A Hyperbolic Rational Model for Unconstrained Non-Linear Optimization

Nidhal H. Al-Assady Basim A. Hassan

basmabas39@gmail.com

College of Computer Sciences and Mathematics University of Mosul, Iraq

Received on: 07/09/2003 Accepted on: 16/12/2003

ABSTRACT

We consider a class of invariant Hyperbolic scaling of a strictly convex quadratic function, to extend the family of the conjugate gradient methods for solving unconstrained minimization problems. An algorithm is derived and evaluated numerically. The results indicate that, in general, the new algorithm is superior to the classical standard CG-algorithm.

Keywords: A Hyperbolic Rational Model, Conjugate gradient methods.

1. Introduction

A more general model than the quadratic one is proposed in this paper as a basis for a CG algorithm. If \( q(x) \) is a quadratic function, then a function \( f \) is defined as a non-linear scaling of \( q(x) \) if the following condition holds:

\[
f = F(q(x)), \quad \frac{dF}{dq} = F' > 0 \quad \text{and} \quad q(x) > 0 \quad \text{............... (1)}
\]

where \( x^* \) is the minimizer of \( q(x) \) with respect to \( x \) [15].

The following properties are immediately derived from the above condition:
i) Every contour line to \( q(x) \) is a contour line of \( f \).
ii) If \( x^* \) is a minimizer of \( q(x) \), then it is a minimizer of \( f \).
iii) That \( x^* \) is a global minimum of \( q(x) \) does not necessarily mean that it is a global minimum of \( f \) [7].

Various authors have published related work in the area:

A conjugate method which minimizers the function
\[
f(x) = (q(x))^\rho, \text{ and } x \in \mathbb{R}^n \text{ in at most steps has been described by Fried [11]}.\]

Another special case, namely
\[
F(q(x)) = \epsilon_1 q(x) + \frac{1}{2} \epsilon_2 q^2 (x)
\]

Where \( \epsilon_1 \) and \( \epsilon_2 \) are scalars, has been investigated by Boland and Kowalik [7].

Another model has been developed by Tassopoulos and Storey [16] as follows:
\[
F(q(x)) = \epsilon_1 q(x) + 1/\epsilon_2 q(x): \epsilon_2 > 0
\]

AL-Assady in [3] developed another model as follows:
\[
F(q(x)) = \ln(q(x))
\]

Al-Bayati [1] has been developed a new rational models which is defined as follows:
\[
F(q(x)) = \epsilon_1 q(x)/1-\epsilon_2 q(x), \quad \epsilon_2 < 0.
\]

Also Al-Bayati, [4] developed extended CG algorithm, which is based on a general logarithmic model
\[
F(q(x)) = \log(\epsilon q(x) - 1), \quad \epsilon > 0
\]

Al-Assady and Huda [2] described their ECG algorithm which is based on the natural log function for the rational \( q(x) \) function
\[
F(q(x)) = \log \left( \frac{\epsilon_1 q(x)}{\epsilon_2 q(x) + 1} \right), \quad \epsilon_2 < 0
\]

Al-Assady and Al-Taai [5] described their ECG algorithm which is based on the natural log function for the rational \( q(x) \) function
\[
F(q(x)) = \sinh(\epsilon q(x))
\]

And Al-Assady and Al-Taai [6] developed a new rational model which is defined as follows:
\[
F(q(x)) = \sin(\epsilon q(x))
\]

2. The New Non Quadratic Model:
In this paper, a new sine hyperbolic model is investigated and tested on a set of standard test function, assumed that condition (1) holds. An extended conjugate gradient algorithm is developed which is based on this new model which scales q(x) by the natural sinh function for the rational q(x) functions.

\[ F(q(x)) = \sinh \left( \frac{\varepsilon_1 q_1(x) + 1}{\varepsilon_2 q_2} \right) \]  

We first observe that q(x) and F(q(x)) given by (2) have identical contours, though with different function values, and they have the same unique minimum point denoted by \( x^* \).

2.1 The Algorithm:

Given \( x_0 \in \mathbb{R}^n \) an initial estimate of the minimizer \( x^* \).

Step (1): set \( d_0 = -g_0 \).

Step (2): For \( i = 1, 2, \ldots \)

\[ x_i = x_{i-1} + \lambda_{i-1} d_{i-1} \]

Where \( \lambda_{i-1} \) is the optimal step size obtained by the line search procedure.

Step (3): compute

\[ p_i = \frac{\left[ f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right]^2 + 1}{f_{i-1} + \sqrt{f_{i-1}^2 + 1}} \left[ \ln \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} - \frac{\varepsilon_1}{\varepsilon_2} \right) \right]^2 \]

Where the derivation of scaling \( p_i \) will be presented below.

Step (4): calculate the new direction

\[ d_i = -g_i + \beta_i d_i \]

where \( \beta_i \) is defined by different formulae according to variation and it is expressed as follows:
\[ \beta_i = \rho_i \left( \frac{\|g_i\|^2}{\|g_{i-1}\|^2} \right) \text{(modified by Fletcher and Reeves, 1964 F/R, [10])} \]

\[ \beta_i = g_i^T (\rho_i g_i - g_{i-1}) d_{i-1}^T (\rho_i g_i - g_{i-1}) \text{(modified by Hestenes and Stiefel 1952, H/S, [12])} \]

\[ \beta_i = g_i^T (\rho_i g_i - g_{i-1}) d_{i-1}^T g_{i-1} \text{(modified by Polak and Ribiera 1969, [13])} \]

\[ \beta_i = \rho_i \left\| g_{i-1} \right\|^2 d_i^T g_i \text{ (modified by Dixon 1972[9])} \]

Conjugate gradient methods are usually implemented by restarts in order to avoid an accumulation of errors affecting the search directions.

It is therefore generally agreed that restarting is very helpful in practice, so we have used the following restarting criterion in our practical investigations. If the new direction satisfies:

\[ d_i^T g_i \geq -0.8 \left\| g_i \right\|^2 \]

Then a restart is also initiated. This new direction is sufficiently downhill. [14].

2.2 The Derivation of \( \rho_i \) for the New Model:

The implementation of the extended CG method has been performed for general function \( F(q(x)) \) of the form of equation (2).

The unknown quantities \( \rho_i \) were expressed in terms of available quantities of the algorithm.

The new \( \sinh \left( \frac{\varepsilon_i q(x) + 1}{\varepsilon_2 q(x)} \right) \) model can now be written as:

\[ f(x) = F(q(x)) = \sinh \left( \frac{\varepsilon_i q(x) + 1}{\varepsilon_2 q(x)} \right) \]

Solving equation (2) for \( q \)

\[ \text{Sinh}^{-1} f(x) = \left( \frac{\varepsilon_i q(x) + 1}{\varepsilon_2 q(x)} \right) \]
\[
\ln \left[ f(x) + \sqrt{f(x)^2 + 1} \right] = \frac{\varepsilon_2 q(x) + 1}{\varepsilon_2 q(x)} \Rightarrow 
\]

\[
\Rightarrow q = \frac{1}{\varepsilon_2} \ln \left[ f(x) + \sqrt{f(x)^2 + 1} \right] - \varepsilon_i 
\]

and using the expression for \( \rho_i = \frac{f_i'}{f_i'} + \frac{f_i'}{f_i'} \)

\[
\rho_i = \frac{\cosh(\varepsilon_i q_{i-1} + 1/\varepsilon_i q_{i-1})}{\cosh(\varepsilon_i q_{i-1} + 1/\varepsilon_i q_{i-1})} \left( -\frac{1}{\varepsilon_i q_{i-1}^2} \right) 
\]

from the above equation we have

\[
\rho_i = \frac{\left[ \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) + 1 \right] \ln \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \frac{\varepsilon_1}{\varepsilon_2}^2}{f_{i-1} + \sqrt{f_{i-1}^2 + 1} \left[ \left( f_i + \sqrt{f_i^2 + 1} \right) + 1 \right] \ln \left( f_i + \sqrt{f_i^2 + 1} \right) - \frac{\varepsilon_1}{\varepsilon_2}^2} 
\]

In terms of the known quantities such a function and gradient values, from

\[
g_i = F_i Q(x_i - x^*) 
\]

\[
g_{i-1} = F_{i-1} Q(x_{i-1} - x^*) 
\]

Where Q is the Hessian Matrix and \( x^* \) is the minimum point, we have:

\[
\rho_i = \frac{\left[ \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) + 1 \right] \ln \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \frac{\varepsilon_1}{\varepsilon_2}^2}{f_{i-1} + \sqrt{f_{i-1}^2 + 1} \left[ \left( f_i + \sqrt{f_i^2 + 1} \right) + 1 \right] \ln \left( f_i + \sqrt{f_i^2 + 1} \right) - \frac{\varepsilon_1}{\varepsilon_2}^2} 
\]
Furthermore

Since

Therefore, we can express $\rho_i$ as follows:

$$\rho_i = \frac{g_{i-1}^T (x_{i-1} + \lambda d_{i-1} - x^*)}{g_i^T (x - x^*)} \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (5)$$

$$g_i^T (x_i - x^*) = g_i^T (x_i + \lambda_i d_i - x^*) = g_i^T (x_i - x^*)$$

From (4) and (5), it follows that:

$$\rho_i = \rho_i \left[ \frac{q_{i-1}}{q_i} \right] + \lambda_i g_i^T d_{i-1} / 2F_i q_i$$

Where

$$q = \sqrt{\frac{\ln(f + \sqrt{f^2 + 1}) - \frac{\epsilon_1}{\epsilon_2}}{\epsilon_2}}$$

and

$$t' = \frac{\left[ f + \sqrt{f^2 + 1} \right]^2 + 1 - \epsilon_2 \ln\left( f + \sqrt{f^2 + 1} - \frac{\epsilon_1}{\epsilon_2} \right)}{2 \left[ f + \sqrt{f^2 + 1} \right]}$$

The quantities $q_{i-1} / q_i$ and $F_i q_i$ can be rewritten as:

$$q_{i-1} = \frac{\ln\left( f_i + \sqrt{f_i^2 + 1} \right) - \frac{\epsilon_1}{\epsilon_2}}{\ln\left( f_i + \sqrt{f_i^2 + 1} \right) - \frac{\epsilon_1}{\epsilon_2}}$$

$$f_i q_i = \frac{\left[ f_i + \sqrt{f_i^2 + 1} \right]^2 + 1 - \ln\left( f_i + \sqrt{f_i^2 + 1} - \frac{\epsilon_1}{\epsilon_2} \right)}{2 \left[ f_i + \sqrt{f_i^2 + 1} \right]}$$
From the definition of $\rho_i$ we have:

\[
\frac{\left[ f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right]^2 + 1}{f_{i-1} + \sqrt{f_{i-1}^2 + 1}} \ln \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \frac{\varepsilon_1}{\varepsilon_2} = 0
\]

Using the following transformation:

\[
\frac{f_i + \sqrt{f_i^2 + 1}}{f_{i-1} + \sqrt{f_{i-1}^2 + 1}} = x, \quad \ln \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \frac{\varepsilon_1}{\varepsilon_2} = y
\]

\[
\ln \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \frac{\varepsilon_1}{\varepsilon_2} = y + w \quad \text{and} \quad \ln \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \ln \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) = w
\]

\[
c = \lambda_{i-1} g_{i-1} d_{i-1}
\]

then \( y = cw/xw + c \)

Therefore

\[
\frac{\varepsilon_1}{\varepsilon_2} = \ln \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \frac{\ln \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \ln \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \lambda_{i-1} g_{i-1} d_{i-1}}{\ln \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \ln \left( f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \lambda_{i-1} g_{i-1} d_{i-1}}
\]

3. Numerical Results and conclusion:

In order to test the effectiveness of the new algorithm which has been used to extent the CG method, a number of test functions have been chosen and solved numerically by utilizing the new and established method.

The same line search was employed for all the methods. This was the cubic interpolation procedure described in [8].
It is found that the NEW method which modifies CG-algorithm is better than the standard algorithm shown in Tables (1) and (2). The new method gives an overall improved over the classical CG Algorithms in the 16 and cases out of 20, respectively. The new method can therefore be considered promising.

Table (1), which uses the H/S formula, presents a comparison between the results of the NEW method and the classical CG-method. So we can show that the NEW method has less (NOI) and (NOF) than the classical CG. Method and NEW method improves the two measures of performances, vis (NOI) and (NOF) (81.46)% and the (79.49) % for the H/S formula.

Table (1) the comparison between the different ECG – method by using H/S formula.

<table>
<thead>
<tr>
<th>Test Function</th>
<th>N</th>
<th>New NOI (NOF)</th>
<th>Classical CG NOI (NOF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUBIC</td>
<td>2</td>
<td>16 (46)</td>
<td>19 (53)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>14 (36)</td>
<td>14 (40)</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>13 (33)</td>
<td>14 (40)</td>
</tr>
<tr>
<td>ROSEN</td>
<td>2</td>
<td>35 (91)</td>
<td>34 (87)</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>18 (55)</td>
<td>17 (52)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>20 (59)</td>
<td>17 (52)</td>
</tr>
<tr>
<td>POWELL</td>
<td>10</td>
<td>25 (64)</td>
<td>35 (89)</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>65 (173)</td>
<td>125 (303)</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>374 (754)</td>
<td>401 (860)</td>
</tr>
<tr>
<td>WOOD</td>
<td>4</td>
<td>27 (61)</td>
<td>26 (60)</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>47 (100)</td>
<td>59 (126)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>54 (114)</td>
<td>103 (213)</td>
</tr>
<tr>
<td>MIELE</td>
<td>20</td>
<td>41 (105)</td>
<td>54 (141)</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>72 (173)</td>
<td>82 (197)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>102 (253)</td>
<td>142 (346)</td>
</tr>
<tr>
<td>CANTRAL</td>
<td>10</td>
<td>21 (115)</td>
<td>20 (135)</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>20 (112)</td>
<td>20 (132)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>18 (101)</td>
<td>20 (132)</td>
</tr>
<tr>
<td>Non Diagonal</td>
<td>40</td>
<td>16 (45)</td>
<td>22 (73)</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>16 (46)</td>
<td>22 (59)</td>
</tr>
</tbody>
</table>
Table (2), which uses the P/R formula, presents a comparison between the results of the NEW method and the classical CG method. So we can show that the NEW method has less (NOI) and (NOF) than the classical CG method and NEW method improves the two measures of performances, vis (NOI) and (NOF) by (73.99)% and the (80.70)% for the P/R formula.

<table>
<thead>
<tr>
<th>Test Function</th>
<th>N</th>
<th>New NOI (NOF)</th>
<th>Classical CG NOI (NOF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUBIC</td>
<td>100</td>
<td>18 (34)</td>
<td>15 (40)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>14 (39)</td>
<td>15 (40)</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>13 (33)</td>
<td>15 (40)</td>
</tr>
<tr>
<td>ROSEN</td>
<td>10</td>
<td>24 (64)</td>
<td>26 (68)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>18 (56)</td>
<td>22 (61)</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>17 (53)</td>
<td>22 (61)</td>
</tr>
<tr>
<td>CANTRAL</td>
<td>10</td>
<td>20 (111)</td>
<td>22 (103)</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>29 (193)</td>
<td>18 (103)</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>23 (160)</td>
<td>22 (157)</td>
</tr>
<tr>
<td>WOOD</td>
<td>4</td>
<td>27 (61)</td>
<td>33 (74)</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>61 (128)</td>
<td>68 (144)</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>83 (172)</td>
<td>85 (178)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>105 (217)</td>
<td>108 (213)</td>
</tr>
<tr>
<td>POWELL</td>
<td>40</td>
<td>52 (128)</td>
<td>71 (155)</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>49 (107)</td>
<td>118 (255)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>105 (240)</td>
<td>119 (252)</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>76 (163)</td>
<td>205 (427)</td>
</tr>
<tr>
<td>Non</td>
<td>60</td>
<td>17 (49)</td>
<td>18 (53)</td>
</tr>
</tbody>
</table>
Appendix

1. Cubic Function:

   \[ F(x) = 100 \left( x_2 - x_1^2 \right)^2 + \left( 1 - x_1 \right)^2, \quad x_0 = (-1.2, -1)^T \]

2. Non-Diagonal Variant of Rosenbrock Function:

   \[ F(x) = \sum_{i=2}^{n} \left[ 100 \left( x_i - x_i^2 \right)^2 + \left( 1 - x_i \right)^2 \right], \quad n > 1, \]

3. Wood Function

   \[ F(x) = \sum_{i=1}^{n/2} \left[ (x_{4i-2} + x_{4i-3})^2 + (1 - x_{4i-3})^2 + 90(x_{4i-1} - x_{3i-1})^2 + (1 - x_{4i-1})^2 + 10 \left( x_{4i-2} - 1 \right)^2 
   + (x_{4i-1})^2 + 19.2 (x_{4i-2} - 1) (x_{4i-1}) \right] \]

   \[ x_0 = (-3.0; 1.0; -3.0; 1.0; \ldots) \]

4. Generalized Powell Quartics Functions:

   \[ F(x) = \sum_{i=1}^{n/2} \left[ (x_{4i-3} + 10 x_{4i-2})^2 + 5 (x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2 x_{4i-1})^4 + 10 (x_{4i-3} - x_{4i})^4 \right] \]

   \[ x_0 = (3.0; -1.0; 0.0; 1.0)^T \]

5. Rosenbrock Function:

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

   \[ x_0 = (-1.2; 1.0; \ldots)^T \]

6. Miele Function:
A Hyperbolic Rational...

\[ F(x) = \sum_{i=1}^{n/4} \left[ \exp(x_{4i-3} - x_{4i-2})^4 + 100(x_{4i-2} - x_{4i-1})^6 + \tan(x_{4i-1} - x_{4i})^4 + x_{4i-3}^8 \right] \]

\[ x_0 = (1.0; 2.0; 2.0; 2.0; \ldots)^T \]

7. Central Function:

\[ F(x) = \sum_{i=1}^{n/4} \left[ \exp(x_{4i-3} - x_{4i-2})^4 + 100(x_{4i-2} - x_{4i-1})^6 + \tan(x_{4i-1} - x_{4i})^4 + x_{4i-3}^8 \right] \]

\[ x_0 = (1.0; 2.0; 2.0; 2.0; \ldots)^T \]

REFERENCES


21


