Iraqi Journal of Statistical Science (20) 2011
The Fourth Scientific Conference of the College of Computer Science & Mathematics
pp [192-201]

A New Type of Conjugate Gradient Method with a Sufficient Descent Property

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

Abstract

This paper presents the development and implementation of a new unconstrained optimization method; based on the inexact line searches. Our new proposed Conjugate Gradient (CG) method always produces descent search directions and has been shown to be a global convergence. Our numerical results are promising in general by implementing ten nonlinear different test functions with different dimensions.

نوع جديد من طرائق التدرج المترافق مع خاصية الانحدار الشديد

المستخلص

يقدم هذا البحث واستعمال طريقة جديدة في مجال الامثلية غير المقيدة وتطويرها تعتمد على الاتجاهات الخطية غير الدقيقة. أعطت هذه الطريقة متجهات بحث انحداري، وبينت إن لها تقارباً شاملا. وأثبتت النتائج العددية كفاءتها باستخدام عشر دوال غير خطية ذوات إبعاد مختلفة.

1. Introduction.

We consider the following unconstrained optimization problem:

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

where R^n denotes an $n-$dimensional Euclidean space and $f: R^n \to R$ is smooth and nonlinear function.

It is a well-known, CG-method which is a line search method that takes the form:

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

where d_k is a descent direction of $f(x)$ at x_k and α_k is a step-size chosen by some kind of line search method and satisfies the Strong Wolfe (SW) conditions:

***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

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta_1 \alpha_k d_k^T g_k \qquad \text{..........(3)}$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \le -\delta_2 d_k^T g_k \qquad \text{..........(4)}$$

with $0 < \delta_1 < \delta_2 < 1$. If x_k is the current iterate, we denote $f(x_k)$ by f_k, $\nabla f(x_k)$ by g_k, $\nabla f(x_k)$ by g_k, respectively. The search direction d_k is generally required to satisfy:

$$g_{k+1}^T d_{k+1} < 0, \qquad \text{..........(5)}$$

which guarantees that d_k is a descent direction of $f(x)$ at x_k [4, 8]. In order to guarantee the global convergence, we sometimes require d_k to satisfy a **sufficient** descent condition:

$$g_{k+1}^T d_{k+1} \le -c \|g_{k+1}\|^2 \qquad \text{..........(6)}$$

where c is a constant [7]. In line search methods, the well-known CG-method has the form (2) in which

$$d_{k+1} = \begin{cases} -g_0 & if\ k = 0 \\ -g_{k+1} + \beta_k d_k & if\ k > 0 \end{cases} \qquad \text{..........(7)}$$

where

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \qquad \text{..........(8)}$$

This method is called FRCG-method [5]. The above mentioned CG-method is equivalent to each other for minimizing strong convex quadratic functions under exact line searches; they have different performance when using them to minimize non-quadratic functions or using inexact line searches. For non-quadratic objective functions, the FRCG method has a global convergence property when exact line searches are used or Strong Wolfe line search [2, 3] is used.

The structure of the paper is as follows. In section (2) we modify the standard FRCG-method and show that the search direction generated by this proposed FRCG-method at each iteration satisfies the sufficient descent condition. Section (3) establishes the global convergence property for the new class of CG-methods with $|\beta_k| \le \beta_k^{FR}$. Section (4) establishes some numerical results to show the effectiveness of the proposed CG-method and Section (5) gives brief conclusions and discussions.

2. Modified Conjugate Gradient Method.

In this section, we propose a modified FRCG-method in which the parameter β_k is defined on the basis of β_k^{FR} as follows :

$$\beta_k^{MFR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} - Min\left\{ \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \frac{\mu \|g_{k+1}\|^2 (g_{k+1}^T d_k)}{\|g_k\|^4} \right\} \qquad \text{..........(9)}$$

where μ is a parameter. Now we present the outline of the new proposed method as follows :

2.1 Outline of the New Algorithm: :

Step 0 : Given $x_0 \in R^n, \varepsilon = 1*10^{-4}$, $\delta_1 \in (0,1)$, $\delta_2 \in (0,1/2)$, $d_0 = -g_0$

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

Step 2 : Set $\beta_k = \beta_k^{MFR} = \dfrac{\|g_{k+1}\|^2}{\|g_k\|^2} - Min\left\{\dfrac{\|g_{k+1}\|^2}{\|g_k\|^2}, \dfrac{\mu\|g_{k+1}\|^2(g_{k+1}^T d_k)}{\|g_k\|^4}\right\}$,

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

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

Step 5 : If $k = n$ go to **Step 1** with new values of x_{k+1} and g_{k+1}, if not, set k=k+1 and continue.

Theorem (2.2)

Consider any iterative CG-method of the form (2) and (7), where $\beta_k = \beta_k^{MFR}$. If $g_k \ne 0$ for all $k \ge 1$, then:

$$g_{k+1}^T d_{k+1} \le -c\|g_{k+1}\|^2 < 0. \qquad ....(10)$$

Proof.

Firstly, for $k = 0$, it is easy to see that (10) is true since $d_0 = -g_0$.
Secondly, assume that:

$$g_k^T d_k \le -c\|g_k\|^2 < 0 \text{ where } 0 < c < 1 \qquad ....(11)$$

holds for k when $k \ge 1$. Multiplying (7) by g_{k+1}^T, we have

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \left[\frac{\|g_{k+1}\|^2}{\|g_k\|^2} - Min\left\{\frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \frac{\mu\|g_{k+1}\|^2(g_{k+1}^T d_k)}{\|g_k\|^4}\right\}\right] g_{k+1}^T d_k \qquad ....(12)$$

If $\dfrac{\|g_{k+1}\|^2}{\|g_k\|^2} < \dfrac{\mu\|g_{k+1}\|^2(g_{k+1}^T d_k)}{\|g_k\|^4}$ then

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 . \qquad ....(13)$$

If $\dfrac{\|g_{k+1}\|^2}{\|g_k\|^2} > \dfrac{\mu\|g_{k+1}\|^2(g_{k+1}^T d_k)}{\|g_k\|^4}$ then

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \left[\frac{\|g_{k+1}\|^2}{\|g_k\|^2} - \frac{\mu\|g_{k+1}\|^2(g_{k+1}^T d_k)}{\|g_k\|^4}\right] g_{k+1}^T d_k$$

$$= -\|g_{k+1}\|^2\left[1 - \frac{g_{k+1}^T d_k}{\|g_k\|^2} + \frac{\mu(g_{k+1}^T d_k)^2}{\|g_k\|^4}\right] \qquad ....(14)$$

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

from (4) and (11), we get:

$$g_{k+1}^T d_{k+1} \leq -\|g_{k+1}\|^2 \left[1 - \frac{\delta_2 c \|g_k\|^2}{\|g_k\|^2} + \frac{\mu(\delta_2 c \|g_k\|^2)^2}{\|g_k\|^4} \right]$$

$$g_{k+1}^T d_{k+1} \leq -\|g_{k+1}\|^2 \left[1 - \delta_2 c + \mu(\delta_2 c)^2 \right] \qquad \text{..........(15)}$$

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

3. Global convergence

In this section, we come to study the global convergence property of the new proposed **Algorithm (2.1)**. For this, we are going to verify that **Algorithm (2.1)** is well defined. For the proof of the global convergence property, the following Assumption is needed.

Assumption (3.1)

i- The level set $L = \{x \in R^n | f(x) \leq f(x_0)\}$ is bounded.

ii- In some neighborhood U and L, $f(x)$ is continuously differentiable and its gradient id Lipschitz continuous, namely, there exists a constant $\mu > 0$ such that:

$$\|g(x_{k+1}) - g(x_k)\| \leq \mu \|x_{k+1} - x_k\|, \ \forall x_{k+1}, x_k \in U . \qquad \text{..........(16)}$$

We will see that it is possible to obtain the global convergence property if the parameter β_k is appropriately bounded in magnitude. We consider a method of the form (2) and (7), where β_k is any scalar such that:

$$|\beta_k| \leq z \beta_k^{FR} , \quad z > 1 \qquad \text{..........(17)}$$

for all $k \geq 2$, and where the step length satisfies the strong Wolfe conditions $(3)-(4)$. Note that Zoutendijk's condition holds in this case. We show that any method of the form (2) and (7) is globally convergent if β_k satisfies (17). For the details of this theorem see [9].

Theorem (3.2)

Suppose that **Assumption (3.1)** holds. Let $\{g_k\}$ and $\{d_k\}$ be generated by **Algorithm (2.1)**, then we have:

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \ . \qquad \text{..........(18)}$$

Proof :

From **Theorem (2.2)** we have $g_k^T d_k < 0$ for all $k+1$. We also have from (4) and **Assumption (3.1, ii)** that:

$$-(1-\delta_2) d_k^T g_k \leq (g_{k+1} - g_k)^T d_k \leq \mu \alpha_k \left\| d_k^{\ 2} \right\| \ . \qquad \text{..........(19)}$$

Thus:

$$\alpha_k \geq -\frac{1-\delta_2}{\mu} \frac{g_k^T d_k}{\|d_k\|^2} \ , \qquad \text{..........(20)}$$

which combining (4), we get:

$$f(x_k) - f(x_{k+1}) \geq -\delta_1 \alpha_k g_k^T d_k \geq \delta_1 \frac{1-\delta_2}{\mu} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \ . \qquad \text{..........(21)}$$

Further, from **Assumption (3.1, i)** we have from **[6]** $\{f(x_k)\}$ is a decreasing sequence and has a bound below in L, and shows $\lim_{k\to\infty} f(x_{k+1}) < \infty$, this shows:

$$\infty > f(x_1) - \lim_{k\to\infty} f(x_{k+1}) = \sum_{k\geq 1} [f(x_k) - f(x_{k+1})] \geq \delta_1 \frac{1-\delta_2}{\mu} \sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} . \ , \qquad \text{..........(22)}$$

We can conclude that (18) holds.

Theorem (3.3)

Suppose that **Assumptions (3.1)** holds. Consider any method of the form (2) and (7), where β_k satisfies (12), and where the step length satisfies the strong Wolfe conditions $(3)-(4)$ with $0 < \delta_1 < \delta_2 < 1$ then:

$$\lim_{k\Rightarrow\infty} \inf \|g_k\| = 0 \quad . \qquad \text{..........(23)}$$

Proof : See [9].

It is natural to ask if the bound $|\beta_k| \leq \beta_k^{FR}$ can be replaced by

$$|\beta_k| \leq c_2 \beta_k^{FR} \qquad \text{..........(24)}$$

where $c_2 > 1$ is some suitable constant.

This theorem suggests the following globally convergent modification of the FR method. Applying relation (24) of the parameter (9) we get:

$$\left|\beta^{MFR}\right| \leq \left|\frac{\|g_{k+1}\|^2}{\|g_k\|^2}\right| + \left|\frac{\mu \|g_{k+1}\|^2}{\|g_k\|^4}\right| \left|g_{k+1}^T d_k\right| \qquad \text{..........(25)}$$

from (4) and (10) we get:

$$\begin{aligned} \left|\beta^{MFR}\right| &\leq \left|\frac{\|g_{k+1}\|^2}{\|g_k\|^2}\right| + \left|\frac{\mu \|g_{k+1}\|^2}{\|g_k\|^4}\right| \delta_2 c \|g_k\|^2 \\ &\leq \frac{\|g_{k+1}\|^2}{\|g_k\|^2} [1 + \mu \delta_2 c] \qquad \text{..........(26)} \\ &\leq c_1 \beta^{FR} \end{aligned}$$

4. Numerical Results

In this section, we have reported some numerical results obtained with the implementation of the new **Algorithm (2.1)** 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 strong Wolfe line search condition (3) and (4) with $\delta_1 = 0.0001$ and $\delta_2 = 0.9$ In all these cases, the stopping criteria is the $\|g_k\| \le 10^{-4}$. The programs are written in Fortran 90. The test functions are commonly used for unconstrained test problems with standard starting points and a summary of the results of these test functions was given in **Table (4.1)**. We tabulate for comparison of these algorithms, the Number Of Function evaluations (NOF) and the Number Of Iterations (NOI) .

Table (4-1)

No.	n	New Algorithm (2.1) NOF (NOI)	FR Algorithm NOF (NOI)
Powell	**100**	208 (102)	209 (102)
	1000	662 (329)	679 (337)
Wood	**100**	548 (218)	864 (319)
	1000	2101 (1003)	7739 (2004)
Miele	**100**	224 (101)	212 (101)
	1000	746 (332)	804 (372)
Cantrel	**100**	151 (27)	152 (27)
	1000	163 (28)	164 (28)
Rosen	**100**	269 (103)	297 (103)
	1000	380 (159)	376 (143)
Wolfe	**100**	99 (49)	99 (49)
	1000	259 (129)	279 (139)
Sum	**100**	68 (14)	68 (14)
	1000	119 (27)	121 (27)
Penalty 2	**100**	207 (101)	207 (101)
	1000	421 (208)	463 (229)
Beale	**100**	75 (37)	75 (37)
	1000	85 (42)	85 (42)
Helical	**100**	214 (105)	234 (115)
	1000	466 (231)	1124 (560)
	Total	**7465 (3345)**	**14251 (4849)**

5. Conclusions and Discussions.

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

Table (4.1) gives a comparison between the new-algorithm and the Fletcher-Reeves (FR) algorithm for convex optimization, this table indicates, see **Table (4.2)**, that the new algorithm saves (52.38)% NOI and (68.98)% NOF, overall against the standard Fletcher-Reeves (FR) algorithm, especially for our selected group of test problems.

Table(4.2): Relative efficiency of the new **Algorithm (2.1)**

Tools	NOI	NOF
FR Algorithm	100 %	100 %
Algorithm (2.1)	47.62 %	31.02 %

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

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*

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

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

7. *Sum of Quartics function*

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

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

8. *Penalty 2 function* :

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

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

9. *Beale function* :

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

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

10. *Helical valley function*:

$$f(x) = 100((x_3 - 10\theta)^2 + (r-1)^2) + x_2^3 \quad where \quad \theta = \begin{cases} (2\Pi)^{-1}\tan(x_2/x_1) & for\ x_1 > 0 \\ 0.5 + (2\Pi)^{-1}\tan(x_2/x_1) & for\ x_1 < 0 \end{cases}$$

$$r = (x_1^2 + x_2^2)^{1/2}$$

$Starting\ point: (-1,0,0, \ldots\ldots\ldots\ldots\ldots)^T$

References

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

[2] Al-Baali, M. (1985) ' New property and global convergence of the Flecher-Reeves method with line searches ' IMA Journal of Numerical Analysis, **5**, pp. 122-124.

[3] Dai, H. Y. and Yuan, Y. (1996) ' Convergence properties of the Flecher-Reeves method ' IMA Journal of Numerical Analysis, **16**, pp. 155-164.

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

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

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

[7] Luo, Z. and Zhu, Z. (2009) ' Convergence of Liu-Storey conjugate method with nonmonotone armijo line search ' International Journal of Pure and Applied Mathematics, **51**,(2), pp. 249-256.

[8] Nocedal, J. and Wright, J.S. (1999) ' Numerical Optimization ' Springer-Verlag, New York, USA.

[9] Nocedal, J. and Gilbert, J.,G. (1992) ' Global convergence properties of conjugate gradient methods for Optimization ' SIAM J. Optimization, **2**(1), pp. 21-42.