A NEW CONJUGATE GRADIENT METHOD WITH STRONGLY GLOBAL CONVERGENCE AND SUFFICIENT DESCENT CONDITION

DONG Xiao-liang¹, HE Yu-bo², KONG Xiang-yu¹, LI Wei-jun³
(¹.School of Mathematics and Information, Beifang University of Nationalities, Yinchuan, 710021, China)
(².Department of Mathematics and Application Mathematics, Huaihua University, Huaihua, 418008, China)
(³.Network Information Technology Center, Beifang University of Nationalities, Yinchuan, 710021, China)

Abstract: In this paper, we study the WYL conjugate gradient method for unconstrained optimization problems. By making use of the modified iterative scheme, the sufficient descent conditions are satisfied at each iteration independent of the line search used. Also, by removing the original restriction on the parameter of the Wolfe conditions, we establish the strongly global convergence property for the general function. Numerical results illustrate that our method is efficient for the test problems.

Keywords: conjugate gradient method; sufficient descent condition; strongly global convergence; Wolfe line search

2010 MR Subject Classification: 90C30; 65K10

1 Introduction

Consider the following unconstrained optimization problem

\[ \min f(x), \quad (1.1) \]

where \( f : \mathbb{R}^n \to \mathbb{R} \) is a smooth and nonlinear function, and its gradient \( g_k = \nabla f(x_k) \) is available. The conjugate gradient (CG) methods represent an important class of unconstrained optimization algorithms with strong local and global convergence properties and modest memory requirements.
The iterative formula of the CG method is given by

\[ x_{k+1} = x_k + s_k, \quad s_k = \alpha_k d_k. \quad (1.2) \]

In (1.2), \( d_k \) is a search direction updated by

\[ d_k = \begin{cases} 
-g_k, & k = 1, \\
-g_k + \beta_k d_{k-1}, & k \geq 2
\end{cases} \quad (1.3) \]

and the steplength \( \alpha_k > 0 \) is commonly chosen to satisfy certain line search conditions. Among them, the so-called Wolfe conditions have attracted special attention in the convergence analyses and the implementations of CG methods, requiring that

\[ f(x_k + \alpha_k d_k) - f(x_k) \leq \rho \alpha_k g_k^T d_k, \quad (1.4) \]
\[ g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \quad (1.5) \]

where \( 0 < \rho < \sigma < 1 \) are often imposed on the line search.

The well-known formulas \( \beta_k \) are Fletcher-Reeves, Hestenes–Stiefel, Polak-Ribi`ere-Polyak and Dai-Yuan formulas, which are specified by

\[ \beta_{k}^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \beta_{k}^{HS} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T y_{k-1}}, \]
\[ \beta_{k}^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}, \quad \beta_{k}^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, \quad (1.6) \]

where \( y_{k-1} = g_k - g_{k-1} \). Please refer to [1] for more details of their convergence properties.

Little is known concerning global convergence of the HS method for nonconvex minimization problems. However, the HS method, which resembles the PRP method in practical computation and is often recommended due to its superior numerical properties.

Recently, various modifications of the HS method received growing interests, in which sufficient descent condition is important in the convergence analysis of CG method.

Hager and Zhang [2] proposed the CG_DESCENT method and proved its convergence with the Wolfe search. Based on the MBFGS method [3], Zhang [4] and Dai [5] introduced modified HZ methods with the Armijo search for nonconvex objective functions. Zhang, Zhou and Li [6] proposed a three-term modified PRP method, in which the property \(-d_k^T g_k = \|g_k\|^2\) always holds without any line searches. To seek the CG direction that is closest to the direction of the scaled memoryless BFGS method, Dai and Kou [7] proposed a family of CGOPT methods. Yu et al. [8, 9] proposed several modified spectral CG methods. For further study on the some recent advances, we may refer to [10–15].

In this paper, we will continue to study the HS-type method. One feature of our method is that our method guarantee sufficient descent condition and strongly global convergence.

The rest of this paper is organized as follows. In the next section, the new method is presented formally. Meanwhile, we prove the proposed method possesses sufficient descent
and strongly global convergence properties with the Wolfe line search. In Section 4, we report numerical comparisons with existing methods by using some test problems.

2 A New CG Method and Its Properties

Recently, Wei, Yao and Huang [16] proposed the MHS, WYL and MLS methods, in which the parameters $\beta_k$ are given by

$$\beta_{MHS}^k = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \beta_{WYL}^k = \frac{g_k^T y_{k-1}}{||y_{k-1}||^2}, \beta_{MLS}^k = -\frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}},$$

where $y_{k-1} = g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}$.

The global convergence of these methods with the strong Wolfe line search was proved for cases that the parameter $\sigma$ in (1.5) is constrained to $\sigma \in [0, \frac{1}{3}]$, $\sigma \in [0, \frac{1}{4}]$ and $\sigma \in [0, \frac{1}{4}]$, respectively. Furthermore, with the same restriction on the parameter $\sigma$, these methods above possessed the sufficient descent condition.

The above discussion prompts naturally us to raise the following question: Is it possible to modify the direction of the MHS method in a suitable way, thereby enlarging its sufficient descent properties and ensuring the global convergence for the general functions?

By taking the theoretical advantage of the MHS method into consideration, we give another method to guarantee sufficient descent condition, in which strongly global convergence is satisfied. With our choices, the additional restriction on the parameter $\sigma$ in (1.5) can be removed. The direction $d_k$ in our method is given by

$$d_{DHS}^k = \begin{cases} -g_k, & d_{k-1}^T y_{k-1} \leq \varepsilon_1 \|y_{k-1}\| \|d_{k-1}\|, \\ -g_k + \beta_{DHS}^k d_{k-1} + \varphi_k g_{k-1}, & d_{k-1}^T y_{k-1} > \varepsilon_1 \|y_{k-1}\| \|d_{k-1}\|, \end{cases}$$

where $\varepsilon_1 > 0$ is a constant.

In (2.2), the parameters $\beta_{DHS}^k$ and $\varphi_k$ are chosen as

$$\beta_{DHS}^k = \frac{g_k^T (g_k - \|g_k\| g_{k-1})}{\max\{d_{k-1}^T y_{k-1}, \lambda |d_{k-1}^T g_k|\}},$$

$$\varphi_k = \frac{g_k^T d_{k-1}}{\max\{d_{k-1}^T y_{k-1}, \lambda |d_{k-1}^T g_k|\} \|g_{k-1}\|},$$

where $\lambda > 1$ and $\varepsilon_1 > 0$ are two given constants.

For convenience, we call our method (2.2) as DHS method in later part of this paper and formally state the steps of this method as follows.

**Algorithm 2.1 (DHS method)**

**Step 1** Choosing constants $\lambda > 1$, $\varepsilon_1 > 0$ and $\varepsilon > 0$. Select an initial point $x_1 \in R^n$, set $k = 1$. 

Step 2 Test a criterion for stopping the iterations. If \(\|g_k\| < \varepsilon\), then stop, otherwise calculate \(d_k\) by (2.2).

Step 3 Determine the steplength by the Wolfe conditions.

Step 4 Calculate the new iterate by \(x_{k+1} = x_k + \alpha_k d_k\), set \(k = k + 1\) and goto Step 2.

The descent property is an indispensable factor in the convergence analysis of CG method. More exactly, if there exists a constant \(c_1 > 0\) such that
\[
d_k^T g_k \leq -c_1\|g_k\|^2, \forall k \in \mathbb{N}
\] (2.5)
holds for all \(k \in \mathbb{N}\), then the so-called sufficient descent condition holds.

Property (2.5) is guaranteed in our method, as proved in the following lemma.

Lemma 2.1 Let \(\{x_k\}\) and \(\{d_k\}\) be generated by the DHS method. Then the direction \(d_k\) satisfies the sufficient descent condition (2.5), which is independent of any line search.

Proof For \(k = 1\), it follows that \(d_1^T g_1 = -\|g_1\|^2\). Now we mainly consider the case where \(\frac{d_{k-1}^T y_{k-1}}{d_{k-1}^T} > \varepsilon_1\|y_{k-1}\||d_{k-1}|\) is satisfied. It follows from (2.2) that
\[
g_k^T d_k = -\|g_k\|^2 + \frac{g_k^T d_{k-1}}{\max\{d_{k-1}^T y_{k-1}, \lambda |d_{k-1}^T g_k|\}}\|g_k\|^2
\]
\[
\leq -\|g_k\|^2 + \frac{|g_k^T d_{k-1}|}{\max\{d_{k-1}^T y_{k-1}, \lambda |d_{k-1}^T g_k|\}}\|g_k\|^2
\]
\[
\leq - \left(1 - \frac{1}{\lambda}\right) \|g_k\|^2.
\] (2.6)

Then the conclusion holds by letting \(c_1 = 1 - \lambda^{-1}\).

3 Convergence Analysis

In this section, we prove the global convergence of the proposed CG method. We need the following assumptions, which are generally assumed in the literature.

Assumption 3.1

Boundedness Assumption: the level set defined by \(\Omega = \{x \in \mathbb{R}^n | f(x) \leq f(x_1)\}\) with \(x_1\) to be the initial point, is bounded;

Lipschitz Assumption: in some neighborhood \(\Omega_0\) of \(\Omega\), the objective function \(f\) is continuously differentiable, and its gradient \(g\) is Lipschitz continuous, namely, there exist a constant \(L > 0\) such that
\[
\|g(x) - g(y)\| \leq L\|x - y\|, \forall x, y \in \Omega_0.
\] (3.1)

Theorem 3.1 Suppose that Assumption 3.1 holds. Let \(\{x_k\}\) and \(\{d_k\}\) be generated by Algorithm 2.1. Then there exists a constant \(\overline{\sigma}\) such that
\[
\|d_k\| \leq \overline{\sigma}\|g_k\|, \forall k \in \mathbb{N}.
\] (3.2)
Proof Since \( d_1 = -g_1 \), the result obviously holds for \( k = 1 \).
If \( k > 1 \), it suffice to consider the case where \( d_{k-1}^T y_{k-1} > \varepsilon_1 \| y_{k-1} \| \| d_{k-1} \| \) holds.
It follows from the definition \( \beta_k^{DHS} \) in (2.3) and the Cauchy inequality that
\[
|\beta_k^{DHS}| \leq \frac{\| g_k - g_{k-1} \| + \| g_k \| - \| g_{k-1} \| \| g_{k-1} \|}{d_{k-1}^T (g_k - g_{k-1})} \\
\leq \frac{2 \| g_k \| \| y_{k-1} \|}{|d_{k-1}^T y_{k-1}|} \leq \frac{2}{\varepsilon_1} \| d_{k-1} \|. 
\]
Also, we can estimate the upper bound for \( |\varphi_k| \), presented by
\[
|\varphi_k| = \left| \frac{g_k^T d_{k-1}}{\max \{ d_{k-1}^T y_{k-1}, \lambda \| d_{k-1}^T g_k \| \} } \right| \cdot \frac{\| g_k \|}{\| g_{k-1} \|} \leq \frac{1}{\lambda} \frac{\| g_k \|}{\| g_{k-1} \|}. \tag{3.3}
\]
Combining this with (3.3) yields
\[
\| d_k \| \leq \| g_k \| + |\beta_k^{DHS}| \| d_{k-1} \| + |\varphi_k| \| g_{k-1} \| \\
\leq \left( 1 + \frac{2}{\varepsilon_1} + \frac{1}{\lambda} \right) \| g_k \|. \tag{3.4}
\]
Let \( \overline{\varepsilon} = 1 + \frac{2}{\varepsilon_1} + \frac{1}{\lambda} \), we complete the proof.

The conclusion of the following theorem, called the Zoutendijk condition, is often used to prove global convergence of nonlinear CG method.

**Theorem 3.2** (see [17]) Suppose that Assumption 3.1 holds. Consider the general CG method, where \( d_k \) is a descent direction and \( \alpha_k \) satisfies the Wolfe conditions. Then we have
\[
\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\| d_k \|^2} < \infty. \tag{3.5}
\]
For the general function, we can develop a strongly global convergence result as follows.

**Theorem 3.3** Suppose that Assumption 3.1 holds. Let \( \{ x_k \} \) be generated by Algorithm 2.1. Then
\[
\lim_{k \to \infty} \| g_k \| = 0. \tag{3.6}
\]

**Proof** The bound for \( \| d_k \| \) in (3.2) coupled with (2.5) indicates that
\[
\sum_{k \geq 1} (\overline{\varepsilon})^{-2} \| g_k \|^2 \leq \sum_{k \geq 1} \| g_k \|^4 \leq \left( 1 - \frac{1}{\lambda} \right)^{-2} \sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\| d_k \|^2} < +\infty. \tag{3.7}
\]
Equation (3.7) leads to (3.6).

**4 Numerical Experiments**

In this section, we provide the implementation detail of the new methods to verify the numerical performance. Our tests problems come form Moré [18].
We stop the iteration if the inequality $||g_k|| \leq 10^{-6}$ is satisfied. We use "F" to denote the numbers of iteration exceed 2000. All algorithms use exactly the same implementation of the strong Wolfe conditions with $\rho = 10^{-3}, \sigma = 0.5$. The parameters in (2.2) are specified by $\varepsilon_1 = 10^{-12}, \mu = 10$.

In order to rank the CG methods, we subsequently use the method of Dai and Ni [19] to compare the performance of our methods with that of the other methods in [16].

First, we compute the total number of the function and its gradient evaluations by the formula $N_{total} = NF + m \times NG$, where $m=5$.

Second, for all our two methods, we evaluate their efficiency with respect to the WYL method in the following way. For each problem $i$, compute the ratio

$$r_i = \frac{N_{total,i}(\text{a method in [16]})}{N_{total,i}(\text{DHS})}$$

and the geometric mean of these ratios over all the test problems $r_{total} = \left( \prod_{i \in S} r_i \right)^{\frac{1}{|S|}}$, where $S$ denotes the set of the problems and $|S|$ the number of elements in $S$.

### Table 4.1: Numerical Results

<table>
<thead>
<tr>
<th>Problem</th>
<th>N</th>
<th>DHS NI/NF/NG/TIME</th>
<th>MHS NI/NF/NG/TIME</th>
<th>WYL NI/NF/NG/TIME</th>
<th>MLS NI/NF/NG/TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROSE 2</td>
<td>36/120/102/0.094</td>
<td>93/254/218/0.03</td>
<td>105/282/248/0.034</td>
<td>61/176/151/0.021</td>
<td></td>
</tr>
<tr>
<td>FROTH 2</td>
<td>19/76/64/0.0086</td>
<td>24/95/79/0.01</td>
<td>39/147/125/0.016</td>
<td>25/102/85/0.01</td>
<td></td>
</tr>
<tr>
<td>BADSCP 2</td>
<td>48/245/219/0.025</td>
<td>F</td>
<td>F</td>
<td>902/2548/2345/0.34</td>
<td></td>
</tr>
<tr>
<td>BADSCB 2</td>
<td>13/56/45/0.0061</td>
<td>F</td>
<td>F</td>
<td>253/1165/836/0.12</td>
<td></td>
</tr>
<tr>
<td>BEALE 2</td>
<td>14/47/37/0.0069</td>
<td>16/57/43/0.0073</td>
<td>17/61/47/0.0073</td>
<td>26/86/69/0.01</td>
<td></td>
</tr>
<tr>
<td>JENSAM 2</td>
<td>13/43/31/0.0061</td>
<td>15/51/39/0.0064</td>
<td>19/63/50/0.0087</td>
<td>18/61/47/0.0085</td>
<td></td>
</tr>
<tr>
<td>HELIX 3</td>
<td>55/149/121/0.026</td>
<td>100/262/221/0.037</td>
<td>134/364/312/0.05</td>
<td>43/121/95/0.016</td>
<td></td>
</tr>
<tr>
<td>BARD 3</td>
<td>28/88/77/0.018</td>
<td>22/69/58/0.014</td>
<td>25/81/68/0.016</td>
<td>37/106/85/0.02</td>
<td></td>
</tr>
<tr>
<td>GAUSS 3</td>
<td>4/9/5/0.0023</td>
<td>4/9/5/0.0018</td>
<td>4/9/5/0.0026</td>
<td>4/9/5/0.0018</td>
<td></td>
</tr>
<tr>
<td>MEYER 3</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>GULF 7</td>
<td>29/134/131/0.059</td>
<td>29/131/130/0.023</td>
<td>31/140/119/0.025</td>
<td>30/122/99/0.022</td>
<td></td>
</tr>
<tr>
<td>BOX 3</td>
<td>30/125/99/0.029</td>
<td>19/67/54/0.011</td>
<td>174/525/456/0.085</td>
<td>99/303/259/0.049</td>
<td></td>
</tr>
<tr>
<td>SING 4</td>
<td>188/588/514/0.07</td>
<td>324/1041/923/0.12</td>
<td>425/1354/1201/0.16</td>
<td>525/1625/1433/0.22</td>
<td></td>
</tr>
<tr>
<td>WOOD 4</td>
<td>79/286/246/0.034</td>
<td>407/1093/989/0.14</td>
<td>377/1093/923/0.13</td>
<td>343/989/830/0.12</td>
<td></td>
</tr>
<tr>
<td>KOWOSB 4</td>
<td>211/650/580/0.1</td>
<td>51/165/146/0.027</td>
<td>57/173/149/0.028</td>
<td>63/231/201/0.034</td>
<td></td>
</tr>
<tr>
<td>BD 4</td>
<td>64/866/372/0.27</td>
<td>F</td>
<td>57/491/256/0.14</td>
<td>53/770/355/0.2</td>
<td></td>
</tr>
<tr>
<td>OSBI 5</td>
<td>294/2574/2290/0.67</td>
<td>F</td>
<td>F</td>
<td>336/1172/1027/0.29</td>
<td></td>
</tr>
<tr>
<td>BIGGS 6</td>
<td>75/264/232/0.046</td>
<td>142/486/429/0.087</td>
<td>113/385/335/0.068</td>
<td>184/542/487/0.11</td>
<td></td>
</tr>
<tr>
<td>OSR2 11</td>
<td>215/559/509/0.31</td>
<td>183/481/426/0.27</td>
<td>217/589/527/0.33</td>
<td>222/586/1354/0.34</td>
<td></td>
</tr>
<tr>
<td>WXAN 20</td>
<td>1039/3521/3102/4.6</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td>ROSEX 10000</td>
<td>39/132/112/2</td>
<td>119/361/320/5.7</td>
<td>64/206/185/3.3</td>
<td>84/264/228/4.1</td>
<td></td>
</tr>
<tr>
<td>SINGX 1000</td>
<td>361/1175/1029/19</td>
<td>1509/5135/4486/78</td>
<td>265/832/733/13</td>
<td>1832/5809/5237/90</td>
<td></td>
</tr>
<tr>
<td>PEN1 20</td>
<td>141/625/524/0.21</td>
<td>100/402/348/0.14</td>
<td>100/404/349/0.14</td>
<td>120/461/404/0.16</td>
<td></td>
</tr>
<tr>
<td>PEN2 20</td>
<td>144/658/580/0.13</td>
<td>150/617/537/0.13</td>
<td>107/443/376/0.087</td>
<td>347/1191/1036/0.25</td>
<td></td>
</tr>
<tr>
<td>VARDIM 100</td>
<td>10/9/59/0.028</td>
<td>7/92/61/0.027</td>
<td>9/84/55/0.032</td>
<td>9/84/55/0.026</td>
<td></td>
</tr>
<tr>
<td>TRG 1000</td>
<td>199/123/109/0.3</td>
<td>50/128/113/0.31</td>
<td>50/128/116/0.31</td>
<td>49/125/108/0.3</td>
<td></td>
</tr>
<tr>
<td>BV 1000</td>
<td>0/1/1/0.022</td>
<td>0/1/1/0.021</td>
<td>0/1/1/0.02</td>
<td>0/1/1/0.02</td>
<td></td>
</tr>
<tr>
<td>IE 500</td>
<td>6/14/9/1.7</td>
<td>6/14/9/1.7</td>
<td>6/14/9/1.7</td>
<td>6/14/9/1.7</td>
<td></td>
</tr>
<tr>
<td>BAND 100</td>
<td>18/54/33/0.038</td>
<td>18/57/35/0.043</td>
<td>19/59/36/0.041</td>
<td>19/59/36/0.041</td>
<td></td>
</tr>
<tr>
<td>LIN 100</td>
<td>1/8/8/0.015</td>
<td>1/8/8/0.014</td>
<td>1/8/8/0.014</td>
<td>1/8/8/0.014</td>
<td></td>
</tr>
<tr>
<td>LIN 500</td>
<td>1/9/9/0.28</td>
<td>1/9/9/0.29</td>
<td>1/9/9/0.28</td>
<td>1/9/9/0.29</td>
<td></td>
</tr>
</tbody>
</table>
Finally, we present the $r_{\text{total}}$ in the Table 4.2:

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>DHS</th>
<th>MHS</th>
<th>WYL</th>
<th>MLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_{\text{total}}$</td>
<td>1.1216</td>
<td>1.1642</td>
<td>1.3124</td>
<td>1.3124</td>
</tr>
</tbody>
</table>

References


一类新的具有充分下降条件和强收敛性的共轭梯度法

董晓宽1,2, 何郁波3, 孔翔宇1, 李卫军1
(1.北方民族大学数学与信息学院, 宁夏 银川 750021)
(2.怀化学院数学与应用数学系, 湖南 怀化 418008)
(3.北方民族大学网络信息技术中心, 宁夏 银川 750021)

摘要: 本文研究了解无约束优化问题的WYL共轭梯度法。利用修正迭代格式, 得到了算法在每步迭代能产生不依赖于搜索条件的充分下降方向。同时, 在原算法中关于Wolfe条件中参数去掉的情况下, 获得了本文算法是强收敛的。数值实验表明本文算法可以有效地求解测试问题。

关键词: 共轭梯度法; 充分下降条件; 强收敛性; Wolfe搜索