References :

[1]. Andrei N. (2008), "An Unconstrained Optimization test function collection." Adv. Model. Optimization .No.10. pp.147-161.

[2]. Barzilai, J. and Borwein, J., (1988), Two Point Step-Size Gradient Method. IMAJ. Numerical. Analysis, Vol.(8), pp. 141-148.

[3]. Basim A. H. and Hameed M.S. ,(2014), "A Modified class of conjugate gradient algorithms based on quadratic model for nonlinear unconstrained optimization." Raf. J. of Comp. and Math.,No.1, pp. 25-37.

[4]. Birgin, F. G. and Martinez, J. M. (2001), "A spectral Conjugate Gradient Method for Unconstrained Optimization", Applied Mathematics and Optimization, No.43,pp. 117-128.

[5]. Dai, Y. and Yuan Y. (1999), "A nonlinear conjugate gradients method with a strong global convergence property" SIAM J. Optim. 10, pp. 177-182

[6]. Fletcher, R. (1987) "Practical Methods of Optimization (second edition)". John Wiley and Sons, New York.

[7]. Fletcher, R. and Reeves C. (1964) "Function minimization by conjugate gradients" Computer J. 7, pp. 149-154.

[8]. Hestenes, M. R. and Stiefel E. L. (1952) "Method of conjugate gradients for solving linear systems" J. Research Nat. Standards 49, pp. 409-436.

[9]. Liu Y. and Storey C. (1991) "Efficient generalized conjugate gradients algorithms" Part 1 : Theory. J. Opti. Theory and Appli. No.69, pp. 129-137.

[10]. Polak, E. and Ribiere, G. (1969) "Note for Convergence Direction Conjugate", Revue Francaise Informant, Reserche. Opertionelle, 16, 35-43.

[11]. Wei Z., Tang C. and Li G. (2007) " New conjugacy conditions and related new conjugate gradient methods for unconstraint optimization" J. of computation and Applied Mathematics, 202, pp.523-539.

[12]. Wei Z., Mo L. and Lu S. (2011) "Some global convergence properties of the Wei-Yao-Liu conjugate gradient with inexact line search "J. Applied Mathematics and computation 217, pp.7132-7137.

[13]. Kafaki S.B., Ghanbari R. and Amiri M.N., (2010), Two new conjugate gradient method based on modified secant equations, Journal of computational and applied Mathematics, 234, 1374-1386.

[14]. Sakaiwa, N., Takano, M. and Yabe, H. (2003), Global Convergence Properties of New Nonlinear Conjugate Gradient Methods for Unconstrained Optimization, Department.

Denschnc	59	41	39	28	40	29
Extended Block Diagonal	944	282	914	281	871	267
Generalized Quad. GQ1	1252	367	1189	357	1192	341
Sincos	870	257	918	279	855	269
Liarwhd (CUTE)	96	52	94	54	93	52
Generalized Quad. GQ2	753	236	771	246	817	274
Total	**9633**	**4006**	**9119**	**3669**	**9112**	**3663**

6. Conclusios :

From the numerical results of the above tables, we say that the results of Table **(5.1)** and Table **(5.2)** give a general comparison between DY and two modified spectral CG-methods taking non linear test function with n=100, 200, … , 1000. This table indicates that the modified methods save (6-8)% NOI and (8-12)% IRS. The percentage performance of the improvements of the Table **(5.1)** and Table **(5.2)** are given by the following table **(5.3)** .

Table (5.3) : Relative efficiency of the new Algorithm

Tools	NOI	IRS
Dai-Yuan method	100%	100%
New Algorithm with θ_k^{OS1}	94.66%	92.11%
New Algorithm with θ_k^{OS2}	92.11%	88.02%

Table (5.1) : Comparison of methods for n= 100, 300, 500, 700, 900

Test Problems	β_k^{DY}		θ_k^{OS1}		θ_k^{OS2}	
	NOI	**IRS**	**NOI**	**IRS**	**NOI**	**IRS**
Extended Rosenbrock	188	160	111	82	94	66
Extended While & Holst	1176	546	1200	528	1154	524
Extended PSC 1	970	860	1079	972	754	645
Extended Maratos	275	248	115	85	125	96
Quadratic QF2	255	83	247	90	252	99
Arwhead	409	132	411	128	393	129
Nondia	602	119	550	113	556	109
Partial Perturbed Quad	404	257	399	254	389	247
Liarwhd	1211	362	1065	319	1073	308
Denschnc	50	37	54	38	51	35
Extended Block Diagonal	851	273	807	256	807	244
Generalized Quad. GQ1	1150	348	1070	307	1096	325
Sincos	893	270	832	247	775	233
Liarwhd (CUTE)	98	55	97	54	92	52
Generalized Quad. GQ2	703	233	666	217	741	257
Total	**9235**	**3983**	**8743**	**3690**	**8352**	**3369**

Table (5.2) : Comparison of methods for n= 200, 400, 600, 800, 1000

Test problems	β_k^{DY}		θ_k^{OS1}		θ_k^{OS2}	
	NOI	**IRS**	**NOI**	**IRS**	**NOI**	**IRS**
Extended Rosenbrock	100	72	135	104	88	54
Extended While & Holst	1314	600	1205	550	1160	518
Extended PSC 1	1108	993	750	631	864	747
Extended Maratos	157	131	231	204	168	136
Quadratic QF2	273	113	266	104	270	136
Arwhead	430	135	378	126	389	122
Nondia	600	110	602	126	646	123
Partial Perturbed Quad	403	254	398	252	391	240
Liarwhd	1274	363	1229	357	1268	355

considered 10 numerical experiments with number of variables $n = 100, 200, \dots, 1000$. We use $\delta_1 = 10^{-4}$ and $\delta_2 = 0.9$ in the line search routine (3)-(4). All these methods terminate when the following stopping criterion is met $\|g_{k+1}\| \le 10^{-6}$.

All codes are written in double precision FORTRAN Language with F90 default complier settings. We record the number of iterations calls (NOI), and the number of restart calls (IRS) for the purpose our comparisons.

Lemma 4.1

Suppose that Assumptions 1 and 2 hold. Consider any conjugate gradient method in the form (2)-(14), where d_{k+1} is a descent direction and α_k is computed using the strong Wolfe line search conditions. If

$$\sum_{k\geq 0} \frac{1}{\|d_{k+1}\|^2} = \infty \, , \qquad \ldots(38)$$

Then we have

$$\lim_{k\to\infty} \inf\|g_k\| = 0 \qquad \ldots(39)$$

Theorem 4.2

Suppose that Assumptions 1 and 2 and the descent condition hold. Consider a conjugate gradient method in the form (2)-(14) with θ_k^{OS1} and θ_k^{DY} as in (6), where α_k is computed from the standard Wolfe line search conditions (3)-(4). Suppose that there exists the positive constant c_1 such that $\theta_k \leq c_1$ for all $k \geq 1$. If the objective function is uniformly convex on S, then $\lim_{k\to\infty}\|g_k\| = 0$.

Proof :

Now, from (14) with θ_k^{OS1} it follows that f is uniformly convex function in S and therefore $y_k^T v_k \geq \mu\|v_k\|^2$. Because the descent condition hold, we have $d_{k+1} \neq 0$. Also, from Assumptions 1 and 2, Proposition 4.1, Lemma 4.2, we have

$$\|d_{k+1}\| = \left\|-\theta_k^{OS1} g_{k+1} + \beta_k^{DY} v_k\right\|$$

$$\leq \left|\theta_k^{OS1}\right|\|g_{k+1}\| = |\beta_k|\|v_k\|$$

$$\leq c_1\|g_{k+1}\| + \left(\frac{\|g_{k+1}\|^2}{|y_k^T v_k|}\right)\|v_k\|$$

$\ldots(40)$

$$\leq c_1\bar{\gamma} + \frac{[\bar{\gamma}]^2}{\mu\|v_k\|^2}\|v_k\|$$

$$\|d_{k+1}\| \leq \left(c_1 + \frac{\bar{\gamma}}{B\mu}\right)\bar{\gamma}$$

This relation shows that

$$\sum_{k\geq 1} \frac{1}{\|d_{k+1}\|^2} \geq \left(\frac{B\mu}{(c_1 B\mu + \bar{\gamma})\bar{\gamma}}\right)\sum_{k\geq 1} 1 = 0$$

$\ldots(41)$

Therefore, from Lemma 4.1 we have $\lim_{k\to\infty} \inf\|g_k\| = 0$, which for uniformly convex function is equivalent to $\lim_{k\to\infty}\|g_k\| = 0$.

Remark : we use similarly technique to classical algorithm θ_k^{OS2}.

5. Numerical Results :

In this section, we reported some numerical results obtained with the implementation of the new methods on a set of unconstrained optimization test problems taken from (Andrei, 2008) [1].

We selected (15) large scale unconstrained optimization test problems. For each test function we have

$$= \frac{1}{4}\|g_{k+1}\|^2 + \frac{(v_k^T g_{k+1})^2\|g_{k+1}\|^2}{(y_k^T v_k)^2}$$

From (27) and (28) we have

$$g_{k+1}^T d_{k+1} \le -\theta_k^{OS1}\|g_{k+1}\|^2 + \frac{1}{4}\|g_{k+1}\|^2 + \frac{(v_k^T g_{k+1})^2\|g_{k+1}\|^2}{(y_k^T v_k)^2} \quad \ldots(30)$$

Since $\frac{(v_k^T g_{k+1})^2\|g_{k+1}\|^2}{(y_k^T v_k)^2}$ approximation zero we get

$$g_{k+1}^T d_{k+1} \le -\left[\theta_k^{OS1} - \frac{1}{4}\right]\|g_{k+1}\|^2 \ldots(31)$$

Remark : we use similarly technique to classical algorithm θ_k^{OS2}.

4. Convergence analysis

In this section we analyze the convergence of the algorithm (2) and (15), where θ_k and β_k are given by (13) and (17) respectively. In the following we consider that

$$g_{k+1} \neq 0, \quad \forall\, k \ge 1 \quad \ldots(32)$$

Otherwise, a stationary point is at hand. We make the following basic assumptions on the objective function.

Definition 4.1

A twice continuously differentiable function f is said to be uniformly convex on the nonempty open convex set S if and only if there exists $M > 0$ such that

$$(g(x) - g(y))^T.(x-y) \ge M\|x-y\|^2,\ x, y \in S \quad \ldots(33)$$

or, equivalently, there exists $r > 0$ such that

$$z^T\nabla^2 f(x) z \ge r\|z\|^2, \quad \forall\, x \in S, \forall\, z \in R^n \quad \ldots(34)$$

See [13].

Assumption 1

The level set $l = \{x: f(x) \le f(x_1)\}$ is bounded ; that is, there exists a constant $B > 0$ such that

$$\|v\| \le B\,,\ \forall\, x \in l \quad \ldots(35)$$

Assumption 2

In some neighborhood N of l $(l \subseteq N)$, f is continuously differentiable, and its gradient is Lipschitz continuous , that is , there exists a constant $L > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x-y\|, \quad \forall x, y \in N \quad \ldots(36)$$

The following proposition is now immediate [13-14]

Proposition 4.1

Under Assumptions 1 and 2 on f, there exists a constant $\bar{\gamma} > 0$ such that

$$\|\nabla f(x)\| \le \bar{\gamma}\,,\quad \forall\, x \in l \quad \ldots(37)$$

Now we can obtain the new conjugate gradient algorithms .

New algorithms :

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 $\alpha_1 = 1/\|g_1\|$.
2. Test for continuation of iterations. If $\|g_{k+1}\| < 10^{-6}$, then stop.
3. Line search. Compute $\alpha_{k+1} > 0$ satisfying the Wolfe line search condition (3) and (4) and update the variables $x_{k+1} = \alpha_k + d_k$.
4. β_k conjugate gradient parameter which defined (6)
5. θ_k is computed as in (19) and (22) where $\theta_k > 1/4$
6. Direction computation. Compute $d_{k+1} = -\theta_k g_{k+1} + \beta_k d_k$. If the restart criterion of Powell $|g_{k+1}^T g_k| \geq 0.2\|g_{k+1}\|^2$, is satisfied, then $d_{k+1} = -\theta_k g_{k+1}$, set $k = k+1$ and continue with step 2 .

3. The sufficient descent condition

In this section we shall introduce the new theorem which is ensure the sufficient descent condition of the new methods (14) with (19) and (23)

Theorem (3.1)

If $\theta_k^{OS1} > 1/4$, then the direction $d_{k+1} = -\theta_k^{OS1} g_{k+1} + \beta_k^{DY} v_k$ satisfies the sufficient descent direction

$$g_{k+1}^T d_{k+1} \leq -\left[\theta_k^{OS1} - \frac{1}{4}\right] \|g_{k+1}\|^2 \text{...(26)}$$

Proof.

Since $d_0 = -g_0$, we have $g_0^T d_0 \leq -\|g_0\|^2 < 0$. Assume by induction that

$$g_k^T d_k \leq -c\|g_k\|^2 < 0 \quad \text{where } 0 < c < 1$$

…(27)

which is a sufficient descent direction. To complete the proof, we have to show that the theorem is true for all $k+1$. Multiplying (14) by g_{k+1}^T, we have

$$g_{k+1}^T d_{k+1} = -\theta_k^{OS1}\|g_{k+1}\|^2 + \beta_k^{DY} g_{k+1}^T v_k$$

$$= -\theta_k^{OS1}\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{y_k^T v_k} g_{k+1}^T v_k$$

…(28)

$$\frac{g_{k+1}^T g_{k+1}(v_k^T g_{k+1})}{y_k^T v_k} = \frac{[(y_k^T v_k) g_{k+1}/\sqrt{2}][\sqrt{2}(v_k^T g_{k+1}) g_{k+1}]}{(y_k^T v_k)^2}$$

$$\leq \frac{\frac{1}{2}\left[\frac{1}{2}(y_k^T v_k)^2\|g_{k+1}\|^2 + 2(v_k^T g_{k+1})^2\|g_{k+1}\|^2\right]}{(y_k^T v_k)^2}$$

…(29)

Multiply (16) by y_k , and after some algebra we get :

$$-\left[\frac{2d_k^T d_k(f_k-f_{k+1})}{(g_k^T d_k)^2}\right] g_{k+1}^T y_k = -\theta_k g_{k+1}^T y_k + \beta_k^{DY} d_k^T y_k \quad \dots(17)$$

$$\theta_k g_{k+1}^T y_k = \left[\frac{2d_k^T d_k(f_k-f_{k+1})}{(g_k^T d_k)^2}\right] g_{k+1}^T y_k + \|g_{k+1}\|^2 \quad \dots(18)$$

$$\theta_k^{OS1} = \left[\frac{2d_k^T d_k(f_k-f_{k+1})}{(g_k^T d_k)^2}\right] + \frac{\|g_{k+1}\|^2}{g_{k+1}^T y_k} \quad \dots(19)$$

2- The second procedure is based on the conjugacy condition.

Substituting (15) into (12), we obtain :

$$d_{k+1}^T y_k = -\left[\frac{2d_k^T d_k(f_k-f_{k+1})}{(g_k^T d_k)^2}\right] g_{k+1} y_k = -g_{k+1}^T v_k \quad \dots(20)$$

On the other hand, if we use the inexact line search with scale CG-method which defined in (14) we get :

$$d_{k+1}^T y_k = -\theta_k g_{k+1}^T y_k + \beta_k^{DY} d_k^T y_k = -t g_{k+1}^T v_k \quad \dots(21)$$

Where t_1 is unknown parameter from equation (20)-(21) we have :

$$-\theta_k g_{k+1}^T y_k + \beta_k^{DY} d_k^T y_k = -t g_{k+1}^T v_k == -\left[\frac{2d_k^T d_k(f_k-f_{k+1})}{(g_k^T d_k)^2}\right] g_{k+1} y_k + g_{k+1}^T v_k \quad ...(22)$$

Put the value of β_k^{DY} in equation (22)

$$\theta_k^{OS2} = \frac{\|g_{k+1}\|^2}{g_{k+1}^T y_k} + (t-1)\frac{g_{k+1}^T v_k}{g_{k+1}^T y_k} + \frac{2d_k^T d_k(f_k-f_{k+1})}{(g_k^T d_k)^2} \quad \dots(23)$$

To find the value of t in (23) we solve it for Newton direction i.e. $-G^{-1} g_{k+1} = -\theta_k g_{k+1} + \beta_k^{DY} d_k$

$$-G^{-1} g_{k+1} = -\left[\frac{\|g_{k+1}\|^2}{g_{k+1}^T y_k} + (t-1)\frac{g_{k+1}^T v_k}{g_{k+1}^T y_k} + \frac{2d_k^T d_k(f_k-f_{k+1})}{(g_k^T d_k)^2}\right] g_{k+1} + \beta_k^{DY} d_k \quad \dots(24)$$

Multiply both sides by y_k and use $v_k = G^{-1} y_k$ we get :

$$-v_k^T g_{k+1} = -\left[\frac{\|g_{k+1}\|^2}{g_{k+1}^T y_k} + (t-1)\frac{g_{k+1}^T v_k}{g_{k+1}^T y_k} + \frac{2d_k^T d_k(f_k-f_{k+1})}{(g_k^T d_k)^2}\right] g_{k+1} y_k + \frac{\|g_{k+1}\|^2}{d_k^T y_k} d_k^T y_k$$

$$-v_k^T g_{k+1} = g_{k+1}^T v_k - t g_{k+1}^T v_k - \frac{2d_k^T d_k(f_k-f_{k+1})}{(g_k^T d_k)^2} g_{k+1} y_k$$

$$-2v_k^T g_{k+1} = -t g_{k+1}^T v_k - \frac{2d_k^T d_k(f_k-f_{k+1})}{(g_k^T d_k)^2} g_{k+1} y_k \quad \dots(25)$$

$$t v_k^T g_{k+1} = 2 g_{k+1}^T v_k - \frac{2d_k^T d_k(f_k-f_{k+1})}{(g_k^T d_k)^2} g_{k+1} y_k$$

$$t = 2 - \frac{2d_k^T d_k(f_k-f_{k+1})}{(g_k^T d_k)^2} \frac{g_{k+1} y_k}{g_{k+1}^T v_k}$$

As above, since $g_{k+1}^T v_k \to 0$ along the iterations, θ_k^{OS1} obtained from the Newton direction paradigm is very similar to θ_k^{OS2} based on the conjugacy condition.

$$H_{k+1}y_k = v_k \quad \ldots(10)$$

from (10) and the search direction d_{k+1} can be calculated in the form

$$d_{k+1} = -H_{k+1}g_{k+1} \quad \ldots(11)$$

we have

$$d_{k+1}^T y_k = -(H_{k+1}g_{k+1})^T y_k = -g_{k+1}^T H_{k+1} y_k = -g_{k+1}^T v_k \quad \ldots(12)$$

By introducing a scaling factor t, Dai and Liao considered a generalized conjugacy condition,

$$d_{k+1}^T y_k = -t g_{k+1}^T v_k\,, \quad t \geq 0\,, \quad \ldots(13)$$

where t is a parameter. In the case $t = 0$, then (13) becomes (9). In case $t = 1$, (13) reduced to (12). Furthermore, if exact line search is used, then $g_{k+1}^T v_k = 0$ holds for all k. It follows that both (12) and (13) coincide with (9) more details can be found in [11].

Another popular method to solving problem in (1) is the spectral gradient method, which was developed originally by Barzilai and Borwein in 1988 [2].

The direction d_{k+1} is given by the following way

$$d_{k+1} = -\theta_k g_{k+1} + \beta_k d_k \quad \ldots(14)$$

where θ_k is scalar parameter which follows to be determined. More details can be found in [4].

2. A Spectral Conjugate Gradient Methods

In this section, we describe the new spectral method whose form is similar to that of [4] but with different parameters and is based on the Quasi-Newton method.

In [3], we compute the Newton search direction by :

$$d_{k+1} = -\left[\frac{2d_k^T d_k(f_k - f_{k+1})}{(g_k^T d_k)^2}\right] g_{k+1} \quad \ldots(15)$$

for more details can be found in [3].

To determine the parameter θ_k in (14) we suggest the following two procedures :

1- When the initial point is near the solution of (1) and the Hessian of function f is nonsingular matrix we know that the Newton direction is the best line search direction. Therefore, to get a good algorithm for solving (1) this is very good motivation to choose the parameter θ_k in such a way that for every $k \geq 1$ the direction d_{k+1} given by (14) is the Newton direction. Therefore, from β_k^{DY} with the equation

$$-\nabla^2 f(x_{k+1})^{-1} g_{k+1} = -\theta_k g_{k+1} + \beta_k^{DY} d_k$$

$$-\left[\frac{2d_k^T d_k(f_k - f_{k+1})}{(g_k^T d_k)^2}\right] g_{k+1} = -\theta_k g_{k+1} + \beta_k^{DY} d_k \quad \ldots(16)$$

1- Introduction :

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

$$min\{f(x) \quad | \quad x \in R^n\} \qquad \dots (1)$$

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

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

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

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta_1 \alpha_k d_k^T g_k \quad .(3)$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \delta_2 d_k^T g_k \qquad \dots(4)$$

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

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

where β_k is a scalar, g_{k+1} denotes $g(x_{k+1})$. Different CG methods correspond to different choices for the scalar β_k. The well Known formulas for β_k such as β_k^{HS} (Hestenes, Stiefel [7]), β_k^{FR} (Fletcher, Reeves [6]), β_k^{PRP}(Polak, Ribiere and Polyak [9]), β_k^{CD} (Fletcher [5]), β_k^{LS}(Liu, Storey [8]), β_k^{DY}(Dai, Yuan [4]) can be found in many related literatures. In practical computations, the DY method with

$$\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{y_k^T d_k} \qquad \dots(6)$$

performs more effective more details can be found in [12].

It is well known that linear conjugate gradient methods generate a sequence of search directions d_{k+1} such that the following cojugacy conditions holds :

$$d_i^T H d_j = 0 \, , \quad \forall \, i \ne j \qquad \dots(7)$$

where H is the Hessian of the objective function. For general nonlinear function f, we know by the mean value theorem that there exists some $t \in (0,1)$ such that

$$d_{k+1}^T y_k = \alpha_k d_{k+1}^T \nabla^2 f(x_k + t\alpha_k d_k) \quad ..(8)$$

Therefore, it is reasonable to replace (8) with the following conjugacy condition :

$$d_{k+1}^T y_k = 0 \qquad \dots(9)$$

Recently, extensions of (9) have been studied in [5,8] that are based on the standard secant equation

Two Modified Spectral Conjugate Gradient methods for Optimization

Osama Mohammed Taher Wais

Dep. of Mathematic

College of Basic Education

University of Mosul , Iraq

(received in 13\2\2018., accepted in 17\4\2018)

Abstract:

This paper presents two modified spectral conjugate gradient methods which are designed for solving nonlinear unconstrained optimization problems. The presented methods have sufficient descent properties . We prove that the methods is globally convergent. Experimental results indicate that the new proposed methods more efficient than the Dai and Yuan - method .

طريقتان متطورتان من طرائق التدرج المترافق الطيفية في الأمثلية

الملخص:

في هذا البحث تم استحداث طريقتين جديدتين من طرائق التدرج المترافق الطيفية لحل مسائل الأمثلية اللاخطية وغير المقيدة. تمتلك الطرائق المقدمة خصائص الانحدار الكافي . كما تم اثبات التقارب الشامل للطرائق . النتائج العددية أثبتت كفاءة الطريقتين مقارنة بطريقة داي –يوان.