Iraqi Journal of Statistical Sciences (26) 2014
PP[12-24]

A New sufficient descent Conjugate Gradient Method for Nonlinear Optimization

Dr. Basim A. Hassan * Omer M. Esmaeel **

Abstract

In this paper, a new conjugate gradient method based on exact step size which produces sufficient descent search direction at every iteration is introduced. We prove its global convergence, and give some results to illustrate its efficiency by comparing with the Polak and Ribiere method.

طريقة جديدة للتدرج المترافق ذات الانحدار الكافي في الأمثلية اللاخطية

الملخص .

في هذا البحث قدمت طريقة جديدة للتدرج المترافق المعتمدة على طول الخطوة المضبوطة وقد أثبتت الطريقة أن لها إتجاه بحث ذي إنحدار كاف عند كل تكرار. كما اثبتت التقارب الشامل، فضلا عن إعطاء بعض النتائج العددية لتوضيح كفاءتها مقارنة بطريقة بولاك و ريبي .

* *Assistant Prof / Department of Math / College of Computer Science and Mathematics / University of Mosul.*
** *Researcher / Department of Math / College of Computer Science and Mathematics / University of Mosul.*

Received Date 9/9/2013 ——— Accept Date 24/11/2013

1. Introduction

We will refer to the problem

$$\text{minimize } f(x), \ x \in R^n \quad \text{..........(1)}$$

where $f : R^n \to R$ is a smooth function with a continuous gradient $g : R^n \to R^n$, which is assumed to be available.

In connection with problem (1) we consider conjugate gradient algorithms of the form

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

with

$$d_{k+1} = -g_{k+1} + \beta_k d_k \quad \text{..........(3)}$$

where x_0 is a given initial point, α_k is the steplength along d_{k+1} and β_k is suitable scalar. When $f(x)$ is a strictly convex quadratic function, that is when

$$f(x) = \frac{1}{2} x^T G x + cx \quad \text{..........(4)}$$

where G is a symmetric positive definite matrix, algorithm $(2)-(3)$ is uniquely defined by computing α_k as the one dimensional minimize of $f(x_k + \alpha_k d_k)$, that is

$$\alpha_k = \frac{-g_k^T d_k}{d_k^T G d_k} \quad \text{..........(5)}$$

and by letting

$$\beta_{k+1}^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \ . \quad \text{..........(6)}$$

Algorithm $(2)-(3)$, with α_k and β_k defined as in $(5)-(6)$ is the well known conjugate gradient method of Hestenes and Stiefel (HS) [2], which determines the minimizer of (4) in n iterations at most. More details can be found in **[6]**.

Various extensions to the general (non quadratic) case have been proposed, by replacing (5) with a one dimensional search and by deriving formulas for the computation of β_k that do not contain explicitly the Hessian matrix of f, but reduce to (6) when f is quadratic. The best known formulas are the Fletcher-Reeves (FR) [3] formula

$$\beta_{k+1}^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \qquad \text{..........(7)}$$

and the Polak-Ribiere-Polyak (PRP) [12] formula

$$\beta_{k+1}^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2} \qquad \text{..........(8)}$$

but other different formulas can be considered (see, [4], [5] and [13])

The global convergence properties of the FR, PRP and HS methods without regular restarts have been studied by many researchers, including Zoutendijk [7], Al-Baali [8] and Gilbert and Nocedal [9]. The conjugate gradient method with regular restart was stated in [10]. To establish the convergence results of these methods, it is usually required that steplength α_k should satisfy the strong Wolf conditions:

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k \qquad \text{..........(9)}$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq \sigma \left| g_k^T d_k \right| \qquad \text{..........(10)}$$

where $0 < \delta < \sigma < 1$. On the other hand, many other numerical methods for unconstrained optimization are proved to be convergent under the standard Wolfe conditions:

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k \qquad \text{..........(11)}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k \qquad \text{..........(12)}$$

For example, see Nocedal and Wright [10].

The paper is organized as follows. In section (1) is the introduction. In section (2) we present the new formula β_k^{New} and the descent algorithm. Section (3) shows that the search direction generated by this proposed algorithm at each iteration satisfies the sufficient descent condition. Section (4) establishes the global convergence analysis for uniformly convex function property for the new formula β_k^{New}. Section (5) establishes some numerical results to show the effectiveness of the proposed CG-method and Section (6) gives a brief conclusions and discussions.

2. A New Conjugate Gradient Method :

In this section, we derive a new conjugate gradient method based on steplength which is defined in (5). From (5) and (2), we get :

$$x_{k+1} - x_k = -\frac{g_k^T d_k}{d_k^T G d_k} d_k$$
$$v_k = -\frac{g_k^T d_k}{d_k^T G d_k} d_k \qquad \text{..........(13)}$$

Multiplying (13) by s_k^T where $s_k \in R^n$ is any vector such that $s_k^T v_k \neq 0$ we get :

$$s_k^T v_k = -\frac{g_k^T d_k}{d_k^T G d_k}(s_k^T d_k) \qquad \text{..........(14)}$$
$$d_k^T G d_k \,(s_k^T v_k) = -g_k^T d_k \,(s_k^T d_k)$$

Now assume that we want a matrix $G = \delta_{k+1}\, I$, and which satisfies $\delta_{k+1} v_k = y_k$.

from $G = \delta_{k+1}\, I$, and (14) we get

$$d_k^T G d_k (s_k^T v_k) = -g_k^T d_k (s_k^T d_k)$$
$$d_k^T \delta_{k+1} I_k d_k (s_k^T v_k) = -g_k^T d_k (s_k^T d_k)$$
$$d_k^T \delta_{k+1} d_k (s_k^T v_k) = -g_k^T d_k (s_k^T d_k)$$
$$\delta_{k+1} d_k^T d_k (s_k^T v_k) = -g_k^T d_k (s_k^T d_k)$$
$$\delta_{k+1} = -\frac{g_k^T d_k (s_k^T d_k)}{d_k^T d_k (s_k^T v_k)} \qquad \text{..........(15)}$$

This formula defines the most popular Barzilai-Borwin method [14]. Namely method for unconstrained minimization is of the form (2), at each iteration,

$$d_{k+1} = -\frac{1}{\delta_{k+1}}\, g_{k+1} \;. \qquad \text{..........(16)}$$

Whereas in the case of the conjugate gradient (CG) method, we have $d_k = -g_k + \beta_k d_{k-1}$ and thus :

$$\delta_{k+1}^{-1} = \frac{d_k^T d_k (s_k^T v_k)}{g_k^T g_k (s_k^T d_k)} \qquad \text{..........(17)}$$

For the new algorithm, we implemented numerical calculations for s_k with different of the vector, for example $s_k = d_k$. Then the direction $d_{k+1} = -\delta_{k+1}^{-1}\ g_{k+1}$ can be written as :

$$d_{k+1} = -\left(\frac{d_k^T v_k}{g_k^T g_k}\right) g_{k+1} \qquad \text{..........(18)}$$

Since Newton direction are conjugate gradient with exact line searches we get :

$$-\left(\frac{d_k^T v_k}{g_k^T g_k}\right) g_{k+1}^T y_k = -g_{k+1}^T y_k + \beta_k d_k^T y_k \qquad \text{..........(19)}$$

then we have

$$\beta_k = \left(1 - \frac{d_k^T v_k}{g_k^T g_k}\right) \frac{g_{k+1}^T y_k}{d_k^T y_k} \qquad \text{..........(20)}$$

$$d_{k+1} = -g_{k+1} + \left(1 - \frac{d_k^T v_k}{g_k^T g_k}\right) \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k \qquad \text{..........(21)}$$

Now we can obtain the new descent conjugate gradient algorithms, as follows :

The Descent Algorithm

Step 1. Initialization: Select $x_1 \in R^n$ and the parameters $0 < \delta_1 < \delta_2 < 1$. Compute $f(x_1)$ and g_1. Consider $d_1 = -g_1$ and set the initial guess $\alpha_1 = 1/\|g_1\|$.

Step 2.Test for continuation of iterations. If $\|g_{k+1}\| \le 10^{-6}$, then stop. else step3.

Step 3. Line search: Compute $\alpha_{k+1} > 0$ satisfying the Wolfe line search condition (11) and (12) and update the variables $x_{k+1} = x_k + \alpha_k d_k$.

Step 4. β_k conjugate gradient parameter which defined in (20).

Step 5. Direction computation. Compute $d_{k+1} = -g_{k+1} + \beta_k d_k$. If the restart criterion of Powell $\left|g_{k+1}^T g_k\right| \ge 0.2\,\|g_{k+1}\|^2$, is satisfied, then set $d_{k+1} = -g_{k+1}$ otherwise define $d_{k+1} = d$. Compute the initial guess $\alpha_k = \alpha_{k-1}\|d_{k-1}\| / \|d_k\|$, set $k = k+1$ go to with step2 .

3. The Sufficient Descent Property :

Below we have to show the sufficient descent property for our proposed new conjugate gradient methods, denoted by β_k^{New} . For the sufficient descent property to be hold :

$$g_{k+1}^T d_{k+1} \le -c\|g_{k+1}\|^2 \quad \text{for} \quad k \ge 0 \text{ and } c > 0 \qquad ....(22)$$

Assumption(1):

Assume f is bounded below in the level set $S = \{x \in R^n : f(x) \le f(x_\circ)\}$; In some
neighborhood N of S, f is continuously differentiable and its gradient is
Lipshitz continuous, there exist $L > 0$ such that :

$$\|g(x) - g(y)\| \le L\|x - y\| \quad \forall \text{x, y} \in \text{N} \ . \qquad ....(23)$$

More details can be found in **[13]**.

Theorem (3.1) :

If $\frac{d_k^T v_k}{g_k^T g_k} = \mu \ge 1$ then the search direction (3) and β_k^{New} given in equation (20), with condition (22) will hold for all $k \ge 1$.

Proof :

Since $d_0 = -g_0$, we have $g_0^T d_0 = -\|g_0\|^2$, which satisfy (22). Multiplying (21) by g_{k+1}, we have

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \left[1 - \frac{d_k^T v_k}{g_k^T g_k}\right] \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k^T g_{k+1} \qquad ....(24)$$

$$= -\|g_{k+1}\|^2 + [1 - \mu] \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k^T g_{k+1} \qquad ....(25)$$

yielding

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + (1 - \mu) \frac{g_{k+1}^T y_k}{v_k^T y_k} v_k^T g_{k+1} \qquad ....(26)$$

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + (1-\mu)\frac{g_{k+1}^T y_k}{(v_k^T y_k)^2}(v_k^T y_k)v_k^T g_{k+1} \qquad \text{..........(27)}$$

Applying the inequality $w^T v \le \frac{1}{2}(\|w\|^2 + \|v\|^2)$ to the second term of the right hand side of the above equality, with $w = (y_k^T v_k)g_{k+1}$ and $v = (g_{k+1}^T v_k)y_k$ we get :

$$d_{k+1}^T g_{k+1} \le -\|g_{k+1}\|^2 + \frac{(1-\mu)}{(v_k^T y_k)^2}\left(\frac{1}{2}\left[\|g_{k+1}\|^2 (y_k^T v_k)^2 + (g_{k+1}^T v_k)^2 (\|y_k\|^2)\right]\right) \qquad \text{..........(28)}$$

$$\begin{aligned} d_{k+1}^T g_{k+1} &\le \left[\frac{(1-\mu)}{2} - 1\right]\|g_{k+1}\|^2 + \frac{(1-\mu)}{2(v_k^T y_k)^2}(g_{k+1}^T v_k)^2 \|y_k\|^2 \\ &\le \left[\frac{1}{2} - \frac{\mu}{2} - 1\right]\|g_{k+1}\|^2 + \frac{(1-\mu)}{2(v_k^T y_k)^2}(g_{k+1}^T v_k)^2 \|y_k\|^2 \end{aligned} \qquad \text{..........(29)}$$

from (29) we get :

$$\begin{aligned} d_{k+1}^T g_{k+1} &\le \left[-\frac{1}{2} - \frac{\mu}{2}\right]\|g_{k+1}\|^2 + \frac{(1-\mu)}{2(v_k^T y_k)^2}(g_{k+1}^T v_k)^2 \|y_k\|^2 \\ &\le -\left[\frac{1}{2} + \frac{\mu}{2}\right]\|g_{k+1}\|^2 + \frac{(1-\mu)}{2(v_k^T y_k)^2}(g_{k+1}^T v_k)^2 \|y_k\|^2 \end{aligned} \qquad \text{..........(30)}$$

$$\le -\|g_{k+1}\|^2\left(\frac{1}{2} + \frac{\mu}{2}\right) + \frac{(g_{k+1}^T s_k)^2}{(s_k^T y_k)^2}\left(\frac{1}{2}(1-\mu)\|y_k\|^2\right) \qquad \text{..........(31)}$$

Therefore, when $\frac{1}{2} + \frac{\mu}{2} > 0$ and $1-\mu < 0$, we get

$$\begin{aligned} d_{k+1}^T g_{k+1} &\le -\left(\frac{1}{2} + \frac{\mu}{2}\right)\|g_{k+1}\|^2 \\ &\le -c\|g_{k+1}\|^2 \end{aligned} \qquad \text{..........(32)}$$

where

$$c = \frac{1}{2} + \frac{\mu}{2} \cdot \qquad \text{..........(33)}$$

4. Convergence analysis for uniformly convex function :

Next we will show that CG method with β_k^{New} converges globally. We study the convergence of suggested methods using uniformly convex function, then there exists a constant $\eta > 0$ such that

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \eta \|x - y\|^2 \text{ , for any } x, y \in S \qquad ....(34)$$

or equivalently

$$y_k^T s_k \geq \eta \|s_k\|^2 \quad \text{and} \quad \eta \|s_k\|^2 \leq y_k^T s_k \leq L \|s_k\|^2 \qquad ....(35)$$

On the other hand, under Assumption(1), It is clear that there exist positive constants B, such

$$\|x\| \leq B \text{ , } \forall x \in S \qquad ....(36)$$

Proposition:

Under Assumption1 and equation (36) on f, there exists a constant $\bar{\gamma} > 0$ such

that

$$\|\nabla f(x)\| \leq \bar{\gamma} \text{ , } \forall x \in S \qquad ....(37)$$

Lemma(1):

Suppose that Assumption(1) and equation (36) hold. Consider any conjugate gradient method in from (2) and (3), where d_k is a descent direction and α_k is obtained by the Wolfe line search. If

$$\sum_{k>1} \frac{1}{\|d_{k+1}\|^2} = \infty \qquad ....(38)$$

then we have

$$\lim_{k \to \infty} \left(\inf \|g_k\| \right) = 0 . \qquad ....(39)$$

More details can be found in **[11]**.

Theorem (4.1):

Suppose that Assumption (1) and equation (36) and the descent condition hold. Consider a conjugate gradient method in the form $(2)-(3)$ with β_k^{New} as in

(20), where α_k is computed from Wolf line search condition (11) and (12). If the objective function is uniformly convex on S, then $\lim\limits_{k\to\infty}\inf\|g_k\|=0$.

Proof :

Firstly, we need simplify our new β_k^{New}, So that our convergence proof will be much easier. Subsisting (20) into (21), we obtain :

$$\begin{aligned}\|\mathrm{d}_{k+1}\| &= \left\| -g_{k+1}+(1-\mu)\frac{g_{k+1}^T y_k}{d_k^T y_k}d_k\right\| \\ &= \left\| -g_{k+1}+\frac{g_{k+1}^T y_k}{d_k^T y_k}d_k-\mu\frac{g_{k+1}^T y_k}{d_k^T y_k}d_k\right\| \qquad \text{..........(40)} \\ &= \left\| -g_{k+1}+\frac{g_{k+1}^T y_k}{y_k^T s_k}s_k-\mu\frac{g_{k+1}^T y_k}{y_k^T s_k}s_k\right\|\end{aligned}$$

$$\begin{aligned}\|\mathrm{d}_{k+1}\| &\le \|g_{k+1}\|+\frac{\|g_{k+1}\|L\|s\|^2}{\eta\|s\|^2}-\mu\frac{\|g_{k+1}\|L\|s\|^2}{\eta\|s\|^2} \qquad \text{..........(41)} \\ &\le \|g_{k+1}\|\left(1+\frac{L}{\eta}-\mu\frac{L}{\eta}\right)\end{aligned}$$

$$\|\mathrm{d}_{k+1}\| \le \left(\frac{\eta+(1-\mu)L}{\eta}\right)\bar{\gamma} \qquad \text{..........(42)}$$

This relation shows that

$$\sum_{k\ge1}\frac{1}{\|d_k\|^2}\ge\left(\frac{\eta}{\eta+(1-\mu)L}\right)^2\frac{1}{\bar{\gamma}^2}\sum_{k\ge1}1=\infty \qquad \text{..........(43)}$$

Therefore, from **Lemma 1** we have $\lim\limits_{k\to\infty}(\inf\|g_k\|)=0$, which for uniformly convex function equivalent to $\lim\limits_{k\to\infty}\|g_k\|=0$.

5. Numerical Results :

In this section, we reported some numerical results obtained with the implementation of the new methods on a set of unconstrained optimization test problems taken from (Andrie, 2008) **[1]**.

We selected (15) large scale unconstrained optimization test problems. For each test function we have considered 10 numerical experiments with number of variables n=100,1000. We use $\delta_1 = 10^{-4}$ and $\delta_2 = 0.9$ in the line search routine $(3)-(4)$. All these methods terminate when the following stopping criterion is met $\|g_{k+1}\| \le 10^{-6}$.

All codes are written in double precision FORTRAN Language with F90 default compiler settings. We record the number of iterations calls (NOI), and the number of restart calls (IRS) for the purpose our comparisons. If NOI exceeded 2000 then denote **F***.

Table (5.1) Comparison of the algorithms for $n = 100$

Test Problems	PR - algorithm		New algorithm	
	NOI	IRS	NOI	IRS
Extended White and Holst	38	16	31	17
Extended Beale	47	26	14	7
DENSCHNC (CUTE)	17	8	18	11
Diagonal 3	167	107	147	89
Extended Tridiagonal 1	21	9	26	13
Extended Maratos	94	34	65	30
Extended Wood	81	37	55	17
Extended Quadratic Penalty	35	13	31	13
ARWHEAD (CUTE)	10	5	11	7
Partial Perturbed Quadratic	85	28	83	24
LIARWHD (CUTE)	25	13	21	12
DENSCHNA (CUTE)	23	13	21	10
DENSCHNF (CUTE)	22	19	21	18
Extended Block-Diagnal	122	62	18	10
Generalized Quadratic GQ2	38	12	37	13
	825	402	599	291

Table (5.2) Comparison of the algorithms for $n = 1000$

Test Problems	PR algorithm		New algorithm	
	NOI	IRS	NOI	IRS
Extended White and Holst	348	317	34	20
Extended Beale	38	19	16	10
DENSCHNC (CUTE)	128	66	13	9
Diagonal 3	F*	F*	1486	1324
Extended Tridiagonal 1	45	21	32	16
Extended Maratos	98	36	77	36
Extended Wood	73	35	64	22
Extended Quadratic Penalty	51	20	37	20
ARWHEAD (CUTE)	39	23	9	7
Partial Perturbed Quadratic	506	264	330	76
LIARWHD (CUTE)	48	33	21	12
DENSCHNA (CUTE)	25	14	19	11
DENSCHNF (CUTE)	23	20	22	19
Extended Block-Diagnal	130	66	16	9
Generalized Quadratic GQ2	112	55	39	14
	1664	989	729	281

6. Conclusions and Discussions :

In this paper, we have proposed a new nonlinear CG- algorithms based on steplength defined by (20) under some assumptions the new algorithm has been shown to be globlly convergent for uniformly convex, functions and satisfies the sufficient descent property. The computational experiments show that the new kinds given in this paper are successful .

Table (5.1) gives a comparison between the new-algorithm and the Polak-Ribiere (PR) algorithm for convex optimization, this table indicates, see **Table (6.1)**, that the new algorithm saves (53.35)% NOI and (41.12)% IRS, overall against the standard Polak-Ribiere (PR) algorithm, especially for our selected group of test problems.

Table(6.1): Relative efficiency of the new Algorithm

Tools	NOI	IRS
PR Algorithm	100 %	100 %
New Algorithm	46.65 %	58.88 %

References

[1] Andrei N. (2008). An Unconstrained Optimization test function collection. Adv. Model. Optimization . 10. pp.147-161.

[2] M. R. Hestenes and E. Stiefel, (1952). Methods of conjugate gradients for solving linear systems, Journal of Research of National Burean of Standard,49, pp. 409-436.

[3] R. Fletcher and C. M. Reeves,(1964). Funtion minimization by conjagate gradients, Computer Journal 7, pp. 149-154.

[4] R. Fletcher, (1989). Practical Method of Optimization (2nd Edition), John Wiley & Sons, New York .

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

[6] Lucidi S, and Grippe L. (1995) ' A global convergence vaesion of the Polak and Ribiere conjugate gradient method ' Dipartimento di Information Sistemistica, Universite di Roma ' La Sapienza ' , Roma, Italy, pp. 1-15.

[7] G. Zoutendijk, 1970. Nonlinear Porgramming, computational methods, in: Integr and Nonlinear Programming, North-Holland, Amsterdam, pp. 37-86.

[8] M. Al-Baali, (1985). Descent property and global convergence of the Flecher-Reeves method with inexact line search, IMA Journal of Numerical Analysis 5, pp. 121-124.

[9] J. C. Gilbert and J. Nocedal, (1992). Global convergence properties of conjugate gradient methods for optimization, SIAM Journal on Optimization 2, pp. 21-42.

[10] J. Nocedal and S. J. Wright, (1999). Numerical Optimization, Springer Series in Operations Research, Springer Verlag, New York.

[12] E. Polak and G. Ribiere, (1969). Note sur la Convergence de Directions Conjugate, Revue Francaise Informant, Reserche. Opertionelle16, pp. 35-43.

[13] Y. H. Dai and Y. Yuan, (1999). A nonlinear conjugate gradient method with a strong global convergence property, SIAM J. optimization, pp. 177-182.

[14] M. Raydan, (1997), The Barzilai and Borwein, conjugate gradient for the large scale unconstrained minimization problem, SIAM J. on optimization, 6, pp.26-33.