

A Hyperbolic Rational Model for Unconstrained Non-Linear Optimization

Nidhal H. Al-Assady

Basim A. Hassan

College of Computer sciences and Mathematics
University of Mosul

Received on: 7/9/2003

Accepted on: 16/12/2003

الملخص

في هذا البحث تم بناء نموذج اكثر عمومية من النموذج التربيعي للأمتالية غير المشروطة وهذا النموذج يمثل القطع الزائد ويعتبر هذا النوع أحد الأصناف ذات المقياس الثابت للدالة التربيعية المحددة . لغرض توسيع وتطوير طرق المتجهات المترافقة في حل مسائل الامتالية غير المشروطة تم اشتقاق خوارزمية جديدة وقيمت عدديا . وان النتائج التي تم التوصل إليها أثبتت أن الخوارزمية الجديدة هي اكثر كفاءة من الخوارزمية القياسية المعرفة في هذا المجال .

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.

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 :

$$f = F(q(x)), dF/dq = F' > 0 \text{ and } q(x) > 0 \quad \dots\dots\dots (1)$$

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 minimizes the function

$f(x) = (q(x))^r$, and $x \in R^n$ in at most steps has been described by Fried [11]. Another special case, namely

$$F(q(x)) = e_1 q(x) + \frac{1}{2} e_2 q^2(x)$$

Where ϵ_1 and ϵ_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: $F(q(x)) = \epsilon_1 q(x)/1 - \epsilon_2 q(x)$, $e_2 < 0$.

Also Al-Bayati, [4] developed extended CG algorithm, which is based on a general logarithmic model

$$F(q(x)) = \log(eq(x) - 1), e > 0$$

Al-Assady and Huda [2] described their ECG algorithm which is based on the natural log function for the rational $q(x)$ function

$$F(q) = \log \left[\frac{e^{e_1 q(x)}}{e^{e_2 q(x)} + 1} \right], \epsilon_2 < 0$$

Al-Assady and Al-Ta'ai [5] described there 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-Ta'ai [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.

$$F(q(x)) = \sinh(\varepsilon_1 q_1(x) + 1/\varepsilon_2 q_2) \dots\dots\dots(2)$$

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^* .

2.1 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 $i = 1, 2, \dots$

$$\text{Compute } x_i = x_{i-1} + \lambda_{i-1} d_{i-1}$$

Where λ_{i-1} is the optimal step size obtained by the line search procedure.

Step (3) : compute

$$r_i = \left[\frac{\left[\left[f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right]^2 + 1 \right] \left[\ln \left(f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \frac{e_1}{e_2} \right]^2}{f_{i-1} + \sqrt{f_{i-1}^2 + 1}} \right. \\ \left. \frac{\left[\left[f_i + \sqrt{f_i^2 + 1} \right]^2 + 1 \right] \left[\ln \left(f_i + \sqrt{f_i^2 + 1} \right) - \frac{e_1}{e_2} \right]^2}{f_i + \sqrt{f_i^2 + 1}} \right]$$

Where the derivation of scaling p_i will be presented below.

Step (4) : calculate the new direction

$$d_i = -g_i + b_i d_{i-1}.$$

where b_i is defined by different formulae according to variation and it is expressed as follows:

$$b_i = r_i (\|g_i\|^2 / \|g_{i-1}\|^2) \{\text{modified by Fletcher and Reeves, 1964 F/R,[10]}\}$$

$$\left[b_i = g_i^T (r_i g_i - g_{i-1}) / d_{i-1}^T (r_i g_i - g_{i-1}) \right] \{\text{modified by Hestenes and Stiefle 1952, H/S,[12]}\}$$

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

$$b_i = r_i \|g_{i+1}\|^2 / d_i^T g_i \{\text{modified by Dixon 1972[9]}\}$$

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

$$\mathbf{d}_i^T \mathbf{g}_i \geq -0.8 \|\mathbf{g}_i\|^2$$

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

2.2 The Derivation of r_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 r_i were expressed in terms of available quantities of the algorithm.

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

$$f(x) = F(q(x)) = \sinh\left(\frac{e_1 q(x)+1}{e_2 q(x)}\right)$$

Solving equation (2) for q

$$\sinh^{-1} f(x) = \left(\frac{e_1 q(x)+1}{e_2 q(x)} \right)$$

$$\ln \left[f(x) + \sqrt{f(x)^2 + 1} \right] = \frac{e_1 q(x)+1}{e_2 q(x)} \quad \text{p}$$

$$\text{p } q = \frac{1}{e_2} \ln \left[f(x) + \sqrt{f(x)^2 + 1} \right] - e_1$$

and using the expression for $r_i = f'_{i-1} / f'_i$

$$r_i = - \frac{\cosh(e_1 q_{i-1} + 1/e_2 q_{i-1}) \left(-1/e_2 q_{i-1}^2 \right)}{\cosh(e_1 q_i + 1/e_2 q_i) \left(-1/e_2 q_i^2 \right)}.$$

from the above equation we have

$$r_i = \left[\frac{\left[\left[f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right]^2 + 1 \right] \left[\ln \left(f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \frac{e_1}{e_2} \right]^2}{f_{i-1} + \sqrt{f_{i-1}^2 + 1}}}{\left[\left[f_i + \sqrt{f_i^2 + 1} \right]^2 + 1 \right] \left[\ln \left(f_i + \sqrt{f_i^2 + 1} \right) - \frac{e_1}{e_2} \right]^2}{f_i + \sqrt{f_i^2 + 1}} \right] \dots\dots\dots(3)$$

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:

$$r_i = \left[\frac{\left[\left[f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right]^2 + 1 \right] \left[\ln \left(f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \frac{e_1}{e_2} \right]^2}{f_{i-1} + \sqrt{f_{i-1}^2 + 1}}}{\left[\left[f_i + \sqrt{f_i^2 + 1} \right]^2 + 1 \right] \left[\ln \left(f_i + \sqrt{f_i^2 + 1} \right) - \frac{e_1}{e_2} \right]^2}{f_i + \sqrt{f_i^2 + 1}} \right]$$

Furthermore

$$\begin{aligned} g_i^T (x_i - x^*) &= g_i^T (x_i + l_i d_i - x^*) \\ &= g_i^T (x_i - x^*) \end{aligned}$$

Since $g_i^T d_{i-1} = 0$

Therefore, we can express r_i as follows: -

$$r_i = \frac{g_{i-1}^T (x_{i-1} + l_{i-1} d_{i-1} - x^*)}{g_i^T (x - x^*)} \dots\dots\dots(5)$$

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

$$r_i = r_i \left[\frac{q_{i-1}}{q_i} \right] + l_{i-1} g_{i-1}^T d_{i-1} / 2 F_i' q_i$$

$$\text{Where } q = \frac{1}{e_2 \left[\ln(f + \sqrt{f^2 + 1}) - \frac{e_1}{e_2} \right]}$$

$$\text{and } f' = \frac{\left[\left[f + \sqrt{f^2 + 1} \right]^2 + 1 \right] - e_2 \left[\ln(f + \sqrt{f^2 + 1}) - \frac{e_1}{e_2} \right]^2}{2 \left[f + \sqrt{f^2 + 1} \right]}$$

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

$$\frac{q_{i-1}}{q_i} = \frac{\ln[f_i + \sqrt{f_i^2 + 1}] - \frac{e_1}{e_2}}{\ln[f_{i-1} + \sqrt{f_{i-1}^2 + 1}] - \frac{e_1}{e_2}}$$

$$f_i' q_i = \frac{\left[\left[f_i + \sqrt{f_i^2 + 1} \right]^2 + 1 \right] \left[\ln(f_i + \sqrt{f_i^2 + 1}) - \frac{e_1}{e_2} \right]}{2 \left[f_i + \sqrt{f_i^2 + 1} \right]}$$

From the definition of r_i we have:

$$\left[\frac{\left[\left[f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right]^2 + 1 \right] \left[\ln(f_{i-1} + \sqrt{f_{i-1}^2 + 1}) - \frac{e_1}{e_2} \right]^2}{\frac{f_{i-1} + \sqrt{f_{i-1}^2 + 1}}{\left[\left[f_i + \sqrt{f_i^2 + 1} \right]^2 + 1 \right] \left[\ln(f_i + \sqrt{f_i^2 + 1}) - \frac{e_1}{e_2} \right]^2}} \right] =$$

$$\left[\frac{f_i + \sqrt{f_i^2 + 1}}{\left[\left[f_i + \sqrt{f_i^2 + 1} \right]^2 + 1 \right] \left[\ln(f_i + \sqrt{f_i^2 + 1}) - \frac{e_1}{e_2} \right]^2} \right]$$

$$\left[\frac{\left[\left[f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right]^2 + 1 \right] \left[\ln \left(f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \frac{e_1}{e_2} \right]}{f_{i-1} + \sqrt{f_{i-1}^2 + 1}} \right] - \frac{(l_{i-1} g_{i-1}^T d_{i-1})}{\left[\left[f_i + \sqrt{f_i^2 + 1} \right]^2 + 1 \right] \left[\ln \left(f_i + \sqrt{f_i^2 + 1} \right) - \frac{e_1}{e_2} \right]} \frac{f_i + \sqrt{f_i^2 + 1}}{f_i + \sqrt{f_i^2 + 1}}$$

Using the following transformation:

$$\frac{\left[f_i + \sqrt{f_i^2 + 1} \right]^2 + 1}{f_i + \sqrt{f_i^2 + 1}} = x, \quad \ln \left[f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right] - \frac{e_1}{e_2} = y$$

$$\ln \left[f_i + \sqrt{f_i^2 + 1} \right] - \frac{e_1}{e_2} = y + w \quad \text{and} \quad \ln \left[f_i + \sqrt{f_i^2 + 1} \right] - \ln \left[f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right] = w$$

$$c = l_{i-1} g_{i-1}^T d_{i-1}$$

then $y = cw/xw + c$

Therefore

$$\frac{e_1}{e_2} = \ln \left[f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right] - \frac{\left[\ln \left(f_i + \sqrt{f_i^2 + 1} \right) - \ln \left(f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) \right] \left[-l_{i-1} g_{i-1}^T d_{i-1} \right]}{\left[\left[f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right]^2 + 1 \right] \left[\ln \left(f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) - \ln \left(f_{i-1} + \sqrt{f_{i-1}^2 + 1} \right) \right] + l_{i-1} g_{i-1}^T d_{i-1}}$$

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.

Test Function	N	New NOI (NOF)	Classical CG NOI (NOF)
CUBIC	2	16 (46)	19 (53)
	100	14 (36)	14 (40)
	400	13 (33)	14 (40)
ROSEN	2	35 (91)	34 (87)
	40	18 (55)	17 (52)
	100	20 (59)	17 (52)
POWELL	10	25 (64)	35 (89)
	60	65 (173)	125 (303)
	400	374 (754)	401 (860)
WOOD	4	27 (61)	26 (60)
	40	47 (100)	59 (126)
	100	54 (114)	103 (213)
MIELE	20	41 (105)	54 (141)
	40	72 (173)	82 (197)
	100	102 (253)	142 (346)
CANTRAL	10	21 (115)	20 (135)
	80	20 (112)	20 (132)
	100	18 (101)	20 (132)
Non Diagonal	40	16(45)	22(73)
	400	16(46)	22(59)
Total	NOI (NOF)	1015 (2536)	1246 (3190)

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 (2) the comparison between the different ECG – method by using P/R formula.

Test Function	N	New NOI (NOF)	Classical CG NOI (NOF)
CUBIC	100	18 (34)	15 (40)
	200	14 (39)	15 (40)
	400	13 (33)	15 (40)
ROSEN	10	24 (64)	26 (68)
	200	18 (56)	22 (61)
	400	17 (53)	22 (61)
CANTRAL	10	20 (111)	22 (103)
	60	29 (193)	18 (103)
	400	23 (160)	22 (157)
WOOD	4	27 (61)	33 (74)
	40	61 (128)	68 (144)
	80	83 (172)	85 (178)
	100	105 (217)	108 (213)
POWELL	40	52 (128)	71 (155)
	80	49 (107)	118 (255)
	100	105 (240)	119 (252)
	200	76 (163)	205 (427)
Non Diagonal	60	17 (49)	18 (53)
	200	13 (41)	25(68)
	400	15 (44)	24 (68)
Total	NOI (NOF)	774 (2093)	1046 (2592)

Appendix

1. Cubic Function :

$$F(\mathbf{x}) = 100(\mathbf{x}_2 - \mathbf{x}_1^3)^2 + (1 - \mathbf{x}_1)^2, \quad \mathbf{x}_0 = (-1.2, -1.)^T$$

2. Non – Diagonal Variant of Rosenbrock Function :

$$F(\mathbf{x}) = \sum_{i=2}^n \left(100(\mathbf{x}_i - \mathbf{x}_{i-1}^2)^2 + (1 - \mathbf{x}_{i-1})^2 \right), \quad \mathbf{n} > 1,$$

3. Wood Function

$$F(x) = \sum_{i=1}^{n/4} 100 \left(x_{4i-2} + x_{4i-3}^2 \right)^2 + \left(1 - x_{4i-3} \right)^2 + 90 \left(x_{4i} - x_{4i-1}^2 \right)^2 + \left(1 - x_{4i-1} \right)^2 + 10.1 \left(x_{4i-2} - 1 \right)^2 \\ + \left(x_{4i} - 1 \right)^2 + 19.8 \left(x_{4i-2} - 1 \right) \left(x_{4i} - 1 \right) \\ x_0 = (-3.0; 1.0; -3.0; -1.0; \dots)^\top$$

4. Generalized Powell Quartics Functions :

$$F(x) = \sum_{i=1}^{n/4} \left(x_{4i-3} + 10 x_{4i-2} \right)^2 + 5 \left(x_{4i-1} - x_{4i} \right)^2 + \left(x_{4i-2} - 2 x_{4i-1} \right)^4 + 10 \left(x_{4i-3} - x_{4i} \right)^4 \\ x_0 = (3.0; -1.0; 0.0; 1.0)^\top$$

5. Rosenbrock Function :

$$F(x) = \sum_{i=1}^{n/2} 100 \left(x_{2i} - x_{2i-1}^2 \right)^2 + \left(1 - x_{2i-1} \right)^2 \\ x_0 = (-1.2; 1.0; \dots)^\top$$

6. Miele Function :

$$F(x) = \sum_{i=1}^{n/4} \left[\exp \left(x_{4i-3} \right) - x_{4i-2} \right]^2 + 100 \left(x_{4i-2} - x_{4i-1} \right)^6 + \\ \left[\tan \left(x_{4i-1} - x_{4i} \right) \right]^4 + x_{4i-3}^8 + \left(x_{4i-1} \right)^2, \\ x_0 = (1.0; 2.0; 2.0; 2.0, \dots)^\top$$

7. Cantral Function :

$$F(x) = \sum_{i=1}^{n/4} \left[\exp \left(x_{4i-3} \right) - x_{4i-2} \right]^4 + 100 \left(x_{4i-2} - x_{4i-1} \right)^6 + \\ \left[a \tan \left(x_{4i-1} - x_{4i} \right) \right]^4 + x_{4i-3}^8. \\ x_0 = (1.0; 2.0; 2.0; 2.0, \dots)^\top$$

REFERENCES

- [1] Al – Bayati, A.Y. (1993) “New Non – Quadratic Model Unconstrained Non- linear optimization method”, Natural and Applied Series, Jordan: Ma’th University, **Mutah Journal for Research and Studies**. Vol. 8, No.1, PP 133-155.
- [2] Al-Assady N. H. and k. M. Huda, (1997) “A rational logarithmic model for Unconstrained Nonlinear Optimization”, Iraq: Mosul University, **Raf. Journal Sci.**, Vol. 8, No. 2, PP 107-117;.
- [3] Al- Assady, N. H. (1991) “New QN and PCG algorithms Based on Non Quadratic Properties”, Iraq: Mosul University, **J. Ed- and Sc.**, Vol.8, No.12, pp.130-140.
- [4] Al- Bayati, A.Y. (1995) “New Extended CG. Methods for Non Linear Optimization, Natural and Applied Series, Jordan: Ma’th University, **Mu’th Journal for search and studies**, Vol. 10, No. 6, pp. 69 –87.
- [5] Al-Assady, N. H. and B. A. Al-Ta’ai (2004) “A Hyperbolic Model for Unconstrained Non-Linear Optimization” , Iraq: **Raf. Jour. Sci.**, Vol.15, No.1, pp.109-116
- [6] Al-Assady , N. H. , and Al-Ta’ai B.A.(2004) “A New Non-Quadratic Model For Unconstrained Non-Linear Optimization” , Iraq: **Raf. Jour. of Comp. and math.**, Vol.1, No.2, pp.17-28
- [7] Boland, W. R and J.S. kowalik, (1979) “ Extended Conjugate Gradient method with Restarts”, **Journal of Optimization Theory and Application**, Vol.28, pp. 1-9.
- [8] Bunday, B. (1984) **Basic Optimization methods**, London: Edward Arnold.
- [9] Dixon, L.C.W. (1972) **Nonlinear Optimization**, London: English University Press.
- [10] Flecher, R. And C.M. Reeves, (1964) “Function Minimization by conjugate Gradient” **Computer journal**, Vol.7, pp. 149-154.
- [11] Fried, I. (1971) “N-step Conjugate Gradient Minimization Scheme for Non – Quadratic Function”, **AIAA Journal**, Vol.19, pp. 2286-2287.

- [12] Hestences, M.R. and E. stiefel, (1952) "Methods of conjugate Gradient for solving Linear systems", **j. Research National Bureau of Standards**, Vol.49, pp.409-456.
- [13] Polak, E. and G. Ribiere, (1969) "Note Surla Convergence des methods de directions conjugate", **Rev. Fr infr, Rech Oper**, 16, R1.
- [14] Powell, M.J.D. (1977) "Restart procedures for the conjugate gradient method", **Math. Programming**, Vol.12, pp.241-254.
- [15] Spedicato, E, (1976) **A variable Metric Method for Function Minimization**, Perived from Invariance and App 1., 20.
- [16] Tassopouls, A. and C. Storey, (1984) "Use of Non, Quadrate Model in A conjugate Gradient Model of Optimization with inexact line search", **JOTA Journal, Theory and Applications**, Vol.43, pp.353-730.