\langle g_k, d_k \rangle|}{\|d_k\|^2}\}.$$

So,(8) is hold.
Lemma 3.3 Assume that sequence x_k is generated by algorithm 2.2, so that:

$$\begin{aligned} f(x_k) &\leq f(x_0) + \lambda\gamma \sum_{r=0}^{k-2} \alpha_r \langle g(x_r), d_r \rangle + \gamma \alpha_{k-1} \langle g(x_{k-1}), d_{k-1} \rangle \\ &\leq f(x_0) + \lambda\gamma \sum_{r=0}^{k-1} \alpha_r \langle g(x_r), d_r \rangle. \end{aligned} \tag{10}$$

Proof: We prove by induction.

if $k = 1$, by (6) and $\lambda \leq 1$, we have

$$f(x_1) \leq f(x_0) + \lambda \alpha_0 \langle g(x_0), d_0 \rangle \leq f(x_0) + \gamma \lambda \alpha_0 \langle g(x_0), d_0 \rangle.$$

Assume (10) is hold for 1 , 2 $\cdots$ k , we can think of this problem from two cases :

case 1 : $\max[f(x_k), \sum_{r=0}^{m_k-1} \lambda_{kr} f(x_{k-r})] = f(x_k)$, by (6),we have

$$\begin{aligned} f(x_{k+1}) = f(x_k + \alpha_k d_k) &\leq f(x_k) + \gamma\alpha_k \langle g(x_k), d_k\rangle. \\ &\leq f(x_0) + \lambda\gamma \sum_{r=0}^{k-1} \alpha_r \langle g(x_r), d_r\rangle + \gamma\alpha_k \langle g(x_k), d_k\rangle. \\ &\leq f(x_0) + \lambda\gamma \sum_{r=0}^{k} \alpha_r \langle g(x_r), d_r\rangle \end{aligned}$$

case 2 : $\max[f(x_k), \sum_{r=0}^{m_k-1} \lambda_{kr} f(x_{k-r})] = \sum_{r=0}^{m_k-1} \lambda_{kr} f(x_{k-r})$,Let$q = \min[k, M-1]$ by (6) we have

$$\begin{aligned} f(x_{k+1}) = f(x_k + \alpha_k d_k) &\leq \sum_{p=0}^{q} \lambda_{kp} f(x_{k-p)}) + \gamma\alpha_k \langle g(x_k), d_k\rangle. \\ &\leq \sum_{p=0}^{q} \lambda_{kp} [f(x_0) + \lambda\gamma \sum_{r=0}^{k-p-2} \alpha_r \langle g(x_r), d_r\rangle + \\ &\gamma\alpha_{k-p-1} \langle g(x_{k-p-1}), d_{k-p-1}\rangle] + \gamma\alpha_k \langle g(x_k), d_k\rangle. \end{aligned}$$

impose (1,2 $\cdots$,q)$\times$(1,2 $\cdots$, $k-q-2$)$\subset \{(p, r) : 0 \leq p \leq q, 0 \leq r \leq k-q-2\}$, $\sum_{p=0}^{q} \lambda_{kp} = 1$, $\lambda_{kp} \geq \lambda$, we have

$$\begin{aligned} f(x_{k+1}) &\leq f(x_0) + \lambda \sum_{r=0}^{k-q-2} (\sum_{p=0}^{q} \lambda_{kp}) \alpha_r \langle g(x_r), d_r\rangle \\ &+ \gamma \sum_{p=0}^{q} \lambda_{kp} \alpha_{k-p-1} \langle g(x_{k-p-1}), d_{k-p-1}\rangle + \gamma\alpha_k \langle g(x_k), d_k\rangle. \\ &\leq f(x_0) + \lambda\gamma \sum_{r=0}^{k-q-2} \alpha_r \langle g(x_r), d_r\rangle + \lambda\gamma \sum_{r=k-p-1}^{k-1} \alpha_r \langle g(x_r), d_r\rangle \\ &+ \gamma\alpha_k \langle g(x_k), d_k\rangle \\ &= f(x_0) + \lambda\gamma \sum_{r=0}^{k-1} \alpha_r \langle g(x_r), d_r\rangle + \gamma\alpha_k \langle g(x_k), d_k\rangle \\ &\leq f(x_0) + \lambda\gamma \sum_{r=0}^{k} \alpha_r \langle g(x_r), d_r\rangle \end{aligned}$$

end.

Theorem 3.1 Assume that x_k and d_k is generated by algorithm 2.2 and A , B holds ,and then

$$\lim_{k\to\infty} \langle g(x_k), d_k\rangle = 0. \tag{11}$$

Proof:

Assume that there exist a boundless sequence index set K, and there exist $\varepsilon > 0$, which satisfy $\langle g(x_k), d_k\rangle \leq -\varepsilon$. for any $k \in K$, impose the lemma 3.3 , for any $k \in K$, we have :

$$-\lambda\gamma \sum_{r=0 r\in K}^{k-1} \alpha_r \langle g(x_r), d_r\rangle \leq f(x_0) - f(x_k), \tag{12}$$

impose the lemma 3.1 , $g_k^T d_k = -\|g_k\|^2$, which means

$$-\frac{\langle g(x_k), d_k\rangle}{\|g_k\|^2} = -1, \tag{13}$$

with (8) (12), (13) we have:

$$\begin{aligned} f(x_0) - f(x_k) &\geq -\lambda\gamma \sum_{r=0,r\in K}^{k-1} \alpha_r \langle g(x_r), d_r\rangle \\ &\geq \lambda\gamma\varepsilon \sum_{r=0,r\in K}^{k-1} \alpha_r \\ &\geq \lambda\gamma\varepsilon \sum_{r=0,r\in K}^{k-1} \min\{1, \frac{(1-\gamma)\beta}{L} \cdot \frac{|\langle g(x_r), d_r\rangle|}{\|d_r\|^2}\} \\ &\geq \lambda\gamma\varepsilon \sum_{r=0,r\in K}^{k-1} \min\{1, \frac{(1-\gamma)\beta}{L}\} \end{aligned}$$

since $f(x)$ is bounded below , let $k \to \infty (k \in K)$, we have

$$\infty \geq f(x_0) - f(x_k) \to \infty.$$

paradoxically, so (11) holds
end.

Theorem 3.2 Assume that sequence $\{x_k\}$ is generated by algorithm 2.2, and then we have

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

Proof:
by Lemma 3.1 and theorem 3.1, we have

$$0 \geq \lim_{k\to\infty} -\|g_k\|^2 \geq \lim_{k\to\infty} \langle g(x_k), d_k\rangle = 0$$

which implies

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

end.

4 Numerical Results

Here, we provide some numerical experiments to show the performance of the two proposed algorithms. in the algorithm 2.1, as we are known, $\mu = 0$ is chosen by nonmonotone and $\mu = 1$ is chosen by monotone. we set the parameters: $\beta = 0.5$, $\delta = 0.2$, $\lambda = 1$, $\mu = 0.8$, $M = 10$. and we explore the following value for the parameters of algorithm 2.2: $\beta = 0.5$, $\delta = 0.2$, $\lambda = 1$, $M = 10$.

The algorithms has been tested on the following set of problems, for all codes the termination criterion was $\|g_k\| \leq 10^{-5}$, we compute the number n_l of line search,and the value $f(\widehat{x})$ of the objective function at the solution found $\widehat{x}$, we denote 3.629×10^{-9} by $3.629(-9)$ in the table.

$Problem1.f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$

$$x_0 = [-1.2, 1]', x^* = [1, 1], f(x^*) = 0.$$

$Problem2.f(x) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2 + (x_3 - 1)^2 + 90(x_3^2 - x_4)^2 + 10.1[(x_2 - 1)^2 + (x_4 - 1)^2] + 19.8(x_2 - 1)(x_4 - 1)$

$$x_0 = [-3, -1, -3, -1]', x^* = [1, 1, 1, 1], f(x^*) = 0.$$

$Problem3.f(x) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4$

$$x_0 = [3, -1, 0, 1]', x^* = [0, 0, 0, 0], f(x^*) = 0.$$

$Problem4.f(x) = 100(x_2 - x_1^3)^2 + (1 - x_1)^2$

$$x_0 = [-1.2, -1]', x^* = [1, 1], f(x^*) = 0.$$

$Problem5.f(x) = (x_1 + 10 * x_2)^4 + 5 * (x_3 - x_4)^4 + (x_2 - 2 * x_3)^4 + 10 * (x_1 - 10 * x_4)^4$

$$x_0 = (2, 2, -2, -2)^T, x^* = (0, 0, 0, 0)^T, f^* = 0.$$

$Problem6.f(x) = (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_3 - 1)^2 + (x_4 - 1)^4 + (x_5 - 1)^6$

$$x_0 = (2, 2, 2, 2, 2)^T, x^* = (1, 1, 1, 1, 1)^T, f^* = 0.$$

Table 1: results for algorithm 1 with $\mu = 0.8$

Alg 1	pro 1	pro 2	pro 3	pro 4	pro 5	pro 6
n_l	272	433	294	269	357	121
$f(\widehat{x})$	3.629(-9)	2.550(-11)	8.845(-9)	2.800(-9)	1.337(-6)	2.072(-8)

Table 2: varying μ

μ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
problem 5	578	437	516	368	489	232	432	204	357	192	230
problem 6	478	315	330	271	230	191	173	138	121	90	98

Table 3: algorithm 2 with varying M

M	1	2	3	4	5	6	7	8	9
problem 5	230	70	75	113	101	102	301	140	145

In the table 2, we can obtain a better convergence when $\mu \in [0, 1]$, especially the choice of $\mu = 0.9$.

In table 3, the numerical performance is highly dependent on the choice of M. almost the choice of M is better than the choice of $M = 1$ which is monotone Armijo line search. Although not all nonmonotone line search perform better than monotone line search, we can find a better.

ACKNOWLEDGEMENTS. This work is supported by the National Natural Science Foundation of China (No. 11101115), the Natural Science Foundation of Hebei Province (No. 2014201033), the Key Research Foundation of Education Bureau of Hebei Province, (No.ZD2015069) and the Science and Technology project of Hebei province (No. 13214715).

References

[1] J.C. Gilbert, J. Nocedal. Global convergence properties of conjugate gradient methods for optimization, SIAM J. Optim. 2(1) 1992, 21-42.

[2] Fletcher, Reeves, C. M, Function Minimization by Conjugate gradients, Computer Journal, 1964, Vol7, 149-154.

[3] Fletcher, R, Practical Methods of Optimization, Wiley, New YorK,1987, Vol I, Unconstrained Optimization, 2nd edition.

[4] Dai, Y.H. and Yuan, Y, A nonlinear conjugate gradient methods with a strong global convergence property, SIAM J Optim. 10, 1999 , 177-182.

[5] M. Raydan, The Barzilai and Borwein gradient method for large scale unconstrained minimization problem, SIAM J. Optim.,7 1996, 26-33

[6] Birgin E. G., Chambouleyron I. and Martinez J. M, Estimation of optical constants of thin films using unconstrained optimization, Journal of Computational Physics 151 1999, 862-880

[7] Glunt W., Hayden T. L. and Raydan M, Molecular conformations from distance matrices, Journal of Computational Chemistry 14 1993, 114-120

[8] Birgin, E. and Martinez, J. M,. A spectral conjugate gradient method for unconstrained optimization, Applied Mathematics and Optimization. 43 2001, 117-128

[9] Hu Chaomin, Wan Zhong, and Wang Xu, a new nonmonotone spectral conjugate gradient algorithm, School of Mathematics and Statistics,Central South University,Changsha 4100832013

[10] Keyvan Amini, Masoud Ahookhosh and Hadi Nosratipour. An inexact line search approach using modified nonmonotone strategy for unconstrained optimization, Numer Algor 2014, 49-78.

[11] Hongchao Zhang and William W. Hager. A nonmonotone line search technique and its application to unconstrained optimization, SIAM J. Optim. 14, 2004, 1043-1056.

[12] L. Grippo, F. Lampariello, and S. Ludidi, A nonmonotone line search technique for Newton's method, SIAM J. Numer. Anal. 23, 1986, 707-716.

[13] Chamberlain, R.M, Powell, M.J.D, Lemarechal, C, and Pedersen, H.C, The watchdog technique for forcing convergence in algorithm for constrained optimization, Math. Program. Stud.16, 1982, 1-17.

[14] P. L. Toint, An assessment of nonmonotone line search techniques for unconstrained optimization, SIAM J. Sci. Comput.,171996, 725-739.

[15] Zhensheng Yu and Dingguo Pu. A new nonmonotone line search technique for unconstrained optimization, Computational and Applied Mathematics, 219 2008, 134-144.

Received: July, 2015