International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 3 Issue 4, April-2014, pp: (179-186), Impact Factor: 1.252, Available online at: www.erpublications.com

Extended Spectral Nonlinear Conjugate Gradient methods for solving unconstrained problems

Dr. Basim A. Hassan
Asst. Professor, College of Computers Sciences and Math., University of Mosul, Iraq

Abstract: In this paper, we present extension forms of Dai, Yuan (DY), Fletcher, Reveres (FR) and Conjugate Descent (CD) CG algorithms. The extended method have the sufficient descent and globally convergence properties under certain conditions. These new algorithms are tested on some standard test functions and compared with the original FR algorithm showing considerable improvements over all these comparisons.

Keywords: Conjugate gradient method, Spectral Conjugate gradient method, sufficient descent property, global convergent methods.

Introduction

The nonlinear conjugate gradient (CG) method is designed to solve the following unconstrained optimization problem

$$\min\left\{ f(x) \;\middle|\; x \in R^n \right\} \qquad \text{..........(1)}$$

where $f : R^n \to R$ is a continuously differentiable nonlinear function whose gradient is denoted by g. Due to its simplicity and its very low memory requirement, the CG method has played a special role for solving large scale nonlinear optimization problems. The iterative formula of the CG method is given by

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

where $\alpha_k > 0$ is a step length which is computed by carrying out a line search and satisfies the standard Wolfe (SW) conditions :

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta_1 \alpha_k d_k^T g_k \qquad \text{..........(3)}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \delta_2 d_k^T g_k \qquad \text{..........(4)}$$

with $0 < \delta_1 < \delta_2 < 1$, and d_{k+1} is the search direction defined by

$$d_{k+1} = \begin{cases} -g_1 & k = 1, \\ -g_{k+1} + \beta_k d_k & k > 1, \end{cases} \qquad \text{..........(5)}$$

where d_k is a descent direction. Different conjugate gradient algorithms correspond to different choices for the scalar parameter β_k see [7]. The well-known formula of β_k are given by

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}, \qquad \text{..........(6)}$$

$$\beta_k^{CD} = \frac{g_{k+1}^T g_{k+1}}{-g_k^T d_k}, \qquad \text{..........(7)}$$

$$\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{y_k^T d_k}, \qquad \text{..........(8a)}$$

which are called Fletcher and Reeves (FR) [4], Conjugate Descent (CD) [3] and Dai and Yuan (DY) [2] , respectively. In fact, utilizing (5), β_k^{DY} can be rewritten as :

$$\beta_k^{DY} = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}. \qquad \text{..........(8b)}$$

In [6] modified conjugate gradient methods are given by the rule

$$d_{k+1} = -\left(1 + \beta_k \frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2}\right) g_{k+1} + \beta_k d_k \qquad \text{..........(9)}$$

where β_k is one of the values in (6) or (7) or (8).

Zhang and Wang [8] proposed a general form of conjugate gradient methods are given by the rule

$$d_{k+1} = -\left(\xi + \beta_k \frac{y_k^T d_k}{\|g_{k+1}\|^2}\right) g_{k+1} + \beta_k d_k \qquad \text{..........(10)}$$

where β_k is one of the values in (6) or (7) or (8).

The paper is organized as follows. In section (1) is the introduction. In section (2) we present the extended spectral method and algorithm. Section (3) show that the search direction generated by this proposed algorithm at each iteration satisfies the sufficient descent condition and establishes the global convergence analysis. Section (4) establishes some numerical results to show the effectiveness of the proposed CG-method and Section (5) gives a brief conclusions .

Extended spectral conjugate gradient method and algorithm

In this paper we suggest a new type of spectral conjugate gradient methods for solution of the $\min f(x)$. In [5] we consider a condition that a descent search direction is generated, and we extend the DY method. We make such a direction inductively. Suppose that the current search direction d_k is a descent direction, namely, $g_k^T d_k < 0$ at the k^{th} iteration. Now we need to find a β_k that produces a descent search direction d_{k+1}. This requires that

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_k g_{k+1}^T d_k . \qquad \text{..........(11)}$$

Letting γ_{k+1} be a positive parameter, we define

$$\beta_k = \frac{\|g_{k+1}\|^2}{\tau_{k+1}} . \qquad \text{..........(12)}$$

Equation (11) is equivalent to

$$\tau_{k+1} > g_{k+1}^T d_k . \qquad \text{..........(13)}$$

Taking the positively of γ_{k+1} in to consideration, we have

$$\tau_{k+1} > \max \{ g_{k+1}^T d_k ,\ 0 \}. \qquad \text{..........(14)}$$

Therefore if condition (14) is satisfied for all k, the conjugate gradient method with (12) produces a descent search direction at every iteration. From (12) we can get various kinds of conjugate gradient method by choosing various τ_{k+1}.

Hideaki and Yasushi proposed a new conjugate gradient method which was obtained by modifying the DY method and called MDY method. A nice property of the MDY method is that it generates sufficient descent directions. The parameter β_k in MDY method is given by

$$\beta_k^{MDY} = \frac{g_{k+1}^T g_{k+1}}{\tau_{k+1}} \qquad \text{..........(15)}$$

where

$$\tau_{k+1} = \frac{2}{\alpha_k}(f_k - f_{k+1}) \qquad \text{..........(16)}$$

The definition of search direction and Formula (15) ensure that

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_k^{MDY} g_{k+1}^T d_k = (-\tau_{k+1} + g_{k+1}^T d_k)\beta_k^{MDY}, \qquad \text{..........(17)}$$

and hence,

$$0 < \beta_k = \frac{-g_{k+1}^T d_{k+1}}{\tau_{k+1} - g_{k+1}^T d_k} < \frac{-g_{k+1}^T d_{k+1}}{-g_k^T d_k} = \frac{\psi_{k+1}}{\psi_k}. \quad \text{..........(18)}$$

performs more effective More details can be found in [5].

Let us try to derive a new type method we need the next direction d_{k+1} to be descent. Assume that $\beta_k > 0$. By this, we have for any $\xi \in (0, 1]$, and from (13) the following inequality holds :

$$\begin{aligned} \beta_k \tau_{k+1} &> \beta_k g_{k+1}^T d_k \\ \beta_k \tau_{k+1} &> -\xi \|g_{k+1}\|^2 + \beta_k g_{k+1}^T d_k \ , \end{aligned} \quad \text{..........(19)}$$

i.e. ,

$$-\xi \|g_{k+1}\|^2 - \beta_k \tau_{k+1} + \beta_k g_{k+1}^T d_k < 0. \quad \text{..........(20)}$$

Now we can rewrite the above inequality as

$$g_{k+1}^T \left[-\left(\xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|^2} \right) g_{k+1} + \beta_k d_k \right] < 0. \quad \text{..........(21)}$$

Hence, we obtain our new directions as follows :

$$d_{k+1} = -\left(\xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|^2} \right) g_{k+1} + \beta_k d_k . \quad \text{..........(22)}$$

Then we can rewrite (22) as

$$d_{k+1} = -\varphi_k \, g_{k+1} + \beta_k d_k . \quad \text{..........(23)}$$

where

$$\varphi_k = \xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|^2}. \quad \text{..........(24)}$$

This method includes the Zhang and Wang (ZW) method as a special case. By setting $\tau_{k+1} = y_k^T d_k$, direction (22) reduces to the Zhang and Wang (ZW) method which defined in (10).

Now we can obtain the a new conjugate gradient algorithms, as follows :
The New Algorithm (2.1)

Step 1. Initialization : Select $x_1 \in R^n$ and the parameters $0 < \delta_1 < \delta_2 < 1$. Compute $f(x_1)$ and g_1. Consider $d_1 = -g_1$ and set the initial guess $\alpha_1 = 1/\|g_1\|$.

Step 2. Test for continuation of iterations. If $\|g_{k+1}\| \le 10^{-6}$, then stop. else step3.

Step 3. Line search : Compute $\alpha_{k+1} > 0$ satisfying the Wolfe line search condition (6) and update the variables $x_{k+1} = x_k + \alpha_k d_k$.

Step 4. Conjugate gradient parameter which defined in (6) or (7) or (8).

Step 5. Direction computation d_{k+1} which defined in (22).If the restart criterion of Powell $|g_{k+1}^T g_k| \ge 0.2 \, \|g_{k+1}\|^2$, is satisfied, then set $d_{k+1} = -g_{k+1}$ otherwise define $d_{k+1} = d$.

Global Convergence

In this section, we establish convergence of the proposed method, the following assumptions for the objective function are needed.

Assumption (3.1)

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

ii- In some neighborhood U of L, $f(x)$ is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $\mu > 0$ such that

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

Assumption 3.1 imply that there exists a positive constant γ such that

$$\|g_{k+1}\| \le \gamma \quad , \quad \forall x \in U \ . \qquad ....(26)$$

Here we have to present sufficient descent property.

Theorem 1

Let $\{x_{k+1}\}$ and $\{d_{k+1}\}$ be generated by (2) and (22), where α_k satisfies Wolfe line search conditions, then holds of the sufficient descent property

$$g_{k+1}^T d_{k+1} < -c\|g_{k+1}\|^2 \ . \qquad ....(27)$$

Proof :

Then conclusion can be proved by induction. When $k=0$, we have $g_0^T d_0 < -\|g_0\|^2 < 0$. Suppose that $g_k^T d_k < -c\|g_k\|^2$. From (13) and (22) we have

$$g_{k+1}^T d_{k+1} = g_{k+1}^T\left[-\left(\xi+\beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|^2}\right)g_{k+1}+\beta_k d_k\right]$$

$$= -\xi\|g_{k+1}\|^2 - \beta_k\tau_{k+1} + \beta_k g_{k+1}^T d_k \qquad ....(28)$$

$$\le -\xi\|g_{k+1}\|^2 - \beta_k g_{k+1}^T d_k + \beta_k g_{k+1}^T d_k$$

$$g_{k+1}^T d_{k+1} \le -\xi\|g_{k+1}\|^2 \le -c\|g_{k+1}\|^2 \ . \qquad ....(29)$$

where $c=\xi$. Thus the theorem is proved.

The following Lemma [9] is the result for general iterative methods :

Lemma 1

Suppose that Assumption 2.2 is satisfied and consider any method with Eq. (2), where α_k satisfies Eqs. (11) and (12) . Then,

$$\sum_{i=1}^{\infty}\frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} < \infty . \qquad ....(30)$$

From the previous analysis, we can get the following global convergence result for new Algorithm.

Theorem 2

Suppose that Assumption (3.1) holds, and these methods have the satisfies sufficient descent condition with $c=\xi$. Then these method are globally convergent, one has

$$\liminf_{k\Rightarrow\infty}\|g_{k+1}\|=0 \quad \text{or} \quad \sum_{k=1}^{\infty}\frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} < +\infty . \qquad ....(31)$$

Proof :

Now we will prove global convergence. We suppose that the theorem is not true. Suppose by contradiction that there exists $\varepsilon_1 > 0$ such that

$$d_0 = -g_0 \quad \text{and} \quad \|g_{k+1}\| > \varepsilon_1 \qquad ....(32)$$

By squaring the two sides of (22) and transferring and trimming, we get :

$$\|d_{k+1}\|^2 = \beta_k^2\|d_k\|^2 - \left(\xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|^2}\right)^2 \|g_{k+1}\|^2 - 2\left(\xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|^2}\right) d_{k+1}^T g_{k+1}. \quad \text{..........(33)}$$

Dividing the previous in equation by $(d_{k+1}^T g_{k+1})^2$, we get :

$$\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} = \frac{\beta_k^2\|d_k\|^2}{(d_{k+1}^T g_{k+1})^2} - \left(\xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|^2}\right)^2 \frac{\|g_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} - \frac{2}{(d_{k+1}^T g_{k+1})}\left(\xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|^2}\right) \quad \text{..........(34)}$$

$$\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} = \frac{\beta_k^2\|d_k\|^2}{(d_{k+1}^T g_{k+1})^2} - \left(\xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|^2}\right)^2 \frac{\|g_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} - \frac{2}{(d_{k+1}^T g_{k+1})}\left(\xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|^2}\right) - \frac{1}{\|g_{k+1}\|^2} + \frac{1}{\|g_{k+1}\|^2} \quad \text{..........(35)}$$

$$\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} = \frac{\beta_k^2\|d_k\|^2}{(d_{k+1}^T g_{k+1})^2} - \left[\frac{1}{(d_{k+1}^T g_{k+1})}\left(\xi + \beta_k \frac{\tau_{k+1}}{\|g_{k+1}\|^2}\right) + \frac{1}{\|g_{k+1}\|}\right]^2 + \frac{1}{\|g_{k+1}\|^2} \quad \text{..........(36)}$$

$$\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} \le \frac{\beta_k^2\|d_k\|^2}{(d_{k+1}^T g_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2} \quad \text{..........(37)}$$

a. When $\beta_k = \beta_k^{DY}$. Then by $(8b)$

$$\beta_k^{DY} = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k} \quad \text{..........(38)}$$

and applying (37) , we have

$$\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} \le \frac{\|d_k\|^2}{(d_k^T g_k)^2} + \frac{1}{\|g_{k+1}\|^2} \quad \text{..........(39)}$$

Noting that

$$\frac{\|d_1\|^2}{(d_1^T g_1)^2} = \frac{1}{\|g_1\|^2}, \quad \text{..........(40)}$$

With this, from (32) we have

$$\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} \le \sum_{i=1}^{k+1} \frac{1}{\|g_{i+1}\|^2} \le \frac{k}{\varepsilon_1^2} . \quad \text{..........(41)}$$

Then we get

$$\frac{(d_{k+1}^T g_{k+1})^2}{\|d_{k+1}\|^2} \ge \frac{\varepsilon_1^2}{k} . \quad \text{..........(42)}$$

Which indicates

$$\sum_{k=1}^{\infty}\frac{(d_{k+1}^T g_{k+1})^2}{\|d_{k+1}\|^2} \geq \sum_{k=1}^{\infty}\frac{\varepsilon_1^2}{k} = \infty \quad . \qquad ....(43)$$

This is a contradiction to the (30) .

b. When $\beta_k = \beta_k^{FR}$. Then from (37) and sufficient descent condition with $c = \xi$.

$$\begin{aligned}\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} &\leq \frac{\|g_{k+1}\|^4}{\|g_k\|^4}\frac{\|d_k\|^2}{(d_{k+1}^T g_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2} \\ &\leq \frac{\|g_{k+1}\|^4}{\|g_k\|^4}\frac{\|d_k\|^2}{c^2\|g_{k+1}\|^4} + \frac{1}{\|g_{k+1}\|^2} \qquad ....(45) \\ &\leq \frac{\|d_k\|^2}{c^2\|g_k\|^4} + \frac{1}{\|g_{k+1}\|^2}\end{aligned}$$

we also have

$$\sum_{k=1}^{\infty}\frac{(d_{k+1}^T g_{k+1})^2}{\|d_{k+1}\|^2} \geq \sum_{k=1}^{\infty}\frac{\varepsilon_1^2}{k} = \infty \quad .$$

c. When $\beta_k = \beta_k^{CD}$. Then from (37) and sufficient descent condition with $c = \xi$.

$$\begin{aligned}\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} &\leq \frac{\|g_{k+1}\|^4}{(d_k^T g_k)^2}\frac{\|d_k\|^2}{(d_{k+1}^T g_{k+1})^2} + \frac{1}{\|g_{k+1}\|^2} \\ &\leq \frac{\|g_{k+1}\|^4}{c^2\|g_k\|^4}\frac{\|d_k\|^2}{c^2\|g_{k+1}\|^4} + \frac{1}{\|g_{k+1}\|^2} \qquad ....(46) \\ &\leq \frac{\|d_k\|^2}{c^4\|g_k\|^4} + \frac{1}{\|g_{k+1}\|^2}\end{aligned}$$

we also have

$$\sum_{k=1}^{\infty}\frac{(d_{k+1}^T g_{k+1})^2}{\|d_{k+1}\|^2} \geq \sum_{k=1}^{\infty}\frac{\varepsilon_1^2}{k} = \infty \quad .$$

which contradicts **Lemma 1**. Therefore, we get this theorem.

Numerical Results

In this section, we will report the numerical performance of Algorithm (2.1). We test Algorithm (2.1) by solving the 15 benchmark problems from [1] and compare its numerical performance with that of the other similar method, which include the standard FR conjugate gradient method in [3]. All codes of the computer procedures are written in Fortran. The parameters are chosen as follows :

$$\varepsilon = 10^{-6} \quad , \quad \xi = 0.5 \quad , \quad \delta_1 = 0.001 \quad , \quad \delta_2 = 0.9$$

In Tables 1 and 2, we use the following denotations :

n : the dimension of the objective function.
NOI : the number of iterations.
NOF : the number of function evaluations.
FR : the standard FR conjugate gradient method in [3].
SFR : the new spectral FR method presented in this paper.
SDY : the new spectral DY method presented in this paper.
SCD : the new spectral CD method presented in this paper
F : If NOF or NOI exceeded 2000 then denote F.

Table (4.1) Comparison of the algorithms for $n = 100$

Test problems	FR		SFR		SDY		SCD	
	NOI	NOF	NOI	NOF	NOI	NOF	NOI	NOF
Trigonometric	22	40	20	36	21	39	21	39
Extended Rosenbrock	49	98	53	104	56	113	58	115
Extended White & Holst	69	125	45	90	64	127	69	132
Penalty	47	1001	13	33	12	30	13	32
Generalized Tridiagonal 1	61	904	25	44	26	44	28	48
ExtendedThreeExpo Terms	34	313	14	21	22	34	19	30
Generalized Tridiagonal 2	105	194	44	69	46	73	52	75
Diagonal 4	33	51	17	32	24	48	24	48
Extended Himmelblau	16	31	15	28	13	24	15	28
Broyden Tridiagonal	45	72	31	51	35	56	35	53
DENSCHNA (CUTE)	12	23	25	41	23	36	23	35
DENSCHNF (CUTE)	22	39	23	40	19	33	22	38
Extended Block-Diagonal	28	44	13	24	23	46	23	46
Generalized quartic GQ1	10	22	9	20	9	20	9	20
Generalized quartic GQ2	50	81	39	69	48	76	54	86
Total	603	3038	386	702	441	799	465	825

Table (4.2) Comparison of the algorithms for $n = 1000$

Test problems	FR		SFR		SDY		SCD	
	NOI	NOF	NOI	NOF	NOI	NOF	NOI	NOF
Trigonometric	42	67	33	58	32	57	41	70
Extended Rosenbrock			60	113	65	126	61	122
Extended White & Holst	237	299	58	115	71	136	47	92
Penalty	26	209	16	43	12	30	13	37
Generalized Tridiagonal 1	43	390	30	57	58	993	39	376
ExtendedThreeExpo Terms			12	21	38	583	49	809
Generalized Tridiagonal 2	121	231	72	119	78	126	51	79
Diagonal 4	32	50	17	32	27	54	27	54
Extended Himmelblau	21	37	16	30	14	26	12	23
Broyden Tridiagonal	47	78	42	69	41	68	40	67
DENSCHNA (CUTE)	58	110	18	30	19	31	17	29
DENSCHNF (CUTE)	24	41	22	39	18	31	19	34
Extended Block-Diagonal	31	50	25	48	23	46	23	46
Generalized quartic GQ1	8	20	8	20	8	20	8	20
Generalized quartic GQ2	52	83	39	63	46	75	50	82
Total	742	1665	396	731	447	1247	387	1009

From the above numerical experiments, it is shown that the new algorithms in this paper is promising.

Conclusions and Discussions

In this paper, a new spectral conjugate gradient algorithm has been developed for solving unconstrained minimization problems. Under some mild conditions, the global convergence has been proved. Compared with the other similar algorithm, the numerical performance of the developed algorithm is promising.

Table (5.1) gives a comparison between the new-algorithm (2.1) and the Fletcher and Reeves (FR)-algorithm for convex optimization , this table indicates that the new algorithm (2.1) saves $(58-66)\%$ NOI and $(30-43)\%$

NOF, overall against the standard Fletcher and Reeves (FR)-algorithm, especially for our selected test problems.
Relative Efficiency of the Different Methods Discussed in the Paper.

Tools	NOI	NOF
FR- algorithm	100 %	100 %
SFR- algorithm	41.85 %	69.53 %
SDY- algorithm	33.97 %	43.50 %
SCD- algorithm	36.65 %	61.00 %

References

[1]. Andrie, N. (2008) ' An Unconstrained Optimization Test functions collection' Advanced Modeling and optimization.10, pp.147-161.
[2]. Dai, Y. and Yuan Y. (1999) ' A nonlinear conjugate gradients method with a strong global convergence property' SIAM J. Optimization 10, pp.177-182.
[3]. Fletcher, R. (1987) ' Practical Methods of Optimization (second edition). John Wiley and Sons, New York.
[4]. Fletcher, R. and Reeves C. (1964) ' Function minimization by conjugate gradients ' Computer J. 7, pp.149-154.
[5]. Hideaki I., Yasushi N., (2011). Conjugate gradient methods using value of objective function for unconstrained optimization Optimization Letters, V.6, Issue 5, 941-955.
[6]. Matonoha C., Luksan L. and Vlcek J. (2008) ' Computational experience with conjugate gradient methods for unconstrained optimization ' Technical report no.1038, pp.1-17.
[7]. Yabe H. and M. Takano, (2004). Global convergence properties of nonlinear conjugate gradient methods with modified secant condition, Comput. Optim. Appl., 28, 203–225.
[8]. Zhang Y. and Wang K. , (2012) ' A new general form of conjugate gradient methods with guaranteed descent and strong global convergence properties' , Numer Algor , 60, pp.135–152.
[9]. Zoutendijk, G. : Nonlinear programming. In: Abadie, J. (ed.) Integer and Nonlinear Programming, North-Holland, Amsterdam, 123, pp. 37–86.