A Parameteric Conjugate Gradient Methods for Large Scale Problem

Mingjuan Ma¹, a, Yongpo Zhang, Jiahui Sun, Huiyu Mao

¹Full Teaching and Research section of Mathematics, Aviation University of air force, changchun, 130022

¹157793988@qq.com

Keywords: optimization, spectral-gradient, conjugate gradient, convergence.

Abstract. In this paper, we give out a Parameter. By this Parameter, we get a series of conjugate gradient methods. We give the Algorithm of the Parameteric conjugate gradient. The efficiency of the method can be got by the numerical result of the examples.

Introduction

Optimization problem is widely used in management and engineering. Many numerical problem can be transformed into mathematical model. Using the mathematical method, we can solve the question. So, in our paper, we consider this unconstraint optimization problem:

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

where $$f: \mathbb{R}^n \rightarrow \mathbb{R}$$ is a smooth function.

The general idea for this series problem has the following formula:

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

Where $$\alpha_k$$ is the step-size along the search direction. For the search direction $$d_k$$, we have the form

$$d_k = -g_k + \beta_k d_{k-1}$$

Where $$\beta_k$$ is a scalar, $$g_k = \nabla f(x_k)$$, the gradient of the function $$f(x)$$ at $$x_k$$, while $$k = 1, d_1 = -g_1$$.

Nowadays, the key question is how to define the scalar $$\beta_k$$ for the search direction $$d_k$$. In many literatures [1], [2], [3], we have the well-known formula as follows.

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \beta_k^{DY} = \frac{\|g_k\|^2}{d_k^T y_{k-1}}, \quad \beta_k^{PR} = \frac{\langle g_k, y_{k-1} \rangle}{\|g_{k-1}\|^2}, \quad \beta_k^{CD} = \frac{\|g_k\|^2}{-d_k^T g_{k-1}}, \quad \beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_k^T y_{k-1}},$$

where $$\|\cdot\|$$ denotes Euclidean norm, $$y_{k-1} = g_k - g_{k-1}$$.

In the literature[4], E.G.Birgin give the modification of $$d_k$$, let

$$d_k = -\theta_k g_k + \beta_k d_{k-1},$$

For the concept of $$\theta_k$$, we have the following form: $$\theta_k = \frac{s_k^T s_{k-1}}{s_{k-1}^T y_{k-1}}.$$

Where $$s_{k-1} = x_k - x_{k-1}$$. The parameter $$\theta_k$$ is the inverse of the Raleigh quotient

$$\frac{s_{k-1}^T \int_0^1 \nabla^2 f(x_{k-1} + ts_{k-1})dt s_{k-1}}{s_{k-1}^T s_{k-1}},$$

which lies in the largest and the smallest eigenvalue of the Hessian average

$$\int_0^1 \nabla^2 f(x_{k-1} + ts_{k-1})dt.$$

With the suggestion of [4][5], we give a new form of $$\beta_k$$, which has the form as follows.

$$\beta_k = \theta_k \beta_k^{FR}, \beta_k = \theta_k \beta_k^{HS}, \beta_k = \theta_k \beta_k^{CD}, \beta_k = \theta_k \beta_k^{DY}, \beta_k = \theta_k \beta_k^{PR}.$$ 

For the direction, we have the following form.

$$d_k = -g_k + \beta_k d_{k-1}$$

Here, we give the example of $$\beta_k = \theta_k \beta_k^{HS}$$, and we also give the proof of the convergence. For Other formula, we can have the similar conclusion.
Main note
We give the Assumption condition.

M1. The level set \( A = \{ x \in \mathbb{R}^n : f(x) \leq f(x_i) \} \) is bounded, namely, there exists a constant \( b \) such that \( \| x \| \leq b \), \( \forall x \in A \).

M2. In some neighborhood \( U \) of \( L \), \( f \) is continuously differentiable, and its gradient \( g \) is Lipschitz continuous; namely, there exists a constant \( L \) such that
\[
\| g(x) - g(y) \| \leq L \| x - y \|, \quad \forall x, y \in U.
\]

Under the above assumptions on M1, M2, there exists a constant \( \chi \geq 0 \) such that
\[
\| \nabla f(x) \| \leq \chi, \quad x \in U
\]

For the choice of step-size \( \alpha \), we use the general Wolfe line search condition. That is to say that the step-size \( \alpha \) satisfy
\[
f(x_k + \alpha d_k) - f(x_k) \leq \tau \alpha g_k^T d_k \tag{5}
\]
\[
g(x_k + \alpha d_k)^T d_k > \delta g_k^T d_k, \quad 0 < \tau \leq \delta < 1 \tag{6}
\]

Our methods and the numerical experiment

With the result of the discussion, we give the mail method as follows.

Algorithm A
Step0: Initialization, choose \( 0 < \tau \leq \delta < 1 \);
step1: \( d_1 = -g_1 \), \( k \leftarrow 1 \) if \( g_1 = 0 \) terminate the algorithm, else goto the next step;
Step2: calculate the step-size \( \alpha \), \( \alpha_{\text{init}} = 1 \) if \( \alpha \) satisfy
\[
f(x_k + \alpha d_k) - f(x_k) \leq \tau \alpha g_k^T d_k
\]
\[
g(x_k + \alpha d_k)^T d_k > \delta g_k^T d_k,
\]

Where \( 0 < \tau \leq \delta < 1 \), define \( \alpha = \alpha \), terminate the line research and goto step3, else \( \alpha_{\text{new}} = \sigma \alpha \), \( \sigma \in (0, 1, 0.9) \), \( \alpha \leftarrow \alpha_{\text{new}} \) return to the start of this step and have the check of the condition.
Step3: \( x_{k+1} = x_k + \alpha d_k \), compute \( g_{k+1} = \nabla f(x_{k+1}) \) if \( g_{k+1} = 0 \), terminate the algorithm and putout \( x_{k+1} \), else goto the next step;
Step4: find \( d_{k+1} \), \( d_{k+1} = -g_{k+1} + \beta d_k \), set \( k = k + 1 \) return to step2;
Next, we give the proof of the convergence.

Lemma 3.1 Suppose that the function satisfies the Assumption condition M1 and M2, Lipschitz constant is \( L(x) \), \( 0 < \tau \leq \delta < 1 \), \( d \) is a decent direction at \( x \), then the Wolfe line search condition
\[
f(x_k + \alpha d_k) - f(x_k) \leq \tau \alpha g_k^T d_k
\]
Is satisfied for all the \( \alpha \in [0, \alpha_{\max}] \), where
\[
\alpha_{\max} = \frac{2(\tau - 1)g(x)^T d}{L(x)\| d \|^2}.
\]

This lemma give the existence of the step-size \( \alpha \), next, we give the proof of the convergence in the condition of uniformly convex function.

Lemma 3.2 Suppose that the function satisfies the Assumption condition M1 and M2, \( x_k \) can be calculate by (2), \( \alpha_k \) can be got by (5)(6), \( d_k \) is the decent direction, that is to say \( d_k \) satisfies
\[
d_k^T g_k < 0 \), then, there exist a constant \( n \), for the arbitrary \( k \), we have
\[
f(x_k) - f(x_{k+1}) \geq n\| \alpha d_k \|^2
\]

Lemma 3.3 when \( k \to \infty \), we have the result \( \| x_k - x_{k+1} \| \to 0 \)
**Lemma 3.4** Suppose that the function satisfies the Assumption condition M1 and M2, \( f(x) \) is the uniformly convex function. \( x_k \) can be iterative by (2), \( \alpha_k \) can be got by (5)(6), \( d_k \) is the decent direction, if \( \sum_{i=1}^{\infty} \frac{1}{\|d_k\|^2} = \infty \). Then, we have\[
\lim_{k \to \infty} \inf \|g_k\| = 0
\]

**Theorem 1** Suppose that the function satisfies the Assumption condition M1 and M2, \( f(x) \) is the uniformly convex function. \( x_k \) can be iterative by (2), \( \alpha_k \) can be got by (5)(6), \( d_k \) is the decent direction, \( \beta_k = \theta_k \beta_k^{IR} \), then we have the conclusion \[
\lim_{k \to \infty} \inf \|g_k\| = 0
\]

Next, we give the numerical examples. We also calculate the examples in [6]. In order to give the advantage of our method, we compare our result with the result in [6],[7] in the form of a table. The five test function is as follows.

1. \( f(x) = (x_1 - 2)^4 + (x_2 - 2)^2 + x_2 + 1 \)
   Initial point \( x_0 = [1, 1] \);

2. Rosenbrock function \( f(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2) + (1 - x_i)^2] \)
   Initial point \( x_0 = [-1.2, 1,-1.2,1,\ldots,-1.2,1] \);

3. Wood function \( f(x) = 100(x_1^2 - x_2^2) + (x_1 - 1)^2 + (x_2 - 1)^2 + 90(x_3^2 - x_4) + 10.1[(x_2 - 1)^2 + (x_4 - 1)^2] + 19.8(x_2 - 1)(x_4 - 1) \)
   Initial point \( x_0 = [-3,-1,-3,-1] \);

4. Powell function \( f(x) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^2 + 10(x_1 - x_4)^4 \)
   Initial point \( x_0 = [3,-1,0,1] \);

5. Cube function \( f(x) = 100(x_2 - x_1^3)^2 + (1 - x_1)^2 \)
   Initial point \( x_0 = [-1.2,-1,-1.2,-1] \).

The choice of the parameters are: \( \tau = 0.6, \delta = 0.8 \) For \( \beta_k \) we choose \( \beta_k = \theta_k \beta_k^{IR} \). And the stop criterion is \( \|g_k\| \leq 10^{-6} \).

In the table 1, we conclude the numerical result compared with the result in [7]. In the table 2, with the other form of \( \beta_k \) that is to say \( \beta_k = \theta_k \beta_k^{IS}, \beta_k = \theta_k \beta_k^{CD}, \beta_k = \theta_k \beta_k^{DY}, \beta_k = \theta_k \beta_k^{RP} \), we give the advantage in the number of iteration. In the table 3, we compare the results in our paper with the result in [6].

By the observation of the three tables, we found that, our method is a good method for the large scale problem. For most of the problem, we found that our methods is better than BFGS and the result in [6] in the number of iteration.

### Table 1

<table>
<thead>
<tr>
<th>Problem</th>
<th>dimension</th>
<th>Amijo</th>
<th>Nonmonotone-method</th>
<th>Our result</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>22</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>39</td>
<td>30</td>
<td>34</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>52</td>
<td>44</td>
<td>48</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>40</td>
<td>31</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>34</td>
<td>34</td>
<td>31</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>28</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>$\beta_k = \theta_k \beta_k^{\text{RP}}$</td>
<td>$\beta_k = \theta_k \beta_k^{\text{HS}}$</td>
<td>$\beta_k = \theta_k \beta_k^{\text{CD}}$</td>
<td>$\beta_k = \theta_k \beta_k^{\text{DY}}$</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>$x_1$</td>
<td>(1.9184,-1.1184)</td>
<td>(1.8813,-1.0814)</td>
<td>(1.8814,-1.0814)</td>
<td>(1.9183,-1.1184)</td>
</tr>
<tr>
<td>$x_2$</td>
<td>(1.9467,-1.0560)</td>
<td>(2.0079,-0.9818)</td>
<td>(2.0179,-0.9833)</td>
<td>(2.0299,-0.9792)</td>
</tr>
<tr>
<td>$x_3$</td>
<td>(2.0213,-0.9847)</td>
<td>(1.9965,-1.0007)</td>
<td>(1.9942,-1.0053)</td>
<td>(1.9954,-1.0019)</td>
</tr>
<tr>
<td>$x_4$</td>
<td>(2.0015,-0.9983)</td>
<td>(2.0008,-0.9988)</td>
<td>(2.0012,-0.9989)</td>
<td>(2.0010,-0.9997)</td>
</tr>
<tr>
<td>$x_5$</td>
<td>(2.0004,-1.0013)</td>
<td>(1.9995,-1.0001)</td>
<td>(1.9998,-1.0001)</td>
<td>(1.9998,-1.0001)</td>
</tr>
<tr>
<td>$x_6$</td>
<td>(1.9994,-1.0001)</td>
<td>(2.0000,-1.0000)</td>
<td>(2.0000,-1.0000)</td>
<td>(2.0000,-1.0000)</td>
</tr>
<tr>
<td>$x_7$</td>
<td>(2.0000,-1.0000)</td>
<td>(2.0000,-1.0000)</td>
<td>(2.0000,-1.0000)</td>
<td>(2.0000,-1.0000)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$x_0 = [1,1]$;</th>
<th>BFGS</th>
<th>The result in [6]</th>
<th>$\beta_k = \theta_k \beta_k^{\text{FR}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>(1,-0.5)</td>
<td>(2.5,-0.5)</td>
<td>(1.9183,-1.1184)</td>
</tr>
<tr>
<td>$x_2$</td>
<td>(1.45,-0.3875)</td>
<td>(2.07813,-0.82813)</td>
<td>(2.0236,-0.9859)</td>
</tr>
<tr>
<td>$x_3$</td>
<td>(1.5889,-0.63729)</td>
<td>(1.97052,-0.93611)</td>
<td>(1.9955,-1.0017)</td>
</tr>
<tr>
<td>$x_4$</td>
<td>(1.82541,-0.97156)</td>
<td>(1.96625,-1.0206)</td>
<td>(2.0008,-0.9998)</td>
</tr>
<tr>
<td>$x_5$</td>
<td>(1.94063,-1.04509)</td>
<td>(2.05113,-1.00761)</td>
<td>(1.9999,-1.0000)</td>
</tr>
<tr>
<td>$x_6$</td>
<td>(1.96414,-1.04509)</td>
<td>(2.00732,-0.999473)</td>
<td>(2,-1)</td>
</tr>
<tr>
<td>$x_7$</td>
<td>(1.99521,-1.00174)</td>
<td>(1.99999,-0.99998)</td>
<td>(2,-1)</td>
</tr>
<tr>
<td>$x_8$</td>
<td>(2.00007,-1.00043)</td>
<td>(2,-1)</td>
<td>(2,-1)</td>
</tr>
<tr>
<td>$x_9$</td>
<td>(1.99998,-0.99995)</td>
<td>(2,-1)</td>
<td>(2,-1)</td>
</tr>
<tr>
<td>$x_{10}$</td>
<td>(2.0,-1)</td>
<td>(2,-1)</td>
<td>(2,-1)</td>
</tr>
</tbody>
</table>

References