New Secant Hyperbolic Model for Conjugate Gradient Method

Abbas Al-Bayati **Baan Ahmed**

College of Computer Sciences and Mathematics
University of Mosul

Received on:28/6/2003 **Accepted on:13/12/2004**

الملخص

يتركز هذا البحث على اقتراح نموذج جديد اكثر عمومية من النماذج التربيعية لحل مسائل الامثلية اللاخطية وغير المقيدة والذي يطور خوارزميات التدرج المترافق الكلاسيكية وله نفس خواص خوارزميات التدرج المترافق المطبقة على دالة تربيعية. النموذج الجديد المستخدم في خوارزميات التدرج المترافق الكلاسيكية هو نموذج القاطع الزائدي الذي تم اشتقاقه وحله عدديا" باستخدام بعض الدوال غير الخطية المختارة لاختبار كفاءة الخوارزمية الجديدة. وبصورة عامة النتائج تشير إلى أن الخوارزمية الجديدة هي تحسين للخوارزميات السابقة.

ABSTRACT

New hyperbolic model different from the quadratic ones is proposed for solving unconstrained optimization problems which modify the classical conjugate gradient method. This new model was compared with established methods over a variety of standard non-linear test functions. The numerical results show that the use of non-quadratic model is beneficial in most of the problems considered especially when the dimensional of the problems increases.

1. Introduction:

Hestenes and Stiefel proposed a CG-method in 1952 for solving systems of linear equations. This method is used for unconstrained optimization, where CG methods are applied to quadratic functions. Fletcher and Reeves first did the extension of the CG method for solving nonlinear equation systems and its use in solving general unconstrained minimization problems in 1964.

The basis of all conjugate gradient (CG) methods is to determine a new direction of search using information related only to the gradient of a quadratic objective function. In such away the successive search directions are conjugate with respect to positive definite Hessian matrix G of the quadratic form. At each stage k the direction d_k is obtained by linear combination of the gradient $-g_k$ and the point x_k and the set direction

$d_0,\ldots, d_{k-1}$, such that d_k is conjugate to all previous directions based on the conjugacy coefficients. Therefore, the search direction d_{k+1} is obtained by the following rule:

$$d_{k+1} = -g_{k+1} + b_k d_k \qquad \ldots(1)$$

where b_k is a scalar.

(Bazarra, 2000).

2. Extension of the CG-Methods:

Among the existing conjugate direction algorithms the updating process rarely takes into account any deviation of the objective function from quadratic behaviour. However, quadratic models may not always be adequate to incorporate all the information which might be needed to represent the objective function *f(x)* successfully.

In order to improve the rate of convergence for the minimization algorithms when applied to non-quadratic method rather than quadratic, we introduce in this paper a new algorithm which is invariant to non-linear scaling of quadratic function.

If *q(x)* is a quadratic function then a function *f* is defined as a non-linear scaling *q(x)* if the following condition holds:

$f(x) = F(q(x))$, then $df / dq > 0$, for $x = x^*$

where x^* is the minimizer of *q(x)* (Spedicato, 1976).

Many authors have proposed a special models to determine r_k where r_k is defined as follows:

$$r_k = F'_{k-1} / F'_k$$

where $df / dq = F' > 0$, for $q > 0$ and *f* is an increasing monotonic function, which may better represent the objective function *f* and *q* is a quadratic function.

Fried (1971) has described a CG – method capable of minimizing the function $F(q) = (q(x))^p;\ p > 0,\ x\ \hat{I} R^n$, in at most *n* iterations.

Another special case, namely

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

where e_1 and e_2 are scalers had been investigated by Boland et. al. (1979).

Another model has been developed by Tassopoulos and Storey – T/S (1984), as follows:

$$F(q(x)) = \frac{(e_1 q(x) + 1)}{e_2 q(x)}, e_2 < 0 \qquad \ldots(2)$$

Al-Assady (1991) developed a model as follows:

$$F(q(x)) = \ln(q(x)) \qquad \ldots(3)$$

Al-Bayati (1993) has developed a new rational models which is defined as follows:

$$F(q(x)) = \frac{e_1 q(x)}{1 - e_2 q(x)}, \quad e_2 < 0 \qquad \ldots(4)$$

Also, Al-Bayati (1995) developed an extended conjugate gradient algorithm which is based on a general logarithmic model:

$$F(q(x)) = \log(e \;\; q(x)) - 1 \quad , e > 0 \qquad \ldots(5)$$

Al-Assady and Huda (1997) described the extended CG – algorithm which is based on the natural log function for the rational q(x) function

$$F(q(x)) = Log\left[\frac{e_1 q(x)}{e_2 q(x) + 1}\right], e_2 < 0 \qquad \ldots(6)$$

Al-Assady and Al-Ta'ai (2002) developed an extended conjugate gradient algorithm which is:

$$F(q(x)) = \sin(eq(x)) \qquad \ldots(7)$$

Also, Al-Assady and Al-Ta'ai (2002) suggested a new model that is:

$$F(q(x)) = \sinh(eq(x)) \qquad \ldots(8)$$

3. New Hyperbolic Model for the CG-Methods:

In this paper, a new tangent hyperbolic model is investigated and tested on a set of standard test functions. 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:

$$f = F(q(x)), \quad \frac{df}{dq} = F^{/} > 0 \quad \text{and } q(x) > 0. \qquad \ldots(9)$$

(See Biggs, 1973; Tassopolus and Storey, 1984).

An extended conjugate pair algorithm is developed which is based on this new model which scales *q(x)* by the natural *sech* function for the rational *q(x)* functions

$$F(q(x)) = \sec h(eq(x)) \qquad \ldots(10)$$

We first observe that *q(x)* and *F(q(x))* given by eq.(10) have identical contours, though with different function values, and have the same unique minimum point denoted by x^*.

4. The Outlines of the New Algorithm:

Step (1): $k=1$.

Step (2): Compute $d_k = -g_k$, line search along d_k to get $x_{k+1} = x_k + b_k d_k$.

Step (3): If at x_{k+1} the stopping criterion $\|g_{k+1}\| \le 1\times10^{-5}$ is satisfied, then terminate.

Step (4): Check for restarting criterion if $k=n$ then go to step (1). Else go to step (5).

Step (5): Compute

$$r_k = \frac{\left(\frac{1+\sqrt{1-f_{k-1}^2}}{f_{k-1}} - \frac{f_{k-1}}{1+\sqrt{1-f_{k-1}^2}}\right)\left(\frac{1+\sqrt{1-f_k^2}}{f_k} + \frac{f_k}{1+\sqrt{1-f_k^2}}\right)^2}{\left(\frac{1+\sqrt{1-f_{k-1}^2}}{f_{k-1}} + \frac{f_{k-1}}{1+\sqrt{1-f_{k-1}^2}}\right)^2\left(\frac{1+\sqrt{1-f_k^2}}{f_k} - \frac{f_k}{1+\sqrt{1-f_k^2}}\right)}$$

Step (6): Compute $d_{k+1} = -g_{k+1} + r_k b_k d_k$, where b_k is H/S is defined by:

$$b_k = \frac{y_k^T g_{k+1}}{y_k^T d_k}$$

Step (7): Set $k=k+1$. If $k > 1000$, stop. Else go to step 2.

4. The Derivative of the r_k :

$f(x)=F(q(x))=sech(e\ q(x))$

$$\sec h^{-1}(f) = e\ q \Rightarrow \ln\left(\frac{1+\sqrt{1-f^2}}{f}\right) = e\ q\ , \qquad \dots (11)$$

$$q = \frac{\ln\left[\frac{1+\sqrt{1-f^2}}{f}\right]}{e} \qquad \dots(12)$$

and since $r_k = \frac{f_{k-1}^{/}}{f_k^{/}}$, then

$$r_k = \frac{-\sec h(e\ q_{k-1}(x))\tanh(e\ q_{k-1}(x))}{-\sec h(e\ q_k(x))\tanh(e\ q_k(x))} \qquad \dots(13)$$

$$= \frac{\frac{2}{e^{eq_{k-1}} + e^{-eq_{k-1}}} \cdot \frac{e^{eq_{k-1}} - e^{-eq_{k-1}}}{e^{eq_{k-1}} + e^{-eq_{k-1}}}}{\frac{2}{e^{eq_k} + e^{-eq_k}} \cdot \frac{e^{eq_k} - e^{-eq_k}}{e^{eq_k} + e^{-eq_k}}} \qquad \dots(14)$$

by use eq.(10), we get

$$r_k = \frac{\left(\frac{1+\sqrt{1-f_{k-1}^2}}{f_{k-1}} - \frac{f_{k-1}}{1+\sqrt{1-f_{k-1}^2}}\right)\left(\frac{1+\sqrt{1-f_k^2}}{f_k} + \frac{f_k}{1+\sqrt{1-f_k^2}}\right)^2}{\left(\frac{1+\sqrt{1-f_{k-1}^2}}{f_{k-1}} + \frac{f_{k-1}}{1+\sqrt{1-f_{k-1}^2}}\right)^2\left(\frac{1+\sqrt{1-f_k^2}}{f_k} - \frac{f_k}{1+\sqrt{1-f_k^2}}\right)} \quad \ldots(15)$$

Stability of the new model:

Conjugate direction methods have been proved to produce a successive of conjugate descent directions and hence it is well-known that these directions are stable. (Bazarra, 2000)

5. Numerical Results:

To test the effect of the suggested method, a sample of six problems was solved, in order to compare the new algorithm with the original CG-algorithm by using H/S formula.

For all cases the stopping criterion is taken to be $\|g_{k+1}\| < 1\times10^{-5}$.

The line search routine used was a cubic interpolation which uses function and gradient values which is described in Bunday (1984).

The following table gives the comparison between the results of the suggested algorithm with standard CG-algorithm, the results in this table indicate that the new algorithm is superior to the standard CG- algorithm with respect to the total of number of iteration NOI and number of function evaluation NOF.

Table (1)
Comparative Performance of the Two Algorithms for Group of Test Function

Test function	N	*CG Algorithm* NOI (NOF)	New Algorithm NOI (NOF)
(1)	100 1000	105 (244) 121 (317)	98 (202) 110 (300)
(2)	100	40 (99)	28 (62)
(3)	1000	45 (101)	27 (67)
(4)	100 1000	35 (64) 37 (71)	35 (64) 34 (70)
(5)	100 1000	13 (35) 14 (35)	11 (29) 14 (81)
(6)	100	50 (112)	41 (96)
Total		460 (1078)	399 (971)

6. Appendix:

The test functions

1- Generalized Powell Function:

$$f=\sum_{i=1}^{n/4}[(x_{4i-3}-10x_{4i-2})^2+5(x_{4i-1}-x_{4i})^2+(x_{4i-2}-2x_{4i-1})^4+10(x_{4i-3}-x_{4i})^4],$$

$x_0=(3,-1,0,1;\ldots)^T$.

2- Generalized Rosenbrock Function:

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

$x_0=(-1.2,1;\ldots)^T$.

3- Generalized Sum of Quadratics Function:

$$f=\sum_{i=1}^{n}(x_i-i)^4,$$

$x_0=(2;\ldots)^T$.

4- Generalized Dixon Function:

$$f=\sum_{i=1}^{n}[(1-x_i)^2+(1-x_n)^2+\sum_{i=1}^{n-1}(x_i^2-x_{i-1})^2],$$

$x_0=(-1;\ldots)^T$.

5- Generalized Cubic Function:

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

$x_0=(-1.2,1;\ldots)^T$

6- Generalized Wood Function:

$$f=\sum_{i=1}^{n/4}100[(x_{4i-2}-x_{4i-3}^2)^2]+(1-x_{4i-3})^2+9(x_{4i}-x_{4i-1}^2)^2+(1-x_{4i-1}^2)^2 .$$

$$+10.1[(x_{4i-2}-1)^2+(x_{4i}-1)^2]+19.8(x_{4i-2}-1)^2(x_{4i}-1),$$

$x_0=(-3,-1,-3,-1;\ldots)^T$.

REFERENCES

[1] Al-Assady, N.H. (1991), “New QN and PCG Algorithms Based on Non-Quadratic Properties”, J.Ed. and Sc.,12.

[2] Al-Assady, N.H. and Huda, K.M. (1997) “A Rotational Logarithmic Model for Unconstratined Nonlinear Optimization”, Rafideen Science, J. Vol. 8, No. 2, pp. 107-117, Mosul University, Iraq.

[3] Al-Assady, N.H. and Al-Ta’ai, B.A. (2002), “A New Non-Quadratic Model for Unconstrained Nonlinear Optimization” University of Mosul, College of Computer and Mathematical Science, No. 48.

[4] Al-Assady, N.H. and Al-Ta’ai, B.A. (2002) “Hyperbolic Model for Unconstrained Nonlinear Optimization” University of Mosul, College of Computer and Mathematical Science, No. 48.

[5] Al-Bayati, A.Y. (1993) “A New Non-Qudratic Model for Unoconstrained Nonlinear Optimization Method”, Natural and Applied Series, Mu’tah Journal for Research and Studeis, Mu’tah University, Jordan, Vol. 8, No. 1, pp. 133-155.

[6] Al-Bayati, A.Y. (1995) “New Extended CG-Methods for Non-linear Optimization”, Natural and Applied Series, Mu’tah Journal for Research and Studies, Mut’ah University, Jordan, Vol. 10, No. 6, pp. 69-87.

[7] Bazaraa, M.S.; Shetty, C.M. and Sherali, H.D. (2000) “ Nonlinear Programming”, England, Universities Press, London.

[8] Biggs, M.C. (1973) “A Note On Minimization Algorithms which Make Use of Non-Quadratic Properties of the Objective Function”, Journal of Institute of Mathematics and Its Application, No. 12, pp. 337-338.

[9] Bunday, B. (1984) “Basic Optimization Methods”, Edward Arnold, Bedford Square, London.

[10] Boland, W.R. and Kowalik, J.S. (1979) “Extended CG-Method with Restarts”, Journal of Opt, Theory and Applications, Vol. 28, pp. 1-9.

[11] Fried, I. (1971) “N-Steps Conjugate Gradient Minimization Scheme for Non-Quadratic Function”, AIAA Journal,Vol. 19, pp. 2286-2287.

[12] Fletcher, R. and Reeves, C.M., (1964) “ Function Minimization by Conjugate Gradient” Computer Journal, vol. 7, pp. 149-154.

[13] Hestenes, M.R., and Stiefel, E. (1952) " Methods of Conjugate Gradient for Solving Linear Systems", J. Res. N.B.S., 49.

[14] Spedicato, E. (1976) "A Variable Metric Method for Function Minimization Derived from Invariance to Nonlinear Scaling", Journal of Opt. Theory and Application, Vol. 20.

[15] Tassopoulos, A. and Storey, C. (1984) "Use of Non-Quadratic Model in a Conjugate Gradient Model of Optimization with Inexact Line Search", JOTA, Vol. 43.