A Modified Conjugate Gradient Method with Armijo-Type Linear Search

Qi Liu, Zhijun Luo (Corresponding author) and Guohua Chen, School of Mathematics and Finance, Hunan University of Humanities, Science and Technology, Loudi, 417000, P.R.China

Abstract—In this paper, we have presented a new conjugate gradient algorithm based on the Armijo-type line search for solving unconstrained optimization problems. Under some suitable conditions, we proved the global convergence of the algorithm. The numerical results show that the proposed methods are effective.

Keywords—Unconstrained optimization; Conjugate gradient method; Armijo-type line search; Global convergence

I. INTRODUCTION

Various numerical methods have been introduced to solve unconstrained optimization problems as they appear in engineering, mathematics and computer science. The conjugate gradient method is an efficient algorithm for the numerical solution of unconstrained optimization problems. In this paper, we consider the following unconstrained optimization problems

$$\min f(x)$$

where $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$ is a smooth, nonlinear function whose gradient will be denoted by $g(x)$. The conjugate gradient method for solving problem (1) always generates a sequence $\{x_k\}$ as

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \ldots,$$

(2)

where $\alpha_k > 0$ is obtained by line search and the directions $d_k$ are generated as

$$d_k = \begin{cases} -g_k, & \text{if } k = 0 \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geq 1, \end{cases}$$

(3)

where $g_k = \nabla f(x_k)$, and $\beta_k$ is a scalar, different conjugate gradient algorithms correspond to different choices for the scalar parameter $\beta_k$. There are some well-known formulas for $\beta_k$ which are given below (see [1-6]):

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

$$\beta_k^{sr} = \frac{g_k^T y_{k-1}}{g_{k-1}^T y_{k-1}}, \quad \beta_k^{ns} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \quad \beta_k^{ts} = \frac{\|g_k\|^2}{d_k^T g_k},$$

or by other formulae (where $y_{k-1} = g_k - g_{k-1}$, the symbol $\| \|$ be the Euclidean norm). The stepsize $\alpha_k$ is determined by exact or inexact line search. The Armijo-type line search: Let $s > 0$ be a constant, $\mu \in (0, 1), \rho \in (0, 1)$, choose $\alpha_k$ to be the largest $\alpha$ in the set $\{s, s\rho, s\rho^2, \ldots\}$ such that

$$f_k - f(x_k + \alpha_k d_k) \geq -\alpha_k g_k^T d_k.$$

In this paper, we devote to the global convergence of a modified Conjugate Gradient Method for unconstrained optimization problems. This method uses a new Armijo-type line search which allows one to find a larger accepted step size and possibly reduces the function evaluations at each iteration. It can guarantee the global convergence of optimal solution under some mild conditions.

II. DESCRIPTION OF ALGORITHM

We first assume that

H2.1

i) The objective function $f(x)$ is continuously differentiable and has a lower bound on the level set $L_0 = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$, where $x_0$ is the starting point.

ii) The gradient $g(x)$ of $f(x)$ is Lipschitz continuous in some neighborhood $U$ of $L_0$, namely, there exists a constant $L > 0$ such that

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

Throughout this paper we suppose that the Lipschitz constant $L$ of $g(x)$ is a known priori or easy to estimate in practical computation. There are some estimations $L_k$ for the Lipschitz constant $L$ [6-7]. Motivated by [6], we propose the following New Armijo-type line search:
Given \( \mu \in \left( 0, \frac{1}{2} \right), \beta \in (0,1), \epsilon \in (0,1) \), set 
\[
l_k = \frac{1-c}{L} \frac{g_k^T \left( y_k - g_{k-1} \right) + \epsilon^2 \|d_k\|^2}{\|d_k\|}
\]
and \( \alpha_k \) is the first \( \alpha \) in 
\[
\left\{l_k \mid l_k \beta \rho, l_k \beta^2, \ldots \right\}
\]
such that : 
\[
f_k - f(x_k + \alpha d_k) \geq - \alpha \mu g_k^T d_k, \tag{4}
\]
where 
\[
L_k = \max \left( \frac{\|y_{k-1}\|}{\|x_{k-1}\|} \right).
\]

Now we state our algorithm as follows.

**Algorithm A:**

**Step 0** Initialization:

Given a starting point \( x_0 \in \mathbb{R}^n \), choose parameters 
\( 0 < \epsilon << 1, \mu \in \left( 0, \frac{1}{2} \right), \rho \in (0,1), \epsilon \in (0,1) \), 
\( c \in (0,1), \epsilon \in (0,1) \). Set \( k := 0 \);

**Step 1** If \( \|g_k\| < \epsilon \), STOP, else go to Step 2;

**Step 2** Compute the search direction \( d_k \) by (3), where 
\[
\beta_k = \frac{g_k^T \left( g_k - g_{k-1} \right)}{\|y_{k-2}^T d_{k-2}\| \|d_{k-1}\| + \epsilon \|d_{k-1}\|^2}.
\]

**Step 3** \( \alpha_k \) is defined by the new Armijo-type line search (4).

**Step 4** Let \( x_{k+1} = x_k + \alpha_k d_k \), \( k := k + 1 \), and go to Step 2.

**Lemma 2.1** Assume that H2.1 hold, the infinite sequence \( \{x_k\} \) generated by Algorithm A, then |

1) there exist \( m_0 > 0 \) and \( M_0 > 0 \) such that 
\( m_0 \leq L_k \leq M_0 \).

2) for \( k \geq 1 \), if 
\[
\alpha_k \leq \frac{1-c}{L} \frac{g_k^T \left( y_k - g_{k-1} \right) + \epsilon^2 \|d_k\|^2}{\|d_k\|},
\]
then, 
\[
g_k^T d_{k+1} \leq -c \|g_k\| \|d_k\|^2.
\]

3) \( \|d_k\| \leq \left( 1 + \frac{\left( 1-c \right) L_k}{m_0} \right) \|g_k\|, \forall k. \)

**III. GLOBAL CONVERGENCE OF ALGORITHM**

In this section, we analyze the global convergence of the Algorithm.

**Lemma 3.1.** Assume that H2.1 hold, then the new Armijo-type linear search is well-defined.

**Proof.** 1) If \( \alpha_k = \frac{1-c}{L} \frac{g_k^T \left( y_k - g_{k-1} \right) + \epsilon^2 \|d_k\|^2}{\|d_k\|} \), then we have 
\[
\alpha_k \leq \frac{1-c}{L} \frac{g_k^T \left( y_k - g_{k-1} \right) + \epsilon^2 \|d_k\|^2}{\|d_k\|} \geq \frac{1-c}{L} \epsilon > 0.
\]

2) If \( \alpha_k < \frac{1-c}{L} \frac{g_k^T \left( y_k - g_{k-1} \right) + \epsilon^2 \|d_k\|^2}{\|d_k\|} \), then, set 
\( \alpha = \rho^{-1} \alpha_k \), the following inequality does not hold, 
\[
f_k - f(x_k + \alpha d_k) \geq - \alpha \mu g_k^T d_k,
\]
i.e. 
\[
f_k - f(x_k + \alpha d_k) < - \alpha \mu g_k^T d_k.
\]

Considering mean value theorem on the left-hand side of the above inequality, there exists \( t_k \in (0,1) \), such that 
\[
- \alpha g_k^T (x_k + t_k d_k) d_k < - \alpha \mu g_k^T d_k,
\]
i.e. 
\[
g_k^T (x_k + t_k d_k) d_k > \mu g_k^T d_k.
\]

By H2.1, and Lemma 2.1, we have 
\[
L_k \alpha \|d_k\|^2 \geq g_k^T (x_k + t_k d_k) - g_k \|d_k\|^2
\]
\[
\geq g_k^T (x_k + t_k d_k) d_k \geq -(1-\mu) g_k^T d_k
\]
\[
\geq c (1-\mu) \|g_k\|^2.
\]
From Lemma 2.1, we can get that 
\[
\alpha_k \geq \frac{c \left( 1-\mu \right) \|g_k\|^2}{L_k \alpha} \geq \left( 1 + \frac{\left( 1-c \right) L_k}{m_0} \right)^{-2} > 0.
\]

**Theorem 3.1** Under the assumption H2.1, the sequence \( \{x_k\} \) generated by Algorithm A is global convergent. That is 
\( \lim_{k \to \infty} \|g_k\| = 0 \).

**Proof.** Set \( \eta_0 = \inf_{\forall k} \{ \alpha_k \} \), if \( \eta_0 > 0 \), then we have 
\[
f_k - f(x_k + \alpha d_k) \geq - \alpha \mu g_k^T d_k \geq \mu \eta_0 c \|g_k\|^2.
\]

By H2.1 we have 
\( \sum_{k=0}^{\infty} \|g_k\|^2 < +\infty \), thus we obtain 
\( \lim_{k \to \infty} \|g_k\| = 0 \).

If the conclusion doesn’t hold, assume that \( \eta_0 = 0 \).
Then, there exists an infinite subset \( K \subseteq \{0,1,2,\ldots\} \) such that 
\( \lim_{k \to \infty} \alpha_k = 0 \), 
(5)

By Lemma 2.1 and Lemma 3.1 we obtain 
\[
l_k = \frac{1-c}{L} \frac{g_k^T \left( y_k - g_{k-1} \right) + \epsilon^2 \|d_k\|^2}{\|d_k\|} > 0.
\]
Therefore, there is a $k'$ such that $\rho^{-1} \alpha_k \leq l_k$, $\forall k \geq k', k \in K$. Let $\alpha = \rho^{-1} \alpha_k$, by Lemma 3.1 we obtain 

$$\alpha_i \geq \frac{c_1 (1-\mu) \|g_i\|}{L} \geq \frac{c_1 (1-\mu) \left(1 + \frac{(1-c)L}{m_0}\right)^{-2}}{L} > 0$$

$k \geq k', k \in K$.

Which contradicts (5). This shows that $\eta_0 > 0$. The whole proof is completed.

IV. NUMERICAL EXPERIMENTS

In the following numerical experiment, the code of the proposed algorithm is written by using MATLAB 7.0. We stop the iteration if the inequality $|g_i| \leq 10^{-6}$ is satisfied. The problems that we tested are from [8] and [9]. Table 1 show the computation results, where the columns have the following meanings:

<table>
<thead>
<tr>
<th>Problem</th>
<th>Dim</th>
<th>Algorithm A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penalty1</td>
<td>4</td>
<td>5/128/89</td>
</tr>
<tr>
<td>Trigonometric</td>
<td>100</td>
<td>18/72/67</td>
</tr>
<tr>
<td>Brown</td>
<td>2</td>
<td>13/148/51</td>
</tr>
<tr>
<td>Axis hyper</td>
<td>500</td>
<td>183/6168/705</td>
</tr>
<tr>
<td>DeJong</td>
<td>2</td>
<td>3/18/13</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>4/24/15</td>
</tr>
</tbody>
</table>

ACKNOWLEDGMENT

This work was supported in part by Scientific Research Fund of Hunan Provincial Education Department (NO. 12A077 and 14C0609), and the Educational Reform Research Fund of Hunan University of Humanities, Science, and Technology(no. RKJGY1526).

REFERENCES


