Journal of Multidisciplinary Modeling and Optimization 1(1) (2018), 41-51.

The New Algorithm Form of the Fletcher – Reeves Conjugate Gradient Algorithm

Basim Abbas HASSAN[1], and Hameed Mohammed Sadeq[2]

[1] Department of Mathematics,
Mosul University, Mosul, Iraq
Basimabas39@gmail.com

[2]Apartment of Nineveh
Ministry of Education of Iraq, Mosul, Iraq
hameedsadiq1975@gmail.com

Abstract —*In this paper, we propose a new spectral form of the Fletcher – Reeves conjugate gradient algorithm for solving unconstrained optimization problems which has sufficient descent direction. We prove the global convergent of these algorithms under Wolf line search conditions. We presented some numerical result and comparison with Fletcher – Reeves algorithm.*

Keywords: Conjugate gradient methods, Descent property, Global convergence.
Mathematics Subject Classification: 65K10, 80M50, 90C26.

1 Introduction

Let $f : R^n \to R$ be continuously differentiable. Consider the unconstrained optimization problem

$$\min f(x)\ ,\ \mathrm{x} \in \mathrm{R}^{\mathrm{n}} \tag{1}$$

We use $g(x)$ to denote the gradient of f at x. We are concerned with the conjugate gradient methods for solving (1). Let $x_\circ$ be the initial guess of the solution of problem (1). A conjugate gradient method generates sequence of iterates by letting

$$x_{k+1} = x_k + \alpha_k d_k, \quad \mathrm{k} = 0,1,..., \tag{2}$$

Where the step length α_k is obtained by carrying out some line search, and the direction d_k is defined by

$$d_{k+1} = \begin{cases} -g_{k+1} & \text{if } \mathrm{k} = 0 \\ -g_{k+1} + \beta_k d_k & \text{if } \mathrm{k} > 0 \end{cases} \tag{3}$$

where β_k is a scalar. If $f(x)$ is a strictly convex quadratic function and if α_k is the exact one-dimensional minimizer, $(2)-(3)$ is called the linear conjugate gradient method. On the other hand, $(2)-(3)$ is called the nonlinear conjugate gradient method for general unconstrained optimization. Some well-known formulae for β_k are as follows.

$$\beta_{k+1}^{CD}=\frac{\|g_{k+1}\|^2}{d_k^T g_k}, \tag{4}$$

$$\beta_{k+1}^{PRP}=\frac{g_{k+1}^T y_k}{\|g_k\|^2}, \tag{5}$$

$$\beta_{k+1}^{HS}=\frac{g_{k+1}^T y_k}{d_k^T y_k}, \tag{6}$$

Here, (CD) denotes the Conjugate Descent [1], (PR) denotes the Polak and Ribiere [2], (HS) denotes the Hestenes and Stiefel [3]. The Fletcher-Reeves(FR) method [4] is famous conjugate gradient method. In the FR method, the parameter β_k is specified by

$$\beta_{k+1}^{FR}=\frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \tag{7}$$

where g_k is the abbreviation of $g(x_k)$ and $\|\cdot\|$ stands for Euclidean norm of vectors.

We see from (3) that for each $k\geq 1$, the directional derivative of f at x_k along direction d_k is given by

$$g_{k+1}^T d_{k+1}=-\|g_{k+1}\|^2+\beta_k^{FR} g_{k+1}^T d_k. \tag{8}$$

It is clear that if exact line search is used, then we have for any $k\geq 0$,

$$g_k^T d_k=-\|g_k\|^2<0. \tag{9}$$

Consequently, vector d_k is a descent direction of f at x_k. Zoutendijk [5] proved that the FR method with exact line search is globally convergent. Another line search that ensures descent property of d_k is the Wolf line search, that is, α_k satisfies the following two inequalities

$$f(x_k+\alpha_k d_k)-f(x_k)\leq \delta\alpha_k g_k^T d_k, \tag{10}$$

and

$$g_{k+1}^T d_k \geq \sigma\, g_k^T d_k, \tag{11}$$

where $0<\delta<\sigma<\frac{1}{2}$. Al-Baali [6] showed the global of FR method with the strong line search, Dai and Yuan [7] extended this result to $\sigma=\frac{1}{2}$. Recently, Birgin and Matinez [8] proposed a spectral conjugate gradient method by combining conjugate gradient method and spectral method [9] in the following way:

$$d_{k+1} = -\theta_k g_{k+1} + \beta_k d_k, \tag{12}$$

where θ_k is parameter, Li zhang, Weijun Zhou and Donghui Li [10].

The paper is organized as follows. Section (1) is the introduction. In Section (2) new spectral form for FR non-linear conjugate gradient algorithm is suggested. The sufficient descent condition are presented in section (3). In section (4) global convergence of new spectral conjugate gradient methods. Numerical results are reported in section (5).

2 New spectral Conjugate Gradient method

In this section, we describe the modified FR method whose form is similar to that of [8] but with different parameters θ_k and β_k. An important feature of the CG method is that satisfies conjugacy condition:

$$d_{k+1}^T y_k = 0 \tag{13}$$

which is independent of the objective function $f(x)$ is convex quadratic and line search is exact [11]. Let the search direction be defined by

$$d_{k+1} = -\theta_{k+1} g_{k+1} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} d_k \tag{14}$$

where θ_k is parameter. We also assume that the search direction (14) satisfies the relation (9) i.e.

$$d_{k+1}^T g_{k+1} = -\theta_{k+1} g_{k+1}^T g_{k+1} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} d_k^T g_{k+1} < 0 \tag{15}$$

$$-\theta_{k+1} + \frac{d_k^T g_{k+1}}{\|g_k\|^2} < 0 \tag{16}$$

or

$$\theta_{k+1} = \lambda + \frac{d_k^T g_{k+1}}{\|g_k\|^2}, \ \lambda > 0 \tag{17}$$

then we have

$$\theta_{k+1} = \frac{\lambda \|g_k\|^2 + d_k^T g_{k+1}}{\|g_k\|^2} \tag{18}$$

Now by using conjugacy condition (14), we find value of λ then

$$y_k^T d_{k+1} = -\theta_{k+1} y_k^T g_{k+1} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} y_k^T d_k = 0 \tag{19}$$

Put the value of θ_k in equation (19), we can then obtain

$$\lambda\, y_k^T g_{k+1} = -\frac{d_k^T g_{k+1}}{\|g_k\|^2} y_k^T g_{k+1} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} y_k^T d_k \tag{20}$$

From (20), we have

$$\lambda = \frac{\|g_{k+1}\|^2 y_k^T d_k - d_k^T g_{k+1}(y_k^T g_{k+1})}{\|g_k\|^2 (y_k^T g_{k+1})} \tag{21}$$

Substituting (21) into (17), we get

$$\theta_{k+1}^{BHS} = \frac{\|g_{k+1}\|^2 (y_k^T d_k)}{\|g_k\|^2 (y_k^T g_{k+1})} \tag{22}$$

Therefore the new spectral FR search direction is

$$d_{k+1} = -\theta_{k+1}^{BHS} g_{k+1} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} d_k \tag{23}$$

3 The Descent Property and Descent Algorithm

In this section, we give a general condition on the spectral θ_k and show that such a condition can ensure the descent property of the conjugate gradient method in the case of Wolfe line searches. Furthermore, the sufficient descent condition, namely

$$g_{k+1}^T d_{k+1} \le -c\|g_{k+1}\|^2 \quad \text{for} \quad k \ge 0 \text{ and } c > 0 \tag{24}$$

Theorem 3.1. Consider the algorithm defined in (2) where d_k computed from (22) and (23). Assume that step size α satisfies the Wolf Condition (10) and (11). Then the search directions d_k generated by the new method algorithm are descent for all k provided $y_k^T g_{k+1} > 0$.

$$d_{k+1}^T g_{k+1} = -\lambda\|g_{k+1}\|^2 \tag{25}$$

Proof :

For initial direction (k=0) we have:

$$d_1 = -g_1 \Rightarrow d_1^T g_1 = -\|g_1\|^2 < 0 \tag{26}$$

Now let the theorem be true for all k, i.e.

$$d_k = -g_k \Rightarrow d_k^T g_k = -\|g_k\|^2 < 0 \tag{27}$$

To complete the proof, we have to show that the theorem is true for all k+1.

Multiplying (12) by g_{k+1}^T, we have

$$d_{k+1}^T g_{k+1} = -\theta_k^{BHS} \|g_{k+1}\|^2 + \beta_k^{FR} d_k^T g_{k+1}$$

$$= -\theta_k^{BHS} \|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} d_k^T g_{k+1} \tag{28}$$

Substituting (18) in to (28), we obtain:

$$d_{k+1}^T g_{k+1} = -\left(\frac{\lambda\|g_k\|^2 + d_k^T g_{k+1}}{\|g_k\|^2}\right)\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} d_k^T g_{k+1} \tag{29}$$

Then, we have

$$d_{k+1}^T g_{k+1} = -\lambda \|g_{k+1}\|^2 \tag{30}$$

For convenience, we summarize the above method as the following algorithm:

Algorithm(The new method)

Step 0: Given an initial starting point $x_\circ \in R^n$ and $\varepsilon = 10^{-6}$, consider $d_\circ = -g_\circ, \alpha_\circ = \frac{1}{\|g_\circ\|}$

and $k = 0$.

Step 1: Test for convergence, If $\|g_k\| < \varepsilon$ stop x_k is optimal Else go to step 2.

Step 2: Compute α_k satisfying the Wolfe line search and update the variable

$x_{k+1} = x_k + \alpha_k d_k$ and compute f_{k+1} , g_{k+1} , y_k and s_k.

Step 3: Direction computation: compute θ_{k+1} from (22) and set

$d_{k+1} = -\theta_{k+1} g_{k+1} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} d_k$. If Powell restart is satisfied then $d_{k+1} = -\theta_{k+1} g_{k+1}$ Else

$d_{k+1} = d$, compute initial guess for $\alpha_{k+1} = \alpha_k \left(\frac{\|d_k\|}{\|d_{k-1}\|}\right)$ and set $k = k+1$ go to step 1.

4 Global convergence property

In order to establish the global convergence of the proposed method. We assume that the following assumption always holds.

Assumption(1):

i- The level set $S = \{x \in R^n : f(x) \leq f(x_\circ)\}$ is bounded, namely, there exists a constant $B > 0$ such that

$$\|x\| \leq B \quad \text{for all} \quad x \in S \tag{31}$$

ii- In some neighbourhood N of S, f is continuously differentiable, and its gradient is Lipschitz continuous, namely, there exist $L > 0$ such that:

$$\|g(x) - g(y)\| \leq L\|x - y\| \quad \forall \mathrm{x}, \mathrm{y} \in \mathrm{N}. \tag{32}$$

Lemma 1. Suppose assumption(1) holds. Consider any iteration of (2) and (12), where d_k satisfies $g_k^T d_k < 0$ for $k \in N^+$ and α_k satisfies the Wolf line search. Then

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \ . \tag{33}$$

More details can be found in [12].

Now, we give the following Theorem of global convergence for the spectral FR conjugate gradient method.

Theorem 4.1. Consider the spectral FR conjugate gradient Algorithm, suppose that Assumptions hold. Then

$$\lim_{k \to \infty} \inf \|g_k\| = 0. \tag{34}$$

Proof :

Suppose by contradiction that there exists a real number $\varepsilon > 0$ such that $\|g_{k+1}\| > \varepsilon$ for all $k = 1,2,3,\ldots$.

Squaring the both terms of $d_{k+1} + \theta_k^{BHS} g_{k+1} = \beta_k d_k$ we get:

$$\|d_{k+1}\|^2 + 2\theta_k^{BHS} d_{k+1}^T g_{k+1} + (\theta_k^{BHS})^2 \|g_{k+1}\|^2 = \beta_k^2 \|d_k\|^2. \tag{35}$$

From(35), we have:

$$\|d_{k+1}\|^2 = \beta_k^2 \|d_k\|^2 - 2\theta_k^{BHS} d_{k+1}^T g_{k+1} - (\theta_k^{BHS})^2 \|g_{k+1}\|^2 \tag{36}$$

Dividing both sides of (36) by $\|g_{k+1}\|^4$, by (7), (33) and $\|g_{k+1}\| > \varepsilon$ we get:

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} = \frac{(\beta_k^{FR})^2\|d_k\|^2 - 2\theta_k^{BHS} d_{k+1}^T g_{k+1} - (\theta_k^{BHS})^2\|g_{k+1}\|^2}{\|g_{k+1}\|^4}$$

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} = \left(\frac{\|g_{k+1}\|^2}{\|g_k\|^2}\right)^2 \frac{\|d_k\|^2}{\|g_{k+1}\|^4} - \frac{2\theta_k^{BHS} d_{k+1}^T g_{k+1}}{\|g_{k+1}\|^4} - \frac{(\theta_k^{BHS})^2\|g_{k+1}\|^2}{\|g_{k+1}\|^4}$$

$$= \frac{\|\mathrm{d}_k\|^2}{\|g_k\|^4} + \frac{2\theta_k^{BHS}}{\|g_{k+1}\|^2} - \frac{(\theta_k^{BHS})^2}{\|g_{k+1}\|^2} \tag{37}$$

$$= \frac{\|\mathrm{d}_k\|^2}{\|g_k\|^4} - \frac{1}{\|g_{k+1}\|^2}\left((\theta_k^{BHS})^2 - 2\theta_k^{BHS} + 1 - 1\right)$$

$$= \frac{\|\mathrm{d}_k\|^2}{\|g_k\|^4} - \frac{(\theta_k^{BHS} - 1)^2}{\|g_{k+1}\|^2} + \frac{1}{\|g_{k+1}\|^2}$$

$$\le \frac{\|\mathrm{d}_k\|^2}{\|g_k\|^2} + \frac{1}{\|g_{k+1}\|^2} \le \sum_{i=0}^{k} \frac{1}{\|g_{i+1}\|} \le \frac{k}{\varepsilon^2} \; .$$

(38)

From (30) and (33) we get:

$$\lambda^2 \sum_{k\ge 0} \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \le \sum \frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} < \infty \tag{39}$$

Thus,

$$\sum_{k\ge 0} \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} < \infty \tag{40}$$

From (38), we know

$$\sum_{k\ge 0} \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \ge \varepsilon^2 \sum_{k\ge 1} \frac{1}{k} = \infty \tag{41}$$

which contradicts [40], so $\lim_{k\to\infty}\|g_k\| = 0$.

5 Numerical Results

In this section, we compare the performance of the new algorithm with one CG-algorithm, the Fletcher-Reeves (FR) algorithm which is one of the best and well-know CG-algorithms. The codes were written in fortran language with f77 and double precision. Our experiment are performed on a set of (15) nonlinear unconstrained test problems. We have considered numerical experiment with the number of variable $n = 100 - 1000$. We use $\delta = 0.001$ and $\sigma = 0.9$ in line search, we stop the iteration if the inequality $\|g_{k+1}\| < 10^{-6}$ is satisfied. We record the number of iteration (NOI), and the number of function evaluation (NOF), and the number of restart (NOR). Comparing the new method with Fletcher – Revees method, we find that for some problems new method really performs much better than Fletcher method. Table (1) and (2) shows the details of numerical results for Fletcher – Revees (FR) and our algorithm.

Table 1: Comparison of the algorithms for $n = 100$

Test Problems	*FR* algorithm			*New* algorithm with θ_k^{BHS}		
	NOI	NOR	NOF	NOI	NOR	NOF
DenschnF	22	19	38	19	17	33
Extended Rosenbrock	47	18	93	42	22	86
Nondia	13	7	25	11	6	22
Extended Tridiagonal 2	40	18	65	41	14	64
Liarwhd	23	11	45	17	9	33
Extended While & Holst	43	18	88	36	20	76
Extended Quadratic Penalty QP2	32	12	65	28	15	60
Arwhead	9	4	18	9	6	36
DenschnB	12	7	25	7	5	15
Generalized Tridiagonal 2	37	8	67	37	11	59
Generalized Quadratic GQ	11	6	24	7	5	16
Extended PSC 1	15	9	31	8	6	17
Partial Perturbed Quadratic	74	21	123	68	26	109
Sincos	15	9	31	8	6	17
Engval 1	34	16	57	28	10	49
	427	183	795	366	178	692

Table 2: Comparison of the algorithms for $n = 1000$

Test Problems	*FR* algorithm			*New* algorithm with θ_k^{BHS}		
	NOI	NOR	NOF	NOI	NOR	NOF
DenschnF	22	21	37	19	17	33
Extended Rosenbrock	78	45	131	42	21	96
Nondia	15	7	29	12	7	25
Extended Tridiagonal 2	43	23	68	41	20	67
Liarwhd	27	11	55	19	10	42
Extended While & Holst	46	19	92	37	19	80
Extended Quadratic Penalty QP2	53	22	116	36	20	87
Arwhead	12	7	82	8	6	56
DenschnB	11	7	23	7	5	15
Generalized Tridiagonal 2	73	27	115	67	25	102
Generalized Quadratic GQ	9	5	22	7	5	18
Extended PSC 1	8	6	17	7	5	15
Partial Perturbed Quadratic	370	88	616	249	66	407
Sincos	8	6	17	7	5	15
Engval 1	142	126	3616	132	118	3565
	917	420	5036	690	349	4623

6 Conclusions and Discussions

In this paper, we have derived a new spectral conjugate gradient method for solving unconstrained minimization problems. It is shown in previous section that the new spectral CG method converges under some assumption using Wolf line search condition and satisfies the sufficient descent property. The computational experiments show that the new algorithm given in this paper are successful.

Table(3) gives the new algorithm saves(15 - 25)% NOI ,(3 - 17)% NOF, and (9 - 12)% IRS, overall against the standard FR algorithm, especially for our selected test functions. These results are shown in following table:

.

Table 3: Relative efficiency of the new Algorithm $(n=100)$

Tools	NOI	NOF	IRS
FR Algorithm	100 %	100 %	100 %
New Algorithm with θ_k^{BHS}	85.71%	97.26%	88.04%

Table 4: Relative efficiency of the new Algorithm $(n=1000)$

Tools	NOI	NOF	IRS
FR Algorithm	100 %	100 %	100 %
New Algorithm with θ_k^{BHS}	75.24%	83.09%	91.79%

References

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

[2] E. Polak and G. Ribiere, Note for Convergence Direction Conjugate, Revue Francaise Informant, Reserche Opertionelle, 3 1969, 35-43.

[3] M. R. Hestenes and E. L. Stiefel, Method of conjugate gradients for Solving linear systems, Journal National Standards, 49 1952, 409-436.

[4] R. Fletcher and C. Reeves, Function minimization by conjugate gradients, Comput. J., 7 1964, 149-154.

[5] G. Zoutendijk, Nonlinear programming computational algorithms, In “Integer and Nonlinear programming”, North-Holland, Amsterdam, 1970, 37-86.

[6] M. Al-Baali, Descent property and global convergence of Flecher-Reeves with inexact line search, IMA J. Anal., 5 1985, 121-124.

[7] Y. Dai and Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, SIAM J. Optimization., 10 1999, 177-182.

[8] E. Birgin and J. M. Martinez, A spectral conjugate gradient method for unconstrained optimization, App. Math. Optim., 43 2001, 117-128.

[9] M. Raydan, The Barzilain and Borwein gradient method for the large unconstrained minimization problem, SIAM J. Optim., 7 1997, 26-33.

[10] Z. Li, Z. Weijun and L. Donghui, Global convergence of a modified Fletcher Reeves conjugate gradient method with Armijo-type line, 2006, 561-572.

[11] Y. Dai and Z. Liao, New conjugate conditions and related nonlinear conjugate gradient methods, Appl. Math. Optim., 43 2001, 87-101.

[12] Z. Li, New versions of the Hestenes-Stiefel nonlinear conjugate gradient method based on the secant condition for optimization, Comput. Appl. Math., 28 2009, 111-133.