Bol. Soc. Paran. Mat. (3s.) **v. 38** 7 (2020): 227–231.
©SPM –ISSN-2175-1188 ON LINE ISSN-00378712 IN PRESS
SPM: www.spm.uem.br/bspm doi:10.5269/bspm.v38i7.46490

Global Convergence of Conjugate Gradient Method in Unconstrained Optimization Problems

Huda Y. Najm, Eman T. Hamed and Huda I. Ahmed

ABSTRACT: In this study, we propose a new parameter in conjugate gradient method. It is shown that the new method fulfills the sufficient descent condition with the strong Wolfe condition when inexact line search has been used. The numerical results of this suggested method also shown that this method outperforms to other standard conjugate gradient method.

Key Words: Unconstrained optimization, Conjugate Gradient method, Inexact line search, Global convergence.

Contents

1. Introduction

Regarding the following problem:

$$\min f(x), \quad x \in R^n \tag{1.1}$$

where f is a continuously differentiable function. The non linear conjugate gradient methods are efficient to solve this problem by iterative method at the $(k+1)$ iteration by the following iteration form:

$$x_{k+1} = x_k + \alpha_k d_k \tag{1.2}$$

Where the step length $\alpha_k > 0$ and d_k denoted by the search direction:

$$d_k = \begin{cases} -g_k & for \quad k = 1 \\ -g_k + \beta_k d_{k-1} & for \quad k > 1 \end{cases} \tag{1.3}$$

where β_k is a scalar conjugacy coefficient , there are some well-known formulas of his scalar such as:Hestenes-Stiefel (HS) [1],Fletcher-Reeves (FR) [2], Polak-Ribière

2010 *Mathematics Subject Classification:* 65k10.
Submitted February 04, 2019. Published September 01, 2019

(PR) [3],Conjugate Descent-Fletcher (CD) [4], Liu-Storey (LS) [5] and Dai - Yuan (DY) [6]. The convergence behavior of conjugate gradient methods are different. There are many convergence results with some line search conditions has been widely studied, there method can guarantee the descent property of each direction which provided the step length computed by carrying out a line search and its satisfies the strong Wolfe conditions.

2. New formula for β_k and the algorithm

In this section, a new coefficient of conjugate gradient using Hestenes- Steifel formula in the original numerator , and the denominator is introduced in the following formula:

$$\beta_k^{new} = \frac{g_k^T(g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1} + \mu \mid g_k^T d_{k-1} \mid}, \tag{2.1}$$

where $y_{k-1}(g_k - g_{k-1})$ and μ is positive constant.

Algorithm (2.1)

Step 1 : For the initial point $x_1 \in R^n$,Set $d_1 = -g_1$,$k = 1$,if $\|g_1\| \leq \varepsilon$, then stop.

Step 2: Set $d_k = -g_k$

Step 3 : Find $\alpha_k > 0$ satisfying the Strong Wolfe Conditions.

Step 4: Let $x_{k+1} = x_k + \alpha_k d_k$ and If $\|g_1\| \leq \varepsilon$ then stop .

Step 5 : Compute β_k by the new formula (6) where $\mathrm{y_{k-1}} = (\mathrm{g_k} - \mathrm{g_{k-1}})$ and $\mu > 1$, then generate d_k by (3).

Step 6 : If $k = n$ or $\frac{|g_k^T(g_k|}{\|g_k\|^2} \geq 0.2$ Powell restart [12],then go to step 2.

Step 7: Set $k := k + 1$, go to Step 4.

3. Sufficient Descent Property and Global Convergence Analysis

We make the following basic assumptions on the objective function in order to establish the global convergence results for the new algorithm:

Assumption (3.1) (see [14])

(i) f is bounded below on the level set $\Omega = \{x \in R^n : f(x) \leq f(x_0)\}$

(ii) In some neighborhood Ω_0 of Ω, f is differentiable and it is gradient $g(x)$ is Lipschitz continuous , namely ,there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq \|x - y\|, \forall x, y \in \Omega \tag{3.1}$$

Under these assumptions, there exists a constant $\varepsilon > 0$ such that

$$\|g_k\| \leq \varepsilon, \forall k \tag{3.2}$$

Lemma (3.2) Suppose that Assumption (3.1) holds, let the sequence $\{x_k\}$ generated by the algorithm (2.1) and the step length α_k satisfies Wolfe conditions, then

$$g_k^T d_k \leq -c\|g_k\|^2 \tag{3.3}$$

Proof: For the initial direction $k = 1$, since $d_0 = -g_0$, then $g_0^T \leq -\|g_0\|^2$ which satisfied (8). For some $k > 1$ and by using (3) and (5), we get

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \beta_k^{new} g_k^T d_{k-1} \\ &= -\|g_k\|^2 + \frac{g_k^T(g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1} + \mu \mid g_k^T d_{k-1} \mid} g_k^T d_{k-1} \\ &= -\|g_k\|^2 + \frac{g_k^T(g_k - g_{k-1})}{y_{k-1}^T d_{k-1} + \mu \mid g_k^T d_{k-1} \mid} g_k^T d_{k-1}. \end{aligned} \tag{3.4}$$

We have

$$d_{k-1}^T y_{k-1} = g_k^T d_{k-1} - g_{k-1}^T d_{k-1} \geq \sigma g_{k-1}^T d_{k-1} - g_{k-1}^T d_{k-1} = (\sigma - 1) g_{k-1}^T d_{k-1} > 0,$$

i.e $d_{k-1}^T y_{k-1} > 0$, and from the strong Wolfe conditions $\sigma g_k^T d_k \leq g(x_k + \alpha_k d_k)^T d_k \leq -\sigma g_k^T d_k$, we get

$$\begin{aligned} g_k^T d_k &\leq -\|g_k\|^2 + \frac{\|g_k\|^2}{\mu\sigma g_{k-1}^T d_{k-1}}(-\sigma g_{k-1}^T d_{k-1}) - \frac{g_k^T g_{k-1}}{\mu\sigma g_{k-1}^T d_{k-1}}(-\sigma g_{k-1}^T d_{k-1}) \\ &\leq -\|g_k\|^2 - \frac{\|g_k\|^2}{\mu} + \frac{g_k^T g_{k-1}}{\mu} \end{aligned} \tag{3.5}$$

From Powell restart condition $\mid g_k^T g_{k-1} > 0.2\|g_k\|^2$, [7] we have:

$$\begin{aligned} g_k^T d_k &\leq -\|g_k\|^2 - \frac{\|g_k\|^2}{\mu} - \frac{(0.2)\|g_k\|^2}{\mu} \\ &\leq -\left(1 + \frac{1}{\mu} + \frac{(0.2)}{\mu}\right)\|g_k\|^2 \\ &\leq -c\|g_k\|^2, \end{aligned} \tag{3.6}$$

where $c = 1 + \frac{1}{\mu} + \frac{(0.2)}{\mu}$, and $\mu > 1$.

Theorem (3.3) Consider the iteration method $x_{k+1} = x_k + \alpha_k d_k$ where d_k defined by (3),(5) and suppose that Assumption (3.1) holds. Then the new algorithm either stops at stationary point

$$\lim_{k \to \infty} inf\|g_k\| = 0 \tag{3.7}$$

4. Numerical Results

In this section, the main idea to report the performance of the new method on a set of test problems of (35) nonlinear unconstrained problems by using Fortran language. These test problems are contributed in CUTE, we can found the details of the test functions, in [8] and [9]. For each test function, the number of variables $n = 100, 200, ..., 500$. In order to evaluate the reliability of the new proposed method. Numerical Results by compare between the new-CG Method and Standard HS

Method by depend the following tools, n : dimension of the problem , $iter$: number of iterations, irs: number of restart , $fgcnt$:number of function and gradient evaluations., $time$: total time required to complete the evaluation process. $fxnew$: the value of function. And $gnorm$: the minimum gradient values, but we do not give the results of all test function due to page limit, see Table (1).

Table 1: Compare Numerical Results for NEW-CG Method and Standard HS- CG Method

Function test		Standard HS- CG Method		NEW-CG Method	
	N	fgcnt	fgcnt	fgcnt	fgcnt
DIXMAANA (CUTE)	100	7	14	7	14
	200	7	14	7	14
	300	5	10	5	10
	400	6	12	6	12
	500	6	12	6	12
LIARWHD (CUTE)	100	21	40	21	41
	200	16	36	16	36
	300	16	37	17	37
	400	17	39	17	39
	500	19	43	17	39
Extended Freudenstein & Roth	100	21	374	13	29
	200	14	29	14	31
	300	25	398	13	28
	400	48	947	16	37
	500	12	27	15	32
Extended Trigonometric Function	100	18	33	18	32
	200	22	39	21	37
	300	60	97	58	89
	400	26	49	24	45
	500	60	99	54	86

Function test		Standard HS- CG Method		NEW-CG Method	
	N	fgcnt	fgcnt	fgcnt	fgcnt
NONDIA (Shanno-78) (CUTE)	100	13	25	12	23
	200	9	17	9	17
	300	9	17	11	22
	400	9	17	10	20
	500	10	19	12	24
DIXMAANB (CUTE)	100	10	18	10	18
	200	11	19	11	19
	300	11	19	11	19
	400	11	19	11	19
	500	11	19	11	19
Extended Powell	100	66	125	48	92
	200	53	99	76	147
	300	66	122	58	114
	400	43	79	83	152
	500	58	112	73	135

5. Conclusion

We have proposed a new β_k, also we have provided proof the global convergence. The effectiveness of the new proposed method β_k^{new} has the good performance compared with other standard conjugate gradient method depended on the selected list of test functions problems. Comparison in Total for all Function Test of new-CG Method against HS Method for 35 Test Problems:

References

1. N. Andrei, Advanced Modeling and optimization **10**, 147–161 (2008).
2. R. Fletcher and C. M. Reeves, The computer journal **7**, 149–154 (1964).
3. E. Polak and G. Ribiere, Revue française d'informatique et de recherche opérationnelle. Série rouge **3**, 35–43 (1969).
4. R. Fletcher, *Practical methods of optimization* (John Wiley & Sons, 2013).
5. Y. Liu and C. Storey, Journal of optimization theory and applications **69**, 129–137 (1991).
6. Z. Dai and F. Wen, Numerical Algorithms **59**, 79–93 (2012).
7. M. J. D. Powell, Mathematical programming **12**, 241–254 (1977).
8. I. Bongartz, A. R. Conn, N. Gould, and P. L. Toint, ACM Transactions on Mathematical Software (TOMS) **21**, 123–160 (1995).
9. N. Andrei, Optimization methods and software **24**, 89–104 (2009).
10. Y.-H. Dai and Y. Yuan, SIAM Journal on optimization **10**, 177–182 (1999).
11. M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, Vol. 49 (NBS Washington, DC, 1952).
12. G. Li, C. Tang, and Z. Wei, Journal of Computational and Applied Mathematics **202**, 523–539 (2007).
13. Z. Wei, G. Li, and L. Qi, Applied Mathematics and Computation **175**, 1156–1188 (2006).
14. H. Yabe and N. Sakaiwa, Journal of the Operations Research Society of Japan **48**, 284–296 (2005).

H. Y. Najm,
Department of Mathematics, College of Science,
University of Duhok, Iraq.
E-mail address: `huda.najm@uod.ac`

and

Eman T. Hamed,
Department of Operation Research and Intelligent Techniques,
College of Computer Sciences and mathematics,
University of Mosul, Iraq.
E-mail address: `emantarik84@gmail.com`

and

Huda I. Ahmed,
Deepmala,
Department of Operation Research and Intelligent Techniques,
College of Computer Sciences and mathematics,
University of Mosul, Iraq.
E-mail address: `hudaea72@gmail.com`