Journal of Zankoy Sulaimani
Part-A- (Pure and Applied Sciences)
http://jzs.univsul.edu.iq

A New Spectral on the Gradient Methods for Unconstrained Optimization Minimization Problem

Basim A. Hassan[1] and Hawraz N. Jabbar*[2]

[1]Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, Iraq.
[2]Department of Mathematics, College of Sciences, University of Kirkuk, Iraq.
*Corresponding author's e-mail: hawrazmath@gmail.com

Article info

Original: 3 February 2020
Revised: 21 August 2020
Accepted: 26 September 2020
Published online: 20 December 2020

Key Words:
Unconstrained Optimization, Conjugate Gradient Methods, Large Scale Methods.

Abstract

One simple and well known method for minimizing the functions is a spectral conjugate gradient method. In this paper, we derive a new spectral on the method of gradient, which can give a new path of search. The new spectral method holds the property of descent and we have shown that the spectral method is convergent globally. The empirical results show that for the test problems, the given approach is competitive with the other conjugate gradient methods.

1. Introduction

We will consider the unconstrained optimization minimization problem as follows:

$$\min_{x\in R^n} f(x) \tag{1}$$

where $f: R^n \to R^l$ is a smooth function whose gradient is for the sake of simplicity. For details see [17].

The best kinds of iterative methods are the conjugate gradient methods to solve unconstrained large-scale optimization problems since they do not require matrix usage and are typically highly efficient. For details see [16].

In conjugate gradient, a sequence $\{x_k\}$, effective solution is proved by the following iteration:

$$x_0 \in R^n, x_{k+1} = x_k + \alpha_k d_k \tag{2}$$

where $\alpha_k > 0$ is a step-size and d_k is direction. The search direction generated by:

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

where β_k is called the CG parameter. The step size α_k satisfies some specific rules of the line, such as the traditional quest for Wolfe line:

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

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

where $0 < \delta < \sigma < 1$, see in [13].

Numerical performance depends on the choice of β_k, it has been focused on in the literature. Some efficient and effective parameters, for example, the ([5-8], [10-11], and [15]).

The most well-known β_k expressions of methods conjugate gradient is Fletcher–Reeves method [7], in which is defined by:

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \tag{6}$$

Different from the category conjugate gradient method, in a spectral conjugate gradient method, the search direction d_{k+1} is defined as follows:

$$d_{k+1} = -\vartheta_k g_{k+1} + \beta_k d_k \tag{7}$$

where ϑ_k is called a spectral coefficient. It is easy to see that (7) reduces to (3) if $\vartheta_k = 1$. More details can be found in [9].

In this contribution, improvements are made to nonlinear CG methods that boost the efficacy of the basic ones. The following modifications, using spectral coefficients:

$$\theta_k^{DY} = \theta_k^{HS} = \frac{y_k^T d_k}{y_k^T d_k} = 1, \qquad \theta_k^{PR} = \theta_k^{FR} = \frac{y_k^T d_k}{g_k^T g_k}, \qquad \theta_k^{LS} = \theta_k^{CD} = \frac{y_k^T d_k}{|g_k^T d_k|} \tag{8}$$

More details can be found in [12].

Conjugate gradient methods have several global convergence findings and these are checked in [8, 14].

The purpose of this paper is to derive spectral coefficient and study its convergence property and numerical results and a discussion.

2. A new spectral conjugate gradient method

One of the essential conjugate gradient methods has been solved open problems in [2] with the following CG parameter:

$$\beta_k^B = \frac{g_{k+1}^T g_{k+1}}{(f(x_k) - f(x_{k+1})/\alpha_k) - (g_k^T d_k/2)} \tag{9}$$

based on a parameter conjugate gradient, we derive a new spectral conjugate gradient.

Within this section, we will state the idea of proposing a new method of spectral conjugate of gradients and develop a new algorithm.

Therefore, parameter conjugate gradient is satisfying the relation:

$$\beta_k \leq \frac{-g_{k+1}^T d_{k+1}}{-g_k^T d_k} \tag{10}$$

then it can be regarded as:

$$\beta_k g_k^T d_k = g_{k+1}^T d_{k+1} \tag{11}$$

$$g_{k+1}^T d_{k+1} = \frac{g_{k+1}^T g_{k+1}}{(f(x_k) - f(x_{k+1})/\alpha_k) - (g_k^T d_k/2)} g_k^T d_k$$

It follows from the meaning of y_k we get:

$$y_k^T d_k = g_{k+1}^T d_k - g_k^T d_k \tag{12}$$

Let $\psi_k = (f(x_k) - f(x_{k+1})/\alpha_k) - (g_k^T d_k/2)$. It is obvious that:

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

$$= -\frac{y_k^T d_k}{\psi_k} g_{k+1}^T g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{\psi_k}\, g_{k+1}^T d_k \tag{13}$$

$$g_{k+1}^T d_{k+1} = -\theta_k g_{k+1}^T g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{\psi_k} g_{k+1}^T d_k \tag{14}$$

where

$$\theta_k^{BBA} = \frac{y_k^T d_k}{(f(x_k) - f(x_{k+1})/\alpha_k) - (g_k^T d_k/2)} \tag{15}$$

which we called BBA method. Based on the (15), which can give a new path of search:

$$d_{k+1} = -\theta_k^{BBA} g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{(f(x_k) - f(x_{k+1})/\alpha_k) - (g_k^T d_k/2)} d_k \tag{16}$$

Utilizing the process above, a concrete method is presented as follows:

2.1 Outline of the the new Algorithm (BBA Algorithm):

Step 0: Chose $\boldsymbol{x_0} \in \boldsymbol{R^n},\ \boldsymbol{0 < \delta_1 < \delta_2 < 1}$, Set $\boldsymbol{d_1 = -g_1}$.

Step 1: Stop when $\|\boldsymbol{g_{k+1}}\| \leq \boldsymbol{10^{-6}}$, else continue.

Step 2: Compute $\boldsymbol{\beta_k}$ is a parameter which is defined in (9) with $\boldsymbol{\theta_k^{BBA}}$ defined in (15).

Step 3: Set $\boldsymbol{x_{k+1} = x_k + \alpha_k d_k}$ such that the inequalities (4) and (5) are satisfied.

Step 4: Evaluate $\boldsymbol{d_{k+1} = -\theta_k g_{k+1} + \beta_k d_k}$.

Step 5: Go to Step 1 with new values of $\boldsymbol{x_{k+1}}$ and $\boldsymbol{g_{k+1}}$.

The next theorem shows that the new Algorithm has descent property.

Theorem 1

Let sequence $\{x_0\}$ be obtained by Algorithm 2.1, then $g_{k+1}^T d_{k+1} \leq -c\|g_{k+1}\|^2$ for all k.

Proof :

So $\boldsymbol{d_0 = -g_0}$ we have $\boldsymbol{g_0^T d_0 = -\|g_0\|^2 < 0}$. Let $\boldsymbol{g_k^T d_k < -c_1\|g_k\|^2}$ for all $\boldsymbol{k \in n}$. Multiplying (16) by $\boldsymbol{g_{k+1}}$ we have :

$$g_{k+1}^T d_{k+1} = -\theta_k^{BBA} g_{k+1}^T g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{(f(x_k) - f(x_{k+1})/\alpha_k) - (g_k^T d_k/2)} g_{k+1}^T d_k \tag{17}$$

$$\begin{aligned} g_{k+1}^T d_{k+1} &= \frac{g_{k+1}^T g_{k+1}}{(f(x_k) - f(x_{k+1})/\alpha_k) - (g_k^T d_k/2)} [-y_k^T d_k + g_{k+1}^T d_k] \\ &= \frac{\|g_{k+1}\|^2}{(f(x_k) - f(x_{k+1})/\alpha_k) - (g_k^T d_k/2)} g_k^T d_k \\ &= \frac{g_k^T d_k}{(f(x_k) - f(x_{k+1})/\alpha_k) - (g_k^T d_k/2)} \|g_{k+1}\|^2 \end{aligned} \tag{18}$$

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

$$g_{k+1}^T d_{k+1} < -c_1 \frac{\|g_k\|^2}{(f(x_k) - f(x_{k+1})/\alpha_k) - (g_k^T d_k/2)} \|g_{k+1}\|^2$$

where $\boldsymbol{c = -c_1 \frac{\|g_k\|^2}{(f(x_k) - f(x_{k+1})/\alpha_k) - (g_k^T d_k/2)}}$. Since $\boldsymbol{c_1},\ \boldsymbol{(f(x_k) - f(x_{k+1})/(\alpha_k - g_k^T d_k/2)}$ and $\|\boldsymbol{g_k}\|^2$ are positive, we have $\boldsymbol{c}$ is also positive value.

$$g_{k+1}^T d_{k+1} \leq -c\|g_{k+1}\|^2 \tag{19}$$

3. Global convergence

We'll be studying the global convergence of Algorithm 2.1 in this section. We first state the following mild hypotheses which are needed in this paper to prove the main findings.

Assumption 1

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

ii- In some neighborhood $\boldsymbol{U}$ and $\boldsymbol{L}$, $\boldsymbol{f(x)}$ is continuously differentiable and its gradient id Lipschitz continuous, namely, there exists a constant $\boldsymbol{\mu_1 > 0}$ such that :

$$\|g(x_{k+1}) - g(x_k)\| \leq \mu_1 \|x_{k+1} - x_k\|, \forall x_{k+1}, x_k \in U \tag{20}$$

More details can be found in [8].

The result of the following lemma, which is also called the Zoutendijk condition, is used to prove the global convergence of the algorithms proposed. Zoutendijk originally gave this [18].

Lemma 1

Let assumptions (i) and (ii) holds and consider the CG method (1) and (3) with $\boldsymbol{d_{k+1}}$ satisfies $\boldsymbol{g_{k+1}^T d_{k+1} \leq 0}$ and $\boldsymbol{\alpha_k}$ satisfies the inequalities (4-5). Then

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_2\|^2} < +\infty \tag{21}$$

The theorem below points out the global convergence of the methods proposed.

Theorem 2

Let assumptions (i) and (ii) holds and sequences $\{\boldsymbol{g_{k+1}}\}$ and $\{\boldsymbol{d_{k+1}}\}$ be generated by Algorithm 2.1. Then

$$\mathbf{lim}_{k\to\infty}\ \mathbf{inf}\ \|\boldsymbol{g_k}\| = \mathbf{0}. \tag{22}$$

Proof :

Lemma 1 holds all according to the specified conditions. We'll be having the inference by implication in the following. Suppose a positive constant exists $\varepsilon_1 > 0$, by contradiction

$$\|g_{k+1}\| > \varepsilon_1 \tag{23}$$

On the one hand, rewriting (18) as follows

$$d_{k+1} + \theta_k^{BBA} g_{k+1} = \beta_k^B d_k \tag{24}$$

and squaring both side of it, we get:

$$\|d_{k+1}\|^2 + (\theta_k^{BBA})^2\|g_{k+1}\|^2 + 2\theta_k^{BBA} d_{k+1}^T g_{k+1} = (\beta_k^B)^2\|d_k\|^2 \tag{25}$$

From (25), we get

$$\|d_{k+1}\|^2 = (\beta_k^B)^2\|d_k\|^2 - 2\theta_k^{BBA} d_{k+1}^T g_{k+1} - \left(\theta_k^{BBA}\right)^2\|g_{k+1}\|^2 \tag{26}$$

From the above equation and (10), we have:

$$\|d_{k+1}\|^2 \leq \left(\frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}\right)^2\|d_k\|^2 - 2\theta_k^{BBA} d_{k+1}^T g_{k+1} - \left(\theta_k^{BBA}\right)^2\|g_{k+1}\|^2 \tag{27}$$

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

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

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

Using (28) on recurrence, and note that $\|d_1\|^2 = -g_1^T d_1 = \|g_1\|^2$, we get :

$$\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} \leq \sum_{i=1}^{k} \frac{1}{\|g_i\|^2} \tag{29}$$

Then we get from (28) and (23) that

$$\frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} \geq \frac{\varepsilon_1^2}{k} \tag{30}$$

which indicates

$$\sum_{k=1}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} \geq \sum_{k=1}^{\infty} \frac{\varepsilon_1^2}{k} = \infty \tag{31}$$

This contradicts the condition (21) of Zoutendijk. Hence holds the conclusion (22).

4. Numerical experiments

We present the results of our numerical experiments in this section to compare the algorithm presumed by FR in [6] and the algorithm 2.1 in this paper.

To equate this paper's methods with other classical ones, we will test the gradient errors for the algorithms to calculate the stop criterion $\|g_{k+1}\| \leq 10^{-6}$. Usually, when the gradient norm satisfies the following inequality, we must compel the iteration to end. In the search step for the Wolfe line we must select the parameters as follows, $\delta = 0.001$ and $\sigma = 0.9$.

Our studies were carried out unrestrictedly on a range of 30 Optimization test problems with the Andrei set dimensions 100 and 1000, please see Andrei[1] for information. Various test functions were observed in [3 and 4]. The comparing data includes number of iterations (IN), number of restart (NR) and number of evaluations of functions (NF).

Table 1: Numerical Results of BBA-Algorithm and FR- algorithm .

		BBA algorithm			FR algorithm		
P. No.	n	NI	NR	NF	NI	NR	NF
Extended White & Holst	100	34	18	68	43	18	88
	1000	33	17	71	46	19	92
Extended Beale	100	14	8	27	32	15	52
	1000	13	8	26	22	10	42
Penalty	100	11	7	28	10	6	27
	1000	22	13	47	24	16	191
Perturbed Quadratic	100	83	29	125	95	33	150
	1000	339	91	533	349	95	568
Generalized Tridiagonal 1	100	22	6	44	25	11	43
	1000	26	6	54	46	28	741
Extended Tridiagonal 1	100	10	5	21	32	13	64
	1000	15	7	29	77	46	129
Generalized Tridiagonal 2	100	40	15	61	37	8	67
	1000	64	27	101	73	27	115
Extended PSC1	100	8	6	17	15	9	31
	1000	7	5	15	8	6	17
Quadratic Diagonal P.	100	49	10	87	124	41	231
	1000	179	38	311	445	196	711
Extended Wood	100	31	12	59	71	35	110
	1000	26	10	51	47	15	84
Extended Hiebert	100	80	51	176	101	40	217
	1000	80	51	173	101	40	214
Extended Quadratic	100	24	12	51	32	12	65
Penalty	1000	37	20	90	53	22	116
ARWHEAD (CUTE)	100	8	4	16	9	4	18
	1000	8	6	56	12	7	82
NONDIA (CUTE)	100	11	6	21	13	7	25
	1000	15	8	30	15	7	29
DIXMAANE (CUTE)	100	85	26	133	121	65	218
	1000	253	77	394	345	169	634
Total		1627	599	2921	2423	1020	5171

5. Conclusions

The spectral conjugate gradient is simple and well known method for minimizing the functions. We derive a new spectral conjugate gradient which satisfies the sufficient descent condition and is the globally convergent.

Numerical tests are performed through a wide range of normal test functions. Computational findings indicate that the total number of iterations, the total number of restart and function evaluations respectively decreased by 32%, 41% and 43%

Table 2 : Relative efficiency of the new Algorithm

Tools	NI	NR	NF
FR Algorithm	100 %	100 %	100 %
BBA Algorithm	67.14 %	58.72 %	56.48 %

Acknowledgments

The authors are very grateful to the University of Mosul / College of Computers Sciences and Mathematics and University of Kirkuk / College of Sciences for their provided facilities, which helped to improve the quality of this work.

References

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

[2] Basim A. Hassan, *"A new formula for conjugate parameter computation based on the quadratic model*". Indonesian Journal of Electrical Engineering and Computer Science. Vol. 3, pp. 954-961. (2019).

[3] Basim A. Hassan. "*A Globally Convergence Spectral Conjugate Gradient Method for Solving Unconstrained Optimization Problems and Mohammed*". Raf. J. of Comp. & Math's. Vol. 10, pp. 21-28. (2013).

[4] Basim A. Hassan. "*Development a Special Conjugate Gradient Algorithm for Solving Unconstrained Minimization Problems*". Raf. J. of Comp. & Math's. Vol. 9, pp. 73-84. (2012).

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

[6] Fletcher, R. *"Practical Method of Optimization".* 2nd Edition. John Wiley and Sons. New York. (1989).

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

[8] Hager and W.W. and Zhang H. "*A new conjugate gradient method with guaranteed descent and an efficient line search".* SIAM J. Optim. Vol. 16, pp. 170–192. (2005).

[9] Hu C. M. and Wan Z. "*An extended spectral conjugate gradient method for unconstrained optimization problems".* British Journal of Mathematics & Computer Science. Vol. 3, No. 2, pp.86-98. (2013).

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

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

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

[13] Ming L., Hongwei L. and Zexian L. "*A new family of conjugate gradient methods for unconstrained optimization*". J. Appl. Math. Comput. (2017).

[14] Narushima Y. and Yae H. *"A survey of sufficient descent conjugate gradient methods for unconstrained optimization*". SUT J. Math. Vol. 50, pp. 167–203. (2014).

[15] Polak, E. and Ribiere, G. "*Note sur la Convergence de Directions Conjugate"*. Revue Francaise Informant. Reserche. Opertionelle. Vol. 16, pp. 35-43. (1969).
[16] Rao S. *"Engineering Optimization Theory and Practice"*. 4th edition, John Wiley & Sons Inc., New Jersey, Canada. (2009).
[17] Zhongbo S., Hongyang L., Jing W. and Yantao T. *"Two modified spectral conjugate gradient methods and their global convergence for unconstrained optimization"*. International Journal of Computer Mathematics, (2017).
[18] Zoutendijk, G. "*Nonlinear programming, computational methods"*. (In: Abadie, J. (eds.) Integerand Nonlinear Programming' . North-Holland, Amsterdam) .pp. 37–86. (1970).