New Extended CG Algorithm For Non-Linear Optimization

Abbas H.Taqi and Adham A. Ali
College of Science-University of Kirkuk

Abstract

This paper presents the development and implementation of a new algorithm based on non-quadratic rational function model.The derivation of the new algorithm is based on aquadratic function with exact line searches and evaluated numerically against the standard CG-algorithm by using(25) non-linear test functions with different dimensions. The numerical results indicate that the new algorithm is found to be superior to the standard CG algorithm.

Introduction:

Conjugate gradient methods are iterative methods which generate a sequence of approximations to minimize a function f(x).The methods are based on an important concept of conjugating two vectors $x_1 \varepsilon R^n$ and $x_2 \varepsilon R^n$,that are said to be conjugate directions with respect to G if $x_1^T G x_2 = 0$ where G is Hession matrix. The notion of conjugate directions is a generalization of the concept of orthogonality (conjugate vectors x_1 and x_2 are orthogonal when G is n×n identity matrix I). Several algorithms have been proposed in literature for generating conjugate directions of quadratic forms.The first conjugate gradient method was published by Hestense and Stiefel in (Hajitharwat,H. and AI-Bayati,A.Y.,2005), for solving a system of linear algebraic equations. Fletcher and Reeves (Bunday,1984) were the first among others scholars, to use this technique to minimize a non linear function of several variables.

Definition:

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

$$f = F(q(x)), \frac{dF}{dq} = \overline{F} > 0 \text{ and } q(x) > 0 \qquad ...(1)$$

where x^* is the minimizer of q(x) with respect to x (Taqi,& AI-Assady,2000).

The following properties are immediately derived from the above condition :

i) Every contour line of 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.

Many authors have published related work in this area:
i) A CG methods which minimize the following polynomial model

$$f(x) = ((q(x))^{\rho}, \rho > 0 \quad x \varepsilon R^n \qquad ...(2)$$

In at most n steps has been proposed in (Fried, 1971).
ii) Two CG methods which minimize the following polynomial model.

$$f(x) = \varepsilon_1 q(x) + \frac{1}{2}\varepsilon_2 q^2(x) \qquad ...(3)$$

where ε_1 and ε_2 are scalars, have been investigated with two different restarting critieriain (Boland , Kamagnia & Kowalik , 1977a ; Boland & kowalik , 1979b).
iii) Also two different rational models have been developed in (Taqi & AI-Assady, 2000) and (Tassopoulos & Storey, 1984a), namely

$$f(x) = \frac{\varepsilon_1 q(x) + 1}{\varepsilon_2 q(x)}, \varepsilon_2 > 0 \qquad ...(4)$$

and

$$f(x) = \frac{\varepsilon q(x)}{1 + q(x)}, \varepsilon > 0 \qquad ...(5)$$

iv) Another new CG method which based on general logarithmic model.

$$f(x) = \varepsilon(\log q(x) - 1), \varepsilon > 0 \qquad ...(6)$$

have been implemented by Al-Bayati (Al-Bayati, 1995).
v) And Taqi A. and Al-Assady (Tassopoulos & Storey,1984b) described their ECG algorithm which based on the natural log function for the rational q(x) function:

$$F(q(x)) = \log \frac{\varepsilon_1 q^r(x)}{\varepsilon_2 q^r(x) + 1}; r > 0 \quad \varepsilon_2 < 0 \qquad ...(7)$$

where ε_1 and ε_2 are scalars.
vi) Also Al-Mashhadany H (Al-Mashhadany, 2002) has been developed a new rational models which is defined as following:

$$F(q(x)) = \exp(\frac{\varepsilon_1 q(x)}{\varepsilon_2 q(x) - 1}); \quad \varepsilon_2 < 0 \qquad ...(8)$$

vii) Finally, another specific rational model was considered by Haji Tharwat H. (Hajitharwat & AI-Bayati,2005) which is defined as following:

$$F(q(x)) = \sinh^{-1}(q(x)) \qquad ...(9)$$

In this paper a new extended CG method (ECG) is investigated and tested on a set of some standard test functions. The new model is given as:

$$f(x) = \cos^{-1}\left(\frac{1-\varepsilon_1 q(x)}{\varepsilon_2 q(x)}\right) \qquad ...(10)$$

where ε_1 and ε_2 are scalars.

It is assumed that: $\frac{dF}{dq} = f' > 0$ for $q > 0$ ...(11)

holds

However, we first observe that q(x) and f(x) given in the above new model have identical contours through with different function values, and they have the same unique minimum point x^* .

The derivation of new ECG –method:

The key element of the modified algorithms is the determination of the expression ρ_i where:

$$\rho_i = \frac{f'_{i-1}}{f'_i} \qquad ...(12)$$

It is first assumed that neither ε_1 and ε_2 is zero in (10), solving (10) for q(x), then:

$$q = \frac{1}{\varepsilon_2(\cos f + \frac{\varepsilon_1}{\varepsilon_2})} \qquad ...(13)$$

and using the expression for ρ_i

$$\rho_i = \frac{f'_{i-1}}{f'_i} = \left[\frac{\sin f_i}{\sin f_{i-1}}\right]\left[\frac{\cos f_{i-1} + \frac{\varepsilon_1}{\varepsilon_2}}{\cos f_i + \frac{\varepsilon_1}{\varepsilon_2}}\right]^2 \qquad ...(14)$$

Where $f'_i = \frac{\varepsilon_2(\cos f_i + \frac{\varepsilon_1}{\varepsilon_2})^2}{\sin f_i}$ and $f'_{i-1} = \frac{\varepsilon_2(\cos f_{i-1} + \frac{\varepsilon_1}{\varepsilon_2})^2}{\sin f_{i-1}}$

the quantity which has be determined explicitly is $\frac{\varepsilon_1}{\varepsilon_2}$ during every iteration $\frac{\varepsilon_1}{\varepsilon_2}$ must be evaluated as a function of known available quantities.

From the relation: $g_i = f'_i G(x_i - x^*)$ ...(15-a)

$$g_{i-1} = f'_{i-1} G(x_{i-1} - x^*) \qquad ... (15\text{-b})$$

where G is Hessian matrix and x^* is the minimum point from the above system we have:

$$\rho_i = \frac{f'_{i-1}}{f'_i} = \frac{g_{i-1}^T(x_i - x^*)}{g_i^T(x_{i-1} - x^*)} \qquad ...(16)$$

where g_i^T is the tranpose of g_i furthermore.

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

and

$$g_i^T(x_{i-1} - x^*) = g_i^T(x_i - \lambda_{i-1} d_{i-1} - x^*) = g_i^T(x_i - x^*)$$

since $g_i^T d_{i-1} = 0; (ELS)$

therefore we can express ρ_i as follows:

$$\rho_i = \frac{\left[g_{i-1}^T(x_{i-1} - x^*) + \lambda_{i-1} g_{i-1}^T d_{i-1}\right]}{g_i^T(x_i - x^*)} \qquad ...(17)$$

from (12) and (15), we get

$$\rho_i = \frac{\left[f'_{i-1}(x_{i-1} - x^*)^T G(x_{i-1} - x^*) + \lambda_{i-1} g_{i-1}^T d_{i-1}\right]}{f'_i(x_i - x^*)^T G(x_i - x^*)}$$

therefore

$$\rho_i = \frac{\left(2 f'_{i-1} q_{i-1} + \lambda_{i-1} g_{i-1}^T d_{i-1}\right)}{2 f'_i q_i} = \rho_i \frac{q_{i-1}}{q_i} + \frac{\lambda_{i-1} g_{i-1}^T d_{i-1}}{2 f'_i q_i} \qquad ...(18)$$

where q_i is the quadratic function defined by:

$$q_i = \frac{1}{2}(x_i - x^*)^T G(x_i - x^*)$$

The quantities $(\frac{q_{i-1}}{q_i})$ and $f'_i q_i$ can be rewritten as:

$$\frac{q_{i-1}}{q_i} = \frac{\cos f_i + \frac{\varepsilon_1}{\varepsilon_2}}{\cos f_{i-1} + \frac{\varepsilon_1}{\varepsilon_2}} = \frac{1}{\sqrt{\rho_i}} \sqrt{\frac{\sin f_i}{\sin f_{i-1}}} \qquad ...(19)$$

$$f'_i q_i = \frac{\left(\cos f_i + \frac{\varepsilon_1}{\varepsilon_2}\right)}{\sin f_i} \qquad ...(20)$$

Substituting equations (19) and (20) in equation (18) gives:

$$\rho_i^{New2} = \sqrt{\rho_i} \sqrt{\frac{\sin f_i}{\sin f_{i-1}}} + \frac{n \sin f_i}{\cos f_i + \frac{\varepsilon_1}{\varepsilon_2}} \qquad ...(21)$$

where $n=\dfrac{\lambda_{i-1}g_{i-1}^{T}d_{i-1}}{2}$

From equations (14) and (21), it follows that:

$$\left[\frac{\sin f_i}{\sin f_{i-1}}\right]\left[\frac{\cos f_{i-1}+\frac{\varepsilon_1}{\varepsilon_2}}{\cos f_i+\frac{\varepsilon_1}{\varepsilon_2}}\right]^2=\sqrt{\frac{\sin f_i}{\sin f_{i-1}}}\left[\frac{\cos f_{i-1}+\frac{\varepsilon_1}{\varepsilon_2}}{\cos f_i+\frac{\varepsilon_1}{\varepsilon_2}}\right]\sqrt{\frac{\sin f_i}{\sin f_{i-1}}}+\frac{n\sin f_i}{\cos f_i+\frac{\varepsilon_1}{\varepsilon_2}} \qquad ...(22)$$

$$\Rightarrow\left[\frac{\cos f_{i-1}+\frac{\varepsilon_1}{\varepsilon_2}}{\cos f_i+\frac{\varepsilon_1}{\varepsilon_2}}\right]^2=\frac{\cos f_{i-1}+\frac{\varepsilon_1}{\varepsilon_2}}{\cos f_i+\frac{\varepsilon_1}{\varepsilon_2}}+\frac{n\sin f_{i-1}}{\cos f_i+\frac{\varepsilon_1}{\varepsilon_2}} \qquad ...(23)$$

Equation (23) can be rewritten as:

$$\left(\cos f_{i-1}+\frac{\varepsilon_1}{\varepsilon_2}\right)^2=\left(\cos f_{i-1}+\frac{\varepsilon_1}{\varepsilon_2}\right)\left(\cos f_i+\frac{\varepsilon_1}{\varepsilon_2}\right)+n\sin f_{i-1}\left(\cos f_i+\frac{\varepsilon_1}{\varepsilon_2}\right) \qquad ...(24)$$

By solving the equation (24), we have:

$$\frac{\varepsilon_1}{\varepsilon_2}=\frac{\cos f_{i-1}(\cos f_i-\cos f_{i-1})+n\sin f_{i-1}\cos f_i}{(\cos f_{i-1}-\cos f_i-n\sin f_{i-1})} \qquad ...(25)$$

Using the following transformation:

$$w=\cos f_i-\cos f_{i-1}$$

The equation (25) can be rewritten as:

$$\frac{\varepsilon_1}{\varepsilon_2}=\frac{w\cos f_{i-1}+n\sin f_{i-1}\cos f_i}{-(w+n\sin f_{i-1})} \qquad ...(26)$$

Substituting equation (26) in equation (14) we obtain:

$$\rho_i=\left[\frac{\sin f_i}{\sin f_{i-1}}\right]\left[\frac{-n\sin f_{i-1}}{w}\right]^2$$

where $n=\dfrac{\lambda_{i-1}g_{i-1}^{T}d_{i-1}}{2}$, $w=\cos f_i-\cos f_{i-1}$

The outlines of the new algorithm:

Given $x_0\varepsilon R^n$ an initial estimate of the minimizer x^* and scalar ε.

(i): Set $d_\circ=-g_\circ$

(ii): For i = 1,2,…

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.

(iii): Define

$$n = \frac{\lambda_{i-1} g_{i-1}^T d_{i-1}}{2}$$

$$w = \cos f_i - \cos f_{i-1}$$

(iv): Compute

$$\rho_i = \left[\frac{\sin f_i}{\sin f_{i-1}}\right]\left[\frac{-n \sin f_{i-1}}{w}\right]^2$$

(vi): Calculate the new direction

$$d_i = -g_i + \beta_i d_{i-1}$$

where β_i is defined as follows:

$$\beta_i = \frac{g_i^T(\rho_i g_i - g_{i-1})}{[d_{i-1}^T(\rho_i g_i - g_{i-1})]}$$ modified H/S in (Hestences& Stiefle, 1952)

$$\beta_i = \frac{g_i^T(\rho_i g_i - g_{i-1})}{(g_{i-1}^T g_{i-1})}$$ modified P/r in (Polak& Ribier,1969)

(vii) Check for convergence

If $\|g_i\| \le \varepsilon$, then stop, else go to step (viii)

(viii) Check for restarting criterion

If i = n, set i = 0 and $x_0 = x_n$ then go to step (i)

Else set i = i +1 and go to step (ii)

Numerical Computation:

To test the effectiveness of ECG-method, a number of standard test functions were solved in order to compare the new algorithm with the standard CG method the identical linear search was used, namely a cubic fitting procedure described in Bunday (Bunday, 1984).Finally the convergence criterion used in each case is that $\|g_i\| \le 5 \times 10^5$ for this ECG–method with E/S all computations in double precision arithmetic are performed by using personal computer (Pentium iv), all programs are written in Fortran language. All the results given in the tables specifically count the number of function calls (NOF) and the number of the iterations call (NOI). Results in table 1 and 2 give the comparison of new ECG with standard CG method.

Table.1

Test Function	n	H/S	
		ECG NOI(NOF)	Standard CG NOI(NOF)
Powell	100 300 500 1000	87(188) 119(246) 77(165) 90(191)	129(263) 328(661) 458(921) 558(1121)
Cantral	4 100 1000	27(155) 18(116) 23(151)	33(230) 19(137) 20(152)
Rosen	300 1000	17(41) 30(65)	24(59) 24(59)
Cubic	100 500 1000	13(33) 13(33) 13(33)	14(37) 14(37) 14(37)
OSP	4 100 500	5(21) 21(77) 100(280)	5(24) 17(67) 93(246)
Beale	100 300 1000	10(25) 10(25) 10(25)	8(18) 8(18) 8(18)
Non Diagonal	4 300 500	33(75) 22(61) 39(82)	21(50) 24(60) 25(62)
Wolfe	100 300 500 1000	838(1761) 44(89) 49(102) 48(97)	1011(2099) 53(107) 56(113) 70(141)
Total NOI(NOF)		1756(4137)	3034(6737)

Table .2

Test Function	n	P/R ECG NOI(NOF)	P/R Standard CG NOI(NOF)
Powell	100 300 500 1000	79(161) 111(228) 139(289) 138(293)	118(665) 322(665) 506(1016) 1006(2016)
Cantral	20 80 100	14(79) 14(79) 18(139)	19(119) 19(119) 19(119)
Rosen	4 300 1000	29(70) 20(43) 18(43)	30(72) 26(62) 26(62)
Cubic	100 500 1000	23(52) 23(52) 23(52)	14(37) 14(37) 14(37)
OSP	4 20 500	5(21) 16(58) 106(304)	5(23) 15(61) 127(385)
Beale	100 500 1000	10(26) 10(26) 10(26)	8(20) 8(20) 8(20)
Non Diagonal	10 300	17(39) 30(69)	24(59) 26(60)
Wolfe	100 300 500 1000	806(1670) 44(89) 49(108) 50(105)	907(1853) 54(109) 58(117) 72(145)
Total NOI(NOF)		1802(4121)	3445(7898)

From comparing new algorithm ECG with standard CG method using (H/S) formula, see table (1) we obtained the following results:

Table.3

Measure	Standard H/S-CG	ECG
NOI	100%	89%
NOF	100%	90%

It is clear from the table 3, that the New algorithm (ECG) improve the standard H/S-CG algorithm in about (11%) NOI and (10%) NOF.And from comparing New algorithm (ECG) with standard CG-method using (P/R) formula, see table (2) we obtained the following results:

Table .4

Measure	Standard P/R-CG	ECG
NOI	100%	91%
NOF	100%	92%

It is clear from the table 4, that the New algorithm (ECG) improve the standard P/R-CG algorithm in about (9%) NOI and (8%) NOF.

References:

- Al-Bayati, A. Y.,(1995):New Extended CG. Methods for Non Linear Optimization, Natural and Applied Series, Mu’tah Journal for Research and Studies. Mu’tah University, Jordan Vol. 10,pp:69-87.
- Al-Mashhadany,H.,(2002):A Rational Exponential Model for Unconstrained Non-Linear Optimization,Qatar University.Sc.J,pp:5-16.
- Boland, W. R.; Kamagnia, E. R. and Kowalik, J. S.,(1977a):A Conjugate Gradient Optimization Method in Variant to Non-Linear scalling, Journal of Optimization Theory and Application,Vol.27,pp:221-230.
- Boland,W. R . and kowalik , J . S . , (1979b) : Extended Conjugate Gradient Method with Restarts,Journal of Opt.Theory and Application,Vol.28 , pp:1-9.
- Bunday, B . , (1984): Basic Optimization Methods , Edward Arnold , Bedford Square , London.
- Fried, I.,(1971):N-Step Conjugate Gradient Minimization Scheme for Non - Quadratic Function, AIAA J., a.
- Hajitharwat, H. and AI-Bayati , A. Y.,(2005):Investigation in general Optimization Algorithms based on Quadratic and Non-Quadratic Models,Ph.D. thesis University of Mosul.
- Hestences, M. R. and Stiefle, E.,(1952):Methods of Conjugate Gradient for Solving Linear Systems, J.Res. N. B. S., Vol.49.
- Polak, E. and Ribier, G.(1969): Note Surla Convergence Des Methods De Directions Conjugate, Rev. Fr. infr, Rech Oper,Vol.16,R.
- Taqi, A. H and AI-Assady, N. H.,(2000):An Investigation of Numerical Techniques for Solving of Optimization Problems, Ph.D Thesis University of Mosul.

- Tassopoulos, A. and Storey, C.,(1984a):A Conjugate Direction Method Based on Non-Quadratic Model, Journal of Optimization theory and Applications,Vol.43, pp:371-381.
- Tassopoulos, A. and Storey, C., (1984b):Use of a New-Quadratic Model in a CG searches, JOTA,V0l.43.

Test functions:

1. Beale Function:

$$f = \left(1.5 - x_1(1-x_2^2)\right) + \left(2.25 - x_1(1-x_2^2)\right)^2 + \left[2.625 - x_1(1-x_2^9)^2\right]$$

$$x_\circ = (0,0)^T$$

2. Cubic Function:

$$f = 100(x_2 - x_1^3)^2,$$

$$x_\circ = (-1.2,1)^T$$

3. Generalized Central Function:

$$f = \sum_{i=1}^{\frac{n}{4}} [\exp(x_{4i-3}) - x_{4i-2}]^2 + 100(x_{4i-2} - x_{4i-1})^6 + \left[[a\tan(x_{4i-1} - x_{4i})]^4 + x_{4i-3}^8\right]$$

$$x_\circ = (1,2,2,2)^T$$

4. Oren and Spedicato Power Function (OSP):

$$f = \sum_{i=1}^{n} \left(ix_i^2\right)^r, r = 2$$

$$x_\circ = (1,....)^T$$

5. Rosenbrock Function:

$$f = \sum_{i=1}^{\frac{n}{2}} \left(100(x_{2i} - x_{2i}^2) + (1 - x_{2i-1})^2\right)$$

$$x_\circ = (-1.2,1)^T$$

6. Generalized Powell Function:

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

$$x_\circ = (3,-1,0,1,.....)^T$$

7. Non –Diagonal variant of Rosenbrock Function:

$$f = \sum_{i=1}^{n} 100\left(x_i - x_\circ^2\right)^2 + (1 - x_\circ)^2$$

$$x_\circ = (-1,.....)^T$$

8. Wolfe Function:

$$f = \left[-x_1(3 - \frac{x_1}{2}) + 2x_2 - 1\right]^2 + \sum_{i=1}^{n-1}\left[x_{i-1} - x_i(3 - \frac{x_i}{2}) + 2x_{i-1} - 1\right]^2 + \left[x_{n-1} - x_n(3 - \frac{x_n}{2}) - 1\right]^2$$

$$x_\circ = (-1,............)^T$$

توسيع جديد لخوارزمية التدرج المترافق في الامثلية اللاخطية

عباس حسن تقي و ادهم عبدالوهاب علي
كلية العلوم ــ جامعة كركوك

الخلاصة

يقدم هذا البحث تطوير واستعمال خوارزمية جديدة في مجال الأمثلية غير المقيدة والتي تعتمد علــى أحــد النماذج النسبية غير التربيعية. تم اشتقاق هذه الخوارزمية بالاعتماد على الدالة التربيعية وعلى أساس خط البحث التام وحسبت عددياً مقارنة بخوارزمية CG القياسية، باستخدام (25) دالة غير خطية من ذوات الأبعاد المختلفة. النتائج العددية التي توصلنا إليها ، تشير إلى كفاءة الخوارزمية الجديدة بعد مقارنتها مع بعض الخوارزميات في هذا البحث وباستعمال عدد كبير من الدوال اللاخطية.