TELKOMNIKA, Vol.11, No.1, January 2013, pp. 544~549
ISSN: 2302-4046

Global Convergence of A Kind of Conjugate Gradient Method

Can Li[1], Ling Fang[2], Xianglian Cao[1]
[1]College of Mathematics, Honghe University, Mengzi, China
[2]Department of Foundation Studies, Logistical Engineering University of PLA, Chongqing,China
*Corresponding author, e-mail: lican841014@163.com

Abstract

The conjugate gradient method is welcome method for solving optimization problems due to its simplicity and low storage. In this paper, we propose a kind of conjugate gradient method. The presented method possesses the sufficient descent property under the strong Wolfe line search. Under mild conditions, we prove that the method with strong Wolfe line search is globally convergent even if the objective function is nonconvex. At the end of this paper, we also present numerical experiments to show the efficiency of the proposed method.

Keywords*: unconstrained optimization problem, conjugate gradient method, strong wolfe line search, global convergence.*

1. Introduction

Optimization is an important tool in many areas[1-3], such as engineering, production management, economy etc.. In this paper, we are interested to consider the following unconstrained optimization problem

$$\min f(x), \tag{1}$$

where $f : R^n \to R$ is continuously differentiable and its gradient $g(x) = \nabla f(x)$ is available.The conjugate gradient method is a powerful line search method for solving (1) because of its simplicity and its very low memory requirement, especially for the large scale optimization problems. The following iterative formula is often used by the nonlinear conjugate gradient method

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

where α_k is a steplength and d_k is a search direction defined by

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

where $g_k = g(x_k)$, β_k is a scalar which determines the different conjugate gradient methods. Well-known conjugate gradient methods are the Fletcher-Reeves method, Polak-Ribiere-Polyak method, Conjugate –Descent method and Dai-Yuan method etc..

The convergence behavior of the different conjugate gradient methods with some line search conditions has been widely studied, including Armijo [4], Al-baali [5], Dai [6] and Zhang [7-8] etc.. There are many convergence results of the conjugate gradient methods [6-11]. The sufficient descent condition

$$g_k^{\mathrm{T}} d_k \le -c \| g_k \|^2, \forall k,$$

where $c > 0$ is a constant, is crucial to insure the global convergence of the conjugate gradient methods.

In the next section, the algorithm and its property are stated. In section 3, the global convergence of the new method is proved. In section 4, we report some numerical results to test the proposed method.

Received October 1, 2012; Revised December 8, 2012; Accepted December 14, 2012

2. Algorithm

Now we describe our algorithm as follows.

Algorithm 1:

Step 1. Given $x_1 \in R^n, \varepsilon \geq 0,\ 0 < \delta < \sigma < 1$, Let $d_1 = -g_1, k := 1$;

Step 2. If $\|g_k\| \leq \varepsilon$, then stop, otherwise go to Step 3;

Step 3. Computer d_k by (3), where β_k defined by

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

where $\lambda \in [0,+\infty), \mu \in (\lambda,+\infty)$;

Step 4. Determined α_k by strong Wolfe line search satisfying

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{5}$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq -\sigma g_k^T d_k, \tag{6}$$

where $0 < \delta < \sigma < 1$;

Step 5. Let $x_{k+1} = x_k + \alpha_k d_k$;

Step 6. Set $k := k+1$, and go to Step 2.

Theorem 1 Suppose the sequences $\{x_k\}$ and $\{d_k\}$ are generated by the Algorithm 1, then

$$g_k^T d_k < -\frac{\lambda}{\mu}\|g_k\|^2, \forall k \geq 1. \tag{7}$$

Proof (1) If $k=1$, $d_1 = -g_1$, $g_1^T d_1 = -\|g_1\|^2$. It is obviously that (7) holds.

(2) Assuming now that (7) is true for some $k-1$, namely

$$g_{k-1}^T d_{k-1} < -\frac{\lambda}{\mu}\| g_{k-1} \|^2 < 0. \tag{8}$$

We show that (7) continue to hold for k .

Since $\mu - \lambda > 0$, we get from (4), (8) and μ >0 that $\beta_k > 0$. By the definition of d_k, we have from (6) and $0 < \sigma < 1$ that

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \beta_k g_k^T d_{k-1} \\ &\leq -\|g_k\|^2 + \frac{(\mu-\lambda)\|g_k\|^2}{(1+\mu)\|g_{k-1}\|^2 - \mu g_{k-1}^T d_{k-1}} \left| g_k^T d_{k-1} \right| \\ &\leq -\|g_k\|^2 + \frac{-(\mu-\lambda)\sigma g_{k-1}^T d_{k-1}}{(1+\mu)\|g_{k-1}\|^2 - \mu g_{k-1}^T d_{k-1}} \|g_k\|^2 \\ &\leq -\|g_k\|^2 + \frac{\sigma(\mu-\lambda)}{\mu}\|g_k\|^2 \\ &< -\|g_k\|^2 + \frac{\mu-\lambda}{\mu}\|g_k\|^2 \\ &= -\frac{\lambda}{\mu}\|g_k\|^2 . \end{aligned}$$

This implies that d_k provides a descent direction of f at x_k .

3. Global Convergence

In this section, we prove the global convergence of Algorithm 1 under the following assumption.

Assumption A

(1)The level set $\Omega = \{x \in R^n \mid f(x) \le f(x_1)\}$ is bound.

(2)In some neighborhood N of Ω, $f(x)$ is continuously differentiable and its gradient $g(x)$ is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \le L\|x - y\|, \forall x, y \in N.$$

In the latter part of the paper, we always suppose the conditions Assumption A hold.

Lemma 1[12] Let $\{x_k\}$ be generated by (2) and (3), $\{d_k\}$ satisfy $g_k^{\mathrm{T}} d_k \le 0$ and α_k be determined by strong Wolfe line search. Then the Zoutendijk condition holds

$$\sum_{k=1}^{\infty} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} < +\infty. \tag{9}$$

Lemma 2 Let $\{x_k\}$ be generated by Algorithm 1, then

$$\frac{1 - 2\sigma + \sigma^k}{1 - \sigma} \le \frac{-g_k^T d_k}{\|g_k\|^2} \le \frac{1 - \sigma^k}{1 - \sigma} \qquad \forall k \ge 1. \tag{10}$$

Proof (1) If $k = 1$, $d_1 = -g_1$, $g_1^T d_1 = -\|g_1\|^2$. It is obviously that (10) holds.

(2) Assuming that (10) is true for some $k - 1$, namely

$$\frac{1 - 2\sigma + \sigma^{k-1}}{1 - \sigma} \le \frac{-g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} \le \frac{1 - \sigma^{k-1}}{1 - \sigma}. \tag{11}$$

We show that (10) continue to hold for k . By the definition of d_k , we have from (6) , (11) and μ >0 that

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

$$\le 1+\sigma\frac{1-\sigma^{k-1}}{1-\sigma}=\frac{1-\sigma^{k}}{1-\sigma}\ .$$

Repeating the process, we have

$$\frac{-g_k^T d_k}{\|g_k\|^2}\ge 1-\frac{-(\mu-\lambda)\sigma g_{k-1}^T d_{k-1}}{(1+\mu)\|g_{k-1}\|^2-\mu g_{k-1}^T d_{k-1}}$$

$$\ge 1-\sigma\frac{\mu-\lambda}{1+\mu}\frac{-g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2}$$

$$\ge 1-\sigma\frac{\mu}{1+\mu}\frac{-g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2}$$

$$\ge 1-\sigma\frac{-g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2}$$

$$\ge 1-\sigma\frac{1-\sigma^{k-1}}{1-\sigma}$$

$$=\frac{1-2\sigma+\sigma^{k}}{1-\sigma}\ .$$

Theorem 2 Let $\{x_k\}$ be generated by Algorithm 1 and $\sigma\in(0,\frac{1}{2})$, then

$$\liminf_{k\to\infty}\|g_k\|=0\,. \tag{12}$$

Proof We have from (10) and $\sigma\in(0,\frac{1}{2})$ that

$$0<\frac{1-2\sigma}{1-\sigma}\le\frac{-g_k^T d_k}{\|g_k\|^2}\le\frac{1}{1-\sigma}\,. \tag{13}$$

This together with (6), we obtain

$$-g_{k+1}^T d_k\le-\sigma g_k^T d_k\le\frac{\sigma}{1-\sigma}\|g_k\|^2. \tag{14}$$

By the definition of β_k , we get from (7) and $\mu>0$ that

$$\beta_k\le\frac{(\mu-\lambda)\|g_k\|^2}{(1+\mu)\|g_{k-1}\|^2}$$

$$\le\frac{\mu}{1+\mu}\frac{\|g_k\|^2}{\|g_{k-1}\|^2}$$

$$<\frac{\|g_k\|^2}{\|g_{k-1}\|^2}\,. \tag{15}$$

Thus, using (14), we have

$$-2\beta_{k+1}g_{k+1}^T d_k\le\frac{2\sigma}{1-\sigma}\|g_{k+1}\|^2. \tag{16}$$

By the definition of d_k , we have from (15) , (16) that

$$\|d_{k+1}\|^2=\|g_{k+1}\|^2+(\beta_{k+1})^2\|d_k\|^2-2\beta_{k+1}g_{k+1}^T d_k$$

$$\le \|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^4}{\|g_k\|^4}\|d_k\|^2 + \frac{2\sigma}{1-\sigma}\|g_{k+1}\|^2$$

$$= \frac{\|g_{k+1}\|^4}{\|g_k\|^4}\|d_k\|^2 + \frac{1+\sigma}{1-\sigma}\|g_{k+1}\|^2.$$

Let

$$p_k = \frac{\|d_k\|^2}{\|g_k\|^4},$$

then

$$p_{k+1} \le p_k + \frac{1+\sigma}{1-\sigma}\frac{1}{\|g_{k+1}\|^2}. \tag{17}$$

For the sake of contradiction, we suppose that the conclusion is not true. Then there exists a constant $\varepsilon > 0$ such that

$$\|g_k\| \ge \varepsilon \qquad \forall k \ge 1. \tag{18}$$

With (17), we get

$$p_{k+1} \le p_1 + \lambda k,$$

where $\lambda = \dfrac{1+\sigma}{(1-\sigma)\varepsilon^2}$. Furthermore, we have from (13) that

$$\sum_{k=1}^{\infty}\frac{(g_k^T d_k)^2}{\|d_k\|^2} = \sum_{k=1}^{\infty}\left(\frac{-g_k^T d_k}{\|g_k\|^2}\right)^2 \frac{\|g_k\|^4}{\|d_k\|^2}$$

$$= \sum_{k=1}^{\infty}\left(\frac{-g_k^T d_k}{\|g_k\|^2}\right)^2 \frac{1}{p_k}$$

$$\ge \sum_{k=1}^{\infty}\frac{t}{p_1 + \lambda(k-1)} = \infty,$$

where $t = \left(\dfrac{1-2\sigma}{1-\sigma}\right)^2$. This contradicts Zoutendijk condition (9).

4. Numerical Experiments

In this section, we report some results of the numerical experiments. We test Algorithm 1 and compare its performance with those of PRP method whose results be given by [13].

AL1SWP: Algorithm 1 with SWP line search, where $\lambda = 0.2$, , $\mu = 0.5$, $\delta_1 = 0.01$ and $\delta_2 = 0.1$.

PRPSWP[13]: The PRP formula with SWP line search, where $\delta_1 = 0.01$ and $\delta_2 = 0.1$.

The termination condition is $\| g_k \| \le 10^{-5}$ or the iteration number exceeds 2×10^4 or the function evaluation number exceeds 3×10^5.

In the following tables, the numerical results are written in the form NI/NF/NG, where NI, NF, NG denote the number of iteration, function and gradient evaluations respectively. Dim denotes the dimension of the test problems.

From the numerical results, we can see that the proposed method performs better than the PRP method for some problems.

Table 1. Numerical Results

Problem	Dim	AL1SWP NI/NF/NG	PRPSWP NI/NF/NG
rose	2	42/156/52	29/502/65
froth	2	18/49/22	12/30/20
jensam	2	11/33/17	-/-/-
helix	3	45/155/55	49/255/83
bard	3	27/130/31	23/98/37
gulf	3	1/2/2	1/2/2/
box	3	28/91/39	-/-/-
wood	4	197/308/222	337/2125/599
kowosb	4	36/138/37	62/361/105
bd	4	138/597/207	-/-/-
osb1	5	28/44/34	1/51/2
biggs	6	142/247/159	121/495/197
osb2	11	338/612/424	293/1372/480
watson	20	561/762/565	990/2773/1567
rosex	50	37/53/53	31/533/76
pen2	50	102/239/136	906/4057/1585
vardim	50	9/271/179	10/52/36
trig	100	38/241/71	46/342/87
ie	500	5/8/6	6/13/8
rrid	200	36/241/142	30/66/36
band	50	16/117/70	18/183/24
band	200	16/19/19	19/283/27

References

[1] Hu W, Abubakar A, Habashy T M, Liu J. Preconditioned nonlinear conjugate gradient method for frequency domain full-waveform seismic inversion. *Geophysical Prospectiing.* 2011; 59(3):477-491.

[2] Seman K, Puspita F M, Taib B M, Shafii Z. An improved optimization model of internet charging scheme in muti service networks. *TELKOMNIKA.* 2012; 10(3):592-598.

[3] Asady B, Mansouri P, Gupta N. The modify version of artificial bee colony algorithm to solve real optimization problems. *International Journal of Electrical and Computer Engineering.* 2012; 2(4):473-480.

[4] Armijo L. Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics.* 1966; 16(1):1-3.

[5] Al-baali A. Descent property and global convergence of the Fletcher-Reeves method with inexact line search. *IMA Journal of Numerical Analysis.* 1985; 5(1):121-124.

[6] Dai Y H, Yuan Y. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM J. Optim.*.1999; 10(1): 177-182.

[7] Zhang L, Zhou W J, Li D H. Global convergence of a modified Fletcheer-Reeves conjugate gradient method with Arnijo-type line search. *Numerical Mathematics.* 2006; 104(4):561-572.

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

[9] Andrei N. Another hybrid conjugate gradient algorithm for uncontrained optimization. *Numerical Algorithm.* 2008; 47(2):143-156.

[10] Andrei N. Hybrid conjugate gradient algorithm for uncontrained optimization. *Journal of Optimization Theory and Applications.* 2009; 141(2):249-264.

[11] Birgin E, Martinez J M. A spectral conjugate gradient method for unconstrained optimization. *Appl.Math.Optim.*. 2001; 43(2):117-128.

[12] Zoutendijk G. Nonlinear programming computational methods. *Amsterdam: North-Holland.* 1970, 37-86.

[13] Wei Z X, Yao S W, Liu L Y. The convergence properties of some new conjugate gradient methods. *Applied Mathematics and Computation.* 2006; 183(2):1341-1350.