Research Article

Two Versions of the Spectral Nonlinear Conjugate Gradient Method

Basim A. Hassan, Haneen A. Alashoor
Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, IRAQ
*Correspondent Author Email: basimabas39@gnail.com

Article Info

Received
23/Mar./2017

Accepted
8/Nov./2017

Abstract
The nonlinear conjugate gradient method is widely used to solve unconstrained optimization problems. In this paper the development of different versions of nonlinear conjugate gradient methods with global convergence properties proved. Numerical results indicated that the proposed method is very efficient.

Keywords: Conjugate gradient, Spectral conjugate gradient, Descent condition, Global convergence, Numerical results.

الخلاصة
تستخدم طريقة التدرج المترافق بشكل واسع في حل مسائل الأمثلية غير المقيدة. تم في هذا البحث, تطوير النسخ المختلفة لطرق التدرج المترافق غير الخطية بخصائص تقارب شامل تم إثباتها. أشارت النتائج العددية بان الطرق المقترحة كفوءة جدا.

Introduction

Conjugate gradient methods are a class of very important methods for minimizing smooth functions, especially when the dimension is large [7]. They are considered to be conjugate direction methods which lie between the method of steepest descent and Newton's method. Their principal advantage is that they don't require the storage of any matrices as in Newton's method, or as in quasi-Newton methods, and they are designed to converge faster than the steepest descent method. Conjugate gradient methods converge in at most n iterations for unconstrained quadratic optimization problems in R^n when using exact line searches.In this paper, we consider the following unconstrained optimization problem:

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

where f is smooth and its gradient g is available. Conjugate gradient methods are very efficient for solving large-scale unconstrained optimization problems (1). The iterates of conjugate gradient methods are obtained by :

$$x_{k+1} = x_k + \alpha_k d_k \ , \quad k = 0,1,2,....... \qquad (2)$$

where d_k is search direction and α_k is a positive scalar and called the step length. Line search in the conjugate gradient algorithms often is based on the standard Wolfe (SW) conditions.

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k \qquad (3)$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k \qquad (4)$$

where $0 < \delta < \sigma$ .More performance profile, is given in [4].

The Conjugate gradient method generates a search direction that is mutually conjugate to the previous search directions, with respect to a given positive definite matrix H , and finds the optimal point in that direction using a line search technique. Two search directions d_{i+1} and d_{j+1} are said to be mutually conjugate with respect to H if the following condition is satisfied:

$$d_{i+1} H d_{j+1} = 0 \quad , i \neq j \qquad (5)$$

In other words, the next search direction is calculated as a linear combination of the previous direction and the current gradient, in such a way that the minimization steps in all previous directions are not interfered with. The search direction can be determined as follows :

$$d_{k+1} = -g_{k+1} + \beta_k d_k \qquad (6)$$

where β_k is a parameter to be determined so that d_{k+1} becomes the $k+1^{th}$ conjugate direction. There are various ways for computing β_k .The most well known conjugate gradient methods include the Fletcher-Reeves (FR) method [5], the Hestenes-Stiefel (HS) method [6] and the Polak-Ribi`ere (PR) method [8]. The update parameters of these methods are respectively specified as follows :

$$\beta_{k+1}^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k},\ \beta_{k+1}^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2},\ \beta_{k+1}^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}. \quad (7)$$

In [2,3], Basim et al. proposed a new nonlinear conjugate gradient methods called the **BSQ** and **BSI** methods with the parameters β_k given by :

$$\beta_k^{BSQ} = \frac{g_{k+1}^T g_{k+1}}{\xi_{k+1}}, \quad \beta_k^{BSI} = \frac{g_{k+1}^T g_{k+1}}{\delta_{k+1} d_k^T v_k} \quad (8)$$

where

$$\xi_{k+1} = \alpha_k (g_k^T d_k)^2 / 2(f_k - f_{k+1}),\ \delta_{k+1} = \|y_k\| / \|v_k\|. \quad (9)$$

In fact, utilizing (6), β_k^{BSQ} and β_k^{BSI} may be rewritten as :

$$0 < \beta_k \le \frac{-g_{k+1}^T d_{k+1}}{-g_k^T d_k}. \quad (10)$$

Zhang et al. [10] proposed a modified FR method (called MFR), in which the direction d_{k+1} is defined by

$$d_{k+1} = -\theta_k^{MFR} g_{k+1} + \beta_k^{FR} d_k, \quad (11)$$

where

$$\theta_k^{MFR} = \frac{y_k^T d_k}{g_k^T g_k}. \quad (12)$$

and $y_k = g_{k+1} - g_k$. Based on the idea of them, we propose a new spectral conjugate gradient method without line search.

In the next section, we present the spectral **BSQ** and **BSI** conjugate gradient methods. In section 3, Some mild conditions are also given and the global convergence will be given. Some numerical results are reported in last section.

Two versions of the spectral conjugate gradient method :

Conjugate gradient algorithm (1) and (5) with exact line search always satisfy the condition $g_{k+1}^T d_k = -\|g_{k+1}\|^2$ which is in a direct connection with the sufficient descent condition

$$g_{k+1}^T d_k \le -w\|g_{k+1}\|^2 \quad (13)$$

for some positive constant $w > 0$. Observe that w is an arbitrary positive constant.

In this paper, we take a little modification to the **BSQ** and **BSI** methods such that the direction generated by the modified BSQ and BSI methods is always a descent direction.

Let the iterative direction d_{k+1} satisfy the sufficient descent condition ,we only need to guarantee the following equality hold :

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

From above equation and (10) we get :

$$g_{k+1}^T d_{k+1} = \beta_k^{BSQ} g_k^T d_k \le -w\|g_{k+1}\|^2$$
$$\frac{\|g_{k+1}\|^2}{\xi_{k+1}} g_k^T d_k \le -w\|g_{k+1}\|^2 \quad (14)$$

Since

$$y_k^T d_k = g_{k+1}^T d_k - g_k^T d_k \quad (15)$$

From (15) and (14) we get :

$$-\frac{\|g_{k+1}\|^2}{\xi_{k+1}}\left[\frac{y_k^T d_k}{\xi_{k+1}}\xi_{k+1} - g_{k+1}^T d_k\right] \le -w\|g_{k+1}\|^2$$
$$-\frac{y_k^T d_k}{\xi_{k+1}}\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\xi_{k+1}} g_{k+1}^T d_k \le -w\|g_{k+1}\|^2 \quad (16)$$
$$-\left[\frac{y_k^T d_k}{\xi_{k+1}} - w\right]\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\xi_{k+1}} g_{k+1}^T d_k \le 0$$

Since, w is a constant parameter, let w be defined by :

$$w = \rho \frac{g_{k+1}^T d_k}{\xi_{k+1}} \quad (17)$$

where is constant and $\rho \in [0, 1]$ .Then, we have :

$$\theta_k^{EBSQ} = \frac{y_k^T d_k}{\xi_{k+1}} - \rho \frac{g_{k+1}^T d_k}{\xi_{k+1}}. \quad (18)$$

Thus, we obtain the following iterative direction :

$$d_{k+1} = -\left[\frac{y_k^T d_k}{\xi_{k+1}} - \rho \frac{g_{k+1}^T d_k}{\xi_{k+1}}\right] g_{k+1} + \beta_k^{BSQ} d_k. \quad (19)$$

A similar result holds for the **BSI** formula. We give the specific form of the proposed spectral conjugate gradient method as follows :

$$\theta_k^{EBSI} = \frac{y_k^T d_k}{\delta_{k+1} d_k^T v_k} - \rho \frac{g_{k+1}^T d_k}{\delta_{k+1} d_k^T v_k}. \quad (20)$$

Then, we can propose the following new spectral conjugate gradient methods (**EBSQ** and **EBSI**) :

New 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.

Step 3. Line search. Compute $\alpha_{k+1} > 0$ satisfy-ying the Wolfe line search condition (4) and (5) and update the variables $x_{k+1} = x_k + \alpha_k d_k$.

Step 4. Set $\beta_k = \beta_k^{BSQ}$ *or* β_k^{BSI} with θ_k^{EBSQ} *or* θ_k^{EBSI} respectively.

Step 5. Direction computation. Compute $d_{k+1} = -\theta_k g_{k+1} + \beta_k d_k$. If the restart criterion of Powell $|g_{k+1}^T g_k| \ge 0.2\,\|g_{k+1}\|^2$, is satisfied, then set

$d_{k+1} = -\theta_k \; g_{k+1}$ otherwise set $k = k+1$ and continue with **step2**.

The global Convergence of EBSQ

First, we will give the following assumptions on the objective function $f(x)$, which have been used often in literatures to analyze the global convergence of the conjugate gradient methods with inexact line searches.

We assume that f is strongly convex and is Lipschitz continuous on the level set :

$$L = \left\{ x \in R^n \middle| f(x) \le f(x_0) \right\} \tag{21}$$

That is, there exist constants $\mu > 0$ and $c > 0$ such that :

$$(\nabla f(x) - \nabla f(y))^T (x - y) \ge \mu \|x - y\|^2 \tag{22}$$

and

$$\|\nabla f(x) - \nabla f(y)\| \le c\|x - y\|, \tag{23}$$

for all x and y from L_0 [11].

From the definition of θ_k^{EBSQ} *or* θ_k^{EBSI} , it is easy to prove that the search direction d_{k+1} satisfy the sufficient descent condition holds with $w < \rho$.

In the following we state a lemma, often called the Zoutendijk condition, is used to prove the global convergence of the proposed algorithms. It was originally given by Zoutendijk [9,12].

Lemma (1) :

Suppose that Assumption holds. And x_k is given by the Algorithm, then :

$$\sum_{k \ge 1} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty \text{ or } \sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty . \tag{24}$$

Theorem (1) :

Suppose that assumption hold. Consider any method of the form (2) and (11), where θ_k computed by (20) and α_k satisfied the Wolfe line searches. Then,

$$\lim_{k \to \infty} \inf \|g_k\| = 0 . \tag{25}$$

Proof :

Suppose by contradiction that there exists a positive constant $\varepsilon > 0$, such that :

$$\|g_{k+1}\| > \varepsilon, \tag{26}$$

From (6), we have :

$$\|d_{k+1}\|^2 = (\beta_k^{BSQ})^2 \|d_k\|^2 - 2\theta_k g_{k+1}^T d_{k+1} - \theta_k^2 \|g_{k+1}\|^2 \tag{27}$$

From the above equation and (12), we have :

$$\|d_{k+1}\|^2 \le \left(\frac{g_{k+1}^T d_{k+1}}{g_k^T d_k} \right)^2 \|d_k\|^2 - 2\theta_k^{EBSQ} d_{k+1}^T g_{k+1} - (\theta_k^{EBSQ})^2 \|g_{k+1}\|^2 \tag{28}$$

Dividing the both inequalities by $(g_{k+1}^T d_{k+1})^2$, yields to :

$$\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} \le \frac{\|d_k\|^2}{(g_k^T d_k)^2} - (\theta_k^{EBSQ})^2 \frac{\|g_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} - 2\theta_k^{EBSQ} \frac{1}{d_{k+1}^T g_{k+1}}$$

$$\le \frac{\|d_k\|^2}{(g_k^T d_k)^2} - \left(\theta_k^{EBSQ} \frac{\|g_{k+1}\|}{c\|g_{k+1}\|^2} + \frac{1}{\|g_{k+1}\|} \right)^2 + \frac{1}{\|g_{k+1}\|^2} \tag{29}$$

$$\le \frac{\|d_k\|^2}{(g_k^T d_k)^2} + \frac{1}{\|g_{k+1}\|^2}$$

Using (32) recursively and noting that $\|d_1\|^2 = -g_1^T d_1 = \|g_1\|^2$ .Then :

$$\frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \le \sum_{i=1}^{k} \frac{1}{\|g_i\|^2} \; . \tag{30}$$

Then we get from (33) and (27) that :

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \frac{\varepsilon_1^2}{k}, \tag{31}$$

which indicates :

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \sum_{k=1}^{\infty} \frac{\varepsilon_1^2}{k} = \infty \tag{32}$$

This contradicts the Zoutendijk condition (25). Therefore the conclusion (26) holds. Its proof is similar to that of **EBSI**.

Remark 3.1

Global convergence property and descent property of **SBSI** algorithm are similar to those of **SBSQ** Algorithm..

Numerical Results

This section reports some numerical experiments. The test problems with the given initial points can be found at which were collected by Neculai Andrei. Some of the test problems are from the CUTE collection established by Bongartz, Conn, Gould and Toint [1].

The stop criterion is as follows : the program is stopped if the inequality $\|g_{k+1}\| \le 10^{-6}$ is satisfied. All codes were written in Fortran 90.

All these algorithms are implemented with the standard Wolfe line search conditions with $\delta_1 = 0.001$ and $\delta_2 = 0.9$. We tested the conjugate gradient algorithms with the following β_k **:**

1. FR : The Fletcher-Reeves method

2. BSQ with EBSQ : The β_k^{BSQ} with θ_k^{EBSQ}.

3. BSI with EBSI : The β_k^{BSI} with θ_k^{EBSI} .

Our numerical results are listed in the form NOI / IRS, where NOI denotes the number of iterations, and the number of restart IRS.

From Tables 1, and 2, we draw a conclusion that new Algorithms performs better than the FR method for the most tested problems under only Wolfe line search. Therefore, the proposed method is promising and comparable to the FR method and comparable to the FR method.

Table 1 : Comparison of methods for n= 100

Test problems	FR algorithm		EBSI		EBSQ	
	NOI	IRS	NOI	IRS	NOI	IRS
Extended Three Expo Terms	15	6	20	11	20	11
Quadratic Diagonal Perturbed	124	41	52	8	51	8
Extended Powell	180	60	77	26	74	24
Extended Tridiagonal 2	40	18	35	11	35	11
Raydan 1	102	52	98	36	94	34
Partial Perturbed Quadratic	74	21	85	26	85	26
DIXMAANE (CUTE)	121	65	83	27	83	27
EDENSCH (CUTE)	69	50	49	33	40	24
STAIRCASE S1	671	338	559	162	470	135
Perturbed Quadratic	95	33	98	32	98	32
Extended Cliff	F	F	9	7	7	6
Extended Maratos	89	32	70	37	70	37
NONDIA (CUTE)	13	7	15	8	13	7
Extended Block-Diagonal BD2	122	62	13	8	13	8
ENGVAL1 (CUTE)	34	16	27	10	45	27
Total	1749	801	1281	435	1191	411

Table 2: Comparison of methods for n= 1000

Test problems	FR algorithm		EBSI		EBSQ	
	NOI	IRS	NOI	IRS	NOI	IRS
Extended Three Expo Terms	127	117	78	74	14	9
Quadratic Diagonal Perturbed	445	196	183	31	158	34
Extended Powell	F	F	96	30	89	27
Extended Tridiagonal 2	43	23	50	29	41	19
Raydan 1	F	F	385	237	F	F
Partial Perturbed Quadratic	370	88	251	63	257	65
DIXMAANE (CUTE)	345	169	273	80	231	70
EDENSCH (CUTE)	98	82	150	135	98	82
STAIRCASE S1	F	F	F	F	F	F
Perturbed Quadratic	349	95	384	112	341	98
Extended Cliff	60	31	12	9	10	8
Extended Maratos	107	40	69	34	65	32
NONDIA (CUTE)	15	7	12	7	18	10
Extended Block-Diagonal BD2	130	66	10	7	10	7
ENGVAL1 (CUTE)	142	126	131	119	76	63
Total	2231	1040	1603	700	1319	497

F : The algorithm fail to converge**.**

Conclusion

In this paper, we have derived a new spectral nonlinear conjugate gradient methods based on our sufficient descent condition. Some numerical results have been reported, which showed the effectiveness of our method with the parameter ρ.

Moreover, we would like to find optimal values of parameter ρ in theory and in practical computation.

Table 1 and 2 gives a comparison between the a class of new descent gradient methods and the Fletcher–Reeves (FR) method taking nonlinear test function with n=100,1000. This table indicates that the new descent-type methods saves (27-36) % NOI and (38-50) % IRS, especially for our selected test problems. The Percentage Performance of the improvements of the Table 1 and Table 2 are given by the following Table 3. Relative Efficiency of the Different Methods Discussed in the Paper.

Table 3: Relative efficiency of the new Algorithm

Tools	NOI	IRS
Fletcher–Reeves (FR) method	100 %	100 %

EBSQ method	63.06 %	49.26 %
EBSI method	72.46 %	61.65 %

References

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

[2] Al-Baali, "Descent property and global convergence of the Fletcher–Reeves method with inexact line search," IMA J. Numer. Anal. 5 pp. 121–124, 1985.

[3] Birigin, F. G. and Martinez, J. M., "A Spectral Conjugate Gradient Method for Unconstrained Optimization," Applied Mathematics and Optimization, 43, 117-128. 2001.

[4] Basim A. H. and Haneen A. A., "New Nonlinear Conjugate Gradient Formulas for Solving Unconstrained Optimization Problems," Al-Mustansiriyah Journal of Science, 3, pp. 82-88. 2016.

[5] Basim A. H. and Haneen A. A., "A New Nonlinear Conjugate Gradient Method Based on the Scaled Matrix ," 2015.

[6] Dolan E. and More J.,"Benchmarking optimization software with performance profiles ," Math. Programming 91, pp. 201-213, 2002.

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

[8] Hager W. W. and Zhang. H. "A surrevy of nonlinear conjugate gradient methods "Paaific Journal of optimization. 2006.

[9] Zhang, L.W. Zhou, D. Li,"Global convergence of a modified Fletcher–Reeves conjugate gradient method with Armijo-type line search," Numer. Math. 104, PP. 561–572, 2006.

[10] Zoutendijk, G.,"Nonlinear programming, computational methods. In: Abadie, J. (eds.) Integerand Nonlinear Programming, " North-Holland, Amsterdam .pp. 37–86, 1970.