Research Article Open Access

A Modified Spectral Methods for Large-Scale Uncon-Strained

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

Article Info

Received
1/Mar./2017

Accepted
17/Oct./2017

Abstract
A modified spectral methods for solving unconstrained optimization problems based on the formulae are derived which are given in [4, 5]. The proposed methods satisfied the descent condition. Moreover, we prove that the new spectral methods are globally convergent. The Numerical results show that the proposed methods effective by comparing with the FR- method.

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

الخلاصة
اشتقاق الطرائق الطيفية المطورة لحل مسائل الامثلية غير المقيدة معتمدة على صيغ الذي معطى في **[4,5]**.الطرائق الطيفية المقترحة تحقق شرط الانحدار .بالإضافة الى ذلك تم أثبات التقارب الشامل للطرق الطيفية الجديدة. وقد أظهرت النتائج العددية فعالية الطرق المقترحة مقارنةً بطريقة **-FR** .

Introduction
Unconstrained optimization problems have extensive applications, for example, in petroleum exploration, transportation, and other domains. However, the amount of necessary calculation also grows exponentially with the increasing scale of the problem. Therefore, it is required to develop new methods to solve the large-scale unconstrained optimization problems. For solving large - scale unconstrained optimization problems:

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

where $f: R^n \to R$ is a continuously differentiable function, bounded from below, of the most elegant and probably the simplest methods are the conjugate gradient methods. For solving such problems, staring from an initial guess $x_0 \in R^n$, a nonlinear conjugate gradient method, generates a sequence $\{x_k\}$ as :

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

where α_k is a positive scalar and called the step length which is determined by some line search, and d_k are generated as :

$$d_{k+1} = -g_{k+1} + \beta_k d_k \tag{3}$$

where β_k is a parameter such that when applied to minimize a strictly convex quadratic function, the directions d_{k+1} and d_k are conjugate with respective to the Hessian of the objective function. Different conjugate gradient algorithms corresponds to different choices for the scalar parameter β_k. Line search in the conjugate gradient algorithms often is based on the Standard Wolfe (SW) conditions.

$$f(x_k) - f(x_k + \alpha_k d_k) \ge -\delta \alpha_k g_k^T d_k \tag{4}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k \tag{5}$$

where d_k is supposed to be a descent direction and with $0 < \delta < \sigma$ and More performance profile, is given in [6].

The Fletcher–Reeves (FR) method is a well-known conjugate gradient method. In the FR method, the parameter β_k is specified by :

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

More details can be found in [7]. Zoutendijk [10] proved that the FR method with exact line search is globally convergent. Under the strong Wolfe line search, Al-Baali [2] proved the global convergence of the FR method.

Birgin and Martinez [3] proposed another kind of conjugate gradient method, called spectral conjugate gradient method. Let $y_k = g_{k+1} - g_k$ and $v_k = x_{k+1} - x_k$. Then, the direction in this method was defined by :

$$d_{k+1} = -\theta_k g_{k+1} + \beta_k d_k \quad (7)$$

where $\theta_k = v_k^T v_k / y_k^T v_k$ is the spectral gradient, $\beta_k = (\theta_k y_k - v_k)^T g_k / d_k^T y_k$.

Zhang et al. [9] proposed a modified FR method (called MFR), in which the direction d_{k+1} is defined by

$$d_{k+1} = -\theta_k^{MFR} g_{k+1} + \beta_k^{FR} d_k, \quad (8)$$

where

$$\theta_k^{MFR} = \frac{y_k^T d_k}{g_k^T g_k}. \quad (9)$$

which is a descent direction independent of the line search.

In this paper, we are going to develop a modified spectral method. The search direction, generated by the method at each iteration, satisfies the descent condition. We are also going to establish the global convergence of the proposed algorithm with the Wolfe line search.

A new modified methods

In mid of 2016, Basim and Haneen proposed a modified nonlinear conjugate gradient formulae which is simple and easy to be used. These formula are called the BSQ and BSI and are given by :

$$\beta_k^{BSQ} = \frac{g_{k+1}^T g_{k+1}}{\xi_{k+1}}, \quad \beta_k^{BSI} = \frac{g_{k+1}^T g_{k+1}}{\delta_{k+1} d_k^T v_k} \quad (10)$$

where

$$\xi_{k+1} = \alpha_k (g_k^T d_k)^2 / 2(f_k - f_{k+1}), \ \delta_{k+1} = \|y_k\| / \|v_k\|. \quad (11)$$

In process of studying the BSQ and BSI methods, Basim and Haneen [4,5] established some convergence results applicable to any methods for which β_k can be expressed as a ratio :

$$0 < \beta_k \le \frac{-g_{k+1}^T d_{k+1}}{-g_k^T d_k}. \quad (12)$$

Our idea originates from (12):

$$\beta_k^{BSQ} \, g_k^T d_k = g_{k+1}^T d_{k+1}$$

$$g_{k+1}^T d_{k+1} = \frac{g_{k+1}^T g_{k+1}}{\xi_{k+1}} \, g_k^T d_k \quad (13)$$

Since

$$y_k^T d_k = g_{k+1}^T d_k - g_k^T d_k \quad (14)$$

From (13) and (14) we get:

$$g_{k+1}^T d_{k+1} = -\frac{g_{k+1}^T g_{k+1}}{\xi_{k+1}} \left[\frac{y_k^T d_k}{\xi_{k+1}} \xi_{k+1} - g_{k+1}^T d_k \right]$$

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

Hence

$$g_{k+1}^T d_{k+1} = -\theta_k^{SBSQ} g_{k+1}^T g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{\xi_{k+1}} g_{k+1}^T d_k \quad (16)$$

where

$$\theta_k^{SBSQ} = \frac{y_k^T d_k}{\xi_{k+1}}. \quad (17)$$

Then the search direction can be written as :

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

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

$$\theta^{SBSI} = \frac{y_k^T d_k}{\delta_{k+1} d_k^T v_k}. \quad (19)$$

New Algorithm and Descent property

In this section, we give the specific form of the proposed scaled conjugate gradient algorithm as follows and prove its descent property.

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

Outline of the SBSQ and SBSI Algorithms:

Step 0: Given $x_0 \in R^n$, $\varepsilon = 0.0001$, $\delta_1 \in (0, 1)$ and $\delta_2 \in (0, 1/2)$.

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

Step 2 : Set $\beta_k = \beta_k^{BSQ}$ *or* β_k^{BSI} with θ_k^{SBSQ} *or* θ_k^{SBSI} respectively.
Step 3 : Set $x_{k+1} = x_k + \alpha_k d_k$,(Use SW-conditions to compute α_k).
Step 4 : Compute $d_{k+1} = -\theta_k g_{k+1} + \beta_k d_k$,

Step 5 : Go to Step 1 with new values of x_{k+1} and g_{k+1}. **The** following theorem shows that the modified Algorithms 3.1 possesses descent property.

Theorem 1.
Suppose the direction d_{k+1} is obtained by (9), then we have

$$g_{k+1}^T d_{k+1} \le -c\|g_{k+1}\|^2 \tag{20}$$

holds for all $k \ge 1$.
Proof :
Since $d_0 = -g_0$ we have $g_0^T d_0 = -\|g_0\|^2 < 0$. Suppose that $g_k^T d_k < -c_1\|g_k\|^2$ for all $k \in n$. Multiplying (18) by g_{k+1}, we have :

$$g_{k+1}^T d_{k+1} = -\theta^{SBSQ} g_{k+1}^T g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{\xi_{k+1}} g_{k+1}^T d_k \tag{21}$$

$$g_{k+1}^T d_{k+1} = \frac{g_{k+1}^T g_{k+1}}{\xi_{k+1}}\left[-y_k^T d_k + g_{k+1}^T d_k\right]$$

$$= \frac{\|g_{k+1}\|^2}{\xi_{k+1}} g_k^T d_k \tag{22}$$

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

Since $g_k^T d_k < -c_1\|g_k\|^2$, then we have :

$$g_{k+1}^T d_{k+1} < -c_1 \frac{\|g_k\|^2}{\xi_{k+1}}\|g_{k+1}\|^2 .$$

where $c = c_1 \frac{\|g_k\|^2}{\xi_{k+1}}$. Since c_1, ξ_{k+1} and $\|g_k\|^2$ are positive values, then c is also positive value.

$$g_{k+1}^T d_{k+1} \le -c\|g_{k+1}\|^2 . \tag{23}$$

Global convergence

In order to prove the global convergence of algorithm (3.1), we assume that the objective function satisfied the following assumptions 1.
Assumption 1

i- The level set $L = \{x \in R^n \mid f(x) \le f(x_0)\}$ is bounded.
ii- In some neighborhood U and L, $f(x)$ is continuously differentiable and the gradient is Lipschitz continuous, namely, there exists a constant $\mu_1 > 0$ such that :

$$\|g(x_{k+1}) - g(x_k)\| \le \mu_1\|x_{k+1} - x_k\|,\ \forall x_{k+1}, x_k \in U \tag{24}$$

More details can be found in [8].
The conclusion of the following lemma, often called the Zoutendijk condition, is used to prove the global convergence of the proposed algorithms. It was originally given by Zoutendijk [10].

Lemma
Suppose that Assumption 1 holds. Consider any method of the form $(1)-(3)$, where d_{k+1} is a descent search direction and α_k satisfies the Wolfe line search. Then :

$$\sum_{k=1}^{\infty}\frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty . \tag{25}$$

The following theorem establishes the global convergence of the proposed methods. More performance profile, is given in [10].

Theorem 2
Suppose that Assumption 1 holds. Let the sequences $\{g_{k+1}\}$ and $\{d_{k+1}\}$ be generated by Algorithm 3.1, and let the α_k be determined by the Wolfe line search (4) and (5). Then :

$$\liminf_{k\to\infty}\|g_k\| = 0 . \tag{26}$$

Proof :
According to the given conditions, **Lemma 4.1** all hold. In the following, we will obtain the conclusion (26) by contradiction. Suppose by contradiction that there exists a positive constant $\varepsilon_1 > 0$ such that :

$$\|g_{k+1}\| > \varepsilon_1, \tag{27}$$

From equation (18), we get :

$$d_{k+1} + \theta_k^{SBSQ} g_{k+1} = \beta_k^{BSQ} d_k , \tag{28}$$

and squaring both side of it, we get :

$$\|d_{k+1}\|^2 + (\theta_k^{SBSQ})^2\|g_{k+1}\|^2 + 2\theta_k^{SBSQ} d_{k+1}^T g_{k+1} = (\beta_k^{BSQ})^2\|\text{(} \tag{29}$$

from (29), we get :

$$\|d_{k+1}\|^2 = (\beta_k^{BSQ})^2 \|d_k\|^2 - 2\theta_k^{SBSQ} d_{k+1}^T g_{k+1} - (\theta_k^{SBSQ})^2 \|g_{k+1}\| \quad (30)$$

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

$$\|d_{k+1}\|^2 \le \left(\frac{g_{k+1}^T d_{k+1}}{g_k^T d_k} \right)^2 \|d_k\|^2 - 2\theta_k^{SBSQ} d_{k+1}^T g_{k+1} - (\theta_k^{SBSQ})^2 \|g_{k+1}\| \quad (31)$$

Dividing the both inequality by $(g_{k+1}^T d_{k+1})^2$, we have:

$$\begin{aligned} \frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} &\le \frac{\|d_k\|^2}{(g_k^T d_k)^2} - (\theta_k^{SBSQ})^2 \frac{\|g_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} - 2\theta_k^{SBSQ} \frac{1}{d_{k+1}^T g_{k+1}} \\ &\le \frac{\|d_k\|^2}{(g_k^T d_k)^2} - (\theta_k^{SBSQ})^2 \frac{\|g_{k+1}\|^2}{c^2 \|g_{k+1}\|^4} - 2\theta_k^{SBSQ} \frac{1}{c\|g_{k+1}\|^2} - \frac{1}{\|g_{k+1}\|^2} + \frac{1}{\|g_{k+1}\|^2} \\ &\le \frac{\|d_k\|^2}{(g_k^T d_k)^2} - \left(\theta_k^{SBSQ} \frac{\|g_{k+1}\|}{c\|g_{k+1}\|^2} + \frac{1}{\|g_{k+1}\|} \right)^2 + \frac{1}{\|g_{k+1}\|^2} \\ &\le \frac{\|d_k\|^2}{(g_k^T d_k)^2} + \frac{1}{\|g_{k+1}\|^2} \end{aligned} \quad (32)$$

Using (32) recursively and noting that $\|d_1\|^2 = -g_1^T d_1 = \|g_1\|^2$, we get :

$$\frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \le \sum_{i=1}^{k} \frac{1}{\|g_i\|^2} \quad . \quad (33)$$

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

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \frac{\varepsilon_1^2}{k}, \quad (34)$$

which indicates :

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \sum_{k=1}^{\infty} \frac{\varepsilon_1^2}{k} = \infty \quad (35)$$

This contradicts the Zoutendijk condition (25). Therefore the conclusion (26) holds.

Remark

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

Numerical Results

This section presents the performance of FORTRAN 90 implementation to the algorithm FR by using a set of well-known unconstrained optimization test functions, for each function we have considered numerical experiments with the number of variables n = 100 and 1000. Test problems are from [1]. We compared the perform acne of the algorithm FR with famous formula FR, which they defined in (6).All these algorithms are implemented with the standard Wolfe line search conditions with $\delta_1 = 0.001$ and $\delta_2 = 0.9$, the stopping condition defined by $\|g_{k+1}\| \le 10^{-6}$. The comparison includes the following :

NOI : number of iterations .

IRS : number of restart .

From Tables 1, and 2, we draw a conclusion that Algorithms 3.1 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.

Table 1 : Comparison of methods for n= 100

Test problems	FR algorithm		SBSI		SBSQ	
	NOI	IRS	NOI	IRS	NOI	IRS
Extended Three Expo Terms	15	6	17	10	17	10
Quadratic Diagonal Perturbed	124	41	49	10	45	7
Extended Powell	180	60	88	29	78	26
Extended Tridiagonal 2	40	18	38	14	38	15
Raydan 1	102	52	74	27	73	26
Partial Perturbed Quadratic	74	21	87	24	87	24
DIXMAANE (CUTE)	121	65	85	26	86	27
EDENSCH (CUTE)	69	50	26	10	63	47

STAIRCASE S1	671	338	399	107	509	134
Perturbed Quadratic	95	33	83	29	83	29
Extended Cliff	F	F	11	9	11	9
Extended Maratos	89	32	78	39	75	37
NONDIA (CUTE)	13	7	17	9	13	7
Extended Block-Diagonal BD2	122	62	12	8	12	8
ENGVAL1 (CUTE)	34	16	24	7	24	7
Total	1749	801	1077	349	1203	404

Table 2: Comparison of methods for n= 1000

Test problems	FR algorithm		SBSI		SBSQ	
	NOI	IRS	NOI	IRS	NOI	IRS
Extended Three Expo Terms	127	117	11	7	21	17
Quadratic Diagonal Perturbed	445	196	177	37	187	36
Extended Powell	F	F	86	26	89	29
Extended Tridiagonal 2	43	23	38	14	48	26
Raydan 1	F	F	457	265	F	F
Partial Perturbed Quadratic	370	88	253	60	260	65
DIXMAANE (CUTE)	345	169	230	69	233	69
EDENSCH (CUTE)	98	82	76	62	29	13
STAIRCASE S1	F	F	F	F	F	F
Perturbed Quadratic	349	95	347	93	334	86
Extended Cliff	60	31	16	11	17	12
Extended Maratos	107	40	166	111	70	34
NONDIA (CUTE)	15	7	15	7	15	8
Extended Block-Diagonal BD2	130	66	12	7	12	7
ENGVAL1 (CUTE)	142	126	88	75	109	96
Total	2231	1040	1429	553	1335	469

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

Conclusion

Through this relation which is defined in (12), we have been presented a new spectral methods that estimates spectral conjugate gradient. Our scheme is simple and able to enhance the performance of the gradient-type methods with minimal storage. SBSQ and SBSI methods gives the best results. SBSQ method give worse results than SBSI method.

Table 1 and 2 gives a comparison between the a new spectral methods and the Fletcher–Reeves (FR) method with the number of variable n=100, 1000. This table indicates that the new spectral methods saves (36 – 37) % NOI and (51 - 52) % 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 is discussed in the paper.

Table 3: Relative efficiency of the new Algorithm

Tools	NOI	IRS
Fletcher–Reeves (FR) method	100 %	100 %
SBSO method	63.76 %	47.41 %
SBSI method	62.96 %	48.99 %

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 profilesm" 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.