A New Family of Conjugate Gradient Method

with Armijo Line search

Min Sun
Department of Mathematics and Information Science
Zaozhuang University, Shandong 277160; China
sunmin_2008@yahoo.com.cn

Jing Liu
School of Mathematics and Statistics
Zhejiang University of Finance and Economics, Hangzhou 310018; China

Abstract

A new class of conjugate method is analyzed in this paper. The Polak-Ribiere-Polyak and Liu-Storey conjugate gradient methods are special cases of the new class of conjugate gradient method. It adopts a new Armijo line search technique, and can get a more larger stepsize than classical Armijo line search. At each iteration, the first stepsize is self-adaptive. The search direction generated by the method at each iteration satisfies the sufficient descent condition. Under the Lipschitz continuity of the underlying function, we prove the global convergence of the method. Preliminary numerical results show that our method is efficient.

Mathematics Subject Classification: 49D27, 90C30

Keywords: Unconstrained optimization, conjugate descent method, global convergence

1 Introduction

In this study we consider the unconstrained minimization problem

\[ \min f(x) \quad x \in \mathbb{R}^n \]  \hspace{1cm} (1)
where \( f : R^n \rightarrow R \) is a smooth function, and \( R^n \) denotes an \( n \)-dimensional Euclidean space.

Generally, a line search method takes the form
\[
x_{k+1} = x_k + \alpha_k d_k, \quad k = 1, 2, 3, \ldots
\]
where \( x_k \in R^n \) is the current iterative, \( d_k \) is a descent direction of \( f(x) \) at \( x_k \), and \( \alpha_k \) is a step size. For convenience, we denote \( \nabla f(x_k) \) by \( g_k \), \( f(x_k) \) by \( f_k \), \( \nabla^2 f(x_k) \) by \( G_k \). If \( G_k \) is available and inverse, then \( d_k = -G_k^{-1} g_k \) leads to Newton method and \( d_k = -g_k \) results in the steepest descent method.

Conjugate gradient method is a very efficient line search method for solving large unconstrained problems, due to its lower storage and simple computation. Conjugate gradient method has the form (2) in which
\[
d_k = \begin{cases} -g_k, & k = 0 \\ -g_k + \beta_k d_{k-1}, & k \geq 1 \end{cases}
\]
Various conjugate gradient methods have been proposed, and they are mainly differ in the choice of the parameter \( \beta_k \). Some well-known formulas for \( \beta_k \) are given below:

\[
\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}, \quad \beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1})},
\]
\[
\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T (g_k - g_{k-1})}, \quad \beta_k^{CD} = \frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}}, \quad \beta_k^{LS} = \frac{g_k^T (g_k - g_{k-1})}{-g_{k-1}^T d_{k-1}}.
\]

where \( \| \cdot \| \) denotes the \( l_2 \)-norm. The corresponding conjugate gradient methods can be abbreviated as FR, PRP, HS, DY, CD and LS methods. Although these methods are identical while \( f \) is a strong convex quadratic function and line search is exact, they have different performances when applied to minimizing general nonlinear functions with inexact line searches.

Quite recently, Shi et.al.[1] proposed a new line search(SG method), in which \( \beta_k \) is defined by

\[
\beta_k^{SG} = \frac{g_k^T (g_k - g_{k-1})}{(1-u)\|g_{k-1}\|^2 - ud_{k-1}^T g_{k-1}}
\]

where \( u \in [0,1] \) is a constant. Obviously, \( \beta_k^{SG} = \beta_k^{PRP} \) for \( u = 0 \), and \( \beta_k^{SG} = \beta_k^{LS} \) for \( u = 1 \). Zhou[2] proved that LS method is global convergent under two Armijo line searches, and at the same time LS method is sufficient descent. The line search in [1] is very complex which has two inequalities, and its second inequality is used to generate sufficient descent direction. The line search in [2] is very simple compared to that in [1] although the method in [2] is a special case of the
method in [1]. In this paper, we will study whether or not the method proposed in [1] is global convergent under the line search in [2].

We first assume that

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

**H2:** $g(x)$ is Lipschitz continuous on an open convex set $B$ that contains the level set $L_0$, and the Lipschitz constant is denoted by $L$.

### 2 Modified SG Method

Now we give the modified SG method.

**Step 1** Given parameters $\mu \in (0,1), \rho \in (0,1), c \in (0,1/2), u \in [0,1]$, and chose an initial point $x_0 \in \mathbb{R}^n$. Set $k := 0$;

**Step 2** Computing $g_k$. If $\|g_k\| = 0$, stop; else, go to Step 3;

**Step 3** Set $s_k = \frac{1-c}{L} (1-u) \|g_k\|^2 - u g_k^T d_k$, and chose $\alpha_k$ is the largest $\alpha$ in $\{s_k, s_k, s_k, s_k^2, \ldots\}$ such that

$$f(x_k + \alpha d_k) \leq f(x_k) + \mu \alpha \|g_k\|^2 + \frac{1}{2} \alpha L \|d_k\|^2.$$

(3)

**Step 4** Set $x_{k+1} = x_k + \alpha_k d_k$ where $d_k$ is defined by

$$d_k = \begin{cases} -g_k, & k = 0 \\ -g_k + \beta_{k-1} g_k, & k \geq 1 \end{cases}$$

(4)

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

**Remark 1.** Line search (3) can be viewed as a combination of the two line searches in [2].

In the following, we assume an infinite sequence $\{x_k\}$ is generated, otherwise, the modified SG method stops at a stationary point of problem (1).

**Lemma 1.** Assume that (H1) and (H2) hold and the modified SG method generates an infinite sequence $\{x_k\}$, then we have

$$(c-2) \|g_k\|^2 \leq g_k^T d_k \leq -\|g_k\|^2.$$

(5)

**Proof.** For $k = 0$, $g_0^T d_0 = -\|g_0\|^2$, (5) holds.

For $k \geq 1$, from (4), we have
\[ g_k^T d_k = -\|s_k\|^2 + \beta_k^{SG} s_k^T d_{k-1}, \]
then
\[ |g_k^T d_k + \|g_k\|^2| \leq \frac{\|g_k\|^2 \|g_k - g_{k-1}\|}{(1 - u)\|g_{k-1}\|^2 - ud_{k-1}^T g_{k-1}} \|d_{k-1}\| \]
\[ \leq \frac{L\alpha_{k-1} \|g_k\|^2}{(1 - u)\|g_{k-1}\|^2 - ud_{k-1}^T g_{k-1}} \|d_{k-1}\|^2 \]
\[ \leq \frac{Ls_{k-1} \|g_k\|^2}{(1 - u)\|g_{k-1}\|^2 - ud_{k-1}^T g_{k-1}} \|d_{k-1}\|^2 = (1 - c)\|g_k\|^2. \]
thus (5) holds. The proof is finished.

**Remark 2.** From (5) and \( c \in (0, 1/2) \) we have, \( d_k \) is a sufficient descent direction and the line search (3) is finitely stopped.

**Lemma 2.** The search direction \( d_k \) satisfies
\[ \|d_k\| \leq (2 - c)\|g_k\|. \]

**Proof.** For \( k = 0 \), \( \|d_0\| = \|g_0\| \leq (2 - c)\|g_0\| \), (6) obviously holds.

For \( k \geq 1 \), from (4) we have
\[ \|d_k\| = \|g_k + \beta_k^{SG} d_{k-1}\| \leq \|g_k\| + \frac{|g_k^T (g_k - g_{k-1})|}{(1 - u)\|g_{k-1}\|^2 - ud_{k-1}^T g_{k-1}} \|d_{k-1}\| \]
\[ \leq \left(1 + \frac{L\alpha_{k-1}}{(1 - u)\|g_{k-1}\|^2 - ud_{k-1}^T g_{k-1}}\right) \|g_k\| = (2 - c)\|g_k\|. \]
thus (6) holds. The proof is finished.

Set \( \eta_i = 1 - \frac{1 - c}{2} \left(\frac{1 - u}{c} + u\right) \). In the following, we assume that \( \eta_i > 0 \).

**Remark 3.** If \( c = 3/8 \), for any \( u \in [0, 1] \), we have
\[ \eta_i = 1 - \frac{1 - c}{2} \left(\frac{1 - u}{c} + u\right) = 1 - \frac{1}{2} \left(\frac{5}{8} - \frac{8}{3} u + u\right) = \frac{1}{6} + \frac{25}{48} u > 0. \]
Thus the above assumption is available.

### 3 Global Convergent

**Theorem 1.** Assume that (H1) and (H2) hold and the modified SG method
A new family of conjugate gradient method generates an infinite sequence \{x_k\}, then we have
\[\lim_{k \to \infty} \|g_k\| = 0.\]

**Proof.** Firstly we prove that there exists a constant \(\eta_0 > 0\), such that
\[\inf_{k \geq 0} \alpha_k > \eta_0.\] (7)

Suppose there exists \(K \subseteq \{0,1,2,...\}\), such that \(\alpha_k \to 0\) \((k \in K, k \to \infty)\). From (5)(6), we have
\[s_k = \frac{1-c}{L} \frac{(1-u)g_k^T d_k - u \gamma k d_k}{\|d_k\|^2} \geq \frac{1-c}{L} \frac{(1-u+uc)g_k^T d_k}{\|d_k\|^2} \geq \frac{(1-c)(1-u+uc)}{L(2-c)^2} > 0,\]
thus there exists \(k',\) such that
\[\alpha_k / \rho \leq s_k, \forall k \geq k', k \in K.\]

From the line search rule of (3), we have \(\alpha = \alpha_k / \rho\) doesn't satisfies (3), that is
\[f(x_k + \alpha d_k) > f(x_k) + \mu \alpha [g_k^T d_k + \frac{1}{2} \alpha L \|d_k\|^2].\]
from the mean value theorem, we have
\[\alpha g(x_k + \alpha d_k)^T d_k = f(x_k + \alpha d_k) - f(x_k) > \mu \alpha [g_k^T d_k + \frac{1}{2} \alpha L \|d_k\|^2] > \mu g_k^T d_k\]
that is
\[g(x_k + \alpha \theta_k d_k)^T d_k - \mu g_k^T d_k > 0.\]
From (5) again, we get
\[L \alpha \|d_k\|^2 \geq \|g(x_k + \alpha \theta_k d_k) - g_k\|^2 d_k > - (1-\mu) g_k^T d_k \geq c(1-\mu) \|g_k\|^2.\]
thus, from (6), we obtain
\[\alpha_k = \alpha \rho > \frac{c(1-\mu) \|g_k\|^2}{L \|d_k\|^2} \geq \frac{c(1-\mu)}{L(2-c)^2}.\]
This contradicts with the assumption \(\alpha_k \to 0\) \((k \in K, k \to \infty)\), and thus (7) holds.

From (3)(5)(8), we have
\[f(x_k) - f(x_{k+1}) \geq - \mu \alpha_k [g_k^T d_k + \frac{1}{2} \alpha_k L \|d_k\|^2] \geq - \mu \alpha_k [g_k^T d_k + \frac{1}{2} s_k L \|d_k\|^2] \]
\[= - \mu s_k [g_k^T d_k + \frac{1}{2} (1-c)(1-u) \|g_k\|^2 - u g_k^T d_k)]\]
\[\geq - \mu \alpha_k [1 - \frac{1}{2} (1-c)(1-u/c + u)] g_k^T d_k \geq c \mu \eta \gamma \|g_k\|^2.\]

From (H1), we have the desired result. The whole proof is completed.
4 Numerical Reports

In the following numerical experiments, the parameters are set as
\[ \mu = 0.2, \rho = 0.87, c = 0.1. \]
The program was coded in Matlab7.1, and tested on PIV2.8GHz personal computer. \( L \) was estimated by the method in [4].

\textbf{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 \( x_0 = (1,1,1,1) \).

\textbf{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 \( n = 100 \).

Table 1 Numerical of iteration and number of functional and gradient evaluations

<table>
<thead>
<tr>
<th>Problem</th>
<th>( u )</th>
<th>Ite.</th>
<th>Num. of Fun.</th>
<th>Num. of Gra.</th>
<th>CPU</th>
<th>( | g(x_k) | )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1</td>
<td>52</td>
<td>105</td>
<td>77</td>
<td>0.01</td>
<td>7.7118e-007</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>53</td>
<td>107</td>
<td>78</td>
<td>0.01</td>
<td>8.1831e-007</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>55</td>
<td>111</td>
<td>79</td>
<td>0.01</td>
<td>7.7724e-007</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>57</td>
<td>115</td>
<td>79</td>
<td>0.01</td>
<td>9.1926e-007</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>60</td>
<td>121</td>
<td>82</td>
<td>0.01</td>
<td>9.9777e-007</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>64</td>
<td>129</td>
<td>86</td>
<td>0.01</td>
<td>9.4837e-007</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>31</td>
<td>95</td>
<td>62</td>
<td>0.01</td>
<td>8.5242e-007</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>30</td>
<td>85</td>
<td>60</td>
<td>0.01</td>
<td>9.7246e-007</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>30</td>
<td>73</td>
<td>60</td>
<td>0.01</td>
<td>9.4737e-007</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>31</td>
<td>67</td>
<td>62</td>
<td>0.01</td>
<td>7.5628e-007</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>30</td>
<td>60</td>
<td>44</td>
<td>0.01</td>
<td>9.4653e-007</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>34</td>
<td>68</td>
<td>51</td>
<td>0.01</td>
<td>9.0686e-007</td>
</tr>
</tbody>
</table>

From the preliminary numerical results, we have for Problem 1, the new method is quite efficient when \( u \) is little, and for Problem 2 the results are not sensitive to the parameter \( u \), which shows that the new method is robust.
A new family of conjugate gradient method

References


Received: November, 2009