The spectral form of the Dai-Yuan conjugate gradient algorithm

الصيغة الطيفية لخوارزمية التدرج المترافق ل داي-يوان

Prof. Abdul-Ghafoor J. Salem.   Assistant Prof. Khalil K. Abbo

Department of Mathematics

College of computers sciences and mathematics

Mosul University

Abstract

Conjugate gradient (CG) methods comprise a class of unconstrained optimization algorithms which characterized by low memory requirements and strong local and global convergence properties. Most of CG methods don't always generate a descent search directions, so the descent or sufficient descent condition is usually assumed in the analysis and implementations. By assuming a descent and pure conjugacy conditions a new version of spectral Dai-Yuan (DY) non-linear conjugate gradient method introduced in this article. Descent property for the suggested method is proved and numerical tests and comparisons with other methods for large-scale unconstrained problems are given.

الملخص

طرائق المتجهات المترافقة (CG) تشكل صنف من خوارزميات ألأمثليه غير المقيدة وتتميز هذه الطرائق بأنها لاتحتاج الى خزن مصفوفات وكذلك لها خاصية التقارب المحلي والمطلق. اغلب طرائق (CG) لأتولد متجهات بحث انحدارية دائما لذلك خاصية الانحدار عادة يفرض عند تحليل وتمثيل هذه الخوارزميات. بفرض خاصيتي الانحدار والترافق الخالص اقترحنا صيغة طيفية جديدة لخوارزمية داي-يوان للمتجهات المترافقة غير الخطية, تم برهان خاصية الانحدار للخوارزمية المقترحة وكذلك تم مقارنتها عمليا مع خوارزميات اخرى في نفس المجال.

1-Intrduction

The non-linear conjugate gradient (CG) method is a very useful technique for solving large scale unconstrained minimization problems and has wide applications in many fields [9]. This method is an iterative process which requires at each iteration the current gradient and previous direction, which is characterized by low memory requirements and strong local and global convergence properties [3 and 12].

In this paper, we focus on conjugate gradient methods applied to the non-linear unconstrained minimization problem:

$$\min f(x), \quad x \in R^n . \qquad \text{.........(1)}$$

Where $f : R^n \to R$ is continuously differentiable function and bounded below. A conjugate gradient method generates a sequence x_k, $k \geq 1$ starting from an initial guess $x_1 \in R^n$, using the recurrence

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

Where the positive step size α_k is obtained by a line search, and the directions d_k are generated by the rule:

$$d_{k+1} = -g_{k+1} + \beta_k d_k , \quad d_1 = -g_1 \qquad \text{.........(3)}$$

Where $g_k = \nabla f(x_k)$, and let $y_k = g_{k+1} - g_k$ and $s_k = x_{k+1} - x_k$, here β_k is the CG update parameter. Different CG methods corresponding to different choice for the parameter β_k see [1, 4 and 10]. The first CG algorithm for non-convex problems was proposed by Fletcher and Revees (FR) in 1964 [11], which defined as

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}. \qquad \text{.........(4)}$$

We know that the other equivalents forms for β_k are Polack-Ribeir (PR) and Hestenes- Stiefel (HS) for example

$$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{g_k^T g_k}. \quad \text{and} \quad \beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}. \qquad \text{.........(5)}$$

Although all the above formulas are equivalent for convex quadratic functions, but they have different performance for non-quadratic functions, the performance of a non-linear CG algorithm strongly depends on coefficient β_k. Dai and Yuan (DY) in [6] proposed a non-linear CG method (2) and (3) with β_k defined as

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

Which generates a descent search directions

$$d_k^T g_k < 0. \qquad \text{.........(7)}$$

At every iteration k and convergence globally to the solution if the following Wolfe conditions are used to accept the step-size α_k [2]:

$$f(x_k + \alpha_k d_k) \le f(x_k) + c_1 \alpha_k g_k^T d_k \qquad \text{.........(8)}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge c_2 g_k^T d_k \qquad \text{.........(9)}$$

Where $0 < c_1 < c_2 < 1$. Condition (8) stipulates a decrease of f along d_k if (7) satisfied. Condition (9) is called the curvature condition and it's role is to force α_k to be sufficiently far a way from zero [12]. Which could happen if only condition (8) were to be used. Conditions (8) and (9) are called standard Wolfe conditions (SDWC). Notice that if equation (8) satisfied then always there exists $\bar{\alpha} > 0$ such that for any $\alpha_k \in [0, \bar{\alpha}]$ the conditions (8) and (9) will be satisfied according to the theorem (1) given later. If we wish to find a point α_k, which is closer to a solution of the one dimensional problem

$$\operatorname*{Min}_{\alpha>0} \Phi(\alpha) = \min_{\alpha>0} f(x_k + \alpha\, d_k) \qquad \text{.........(10)}$$

Than a point satisfying (8) and (9) we can impose on α_k the strong Wolfe conditions (STWC):

$$f(x_k + \alpha_k d_k) \le f(x_k) + c_1 \alpha_k g_k^T d_k \qquad \text{.........(11)}$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \le c_2 \left| g_k^T d_k \right| \qquad \text{.........(12)}$$

Where $0 < c_1 < c_2 < 1$. In contrast to (SDWC) $g_{k+1}^T d_k$ cannot be arbitrarily large [12]. The (STWC) with the sufficient descent property

$$d_k^T g_k < -c \|g_k\|., \; c \in (0,1) \qquad \text{.........(13)}$$

Widely used in the convergence analysis for the CG methods.

Theorem (1): Assume that f is continuously differentiable and that is bounded below along the line $x = x_k + \alpha\, d_k$, $\alpha \in (0,\infty)$. Suppose also that d_k is a direction of descent (7) is satisfied if $0 < c_1 < c_2 < 1$ then there exist nonempty intervals of step lengths satisfying the (SDWC) and (STWC) conditions. For proof see [12].

The Fletcher-Revees (FR) and Dai-Yuan (DY) methods have common numerator $g_{k+1}^T g_{k+1}$. One theoretical difference between these methods and other choices for the update parameter β_k is that the global convergence theorems only require the Lipschitz assumption not the bounded ness assumption [9].

The global convergence for the methods with $g_{k+1}^T g_{k+1}$ in the numerator of β_k established with exact and inexact line searches for general functions [2, 7, and 13]. Despite the strong convergence theory that has been developed for methods with $g_{k+1}^T g_{k+1}$ in the numerator of β_k, these methods are all susceptible to jamming, that is they begin to take small steps without making significant progress to the minimum [9]. On the other hand the convergence of the methods with $g_{k+1}^T y_k$ in the numerator

(PR) and (HS) for general non-linear function are uncertain, in general the performance of these methods is better than the performance of the methods with $g_{k+1}^T g_{k+1}$ in the numerator of β_k see [9], but they have weaker convergence theorems.

This paper is organized as follows in section 2 new spectral form for DY non-linear conjugate gradient algorithm suggested. In section 3 we will show that our algorithm satisfies sufficient descent condition for every iteration. Section 4 presents numerical experiments and comparisons.

2. New spectral form for Dai and Yuan CG method

An attractive feature of the CG method is that the following (pure conjugacy condition)

$$y_k^T d_{k+1} = 0 \qquad \text{.........(14)}$$

Is always holds if the objective function $f(x)$ is convex quadratic and line search is exact [8]. In this section we use the relation (7) and (14) to derive new spectral DY conjugate gradient method. Consider the search direction of the form

$$d_{k+1} = -\gamma_{k+1} g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k} s_k \,, \quad d_1 = -g_1 \qquad \text{.........(15)}$$

Where γ_k is parameter. Assume that the search direction in (15) satisfies the relation (7) i . e

$$d_{k+1}^T g_{k+1} = -\gamma_{k+1} \, g_{k+1}^T g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k} s_k^T g_{k+1} < 0 \qquad \text{.........(16)}$$

$$-\gamma_{k+1} + \frac{s_k^T g_{k+1}}{y_k^T s_k} < 0$$

or

$$\gamma_{k+1} = \frac{s_k^T g_{k+1}}{y_k^T s_k} + c, \quad c > 0$$

Then

$$\gamma_{k+1} = \frac{s_k^T g_{k+1} + c\, y_k^T s_k}{y_k^T s_k} \qquad \text{.........(17)}$$

To find the value of c, we use the pure conjugacy condition (14) i e $y_k^T d_{k+1} = 0$ then

$$y_k^T d_{k+1} = -\gamma_{k+1}\ y_k^T g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k}\ y_k^T s_k = 0 \qquad (18)$$

With simple algebra we get

$$c = \frac{g_{k+1}^T g_{k+1}\ y_k^T s_k - s_k^T g_{k+1}\ y_k^T g_{k+1}}{y_k^T g_{k+1}\ y_k^T s_k} \qquad (19)$$

Equations (17) and (19) gives

$$\gamma_{k+1} = \frac{g_{k+1}^T g_{k+1}}{y_k^T g_{k+1}} \qquad (20)$$

Therefore the new spectral DY search direction is

$$d_{k+1} = -\gamma_{k+1}\ g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k}\, s_k \qquad \text{.........(21)}$$

With γ_{k+1} defined in the equation (17)

Algorithm (spectral form of DY. SPDY say)

Step (1): Choose an initial starting point $x_1 \in R^n$ *and* $\varepsilon > 0$, consider

$$d_1 = -g_1\ ,\quad \alpha_1 = \frac{1}{\|g_1\|}\ ,\ \ and\ \ k = 1$$

Step(2): Test for convergence. If $\|g_k\| < \varepsilon$ stop x_k is optimal

Else go to step(3)

Step(3): Compute α_k satisfying the (SDWC) or (STWC) 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(4): Direction computation: compute γ_{k+1} from (20) and set

$d = -\gamma_{k+1} \; g_{k+1} + \dfrac{g_{k+1}^T g_{k+1}}{y_k^T s_k} s_k$. If Powell restart is satisfied then

$d_{k+1} = -\gamma_{k+1} g_{k+1}$

Else $d_{k+1} = d$, compute initial guess for $\alpha_{k+1} = \alpha_k (\dfrac{\|d_k\|}{\|d_{k-1}\|})$ and

Set $k = k+1$ go to step(2)

3. Descent property of the SPDY algorithm

An important feature for any minimization algorithm is the descent (7) or the sufficient descent (13) property. In this section we proof that our suggested new algorithm (SPDY) generates a sufficient descent directions for each iteration k.

Theorem (1):

Suppose that the step-size α_k satisfies the standard Wolfe conditions (SDWC), consider the search directions d_k generated from (21) where γ_{k+1} computed from (20) then the search directions d_{k+1} satisfies the sufficient descent condition (13) for all k.

Proof

The proof is by induction. If k=1 then $d_1^T g_1 = -g_1^T g_1 = -\|g_1\| < 0$ then the sufficient descent holds with c=1, know let $s_k^T g_k < -c\|g_k\|$ to proof for k+1, multiply (21)by g_{k+1}^T to get

$$g_{k+1}^T d_{k+1} = -\gamma_{k+1} \; g_{k+1}^T g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k} \; g_{k+1}^T s_k \qquad (22)$$

Note that from second standard Wolfe condition (9)we have

(i) $\quad y_k^T s_k \geq (c_2 - 1)\, s_k^T g_k > 0 \qquad (23)$

(ii)

$$s_k^T g_{k+1} = s_k^T g_{k+1} - s_k^T g_k + s_k^T g_k \le s_k^T g_{k+1} - s_k^T g_k = s_k^T y_k$$

$$\therefore \ \frac{s_k^T g_{k+1}}{s_k^T y_k} \le 1$$

From (20), (2.9) and (23) we get

$$g_{k+1}^T d_{k+1} \le -\frac{g_{k+1}^T g_{k+1}}{y_k^T g_{k+1}} \ g_{k+1}^T g_{k+1} + g_{k+1}^T g_{k+1}$$

$$= -(\frac{g_{k+1}^T g_{k+1}}{y_k^T g_{k+1}} - 1)\, g_{k+1}^T g_{k+1} = -(\frac{g_{k+1}^T g_k}{y_k^T g_{k+1}})\, g_{k+1}^T g_{k+1}$$

Applying the inequality $u^T v \le \frac{1}{2}(\|u\|^2 + \|v\|^2)$ then

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

$$whre \quad c = \frac{\|g_{k+1}\|^2 + \|g_k\|^2}{\|y_k\|^2 + \|g_{k+1}\|^2}$$

4. Numerical results and comparisons

In this section we present the computation performance of a FORTRAN implementation of the SPDY, DY and FR algorithms on a set of unconstrained optimization test problems. We selected (15) large-scale unconstrained optimization test problems in extended or generalized form from [5]. For each function we have considered n=100, 1000 (where n is the number of variables). All algorithms implement the standard Wolfe line search conditions with $c_1 = 0.0001$ and $c_2 = 0.9$ and same stopping criterion $\|g_k\|_\infty < 10^{-6}$, where $\|\ .\ \|_\infty$ is the maximum absolute component of a vector.

The comparison of algorithms are given in the following context. We say that, in the particular problem i the performance of Algorithm(Alg1) was better than the performance of Alg2 if the number of iterations (iter) or the number of function-gradient evolutions (fg) or the number of restartes (irs) of Alg1 was less than the number of (iter) or (fg) or the (irs) corresponding to Alg2 respectively. Table(1) and table(2) shows the details of numerical results for the Fletcher-Revees (FR), Dai-Yuan (DY) and our algorithm (SPDY).

Table (1) Comparison of the algorithms for n=100

Test Problems	**FR**			**DY**			**SPDY**		
	Iter	**fg**	**irs**	**Iter**	**fg**	**irs**	**Iter**	**fg**	**irs**
Trigonometric	18	34	10	18	34	9	18	33	10
Ex. Rosenbrock (CUTE)	41	84	22	40	81	24	34	72	18
Ex. White & Holst	36	76	20	34	68	18	31	67	17
Perturbed Quadratic	101	154	31	82	123	29	95	145	29
Diagonal 2	67	107	23	59	100	17	55	93	18
Hager	28	46	11	27	45	12	25	41	10
Generalized Tridiagonal 2	36	57	11	39	59	15	40	60	16
Extended Powell	59	113	20	72	136	25	66	125	16
Extended BD1	42	70	39	52	86	51	44	75	41
Extended Maratos	70	160	36	68	151	34	64	151	29
Ex. Quad. Penalty QP2	28	60	15	24	51	12	23	54	12
Partial Perturbed Quad.	74	114	26	84	132	23	75	113	21
Almost Perturbed Quad	84	133	21	98	153	31	101	152	32
Tridiago. Perturbed Quad.	105	168	35	105	164	33	95	151	24
ENGVAL1 (CUTE)	27	47	9	23	44	5	29	50	11
Total	816	1423	329	825	1427	335	795	1382	304

Table (2) Comparison of the algorithms for n=1000

Test Problems	FR			DY			SPDY		
	Iter	fg	irs	Iter	fg	irs	Iter	fg	irs
Trigonometric	29	53	19	32	57	19	29	52	18
Ex. Rosenbrock (CUTE)	40	92	20	38	83	21	34	75	18
Ex. White & Holst	36	76	18	32	69	17	26	53	13
Perturbed Quadratic	284	437	83	326	519	85	338	527	98
Diagonal 2	219	360	71	189	313	56	190	315	60
Hager	278	496	248	285	510	255	188	218	159
Generalized Tridiagonal 2	64	98	25	63	99	27	65	100	26
Extended Powell	67	128	22	77	141	24	58	111	17
Extended BD1	53	87	53	53	87	53	53	87	53
Extended Maratos	70	155	36	70	160	34	63	147	29
Ex. Quad. Penalty QP2	36	87	20	37	90	20	32	85	20
Partial Perturbed Quad.	225	373	56	240	391	63	203	335	47
Almost Perturbed Quad	323	503	88	316	502	95	290	451	80
Tridiago. Perturbed Quad.	406	628	114	332	510	87	333	525	95
ENGVAL1 (CUTE)	104	180	90	113	202	101	50	98	25
total	2234	3753	963	2203	3733	957	1952	3179	758

References

[1] Abbas Y. Al-Bayati, Abdul-Ghafoor J. and Khalil A. (2009) 'Two- Version of conjugate gradient algorithm Based on conjugacy Conditions for unconstrained optimization'. American J. of Economics and Business Administration 1, (2).

[2] Al-Baali M. (1985) 'Descent property and global convergence of the Fletcher and Revees Method with inexact line search'. SIAM J. Numer. Anal 5

[3] Andrei, N.(2010) ' Accelerated scaled memory less BFGS preconditioned conjugate Gradient algorithm for unconstrained optimization' European J. Of Operational Research. 204

[4] Andrei, N. (2008) '40 conjugate gradient algorithm for unconstrained Optimization' ICI. Technical Report No. 13108.

[5] Andrei, N.(2008) An unconstrained optimization test functions Collection, Advanced Modeling and Optimization, 10.

[6] Dai, Y. and Yuan, Y. (1999) 'A Nonlinear conjugate gradient method with a strong global convergence property'. SIAM J. Optim.,10.

[7] Dai Y. (2001) 'New properties of anon-linear conjugate gradient method' Numer. Math. 89.

[8] Dai,Y. and Liao, L. (2001) 'New conjugacy conditions and related non-linear CG methods' Appl. Math optim.,(43). Spring-verlag, New York.

[9] Hager, W. and Zhang, H. (2006) 'A survey of non-linear conjugate gradient methods'

Pacific Journal of Op., 2.

[10] Khalil A. (2008)
'New CG method for large-scale unconstrained optimization based on Nazareth theorem'. Iraq J. of Statistical science, (13).

[11] Fletcher R. and Reeves G. (1964)
' Function minimization by conjugate gradients'
computer. Journal , Vol.(7).

[12] Pytlak R. (2009)
'Conjugate Gradient Algorithms in Non-convex Optimization'
Non-convex Optimization and It's Applications. V. 89, Springer-V

[13] Zoutendijk, G. (1970)
'Nonlinear programming, computational methods'
In Integer and Nonlinear Programming, J. Abadie(Ed),
North- Holand