New Projection Matrix for the Stander Conjugate Gradient Method

Dr. ABBAS Y. AL – BAYATI      Dr. EMAN T. HAMED
Dr. Hamsa Th. Chilmerane

Abstract:

In this paper, we have derived anew proposed algorithm for conjugate gradient method based on a projection matrix. This Algorithm satisfies the sufficient descent condition and the globally converges. Numerical comparisons with a standard conjugate gradient algorithm show that this algorithm very effective depending on the number of iterations and the number of functions evaluation.

Keywords: unconstrained optimization, a projection matrix, inexact line search, wolf line search, global convergence.

1-Introduction:

Let us consider the nonlinear unconstrained optimization problem

\[ \min\{f(x) : x \in R^n\} \]

Where \( f \) is smooth and its gradient \( g \) is available.
New Projection Matrix for the Stander Conjugate Gradient Method

Conjugate gradient methods are efficient for solving (1), especially when the dimension $n$ is large. The iterates of conjugate gradient methods for solving (1) are obtained by

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

where $\lambda_k$ is a steplength, which is computed by carrying out some line search, and $d_k$ is the search direction defined by

$$
\begin{bmatrix}
d_k = -g_k \\
d_{k+1} = -g_{k+1} + \beta_k d_k
\end{bmatrix}
$$

where $\beta_k$ is a scalar. Some well-known conjugate gradient methods include the Hestenes–Stiefel (HS) method, Fletcher–Reeves (FR) method, the Polak–Ribière–Polyak (PRP) method, and the Dai–Yuan (DY) method and Al-Bayati & Al-Assady. The parameters $\beta_k$ of these methods are specified as follows

$$
\beta_k^{HS} = \frac{g_k^T y_k}{d_k^T y_k} \quad \text{(Hestenes-Stiefel,1952)}
$$

$$
\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}^T}{g_k^T g_k} \quad \text{(Fletcher-Reeves (FR),1969)}
$$

$$
\beta_k^{PR} = \frac{g_k^T y_k}{g_k^T y_k} \quad \text{(Polak- Ribière (PR))}
$$

$$
\beta_k^{BA} = \frac{-y_k^T y_k}{d_k^T y_k} \quad \text{(Al-Bayati & Al-Assady,1986)}
$$

$$
\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}^T}{d_k^T y_k} \quad \text{(Dai-Yuan (DY),1999)}
$$
The stepsize $\lambda_k$ is usually chosen to satisfy certain line search conditions. Among them, the so-called strong Wolfe line search conditions require that, the weak Wolfe-conditions:

$$f(x_k + \lambda_k d_k) - f(x_k) \leq \delta \lambda_k g_k^T d_k \quad \ldots \quad (4)$$

$$g(x_k + \lambda_k d_k)^T d_k \geq \sigma g_k^T d_k \quad \ldots \quad (5)$$

the strong Wolfe-conditions:

$$f(x_k + \lambda_k d_k) - f(x_k) \leq \delta \lambda_k g_k^T d_k \quad \ldots \quad (6)$$

$$|g(x_k + \lambda_k d_k)^T d_k| \leq -\sigma g_k^T d_k \quad \ldots \quad (7)$$

where $\delta \in (0, \frac{1}{2})$ and $\sigma \in (0,1)$

Beale’s three-term restart conjugate gradient method is a well-known three-term conjugate gradient method in which

$$d_{k+1} = -g_{k+1} + \beta_k d_k + \gamma_k d_l \quad \ldots \quad (8)$$

where $1 \leq l < k$, $\gamma_k = g_k^T y_l / d_l^T y_l$. Another well-known method is Nazareth’s three-term recurrence, where

$$d_{k+1} = -y_k + \frac{y_{k-1}^T y_{k-1}}{y_{k-1}^T d_{k-1}} d_{k-1} + \frac{y_k^T y_k}{y_k^T d_k} \quad d_k \quad \ldots \quad (9)$$

Zhang, Li, Zhou, Weijun, (2007) and Zhang, L., Weijun Zhou, (2007) proposed a three-term PRP conjugate gradient method (TTPRP) and a three-term FR conjugate gradient method (TTFR), respectively, that is,

$$d_{k+1}^{TTPRP} = -g_{k+1} + \beta_k^{PRP} d_k - \theta^1_k y_k \quad \ldots \quad (10)$$

$$d_{k+1}^{TTFR} = -g_{k+1} + \beta_k^{FR} d_k - \theta^2_k g_{k+1} \quad \ldots \quad (11)$$

$$\theta^1_k = \frac{g_{k+1}^T d_k}{\|g_k\|} \quad \text{and} \quad \theta^2_k = \frac{d_k^T y_k}{\|g_k\|^2}$$

[110]
New Projection Matrix for the Standard Conjugate Gradient Method

2. New proposed method

In this paper, we will get new projection matrix from three-term CG-algorithm as follows:

\[ d_{k+1} = -g_{k+1} + \beta^k R \cdot d_k - \beta^k R \cdot \frac{g^T_{k+1} d_k}{d^T_k y_k} y_k \]  
\[ \text{(12)} \]

If we use the exact line search and, then our method (12) becomes the nonlinear conjugate gradient method (3)

\[ d_{k+1} = -g_{k+1} + \beta^k R \cdot (I - \frac{y_k g^T_{k+1}}{d^T_k y_k}) d_k \]  
\[ \text{(13)} \]

Where the matrix \( I - \frac{y_k g^T_{k+1}}{d^T_k y_k} \) is also a projection matrix into the orthogonal complement of \( \text{Span} \{ g_{k+1} \} \)

3. The New Algorithm:

Step 1: For the initial point \( x_0 \), \( x_i \in \mathbb{R}^n, \epsilon \), Set \( d_i = -g_i \), \( k = 1 \), if \( \| g_i \| \leq \epsilon \), then stop.
Step 2: Set \( d_k = -g_k \)
Step 3: Find \( \lambda_k > 0 \) satisfying the strong wolf conditions.
Step 4: Let \( x_{k+1} = x_k + \lambda_k d_k \) and If \( \| g_{k+1} \| \leq \epsilon \) then stop.

Step 5: Compute the search direction \( d_{k+1} \) by (13)

Step 6: If \( k = n \) or \( \frac{g^T_k g_{k+1}}{\| g_{k+1} \|^2} \geq 0.2 \), then go to step 2.
Step 7: Set \( k: = k + 1 \), go to Step 3.

4. Global Convergence Properties for the new Suggestion algorithm:

In this section, we will study the convergence of the new proposed method depending by the following assumption

Assumption(A):
(i) The level set $S = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ is bounded.

(ii) In a neighborhood $N$ of $S$, the function $f$ is continuously differentiable and its gradient is Lipschitz continuous, i.e there exists a constant $L > 0$ such that

$$\|\nabla f(x) - f(y)\| \leq L \|x - y\| \quad \text{for all } x, y \in N ..(14)$$

We can get from assumption (A) that there exists positive constant $\psi > 0$, such that:

$$\|g(x)\| \leq \psi \quad \forall x \in S \quad \text{ ...(15)}$$

**Lemma (1).** Suppose that the assumption (A) hold and consider any conjugate gradient method (2) and (3), where is a descent direction $d_k$ and $\lambda_k$ is obtained by the strong Wolfe line search

If

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} = \infty \quad \text{ .................(16)}$$

Then

$$\liminf_{k \to \infty} \|g_k\| = 0 \quad \text{ .....................................(17)}$$

(Dai, Y.H., et al, 1999)

**Lemma:**

Suppose the assumption (A) hold , let the sequence $\{x_k\}$ generated by (2) and the step length $\lambda_k$ satisfies wolf conditions, then the direction which is define in (13)is satisfied sufficient condition

**Proof:**

By multiply both side of (13) by $g_{k+1}^T$ and dividing by $\|g_{k+1}\|^2$ we get :

$$\frac{g_{k+1}^T d_k + \|g_{k+1}\|^2}{\|g_k\|^2} = \beta_k g_{k+1}^T d_k - \beta_k (g_{k+1}^T y_k)(g_{k+1}^T d_k) \quad \text{ .................(18)}$$

By using strong wolf condition we get:

$$\frac{g_{k+1}^T d_k + \|g_{k+1}\|^2}{\|g_k\|^2} \leq \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \left( \frac{-\sigma d_k^T g_k}{\|g_k\|^2} \right) + \left( g_{k+1}^T y_k \right)(\sigma d_k^T g_k) \quad \text{ .................(19)}$$

[112]
New Projection Matrix for the Stander Conjugate Gradient Method

Since \( g_k^T y_k \leq \| g_k \| \| y_k \| \)

\[
\frac{g_k^T d_{k+1} + \| g_{k+1} \|^2}{\| g_{k+1} \|^2} \leq -\sigma d_k^T g_k + \| g_{k+1} \| \| y_k \| \sigma d_k^T g_k \]
\[
\frac{g_k^T d_{k+1} + \| g_{k+1} \|^2}{\| g_{k+1} \|^2} \leq -\sigma d_k^T y_k - \frac{\| y_k \| \sigma d_k^T y_k}{(\sigma + 1) \| g_k \|^2} \]
\[
\frac{g_k^T d_{k+1} + \| g_{k+1} \|^2}{\| g_{k+1} \|^2} \leq \frac{\| \sigma d_k^T y_k \}}{(\sigma + 1) \| g_k \|^2} = c \]

where \( c \) is small positive constant

\( \therefore g_k^T d_{k+1} \leq -(1-c) \| g_{k+1} \|^2 \)

the proof is complete.

**Theorem (2):**

(Global convergence for new proposed method):

Consider the iteration method which is define in (2) where \( d_k \) defined by,(13) and suppose the assumption A holds. Then the new algorithm either stops at stationary point i.e. \( \| g_k \| = 0 \) or \( \liminf_{k \to \infty} \| g_k \| = 0 \)

**Proof:**

Form (12),we get

\[
d_{k+1} = -g_{k+1} + \beta_k^{FR} d_k - \beta_k^{FR} \frac{g_k^T d_k}{d_k^T y_k} y_k
\]

\[
\| d_{k+1} \| \leq \| g_{k+1} \| + \| \beta_k^{FR} \| \| d_k \| + \| \beta_k^{FR} \| \frac{g_k^T d_k}{d_k^T y_k} \| y_k \|
\]

\[
\| d_{k+1} \| \leq \| g_{k+1} \| + \| \beta_k^{FR} \| \| d_k \| + \| \beta_k^{FR} \| \frac{\| g_k \|^2}{d_k^T y_k} \| y_k \|
\]

\[
\sum_{k=1}^{\infty} \frac{1}{\| d_{k+1} \|^2} \geq \sum_{k=1}^{\infty} \frac{1}{\psi} = \sum_{k=1}^{\infty} 1 = \infty
\]

\[
\sum_{k=1}^{\infty} \frac{1}{\| d_{k+1} \|^2} = \infty
\]

by using lemma(1) we get :

\( \lim_{k \to \infty} \| g_k \| = 0 \)

**5: Numerical experiments**

In this section, we will test the feasibility and effectiveness of the Algorithm 2.1. The algorithm is implemented in Fortran77 code using
double precision arithmetic and Comparison our new algorithm with standard three term CG-algorithm in Table (1)

overall the calculation and for different dimension for \((100 \leq n \leq 5000)\), all the algorithms in this paper use the same ILS strategy.

All the results are obtained using (Pentium 4 computer). All programs are written in FORTRAN 90 language and for all cases the stopping criterion taken to be:

\[\|g_{k+1}\| \leq 10^{-5}\]

The comparative performance for all of these algorithms is evaluated by considering number of function Evaluations \(NOF\) and number of iterations \(NOI\).

Table (1) Comparison of our new algorithm with standard FR CG-algorithm.

<table>
<thead>
<tr>
<th>Test fun.</th>
<th>Dim.</th>
<th>FR. Algorithm</th>
<th>New Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NOI</td>
<td>NOF</td>
</tr>
<tr>
<td>Powell</td>
<td>100</td>
<td>50</td>
<td>136</td>
</tr>
<tr>
<td>Central</td>
<td>100</td>
<td>33</td>
<td>222</td>
</tr>
<tr>
<td>Edger</td>
<td>100</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>Cubic</td>
<td>100</td>
<td>16</td>
<td>44</td>
</tr>
<tr>
<td>Wolfe</td>
<td>100</td>
<td>49</td>
<td>99</td>
</tr>
<tr>
<td>Sum</td>
<td>100</td>
<td>12</td>
<td>64</td>
</tr>
<tr>
<td>Wood</td>
<td>100</td>
<td>29</td>
<td>67</td>
</tr>
<tr>
<td>Miele</td>
<td>100</td>
<td>46</td>
<td>146</td>
</tr>
<tr>
<td>Rosen</td>
<td>100</td>
<td>22</td>
<td>55</td>
</tr>
<tr>
<td>Recip</td>
<td>100</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>Powell</td>
<td>1000</td>
<td>54</td>
<td>164</td>
</tr>
<tr>
<td>Central</td>
<td>1000</td>
<td>40</td>
<td>312</td>
</tr>
<tr>
<td>Edger</td>
<td>1000</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>Cubic</td>
<td>1000</td>
<td>16</td>
<td>44</td>
</tr>
<tr>
<td>Wolfe</td>
<td>1000</td>
<td>64</td>
<td>129</td>
</tr>
<tr>
<td>Sum</td>
<td>1000</td>
<td>21</td>
<td>110</td>
</tr>
<tr>
<td>Wood</td>
<td>1000</td>
<td>29</td>
<td>67</td>
</tr>
</tbody>
</table>

[114]
New Projection Matrix for the Stander Conjugate Gradient Method

<table>
<thead>
<tr>
<th></th>
<th>1000</th>
<th>53</th>
<th>180</th>
<th>45</th>
<th>144</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miele</td>
<td>1000</td>
<td>22</td>
<td>55</td>
<td>22</td>
<td>55</td>
</tr>
<tr>
<td>Rosen</td>
<td>1000</td>
<td>5</td>
<td>16</td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>Recip</td>
<td>1000</td>
<td>56</td>
<td>168</td>
<td>43</td>
<td>128</td>
</tr>
<tr>
<td>Powell</td>
<td>10000</td>
<td>45</td>
<td>384</td>
<td>34</td>
<td>292</td>
</tr>
<tr>
<td>Edger</td>
<td>10000</td>
<td>6</td>
<td>15</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>Cubic</td>
<td>10000</td>
<td>16</td>
<td>44</td>
<td>16</td>
<td>46</td>
</tr>
<tr>
<td>Wolfe</td>
<td>10000</td>
<td>118</td>
<td>238</td>
<td>131</td>
<td>266</td>
</tr>
<tr>
<td>Sum</td>
<td>10000</td>
<td>32</td>
<td>161</td>
<td>35</td>
<td>165</td>
</tr>
<tr>
<td>Wood</td>
<td>10000</td>
<td>29</td>
<td>67</td>
<td>30</td>
<td>69</td>
</tr>
<tr>
<td>Miele</td>
<td>10000</td>
<td>53</td>
<td>180</td>
<td>53</td>
<td>182</td>
</tr>
<tr>
<td>Rosen</td>
<td>10000</td>
<td>22</td>
<td>55</td>
<td>22</td>
<td>55</td>
</tr>
<tr>
<td>Recip</td>
<td>10000</td>
<td>6</td>
<td>18</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>Total</td>
<td>961</td>
<td>3286</td>
<td>903</td>
<td>3039</td>
<td></td>
</tr>
</tbody>
</table>

References
Wanyou Cheng y Dong-Hui Li (2010)z "A Two Term PRP Based Descent Method " College of Mathematics and Econometrics, Hunan University.


Appendix
1. Generalized Central Function:
\[ f(x) = \sum_{i=1}^{n/4} \left[ \left( \exp(x_{i+3}) - x_{i+2} \right)^4 + 100(x_{i+2} - x_{i+1})^6 + \right] \] 
\[ \arctan(x_{i+1} - x_i))^4 + x_{i+3}, \] 
\[ x_0 = (1, 2, 2; \ldots)^T. \]

2. Extended Wood Function
\[ f(x) = \sum_{i=1}^{n/4} \left[ 100(x_{i+3}^2 - x_{i+2})^2 + (x_{i+2} - 1)^2 + 90(x_{i+1} - x_i)^2 + \right] 
\[ (1 - x_{i+1})^2 + 10.1((x_{i+2} - 1)^2 + (x_{i+1} - 1)^2) + 19.8(x_{i+2} - 1)(x_i - 1)), \] 
\[ x_0 = (-3, -1, -1, \ldots, -3, -1)^T. \]

3. Generalized Powell Function:
\[ f(x) = \sum_{i=1}^{n/4} \left[ \left( x_{i+3} + 10x_{i+2} \right)^2 + 5(x_{i+2} - x_i)^2 + \right] 
\[ \left( x_{i+2} - 2x_{i+1} \right)^4 + 10(x_{i+3} - x_i)^4 \right] \] 
\[ x_0 = (3, -1, 0, 1, \ldots, 3, -1, 0, 1)^T. \]

4. Sum of Quadritics (SUM) function:
\[ f(x) = \sum_{i=1}^{n} (x_i - i)^i . \] 
\[ x_0 = (1, \ldots)^T. \]

5. Wolfe Function:
New Projection Matrix for the Stander Conjugate Gradient Method

\[
f(x) = [-x_i (3 - x_i / 2) + 2x_2 - 1]^2 + \sum_{i=1}^{n-1} \left[ x_i - x_i (3 - x_i / 2 + 2x_{i+1} - 1) \right]^2 \
\]

\[
x_0 = (-1, \ldots)^T
\]

6. Rosenbrock Function:

\[
f(x) = \sum_{i=1}^{n/2} 100(x_{2i} - x_{2i-1}^2)^2 + (1 - x_{2i-1})^2
\]

\[
x_0 = (-1.2, 1, \ldots)^T
\]

7. Generalized Recip Function:

\[
f(x) = \sum_{i=1}^{n/3} \left[ (x_{3i-1} - 5)^2 + x_{9i-1}^2 + x_{3i}^2 \right]
\]

\[
X_0 = (2.5, 1, \ldots, 2, 5, 1)^T
\]

8. Miele Function:

\[
f(x) = \sum_{i=1}^{n/4} (\exp(x_{4i-3}) + 10x_{4i-2})^2 + 100(x_{4i-2} + x_{4i-1})^6
\]

\[
+ (\tan(x_{4i-1} - x_{4i}))^4 + (x_{4i-3})^8 + (x_{4i} - 1)^2,
\]

\[
x_0 = (1, 2, 2, \ldots, 1, 2, 2)^T
\]

9. Generalized Edger Function:

\[
f(x) = \sum_{i=1}^{n/2} (x_{2i-1} - 2)^4 + (x_{2i-1} - 2)^2 x_{2i}^2 + (x_{2i} + 1)^2
\]

\[
x_0 = (1, 0, \ldots, 1, 0)^T
\]

10. Generalized Cubic Function:

\[
f(x) = \sum_{i=1}^{n/2} \left[ 100(x_{3i} - x_{3i-1}^3)^2 + (1 - x_{2i-1})^2 \right]
\]

\[
x_0 = (-1.2, 1, \ldots, -1.2, 1)^T
\]