A Modified Conjugate Gradient Method and Its Global Convergence

Can Li

College of Mathematics, Honghe University, Mengzi 661199, China

***Abstract* — Conjugate gradient method is an important method for solving large-scale unconstrained nonlinear optimization problems. In this paper, we take a little modification to CD conjugate gradient method and a modified CD conjugate gradient method is proposed. The presented method has sufficient descent properties. Under mild conditions, we prove that the method with strong Wolfe line search is globally convergent. At the end of this paper, we also present numerical experiments to show the efficiency of the proposed method.**

Keywords - unconstrained optimization; conjugate gradient method; sufficiently descent property; line search; global convergence

I. INTRODUCTION

Consider the unconstrained optimization problem

$$\min f(x) \quad , \quad x \in R^n \quad ,$$

(1)

where $f : R^n \to R$ is a continuously differentiable function. The conjugate gradient method for solving (1) is given by

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

(2)

where α_k is the step length, and d_k is search direction defined by

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

(3)

where $g_k = \nabla f(x_k)$, β_k is a parameter.

Generally, different conjugate gradient methods correspond to the different ways to compute β_k. Over the years, some famous CG methods have been widely studied, such as Fletcher–Reeves (FR)method , Hestenes–Stiefel (HS) method , Polak–Ribière –Polyak (PRP) method , Liu–Storey (LS) method , and Dai–Yuan (DY) method , Conjugate-Descent (CD) method, etc. In the CD method [1], the parameter β_k is specified by

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

where $\|\cdot\|$ stands for the Euclidean norm of vectors.

The conjugate descent (CD) method ensures a descent direction for general functions if the line search satisfies the Wolfe conditions, namely

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k g_k^T d_k ,$$

(4)

and

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

(5)

where $0 < \delta < \sigma < 1$. But, the global convergence of the method is proved [2] only for the case when the line search satisfies (4) and

$$\sigma g_k^T d_k \le g(x_k + \alpha_k d_k)^T d_k \le 0.$$

(6)

An example is constructed in [2] showing that the CD conjugate descent method need not converge with α_k satisfying (4) and

$$\sigma_1 g_k^T d_k \le g(x_k + \alpha_k d_k)^T d_k \le -\sigma_2 g_k^T d_k .$$

(7)

where $\sigma_2 > 0$. In [3], Dai and Yuan proposed a family of globally convergent conjugate methods, in which

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

where $0 \le \lambda \le 1$ is a parameter and $y_{k-1} = g_k - g_{k-1}$, and proved that the family of methods using line searches that satisfy (4) and (7) converges globally if the parameters σ_1, σ_2 and λ are such that $\sigma_1 + \sigma_2 = \lambda^{-1}$. Based on the above research, Sellami, Laskri and Benzine [4] proposed a new two-parameter family of conjugate gradient methods in which

$$\beta_k = \frac{(1-\lambda_k)\|g_k\|^2 + \lambda_k(-g_k^{\mathrm{T}} d_k)}{(1-\lambda_k-\mu_k)\|g_{k-1}\|^2 + (\lambda_k+\mu_k)(-g_{k-1}^{\mathrm{T}} d_{k-1})},$$

where $\lambda_k \in [0,1]$ and $\mu_k \in [0, 1-\lambda_k]$ are parameters. The two-parameter family of methods with the Wolfe line search is shown to ensure the descent property of each search direction. Some general convergence results are also established for the two-parameter family of methods.

In this paper, we present our new β_k and algorithm in Section 2. The global convergence and the proof of the proposed method are presented in Section 3.

II. ALGORITHM

In this section, we take a little modification to CD conjugate gradient method. In the search direction, we define the parameter β_k as follow

$$\beta_k^*(\lambda,\mu)=\frac{(\mu-\lambda)\|g_k\|^2}{(1+\mu-\lambda)\|g_{k-1}\|^2-\mu g_{k-1}^T d_{k-1}}, \tag{8}$$

where $\lambda\in[0,+\infty),\mu\in(\lambda,+\infty)$. The corresponding method is described as below.

Algorithm 1(the modified CD conjugate gradient method **).**

Step1:Given constants $\varepsilon>0$, $0<\delta<\sigma<1$, $\lambda\in[0,+\infty)$, $\mu\in(\lambda,+\infty)$, choose an initial point $x_1\in D$, let $d_1=-g_1$ and $k:=1$.

Step2: If $\|g_k\|\le\varepsilon$, stop, get solution x_k. Otherwise go to step 3.

Step3: Compute β_k by (8) and generate d_k by (3).

Step4: Determine step size α_k by strong Wolfe line search

$$f(x_k+\alpha_k d_k)-f(x_k)\le\delta\alpha_k g_k^T d_k,$$

and

$$\left|g(x_k+\alpha_k d_k)^T d_k\right|\le-\sigma g_k^T d_k,$$

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

Step5: Let $x_{k+1}=x_k+\alpha_k d_k$ and $k:=k+1$.Go to step2.

Theorem 1 Let the sequence $\{d_k\}$ be generated by the modified CD conjugate gradient method, then

$$g_k^T d_k<-\frac{\lambda}{\mu}\|g_k\|^2,\forall k\ge1. \tag{9}$$

Proof (1) Since $d_1=-g_1$, we get

$$g_1^T d_1=-\|g_1\|^2.$$

By the definition of β_k, we know that $\lambda\in[0,+\infty)$, $\mu\in(\lambda,+\infty)$, so $-\frac{\lambda}{\mu}>-1$. Thus

$$g_1^T d_1=-\|g_1\|^2<-\frac{\lambda}{\mu}\|g_1\|^2.$$

(2) By the definition of d_k, we get

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

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

Since $\lambda\in[0,+\infty),\mu\in(\lambda,+\infty)$ and $g_{k-1}^T d_{k-1}<0$, we obtain

$$g_k^T d_k\le-\|g_k\|^2+\beta_k^*\left|g_k^T d_{k-1}\right|$$

$$=-\|g_k\|^2+\frac{(\mu-\lambda)\|g_k\|^2}{(1+\mu-\lambda)\|g_{k-1}\|^2-\mu g_{k-1}^T d_{k-1}}\left|g_k^T d_{k-1}\right|.$$

By the second condition of the strong Wolfe line search, we have

$$g_k^T d_k$$

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

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

$$=-\|g_k\|^2+\frac{\sigma(\mu-\lambda)}{\mu}\|g_k\|^2$$

Since $0<\sigma<1$ and $\frac{\mu-\lambda}{\mu}\|g_k\|^2>0,$ we further deduce that

$$g_k^T d_k<-\|g_k\|^2+\frac{\mu-\lambda}{\mu}\|g_k\|^2$$

$$=-\|g_k\|^2+(1-\frac{\lambda}{\mu})\|g_k\|^2=-\frac{\lambda}{\mu}\|g_k\|^2.$$

The theorem implies that the search direction produced by the modified CD conjugate gradient method is always descent direction.

III. GLOBAL CONVERGENCE

In this section, we prove the global convergence of the modified CD conjugate gradient method. To establish the global convergence theorem of the proposed method, we assume that the objective function satisfies the following conditions.

Assumption A

(1) The level set $\Omega=\{x\in R^n \mid f(x)\le f(x_0)\}$ is bounded.

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

$$\|\nabla f(x)-\nabla f(y)\|\le L\|x-y\|,\ \forall x,y\in N. \tag{10}$$

In the latter part of the paper, we always suppose that the conditions in Assumption A hold. The following important result was obtained by Zoutendijk [5] and Wolfe [6].

Theorem 2 Consider that Assumption A is true. Consider any iteration method of the form (2) and (3), where d_k satisfies $g_k^T d_k \le 0$ and α_k is obtained by the Wolfe line search. Then

$$\sum_{k\ge1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty . \quad (11)$$

The above theorem holds when α_k is obtained by the strong Wolfe line search.

Theorem 3 Let the sequence $\{x_k\}$ be generated by the modified CD conjugate gradient method, then

$$\frac{1-2\sigma+\sigma^k}{1-\sigma} \le \frac{-g_k^T d_k}{\|g_k\|^2} \le \frac{1-\sigma^k}{1-\sigma} . \quad (12)$$

Proof By the definition of d_k and β_k, we get

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

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

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

$$= 1+\frac{-(\mu-\lambda) g_k^T d_{k-1}}{(1+\mu-\lambda)\|g_{k-1}\|^2-\mu g_{k-1}^T d_{k-1}} . \quad (13)$$

The second condition of the strong Wolfe line search can be written as

$$\sigma g_{k-1}^T d_{k-1} \le -g_k^T d_{k-1} \le -\sigma g_{k-1}^T d_{k-1}. \quad (14)$$

We get from (11) and (14) that

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

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

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

Since $\frac{-g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} \le \frac{1-\sigma^{k-1}}{1-\sigma}$, we get

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

In the same way, we can deduce the other side of the inequality as follow.

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

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

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

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

Theorem 4 Let the sequence $\{x_k\}$ be generated by the modified CD conjugate gradient method and $\sigma \in (0,\frac{1}{2})$, then

$$\lim_{k\to\infty}\inf\|g_k\| = 0 . \quad (15)$$

Proof we get from Theorem 2 that

$$0 < \frac{1-2\sigma}{1-\sigma} \le \frac{-g_k^T d_k}{\|g_k\|^2} \le \frac{1}{1-\sigma} . \quad (16)$$

Then we have

$$-g_k^T d_k \le \frac{\|g_k\|^2}{1-\sigma} . \quad (17)$$

By the definition of β_k and $-\mu g_{k-1}^T d_{k-1} > 0$, $\lambda \ge 0$, we get

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

$$\le \frac{(\mu-\lambda)\|g_k\|^2}{(1+\mu-\lambda)\|g_{k-1}\|^2} < \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \quad (18)$$

Squaring both sides of $d_{k+1} = -g_{k+1}+\beta_{k+1}^* d_k$, we have

$$\|d_{k+1}\|^2 = \|g_{k+1}\|^2 + (\beta_{k+1}^*)^2\|d_k\|^2 - 2\beta_{k+1}^* g_{k+1}^T d_k \quad (19)$$

Combining (14) and (17), we obtain

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

Both sides of the inequality are multiplied by $2\beta_{k+1}^*$ simultaneously, we get from (18) that

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

Substituting it to (19), we have

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

$$\le \frac{\|g_{k+1}\|^4}{\|g_k\|^4}\|d_k\|^2 + \frac{1+\sigma}{1-\sigma}\|g_{k+1}\|^2. \tag{20}$$

Let $t_k = \frac{\|d_k\|^2}{\|g_k\|^4}$, we deduce that

$$t_{k+1} \le t_k + \frac{1+\sigma}{1-\sigma}\frac{1}{\|g_{k+1}\|^2}. \tag{21}$$

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

$$\|g_k\| \ge \varepsilon, \forall k \ge 1. \tag{22}$$

Combining (21) and (22)，we obtain

$$t_{k+1} \le t_1 + pk, p = \frac{1+\sigma}{(1-\sigma)\varepsilon^2}. \tag{23}$$

However,

$$\sum_{k\ge1}^{\infty}\frac{(g_k^T d_k)^2}{\|d_k\|^2} = \sum_{k\ge1}^{\infty}\left(\frac{-g_k^T d_k}{\|g_k\|^2}\right)^2\frac{\|g_k\|^4}{\|d_k\|^2}$$

$$= \sum_{k\ge1}^{\infty}\left(\frac{-g_k^T d_k}{\|g_k\|^2}\right)^2\frac{1}{t_k}, \tag{24}$$

We get from (16) that $0 < \frac{1-2\sigma}{1-\sigma} \le \frac{-g_k^T d_k}{\|g_k\|^2}$, then

$$\left(\frac{-g_k^T d_k}{\|g_k\|^2}\right)^2 \ge \frac{(1-2\sigma)^2}{(1-\sigma)^2}. \tag{25}$$

Substituting (23) and (25) to (24), we get

$$\sum_{k\ge1}^{\infty}\frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \sum_{k\ge1}^{\infty}\frac{m}{t_1 + p(k-1)} = +\infty,$$

where $m = \frac{(1-2\sigma)^2}{(1-\sigma)^2}$ is a constant, which contradicts (11). The proof is then complete.

IV. NUMERICAL EXPERIMENTS

In this section, we report the performances of the modified CD conjugate gradient method on a set of test problems and compare them to the PRP method with strong Wolfe line search whose results be given by [7]. The stopping rule is $\| g_k \| \le 10^{-5}$ .The iteration is also terminated if the total iteration number exceeds 2×10^4 and the function evaluation number exceeds 3×10^5 without reaching the stopping rule. In our tests, we choose the parameters $\rho = 0.1$, $\delta = 0.01$, $\mu = 0.5$ and $\lambda = 0.2$.

In the following table, each column has the following meanings:

Problem: the name of the problem;
Dim: the dimension of the problem;
NI: the number of iterations;
NF: the total number of function evaluations;

TABLE I NUMERICAL RESULTS

Problem	Dim	Algorithm 1	PRPSWP
		NI/NF	NI/NF
Rose	2	52/228	29/502
Helix	3	44/235	49/255
Bard	3	14/68	23/98
Gulf	3	1/2	1/2
Kowosb	4	57/201	62/361
Biggs	6	91/532	121/495
Osb2	11	211/769	293/1372
Vardim	50	28/310	10/52
Trig	100	61/404	46/342
Ie	500	3/9	6/13
Lin	1000	1/3	1/3

ACKNOWLEDGMENT

This work is supported by Yunnan Natural Science Foun dation (2014FD053)

REFERENCES

[1] R Fletcher. Practical methods of optimization. Chichester: JohnWiley; 1987.
[2] Y H Dai, Y Yuan, "Convergence properties of the conjugate descent method, "Math. Adv. CMP. Vol. 25, pp. 552–562, 1996.
[3] Y H Dai, Y Yuan. A class of globally convergent conjugate gradient methods. Sci. China Series Math. Ser. A. Vol. 46, pp. 251–261, 2003.
[4] B. Sellami, Y. Laskri & R. Benzine. A new two-parameter family of nonlinear conjugate gradient methods. Optimization, Vol. 64, pp. 993-1009, 2015.
[5] G. Zoutendijk, Nonlinear programming computational methods, in Integer and Nonlinear Programming, J. Abadie (editor), 1970, pp. 37–86.
[6] P. Wolfe, Convergence conditions for ascent method, SIAM Rev.,Vol. 11, pp. 226–235,1969.
[7] Z X Wei, S W Yao, L Y Liu. The convergence properties of some new conjugate gradient methods. Applied Mathematics and Computation. Vol.183,pp.1341-1350,2006.