APPLICATION OF SPECTRAL PRP CONJUGATE GRADIENT PARAMETER FOR UNCONSTRAINED OPTIMIZATION PROBLEMS

Usman Abbas Yakubu\(^{a,b}\), Ibrahim Abdullahi\(^b\), Salisu Murtala\(^c\) and Abba V Mandara\(^d\)

\(^{a,b}\)Department of Mathematics, Yusuf Maitama Sule University, Kano, Nigeria.  
\(^c\)Department of Mathematics, Federal University Dutse, Jigawa, Nigeria.  
\(^d\)Department of Mathematics, University of Maiduguri, Borno, Nigeria.  
\(^*\)usman.abbas84@yahoo.com

Abstract: Conjugate Gradient (CG) method have been utilised to solve nonlinear unconstrained optimization problems because of less storage locations and fewer computational cost in dealing with large-scale problems. In this paper, we present a real life application of spectral PRP Conjugate Gradient method in regression analysis, the proposed method is suitably deriving from the CG search direction without secant condition. Some benchmark functions with several variables have been use to prove the global convergence properties and satisfies sufficient descent condition. The numerical results are certifying by exact line search techniques; the method outperform the prominent least square method.

Keywords: Sufficient descent property; exact line search; spectral CG; global convergence; Regression analysis.

1. INTRODUCTION

One of the most essential features of the CG method is to solve a large number of unconstrained optimization problems within a shortest period and less number of iterations. CG methods require only small storage location and has less computational cost, since they do not use the Hessian matrix or its approximation. The CG method has a rapid global convergent property and satisfies descent condition. The uniqueness of this method is utmost due to its simplicity in both algebraic processes and the development of computer codes. Thus, the method is effective and capable in solving large-scale unconstrained minimization problems [14]. [3] introduced a spectral CG method and they computed their spectral parameter using standard secant equation used by [11]. Spectral CG method combines CG search direction and scalar spectral parameter to form a new search direction, see [1,7,8,9,15,16,13,6,23,19,22,17,12,14], for more details.

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

(1)

where \( f: \mathbb{R}^n \to \mathbb{R} \) is continuous and differentiable, \( g_k \) is a gradient vector of the function and vector \( x_0 \in \mathbb{R}^n \) are solved using the expression (2)

\[
x_{k+1} = x_k + \sigma_k d_k, \quad k = 0,1,2,3,4, \ldots
\]  

(2)

and the vector \( x_k \) stand for a current iteration and \( x_{k+1} \) is a new iteration point, \( \sigma_k > 0 \) represent a step dimension obtained by the line search method given as

\[
\sigma_k = \arg \min_{\sigma > 0} f(x_k + \sigma d_k)
\]  

(3)

also \( d_k \) is a classical search direction.

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

(4)
\[ g_k = \nabla f(x), \text{ parameter } \beta_k \in R \text{ are the gradient vector and CG coefficient respectively. The } \beta_k \text{ of PRP is given by equation (5) } \]

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

where \( g_k \) and \( g_{k-1} \) are gradient vectors at points \( x_k, x_{k-1} \) respectively, and \( \| . \| \) is representing a Euclidian norm. The PRP Conjugate Gradient method is the best method but in general, the convergence analysis for a nonlinear function is uncertain [17].

In this work, spectral PRP Conjugate Gradient method is proposing without using the secant equation and verified its performance with least square method in regression analysis. The essential scheme for the analysis of a statistical data is the regression analysis and utilised in various area of knowledge, that includes engineering economics, and sciences. The analysis is use for forecasting what happen next, and comprehend the relation between dependent and independent variables in real life applications. The dependent variable is denoted by \( y \) and independent is denoted by \( x_j \) where \( j = 1, 2, 3, \ldots, n \) for \( n > 0 \) and \( e \) is an integer constant in the error term. The model is defined by

\[ y = l(x_j + e), \quad \text{for } x_j = x_1, x_2, \ldots, x_n \quad (6) \]

The general model can be defined as

\[ y = \rho_0 + \rho_1 x_1 + \rho_2 x_2 + \ldots + \rho_n x_n + e \quad (7) \]

where \( \rho_0, \rho_1, \rho_2, \ldots, \rho_n \) are the parameters for the regression analysis, the values of the parameters are estimate by using the nonlinear least square method defined by

\[ \min E(\rho) = \sum_{j=1}^{n} (y_i - \rho_0 + y_1 x_{j_1} + y_2 x_{j_2} + \ldots + y_n x_{j_n})^2 \quad (8) \]

where also \( y_i \) is the estimated data of \( i^{th} \) response and \( x_{j_1}, x_{j_2}, \ldots, x_{j_n} \) are \( n \) data estimation of the response variables. The general formula for predicting a data in regression analysis is by calculating the relative error. The error is obtaining by comparing the approximate value and actual value of the data, the formula is describing as follows

\[ \text{Relative error} = \frac{|\text{Exact Value} - \text{Approximate Value}|}{\text{Exact Value}} \quad (9) \]

The least square method determines the best approximation models by comparing the total least square errors acquired. The error is defined as

\[ E_j = (\rho_0 + \rho_1 x) - y_j \]

The approach for fitting the best line through the data would minimize the sum of the residual error squares for all the data available, this problem is the same as the minimization problem for the unconstrained optimization. The numerical optimization techniques to be use in this paper is spectral PRP Conjugate Gradient parameter to find the optimal solution to a real life unconstrained optimization problem.
2. DETAIL DERIVATION OF THE SPECTRAL PRP CONJUGATE GRADIENT METHOD

The spectral CG method was initially introduce by [11], the direction given is \( d_k = -\omega_k b_k + \beta_k z_{k-1} \), where \( z_{k-1} = \sigma_{k-1} d_{k-1} \) and \( \omega_k \) is a spectral scalar parameter. The propose search direction is defined as

\[
d_k = \begin{cases} 
-\omega_k b_k, & \text{if } k = 0 \\
-\omega_k b_k + \beta_k^{PRP} d_{k-1}, & \text{if } k \geq 1 
\end{cases}
\]  

(10)

From the search direction (10), \( d_k = -\omega_k b_k + \beta_k^{PRP} d_{k-1} \) then \( d_k - \beta_k^{PRP} d_{k-1} = -\omega_k b_k \) using the fact that \( d_k = -b_k \) and from equation (10) by substituting equation (5) we have,

\[
\omega_k = 1 - \frac{b_k^T d_{k-1}}{b_k^T b_{k-1}}
\]

(11)

Therefore, the orthogonality of gradients \( b_k^T b_{k-1} = 0 \) and thus \( \varphi_k \) is a new spectral parameter computed by exact line search procedure.

Algorithm 1 (Spectral PRP Conjugate Gradient Method)

Step 1: Given a starting point \( x_0 \in \mathbb{R}^n \) set \( k = 0 \)

Step 2: Compute \( \beta_k \) as given in formula (5) above

Step 3: Compute \( d_k \) given as in (10). If \( \|b_k\| = 0 \), then stop.

Step 4: Compute \( \sigma_k \) given in equation (3).

Step 5: Update the new point as given in the recurrence expression (2).

Step 6: If \( f(x_{k+1}) < f(x_k) \) and \( \|b_k\| < \varepsilon \) then stop, otherwise go to step 1 with \( k = k + 1 \).

The sufficient descent condition ensures that global convergence of iterative procedures or algorithm is achieve. Therefore, the following inequality must hold true.

\[
b_k^T d_k \leq -C \|b_k\|^2 \quad \text{for } k \geq 0 \text{ and } C > 0
\]

(12)

Theorem 1. Suppose a CG method with search direction (10) and \( \beta_k^{PRP} \) given by (5), the condition (12) holds \( \forall k \geq 0 \).

The assumptions below applied to the objective function for the analysis of global convergence properties of a general CG method.

Assumptions 1. (i) A level set \( \Omega = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\} \) is bounded, the function \( f \) is continuously differentiable in a neighbourhood \( N \) of the level set \( \Omega \) and \( x_0 \) is a starting point.

(ii) \( g(x) \) is globally Lipschitz continuous in \( N \) that is \( \exists \) a constant \( L > 0 \), such that \( \|g(x) - g(y)\| \leq L\|x - y\| \) for any \( x, y \in N \).

Lemma 1. Suppose the assumptions 1.1 hold and consider any recurrence expression (2) and search direction (10), \( y_k \) is found in equation (3). Then Zoutendijk condition below holds.

\[
\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty
\]

(13)

Proof of this Lemma is given in Zoutendijk, (1970).
Theorem 2. Suppose the assumptions 1.1 hold, for any \( \{x_k\}, \{d_k\} \) be given as spectral PRP CG method, \( \sigma_k \) are determined by equation (3) and \( \beta_k \) in equation (5). Then

\[
\lim_{k \to \infty} \|g_k\| = 0
\]  

(14)

Theorem 2 is true for a sufficient large \( k \). □

**Detail Description of the Application**

In this segment, the real life problem in Table 1 was obtain from the [15]. The approximate function for nonlinear least square method are formed as follows

\[
f(x) = -0.05690476x^2 + 0.68404762x + 0.13285714
\]

Thus, the function \( f(x) \) is use to approximate the value of \( y \) based on value of \( x \) that is the rate of Divorce within the city from year 2008 to 2016. The least square method can easily transform into unconstrained minimization problems as

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} f(x) &= \sum_{j=1}^{n} \left( (\rho_0 + \rho_1 x_j + \rho_2 x_j^2) - y_j \right)^2 \\
\end{align*}
\]

(15)

The data set in Table 1 shows the rate of Divorce among the young adults aged 18 to 26 in Jigawa state, Nigeria from the year 2008 until 2016. The statistical data collected from HISBAH Board Jigawa State, which conducted yearly in order to reduce the number of divorces that occurs frequently.

From the Table 1, the \( x \)-variable represent the year of the operation while the \( y \)-variable represent the rate of Divorce among the youth in the state. For the data fitting only the data from 2008 to 2015 is been considered, the data for the year 2016 is reserved for the error analysis.

**Table 1. Rate of Divorce in Jigawa state for the Year 2008 to 2016**

<table>
<thead>
<tr>
<th>Number (x)</th>
<th>Yrs.</th>
<th>Rate of Divorce (y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2008</td>
<td>78</td>
</tr>
<tr>
<td>2</td>
<td>2009</td>
<td>135</td>
</tr>
<tr>
<td>3</td>
<td>2010</td>
<td>159</td>
</tr>
<tr>
<td>4</td>
<td>2011</td>
<td>188</td>
</tr>
<tr>
<td>5</td>
<td>2012</td>
<td>195</td>
</tr>
<tr>
<td>6</td>
<td>2013</td>
<td>246</td>
</tr>
<tr>
<td>7</td>
<td>2014</td>
<td>226</td>
</tr>
<tr>
<td>8</td>
<td>2015</td>
<td>181</td>
</tr>
<tr>
<td>9</td>
<td>2016</td>
<td>183</td>
</tr>
</tbody>
</table>

Let the number of data \( x_j \) be the number of years and the value \( y_j \) be the rate of Divorce. The data from 2008 to 2015 are utilize to formulate the nonlinear quadratic model for the least square technique and the corresponding test function of unconstrained optimization problem.

Therefore, careful observation reveals that the data \( x_j \) and the value of \( y_j \) have parabolic relations with the regression function defined by equation (15) and the regression parameters \( \rho_0, \rho_1 \) and \( \rho_2 \).

\[
\begin{align*}
\min_{x \in \mathbb{R}^2} \sum_{j=1}^{n} E_j^2 &= \sum_{j=1}^{n} ((\rho_0 + \rho_1 x + \rho_2 x^2) - y_j)^2 \\
\end{align*}
\]

(16)
Equivalent transformation of the above least squares problem using the data of Table 1 for the nonlinear quadratic unconstrained minimization model given as follows

\[ f(\rho_0, \rho_1, \rho_2) = (8\rho_0 + 36\rho_1 + 204\rho_2 - 14.08)^2 \]  

(17)

The expression (17) is similar to (16) using the data given in Table 1. Therefore, by expanding equation (17) we equally have the following expression

\[ f(\rho_0, \rho_1, \rho_2) = 64\rho_0^2 + 1296\rho_1^2 + 41616\rho_2^2 + 576\rho_0\rho_1 + 3264\rho_0\rho_2 + 14688\rho_1\rho_2 - 225.28\rho_0 - 1013.76\rho_1 - 5744.64\rho_2 + 198.2464 \]  

(18)

The records for the year 2016 are not included from the unconstrained optimization function so that the data for the last year could be utilised to determine the accuracy of the function by computing the relative errors of the predicted data. Therefore, the proposed spectral PRP CG methods are applying to solve the test function by using exact line search technique. Table 2 and Table 3 shows the test results for spectral PRP Conjugate Gradient, HS and SCG methods for some selected initial point.

**Table 2.** Numerical Results for Spectral PRP, SCG and HS Methods based on CPU Time.

<table>
<thead>
<tr>
<th>Initial value</th>
<th>Spectral PRP</th>
<th>SCG</th>
<th>HS</th>
</tr>
</thead>
<tbody>
<tr>
<td>(11,11,11)</td>
<td>97.317</td>
<td>5.58880</td>
<td>0</td>
</tr>
<tr>
<td>(-1,0,-1)</td>
<td>41.3443</td>
<td>4.19763</td>
<td>0</td>
</tr>
<tr>
<td>(-5,-5,-5)</td>
<td>41.3119</td>
<td>14.6685</td>
<td>0.00063</td>
</tr>
<tr>
<td>(-2,-2,-2)</td>
<td>41.3556</td>
<td>4.55623</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 3.** Numerical Results for Spectral PRP, SCG and HS Methods based on Number of Iterations.

<table>
<thead>
<tr>
<th>Initial value</th>
<th>Spectral PRP</th>
<th>SCG</th>
<th>HS</th>
</tr>
</thead>
<tbody>
<tr>
<td>(11,11,11)</td>
<td>1000</td>
<td>2</td>
<td>NaN</td>
</tr>
<tr>
<td>(-1,0,-1)</td>
<td>10000</td>
<td>3</td>
<td>NaN</td>
</tr>
<tr>
<td>(-5,-5,-5)</td>
<td>10000</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>(-2,-2,-2)</td>
<td>1000</td>
<td>2</td>
<td>NaN</td>
</tr>
</tbody>
</table>

To elude matrix inverse in finding the values of \(\rho_0, \rho_1, \rho_2\) the spectral PRP, HS and SCG methods are utilised to solve the unconstrained optimization problem equation (17) using four different categories of initial points presented in the Table 2 and Table 3.

There are two condition of failure in this experiment; one is if the number of iterations is greater than one 10000 that is when the iterative process takes longer time to find the global minimum. The second condition is ‘NaN’ stand for not available based on numerical result. The approximation functions of the CG methods for the rate of Divorce in Jigawa state are summarised in Table 4.

**Table 4.** Approximation Functions for Different Initial Point

<table>
<thead>
<tr>
<th>Initial values</th>
<th>Methods</th>
<th>Approximate Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>(11, 11, 11)</td>
<td>Spectral PRP</td>
<td>[ y = -0.7754x^2 - 69.2073x + 11 ]</td>
</tr>
<tr>
<td></td>
<td>SCG</td>
<td>NaN</td>
</tr>
<tr>
<td></td>
<td>HS</td>
<td>NaN</td>
</tr>
<tr>
<td>(-1, 0, -1)</td>
<td>Spectral PRP</td>
<td>[ y = 0.2142975x^2 + 6.6407718x - 1 ]</td>
</tr>
<tr>
<td></td>
<td>SCG</td>
<td>NaN</td>
</tr>
<tr>
<td></td>
<td>HS</td>
<td>NaN</td>
</tr>
<tr>
<td>(-5, -5, -5)</td>
<td>Spectral PRP</td>
<td>[ y = 0.5243x^2 + 31.9303x - 5 ]</td>
</tr>
<tr>
<td></td>
<td>SCG</td>
<td>NaN</td>
</tr>
</tbody>
</table>
The rate of Divorce in Jigawa state Nigeria are estimated by least square method and the proposed spectral PRP Conjugate Gradient method for nonlinear unconstrained optimization model.

3. Numerical Results

Algorithm 1 have been tested and compared its performance with the recent spectral PRP (RSPRP) developed by [23] and classical PRP CG method [7]. The comparisons is on CPU time and number of iterations on each test function. The stopping criteria used for both methods is $\varepsilon = 10^{-6}$ as suggested by [20] and $\|g_k\| < \varepsilon$. A standard test problem functions in [14] were utilised with four different initial values using MatlabR2015 subroutine programming by Intel® Core™ i5-3317U (1.7GHz with 4 GB (RAM)).

The performances are establishing on CPU time running in seconds and number of iterations respectively. Table 5 shows the list of standard test problems with dimensions and initial points used to test the efficiency of the proposed spectral Conjugate Gradient method. The numerical performance of the proposed algorithms given in Figure 1 and Figure 2 based on a number of iterations and CPU time. The maximum value of the percentage of probability $P_s(t)$ and the solver that reached the solution point foremost are consider the best performing CG method for unconstrained optimization problems.

<table>
<thead>
<tr>
<th>Functions</th>
<th>Dimensions</th>
<th>Initial Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trecanni</td>
<td>2</td>
<td>(3,3), (4,4), (-12,-12), (-25,-25)</td>
</tr>
<tr>
<td>Leon</td>
<td>2</td>
<td>(3,3), (-4,-4), (5,5), (-10,-10)</td>
</tr>
<tr>
<td>Extended Penalty</td>
<td>2,4,10,50</td>
<td>(5,5), (-5,-5), (2,2), (-2,-2)</td>
</tr>
<tr>
<td>Power</td>
<td>2,4,50,100</td>
<td>(5,5), (-5,-5), (100,100), (-100,-100)</td>
</tr>
<tr>
<td>Quadratic QF1</td>
<td>10,100,1000, 10000</td>
<td>(5,5), (-5,-5), (100,100), (-100,-100)</td>
</tr>
<tr>
<td>Ext. Quadratic Penalty QP1</td>
<td>10,100</td>
<td>(8,8), (-8,-8), (5,5), (-5,-5)</td>
</tr>
<tr>
<td>Ext. Quadratic Penalty QP2</td>
<td>10,100</td>
<td>(2,2), (6,6), (8,8), (10,10)</td>
</tr>
<tr>
<td>Himmelblau</td>
<td>10000</td>
<td>(2,2), (-2,-2), (25,25), (-25,-25)</td>
</tr>
<tr>
<td>Freud. &amp; Roth</td>
<td>2,4,10,100,1000, 10000</td>
<td>(7,7), (11,11), (13,13), (25,25)</td>
</tr>
<tr>
<td>White and Holst</td>
<td>2,4,10,100,1000, 10000</td>
<td>(9,9), (-9,-9), (2,2), (5,5)</td>
</tr>
<tr>
<td>Shallow</td>
<td>2,4,10,100,1000, 10000</td>
<td>(200,200), (-200, -200), (-400,-400), (500,500)</td>
</tr>
<tr>
<td>Rosenbrock</td>
<td>2,4,10,100,1000, 10000</td>
<td>(20,20),(40,40),(5,5),(13,13)</td>
</tr>
</tbody>
</table>
Therefore, this work is passionate for the application of our proposed spectral PRP Conjugate Gradient method in comparison with least square method. Utilizing the nonlinear unconstrained optimization model in Table 4, the data for last year in 2016 are been estimated. Then, the relative errors of the data from each models have calculated by comparing the real data and the estimated data. Using equation (9) the efficiency of a particular method can easily be measure.

Table 6. Estimation and Relative Errors for 2016 Data

<table>
<thead>
<tr>
<th>Models</th>
<th>Estimation Point</th>
<th>Relative Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral PRP</td>
<td>1.130602014195</td>
<td>0.3821846916967</td>
</tr>
<tr>
<td>CG</td>
<td>NaN</td>
<td>NaN</td>
</tr>
<tr>
<td>SCG</td>
<td>NaN</td>
<td>NaN</td>
</tr>
<tr>
<td>HS</td>
<td>NaN</td>
<td>NaN</td>
</tr>
<tr>
<td>Least Square</td>
<td>0.1686095216</td>
<td>0.907863649398907</td>
</tr>
</tbody>
</table>
Entirely the computation of errors has made by Microsoft Excel 2016 and MATLAB 2015a subroutine programme. In the Table 6, the model that gives the smallest relative error are consider to be the best model that estimate the rate of Divorce in Jigawa state, Nigeria for the year 2016.

4. CONCLUSION

Sum of relative error for the proposed spectral PRP Conjugate Gradient method have computed based on four different set of the values, in each initial points three set of real numbers are chosen for nonlinear quadratic model. From the Table 6, the average relative error for the predicted data versus the actual data for all the methods are calculated. The relative error for the data generated from nonlinear quadratic models of spectral PRP Conjugate Gradient is smaller compare to least square model; the smallest relative error means the success of the spectral PRP Conjugate Gradient methods.

Conflicts of Interest

There is no any conflict of interest regarding the publication of this article.

References