J Syst Sci Complex (2011) 24: 784–794

A NEW DESCENT MEMORY GRADIENT METHOD AND ITS GLOBAL CONVERGENCE*

Min SUN · Qingguo BAI

DOI: 10.1007/s11424-011-8150-0
Received: 8 April 2008 / Revised: 30 March 2010

Abstract In this article, a new descent memory gradient method without restarts is proposed for solving large scale unconstrained optimization problems. The method has the following attractive properties: 1) The search direction is always a sufficiently descent direction at every iteration without the line search used; 2) The search direction always satisfies the angle property, which is independent of the convexity of the objective function. Under mild conditions, the authors prove that the proposed method has global convergence, and its convergence rate is also investigated. The numerical results show that the new descent memory method is efficient for the given test problems.

Key words Global convergence, memory gradient method, sufficiently descent.

1 Introduction

The objective of this paper is to study the unconstrained minimization problem

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

where R^n denotes an n-dimensional Euclidean space, and $f : R^n \to R^1$ is smooth whose gradient is denoted by g. Line search methods are traditional iterative methods for solving (1). The iterative process of the line search method is given by

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

where d_k is the search direction of $f(x)$ at x_k and the steplength α_k is obtained by carrying out some line search. For convenience, if x_k is the current iterative point, we denote $\nabla f(x_k)$ by g_k, $f(x_k)$ by f_k, $f(x^*)$ by f^*, where x^* is a solution of (1).

Generally, memory gradient methods and conjugate gradient methods are two powerful line search methods for solving large scale problems because they avoid the computation and storage

Min SUN
Department of Mathematics and Information Science, Zaozhuang University, Zaozhuang 277160, *China.*
Email: sunmin_2008@yahoo.com.cn.
Qingguo BAI
School of Management, Qufu Normal University, Rizhao 276826, *China.* Email: qfnubaiqg@163.com.
*This research is supported by the National Science Foundation of China under Grant No. 70971076 and the Foundation of Shandong Provincial Education Department under Grant No. J10LA59.
◇ *This paper was recommended for publication by Editor Shouyang WANG.*

of some matrices associated with Newton type methods. The search direction d_k in conjugate gradient methods is recursively defined by

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

where β_k is a parameter. The search direction d_k in memory gradient method does not have a uniform form. Compared with conjugate gradient method, memory gradient method can use the information of the previous iterations more sufficiently and hence it is helpful to design algorithms with quick convergence rate. In line search method, the search direction d_k is generally required to satisfy

$$g_k^\top d_k < 0, \tag{4}$$

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

$$g_k^\top d_k \leq -c\|g_k\|^2, \tag{5}$$

where $c > 0$ is a constant and $\|\cdot\|$ denotes the Euclidean norm of vectors. Then it is an interesting task to design a conjugate gradient method which possesses this condition. Moreover, the angle property is often used in proving the global convergence of related line search methods[1], that is,

$$\cos\langle -g_k, d_k\rangle = -\frac{g_k^\top d_k}{\|g_k\| \cdot \|d_k\|} \geq \tau, \tag{6}$$

where $\tau > 0$ is a constant. (6) indicates that the angle of $-g_k$ and d_k needs to be less than $\pi/2$.

Descent property (4) is very important for the global convergence of conjugate gradient method. Whether conjugate gradient methods can generate descent directions depends on some line searches, but for memory gradient method[2−3], it is easy to deduce that the search direction satisfies (4) and (5) without the line search.

Recently, Zhang[4−5] proposed a descent conjugate gradient method. The search direction in their method always satisfies

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

which shows that d_k is a sufficiently descent direction unless x_k is a stationary point of (1), but it cannot guarantee the angle property. Shi[6] proposed a descent conjugate gradient method with a line search satisfying the sufficient descent condition in addition to the standard Wolfe conditions. The search direction d_k in [6] satisfies the sufficient descent condition by the line search, and it also satisfies the angle property by the Lipschitz continuous of g. Cheng[7] proposed another descent method and pointed out that no matter β_k taking any form, if d_k is defined by

$$d_k = \begin{cases} -g_k, & k = 1, \\ -\left(1 + \beta_k \dfrac{g_k^\top d_{k-1}}{\|g_k\|^2}\right) g_k + \beta_k d_{k-1}, & k \geq 2, \end{cases} \tag{8}$$

then d_k always satisfies the sufficient descent condition (7). Moreover, some new theoretical and numerical results on conjugate gradient method were addressed in [3–6]. But the question whether we can find a line search method not only guarantees the sufficient descent condition but also satisfies the angle property without utilizing any property of the underlying function f and the line search. The answer is yes.

In this paper, based on [7], a new memory gradient method is proposed. The search direction d_k generated by the new method satisfies sufficient descent condition and angle property, and we give an appropriate initial step size at each iteration which can decrease the function evaluations so as to improve the performance of the new method.

The rest of this paper is organized as follows. In Section 2, we introduce the new memory gradient method. In Section 3, the global convergence and linear convergence rate are analyzed. Numerical results and one conclusion are presented in Section 4 and Section 5, respectively.

2 New Memory Gradient Method

We need to impose the following assumptions on f, which is weaker than assumptions used in [2–8].

H1) The objective function f has lower bound on the level set $L_0 = \{x \in R^n | f(x) \leq f(x_1)\}$.

H2) In some neighborhood N of L_0, g is uniformly continuous on an open convex set B that contains L_0.

Let $\{Q_k\}$ be a sequence of positive definite matrices. We define $\|d_k\|_{Q_k} = \sqrt{d_k^\top Q_k d_k}$.

Now, we begin to describe the new memory gradient method.

Algorithm 1 (A new memory gradient method)

Step 0 Given an initial point $x_1 \in R^n$, $t \geq 1$, $\gamma \in (0,1)$, $\rho \in (0,1)$, a positive definite matrix Q_1 and set $d_1 = -g_1$, $k := 1$.

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

Step 2 Compute d_k by (8), where β_k is defined by

$$\beta_k^N = \frac{g_k^\top \left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1} \right)}{|g_k^\top d_{k-1}| + t\|g_k\| \cdot \|d_{k-1}\|}. \tag{9}$$

Set $s_k = -g_k^\top d_k / \|d_k\|_{Q_k}^2$. Let α_k be the largest α in $\{s_k, \rho s_k, \rho^2 s_k, \cdots\}$ such that

$$f(x_k + \alpha d_k) - f_k \leq \gamma \alpha g_k^\top d_k. \tag{10}$$

Step 3 Update Q_k to obtain Q_{k+1} by some technique. Set $x_{k+1} = x_k + \alpha_k d_k$, $k := k+1$; go to Step 1.

Remark 2.1 From (9), it is obvious that: 1) The scalar β_k^N keeps nonnegative, and this property is independent of the line search used; 2) $\beta_k^N = 0$ if $g_k = g_{k-1}$. Thus, the new direction tends to the steepest direction $-g_k$ if a very little progress is obtained, that is, the new method can restart automatically.

We always choose the initial step size as the one-dimensional minimizer of a quadratic model $\Phi(x) := f_k + \alpha g^\top d_k + \frac{1}{2}\alpha^2 d_k^\top Q_k d_k$, where Q_k can be set to the Hessian matrix or its approximation. This may make the step size α_k more easily be accepted and decrease the function evaluations at each iteration.

3 Global Convergence and Linear Convergence Rate

Throughout this section, we assume that $g_k \neq 0$, for all $k \geq 1$ (otherwise a stationary point has been found). The following lemma implies that d_k provides a sufficient descent direction of f at x_k.

Lemma 3.1 *For all $k \geq 1$, we have*

$$g_k^\top d_k = -\|g_k\|^2.$$

Proof The conclusion is obvious from (8), (9). This completes the proof. ∎

Since d_k is a descent direction of f at x_k, Algorithm 1 is well defined.

Remark 3.1 The parameter β_k^N is inspired by the following analysis. In order to guarantee that d_k satisfies the angle property (6) with some constant $\tau > 0$, from Lemma 3.1 and (6), we only need to guarantee

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

that is, $\|d_k\| \leq \|g_k\|/\tau$. On the other hand, from (8), we have

$$\|d_k\| \leq \|g_k\| + \beta_k \frac{|g_k^\top d_{k-1}| + t\|g_k\| \cdot \|d_{k-1}\|}{\|g_k\|}.$$

Thus, we only need to guarantee

$$\|g_k\| + \beta_k \frac{|g_k^\top d_{k-1}| + t\|g_k\| \cdot \|d_{k-1}\|}{\|g_k\|} \leq \frac{\|g_k\|}{\tau}.$$

Let $\tau = 1/3$, and from the above inequality, we have

$$\beta_k \leq \frac{2\|g_k\|^2}{|g_k^\top d_{k-1}| + t\|g_k\| \cdot \|d_{k-1}\|},$$

which motivates us to design the scalar β_k^N.

Lemma 3.2 *For all $k \geq 1$, we have*

$$\cos\langle -g_k, d_k\rangle \geq \tau = \frac{1}{3}. \tag{11}$$

Proof If $k = 1$ then

$$\cos\langle -g_k, d_k\rangle = \cos\langle -g_k, -g_k\rangle = 1 > \tau.$$

If $k \geq 2$, then, by (8), (9), and $t \geq 1$, we have

$$\begin{aligned}
\|d_k\| &\leq \|g_k\| + \beta_k^N \left(\frac{|g_k^\top d_{k-1}| \cdot \|g_k\|}{\|g_k\|^2} + \|d_{k-1}\|\right) \\
&\leq \|g_k\| + \beta_k^N \frac{|g_k^\top d_{k-1}| + \|g_k\| \cdot \|d_{k-1}\|}{\|g_k\|} \\
&\leq \|g_k\| + \frac{\|g_k\| \|g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}\|}{|g_k^\top d_{k-1}| + t\|g_k\| \cdot \|d_{k-1}\|} \cdot \frac{|g_k^\top d_{k-1}| + t\|g_k\| \cdot \|d_{k-1}\|}{\|g_k\|} \\
&\leq 2\|g_k\| + \frac{\|g_k\|}{\|g_{k-1}\|} \|g_{k-1}\| \\
&= 3\|g_k\|.
\end{aligned}$$

By Lemma 3.1, we get

$$\cos\langle -g_k, d_k\rangle = \frac{\|g_k\|}{\|d_k\|} \geq \frac{1}{3}.$$

This completes the proof. ∎

Remark 3.2 In a similar way, we can get some other scalars which also satisfy the sufficient descent property and the angle property, such as

$$\beta_k = \frac{g_k^\top \left(g_k - \frac{g_k^\top g_{k-1}}{\|g_{k-1}\|^2} g_{k-1}\right)}{|g_k^\top d_{k-1}| + t\|g_k\| \cdot \|d_{k-1}\|},$$

or

$$\beta_k = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|}|g_k^\top g_{k-1}|}{|g_k^\top d_{k-1}| + t\|g_k\| \cdot \|d_{k-1}\|}.$$

Remark 3.3 From the proofs of Lemmas 3.1 and 3.2, we can get that both the sufficient descent property and the angle property are independent of the line search.

The next result shows that Algorithm 1 is globally convergent for general functions.

Theorem 3.1 *If* H1), H2) *hold and Algorithm* 1 *generates an infinite sequence* $\{x_k\}$, *then any cluster point of* $\{x_k\}$ *is a stationary point of* f.

Proof From (10), Lemma 3.1 and H1), we have

$$\lim_{k\to\infty} \alpha_k \|g_k\|^2 = 0. \tag{12}$$

Suppose that $\overline{x}$ is a cluster point of $\{x_k\}$, then there exists an infinite subset $K_0 \subseteq \{2, 3, \cdots\}$ such that

$$\lim_{k\to\infty, k\in K_0} x_k = \overline{x}.$$

We will divide our proof into two cases: $\inf\limits_{k\in K_0} \alpha_k > 0$ and $\inf\limits_{k\in K_0} \alpha_k = 0$. In the first case, we obviously have

$$\|g(\overline{x})\| = \lim_{k\to\infty, k\in K_0} \|g_k\| = 0.$$

If $\inf\limits_{k\in K_0} \alpha_k = 0$, then there exists an infinite subset $K_1 \subseteq K_0$ such that

$$\lim_{k\to\infty, k\in K_1} \alpha_k = 0. \tag{13}$$

If there exists $\varepsilon_0 > 0$ and an infinite subset $K_2 \subseteq K_1$ such that

$$\frac{\|g_k\|^2}{\|d_k\|} \geq \varepsilon_0, \qquad \forall k \in K_2, \tag{14}$$

so conditions (12) and (14) show that

$$\lim_{k\to\infty, k\in K_2} \alpha_k \|d_k\| = 0. \tag{15}$$

By (10), for sufficiently large $k \in K_2$, we have

$$f\left(x_k + \frac{\alpha_k d_k}{\rho}\right) > f(x_k) + \frac{\gamma \alpha_k g_k^\top d_k}{\rho}.$$

Using the mean value theorem in the above inequality, we obtain $\theta_k \in (0, 1)$, such that

$$\left[g\left(x_k + \frac{\theta_k \alpha_k d_k}{\rho}\right) - g_k\right]^\top d_k > (\gamma - 1) g_k^\top d_k.$$

By Lemma 3.1 again, we obtain

$$\begin{aligned}\|g(x_k+\theta_k\alpha_k d_k/\rho)-g_k\| &= \frac{\|g(x_k+\theta_k\alpha_k d_k/\rho)-g_k\|\cdot\|d_k\|}{\|d_k\|}\\ &\geq \frac{[g(x_k+\theta_k\alpha_k d_k/\rho)-g_k]^\top d_k}{\|d_k\|}\\ &\geq (\gamma-1)\frac{g_k^\top d_k}{\|d_k\|} = (1-\gamma)\frac{\|g_k\|^2}{\|d_k\|}.\end{aligned}$$

From H2), (15), and the above inequality, we have

$$\lim_{k\to\infty,k\in K_3}\frac{\|g_k\|^2}{\|d_k\|}=0, \tag{16}$$

which contradicts (14), then we can obtain

$$\lim_{k\to\infty,k\in K_2}\frac{\|g_k\|^2}{\|d_k\|}=0. \tag{17}$$

By Lemma 3.1, Lemma 3.2, and (17), we have

$$0=\lim_{k\to\infty,k\in K_2}\frac{\|g_k\|^2}{\|d_k\|}=-\lim_{k\to\infty,k\in K_2}\frac{\|g_k\|\cdot g_k^\top d_k}{\|d_k\|\cdot\|g_k\|}\geq\lim_{k\to\infty,k\in K_2}\frac{\|g_k\|}{2}=\frac{\|g(\overline{x})\|}{2}.$$

The proof is completed. ∎

If H2) is replaced by the following stronger assumption, we can get a stronger global convergence result.

H2′) In some neighborhood N of L_0, g is Lipschitz continuous on an open convex set Ω that contains L_0, namely, there exists a constant $L>0$ such that for all $x,y\in\Omega$,

$$\|g(x)-g(y)\|\leq L\|x-y\|. \tag{18}$$

We assume that there exist $\lambda_{\min}>0$ and $\lambda_{\max}>0$ such that for all $x\in R^n$,

$$\lambda_{\min}\|x\|^2\leq x^\top Q_k x\leq\lambda_{\max}\|x\|^2.$$

The condition of the following lemma, often called Zoutendijk condition, is used to prove the global convergence of nonlinear conjugate gradient method.

Lemma 3.3 *If* H1), H2′) *hold and Algorithm* 1 *generates an infinite sequence* $\{x_k\}$, *then*

$$\sum_{k=1}^{\infty}\frac{\|g_k\|^4}{\|d_k\|^2}<+\infty. \tag{19}$$

Proof First, we consider the following two cases: $\alpha_k=s_k$ and $\alpha_k<s_k$. In the first case, we obviously have

$$\alpha_k\geq\frac{-g_k^\top d_k}{\lambda_{\max}\|d_k\|^2}=\frac{\|g_k\|^2}{\lambda_{\max}\|d_k\|^2}. \tag{20}$$

If $\alpha_k<s_k$, this implies that α_k/ρ violates (10). As deduced in Theorem 3.1, we have

$$[g(x_k+\theta_k\alpha_k d_k/\rho)-g_k]^\top d_k>(1-\gamma)\|g_k\|^2.$$

By H2′), we have

$$L\alpha_k\|d_k\|^2/\rho > (1-\gamma)\|g_k\|^2.$$

Therefore,

$$\alpha_k \geq \frac{\rho(1-\gamma)\|g_k\|^2}{L\|d_k\|^2}. \tag{21}$$

Letting $N = \min\{1/\lambda_{\max}, \rho(1-\gamma)/L\}$, from (20) and (21), we obtain

$$\alpha_k \geq N\frac{\|g_k\|^2}{\|d_k\|^2}, \quad \forall k. \tag{22}$$

From (10) and (22), we have

$$\gamma N\Big(\frac{\|g_k\|^2}{\|d_k\|}\Big)^2 \leq f_k - f_{k+1},$$

which, together with H1), we get (19). The proof is completed. ∎

Theorem 3.2 *If* H1), H2′) *hold and Algorithm* 1 *generates an infinite sequence* $\{x_k\}$, *then*

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

Proof By Lemma 3.3, we have

$$\lim_{k\to\infty}\frac{\|g_k\|^2}{\|d_k\|} = 0. \tag{23}$$

In the proof of Lemma 3.2, we have

$$\|d_k\| \leq 2\|g_k\|.$$

Therefore,

$$\frac{\|g_k\|^2}{\|d_k\|} \geq \frac{\|g_k\|}{2},$$

which, together with (23), we get the desired result. This completes the proof. ∎

In order to prove the linear convergence rate of the new method, we need the following assumption.

H3) f is uniformly convex and twice continuously differentiable.

In fact, Assumption H3) implies H1), H2), and H2′).

Lemma 3.4[2] *If* H3) *holds, then* f *has the following properties:*

1) f *has a unique minimizer on* R^n, *say* x^*.

2) *There exist* $m > 0$, $M > 0$ *and* $\varepsilon > 0$ *such that*

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

$$m\|x-x^*\| \leq \|g(x)\| \leq M\|x-x^*\|, \quad \forall x \in N(x^*, \varepsilon).$$

The following theorem is inspired by Theorem 4.1 in [6].

Theorem 3.3 *If* H3) *holds, then* $\{x_k\}$ *converges to* x^* *at least R-linearly.*

Proof If H3) holds then there exists k' such that $x_k \in N(x^*, \varepsilon_0)$, $\forall k \geq k$. Without loss of generality, we assume that $x_1 \in N(x^*, \varepsilon_0)$. From (10) and (22), we have

$$f_k - f_{k+1} \geq \frac{\gamma N\|g_k\|^2}{4}. \tag{24}$$

By Lemma 3.1, we have

$$\|g_k\| \cdot \|d_k\| \geq -g_k^\top d_k = \|g_k\|^2,$$

which yields

$$\|d_k\| \geq \|g_k\|. \tag{25}$$

According to Lemma 3.4 2), we have

$$m < L. \tag{26}$$

By Lemma 3.4 2) and (24), we obtain

$$f_k - f_{k+1} \geq \frac{\gamma N \|g_k\|^2}{4} \geq \frac{\gamma N m^2 \|x_k - x^*\|^2}{4} \geq \frac{\gamma N m^2}{2M}(f_k - f^*).$$

By setting

$$\theta = m\sqrt{\frac{\gamma N}{2M}},$$

we have

$$f_k - f_{k+1} \geq \theta^2 (f_k - f^*). \tag{27}$$

Now, we prove that $\theta < 1$. In fact, by the definition of N, (26), and noting $m \leq M$, we have

$$\theta^2 = \frac{\gamma N m^2}{2M} \leq \frac{\gamma \rho (1-\gamma) m^2}{2ML} \leq \frac{\gamma(1-\gamma)\rho m}{L} \leq \rho < 1.$$

By setting

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

obviously $\omega < 1$, we obtain from (27) that

$$f_{k+1} - f^* \leq (1-\theta^2)(f_k - f^*) = \omega^2 (f_k - f^*) \leq \cdots \leq \omega^{2(k-k')}(f_{k'+1} - f^*).$$

By Lemma 3.4 and the above inequality, we have

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

thus,

$$\|x_k - x^*\| \leq \omega^k \sqrt{\frac{2(f_{k'+1} - f^*)}{m\omega^{2(k'+1)}}}$$

and

$$\lim_{k\to\infty} \|x_k - x^*\|^{\frac{1}{k}} \leq \omega < 1.$$

which shows that $\{x_k\}$ converges to x^* at least R-linearly. This completes the proof. ∎

4 Numerical Results

In this section, we provide the implementation details of the new algorithm to verify the theoretical results.

For convenience, we restrict the $Q_k = I$ for all $k \geq 1$, and New Method stands for our new descent memory method with this special choice.

The conjugate gradient method takes the form (3) and

$$\beta_k^{\rm HS} = \frac{g_k^\top (g_k - g_{k-1})}{g_{k-1}^\top d_{k-1}}, \quad \text{(see [9])}; \qquad \beta_k^{\rm FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \text{(see [10])};$$

$$\beta_k^{\rm PRP} = \frac{g_k^\top (g_k - g_{k-1})}{d_{k-1}^\top (g_k - g_{k-1})}, \quad \text{(see [11 − 12])}; \quad \beta_k^{\rm DY} = \frac{\|g_k\|^2}{d_{k-1}^\top (g_k - g_{k-1})}, \quad \text{(see [13])};$$

$$\beta_k^{\rm WYL} = \frac{g_k^\top (g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1})}{\|g_{k-1}\|^2}, \quad \text{(see [14])}.$$

Its corresponding method is called HS, FR, PRP, DY and WYL conjugate gradient method, respectively. The codes were written in Matlab 7.1 and run on a portable computer. For each problem, the limiting number of function evaluations is set to 10000. ‘F’ means the method failed. The stopping criterion is

$$\|g_k\| \le 10^{-6}.$$

Our numerical results are listed in the form NI/NF/NG, where the symbols NI, NF and NG mean the number of iterations, the number of function evaluations and the gradient evaluations, respectively.

Problem 1

$$f(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5(x_1 + x_2) - 21x_3 + 7x_4, \quad x_1 = (1, 1, 1, 1)^\top.$$

Problem 2

$$f(x) = (1 - x_1)^2 + (1 - x_{10})^2 + \sum_{i=1}^{9} (x_i^2 - x_{i+1})^2, \quad x_1 = (-2, \cdots, -2)^\top.$$

Problem 3

$$f(x) = {\rm e}^{x_1} + x_1^2 + 2x_1x_2 + 4x_2^2, \quad x_1 = (1, 1)^\top.$$

Problem 4

$$f(x) = \sum_{i=1}^{n} ({\rm e}^{x_i} - x_i), \quad x_1 = \Big(\frac{n}{n-1}, \frac{n}{n-1}, \cdots, \frac{n}{n-1}\Big)^\top.$$

Problem 5

$$f(x) = 10^{-5} \sum_{i=1}^{n} (x_i - 1)^2 + \Big(\sum_{i=1}^{n} x_i^2 - 0.25\Big)^2, \quad x_1 = (1, 2, \cdots, n)^\top.$$

Problem 6

$$f(x) = \sum_{i=1}^{n/2} ((x_{2i} - x_{2i-1}^2)^2 + (1 - x_{2i-1})^2), \quad x_1 = (-1, 2, 1, \cdots, -1, 2, 1)^\top.$$

Problem 7 $f(x)$ is defined in Problem 5, with another initial point $x_1 = (1, 1, \cdots, 1)^\top$.

From Tables 1–2, New Method is superior to HS method and comparable to PRP method, and it can solve all problems, while DY and FR method cannot solve Problem 2. It can be seen from Table 3 that New Method is comparable to PRP method. PRP method, HS method, WYL method and New Method are all superior to FR method. For Problem 7, the numerical results of all methods are the same.

Table 1 Numerical results of Problems 1–4, $t = 1.1$

P	n	PRP	FR	HS	DY	New Method
P1	4	13/34/14	20/55/21	17/1068/18	20/55/21	10/30/11
P2	10	305/651/306	F	334/592/335	F	229/407/230
P3	2	8/25/9	26/86/27	13/41/14	13/42/14	19/49/20
P4	10	4/6/5	5/6/6	4/6/5	4/6/5	4/6/5
	100	4/6/5	4/6/5	4/6/5	4/6/5	4/6/5
	500	4/6/5	4/6/5	4/6/5	4/6/5	4/6/5
	1000	4/6/5	4/6/5	4/6/5	4/6/5	4/6/5
	10000	4/6/5	4/6/5	4/6/5	4/6/5	4/6/5
	20000	4/6/5	4/6/5	4/6/5	4/6/5	4/6/5

Table 2 Numerical results of Problem 5, $t = 1.46$

n	FR	PRP	DY	HS	New Method
200	8863/9036/8864	8882/9055/8883	73/1245/74	9696/9912/9697	8883/9056/8884
500	6186/6415/6187	6140/6369/6141	44/1307/45	F	6140/6369/6141
1000	4617/4895/4618	4622/4900/4123	43/1318/44	8308/8676/8309	4623/4901/4624

Table 3 Numerical results of Problem 6, $t = 1.1$

n	FR	PRP	HS	WYL	New Method
120	340/1439/341	56/136/57	47/111/48	60/142/61	48/103/49
240	347/1467/348	59/147/60	55/137/56	62/147/63	48/103/49
480	354/1495/355	56/132/57	52/129/53	62/142/63	48/103/49
1000	361/1523/362	48/104/49	48/126/49	66/164/67	60/135/61
2000	367/1548/368	55/132/56	71/190/72	64/156/65	53/118/54
3000	373/1571/374	64/155/65	53/125/54	66/153/67	51/109/52
4000	376/1583/377	66/162/67	61/159/62	73/186/74	58/132/59
5000	379/1595/380	62/143/63	53/139/54	66/162/67	65/154/66

Table 4 Numerical results of Problem 7, $t = 1.46$

n	FR	PRP	HS	New Method
80	8/42/9	8/42/9	8/42/9	8/42/9
160	9/50/10	9/50/10	9/50/10	9/50/10
320	10/59/11	10/59/11	10/59/11	10/59/11
1000	10/64/11	10/64/11	10/64/11	10/64/11
2000	9/68/10	9/68/10	9/68/10	9/68/10
4000	11/80/12	11/80/12	11/80/12	11/80/12
8000	10/85/11	10/85/11	10/85/11	10/85/11
16000	12/98/13	12/98/13	12/98/13	12/98/13

Hence, New Method appears to generate the best direction, on average. The explanation of this behavior is that the choice of β_k^N always keeps nonnegative and independent of line search used, and the generated direction always satisfies the sufficient descent condition, which the direction generated by other conjugate gradient methods does not always satisfy.

5 Conclusions

In this paper, we propose a new descent memory gradient method which possesses the sufficient descent condition without carrying out any line search and satisfies the angle property, and this property is independent of the convexity of the objective function. The computational evidence shows that the performance of our method is comparable to conjugate gradient methods.

It is clear that if exact line search is used, then $g_k^\top d_{k-1} = 0$. In this case, we get another parameter

$$\beta_k = \frac{\|g_k\| - \frac{g_k^\top g_{k-1}}{\|g_{k-1}\|}}{t\|d_{k-1}\|}, \quad t \geq 1.$$

Under inexact line search, whether the conjugate gradient method with this parameter has global convergence? It will be our further research. Moreover, more numerical experiments for large practical problems and for the choice of the constant t should done in the future.

References

[1] Y. X. Yuan and W. Y. Sun, *Theory and Methods of Optimization*, Science Press of China, Beijing, 2002.

[2] Z. J. Shi, A new memory gradient method under exact line search, *Asia-Pacific J. Oper. Res.*, 2003, **20**: 275–284.

[3] Z. J. Shi, A new super-memory gradient method for unconstrained optimization, *Advance in Mathematics*, 2006, **35**: 265–274.

[4] L. Zhang, W. J. Zhou, and D. H. Li, A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence, *IMA Journal of Numerical Analysis*, 2006, **26**: 629–640.

[5] L. Zhang, W. J. Zhou, and D. H. Li, Global convergence of a modified Fletcher-Reeves conjugate method with Armijo-type line search, *Numerische Mathematik*, 2006, **104**: 561–572.

[6] Z. J. Shi and J. Shen, Convergence of Liu-Storey conjugate gradient method, *European Journal of Operational Research*, 2007, **182**: 552–560.

[7] W. Y. Cheng, A two-term PRP-based descent method, *Numerical Functional Analysis and Optimization*, 2007, **28**: 1217–1230.

[8] Z. J. Shi and J. Shen, Convergence of the Polak-Ribière-Polyak conjugate method, *Nonlinear Analysis*, 2007, **66**: 1428–1441.

[9] M. R. Hestenes and E. Stiefel, Methods of conjugate gradient for solving linear systems, *Journal of Research of the National Bureau of Standards*, 1952, **49**: 409–436.

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

[11] E. Polyak and G. Ribière, Note Sur la convergence de méthods de directions conjugées, *Revue Francaise d'Informatique et de Recherche Opérationnelle*, 1969, **16**: 35–43.

[12] B. T. Polyak, The conjugate gradient method in extreme problems, *USSR Computational Mathematics and Mathematical Physics*, 1969, **9**: 94–112.

[13] Y. H. Dai and Y. X. Yuan, A nonlinear conjugate gradient with a strong global convergence properties, *SIAM Journal on Optimization*, 2000, **10**: 177–182.

[14] Z. X. Wei, S. W. Yao, and L. Y. Liu, The convergence properties of some new conjugate gradient methods, *Applied Mathematics and Computation*, 2006, **183**: 1341–1350.

[15] G. H. Yu, L. T. Guan, and Z. X. Wei, A globally convergent Polak-Ribière-Polyak conjugate method with Armijo-type line search, *Numerical Mathematics*, 2006, **15**: 357–366.

[16] Q. Liu, C. Y. Wang, and X. M. Yang, On the convergence of a new hybrid projection algorithm, *Journal of Systems Science & Complexity*, 2006, **19**(13): 423–430.

[17] S. J. Lian, C. Y. Wang, and L. X. Cao, Convergence properties of the dependent PRP conjugate gradient methods, *Journal of Systems Science & Complexity*, 2006, **19**(2): 288–296.

[18] N. Andrei, A Dai-Yuan conjugate gradient algorithm with sufficient descent and conjugacy conditions for unconstrained optimization, *Applied Mathematics Letters*, 2008, **21**: 165–171.