

New CG-Algorithms for the nonquadratic model

Abbas Y. Al-Bayati* **Sudaba A. Mohamed**** **Basim A. Hassan***
* *College Of Computers Sciences and Mathematics -Mosul University*
** *College Of Science- Kirkuk University*

Abstract

In this paper we have investigate a new class of conjugate gradient algorithms for unconstrained non- linear optimization which employ inexact line searches and designed for general use. Some theoretical results are investigated which ensure the local convergence of the new proposed algorithms and they compared numerically with the standard HS-CG algorithm(Hestenes & Stiefel ,1952) using a number of test functions for the dimensions between 2 to 400 with some promising numerical results .

Introduction

Conjugate Gradient (CG) algorithms form a class of algorithms for minimizing a general differentiable function $f(x), x \in R^n$, whose gradient $g(x)$ can be calculated, and based on the following concept of conjugacy:

If Q is a positive definite symmetric $n \times n$ matrix then the directions $d_1, d_2, d_3, \dots, d_k$ where $d_k \neq 0$ and, $k = 1, 2, 3, 4, \dots, n$, are mutually Q -conjugate if $d_k^T Q d_i = 0$ for $i \neq k$

The classical algorithm in this category, proposed by Flectcher and Reeves (Fletcher & Reeves ,1964) based on the following iteration scheme:

$$x_{i+1} = x_i + \lambda_i d_i , i = 1, 2, 3, \dots, m$$

where the scalar λ_i is the smallest positive local minimized of the one-dimensional problem

$$\min_{\lambda} f(x_i + \lambda_i d_i)$$

and d_i is a search direction generated by the equation:

$$\text{for } i = 1, 2, 3, 4, \dots, m$$

$$d_1 = -g_1$$

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

$$\beta_i = \frac{g_{i+1}^T g_{i+1}}{g_i^T g_i} . \quad (\text{AL-Bayati, 1993})$$

Conjugate gradient algorithm with inexact searches

The classical CG-algorithm just discussed are well-known to be sensitive to the exactness of the line searches and this phenomenon can destroy the global efficiency of these algorithms. The higher the dimension of the problem, the bigger is the influence of this phenomenon. In order to improve the local rate of convergence and the efficiency of the classical CG-algorithm several established algorithms, namely (Nazareth ,1977; (Dixon L.C.W. ,1975; Nazareth & Nocedal ,1987) multi-step algorithms have been proposed. They have all shown that such algorithms are able to generate conjugate directions for quadratic models without performing exact searches; they will satisfy the quadratic termination property by using an error vector. Other important algorithms of this type, developed by (Sloboda ,1980;Sloboda ,1982) retain the quadratic termination property without using of an error vector.

Corollary (1) :Metras (Ban ,2000) proposed the following relation which it was suitable for this type of algorithms with complete proof :

$$g_{i+\frac{1}{4}} = \frac{1}{4}(g_{i+1} + 3g_i) \quad .$$

General Back ground

In this section we shall give a brief description of Sloboda algorithms (Sloboda ,1980; Sloboda ,1982) and then we discuss some theoretical properties of these algorithm.

1- The Sloboda algorithm (1980):

Sloboda [Sloboda ,1980) defines a new generalized conjugate gradient algorithm for minimizing a strictly convex function of the general from $f(x) = F(q(x))$

The outline of this algorithm is as follows :

Algorithm (Sloboda ,1980)

Step 1-Set $x_0 \in R^n , d_0 = -g_0$.

Step 2-For $i = 0,1,2,\dots,n-1$ Compute $x_{i+1} = x_i + \lambda_i d_i$

where λ_i is chosen to satisfy the condition of the line search .

Step 3- Compute

$$w_i = \frac{d_i^T g_i}{d_i^T g_{i+\frac{1}{2}}}$$

where $g_{i+\frac{1}{2}} = g(x_i + 1/2\lambda d_i)$ for general function $g_{i+1} = w_i g_{i+\frac{1}{2}} - g_i$ and

Step 4-Compute the new search direction:

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

where β_i is defined (Hestenes & Stiefel , 1952) by :

$$\beta_i = \frac{y_i^T g_{i+1}}{y_i^T d_i} ; y_i = g_{i+1} - g_i$$

Step 5- Check for convergence if $\|g_{i+1}\| \leq \varepsilon$,then stop, else go to step 6.

Step 6- If $i = n + 1$, then set $i = 1$ and go to step 1. Else set $i = i + 1$, and go to step 2.

2-The Sloboda algorithm (1982):

Sloboda (Sloboda ,982) developed an algorithm which generates conjugate directions with imperfect searches and has the quadratic termination property without using an error vector. The outline of this algorithm for general function is as follows:

Algorithm (Sloboda 1982)

Step 1-Set $x_0, g_0 = g_0, d_0 = -g_0$.

Step 2-For $i = 0, 1, 2, \dots, n$ Compute $x_{i+1} = x_i + \lambda_i d_i$

where λ_i is chosen to satisfy the condition of the line search .

Step 3- If $\|g_{i+1}\| \leq \varepsilon$, stop , else go to step 5 .

Step 4- If $i = n + 1$,then set $i = 1$, and go to step 1 , else compute

$$g_{i+1} = (g_{i+1} - g_i) - \frac{(g_{i+1} - g_i)^T g_i}{g_i^T g_i} g_i \dots\dots\dots(1)$$

Step 5- If $\|g_{i+1}\| \leq \varepsilon$, go to step 1 , else set $i = i + 1$,compute

$$d_{i+1} = -g_{i+1} + \frac{(g_{i+1} - g_i)^T g_i}{(g_{i+1} - g_i)^T d_i} d_i$$

set $i = i + 1$ and go to step 3 .

New CG-Algorithms with inexact line searches

In this section, we have to present a new CG-algorithm with inexact line searches for minimizing a function $f(x)$. The new algorithm does not require any error vector and it depends on the vector g_{i+1} defined at the

point $x_{i+\frac{1}{m}}$, $m=1,2,3,\dots,k$, and hence a new g_{i+1}^* which has the property of being orthogonal to g_i . The gradient vector $g_{i+\frac{1}{m}}$ used in this algorithm has the following property:

New Lemma (4.1)

$$\text{Let } \frac{m}{2} g_{i+\frac{1}{m}} = g_{i+\frac{1}{2}} + \frac{(m-2)}{2} g_i, m=1,2,3,\dots,k \dots\dots\dots(2)$$

Then :

$$g_{i+\frac{1}{m}} = \frac{1}{m} (g_{i+1} + (m-1)g_i) \dots\dots\dots(3)$$

proof:

$$\begin{aligned} \frac{m}{2} g_{i+\frac{1}{m}} &= g_{i+\frac{1}{2}} + \frac{(m-2)}{2} g_i \\ &= \frac{1}{2} (g_{i+1} + g_i) + \frac{m}{2} g_i - g_i \\ &= \frac{1}{2} g_{i+1} + \frac{1}{2} g_i + \frac{m}{2} g_i - g_i \\ &= \frac{1}{2} g_{i+1} + \frac{m}{2} g_i - \frac{1}{2} g_i \\ mg_{i+\frac{1}{m}} &= g_{i+1} + mg_i - g_i \\ \therefore g_{i+\frac{1}{m}} &= \frac{1}{m} (g_{i+1} + (m-1)g_i) \end{aligned}$$

Thus we have proved a new relationship to the gradient vector in Sloboda CG-algorithm.

In particular, if $m=4$ then we have corollary (1)

$$g_{i+\frac{1}{4}} = \frac{1}{4} (g_{i+1} + 3g_i)$$

which was defined in (Ban, 2000) as special case from our new formula.

New (1) generalized CG-algorithm which is based on the new defined gradient vector

Step 1-Let $d_0 = -g_0$.

Step 2-For $i=0,1,2,\dots,k$ Compute $x_{i+1} = x_i + \lambda_i d_i$

Where λ_i is chosen to satisfy the condition of the line search .

Step 3- If $\|g_{i+1}\| \leq \varepsilon$, stop , else go to step 4 .

Step 4- If $i = n + 1$,then set $i = 1$, and go to step 1 , else see

$$g_{i+1}^* = (g_{i+\frac{1}{m}} - g_i) - \frac{(g_{i+\frac{1}{m}} - g_i)^T g_i}{g_i^T g_i} g_i \dots\dots\dots(4)$$

Step 5- If $\|g_{i+1}^*\| \leq \varepsilon$, set $i = i + 1$, go to step 2 , else set

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

$$\beta_i^* = \frac{g_{i+\frac{1}{m}}^T g_{i+1}^*}{\frac{1}{m}(g_{i+1} - g_i)^T d_i}$$

where $g_{i+\frac{1}{m}} - g_i = \frac{1}{m}(g_{i+1} - g_i)$
then

$$d_{i+1} = -g_{i+1}^* + \frac{g_{i+\frac{1}{m}}^T g_{i+1}^*}{\frac{1}{m}(g_{i+1} - g_i)^T d_i} d_i \dots\dots\dots(5)$$

New Lemma (4.2)

Let $f(x)$ be a convex function and the vectors d_{i+1} and $g_{i+\frac{1}{m}}$ are defined as in the new modified (2) CG-algorithm for the points x_{i+1} and $x_{i+\frac{1}{m}}$, then the following condition holds :

$$d_{i+1}^T g_{i+\frac{1}{m}} \leq 0 \dots\dots\dots(6)$$

Proof : From eq.(5) we have

$$d_{i+1} = -g_{i+1}^* + \frac{g_{i+\frac{1}{m}}^T g_{i+1}^*}{\frac{1}{m}(g_{i+1} - g_i)^T d_i} d_i$$

Let us multiply this direction by $g_{i+\frac{1}{m}}^T$, then we have :

$$g_{i+\frac{1}{m}}^T d_{i+1} = -g_{i+\frac{1}{m}}^T g_{i+1}^* + \frac{g_{i+\frac{1}{m}}^T g_{i+1}^*}{\frac{1}{m}(g_{i+1} - g_i)^T d_i} g_{i+\frac{1}{m}}^T d_i$$

$$= -g_{i+\frac{1}{m}}^T g_{i+1}^* + \frac{g_{i+\frac{1}{m}}^T g_{i+1}^*}{(g_{i+\frac{1}{m}} - g_i)^T d_i} g_{i+\frac{1}{m}}^T d_i$$

where

$$\begin{aligned} \frac{1}{m}(g_{i+1} - g_i) &= g_{i+\frac{1}{m}} - g_i \\ &= -g_{i+\frac{1}{m}}^T g_{i+1}^* \left[1 - \frac{g_{i+\frac{1}{m}}^T d_i}{(g_{i+\frac{1}{m}} - g_i)^T d_i} \right] \\ &= -g_{i+\frac{1}{m}}^T g_{i+1}^* \left[\frac{g_{i+\frac{1}{m}}^T d_i - g_i^T d_i - g_{i+\frac{1}{m}}^T d_i}{(g_{i+\frac{1}{m}} - g_i)^T d_i} \right] \\ &= -g_{i+\frac{1}{m}}^T g_{i+1}^* \left[-\frac{g_i^T d_i}{(g_{i+\frac{1}{m}} - g_i)^T d_i} \right] \end{aligned}$$

Since $g_i^T d_i < 0$ and $(g_{i+\frac{1}{m}} - g_i)^T d_i > 0$, then $-\frac{g_i^T d_i}{(g_{i+\frac{1}{m}} - g_i)^T d_i} > 0$

Now, to prove that $-g_{i+\frac{1}{m}}^T g_{i+1}^* \leq 0$, as follows from (4) we have

$$g_{i+1}^* = (g_{i+\frac{1}{m}} - g_i) - \frac{(g_{i+\frac{1}{m}} - g_i)^T g_i}{g_i^T g_i} g_i$$

Let us multiplying g_{i+1}^* by $g_{i+\frac{1}{m}}^T$ then:

$$\begin{aligned} g_{i+\frac{1}{m}}^T g_{i+1}^* &= g_{i+\frac{1}{m}}^T (g_{i+\frac{1}{m}} - g_i) - \frac{(g_{i+\frac{1}{m}} - g_i)^T g_i}{g_i^T g_i} g_i^T g_{i+\frac{1}{m}} \\ &= \left\| g_{i+\frac{1}{m}} \right\|^2 - g_i^T g_{i+\frac{1}{m}} - \frac{(g_{i+\frac{1}{m}}^T g_i - \|g_i\|^2)}{\|g_i\|^2} g_i^T g_{i+\frac{1}{m}} \\ &= \frac{\|g_i\|^2 \left\| g_{i+\frac{1}{m}} \right\|^2 - \|g_i\|^2 g_i^T g_{i+\frac{1}{m}} - (g_{i+\frac{1}{m}}^T g_i - \|g_i\|^2)}{\|g_i\|^2} g_i^T g_{i+\frac{1}{m}} \end{aligned}$$

$$= \frac{\|g_i\|^2 \left\| g_{i+\frac{1}{m}} \right\|^2 - \|g_i\|^2 g_i^T g_{i+\frac{1}{m}} - g_{i+\frac{1}{m}}^T g_i g_{i+\frac{1}{m}} - \|g_i\|^2 g_i^T g_{i+\frac{1}{m}}}{\|g_i\|^2}$$

$$= \frac{\|g_i\|^2 \left\| g_{i+\frac{1}{m}} \right\|^2 - g_{i+\frac{1}{m}}^T g_i g_{i+\frac{1}{m}}}{\|g_i\|^2}$$

Again

$$- g_{i+\frac{1}{m}}^T g_{i+1}^* = \frac{\|g_i\|^2 \left\| g_{i+\frac{1}{m}} \right\|^2}{\|g_i\|^2} + \frac{g_{i+\frac{1}{m}}^T g_i g_{i+\frac{1}{m}}}{\|g_i\|^2}$$

Since

$$g_{i+\frac{1}{m}} = \frac{1}{m} (g_{i+1} + (m-1) g_i), \text{ then } \left\| g_{i+\frac{1}{m}} \right\| \leq \frac{1}{m} \|g_{i+1}\| + \frac{(m-1)}{m} \|g_i\| \text{ from the Schwartz}$$

equality. This implies that $- g_{i+\frac{1}{m}}^T g_{i+1}^* \leq 0$. Hence we get that $d_{i+1}^T g_{i+\frac{1}{m}}^* \leq 0$.

The outline of the new (2) CG-algorithm:

Step 1-Set $x_0 \in R^n, d_0 = -g_0$.

Step 2-For $i = 0,1,2,\dots,n-1$ and $m = 8$ Compute $x_{i+1} = x_i + \lambda_i d_i$

Where λ_i is chosen to satisfy the condition of the line search .

Step 3- Compute

$$w_i = \frac{d_i^T g_i}{d_i^T g_{i+\frac{1}{m}}}$$

Where $g_{i+\frac{1}{m}} = \frac{1}{m} (g_{i+1} + (m-1)g_i)$ for general function $g_{i+1}^* = w_i g_{i+\frac{1}{m}} - g_i$

Step 4-Compute the new direction:

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

where β_i is defined by :

$$\beta_i = \frac{y_i^T g_{i+1}}{y_i^T d_i}$$

Step 5- Check for convergence if $\left\| g_{i+1} \right\| \leq \varepsilon$, then stop, else go to step 6.

Step 6-If $i = n$ or $\|g_{i+1}\| \leq \varepsilon$, then go to step 1. Else set $i = i + 1$, and go to step 2.

Numerical Results and Conclusions

The comparison involves Five well-known test functions with twenty different versions (see appendix) with different dimension (2,4,8,10,20,40,60,80,...,400). All the results are obtained using double precision on the (Pentium (4) computer) using programs written in FORTRAN.

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

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

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

The results are reported in Table (1) in terms of the numbers of function evaluations, the number of iterations, the results indicate that the new (2) and $m=8$, algorithm is more efficient. Than the standard CG-algorithm. In this a method we use the restarting criterion (Shareef, 2005) ($d_{i+1} g_{i+1} > 0$ or $N = K$).

The numerical results in Table (1) indicates that the new (2) CG-algorithm improves the standard HS-CG algorithm in about (8.98)% NOI and (7.22)% NOF respectively, for this selected test of nonlinear functions. Note that: We didn't make any numerical computations for new (1) because it was comparable with Solobodas algorithms numerically but the latter has a faster rate of convergence.

Table (1) Comparison of algorithms for $2 \leq N \leq 400$

| Test function | N | CG-algorithm | | new(2),m=8 | |
|---------------|-----|--------------|--------|------------|--------|
| | | NOI | (NOF) | NOI | (NOF) |
| DIXON | 2 | 32 | (67) | 6 | (18) |
| SHALLO | 2 | 48 | (101) | 7 | (25) |
| ROSEN | 4 | 27 | (72) | 27 | (72) |
| NON-DIAGON | 4 | 23 | (65) | 23 | (57) |
| WOOD | 40 | 48 | (101) | 45 | (95) |
| WOLFE | 40 | 47 | (95) | 47 | (95) |
| ROSEN | 60 | 22 | (55) | 21 | (54) |
| CUBIC | 60 | 11 | (32) | 11 | (32) |
| POWELL | 80 | 84 | (193) | 85 | (202) |
| WOLFE | 80 | 49 | (99) | 49 | (99) |
| NON-DIAGON | 100 | 25 | (62) | 22 | (55) |
| WOOD | 100 | 85 | (175) | 74 | (152) |
| POWELL | 100 | 113 | (264) | 114 | (275) |
| SHALLO | 200 | 6 | (17) | 6 | (19) |
| WOOD | 200 | 85 | (175) | 74 | (152) |
| WOLFE | 200 | 51 | (103) | 51 | (103) |
| POWELL | 400 | 415 | (871) | 402 | (860) |
| WOOD | 400 | 86 | (176) | 75 | (154) |
| CUBIC | 400 | 12 | (35) | 12 | (35) |
| WOLFE | 400 | 54 | (109) | 54 | (109) |
| TOTAL | | 1323 | (2867) | 1205 | (2663) |

Percentage performance of new(2) algorithm against the standard CG-algorithm

| Tools | Standard –CG | NEW |
|-------|--------------|-------|
| NOI | 100 | 91.02 |
| NOF | 100 | 92.88 |

Refrences

- AL-Bayati A. ,(1993):Anew Non quadratic Model for Unconstrained Non lineal Optimization , Mu'tah Journal for Research and studies,Vol.8, pp. 131-155.
- Ban A. Metras,(2000) :On conjugate gradient method for solving non-linear programming problems, thesis , university of Mosul .
- Dixon L.C.W., (1975): Conjugate gradient algorithms quadratic termination without linear searches”,Journal of the Instate of Mathematics and its application, Vol.15.
- Fletcher R.and Reeves C., (1964):Function Minimization conjugate gradients ,Computer Journa,Vol.1, pp. 149-154.
- Hestenes H.R. and Stiefel E., (1952) : Method of conjugate gradient for solving linear systems Journal of research of the national Bureau of standards ,Vol. 49, pp.409-436 .
- Nazareth L. (1977):Conjugate direction algorithm without line searches,JOTA, Vol. 23.
- Nazareth L and Nocedal A., (1987):Properties of conjugate gradient algorithm with inexact line searches “,Report 78-1,systems Optima. Laboratory,Dept. O.R.,stanford University.
- Sloboda F., (1980): A generalized extended CG-algorithm”, Numerical mathematics ,Vol.35, pp.223-230.
- Sloboda F., (1982): An imperfect conjugate gredient algorithm ,Aplikace Matematiky,Vol.27.
- Shareef S. G., (2005): On optimally conditional Quasi-Newton” updates .unpublished Thesis of Ph.D.

Appendix

1. *Cubic function* :

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

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

2. *Non – diagonal function* :

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

Starting point : $(-1, \dots)^T$

3. *Rosen brock function* :

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

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

4. *Generalized powell function* :

$$f(x) = \sum_{i=1}^{n/4} (x_{4i-3} - 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-1} - 2x_{4i})^2 + 10(x_{4i-9} - x_{4i})^4 + (x_{4i-2} - 2x_{4i-1} - x_{4i})^2)$$

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

5. *Shallow function* :

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

Starting point : $(-2, -2, \dots)^T$

6. *Dixon function* :

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

Starting point : $(-1, \dots)^T$

7. *Welfe function* :

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

Starting point : $(-1, \dots)^T$

8. *Wood function* :

$$f(x) = \sum_{i=1}^{n/4} 100(x_{4i-2} + x_{4i-3}^2)^2 + (1 - x_{4i-3})^2 + 90(x_{4i} - x_{4i-1}^2)^2 + (1 - x_{4i-1}^2)^2 + 1.0$$

Starting point : $(-3, -1, -3, -1, \dots)^T$

خوارزميات جديدة للتدرج المترافق في نموذج غير تربيعي

عباس يونس البياتي* و سداية عارف محمد** و باسم عباس حسن*
*كلية علوم الحاسبات والرياضيات- جامعة الموصل **كلية العلوم- جامعة كركوك

الخلاصة

في هذا البحث تم التطرق إلى صنف جديد من خوارزميات الاتجاهات المترافقة في الأمثلة اللاخطية باستعمال خط بحث غير تام وصممت للاستخدام العام . أعطيت بعض النتائج النظرية التي تؤكد التقارب الموضوعي للخوارزميات الجديدة المقترحة . كما تمت مقارنة هذه الخوارزميات مع خوارزمية (Hestenes & Stiefel) باستخدام عدد من الدوال الاختباريه وبأبعاد تتراوح من ٢ الى ٤٠٠ مع الحصول على نتائج مشجعة.