A New Conjugate Gradient Method for Finding the Minimum of Non Linear Functions

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

Received on:26/10/2005 **Accepted on:28/2/2006**

ABSTRACT

This paper presents the development and implementation of a new numerical algorithm for solving nonlinear optimization problems. The algorithm is implemented inexact line searches. Powell restarting restart criterion is applied to all the above versions and give dramatic saving in computational efficiency. The results obtained both theoretically and experimentally indicate that in general the new algorithm is superior an standard algorithms using seven nonlinear test-functions with (20) differs dimensions.

Keywords: A new conjugate gradient method, Numerical Results and Conclusions

طريقة التدرج المترافق الجديدة لإيجاد القيمة الصغرى للدول غير الخطية

عباس يونس البياتي **باسم عباس حسن**
كلية علوم الحاسبات والرياضيات
جامعة الموصل

تاريخ قبول البحث: 2006/2/28 **تاريخ استلام البحث: 2005/10/26**

الملخص

تم في هذا البحث تطوير واستعمال خوارزمية جديدة في مجال ألا مثلية غير المقيدة باستخدام خط بحث غير تام .لقد تم استخدام مقياس Powell للإسترجاع على جميع الصيغ المستعملة وكان لذلك اثر كبير في توفير كفاءة للحل. إن النتائج التي تم التوصل أليها سواءاً كانت نظرية أو عملية اثبت إن الخوارزمية الجديدة هي اكثر كفاءة من الخوارزميات الاصلية باستخدام سبعة دوال غير خطية ذات (20) بعد مختلف.

الكلمات المفتاحية : طريقة التدرج المترافق الجديدة , النتائج العددية والاستنتاجات.

1.Introduction

Conjugate gradient methods (CG) are iterative methods, which generate a sequence of approximations to minimize a function $f(x)$. The basis of all CG- methods is to determine new directions of search using information relation only to gradient of a quadratic objective function, in such a way that successive search directions are conjugate with respect to the matrix A of that quadratic form .At each stage i the direction d_i is obtained by combining linearly the gradient at $x_i, -g_i$ and the previous conjugate directions $d_0, d_1, d_2, d_3, ..., d_{i-1}$, where the coefficients of the linear combination is chosen in such a way that d_i is conjugate to all previous directions .There fore, the direction vector d_i is computed recursively by the rule :

$$d_{i+1} = -G_{i+1} + \beta_i d_i \qquad \ldots(1)$$

Where β_i is the conjugate coefficient and it has several values which are introduced from several others as follows:

$$\left.\begin{aligned} \beta_i &= y_i^T G_{i+1} / d_i^T y_i && (Hestenes-Stiefel\ ,[6]\ ) \\ \beta_i &= G_{i+1}^T G_{i+1} / G_i^T G_i && (Fletcher-\mathrm{Re}eves\ ,[5]\ ) \\ \beta_i &= G_{i+1}^T y_i \ / G_i^T G_i && (Polak-Ribiere\ ,[7]\ ) \\ \beta_i &= -G_{i+1}^T G_{i+1} / d_i^T G_i && (Dixon\ ,[4]\ ) \end{aligned}\right\} \qquad \ldots(2)$$

where $y_i = G_{i+1} - G_i$

We can define the classical CG algorithm as follows Beale [3].

Algorithm

Step (1)-Set x_0, ε, n

Step (2)-For $i = 1,\ d_i = -G_i$

Step (3)-Compute $x_{i+1} = x_i + \lambda_i d_i$

Where λ_i is obtained from a line search procedure.

Step (4)-Check for convergence if $\|G_{i+1}\| \le \varepsilon$, then stop.

Step (5)-Compute the new direction

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

Where β_i is the conjugac coefficient.

Step (6)-Check for restarting criterion if $d_{i+1}^T G_{i+1} \succ -0.8\, G_{i+1}^T G_{i+1}$ .then set i=0 and go to step(2) else go to step (3) .

The CG-method is developed in order to overcome the difficulties of the zigzagging method that is very slow in [2].

In order to improve the local rates of convergence and the efficiency of the classical CG method, several established algorithms are discussed by Dixon's [4] asi gradient prediction method, Nazareth's [8] three- term formula & Nazareth and Nocedal's [9] multi step method. They have all shown that such algorithms are able to generate conjugate directions for a quadratic function without performing exact searches, they will satisfy the quadratic termination property by using an error vector. Other important algorithm of this type, developed by Sloboda [11] and Al-Assady and Al-Bayati [1], They satify the quadratic termination property without the use of error vector, some modifications where developed by several authors such as Touati & Story [13] for a such type of algorithms.

2. Development of a New CG –algorithm for the quadratic function:

In this section a new general way for the (CG –F/R) type methods is presented.

This new approach has the property of the quadratic termination even if the line search is not exact, Where G_{i+1} is the gradient of the quadratic function at $x_{i+1} = x_i + \lambda_i d_i$ [12] and let

$$G^*_{I+1} = G_i + \frac{G_i^T G_i}{d_i^T y_i} y_i \qquad \ldots(3)$$

Then the following Lemma is hold.

Lemma (1)

For the quadratic function the term G^*_{i+1} which is defined in (3) is equivalent to that the gradient G_{i+1} is obtained by Hestenes and Stiefel [6].

Proof:

Let we define $G_i = Ax_i - b \quad and \quad d_i = -G_i$ …(4a)

The biorthogonalization process of Hestenes and Stiefel is defined as follows:

$$x_{i+1} = x_i + \lambda_i d_i$$

Wheres $\lambda_i = -G_i^T d_i / d_i^T A d_i$ for exact line searches and A is a symmetric positive definite matrix. Thus

$$G_{i+1} = G_i + \lambda_i A d_i \qquad \ldots(4b)$$

and

$$d_{i+1} = -G_{i+1} + \frac{y_i G_{i+1}}{d_i y_i} d_i$$

Now in order to prove that G_{i+1} which is obtained by H/S with exact line searches algorithm is identical to the term G^*_{i+1} , which is defined in eq.(1) as follows :

From the definition $\lambda_i = -G_i d_i / d_i^T A d_i$ then rewriting eq. (4) , we get

$$G_{i+1} = G_i - \frac{G_i^T d_i}{d_i^T A d_i} A d_i$$

Now multiplying and dividing the second term of the eq. (4) by λ, it becomes as

$$G_{i+1} = G_i - \frac{G_i^T d_i}{d_i^T \lambda A d_i} \lambda A d_i$$

From the definition we have $y_i = G_{i+1} - G_i = \lambda_i A d_i$ then replacing y_i instead of $\lambda_i A d_i$ in the above equation, then we get:

$$G_{i+1} = G_i - \frac{G_i^T d_i}{d_i^T y_i} y_i$$

Thus this equation is identical to the equation which is defined in eq. (1) as:

$$G_{i+1}^* = G_i + \frac{G_i^T G_i}{d_i^T y_i} y_i \quad ,$$

Thus the set of vectors $G_0^*, G_1^*, G_2^*, ..., G_I^*$ and orthogonal as in H/S method in quadratic function. From the above argument we have the following two corollaries:

Corollary (1)

The term G_{i+1}^* which is defined in eq. (3) is used to obtained by the following form:

$$d_{i+1}^* = G_{i+1}^* + \frac{(G_{i+1}^* - G_i)^T G_{i+1}^*}{d_i^{*T}(G_{i+1}^* - G_i)} d_i^* \qquad \ldots(5)$$

Is parallel to that search direction given by H/S algorithm for quadratic function.

Corollary (2)

The search direction which is defined in eq. (5) is descent direction even if for non-quadratic function i.e $d_{i+1}^* G_{i+1}^* \prec 0$.

Proof:

Rewrite the direction in eq. (5)

$$d_{i+1}^* = -G_{i+1}^* + \beta_i^* d_i^*$$

Multiplying this direction by G_{i+1}^* then we have:

$$d_{i+1}^* G_{i+1}^* = -G_{i+1}^* G_{i+1}^* + \beta_i^* d_i^* G_{i+1}^* \qquad \textit{and then} \quad d_{i+1}^* G_{i+1}^* = -\left\| G_{i+1}^* \right\|^2 ,$$

This result is true, because we have $d_i^* G_{i+1}^* = 0,$ it is easy to prove this result, such as follows we have,

$$G_{i+1}^* = G_i + \frac{G_i^T G_i}{d_i^T y_i} y_i \qquad \ldots(6)$$

So $y_i = G_{i+1}^* - G_i$. Now multiply eq.(6) by d_i^T, then we obtain the following result : $d_i^T G_{i+1}^* = d_i^T G_i + \frac{d_i^T G_i}{d_i^T y_i} G_i^T y_i$

From eq (4a) $d_i^T G_{i+1}^* = -\left\| G_i \right\|^2 + \frac{\left\| G_i \right\|^2}{d_i^T y_i} d_i^T y_i = 0$

The New Algorithm

Step (1)-Set $d_0^* = -G_0 = -G_0^*$

Step (2)-For $i = 1, 2, 3, ...$ Compute $x_{i+1} = x_i + \lambda_i d_i^*$

Where λ_i is obtained from the line search procedure.

Step (3)-Check for convergence if $\|G_{i+1}\| \le \varepsilon$, then stop. Otherwise go to step (4) .

Step (4)-Compute G^*_{i+1} as defined in eq. (3).

Step (5)-Compute $the \quad error - term \qquad c_1 = 0$

$$c_{i+1} = c_i + \varepsilon_i d_i \quad , \varepsilon_i = \lambda_i \left(-\frac{d_i^T G_{i+1}}{y_i^T d_i}\right)$$

$$x_{n+1} = x_n + c_n$$

for the purpose of conjugacy condition

Step (6)-Compute the new direction

$$d^*_{i+1} = -G^*_{i+1} + \beta^*_i d^*_i$$

Where

$$\left.\begin{array}{ll}\beta^*_i = y_i^T G^*_{i+1} / d_i^T y_i & (Hestenes - Stiefel \;\; ,1952) \\ \beta^*_i = G^T_{i+1} G^*_{i+1} / G_i^{*T} G^*_i & (Fletcher - \mathrm{Re}eves \;\; ,1964) \\ \beta^*_i = G^{*T}_{i+1} y_i / G_i^{*T} G^*_i & (Polack - Ribiere \;\; ,1969) \\ \beta^*_i = -G^{*T}_{i+1} G^*_{i+1} / d_i^T G^*_i & (Dixon \;\; ,1975)\end{array}\right\} \quad \ldots(7)$$

Step (7)-Check for restarting criterion in [10] $d^T_{i+1} G_{i+1} \succ -0.8\, G^T_{i+1} G_{i+1}$ then set i=0 and go to step(1) else go to step (2) .

This algorithm is identical to the original CG-methods in quadratic function, because of the ortho geniality property and the lemma (1)is holds. For general function this algorithm is reduced to the P/R algorithm even it inexact searches can used as we have that in Corollary (2).

3. Numerical Results and Conclusions:

Several standard test functions are minimized to compare the new algorithm with standard CG-algorithms (see appendix) with different dimensions $(2,4,8,10,20,40,60,............,400)$. All the results are obtained using double precision on the (Pentium (4) computer) using programs written in FORTRAN.

The compression performance of the algorithm are evaluated by considering both the total no. of function evaluations and the total no. of iterations. The stopping criterion is taken to be :

$$\|g_{i+1}\| \prec 1*10^{-5}$$

The line search routine employed is the cubic fitting technique, which uses function and gradient values.

We evaluate for all the algorithms the number of calls of the function evaluations (NOF) , and the number of iterations (NOI). Overall totals are also given for NOF & NOI with each algorithm.

Table (1) contains the numerical results for the new algorithm with (DX) formula & the standard CG-algorithm with the same formula. In this table we see that the new algorithm is more efficient than the standard CG-algorithm, this is obtained from the NOF & NOI of both algorithms.

Table (2) contains the numerical results for the new algorithm with (P/R) formula & the standard CG-algorithm with the same formula. In this table we see that the new algorithm is more efficient than the standard CG-algorithm, this is obtained from the NOF & NOI of both algorithms.

Table (1)

Comparative performance of the two algorithms for group of test function by using β : (DX)

Test function	N	CG-algorithm NOI (NOF)	New-algorithm NOI (NOF)
DIXON	2	15 (33)	9 (22)
NON-DIAGONL	2	35 (80)	10 (48)
DIXON	4	16 (34)	16 (34)
CANTRAL	4	26 (175)	23 (140)
MIELE	4	304 (734)	43 (129)
CUBIC	4	115 (238)	20 (51)
DIXON	10	27 (57)	27 (57)
NON-DIAGONL	10	29 (78)	24 (73)
CANTRAL	20	29 (210)	23 (140)
NON-DIAGONL	40	35 (91)	28 (80)
CANTRAL	40	32 (248)	23 (140)
CUBIC	40	125 (258)	20 (51)
WOLFE	40	55 (111)	55 (111)
NON-DIAGONL	100	118 (259)	25 (70)
CUBIC	100	134 (276)	20 (51)
WOLFE	100	59 (119)	60 (121)
NON-DIAGONAL	200	26 (70)	321 (662)
CANTRAL	400	35 (287)	25 (168)
WOLFE	400	86 (173)	132 (265)
CUBIC	400	143 (294)	20 (51)
TOTAL		1444 (3825)	924 (2521)

Percentage performance of the new algorithm against the standard CG algorithm

Tools	Standard –CG	NEW
NOI	100	63.98
NOF	100	65.90

Table (2)

Comparative performance of the two algorithms for group of test function by using β : (P/R)

Test function	N	CG-algorithm NOI (NOF)	New-algorithm NOI (NOF)
CUBIC	2	15 (36)	12 (33)
NON-DIAGONL	2	11 (33)	10 (31)
DIXON	2	5 (14)	5 (13)
DIXON	4	17 (36)	17 (36)
CANTRAL	4	20 (101)	20 (104)
MIELE	4	304 (706)	146 (333)
POWELL	4	168 (394)	67 (151)
CANTRAL	40	21 (112)	21 (116)
CUBIC	40	14 (36)	12 (33)
NON-DIAGONL	60	22 (61)	19 (54)
WOLFE	80	49 (99)	49 (99)
WOLFE	100	49 (99)	49 (99)
NON-DIAGONL	100	22 (60)	24 (62)
CUBIC	100	14 (36)	12 (33)
MIELE	100	635 (1397)	369 (845)
POWELL	100	339 (736)	76 (171)
CANTRAL	200	21 (112)	21 (112)
CANTRAL	400	21 (112)	21 (116)
CUBIC	400	14 (36)	12 (32)
MIELE	400	223 (907)	478 (1093)
TOTAL		1984 (5011)	1440 (3567)

Percentage performance of the new algorithm against the standard CG-algorithm

Tools	Standard –CG	NEW
NOI	100	72.58
NOF	100	71.18

4. Conclusion: This paper contains a new form for the gradient estimations with the use of inexact line search and error-terms. The new algorithm satisfies the quadratic termination property with ELS and it is very promising for the general functions.

Appendix

1.*Cubic function*:

$$f(x) = 100(x_2 - x_1^3)^2 + (1 - x_1)^2$$

Starting point: $(-1.2, 1, -1.2, 1, ...)^T$

2. *Non – diagonal function*:

$$f(x) = \sum_{i=1}^{n/2} (100(x_i - x_i^3)^2 + (1 - x_i)^2)$$

Starting point: $(-1, ...)^T$

3.*Generalized Cantreal function*:

$$f(x) = \sum_{i=1}^{n/4} (\exp(x_{4i-3}) - x_{4i-2}^3)^2 + 100(x_{4i-2} - x_{4i-1})^6 + ((a\tan(x_{4i-1} - x_{4i})))^4 + x^8{}_{4i-3}$$

Starting point: $(1, 2, 2, 2)^T$

4. *Generalized powell function*:

$$f(x) = \sum (x_{4i-9} - 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-9} - x_{4i})^4$$

Starting point: $(3, 1, 0, 1, ...)^T$

5. *Generalized Miele function*:

$$f(x) = \exp(x_1 - x_2) + 100(x_2 - x_3)^6 + ((\tan(x_3 - x_4)))^4 + x^8{}_1 + x_{4-1}^2$$

Starting point: $(1, 2, 2, 2)^T$

6. *Dixon function*:

$$f(x) = (1 - x_1)^2 + (1 - x_0)^2 + \sum_{i=2}^{9} (x_i - x_{i-1})$$

Starting point: $(-1, ...)^T$

7.*Welfe function*:

$$f(x) = (-x_1(3 - x_1/2) + 2x_2 - 1)^2 + \sum_{i=1}^{n-1} (x_{i-1} - x_i(3 - x_i/2) + 2x_{i+1} - 1)^2 + (x_{n+1} - x_n(3x_n/2) - 1)^2$$

Starting point: $(-1, ...)^T$

REFERENCES

[1] AL-Bayati A.Y. and AL-Assady N.H., (1986), **" Conjugate Gradient Methods ",** Technical Research report, No.(1),School of computer studies, Leeds University ,U.K..

[2] AL-Bayati A.Y. and AL-Assady N.H.,(1996), **" An Investigation in the Restarting Technique "** , AL-Rafidin Journal of Science, Vol.6,pp.22-27.

[3] Beale E.M.L ,(1988) , "**Introduction to optimization " ,** Wiley inter science series in discrete mathematics and optimization

[4] Dixon L.C.w., (1975) , **" Conjugate gradient algorithms quadratic termination without liner searches " ,** Journal of the institute of mathematical and its applications , 15.

[5] Fletcher R. and Reeves C., (1964), **" Function minimization by Conjugate gradient " ,** Computer journal , 7, pp 147-154 .

[6] Hestenes M. and stiefel S. (1952), **" Solving linear system " ,** journal of researches of national of standard , 49, pp. 409-436 .

[7] Polak E. and Ribiere R. (1969), **" Computational method in optimization avnified approach "** Academia press , New Yourk .

[8] Nazareth L.,(1977),"**Conjugate direction algorithm without line searches"**,JOTA,23.

[9] Nazareth L. AND Nocedal J., (1978), **" Properties of conjugate gradient algorithm with inexact line searches",** Report 78-1,systems optimization laboratory,Dept.,O.R.,stanford university.

[10] Powell M.J. (1977),"**Restart procdures for the Conjugate gradient methods",** Mathematical progrmming.12.

[11] Sloboda F., (1982),"**An imperfect Conjugate gradient algorithm",** Aplikace Mathematical, 27.

[12] Shareef S. G., (2005) "On optimally conditional Quasi-Newtion updates .unpublished Thesis of Ph.D.

[13] Touati A. AND Story, (1990)," **Some recent research conjugate gradient method for un constrained optimization "**, J. Fac. Sc. U.A.E. Univ. Val.10 pp.123-131.