Al-Mustansiriyah Journal of Science
ISSN: 1814-635X (print), ISSN:2521-3520 (online)
Volume 30, Issue 1, 2019
DOI: http://doi.org/10.23851/mjs.v30i1.478

Research Article
Open Access

Generalizations of the New Technique for Spectral Conjugate Gradient Methods

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

Article Info

Received
14/11/2017

Accepted
20/10/2018

Published
15/08/2019

Abstract
In this article, we will present a generalization of the new technique for spectral conjugate gradient methods based on the descent condition that by using a simple method to prove the worldwide convergence of the new method without the Wolfe line searches. Depending on our numerical experiments we can confirm that our proposed methods are preferable and in general better to the classical conjugate gradient methods in terms of good organization.

Keywords: Conjugate gradient, Spectral conjugate gradient, Sufficient descent condition, worldwide convergence, Numerical results.

الخلاصة
في هذه المقالة، نقترح تعميم التقنية الجديدة لطرق التدرج المترافق الطيفية المعتمدة على الشرط الانحدار ومن خلال طريقة بسيطة تم إثبات التقارب الشامل للطريقة الجديدة بـدون خط بحث ولف. تجاربنا العددية تشير بان طرقنا المقترحة مفضلة وعموماً أفضل من طرقِ التدرج المترافق الكلاسيكيةِ من ناحية الكفاءةِ.

Introduction

The most well-known minimization technique for unconstrained problems is nonlinear conjugate gradient (CG) method. We take into account solving the unrestrained minimization function:

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

where $f: R^n \to R$ is an uninterruptedly specifiable function, limited from below. The general structure of nonlinear conjugate gradient method can be summarized as follows, preliminary from an initial point $x_1 \in R^n$, the conjugate gradient method yields a cluster $x_k \subset R^n$ such that :

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

where $\alpha_k > 0$ is a step length, received from the line search, and the direction d_k are given by:

$$d_1 = -g_1\ ,\ d_{k+1} = -g_{k+1} + \beta_k d_k \qquad (3)$$

In the previous relation, β_k is the conjugate gradient parameter. Now, we denote $g_k = \nabla f(x_k)$, $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$. More details can be found in [10]. Different conjugate gradient methods communicate to different selected for the formula β_k such as:

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}\ ,\ \beta_k^{CD} = -\frac{g_{k+1}^T g_{k+1}}{g_k^T d_k} \qquad (4)$$

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{y_k^T d_k}\ ,\ \beta_k^{PR} = \frac{g_{k+1}^T y_k}{g_k^T g_k} \qquad (5)$$

The equivalent methods are called Fletcher Reeves [2] (FR) method, Conjugate Descent [3] (CD) method, Polak Ribiere Polyak [8] (PRP) method and Hestenes-Stiefel [4](HS) method, respectively.

The step size α_k is computed by exact or inexact cable search. The Wolfe states is used in familiar:

$$f(x_k) - f(x_k + \alpha_k d_k) \ge -\delta \alpha_k g_k^T d_k \qquad (6)$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k \qquad (7)$$

where d_k is a descent direction and $0 < \delta \leq \sigma < 1$.

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

$$d_{k+1} = -\vartheta_k g_{k+1} + \beta_k d_k, \qquad (8)$$

where ϑ_k is called a spectral coefficient. It is easy to see that (8) reduces to (3) if $\vartheta_k = 1$. Since there are two types of parameters can be suitably chosen to obtain a search direction in (3), it is possible that (3) combines the advantages of spectral method and conjugate gradient method. More details can be found in [6].

In this paper, we will mention our generalizations of the new technique for spectral conjugate gradient methods and its algorithm in part 2. Whereas in part 3, we demonstrate the adequate descent condition and the worldwide convergence realization of the new technique. Whereby in in part 4, we tackle the mathematical results and elaboration and at last we draw a conclusion in the part 5.

Generalizations of the new technique for SCG - methods

One motivation for our new formula for ϑ_k is the descent property of the conjugate descent method.

We assume that the search direction is descent direction for all k steps, and we hope the conclusion also for $k+1$ steps, so we have:

$$g_{k+1}^T d_{k+1} < 0 \qquad (9)$$

From the (10) and multiplying by g_{k+1}^T and from the (18) we have:

$$-\vartheta_k g_{k+1}^T g_{k+1} + \beta_k g_{k+1}^T d_k < 0 \qquad (10)$$

We observe the formula (10), let the form of ϑ_k to be $\vartheta_k = 1 + \eta_k$, and we assume $\beta_k > 0$, we have :

$$\frac{\vartheta_k \|g_{k+1}\|^2}{\beta_k} > g_{k+1}^T d_k$$

$$\frac{(1+\eta_k)\|g_{k+1}\|^2}{\beta_k} > g_{k+1}^T d_k \qquad (11)$$

From (11) we get:

$$\frac{\|g_{k+1}\|^2}{\beta_k} + \frac{\eta_k \|g_{k+1}\|^2}{\beta_k} > g_{k+1}^T d_k \qquad (12)$$

let us select a $d_k^T G v_k \geq 0$ and we get:

$$\frac{\|g_{k+1}\|^2}{\beta_k} + \frac{\eta_k \|g_{k+1}\|^2}{\beta_k} = d_k^T G v_k + g_{k+1}^T d_k \qquad (13)$$

From (12) we let:

$$\frac{\|g_{k+1}\|^2}{\beta_k} = d_k^T G v_k \qquad (14)$$

and

$$\frac{\eta_k \|g_{k+1}\|^2}{\beta_k} = g_{k+1}^T d_k \qquad (15)$$

From (15) and (14) ,we know that:

$$\beta_k = \frac{\|g_{k+1}\|^2}{d_k^T G v_k} \qquad (16)$$

and

$$\vartheta_k^{GB} = 1 + \frac{g_{k+1}^T d_k}{d_k^T G v_k} \qquad (17)$$

In particular, Hideaki and Yasushi (HY) [5] made a modification on the CG method and developed a modified method. Their method expression of the denominator $d_k^T G v_k$, we know that:

$$d_k^T G v_k = 2(f_k - f_{k+1}) / \alpha_k \qquad (18)$$

Accordingly, we give the parameters ϑ_k^{GB} in the above algorithm:

$$\vartheta_k^{GB} = 1 + \frac{g_{k+1}^T d_k}{2(f_k - f_{k+1}) / \alpha_k} \qquad (19)$$

Different choices for $d_k^T G v_k$ lead to different spectral conjugate gradient methods.

Then, we can propose the following generalizations of the new technique for spectral nonlinear conjugate gradient methods (GB - methods):

New Algorithm:

Step 1. Give $x_1 \in R^n$, $\varepsilon \geq 0$. Set $d_1 = -g_1$ and set the initial $\alpha_1 = 1/\|g_1\|$.

Step 2. If $\|g_{k+1}\| \leq 10^{-6}$, then stop, or else be present at Step 3.

Step 3. Find α_k satisfying Wolfe conditions $(6-7)$ and new iterative $x_{k+1} = x_k + \alpha_k d_k$.

Step 4. Compute β_k by (15) with θ_k^{GB} by (16) respectively.

Step 5. Compute direction $d_{k+1} = -\theta_k g_{k+1} + \beta_k d_k$.

Set $k = k+1$ and continue with step2.

Convergent analysis

Within this part, the convergent features of GB will be considered. For a method to converge, it should meet the descent state and the worldwide convergence properties.

Sufficient descent condition

For the sufficient state to hold,

$$g_{k+1}^T d_{k+1} \leq -c\,\|g_{k+1}\|^2, \quad c > 0 \tag{20}$$

Theorem 1.

Consider a CG method with the search direction (3) with β_k and ϑ_k^{GB} given as (16) and (17), then state (20) holds for all $k+1$.

Proof.

If $k=1$, then $g_1^T d_1 \leq -c\,\|g_1\|^2$. Thus, state (20) holds true. We too require to prove that for $k>1$, state (20), will also hold true. From (8), multiply by g_{k+1}^T then:

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\vartheta_k \|g_{k+1}\|^2 + \beta_k g_{k+1}^T d_k \\ &= -\left(1 + \frac{g_{k+1}^T d_k}{d_k^T G v_k}\right)\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{d_k^T G v_k} g_{k+1}^T d_k \\ &= -\|g_{k+1}\|^2 \end{aligned} \tag{21}$$

Furthermore, from above analysis, we also get $\beta_k > 0$. The proof is completed.

Lemma 1.

Consider any method (2), (8), where (15), (16) and the step-size α_k be determined by the Wolfe line search, then:

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \tag{22}$$

The above lemma (1) regularly called Zoutendijk condition. It was originally known by Zoutendijk. [11]

Global convergence properties

In the direction of study, the global convergence properties, first if we chose any value $d_k^T G v_k$ with sufficient descent condition that β_k are always satisfies the relations:

$$0 < \beta_k \leq \frac{-g_{k+1}^T d_{k+1}}{-g_k^T d_k} \tag{23}$$

This formula is very important in our convergence analysis.

The following assumption that may be needed in many theorem proofs:

Assumption (A)

A_0 : $f : R^n \to R^1$ is bounded below.

A_1 : The level set $S = \{x \in R^n \mid f(x) \leq f(x_0)\}$ is bounded, i.e., there exists a positive constant $\zeta > 0$ such that:

$$\|x\| \leq \zeta, \quad \forall\, x \in S \tag{24}$$

A_2 : ∇f satisfied the Lipschitz state namely,

$$\|g(x_{k+1}) - g(x_k)\| \leq L\|x_{k+1} - x_k\|, \ \forall x_{k+1}, x_k \in U \tag{25}$$

where $L > 0$ Lipschitz constant.

Under these assumptions of $f(x)$, there exists a constant $\Gamma \geq 0$ such that:

$$\|g_{k+1}\| \leq \Gamma \tag{26}$$

More details can be found in [9].

Theorem 2.

Suppose that Assumption (A) holds. Consider any method (2), (8), where (15), (16) and the step-length α_k be verify by the Wolfe line search, then:

$$\lim_{k\Rightarrow\infty} \inf \|g_k\| = 0 \qquad (27)$$

Proof:

Suppose by disagreement that near exists a positive constant $\varepsilon > 0$ such that:

$$\|g_{k+1}\| > \varepsilon, \qquad (28)$$

From (8), we have:

$$\|d_{k+1}\|^2 = (\beta_k)^2\|d_k\|^2 - 2\vartheta_k^{GB} g_{k+1}^T d_{k+1} - (\vartheta_k^{GB})^2\|g_{k+1}\|^2 \qquad (29)$$

From the above equation and (23), we have:

$$\|d_{k+1}\|^2 \le \left(\frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}\right)^2 \|d_k\|^2 - 2\vartheta_k^{GB} d_{k+1}^T g_{k+1} - (\vartheta_k^{GB})^2\|g_{k+1}\|^2 \qquad (30)$$

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

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

$$\le \frac{\|d_k\|^2}{(g_k^T d_k)^2} - \left(\vartheta_k^{GB} \frac{\|g_{k+1}\|}{c\|g_{k+1}\|^2} + \frac{1}{\|g_{k+1}\|}\right)^2 + \frac{1}{\|g_{k+1}\|^2} \qquad (31)$$

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

Using (31) recursively and noting that $\|d_1\|^2 = -g_1^T d_1 = \|g_1\|^2$, we obtain:

$$\frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \le \sum_{i=1}^{k} \frac{1}{\|g_i\|^2} \qquad (32)$$

From (31) and (28) we obtain:

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \frac{\varepsilon_1^2}{k}, \qquad (33)$$

which indicates

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

This does not coincide with his (22). So, the result (27) holds.

Numerical results

We implement 15 classical unrestricted optimization questions from [1] successively to check algorithm. These experiments are dome by employing Fortran. In the First place we set $\delta_1 = 0.001$ and $\delta_2 = 0.9$ and if $\|g_{k+1}\| \le 10^{-6}$ then stop. The mathematical results of the tests are placed in Table 1. The initial pillar "Problem" stands for the label of the examined problem. Dim refers to the dimension of the test problems. NI, NF refer to the figure of iterations and function valuations, successively. Out of the mathematical results, it is denoted that the suggested proposed spectral conjugate gradient method is showing excellence.

In Table 1, it can be seen from the comparison given above that the new algorithm in this paper is more efficient than FR method and HY method for solving unconstrained optimization.

Conclusions

In this paper, we contain derived a generalization of the new technique for spectral conjugate gradient based on our descent condition. We have preliminary numerical outcome to show its efficiency. As demonstrated in Section 4, the information mathematical results illustrate that the modified GB performs better than the FR in [2] and HY in [5].

Table 1: Comparison of different CG methods with different test problems and different dimensions.

		FR method		HY method		GB method	
P. No.	n	NI	NF	NI	NF	NI	NF
1	100	19	35	19	36	18	34
	1000	38	65	34	62	36	67
2	100	47	93	42	84	31	64
	1000	78	131	35	73	34	73
3	100	32	52	14	29	13	26
	1000	22	42	16	32	13	25
4	100	10	27	9	25	9	25
	1000	24	191	11	33	11	33
5	100	32	64	13	25	11	23
	1000	77	129	19	34	13	26
6	100	15	25	14	21	8	13
	1000	F	F	F	F	8	15
7	100	71	110	26	50	38	71
	1000	47	84	25	50	29	57
8	100	101	217	85	206	79	195
	1000	101	214	85	203	81	196
9	100	23	45	18	33	19	34
	1000	27	55	19	44	22	47

References

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

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

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

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

[5] Hideaki I., and Yasushi N., (2011)," Conjugate gradient methods using value of objective function for unconstrained optimization Optimization Letters, V.6, Issue 5, pp. 941-955.

[6] Hu C. M. and Wan Z., (2013), " An extended spectral conjugate gradient method for unconstrained optimization problems" British Journal of Mathematics & Computer Science, 3(2), pp.86-98.

[7] Hager W.W.and Zhang H., (2006),' A survey of nonlinear conjugate gradient methods, Pacific journal of Optimization, 2, pp. 35–58.

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

[9] Sneˇzana S. D., (2017), " New Hybrid Conjugate Gradient Method as a Convex Combination of LS and CD methods, Published by Faculty of Sciences and Mathematics, University of Niˇs, Serbia , pp. 1813-1825.

[10] Wang L. and Tian X., (2016),' A PRP-FR hybrid conjugate gradient algorithm for unconstrained optimization. Scholars journal of engineering and technology (SJET), 4, pp. 588-593.

[11] Zoutendijk G., (1970), " Nonlinear programming computational methods’’, in: Abadie J. (Ed.) Integer and nonlinear programming, North Holland, Amsterdam.