Iraqi Journal of Statistical Science (20) 2011
The Fourth Scientific Conference of the College of Computer Science & Mathematics
pp [155-163]

A Spectral Conjugate Gradient Method with Inexact line searches

Dr. Abbas Y. Al-Bayati* **Dr. Basim A. Hassan****

Abstract

In this paper, a new Spectral Conjugate Gradient (SCG) method for solving unconstrained optimization problems; based on inexact line searches is investigated. The search directions of the new CG method are always sufficient descent. The global convergence property of the proposed method has been proved. Finally, we have presented some numerical results to examine the efficiency of the proposed method.

طريقة جديدة للتدرج المترافق الطيفي مع خطوط بحث غير تامة

المستخلص

في هذا البحث تم تقصي طريقة جديدة للتدرج المترافق الطيفي لحل مسائل الأمثلية غير المقيدة والمعتمدة على خطوط بحث غير تامة. يكون اتجاه خطوط البحث لهذه الطريقة الجديدة دائماً منحدرا انحدارا كافيا. تم إثبات صفة التقرب الشامل للطريقة المقترحة. وأخيرا قدمنا بعض النتائج العددية لفحص كفاءة الطريقة المقترحة.

1. Introduction.

We consider the unconstrained optimization problem:

$$\min \left\{ f(x) \;\middle|\; x \in R^n \right\} \qquad ....(1)$$

where $f : R^n \rightarrow R$ is smooth and its gradient g is available.

Conjugate Gradient (CG) methods are very efficient for solving large-scale unconstrained optimization problems. The iterates of the CG methods are obtained by :

$$x_{k+1} = x_k + \alpha_k d_k \qquad ....(2)$$

with

*Prof. \ College of Computers Sciences and Math.\ University of Mosul
**Lecture\\ College of Computers Sciences and Math.\ University of Mosul

Received:1/10 /2011 ____________________Accepted: 21 /12 / 2011

$$d_{k+1} = \begin{cases} -g_1 & k = 1, \\ -g_{k+1} + \beta_k d_k & k > 1, \end{cases} \quad \text{..........(3)}$$

where step-size α_k is positive, $g_k = \nabla f(x_k)$ and β_k is a scalar. In addition, α_k is a step-length which is computed by carrying out some line search procedure. There are several line search rules for choosing the step-size α_k, for example, Wolfe-Powell (WP) rule and Strong Wolfe-Powell (SWP) rule.

In this paper, we have analyzed the results on convergence of the line search methods with the following line search rule **[1]**:

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta_1 \alpha_k d_k^T g_k \quad \text{..........(4)}$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \le -\delta_2 d_k^T g_k \quad \text{..........(5)}$$

with $0 < \delta_1 < \delta_2 < 1$. Some well know formulas are given as follows :

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \quad \text{..........(6)}$$

$$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{g_k^T g_k} \quad \text{..........(7)}$$

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{y_k^T d_k} \quad \text{..........(8)}$$

$$\beta_k^{LS} = -\frac{g_{k+1}^T y_k}{g_k^T d_k} \quad \text{..........(9)}$$

$$\beta_k^{CD} = -\frac{g_{k+1}^T g_{k+1}}{g_k^T d_k} \quad \text{..........(10)}$$

$$\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{y_k^T d_k} \quad \text{..........(11)}$$

Here, (FR) denotes the Fletcher and Reeves (FR) **[**4**]**, (DY) denotes the Dai and Yuan **[**2**]**, (CD) denotes the Conjugate Descent (CD) **[**3**]**, (PR) denotes the Polak and Ribiere **[**8**]**, (HS) denotes the Hestenes and Stiefel **[**5**]**, (LS) denotes the Liu and Storey **[**6**]**, here and throughout, $y_k = g_{k+1} - g_k$.

In **[7]** Matonoha et al proposed a modified Conjugate Descent (CD) method, which was denoted by (MCD) and given**:**

$$d_{k+1} = -\varphi_k^{MCD} g_{k+1} + \beta_k^{CD} d_k, \ \ d_0 = -g_0 \quad \text{..........(12)}$$

where the values φ_k^{MCD}, β_k^{CD} are determined by:

$$\varphi_k^{MCD} = \frac{y_k^T d_k}{\left|g_k^T d_k\right|} \quad , \quad \beta_k^{CD} = \frac{g_{k+1}^T g_{k+1}}{\left|g_k^T d_k\right|}. \qquad \text{..........(13)}$$

In this note, we have introduced a modification to the CG-method of the parameter φ_k defined by:

$$\varphi_k^{MMCD} = \frac{\mu\left|g_{k+1}^T d_k\right| + y_k^T d_k}{\left|g_k^T d_k\right|} \quad . \qquad \text{..........(14)}$$

Note that if we use an exact line search, our modified SCG method (called MMCD) reduces to the method of CD. However, here in this work, we have considered general nonlinear test functions with inexact line searches.

The sufficient descent and global convergence properties of our proposed CG-method will be established later on. Finally, some numerical evidence will be listed to support our findings.

2. New CG-Method with the Convergence Property.

As in [2], we assume that the continuously differentiable function $f(x)$ is bounded in the level set $L = \{x \in R^n | f(x) \le f(x_0)\}$, where x_0 is the starting point; and that $g(x)$ is Lipschitz continuous in L, that, there exists a constant $\mu > 0$ such that:

$$\|g(x_{k+1}) - g(x_k)\| \le \mu\|x_{k+1} - x_k\|, \ \forall x_{k+1}, x_k \in U . \qquad \text{..........(15)}$$

Now we present the outline of the new proposed algorithm as follows :

2.1 Outline of The New Algorithm:

Step 0 : Given $x_0 \in R^n$, $\mu = 1.1$, $\varepsilon = 0.0001$, $\delta_1 \in (0,1)$, $\delta_2 \in (0,1/2)$, $d_0 = -g_0$

Step 1 : Computing g_k ; if $\|g_k\| \le \varepsilon$, then stop ; else continue .

Step 2 : Set $\beta_k = \beta_k^{CD}$, $\varphi_k^{MMCD} = \dfrac{\mu\left|g_{k+1}^T d_k\right| + y_k^T d_k}{\left|g_k^T d_k\right|}$.

Step 3 : Set $x_{k+1} = x_k + \alpha_k d_k$, (Use SW-conditions to compute α_k)

Step 4 : Compute $d_{k+1} = -\varphi_k g_{k+1} + \beta_k d_k$,

Step 5 : If $k = n$, go to **Step 1** with new values of x_{k+1} and g_{k+1}.

Here we have to present several theoretical results and as follows:

Theorem (2.2).

Suppose that d_{k+1} is given by (12) and (14). Then, the following result:

$$g_{k+1}^T d_{k+1} \le -c\|g_{k+1}\|^2 < 0 \qquad \text{..........(16)}$$

holds of the sufficient descent property.

Proof.

Firstly, for $k = 0$, it is easy to see that (16) is true since $d_0 = -g_0$.

Secondly, assume that:

$$g_k^T d_k \le -c\|g_k\|^2 < 0 \text{ where } 0 < c < 1 \quad ....(17)$$

holds for k when $k \ge 1$. Multiplying (12) by g_{k+1}^T, we have:

$$g_{k+1}^T d_{k+1} = -\varphi_k^{MMCD}\|g_{k+1}\|^2 + \beta_k^{CD} g_{k+1}^T d_k$$

$$= -\frac{\mu|d_k^T g_{k+1}| + y_k^T d_k}{|g_k^T d_k|}\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{|g_k^T d_k|} g_{k+1}^T d_k \quad ....(18)$$

from (5) and (17), we get:

$$g_{k+1}^T d_{k+1} = -\frac{\mu|d_k^T g_{k+1}|}{|g_k^T d_k|}\|g_{k+1}\|^2 - \frac{g_{k+1}^T d_k - g_k^T d_k}{|g_k^T d_k|}\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{|g_k^T d_k|} g_{k+1}^T d_k$$

$$\le -\frac{\mu\delta_2 c\|g_k\|^2}{c\|g_k\|^2}\|g_{k+1}\|^2 - \frac{c\|g_k\|^2}{c\|g_k\|^2}\|g_{k+1}\|^2$$

$$= -\frac{\mu\delta_2 c\|g_k\|^2 + c\|g_k\|^2}{c\|g_k\|^2}\|g_{k+1}\|^2 \quad ....(19)$$

$$= -\|g_{k+1}\|^2 \; [1 + \mu\delta_2]$$

$$= -c_1\|g_{k+1}\|^2$$

$$g_{k+1}^T d_{k+1} \le -c_1\|g_{k+1}\|^2$$

This completes the proof.

Lemma (2.3) :

If the sequence $\{x_k\}$ is generated by (2) and (3), the step size α_k satisfies (4) and (5), and d_{k+1} is a descent direction, f is bounded and $g(x)$ is Lipschitz in the level set, then:

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty . \quad ....(20)$$

For the proof see **[9].**

Theorem (2.4).

If $\mu > 0$ in (14), f is bounded and $g(x)$ is Lipschitz in the level set, then our algorithm either terminates at a stationary point or $\liminf_{k\to\infty} \|g_k\| = 0.$

Proof :

If our conclusion does not hold, then there exists a real number $\varepsilon_1 > 0$ such that $\|g_{k+1}\| > \varepsilon_1$ for all $k = 1,2,3,.......$.

Squaring the both terms of $d_{k+1} + \varphi_k^{BLV} g_{k+1} = \beta_k d_k$, we get:

$$\|d_{k+1}\|^2 + (\varphi_k^{BMLV})^2\|g_{k+1}\|^2 + 2\varphi_k^{BMLV} d_{k+1}^T g_{k+1} = \beta_k^2\|d_k\|^2 \quad ....(21)$$

from (21), we get:

$$\|d_{k+1}\|^2 = \beta_k^2\|d_k\|^2 - 2\varphi_k^{BMLV} d_{k+1}^T g_{k+1} - (\varphi_k^{BMLV})^2\|g_{k+1}\|^2 \quad ....(22)$$

Dividing both sides of (22) by $g_{k+1}^T d_{k+1}$, by (16), (17) and $\|g_{k+1}\| > \varepsilon_1$, we have:

$$\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} = \left[\frac{\|g_{k+1}\|^2}{|g_k^T d_k|}\right]^2 \frac{\|d_k\|^2}{(d_{k+1}^T g_{k+1})^2} - (\varphi_k^{BMLV})^2 \frac{\|g_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} - 2\varphi_k^{BMLV} \frac{1}{d_{k+1}^T g_{k+1}}$$

$$\le \left[\frac{\|g_{k+1}\|^2}{c\|g_k\|^2}\right]^2 \frac{\|d_k\|^2}{c^2\|g_{k+1}\|^4} - (\varphi_k^{BMLV})^2 \frac{\|g_{k+1}\|^2}{c^2\|g_{k+1}\|^4} - 2\varphi_k^{BMLV} \frac{1}{c\|g_{k+1}\|^2} \qquad \text{.........(24)}$$

$$\le \left[\frac{\|g_{k+1}\|^2}{c\|g_k\|^2}\right]^2 \frac{\|d_k\|^2}{c^2\|g_{k+1}\|^4} - \left(\varphi_k^{BMLV} \frac{\|g_{k+1}\|}{c\|g_{k+1}\|^2} + \frac{1}{\|g_{k+1}\|}\right)^2 + \frac{1}{\|g_{k+1}\|^2}$$

$$\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} \le \frac{\|d_k\|^2}{c^4\|g_k\|^2} + \frac{1}{\|g_{k+1}\|^2} \le \frac{\|d_k\|^2}{c^4\|g_k\|^2} + \frac{1}{\varepsilon_1^2} \qquad \text{..........(25)}$$

Since $d_1 = -g_1$, so that:

$$\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} < \frac{\|d_1\|^2}{(d_1^T g_1)^2} + \frac{k-1}{\varepsilon_1^2} = \frac{1}{\|g_1\|^2} + \frac{k-1}{\varepsilon_1^2} < \frac{1}{\varepsilon_1^2} + \frac{k-1}{\varepsilon_1^2} = \frac{k}{\varepsilon_1^2} \qquad \text{..........(26)}$$

Thus

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} > \sum_{k=1}^{\infty} \frac{\varepsilon_1^2}{k} = \infty \qquad \text{..........(27)}$$

Which is contrary to **Theorem (2.4)**. Hence, the proof is complete.

3. Numerical Results.

In this section, we reported some numerical results obtained with the implementation of the new CG-method on a set of unconstrained optimization test problems. We have selected (10) large scale unconstrained optimization problems in extended or generalized form, for each test function we have considered numerical experiment with the number of variable n=100-1000. Using the standard Wolfe-Powell line search given in conditions (4) and (5) with $\delta_1 = 0.0001$, $\delta_2 = 0.9$ and $\mu = 1.1$. In all these cases, the stopping criteria is the $\|g_k\| \le 10^{-4}$. The programs were written in Fortran 90. The test functions were commonly used for unconstrained test problems with standard starting points and a summary of the results of these test functions was given in **Table (3.1).** We tabulate for comparison of these algorithms, the Number Of Function evaluations (NOF) and the Number Of Iterations (NOI) .

Table (3.1)

No.	n	CD-algorithm	New-algorithm	MLV -algorithm
		NOF (NOI)	NOF (NOI)	NOF (NOI)
1	100	218 (105)	367 (181)	218 (105)
	1000	2004 (1001)	2006 (1002)	1998 (998)
2	100	899 (154)	218 (107)	751 (293)
	1000	***** (****)	514 (255)	***** (****)
3	100	366 (111)	237 (105)	248 (102)
	1000	971 (357)	864 (375)	690 (279)
4	100	123 (20)	148 (25)	148 (27)
	1000	165 (24)	170 (24)	218 (33)
5	100	2220 (326)	135 (65)	278 (106)
	1000	***** (****)	352 (169)	522 (229)
6	100	101 (50)	77 (38)	99 (49)
	1000	***** (****)	101 (50)	229 (114)
7	100	2005 (302)	94 (40)	229 (103)
	1000	***** (****)	***** (****)	611 (295)
8	100	605 (105)	19 (8)	207 (101)
	1000	***** (****)	15 (6)	461 (228)
9	100	66 (14)	67 (17)	69 (14)
	1000	121 (26)	66 (21)	109 (25)
10	100	75 (37)	135 (67)	85 (42)
	1000	85 (42)	78 (38)	89 (44)
	Total	**10024 (2674)**	**4681 (2113)**	**5436 (2321)**

4. Conclusions and Discussions.

In this paper, we have proposed a new spectral CG method for solving unconstrained minimization problems. The computational experiments show that the new approach given in this paper is successful.

Table (3.1) gives a computational results of the new algorithm against the CD and MCD algorithms for convex optimization, this table indicates that the new algorithm saves about $(75-86)\%$ NOI and $(80-91)\%$ NOF, overall, against the standard CD and MCD algorithms, respectively, especially for our selected test problems. These results are shown in the following tables:

Table (3.2): Relative efficiency of the new algorithm against CD-algorithm.

Tools	NOI	NOF
CD **-algorithm**	100 %	100 %
New-algorithm	25 %	20 %

Table (3.3): Relative efficiency of the new algorithm against MCD-algorithm.

Tools	NOI	NOF
MCD -algorithm	100 %	100 %
New-algorithm	14 %	9 %

Appendix.

1. *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 : $(3,1,0,1,\ldots\ldots)^T$

2. *Generalized wood function* :

$$f(x)=\sum_{i=1}^{n/4} 4(x_{4i-2}-x_{4i-3}^2)^2+(1-x_{4i-3})^2+90(x_{4i}-x_{4i-1}^2)^2+(1-x_{4i-1})^2+ 10.1((x_{4i-2}-1)^2+(x_{4i}-1)^2+19.8((x_{4i-2}-1)+(x_{4i}-1))$$

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

3. *Miele function* :

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

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

4. Cantrell *function* :

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

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

5. *Rosenbrock function* :

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

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

6. *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: $(-1, \ldots\ldots\ldots\ldots)^T$

7. *Non-diagonal function* :

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

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

8. *Penalty* 2 *function* :

$$f(x)=e^{(x(i)-1)^2}+(x(i)^2-0.25)^2$$

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

9. *Sum of Quartics function*

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

Starting point: $(2, \ldots\ldots\ldots\ldots)^T$

10. *Beale function* :

$$f(x)=(1.5-x_1(1-x_2))^2+(2.25-x_1(1-x_2^2)^2+(2.652-x_1(1-x_2^3)^2$$

Starting point : $(0, 0, \ldots\ldots\ldots\ldots)^T$

References.

[1] Andrie, N. (2008) 'An Unconstrained Optimization Test functions collection ' Advanced Modeling and optimization, **10**, pp. 147-161.

[2] Dai, Y. and Yuan Y. (1999) 'A nonlinear conjugate gradients method with a strong global convergence property' SIAM J. Optimization, **10**, pp. 177-182.

[3] Fletcher, R. (1987) ' Practical Methods of Optimization (second edition). John Wiley and Sons, New York, USA.

[4] Fletcher, R. and Reeves C. (1964) 'Function minimization by conjugate gradients ' Computer J, **7**, pp. 149-154.

[5] Hestenes, M. R. and Stiefel E. L. (1952) 'Method of conjugate gradients for solving linear systems' J. Research Nat. Standards, **49**, pp. 409-436.

[6] Liu Y. and Storey C. (1991) ' Efficient generalized conjugate gradients algorithms ' Part 1 : Theory. J. Optimization Theory and Applications, **69**, pp. 129-137.

[7] Matonoha C., Luksan L. and Vlcek J. (2008) 'Computational experience with conjugate gradient methods for unconstrained optimization' Technical Report, No.**1038**, pp. 1-17.

[8] Polak, E. and Ribiere, G. (1969) 'Note for Convergence Direction Conjugate, Revue Françoise Informant, Reserche Opertionelle, **16**, pp. 35-43.

[9] Wei C. and Kai-rong W. (2010) ' Global convergence of a new conjugate gradient for modified Liu-Story formal ' J. of East China Normal University (Natural Science), **1**, pp. 44-51.