# A New Method for Unconstrained Optimization Problem

Zhiguang Zhang

Department of Mathematics, Dezhou University, Dezhou, Shandong 253023, China

E-mail: zhiguangzhang@126.com

### Abstract

This paper presents a new memory gradient method for unconstrained optimization problems. This method makes use of the current and previous multi-step iteration information to generate a new iteration and add the freedom of some parameters. Therefore it is suitable to solve large scale unconstrained optimization problems. The global convergence is proved under some mild conditions. Numerical experiments show the algorithm is efficient in many situations.

Keywords: Unconstrained optimization, Memory gradient method, Global convergence

#### 1. Introduction

Consider the unconstrained optimization problem

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

where  $R^n$  is an *n*-dimensional Euclidean space and  $f : R^n \to R^1$  is a continuously differentiable function. Most of the well-known iterative algorithms for solving (1) take the form

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

where  $d_k$  is a search direction of f(x) at  $x_k$  and  $\alpha_k$  is a positive step-size. Let  $x_k$  be the current iterative point, we denote  $\nabla f(x_k)$  by  $g_k$ ,  $f(x_k)$  by  $f_k$  and  $f(x^*)$  by  $f^*$ , respectively. Let

$$P(x \mid L(x_0)) = \max_{y \in L(x_0)} \|y - x\|, \ P(\partial B \mid L(x_0)) = \inf_{x \in \partial B} P(x \mid L(x_0)),$$

Where 
$$L(x_0) = \{x \in \mathbb{R}^n \mid f(x) \le f(x_0)\}$$
,  $\partial B$  is the border of  $B$ 

Many traditional methods for solving (1) are line search methods such as steepest descent method, Newton-type methods, conjugate gradient methods, etc. Generally, the conjugate gradient method is a useful technique for solving large scale problems because it avoids the computation and storage of some matrices. Memory gradient methods have these good qualities too (e.g., (Cantrell, J.W., 1969)(Miele, A. and Cantrell, J.W., 1969)(Yuan Yaxiang, Sun Wenyu, 1997) etc.).

In order to make full use of the current and previous multi-step iterative information to improve the capability of methods and guarantee them be convergent, some scholars studied memory gradient methods and super-memory gradient methods. These two methods, like conjugate gradient methods, are suitable to solve large scale optimization problems. They are more stable than conjugate gradient methods (e.g., (Cragg, E.E., and Levy, A.V., 1969)(Shi Zhenjun, and Shen J., 2005)(Shi Zhenjun, 2003), etc.), because they use more previous iterative information and add the freedom of selecting parameters. Taking advantage of the line search rule that was presented in (Shi Zhenjun, and Shen J., 2005), this paper presents a new memory gradient method and proves its global convergence under some mild conditions.

The paper is organized as follows. Section 2 describes the new memory gradient algorithm. Section 3 analyzes the global convergence under some mild conditions.

### 2. New Memory Gradient Method

We assume that

(H1) The objective function f(x) has a lower bound on the level set

 $L(x_0) = \{x \in \mathbb{R}^n \mid f(x) \le f(x_0)\}$ , where  $x_0$  is given.

(H2) The gradient  $g(x) = \nabla f(x)$  is Lipschitz continuous in an open convex set *B* that contains the level set  $L(x_0)$ , i.e. there exists L > 0 such that

$$\left\|g(x) - g(y)\right\| \le L \left\|x - y\right\|, \forall x, y \in B,$$
(3)

where B is satisfied with  $P(\partial B | L(x_0)) > 0$ .

Using the line search rule that was presented in (Shi Zhenjun, and Shen J., 2005), we present a new memory gradient algorithm.

New Algorithm:  $0 < \rho < \frac{2}{3}; 0 < \mu < 1; L_0 > 0; x_0 \in \mathbb{R}^n; k := 0$ 

**Step 1** If  $||g_k|| = 0$  then stop! Else go to step 2;

Step 2  $x_{k+1} = x_k + \frac{\alpha_k}{L_k} d_k(\alpha_k)$ , where  $\alpha_k$  is the maximum of  $\alpha \in \{s_k, s_k \rho, s_k \rho^2, \cdots\}$  that is satisfied

$$f_k - f(x_k + \frac{\alpha}{L_k} d_k(\alpha)) \ge -\mu \frac{\alpha}{L_k} [g_k^T d_k(\alpha) + \frac{1}{2} \alpha \|g_k\|^2].$$

$$\tag{4}$$

where

$$d_k(\alpha) = \begin{cases} -g_k, & \text{if } k = 0; \\ -(1-\alpha)g_k + \alpha d_{k-1}, & \text{if } k \ge 1, \end{cases}$$
(5)

with

$$s_{k} = \begin{cases} 1, & \text{if } k = 0; \\ \frac{\rho \|\mathbf{g}_{k}\|^{2}}{\|\mathbf{g}_{k}\|^{2} + |\mathbf{g}_{k}^{T} d_{k-1}|}, & \text{if } k \ge 1. \end{cases}$$
(6)

**Step 3**  $L_{k+1} = \max\left(L_k, \frac{\|g_{k+1} - g_k\|}{\|x_{k+1} - x_k\|}\right);$ 

**Step 4** Let k := k+1 and go to step 1.

Obviously, the algorithm has an important feature that the search direction and step-size are defined at each iteration. It does good to find more suitable search direction and step-size. For simplicity, we denote  $d_k(\alpha_k)$  by  $d_k$  throughout the paper.

**Lemma 1** If (H2) holds, then there exists  $0 < m_0 \le M_0$  such that  $m_0 \le L_k \le M_0$ .

**Lemma 2** For all  $k \ge 0$ , if there exists  $\alpha \in (0, s_k]$ , then  $g_k^T d_k(\alpha) \le -(1-\rho) \|g_k\|^2$ .

**Lemma 3** For all  $k \ge 0$  and  $\alpha \in (0, s_k]$ , we have  $||d_k(\alpha)|| \le \max_{0 \le i \le k} \{||g_i||\}$ .

The above three lemmas are easily proved.

### 3. Global Convergence

Lemma 4 If (H1) and (H2)hold, the new algorithm generates an infinite sequence  $\{x_k\}$ , then  $\{\|g_k\|\}$  and  $\{\|d_k\|\}$  have a bound.

**Proof** By Lemma 3, we only need prove that  $\{\|g_k\|\}$  has a bound. Let  $\delta_k = \max_{0 \le j \le k} \{\|g_j\|\}$ . Suppose  $\{\|g_k\|\}$  has not a bound, then  $\lim_{k \to +\infty} \delta_k = +\infty$ .

Therefore there exists an infinite subset  $N \subset \{0, 1, 2, \dots\}$  such that

$$\delta_k = \left\| g_k \right\|, \forall k \in N, \tag{7}$$

and

$$\|g_k\| = \delta_k \to +\infty (k \in N, k \to +\infty).$$
(8)

By lemma1, lemma 2 and (4), we can obtain

$$f_{k} - f(x_{k} + \frac{\alpha_{k}}{L_{k}}d_{k}) \ge -\mu \frac{\alpha_{k}}{L_{k}}[g_{k}^{T}d_{k}(\alpha_{k}) + \frac{1}{2}\alpha_{k} ||g_{k}||^{2}]$$

$$\geq \mu \frac{\alpha_{k}}{L_{k}} [(1-\rho) \|g_{k}\|^{2} - \frac{1}{2} s_{k} \|g_{k}\|^{2}] \geq \mu \frac{\alpha_{k}}{L_{k}} [(1-\rho) \|g_{k}\|^{2} - \frac{1}{2} \rho \|g_{k}\|^{2}]$$
$$= \mu \frac{\alpha_{k}}{L_{k}} (1-\frac{3}{2} \rho) \|g_{k}\|^{2} \geq \frac{\mu (1-\frac{3}{2} \rho)}{M_{0}} a_{k} \|g_{k}\|^{2}.$$

By (H1), we have

$$\sum_{k=0}^{\infty} \alpha_k \left\| \boldsymbol{g}_k \right\|^2 < +\infty \,, \tag{9}$$

Thus

$$\sum_{k \in \mathbb{N}} \alpha_k \left\| \boldsymbol{g}_k \right\|^2 \le \sum_{k=0}^{\infty} \alpha_k \left\| \boldsymbol{g}_k \right\|^2 < +\infty.$$
(10)

By lemma 3 and (7), when  $\alpha \in (0, s_k]$ , we have

$$\|d_{k}(\alpha)\|^{2} \leq \max_{0 \leq i \leq k} \{\|g_{i}\|^{2}\} \leq \delta_{k}^{2} = \|g_{k}\|^{2}, \forall k \in \mathbb{N}.$$
(11)

From (10) and (11), we obtain

$$\sum_{k \in N} \alpha_k \left\| d_k(\alpha) \right\|^2 < +\infty, \tag{12}$$

therefore

$$\alpha_k \left\| d_k(\alpha) \right\|^2 \to 0 (k \in N, k \to +\infty).$$
<sup>(13)</sup>

By lemma 2 and Cauchy-schwarz inequality, when  $\alpha \in (0, s_k]$ , we have

$$\|g_k\| \cdot \|d_k(\alpha)\| \ge -g_k^T d_k(\alpha) \ge (1-\rho) \|g_k\|^2,$$
  
therefore

$$\left\| d_k(\alpha) \right\| \ge (1-\rho) \left\| g_k \right\|. \tag{14}$$

From (8) and (14), we can obtain  $\|d_k(\alpha)\| \to +\infty (k \in N, k \to +\infty)$ , by (13), we have

$$\alpha_k \left\| d_k(\alpha) \right\| \to 0 (k \in N, k \to +\infty).$$
<sup>(15)</sup>

Thus

$$\lim_{k\in N, k\to\infty}\alpha_k=0$$

It follows, when  $\alpha = \alpha_k / \rho \in (0, s_k]$ , we have

$$f_k - f(x_k + \frac{\alpha}{L_k} d_k(\alpha) < -\mu \frac{\alpha}{L_k} [g_k^T d_k(\alpha) + \frac{1}{2} \alpha \|g_k\|^2]$$

$$f_k - f(x_k + \frac{\alpha}{L_k} d_k(\alpha)) < -\mu \frac{\alpha}{L_k} g_k^T d_k(\alpha) \, .$$

thus

By mean value theorem, there exists 
$$\theta_k \in [0,1]$$
 such that

$$f_k - f(x_k + \frac{\alpha}{L_k} d_k(\alpha) = -\frac{\alpha}{L_k} g(x_k + \theta_k \frac{\alpha}{L_k} d_k(\alpha))^T d_k(\alpha)$$

Therefore

$$g(x_k + \theta_k \frac{\alpha}{L_k} d_k(\alpha))^T d_k(\alpha) > \mu g_k^T d_k(\alpha).$$

By lemma 1, (H2) and (11), we obtain  $(1-\rho)(1-\mu) \|g_k\|^2 \le (1-\mu)(-g_k^T d_k(\alpha))$ 

$$\leq [g(x_{k} + \theta_{k} \frac{\alpha}{L_{k}} d_{k}(\alpha)) - g_{k}]^{T} d_{k}(\alpha)$$

$$\leq \|d_{k}(\alpha)\| \cdot \left\|g(x_{k} + \theta_{k} \frac{\alpha}{L_{k}} d_{k}(\alpha)) - g_{k}\right\|$$

$$\leq \|g_{k}\| \cdot \left\|g(x_{k} + \theta_{k} \frac{\alpha}{L_{k}} d_{k}(\alpha)) - g_{k}\right\|, \quad \forall k \in N,$$
Thus
$$(1 - \rho)(1 - \mu)\|g_{k}\| \leq \left\|g(x_{k} + \theta_{k} \frac{\alpha}{L_{k}} d_{k}(\alpha)) - g_{k}\right\|, \forall k \in N.$$

Thus

By (H2) and (15), we can obtain  $\|g_k\| \to 0 (k \in N, k \to +\infty)$ , which contradicts with (8). The above proof shows that  $\{\|g_k\|\}$  has a bound, therefore  $\{\|d_k\|\}$  has a bound. The proof is completed.

**Theorem** If (H1) and (H2) hold, the new algorithm generates an infinite sequence  $\{x_k\}$ , then

$$\lim_{k \to \infty} \left\| \boldsymbol{g}_k \right\| = 0 \,. \tag{16}$$

Suppose (16) does not hold, then there exists an infinite subsequence  $N_1$  and  $\varepsilon > 0$ , such that Proof  $\|\boldsymbol{g}_k\| \geq \varepsilon, \forall k \in N_1.$ 

From the proof process of lemma 4, we have

$$\varepsilon^{2} \sum_{k \in N_{1}} \alpha_{k} \leq \sum_{k \in N_{1}} \alpha_{k} \left\| g_{k} \right\|^{2} \leq \sum_{k=m}^{\infty} \alpha_{k} \left\| g_{k} \right\|^{2} < +\infty,$$

Therefore  $\alpha_k \to 0 \ (k \in N_1, k \to +\infty)$ . By lemma 4, we obtain that there exists M > 0 such that

$$\left\|\boldsymbol{d}_{k}\right\| \leq \boldsymbol{M}, \forall k \in \boldsymbol{N}.$$
(17)

By lemma 2, lemma 4, Cauchy-schwarz inequality and (17), we can obtain that when  $\alpha = \alpha_k / \rho$ , then

$$(1-\rho)(1-\mu) \|g_{k}\|^{2} \leq (1-\mu)(-g_{k}^{T}d_{k}(\alpha))$$

$$\leq (g(x_{k}+\theta_{k}\frac{\alpha}{L_{k}}d_{k}(\alpha))-g_{k})^{T}d_{k}(\alpha)$$

$$\leq \|d_{k}(\alpha)\| \cdot \|g(x_{k}+\theta_{k}\frac{\alpha}{L_{k}}d_{k}(\alpha))-g_{k}\|$$

$$\leq M \|g(x_{k}+\theta_{k}\frac{\alpha}{L_{k}}d_{k}(\alpha))-g_{k}\|.$$
(18)

Because  $\alpha_k \to 0 (k \in N_1, \ k \to +\infty)$  , by lemma 4 ,we obtain

$$\alpha \left\| d_k(\alpha) \right\| \to 0 (k \in N_1, \ k \to +\infty).$$
<sup>(19)</sup>

By (H2), (18) and (19), we obtain  $\|g_k\| \to 0 (k \in N_1, k \to +\infty)$ , which contradicts with previous supposition. Therefore (16) holds. The proof is completed.

## References

Cantrell, J.W. (1969). Relation between the memory gradient method and the Fletcher-Reeves method, J. Optim. Theory Appl., 4: 67-71.

Cragg, E.E., and Levy, A.V. (1969). Study on a supermemory gradient method for the minimization of functions, J. Optim, Theory Appl., 4: 191-205.

Miele, A. and Cantrell, J.w. (1969). Study on a memory gradient method for the minimization of functions, J. Optim. Theory and Appl., 1 3: 459-470.

Shi Zhenjun. (2003). A new memory gradient method under exact line search, *Asia-Pacific J. Oper. Res.*, 20(2): 275-284.

Shi Zhenjun, and Shen J. (2005). New inexact line search method for unconstrained optimization, J. Optim. Theory and Appl., 127: 425-446.

Yuan Yaxiang, Sun Wenyu. (1997). Theory and methods for optimization, Science press, Beijing.