IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319–765X. Volume 10, Issue 3 Ver. V (May-Jun. 2014), PP 78-82
www.iosrjournals.org

New Iterative Conjugate Gradient Method for Nonlinear Unconstrained Optimization Using Homptopy Technique

Salah Gazi Shareef and Hussein Ageel Khatab
Dept. of Mathematics, Faculty of Science, University of Zakho, Kurdistan Region-Iraq

Abstract: *A new hybrid conjugate gradient method for unconstrained optimization by using homotopy formula , we computed the parameter β_k as aconvex combination of β^{PR} (Polak-Ribiere (PR))[9] and β^{BA} (Al-Bayati and Al-Assady (BA))[1].*

Keywords: *Unconstrained optimization , line search , conjugate gradient method . homotopy formula.*

I. Introduction

Consider the nonlinear following unconstrained optimization problem

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

Where f : $R^n \rightarrow R$ is continuously differentiable function whose gradient is denoted by $g(x)$. For solving this problem , starting from an initial vector $x_0 \in R^n$, a nonlinear gradient algorithms generates a sequence $\{x_k\}$ as:

$$x_{k+1} = x_k + \alpha_k d_k \quad ,k=0,1,\ldots,n \tag{1.2}$$

Where $\alpha_k > 0$ is obtained by line search, the direction d_k is generated by

$$d_k = \begin{cases} -g_k, if & k = 0, \\ -g_k + \beta_{k-1} d_{k-1}, if & k \geq 1, \end{cases} \tag{1.3}$$

and $d_0 = -g_0$,

and $\beta_k \in R$ is a parameter which determines the different conjugate gradient methods . There are some well-known formulas which are given as follows :

$$\beta^{PRP} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{\|g_k\|^2} \tag{1.4}$$

$$\beta^{FR} = \frac{g_{k+1}^T g_{k+1}}{\|g_k\|^2} \tag{1.5}$$

$$\beta^{HS} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{(g_{k+1} - g_k)^T d_k} \tag{1.6}$$

$$\beta^{DY} = \frac{g_{k+1}^T g_{k+1}}{(g_{k+1} - g_k)^T d_k} \tag{1.7}$$

$$\beta^{CD} = \frac{g_{k+1}^T g_{k+1}}{-d_k^T g_k} \tag{1.8}$$

$$\beta^{BA} = \frac{\|y_k\|^2}{d_k^T y_k} \tag{1.9}$$

Where g_{k-1} and g_k are gradients $\nabla f(x_{k-1})$ and $\nabla f(x_k)$ of f(x) at the point x_{k-1} and x_k , respectively , $\|.\|$ denotes the Euclidian norm of vectors. The line search in the conjugate gradient algorithms often is based on the standard Wolfe conditions:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k \tag{1.10}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k \tag{1.11}$$

where $0 < \delta < \sigma < 1$.

The strong Wolfe line search corresponds to: that

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k \tag{1.12}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \leq |\sigma g_k^T d_k| \tag{1.13}$$

where $0 < \delta < \sigma < 1$.[10]

II. Hybrid Conjugate Gradient Algorithms

The hybrid conjugate gradient algorithms are combinations of different conjugate gradient algorithms , mainly with the purpose of avoiding the jamming phenomenon.[2],

The methods of (FR)[5],(DY) [3] and (CD)[4] have strong convergence properties, but they may have modest practical performance due to jamming.

On the other hand, the methods of (PR)[9],(HS)[6] and (LS)[8] may not always be convergent, but the often have better computational performances.[2]

III. New Hybrid conjugate gradient algorithm

The iterates x_0 , x_1, x_2,… of new hybrid conjugate gradient algorithm computed by means of the recurrence ($x_{k+1} = x_k + \propto_k d_k$), where the stepsize $\propto_k > 0$ is determined according to the Wolf line search condition (1.10) and (1.11), and the directions d_k are computed by the rule :

$$d_{k+1} = -g_{k+1} + \beta_k^{NEW} d_k, \tag{3.1}$$

Where

$$\beta_k^{NEW} = (1-\theta_k)\beta^{PR} + \theta_k \beta^{BA} ,$$

$$=(1-\theta_k)\frac{(g_{k+1}^T y_k)}{\|g_k\|^2} + \theta_k \frac{\|y_k\|^2}{d_k^T y_k} \tag{3.2}$$

And θ_k is a scalar parameter satisfying $0 \le \theta_k \le 1$ which will be determined in a specific way to be described later. Observe that if $\theta_k = 0$, then $\beta_k^{NEW} = \beta^{PR}$ and if $\theta_k = 1$, then $\beta_k^{NEW} = \beta^{BA}$, On the other hand if $0 < \theta_k < 1$, then $\beta_k^{NEW\ 1}$ is a convex combination of $\beta^{PR} and\ \beta^{BA}$.

The parameter θ_k is selected in such a way that at every iteration the conjugacy condition is satisfied independently of the line search.

Obviously

$$d_{k+1} = -g_{k+1} + (1-\theta_k)\frac{(g_{k+1}^T y_k)}{\|g_k\|^2} d_k + \theta_k \frac{\|y_k\|^2}{d_k^T y_k} d_k , \tag{3.3}$$

So, multiply both sides of above equation by y_k and we use conjugacy condition($d_{k+1}^T y_k$= 0)

$$0 = -g_{k+1}^T y_k + (1-\theta_k)\frac{(g_{k+1}^T y_k)}{\|g_k\|^2} d_k^T y_k + \theta_k \frac{\|y_k\|^2}{d_k^T y_k} d_k^T y_k , \tag{3.4}$$

Implies that

$$g_{k+1}^T y_k - \frac{(g_{k+1}^T y_k)}{\|g_k\|^2} d_k^T y_k = -\theta_k \frac{(g_{k+1}^T y_k)}{\|g_k\|^2} d_k^T y_k + \theta_k \|y_k\|^2$$

$$\theta_k = \frac{g_{k+1}^T y_k - \frac{(g_{k+1}^T y_k)}{\|g_k\|^2} d_k^T y_k}{-\frac{(g_{k+1}^T y_k)}{\|g_k\|^2} d_k^T y_k + \|y_k\|^2}$$

Finally, we have

$$\theta_k = \frac{(g_{k+1}^T y_k)\|g_k\|^2 - (g_{k+1}^T y_k)(d_k^T y_k)}{\|y_k\|^2 \|g_k\|^2 - (g_{k+1}^T y_k)(d_k^T y_k)} \tag{3.5}$$

Theorem 3.1:- assume that d_k is descent direction and $\propto_k$ in the algorithm (1.2),(3.2)and (3.5) determined by the wolfe line search (1.10)and (1.11). If $0 < \theta_k < 1$, then the d_{k+1} is given by (3.3) is a descent direction.

Proof:-

Multiply both sides of (2.3.3) by g_{k+1}, we have

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

Since $0 < \theta_k < 1$, then

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \frac{\|y_k\|^2}{d_k^T y_k} d_k^T g_{k+1} + \frac{(g_{k+1}^T y_k)}{\|g_k\|^2} d_k^T g_{k+1} \tag{3.6}$$

The prove is complete if the step length $\propto_k$ is chosen by an exact line search which requires $d_k^T g_{k+1} = 0$.

If the step length $\propto_k$ is chosen by an inexact line search which requires $d_k^T g_{k+1} \neq 0$,

We know that the first two terms of equation (3.6) are less than or equal to zero because the algorithm of (PR) is satisfies the descent condition (i.e)

$- \|g_{k+1}\|^2 + \frac{\|y_k\|^2}{d_k^T y_k} d_k^T g_{k+1} \le 0,$

Now , we will explain the third term is less than or equal to zero

We know that

$$\sigma d_k^T g_k \le d_k^T y_k \tag{3.7}$$

Where σ is positive ,multiply both sides by (-1) , we have

$$-\sigma d_k^T g_k \geq -d_k^T y_k \tag{3.8}$$

Implies that

$$\|g_k\|^2 \geq -\frac{1}{\sigma} d_k^T y_k \tag{3.9}$$

Then

$$\frac{(g_{k+1}^T y_k)}{\|g_k\|^2} d_k^T g_{k+1} \leq -\sigma \frac{(g_{k+1}^T y_k)}{d_k^T y_k} d_k^T g_{k+1} \leq -\sigma \|g_{k+1}\|^2 \leq 0.$$

Because $\frac{(g_{k+1}^T y_k)}{d_k^T y_k} d_k^T g_{k+1} \leq \|g_{k+1}\|^2$

Then the proof is completed.■

3.1 The global convergence

The following assumption are often needed to prove the convergence of the nonlinear conjugate gradient method(see [7]).

Assumption:-

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

ii- In some neighborhood N of L , the objective function f is continuously differentiable, and it is gradient is Lipschitz continuous, i.e., there exists a constant $k > 0$ such that $\|\nabla f(x) - \nabla f(\overline{x})\| \leq K\|x - \overline{x}\|$, for all $x, \overline{x} \in N$.

Under above Assumption , there exists a constant μ such that $\|\nabla f(x)\| \leq \mu$, for all $x \in L$,

Lemma :- Let assumption (i) and (ii) hold and consider any conjugate gradient method (3.2) and (3.5) , where d_k is a descent direction and α_k is obtained by the strong Wolfe line search. If

$$\sum_{k\geq 1} \frac{1}{\|d_k\|^2} = \infty, \tag{3.1.1}$$

Then

$$\lim\inf_{k\to\infty} \|g_k\| = 0. \tag{3.1.2}$$

For uniformly convex function which satisfy the above assumptions, we can prove that the norm of d_{k+1} given by () is bounded above. Assume that the function f is uniformly convex function, i.e., there exists a constant $\Gamma \geq 0$, such that for all $x, y \in L$,

$$\left(\nabla f(x) - \nabla f(\overline{x})\right)^T (x - \overline{x}) \geq \Gamma\|x - \overline{x}\|^2 \tag{3.1.3}$$

and the step length α_k is given by the strong Wolfe line search.

$$f(x_{k+1}) \leq f(x_k) + \sigma_1 \alpha_k g_k^T d_k, \tag{3.1.4}$$

$$|g_{k+1}^T d_k| \leq -\sigma_2 g_k^T d_k. \tag{3.1.5}$$

Using lemma the following result can be proved.

Theorem 3.2:- Suppose that the assumption (i) and (ii) hold. Consider the algrithm (1.2),(3.2) and (3.5) , where $0 \leq \theta_k \leq 1$ and α_k is obtained by the strong Wolfe line search (1.12) and (1.13). If d_k is tends to zero and there exists nonnegative constants δ_1 and δ_2 such that

$\|g_k\|^2 \geq \delta_1\|v_k\|^2$, and $\|g_{k+1}\|^2 \leq \delta_2\|v_k\|$

And f is a uniformly convex function, then

$$\lim_{k\to\infty} g_k = 0. \tag{3.1.6}$$

Proof:-

From (3.1.3) it follows that

$$y_k^T v_k \geq \Gamma\|v_k\|^2.$$

Since $0 \leq \theta_k \leq 1$, from uniform convexity and (3.2) we have

$$|\beta_k^{NEW}| = \left|\frac{(g_{k+1}^T y_k)}{\|g_k\|^2} + \frac{\|y_k\|^2}{d_k^T y_k}\right| \leq \left|\frac{(g_{k+1}^T y_k)}{\|g_k\|^2}\right| + \left|\frac{\|y_k\|^2}{d_k^T y_k}\right|$$

$$|\beta_k^{NEW}| \leq \frac{\|g_{k+1}\|\|y_k\|}{\delta_1\|v_k\|^2} + \frac{\|y_k\|^2}{\Gamma\|v_k\|^2}$$

$$|\beta_k^{NEW}| \leq \frac{\mu\|y_k\|}{\delta_1\|v_k\|^2} + \frac{\|y_k\|^2}{\Gamma\|v_k\|^2} \tag{3.1.7}$$

From Lipschitz Condition

$$\|y_k\| \leq K\|v_k\|,$$

We get

$$\|y_k\|^2 \leq K^2\|v_k\|^2$$

Then (3.1.7) gives

$$|\beta_k^{NEW}| \le \frac{\mu K}{\delta_1 \|v_k\|} + \frac{K^2}{\Gamma} \qquad (3.1.8)$$

Hence

$$\|d_{k+1}\| \le \|g_{k+1}\| + |\beta_k^{NEW}| \|d_k\|$$

Then

$$\|d_{k+1}\| \le \mu + \frac{\alpha\mu K}{\delta_1} + \frac{K^2}{\Gamma}\alpha\|v_k\| \qquad (3.1.9)$$

Since $\|v_k\| \to 0$ as $\to \infty$, then there exist a constant $\omega > 0$ such that

$$\|v_k\| \le \omega$$

Hence (3.1.9) gives

$$\|d_{k+1}\| \le \mu + \frac{\alpha\mu K}{\delta_1} + \frac{K^2}{\Gamma}\alpha\omega$$

Which implies that (3.1.1) is true. so, by Lemma we have (3.1.2), which for uniformly convex functions is equivalent to (3.1.6).■

3.2 Algorithm of New hybrid Conjugate Gradient coefficient:

step (1) :- set k=0 , select the initial point x_k .

step(2) :- $g_k = \nabla f(x_k)$, If $g_k = 0$,then stop .

else

set $d_k = - g_k$.

step (3) :- compute $\alpha_k > 0$ satisfying the wolfe lline search condition to minimize $f(x_{k+1})$.

step (4) :- $x_{k+1} = x_k + \alpha_k d_k$.

step (5) :- $g_{k+1} = \nabla f(x_{k+1})$, If $g_{k+1} = 0$,then stop .

step (6):- compute θ_k as in (3.5).

Step (7):-if $0 < \theta_k < 1$, then compute β_k^{NEW} as in (3.2). If $\theta_k \ge 1$, then set $\beta_k^{NEW} = \beta^{BA}$. If $\theta_k \le 0$, then set $\beta_k^{NEW} = \beta^{PR}$.

step (8) :- $d_{k+1} = - g_{k+1} + \beta_k^{NEW} d_k$.

step (9) :- If k=n then go to step 2,

else

k=k+1 and go to step 3.

3.3 Numrical Results:-

This section is devoted to test the implementation of the new methods. We compare the hybrid algorithm first suggestion with standard Polak-Ribiere (PR) and Al-Bayati (BA) ,the comparative tests involve well-known nonlinear problems (standard test function) with different dimension 4 ≤n≤5000, all programs are written in FORTRAN95 language and for all cases the stopping condition is $\|g_{k+1}\|_\infty \le 10^{-5}$ The results are given in table is specifically quote the number of functions NOF and the number of iteration NOI .experimental results in table confirm that the new CG method is superior to standard CG method with respect to the NOI and NOF.

Table

Comparative Performance of the three algorithms (Standard PR,BA and New formula)

Test fun.	N	Standard formula (PR)		Standard formula (BA)		New formula	
		NOI	NOF	NOI	NOF	NOI	NOF
Powell	4	60	162	3493	7010	65	170
	100	119	252	5205	10424	105	276
	500	505	1028	1001	2014	502	1062
	1000	1002	2009	2098	4208	637	1332
	3000	2531	5067	3024	6064	879	1816
	5000	2817	5639	5001	10014	1008	2074
Wood	4	33	74	1021	2395	26	60
	100	103	213	629	1691	103	213
	500	108	223	575	1345	108	223
	1000	110	227	730	1661	109	225
	3000	188	282	750	1646	186	378
	5000	130	267	676	1515	130	267

Cubic	4 100 500 1000 3000 5000	16 15 15 15 15 15	42 40 40 40 40 40	25 46 46 46 49 49	66 109 109 109 105 105	15 14 14 14 14 14	43 40 40 40 40 40
Rosen	4 100 500 1000 3000 5000	22 22 22 22 22 22	63 61 61 61 61 61	68 113 484 502 526 542	247 248 990 1026 1074 1106	23 17 17 17 17 17	66 52 52 52 52 52
Beel	4 100 500 1000 3000 5000	10 9 9 9 9 9	23 23 23 23 23 23	25 40 40 40 40 40	59 89 89 89 89 89	9 9 9 9 9 9	22 23 23 23 23 23
Non Digonal	4 100 500 1000 3000 5000	23 22 24 * * *	59 62 68 * * *	40 105 * * 84 *	115 462 * * 203 *	23 22 22 22 22 22	61 60 59 59 59 59
EDGR	4 100 500 1000 3000 5000	6 6 6 6 6 6	16 16 16 16 16 16	9 33 35 36 37 37	24 70 74 76 78 78	6 6 6 6 6 6	15 15 15 15 15 15
Extended PSC1	4 100 500 1000 3000 5000	36 36 38 42 42 42	78 78 83 88 88 88	66 66 68 68 68 68	220 220 222 222 222 222	34 34 36 36 36 36	77 77 82 82 82 82
Generalized PSC1	4 100 500 1000 3000 5000	60 104 106 108 108 108	166 276 278 282 282 282	120 124 124 124 124 124	352 358 358 358 358 358	60 102 104 104 104 104	166 272 276 276 276 276
Q-Central	4 100 500 1000 3000 5000	25 19 23 23 23 23	163 115 172 172 172 172	45 * * * * *	227 * * * * *	25 20 20 21 21 21	148 132 132 147 147 147

IV. Conclusion

In this paper we have presented a new hybrid conjugate gradient method in which a famous parameter β_k is computed as a convex combination of $\beta_k^{PR}$and β_k^{BA} and comparative numerical performances of a number of well known conjugate gradient algorithms Polak-Ribiere-Polyak (PR) and Al-Bayati and Al-Assady (BA). We saw that the performance profile of our method was higher than those of the well established conjugate gradient algorithms PR and BA .

Refrences

[1]. A.Y. Al-Bayati, and N.H. Al-Assady, Conjugate Gradient Method,Technical Research report, No. 1,school of computer studies , Leeds Uuiversity .

[2]. N. Andrei, Hybrid Conjugate Gradient Algorithm for Unconstrained Optimization , J.Optim. Theory Appl., 141, 2009, 249-264.

[3]. Y.H. Dai, and Y.Yuan, A nonlinear Conjugate Gradient Method With a Strong Global Convergence Property, SIAM J. Optimization, 10,1999, Pp. 177-182.

[4]. R. Fletcher, Practical Methods of Optimization,(Vol. 1: Unconstrined Optimization , John Wiley and Sons, New York, 1987).

[5]. R. Fletcher, and C. Reeves, Function Minimization by Conjugate Gradients, Comput. J.,7, 1964, Pp. 149-154.

[6]. M.R. Hestenes, and E.L.Stiefel, Methods of Conjugate Gradients for Solving Linear Systems, J.Research Nat. Bur. Standards, 1952,Pp. 409- 436,49 .

[7]. T. Hirotaka and Y. Hiroshi ,A Hybrid Conjugate Gradient Method for Unconstrained Optimization , Dep. Of Math. Information Science , Tokyo Univesity of Science , 1-3 Kagurazaka,Shinjuku , Tokyo 162-8601, Ja pan, (2004).

[8]. Y. Liu , and C. Storey, Efficient generalized Conjugate Gradient Algorithms , Part 1: Theory. JOTA, ,69, 1991, Pp.129-137.

[9]. E. Polak and G. Ribiere ,, Note sur Ia convergence de directions Conjugate , Rev. Francaise Informat Recherche Operationelle, 3e Annee , 16,1969,Pp. 35-43.

[10]. P. Wolfe, Convergence conditions for ascent methods, SIAM Rev., 11,1969,226-235.