A Rational Triangle Function as a Model for a Conjugate Gradient Optimization Method.

Abbas Y. Al-Bayati       Basim A. Hassan
profabbasalbayati@yahoo.com basimabas39@gmail.com
College of Computer Sciences and Mathematics
University of Mosul, Iraq

Received on: 12/10/2003 Accepted on: 22/03/2004

ABSTRACT

This paper presents the development and implementation of a new numerical based on a non-quadratic Triangular rational function model. For solving non-linear optimization problem. The algorithm is implemented in one version, employing exact line search. This version is compared numerically against versions of the CG-method. The results indicate that in general the new algorithm is superior to the previous algorithm.

Keywords: Non-quadratic Triangular rational function model, Numerical experiments.

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:

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

where \( x^* \) is the minimizer of q(x) with respect to x [13].
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 [5].

**Various authors have published-related work in the area:**

A conjugate method which minimizes the function 
\[ f(x) = (q(x))^p, \quad x \in \mathbb{R}^n \] 
in at most step has been described by Fried[9].

Another special case, namely 
\[ F(q(x)) = e_1 q(x) + e_2 q^2(x) \] 
Where e_1 and e_2 are scalars, has been investigated by Boland et al, [5].

Another model has been developed by Tassopoulos and Storey, [14] as follows: 
\[ F(q(x)) = e_1 q(x) + 1/e_2 q(x); \quad e_2 > 0 \]

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

Al-Bayat, [1] has developed a new rational model which is defined as follows: 
\[ F(q(x)) = e_1 q(x) / (1-e_2 q(x)) \]

Also Al-Bayati [4] developed an extended CG algorithm which is based on a general logarithmic model 
\[ F(q(x)) = \log(e_1 q(x) - 1), e > 0 \]

And Al-Assady, [2] described there ECG algorithm which is based on the natural log function for the rational q(x) function

\[ F(q) = \log \left[ \frac{e_1 q(x)}{e_2 q(x) + 1} \right], \quad e_2 < 0 \]

In this paper, a new sine model is investigated and tested on a set of standard test function, on the 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.

\[ F(q(x)) = \sin(e_1 q(x) / e_2 q(x) + 1) \quad \ldots \ldots \ldots (2) \]

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. Theorem**

Given an identical starting point x_1, the method of Fletcher and Reeves [8] defined by
\[ d_1 = -g_1 \]
\[ d_{i+1} = -g_{i+1} + \beta_id_i, \quad i \geq 1 \]

\[ \beta_i = \frac{\|g_{i+1}\|^2}{\|g_i\|^2} \]

and \( \| \| \) is the Euclidean norm applied to \( f(x)=q(x) \) and the ECG-method using the following search directions:

\[ d_1^- = -g_1^- \]
\[ d_{i+1}^- = -g_{i+1}^- + \rho_i \beta_id_i^-, \quad i \geq 1 \]

\[ \rho_i = \frac{f_i}{f_{i+1}} \]

\[ \beta_i = \frac{\|g_{i+1}\|^2}{\|g_i\|^2} \]

and applied to \( f(q(x)) \) generate identical conjugate directions (within a positive multiple \( f'_i \)) and the identical sequence of approximations \( x_i \) to the solution \( x^* \) for any function satisfying (1).

It is assumed that the one-dimensional searches are exact. The vectors \( \bar{g}_1, g_i^- \) are gradients of \( f(q(x)) \) at \( x_1 \) and \( x_i \), respectively.

**Proof:**

The theorem is true for \( i=1 \), because

\[ d_1^- = -g_1^- = -f_1'g_1 = f_1'd_1 \]

Now for \( i=2 \), we have

\[ d_2^- = -g_2^- + \rho_2\beta_2d_2^- \]

\[ = -f_2'g_2 + \left( \frac{f_2'}{f_1'} \right) \left( \frac{\|g_2\|^2}{\|g_1\|^2} \right) f_1'd_1 \]

\[ = -f_2'g_2 + \left( \frac{f_2'}{f_1'} \right) \left( \frac{f_1'}{f_1'} \right)^2 \left( \frac{\|g_2\|^2}{\|g_1\|^2} \right) f_1'd_1 \]

\[ = f_2'd_2. \]

Assume that, for \( i \geq 2 \),
\[ d_i = f'_i \left[ -g_{i+1} + \left( \frac{\|g_{i+1}\|^2}{\|g_i\|^2} \right) d_i \right] \]
\[ = f'_i d_i \]

It follows from (4) that
\[ d_{i+1} = -g_{i+1} + \rho_i \beta_i d_i \]
\[ = -f'_{i+1} g_{i+1} + \left( \frac{f'_{i+1}}{f'_i} \right) \left( \frac{f'_i}{\|g_i\|^2} \right)^2 \left( \frac{\|g_{i+1}\|^2}{\|g_i\|^2} \right) f'_i d_i \]
\[ = -f'_{i+1} d_{i+1} \]

Both methods generate the same sequence of approximations \( x_1 \), since isocontour curve of \( q(x) \) and \( f(q(x)) \) are identical. These isocurves differ only by the function values on the corresponding curves, and hence the theorem is proved.

3. 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 equations (2).

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

The new \( \sin \left( \frac{e_1 q(x) + 1}{e_2 q(x)} \right) \) model can now be written as

\[ f(x) = F(q(x)) = \sin \left( \frac{e_1 q(x) + 1}{e_2 q(x)} \right) \]

Solving equation (2) for \( q \)

\[ \sin^{-1} f(x) \left( \frac{e_1 q(x) + 1}{e_2 q(x)} \right) \]

\[ \ln \left[ f(x) + \sqrt{1 - f(x)^2} \right] = \frac{e_1 q(x) + 1}{e_2 q(x)} \quad \Rightarrow q = \frac{1}{e_1 \ln[f(x) + \sqrt{1 - f(x)^2}]} e_i \]

And using the expression for \( p_i = f'_i \cdot 1 / f'_i \)
\[
\rho_i = -\frac{\cos(\varepsilon_i q_{i-1} + 1/\varepsilon_2 q_{i-1}) \left(\frac{-1}{\varepsilon_2 q_{i-1}^2}\right)}{\cos(\varepsilon_i q_i + 1/\varepsilon_2 q_i) \left(\frac{-1}{\varepsilon_2 q_i^2}\right)}
\]

from the above equation we have

\[
\rho_i = \frac{\left|\begin{array}{cc}
if_{i+1} + \sqrt{1-f_i^2} & +1 \\
\ln \left|\begin{array}{c}
if_{i+1} + \sqrt{1-f_i^2} - \frac{\varepsilon_1}{\varepsilon_2}
\end{array}\right|
\end{array}\right|^2}{\left|\begin{array}{cc}
if_{i+1} + \sqrt{1-f_i^2} & +1 \\
\ln \left|\begin{array}{c}
if_{i+1} + \sqrt{1-f_i^2} - \frac{\varepsilon_1}{\varepsilon_2}
\end{array}\right|
\end{array}\right|^2}
\]

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

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

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

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

\[
\rho_i = \frac{\left|\begin{array}{cc}
if_{i+1} + \sqrt{1-f_i^2} & +1 \\
\ln \left|\begin{array}{c}
if_{i+1} + \sqrt{1-f_i^2} - \frac{\varepsilon_1}{\varepsilon_2}
\end{array}\right|
\end{array}\right|^2}{\left|\begin{array}{cc}
if_{i+1} + \sqrt{1-f_i^2} & +1 \\
\ln \left|\begin{array}{c}
if_{i+1} + \sqrt{1-f_i^2} - \frac{\varepsilon_1}{\varepsilon_2}
\end{array}\right|
\end{array}\right|^2}
\]

Furthermore

\[
g_{i+1}^T(x_i - x^*) = g_{i+1}^T(x_{i+1} + \lambda_{i+1} d_{i+1} - x^*)
\]

\[
= g_{i+1}^T(x_{i+1} - x^*) + \lambda_{i+1} g_{i+1}^T d_{i+1} \]

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

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

Since \( g_i^T d_{i-1} = 0 \) therefore, we can express \( \rho_i \) as follows:

\[
\rho_i = \frac{g_{i+1}^T(x_{i+1} + \lambda_{i+1} d_{i+1} - x^*)}{g_i^T(x_i - x^*)}
\]
From (7) and (8), it follows that:

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

Where

\[
q = \frac{1}{\varepsilon_2} \left[ \ln \left( if + \sqrt{1 - f^2} \right) - \frac{\varepsilon_1}{\varepsilon_2} \right]
\]

and

\[
T = \frac{\left[ if + \sqrt{1 - f^2} \right]^2 + 1 - \varepsilon_2 \left[ \ln \left( if + \sqrt{1 - f^2} \right) - \frac{\varepsilon_1}{\varepsilon_2} \right]^2}{2 \left[ if + \sqrt{1 - f^2} \right]}
\]

The quantities \( q_{i-1}/q_i \) and \( f_i q_i \) can be rewritten as:

\[
q_{i-1} = \frac{\ln \left[ if_i + \sqrt{1 - f_i^2} \right]}{q_i} - \frac{\varepsilon_1}{\varepsilon_2}
\]

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

From the definition of \( \rho_i \) we have:

\[
\frac{\left[ if_{i-1} + \sqrt{1 - f_{i-1}^2} \right]^2 + 1}{\left[ if_i + \sqrt{1 - f_i^2} \right]^2 + 1} \ln \left( if_{i-1} + \sqrt{1 - f_{i-1}^2} \right) - \frac{\varepsilon_1}{\varepsilon_2}^2
\]
\[
\begin{align*}
&\left[\left[if_{i+1} + \sqrt{1 - f_{i+1}^2}\right]^2 + 1\right]\frac{\ln\left(if_{i+1} + \sqrt{1 - f_{i+1}^2}\right) - \frac{\epsilon_1}{\epsilon_2}}{if_{i+1} + \sqrt{1 - f_{i+1}^2}} \\
&\frac{\ln\left(if_{i+1} + \sqrt{1 - f_{i+1}^2}\right) - \frac{\epsilon_1}{\epsilon_2}}{if_{i+1} + \sqrt{1 - f_{i+1}^2}} \\
&\left[if_{i+1} + \sqrt{1 - f_{i+1}^2}\right] + 1\right]\frac{\ln\left(if_{i+1} + \sqrt{1 - f_{i+1}^2}\right) - \frac{\epsilon_1}{\epsilon_2}}{if_{i+1} + \sqrt{1 - f_{i+1}^2}}
\end{align*}
\]

Using the following transformation:

\[
\begin{align*}
\frac{\left[if_{i+1} + \sqrt{1 - f_{i+1}^2}\right]^2 + 1}{if_{i+1} + \sqrt{1 - f_{i+1}^2}} &= x, \\
\ln\left(if_{i+1} + \sqrt{1 - f_{i+1}^2}\right) - \frac{\epsilon_1}{\epsilon_2} &= y
\end{align*}
\]

\[
\ln\left(if_{i+1} + \sqrt{1 - f_{i+1}^2}\right) - \frac{\epsilon_1}{\epsilon_2} = y + w \quad \text{and} \quad \ln\left(if_{i+1} + \sqrt{1 - f_{i+1}^2}\right) - \ln\left(if_{i+1} + \sqrt{1 - f_{i+1}^2}\right) = w
\]

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

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

Therefore

\[
\frac{\epsilon_1}{\epsilon_2} = \ln\left(if_{i+1} + \sqrt{1 - f_{i+1}^2}\right) - \frac{\ln\left(if_{i+1} + \sqrt{1 - f_{i+1}^2}\right) - \ln\left(if_{i+1} + \sqrt{1 - f_{i+1}^2}\right) - \frac{\lambda_{i-1}g_{i-1}^Td_{i-1}}{if_{i+1} + \sqrt{1 - f_{i+1}^2}}}{if_{i+1} + \sqrt{1 - f_{i+1}^2}}
\]

4. The Outlines of our New Algorithm Area:

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

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

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

Compute \( x_i = x_{i-1} + \lambda_{i-1}d_{i-1} \)

49
Where $$\lambda_{i-1}$$ is the optimal step size obtained by the line search procedure.

Step (3) : compute

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

Where the derivation of scaling $$\rho_i$$ 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:

$$\beta_i = \rho_i \left( ||g_i||^2 / ||g_{i-1}||^2 \right)$$ \{modified Fletcher and Reeves, 1964 F/R,[8]\}

$$\beta_i = g_i^T \left( \rho_i g_i - g_{i-1} \right) / d_{i-1}^T \left( \rho_i g_i - g_{i-1} \right)$$ \{modified Hestenes an stieffe 1952, H/s[10]\}

$$\beta_i = g_i^T \left( \rho_i g_i - g_{i-1} \right) / d_{i-1}^T g_{i-1}$$ \{modified Polak and Ribiera 1969,[11]\}

$$\beta_i = \rho_i ||g_{i+1}||^2 / d_i^T g_i$$ \{modified Dixon 1972,[7]\}

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 practices, so we have used the following restarting criterion in our practical investigations. If the new direction satisfies:

$$d_i^T g_i \geq -0.8 \|g_i\|^2$$

Then a restart is also initiated. This new direction is sufficiently downhill in Powell [12].

5. The Numerical Experiments:
In order to test the effectiveness of the new algorithm that have used to extend the CG method, a number of 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 Bunday [6].

It is found that the NEW method which modifies CG-algorithm is better than the previous algorithm shown in Tables (1) and (2).

Table (1) which uses the H/S formula, presents a comparison between the results of the NEW methods 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 improve the two measures of performances, vis (NOI) and (NOF) (56.60)% and the (60.16) % for the H/S formula.

Table (1): Comparison between the different ECG – methods by using H/S formula.

| Test Function | N  | New NOI (NOF) | Classical CG NOI (NOF) |
|---------------|----|---------------|------------------------|
| CUBIC         | 2  | 18 (51)       | 19 (53)                |
|               | 200| 12 (35)       | 14 (40)                |
|               | 400| 13 (32)       | 14 (40)                |
| ROSEN         | 2  | 31 (82)       | 34 (87)                |
|               | 10 | 21 (63)       | 26 (71)                |
|               | 100| 19 (56)       | 17 (52)                |
| POWELL        | 60 | 48 (102)      | 125 (303)              |
|               | 80 | 91 (203)      | 112 (303)              |
|               | 400| 221 (537)     | 401 (860)              |
| Non Diagonal  | 40 | 16 (44)       | 22 (73)                |
|               | 60 | 17 (47)       | 22 (61)                |
|               | 100| 16 (46)       | 22 (60)                |
| MIELE         | 40 | 50 (124)      | 82 (197)               |
|               | 200| 147 (338)     | 211 (491)              |
|               | 400| 142 (324)     | 402 (910)              |
| CANTRAL       | 4  | 18 (113)      | 25 (148)               |
|               | 40 | 19 (129)      | 20 (132)               |
|               | 400| 14 (71)       | 20 (132)               |
| SHALLOW       | 40 | 9 (21)        | 9 (20)                 |
| Total         | NOI (NOF) | 930 (2439)   | 1606 (4054)            |

Table (2) which uses the P/R formula, presents a comparison between the results of the NEW methods 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 improve the two measures of performances, vis (NOI) and (NOF) by (49.22)% and the (53.71) % for the P/R formula.

Table (2): Comparison between the different ECG – methods by using P/R formula.

| Test Function | N  | New NOI (NOF) | Classical CG NOI (NOF) |
|---------------|----|---------------|------------------------|
| CUBIC         | 2  | 18 (51)       | 19 (53)                |
|               | 200| 12 (33)       | 15 (40)                |
|               | 400| 11 (32)       | 15 (40)                |
| ROSEN         | 2  | 31 (82)       | 33 (53)                |
|               | 200| 18 (53)       | 22 (61)                |
|               | 400| 18 (54)       | 22 (61)                |
| POWELL        | 80 | 52 (117)      | 118 (255)              |
|               | 200| 117 (240)     | 205 (427)              |
|               | 400| 52 (112)      | 405 (826)              |
| Non Diagonal  | 60 | 17 (49)       | 18 (53)                |
|               | 80 | 15 (43)       | 25 (70)                |
|               | 100| 17 (47)       | 22 (62)                |
| MIELE         | 40 | 56 (155)      | 85 (238)               |
|               | 60 | 56 (133)      | 65 (189)               |
|               | 100| 39 (101)      | 71 (199)               |
| CANTRAL       | 4  | 23 (162)      | 25 (163)               |
|               | 10 | 19 (92)       | 22 (135)               |
|               | 400| 14 (72)       | 22 (157)               |
| SHALLOW       | 10 | 8(21)         | 8(19)                  |
|               | 400| 10(27)        | 8(19)                  |
| Total         | NOI (NOF) | 603 (1676) | 1225 (3120) |
A Rational Triangle Function…

APPENDIX

1. Cubic Function :
   \[ F(x) = 100(x_2 - x_1)^2 + (1 - x_1)^2, \quad x_0 = (-1.2, -1.0)^T \]

2. Non – Diagonal Variant of Rosenbrock Function :
   \[ F(x) = \sum_{i=2}^{n} \left[ 100(x_i - x_{i-1})^2 + (1 - x_{i-1})^2 \right], \quad n > 1, \]

3. SHALLOW Function
   \[ F(x) = \sum_{i=1}^{n} \left[ (x_{2i-1})^2 - (x_{2i})^2 + (1 - x_{2i-1}) \right] \]
   \[ x_0 = (-2.0, -2.0; \ldots)^T \]

4. Generalized Powell Quartics Functions :
   \[ F(x) = \sum_{i=1}^{n} \left[ x_{4i-3} + 10(x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-3} - x_{4i})^4 \right] \]
   \[ x_0 = (3.0; -1.0; 0.0; 1.0)^T \]

5. Rosenbrock Function :
   \[ F(x) = \sum_{i=1}^{n} \left[ (x_{2i-1})^2 - (x_{2i})^2 + (1 - x_{2i-1}) \right] \]
   \[ x_0 = (-1.2; 1.0; \ldots)^T \]

6. Miele Function :
   \[ F(x) = \sum_{i=1}^{n} \left[ \exp(x_{4i-3}) \cdot x_{4i-2}^6 + 100(x_{4i-2} - x_{4i-1})^6 + \right. \]
   \[ \left. \left[ \tan(x_{4i-1} - x_{4i}) \right]^8 + x_{4i-3} \cdot (x_{4i-1})^2 \right], \]
   \[ x_0 = (1.0; 2.0; 2.0; 2.0; \ldots)^T \]

7. Cantral Function :
   \[ F(x) = \sum_{i=1}^{n} \left[ \exp(x_{4i-3}) \cdot x_{4i-2}^6 + 100(x_{4i-2} - x_{4i-1})^6 + \right. \]
   \[ \left. \left[ a \tan(x_{4i-1} - x_{4i}) \right]^8 + x_{4i-3} \cdot \right], \]
   \[ x_0 = (1.0; 2.0; 2.0; 2.0; \ldots)^T \]
REFERENCES

[1] Al – Bayati, A.Y (1993). New Non – Quadratic Model Unconstrained Non- linear optimization method”, Natural and Applied Series. Mutah Journal for Research and Studies. Mu’tah University, Jordan Vol 8, No.1, PP 133-155.

[2] Al-Assady N.H.; Huda, k.M (1997). “A rational logarithmic model for Unconstrained Nonlinear Optimization. Raf. Journal Sci, Vol 8, No 2, PP 107-117; Mosul University. Iraq.

[3] Al-Assady, N.H. (1991). “new QN and PCG algorithms Based on Non Quadratic Properties, J. Ed- and Sc., 12.

[4] Al-Bayati, A.Y., (1995). “New Extended CG. Methods for Non Linear Optimization, “Nataral and Applied Series, Mu’tah Journal for search and studies. Mu’tah university, Jordan Vol. 10, No. 6 PP 69 – 87.

[5] Boland, W. R and kowalik, J.S . (1979). “ Extended Conjugate Gradient method with Restarts”, Journal of Opt. Theory and Application, 28, PP. 1-9.

[6] Bunday, b., (1984). “Basic Optimization methods”, Edward Arnold, Bedford square, London.

[7] Dixon, L.C.W., (1972). “Non linear Optimization”, London, English University Press.

[8] Flecher, R. And Reeves, C.M., (1964) “Function Minimization by conjugate Gradient” Computer journal, 7, PP. 149-154.

[9] Fried, I., (1971). “N-step Conjugate Gradient Minimization Scheme for Non – Quadratic Function,” AIAA Journal, 19, PP. 2286-2287.

[10] Hestences, M.R. and stiefel, E, (1952), “Mechods of conjugate Gradient for solving Linear systems”, j. Res .B.S., 49.

[11] Polak, E. and Ribiere, G., (1969). Note Surla Convergence des methods de directions conjugate; Rev. Fr in fr, Rech Oper, 16, R1.

[12] Pedicato, E, (1976). “A variable Metric Method for Function Minimization Derived from Invariance and App. 1., 20.

[13] Tassopouls, A. and Storey, c., (1984), “Vse of Non Quadrate Model in a conjugate Gradient Model of Optimization with inexact line search”, Jota, 43.