ACTA MATHEMATICA VIETNAMICA
Volume 35, Number 3, 2010, pp. 371–378

CONVERGENCE OF A MODIFIED CONJUGATE DESCENT METHOD IN THE PRESENCE OF ERRORS

MIN SUN

ABSTRACT. The conjugate gradient method is a useful method for solving large-scale minimization problems. Errors may arise because of inexact computation. In this paper we study the global convergence of a modified conjugate descent method with strong Wolfe line search in the presence of errors. Preliminary numerical experiments are given to show the efficiency and robustness of the method.

1. INTRODUCTION

In this paper we study the unconstrained minimization problem which is to find the minimal point x^* of $f(x)$ over R^n, denoted by

$$\min f(x),\ x \in R^n, \tag{1.1}$$

where $f : R^n \to R$ is smooth and its gradient $g(x) = \nabla f(x)$ is available.

Conjugate gradient method is a very useful and powerful method for solving large-scale minimization problems (1.1) due to its smaller storage requirements and simple computation ([1, 2, 6]). Its iteration formula is given by

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.2}$$

where $\alpha_k \geq 0$ is a stepsize obtained by means of one-dimensional line search and d_k is the search direction defined by

$$d_k = \begin{cases} -g_k,\ k = 1, \\ -g_k + \beta_k d_{k-1},\ k \geq 2, \end{cases} \tag{1.3}$$

where $g_k = g(x_k)$ and β_k is a scalar. Varieties of this method differ in the way of selecting β_k. The best-known formulae for β_k are called the Fletcher-Reeves (FR), Polak-Ribiere-Polyak (PRP), Dai-Yuan (DY), and Conjugate Descent (CD)

Received February 17, 2009.
2000 *Mathematics Subject Classification.* 90C30, 90C33.
Key words and phrases. Conjugate descent method, Wolfe line search, Global convergence, Errors.
This work was supported by the Foundation of Shangdong Provincial Eduacation Department No. J10LA59.

formulae, and are given by

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

$$\beta_k^{\mathrm{DY}} = \frac{\|g_k\|^2}{d_{k-1}^\top (g_k - g_{k-1})} = \frac{g_k^\top d_k}{g_{k-1}^\top d_{k-1}}, \ \beta_k^{\mathrm{CD}} = -\frac{\|g_k\|^2}{g_{k-1}^\top d_{k-1}}.$$

The global convergence of conjugate gradient methods can refer to a book [3] and a recent review paper [4]. In [3], Dai pointed out that the global convergence of conjugate gradient method can not be guaranteed if $\beta_k = \beta_k^{\mathrm{CD}}$. Recently, Tang and Shi [8] proposed a memory gradient method which is similar to CD conjugate gradient method and only differs in the selecting of β_k which is defined by

$$\beta_k = \frac{\rho\|g_k\|^2}{\sigma g_{k-1}^\top d_{k-1}}, \tag{1.4}$$

where $\sigma \in (0, 1/2)$ and $\rho \in (1, \sigma/(1+\sigma))$. The parameters σ and ρ played an important role in the global convergence.

In practical computation, errors may arise because of inexact computation of the gradient of $f(x)$ or approximate computation of subproblems. Li and Wang [5] studied DY conjugate gradient method with line search in the presence of errors. The main direction s_k satisfies

$$g_k^\top s_k \le -c_1\|g_k\|^2, \ c_1 > 0, \tag{1.5}$$

and the error term w_k satisfies

$$\|w_k\| \le \gamma_k(q + p\|g_k\|), \ p, q > 0,$$

where

$$\gamma_k = O(1/k), \ \gamma_k > 0.$$

Under the above conditions and $g(x)$ is Lipschitz continuous, they proved either

$$\lim_{k\to\infty} f(x_k) = -\infty,$$

or

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

However, to the best of our knowledge, there is no published analysis for the global convergence of conjugate descent method with errors.

In this paper, motivated by [8] and [5], we study a modified conjugate descent method with errors. β_k in the modified descent method has the following new form

$$\beta_k = \begin{cases} -\dfrac{\rho\|g_k\|^2}{\sigma g_{k-1}^\top d_{k-1}}, \ g_{k-1}^\top d_{k-1} < 0, \\ \\ \dfrac{g_k^\top (g_k - g_{k-1})}{\|g_{k-1}\|^2}, \ \text{otherwise}, \end{cases} \tag{1.6}$$

where σ and ρ are two positive parameters defined below. An important feature in our method is that we can prove the condition (1.5) is always right, and d_k is always a descent direction.

The paper is organized as follows: In Section 2 we propose the modified conjugate descent method with errors and prove the global convergence property of the method under mild conditions. In Section 3 we give some preliminary numerical experiments. Some concluding remarks are given in Section 4.

2. Modified conjugate descent method with errors

For the modified conjugate descent method with errors, we let

$$x_{k+1} = x_k + \alpha_k(s_k - w_k), \tag{2.1}$$

where α_k is stepsize, w_k is accumulative error and main direction s_k is determined by

$$s_k = \begin{cases} -g_k, \ k = 1, \\ -g_k + \beta_k d_{k-1}, \ k \geq 2, \end{cases} \tag{2.2}$$

and

$$d_{k-1} = \begin{cases} s_{k-1} - w_{k-1}, \ g_{k-1}^\top (s_{k-1} - w_{k-1}) \leq 0, \\ -s_{k-1} + w_{k-1}, \ g_{k-1}^\top (s_{k-1} - w_{k-1}) > 0 \end{cases} \tag{2.3}$$

where β_k is determined by (1.6).

s_k, w_k satisfy the following assumptions:

(A) $\|w_k\| \leq \gamma_k(q + p\|g_k\|), \ p, q > 0$.

(B) $0 \leq \gamma_k \leq c_1/k, \ c_1 > 0$.

The following inexact line search is used in this paper.

Search A (Strong Wolfe Search) Choose a $\lambda_k > 0$ such that

$$\begin{cases} f(x_k + \lambda_k d_k) \leq f(x_k) + \rho \lambda_k g_k^\top d_k, \\ |g(x_k + \lambda_k d_k)^\top d_k| \leq -\sigma g_k^\top d_k. \end{cases} \tag{2.4}$$

Now we describe the algorithm formally as follows:

Algorithm 2.1. Set $\sigma \in (0, 1/2)$ and $\rho \in (0, \sigma/(\sqrt{3} + 2\sigma))$, $x_1 \in R^n, k := 1$;

Step 1 Compute g_k. If $\|g_k\| = 0$, then stop, and x_k is a stationary point, else goto Step 2;

Step 2 Let d_k is determined by (**??**).

If $g_k^\top (s_k - w_k) = 0$, let $\alpha_k = 0$ and goto Step 3;

If $g_k^\top (s_k - w_k) \neq 0$, let $\alpha_k = \lambda_k$, where λ_k is also determined by Search A, and goto Step 3;

Step 3 Let $x_{k+1} = x_k + \alpha_k d_k$, $k := k + 1$, and return to Step 1.

In order to establish global convergence, we need the following assumption.

(H1). The level set $L(x_1) = \{x \in R^n | f(x) \leq f(x_1)\}$ with x_1 given is bounded.

(H2). The gradient $g(x)$ of $f(x)$ is Lipschitz continuous on an open convex set B that contains the level set $L(x_1)$, i.e., there exists an $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \ \forall x, y \in B.$$

Under (H1) and (H2), there exists positive c such that

$$\|g(x)\| \leq c, \ \forall x \in L(x_1). \tag{2.5}$$

Remark 2.1. From (B), we have

$$\lim_{k\to\infty} \gamma_k = 0. \tag{2.6}$$

Remark 2.2. In Algorithm 2.1, the search direction d_k is always a descent direction of $f(x)$ at x_k.

Lemma 2.1. [3] *If $\varphi(\alpha) = f(x_k+\alpha d_k)$ is bounded below when $\alpha > 0$, there exists $\alpha > 0$ which satisfies the strong Wolfe line search.*

Lemma 2.2. *For all $k \geq 1$, we have $g_k^\top s_k \leq (\rho - 1)\|g_k\|^2$.*

Proof. In the case $k = 1$, the conclusion is obviously right.

In the case $k \geq 2$. If $g_{k-1}^\top d_{k-1} = 0$, then $x_k = x_{k-1}$, thus

$$g_k^\top s_k = g_k^\top(-g_k + \beta_k d_{k-1}) = g_{k-1}^\top(-g_{k-1} + \beta_k d_{k-1}) = -\|g_{k-1}\|^2 = -\|g_k\|^2.$$

If $g_{k-1}^\top d_{k-1} < 0$, we have

$$g_k^\top s_k \leq -\|g_k\|^2 - \sigma\beta_k g_{k-1}^\top d_{k-1} \leq -\|g_k\|^2 + \frac{\rho\|g_k\|^2}{\sigma g_{k-1}^\top d_{k-1}}\sigma g_{k-1}^\top d_{k-1} = (\rho - 1)\|g_k\|^2.$$

Thus the conclusion holds. □

Remark 2.3. If d_k is orthogonal to g_k, let $w_{k+1} = w_k$, $\alpha_k = 0$ and it follows (2.1) that $x_{k+1} = x_k$. From (**??**) and Lemma 2.2, we have

$$\begin{aligned} g_{k+1}^\top d_{k+1} &= g_k^\top(-g_k + \beta_{k+1} d_k + w_{k+1}) \\ &= -\|g_k\|^\top + \beta_{k+1} g_k^\top d_k + g_k^\top w_k \\ &= -\|g_k\|^2 + g_k^\top(s_k - d_k) \\ &\leq -(2 - \rho)\|g_k\|^2. \end{aligned}$$

Thus d_{k+1} is a descent direction except $\|g_k\| = 0$.

Lemma 2.3. *If $\{x_k\}$ is an infinite sequence generated by Algorithm 2.1, then when $I = \{k|g_k^\top(s_k - w_k) \geq 0\}$ is an infinite set, we have*

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

Proof. It follows from the definition of I that

$$g_k^\top s_k \geq g_k^\top w_k.$$

By Lemma 2.2, we have

$$(1 - \rho)\|g_k\|^2 \leq -g_k^\top w_k \leq \|g_k\|\|w_k\| \leq \gamma_k(q + p||g_k||)\|g_k\|,$$

that is

$$(1-\rho-\gamma_k p)\|g_k\| \le \gamma_k q.$$

By Remark 2.1, we get the desired conclusion. □

In following, we assume that I is a finite set. Without loss of generality, we assume that

$$d_k = s_k - w_k,\ g_k^\top d_k < 0,\ \forall k.$$

Lemma 2.4. [3] *If* (H1), (H2) *hold and* $\{x_k\}$ *is an infinite sequence generated by Algorithm 2.1,* d_k *satisfies* $g_k^\top d_k < 0$*, then we have*

$$\sum_{k=1}^{\infty} \frac{(g_k^\top d_k)^2}{\|d_k\|^2} < +\infty. \tag{2.7}$$

To simplify the narration, we give the following notation

$$t_k = \frac{|g_k^\top d_k|}{\|d_k\|}.$$

Lemma 2.5. *If* (H1), (H2) *hold and* $\{x_k\}$ *is an infinite sequence generated by Algorithm 2.1, and there exists a constant* $\epsilon_0 > 0$ *such that*

$$\|g_k\| \ge \epsilon_0,\ \forall k \ge 1, \tag{2.8}$$

then

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

Proof. From $g_k^\top d_k = -\|g_k\|^2 + \beta_k g_k^\top d_{k-1} - g_k^\top w_k$, we have

$$\begin{aligned}
|g_k^\top d_k| &\ge \|g_k\|^2 - \beta_k |g_k^\top d_{k-1}| - |g_k^\top w_k| \\
&\ge \|g_k\|^2 - \beta_k \sigma |g_{k-1}^\top d_{k-1}| - \|g_k\|\|w_k\| \\
&= \|g_k\|^2 - \beta_k \sigma t_{k-1}\|d_{k-1}\| - \|g_k\|\|w_k\| \\
&\ge \|g_k\|^2 - \sigma t_{k-1}(\|s_k\| + \|g_k\|) - \|g_k\|\|w_k\| \\
&\ge \|g_k\|^2 - \sigma t_{k-1}(\|d_k\| + \|w_k\| + \|g_k\|) - \|g_k\|\|w_k\| \\
&\ge \|g_k\|^2 - \sigma t_{k-1}(\|d_k\| + \frac{\|w_k\|\|g_k\|^2}{\epsilon_0^2} + \frac{\|g_k\|^2}{\epsilon_0}) - \frac{\|g_k\|^2\|w_k\|}{\epsilon_0} \\
&= (1 - \frac{\sigma t_{k-1} + \|w_k\|}{\epsilon_0} - \frac{\sigma t_{k-1}\|w_k\|}{\epsilon_0^2})\|g_k\|^2 - \sigma\|d_k\| t_{k-1}.
\end{aligned}$$

Dividing both sides of the above inequality by $\|d_k\|$, we have

$$t_k + \sigma t_{k-1} \ge (1 - \frac{\sigma t_{k-1} + \|w_k\|}{\epsilon_0} - \frac{\sigma t_{k-1}\|w_k\|}{\epsilon_0^2})\frac{\|g_k\|^2}{\|d_k\|}.$$

From Remark 2.1 and Lemma 2.4, we have

$$\lim_{k\to\infty} t_k = 0,\ \lim_{k\to\infty} \|w_k\| = 0.$$

Thus, for some k sufficiently large

$$t_k + \sigma t_{k-1} \geq \frac{\|g_k\|^2}{2\|d_k\|}. \tag{2.9}$$

From

$$[t_k + \sigma t_{k-1}]^2 \leq 2t_k^2 + 2\sigma^2 t_{k-1}^2,$$

and

$$\sum_{k=1}^{\infty} t_k^2 < +\infty,$$

we have

$$\sum_{k=1}^{\infty} [t_k + \sigma t_{k-1}]^2 < +\infty. \tag{2.10}$$

From (2.9), (2.10), we have

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

The proof is complete. □

Theorem 2.1. *If* (H1), (H2) *hold and* $\{x_k\}$ *is an infinite sequence generated by Algorithm 2.1, then*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{2.11}$$

Proof. We assume that (2.11) is not true, then there exists a constant $\epsilon_0 > 0$ such that (2.8) is true. From Lemma 2.2 and (A), we have

$$\begin{aligned} g_k^\top d_k &\leq (\rho - 1)\|g_k\|^2 + \gamma_k \|g_k\|(q + p\|g_k\|) \\ &\leq (p\gamma_k + \rho - 1)\|g_k\|^2 + \gamma_k \frac{q\|g_k\|^2}{\epsilon_0} \\ &= -[1 - \rho - \gamma_k(p + q/\epsilon_0)]\|g_k\|^2 \end{aligned}$$

thus, from (B), for sufficiently large k, we have

$$g_k^\top d_k \leq -(1 - 2\rho)\|g_k\|^2. \tag{2.12}$$

From (1.2), (1.3), (2.2), we obtain

$$\begin{aligned} \|d_k\|^2 &\leq 3\|g_k\|^2 + 3\beta_k^2\|d_{k-1}\|^2 + 3\|w_k\|^2 \\ &\leq 3\|g_k\|^2 + 3\frac{\rho^2\|g_k\|^4}{\sigma^2(g_{k-1}^\top d_{k-1})^2}\|d_{k-1}\|^2 + 3\gamma_k^2(q + p\|g_k\|)^2 \\ &\leq 3\|g_k\|^2 + 3\frac{\rho^2\|d_{k-1}\|^2}{\sigma^2(1 - 2\rho)^2} + 3\gamma_k^2(q + p\|g_k\|)^2. \end{aligned}$$

Dividing both sides of the above inequality by $\|g_k\|^4$ and let

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

we have

$$
\begin{aligned}
s_k &\le \frac{3}{\|g_k\|^2} + \frac{3\rho^2}{\sigma^2(1-2\rho)^2} s_{k-1} + 3\gamma_k^2 (\frac{q}{\|g_k\|^2} + \frac{p}{\|g_k\|})^2 \\
&\le \frac{3}{\epsilon_0^2} + \frac{3\rho^2}{\sigma^2(1-2\rho)^2} s_{k-1} + 3\gamma_k^2 (\frac{q}{\epsilon_0^2} + \frac{p}{\epsilon_0})^2 \\
&\le w_k + v s_{k-1},
\end{aligned}
$$

where $w_k = 3/\epsilon_0^2 + 3\gamma_k^2(q/\epsilon_0^2 + p\epsilon_0)^2$ and $v = 3\rho^2/\sigma^2(1-2\rho)^2$. From (B) and $0 < \rho < \sigma/(\sqrt{3} + 2\sigma)$, there exists $w > 0$,

$$w_k < w, \ 0 < v < 1.$$

Thus

$$s_k \le v s_1 + w(1 + v + v^2 + \cdots + \cdots + v^{k-2}) \le s_1 + (k-1)w.$$

we have

$$\sum_{k=1}^{\infty} \frac{1}{s_k} \ge \sum_{k=1}^{\infty} \frac{1}{s_1 + (k-1)w} = +\infty.$$

This contradicts Lemma 2.5. Therefore (2.11) is true. □

3. Numerical results

In what follows, we report the numerical results of the new modified conjugate descent method for two standard test problems. The new method is denoted by MCD. The error term w_k is obtained randomly in the case it satisfies assumptions (A)(B). For each problem, the unit of CPU time is second and IT denotes the iteration number when the algorithm terminates.

Parameters used in the algorithm are set as $p = 1$, $q = 0.1$, $c_1 = 1$. All codes are written in Matlab 7.1 and run on a portable computer.

Problem 1.[5]

$$
\begin{aligned}
f(x) = 10(x_1^2 - x_2)^2 + (1 - x_1)^2 + 9(x_4 - x_3^2)^2 + (1 - x_3)^2 + 10.1((x_2 - 1)^2 \\
+ (x_4 - 1)^2) + 19.8(x_2 - 1)(x_4 - 1.
\end{aligned}
$$

The initial point is $(-3, -1, -3, -1)^\top$.

Problem 2.[5]

$$f(x) = \sum_{i=1}^{N/2} ((x_{2i} - x_{2i-1}^2)^2 + (1 - x_{2i-1})^2).$$

The initial point is $(-1, 2, 1, \cdots, -1, 2, 1)^\top$.

Problem 3.[7]

$$f(x) = (x_1 + 10x_2)^4 + 5(x_3 - x_4)^4 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4.$$

The initial point is $(2, 2, -2, -2)^\top$.

For each method, we obtain the corresponding results for $\|g(x_k)\| \le 10^{-6}$, and $3.2119(-7)$ means 3.2119×10^{-7}. The numerical results are listed in Table 1.

Table 1. Numerical results of problems 1-3

Problem	ρ	σ	IT	CPU	$f(x_k)$
1	0.05	0.1	306	0.1903	3.9468(-13)
	0.09	0.2	356	0.2203	2.9511(-13)
	0.1	0.3	474	0.3004	3.5701(-13)
2	0.09	0.2	35	0.2053	3.2782(-12)
	0.15	0.4	40	0.3292	1.5623(-12)
3	0.05	0.1	25	0.0200	7.0888(-6)
	0.09	0.2	24	0.0200	9.9293(-6)
	0.1	0.3	26	0.0200	9.9407(-6)
	0.15	0.4	22	0.0200	6.0837(-6)

From Table 1, MCD is stable and effective for the tested problems. Certainly, more test should probably required.

4. Conclusion

In this paper we proposed a conjugate descent method with errors. When the descent direction is slightly perturbed, the global convergence still holds, and this shows the stability of our new method. For further research, we should investigate the upper bound of the error term w_k, and hope to find a larger upper bound.

References

[1] Y. H. Dai and Y. Yan, Convergence properties of the Fletcher-Reeves method, *IMA J. Numer. Anal.* **16** (1994), 155–164.

[2] Y. H. Dai and Y. Yan, A nonlinear conjugate gradient method with a strong global convergence property, *SIAM J. Optim.* **10** (1999), 177–182.

[3] Y. H. Dai and Y. Yan, *Nonlinear Conjugate Gradient Method*, Shanghai Science and Technology Press, 2000.

[4] W. W. Hager and H.Zhang, A survey of nonlinear conjugate gradient methods, *Pacific J. Optim.* **2** (2006), 35–58.

[5] M. X. Li and C. Y. Wang, Dai-Yuan conjugate gradient method with linesearch in the presence of errors, *Chinese J. Engin. Math.* **23** (5) (2006), 891–900.

[6] Z. J. Shi and J. Shen, Convergence of Liu-Story conjugate gradient method, *European J. Oper. Res.* **182** (2007), 552–560.

[7] Z. J. Shi, A new memory gradient method under exact line search, *Asian Pacific J. Oper. Res.* **20** (2) (2003), 275–284.

[8] J. Y. Tang and Z. J. Shi, A class of global convergent memory gradient methods and its linear convergence rate, *Advance in Math.* **36** (1) (2007), 67–75.

Department of Mathematics and Information Science
Zaozhuang University, Shandong 277160, China

E-mail address: sunmin_2008@yahoo.com.cn