Applied Mathematical Sciences, Vol. 5, 2011, no. 50, 2483 - 2490

An Improved Conjugate Gradient Method and it's Global Convergence

Sui Qidi

Department of Mathematics and Information Science
Zaozhuang University Zaozhuang, Shandong, P.R. China
sunmin_2008@yahoo.com.cn

Abstract. In this paper, we consider the conjugate gradient method for unconstrained optimization problems, and propose an improved conjugate gradient method which direction is always a sufficient descent direction for the objective function, and this property is independent of the line search used. Under mild conditions, the global convergence of the proposed method is proved. Finally, we present some numerical results to show the efficiency of the proposed method.

Keywords: Conjugate gradient method; Sufficiently descent property; Global convergence

I. INTRODUCTION

In this paper, we deal with conjugate gradient methods for solving the following unconstrained minimization problem:

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

where $f: R^n \to R^1$ is a nonlinear function whose gradient is denoted by $g(x)$. Iterative methods are often used for solving unconstrained optimization problems, and they are of the form

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

where $x_k \in R^n$ is the k-th approximation to a solution. α_k is a positive step size and $d_k \in R^n$ is a search direction, or more exactly, is a descent direction of $f(x)$ at x_k. Since Hestenes and Stiefel [1] first proposed a conjugate gradient method for

solving a linear system of equations with a symmetric positive definite coefficient matrix, many authors have paid much attention to study conjugate gradient methods. Recently, conjugate gradient methods are paid attention to as iterative methods for solving large-scale unconstrained optimization problems, because they do not need the storage of matrices. The search direction of conjugate gradient methods is defined by

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

where $\beta_k \in R$ is a scalar, and g_k denotes $g(x_k)$. Some well-known formulas for β_k are given below:

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2}, \beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})}$$

Considerable attentions have been made on the global convergence for the above methods. On the other hand, any new formulas for β_k are proposed recently, such as Wei et al.[4] gave a variant of the PRP method, that is

$$\beta_k = \frac{g_k^T(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1})}{\| g_{k-1} \|^2}.$$

An interesting property of the Wei's method is that is satisfies the Property(*) which was first introduced by Gilbert and Nocedal [5]. This shows that Wei's method possesses nice property as the PRP+ method [5] has. Lately, Zhang [6] take a little modification to Wei's method and construct a modified PRP method as follows,

$$\beta_k = \frac{\| g_k \|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} |g_k^T g_{k-1}|}{\| g_{k-1} \|^2}.$$

Under the strong Wolfe line search, Zhang's method has global convergence. The property of the sufficient descent condition

$$g_k^T d_k \leq -c \| g_k \|^2$$

(where $c > 0$ is a constant) is important for the convergent proof of the descent method. However, the sufficient descent property of Zhang's method depends on the line search used. In this paper, we modified Zhang's method further, and propose a

new descent conjugate gradient method which sufficient descent property is independent of the line search used.

II. ALGORITHM AND IT'S CONVERGENCE

In the global convergence analysis of many iterative methods, the following assumption is often used.

H1: The objective function $f(x)$ is continuously differentiable and has a lower bound on the level set $L_0 = \{x \mid f(x) \le f(x_0)\}$, where x_0 is the initial point..

H2: The gradient $g(x)$ is uniformly continuous on an open convex set B that contains the level set L_0,

Now we propose the improved descent conjugate gradient method as follows:

Algorithm 1: Improved Conjugate Gradient Method

Step 0: Given an initial point $x_0 \in R^n$ and parameters $\gamma \in (0,1), \rho \in (0,1)$, set $k := 0$;

Step 1: If $\|g_k\| = 0$, then stop; else, go to Step 2;

Step 2: Compute the descent direction d_k by

$$d_k = \begin{cases} -g_k, k = 0; \\ -(1+\beta_k \frac{g_k{}^T g_{k-1}}{\| g_k \|^2}) g_k + \beta_k g_{k-1}, k > 0. \end{cases} \tag{3}$$

where β_k is defined by the following formula

$$\beta_k = \frac{\| g_k \|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} \left| g_k^T g_{k-1} \right|}{| g_k{}^T d_{k-1} | + \| g_k \|^2}. \tag{4}$$

Step 3: Set $s_k = -g_k{}^T d_k / \| d_k \|^2$, and α_k is the largest α in $\{s_k, s_k\rho, s_k\rho^2, ...\}$ such that

$$f(x_k + \alpha d_k) \le f(x_k) + \gamma \alpha g_k{}^T d_k. \tag{5}$$

Set $x_{k+1} = x_k + \alpha_k d_k$, $k := k+1$, go to Step 1.

Remark 1. Obviously, if the exactly line search is used, then our parameter β_k (4) reduces to Zhang's parameter. However, our search direction (3) is different from Zhang's direction.

Lemma 1 For any $k \ge 0$, we have $g_k{}^T d_k = -\| g_k \|^2$.

Proof: It's quite easy, thus omitted.

Remark 2 From Lemma 1, the search direction d_k generated by Algorithm 1 is always a descent direction for all k, which implies that the generated $\{f(x_k)\}$ is a descent sequence.

Lemma 2 For all $k \ge 0$, there is a constant $M > 0$, such that

$$\| d_k \| \le M \delta_k. \tag{6}$$

where $\delta_k = \max_{0 \le i \le k} \{\| g_i \|\}$.

Proof: From the definition of β_k, we have

$$\beta_k = \frac{\| g_k \|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} \left| g_k^T g_{k-1} \right|}{| g_k{}^T d_{k-1} | + \| g_k \|^2} \le \frac{\| g_k \|^2 + \frac{\|g_k\|}{\|g_{k-1}\|} \|g_k\| \|g_{k-1}\|}{\| g_k \|^2} = 2$$

and

$$\beta_k = \frac{\| g_k \|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} \left| g_k^T g_{k-1} \right|}{| g_k{}^T d_{k-1} | + \| g_k \|^2} \ge \frac{\| g_k \|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} \|g_k\| \|g_{k-1}\|}{| g_k{}^T d_{k-1} | + \| g_k \|^2} = 0.$$

Therefore, from (3) and Cauchy-Schwartz Inequality, we have

$$\| d_k \| \le \| g_k \| + \beta_k \frac{\| g_k \| \| g_{k-1} \|}{\| g_k \|^2} \| g_k \| + \beta_k \| g_{k-1} \| \le \| g_k \| + 4 \| g_{k-1} \| \le 5 \delta_k.$$

Set $M = 5$ and we get the desired result.

From Lemma 1 and Lemma 2, we can establish the convergent result of

Algorithm 1.The following proof is similar to that of Theorem 1 in [3]. For complement, we will give the detailed proof.

Theorem 1 Let $\{x_k\}$ be generated by Algorithm 1 and Assumption H1, H2 holds, then any cluster point of the sequence $\{x_k\}$ is a stationary point of $f(x)$.

Proof: From (5) and Lemma 1, we have

$$\lim_{k\to+\infty} \alpha_k \| g_k \|^2 = 0. \tag{7}$$

Set $\bar{x}$ is any cluster point of the sequence $\{x_k\}$, then there is a subsequence $\{x_k : k \in K_1\}$ such that

$$\lim_{k\to+\infty, k\in K_1} x_k = \bar{x}.$$

If $\inf_{k\in K_1} \alpha_k > 0$, then from (7), it is easy to get the desired result.

If $\inf_{k\in K_1} \alpha_k = 0$, then there is a subset $K_2 \subseteq K_1$, such that $\lim_{k\to+\infty, k\in K_2} \alpha_k = 0$ .Then, we have the following result

$$\lim_{k\to+\infty, k\in K_2} \| g_k \|^2 / \| d_k \| = 0. \tag{8}$$

In fact, if the above inequality doesn't hold, then there is $\varepsilon > 0$, and subset $K_3 \subseteq K_2$, such that

$$\| g_k \|^2 / \| d_k \| \geq \varepsilon. \tag{9}$$

Therefore, from (7) and (9), we have

$$\lim_{k\to+\infty, k\in K_3} \alpha_k \| d_k \| = 0. \tag{10}$$

From the choice rule of α_k, there is $\bar{k} > 0$, such that for any $k > \bar{k}, k \in K_2$, we get

$$f(x_k + \alpha_k d_k / \rho) > f(x_k) + \gamma \alpha_k g_k^T d_k / \rho.$$

then from the Mean Medium Theorem, we have

$$(g(\xi_k) - g_k)^T d_k \geq (\gamma - 1) g_k^T d_k.$$

where $\xi_k \in (x_k, x_k + \alpha_k d_k / \rho)$. Therefore, for all $k \in K_2$, we have

$$\|g(\xi_k)-g_k\| \geq \frac{\|g(\xi_k)-g_k\|\cdot\|d_k\|}{\|d_k\|} \geq \frac{(g(\xi_k)-g_k)^T d_k}{\|d_k\|} \geq (\gamma-1)\frac{g_k{}^T d_k}{\|d_k\|} = (1-\gamma)\frac{\|g_k\|^2}{\|d_k\|}.$$

Then from (H2) and (10), we have

$$\lim_{k\to+\infty,k\in K_3} \| g_k \|^2 / \| d_k \| = 0. \tag{11}$$

which is contradict with (9) , therefore (8) holds. From Lemma 2 and (8), we have

$$\lim_{k\to+\infty,k\in K_3} \| g_k \|^2 / \delta_k = 0. \tag{12}$$

If the sequence $\{\| g_k \|\}(k \in K_3)$ is unbounded, then we have $\delta_k = \max_{0\leq i\leq k}\{\| g_i \|\} \to \infty(k \in K_3)$, therefore there is an infinite index set $K_4 \subseteq K_3$, such that $\| g_k \| = \delta_k$. Then from (12), we get the desired result. If the sequence $\{\| g_k \|\}(k \in K_3)$ is bounded, then there is a constant $N > 0$, for any $k \in K_3$, we have $\delta_k \leq N$.

Then from (12), we also get the desired result. The proof is complete.

In fact, under the following assumption, we can get the global convergence of Algorithm 1.

(H3) The gradient $g(x)$ is Lipschitz continuous on an open convex set B that contains the level set L_0, and the Lipschitz constant is denoted by L.

Lemma 3 If H1 and H3 hold, then for any $k \geq 0$, we have $\alpha_k \geq -\frac{g_k{}^T d_k}{L\| d_k \|^2}$.

Proof: The proof is similar to Lemma 3 in [6], thus omitted.

Theorem 2 Let $\{x_k\}$ be generated by Algorithm 1 and Assumption H1, H3 holds, then $\lim_{k\to+\infty}\|g_k\| = 0$.

Proof: From Lemma 1, Lemma 3 and the line search rule (5), we have

$$\sum_{k=1}^{\infty}\left(\frac{\| g_k \|^2}{\| d_k \|}\right)^2 < +\infty.$$

Therefore $\lim_{k\to+\infty}\frac{\| g_k \|^2}{\| d_k \|}=0$. The following proof is similar to that of Theorem 1. This completes the proof.

III NUMERICAL REPORTS

In the following numerical experiments, the parameters are set as $\gamma=0.23, \rho=0.85$.

The program was coded in Matlab7.0, and tested on PIV2.8GHz personal computer.

Problem 1. $f(x)=x_1^2+x_2^2+2x_3^2+x_4^2-5(x_1+x_2)-21x_3+7x_4$, the initial point $x_0=(1,1,1,1)$.

Problem 2. $f(x)=\sum_{i=1}^{n}(e^{x_i}-x_i)$, the initial point $x_0=(n/(n-1),...,n/(n-1))$, where $n=100$.

Problem 3. $f(x)=10(x_1^{\,2}-x_2)^2+(1-x_1)^2+9(x_4-x_3^{\,2})^2+(1-x_3)^2$

$$+10.1[(x_2-1)^2+(x_4-1)^2]+19.8(x_2-1)(x_4-1).$$

the initial point $x_0=(-3,-1,-3,-1)$.

Problem 4. $f(x)=(x_1+10x_2)^4+5(x_3-x_4)^4+(x_2-2x_3)^4+10(x_1-x_4)^4$ the initial point $x_0=(2,2,-2,-2)$.

Table 1 Numerical of Iteration and CPU Time

Problem	Ite.	CPU	$\| g(x_k) \|$
1	36	0.01	9.4257e-007
2	28	0.01	8.3562e-007
3	28	0.01	9.8752e-007
4	29	0.01	8.6742e-007

REFERENCES

[1] Hestenes M.R., Stiefel E., Methods of conjugate gradients for solving liner systems. Journal of Research of the National Bureau of Standards, 1952, 49: 409-436.

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

[3] Sun M., A modified super-memory gradient method with angle property. Journal of Shandong University(Natural Science), 2008, 43 ,68-70.

[4] Wei Z.X., Yao S., Liu L., The convergence properties of some new conjugate gradient methods. Applied Mathematics and Computation, 2006, 183, 1341-1350

[5] Gilbert J.C., Nocedal J., Global convergence properties of conjugate gradient methods for optimization, SIAM Journal on Optimization, 1992, 2, 21-42.

[6] Zhang L. An improved Wei-Yao-Liu nonlinear conjugate gradient method for optimization computation. Applied Mathematics and Computation, 2009, 215, 2269-2274

[7] Shi Z.J.. A new Super-memory gradient method for unconstrained optimization Advance in Mathematics, 2006, 35(3), 265-274.

[8] Sun M, Liu J., A new family of conjugate gradient method with Armijo line search Int. Journal of Math. Analysis, 2010, 20,987-993.

Received: February, 2011