Iraqi Journal of Statistical Sciences (25) 2013
The 6th Scientific Conference of the College of Computer Sciences & Mathematics
pp [108-117]

New Projection Matrix for the Stander Conjugate Gradient Method

Dr. ABBAS Y.AL – BAYATI   Dr. EMAN T. HAMED

Dr. Hamsa Th. Chilmerane

Abstract:

In this paper, we have derived anew proposed algorithm for conjugate gradient method based on a projection matrix. This Algorithm satisfies the sufficient descent condition and the globally converges . Numerical comparisons with a standard conjugate gradient algorithm show that this algorithm very effective depending on the number of iterations and the number of functions evaluation.

المصفوفة التساقطيه الجديدة لطريقه التدرج المترافق القياسيه

الخلاصة

تم في هذا البحث اقتراح متجه جديد لطريقة التدرج المترافق يعتمد على المصفوفة التساقطية .تم أثبات أن الخوارزمية الجديدة تحقق شروط الانحدار الكافي والتقارب الشامل ودلت النتائج العددية على أن الطريقة المقترحة جيدة وفعالة بالاعتماد على عدد التكرارات وعدد حسابات الدالة عند مقارنتها من الطريقةالكلاسيكية.

Keywords: unconstrained optimization, a projection matrix, inexact line search ,wolf line search global convergence.

1-Introduction:

Let us consider the nonlinear unconstrained optimization problem

$$\min\{f(x): x \in R^n\} \quad \ldots\ldots\ldots\ldots(1)$$

Where f is smooth and its gradient g is available.

* Prof / Dept. Mathematics / Computer Sciences and Mathematics College / Mosul University

** Assist Prof / Dept. operational research & intelligent techniques/ Computer Sciences and Mathematics College / Mosul University

*** Assist Prof / Dept. Mathematics / Computer Sciences and Mathematics College / Mosul University

Received Date 9/10/2013 ——— Accept Date 21/11/2013

Conjugate gradient methods are efficient for solving (1), especially when the dimension n is large. The iterates of conjugate gradient methods for solving (1) are obtained by

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

where λ_k is a steplength, which is computed by carrying out some line search, and d_k is the search direction defined by

$$\begin{bmatrix} d_k = -g_k & k = 1 \\ d_{k+1} = -g_{k+1} + \beta_k d_k & k \geq 1 \end{bmatrix} \quad \text{.................................(3)}$$

where β_k is a scalar. Some well-known conjugate gradient methods include the Hestenes–Stiefel (HS) method , Fletcher–Reeves (FR) method , the Polak–Ribière–Polyak (PRP) method, and the Dai–Yuan (DY) method and Al-Bayati & Al-Assady . The parameters β_k of these methods are specified as follows

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \qquad \text{(Hestenes-Stiefel,1952)}$$

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \qquad \text{(Fletcher-Reeves (FR),1969)}$$

$$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{g_k^T g_k} \qquad \text{(Polak- Ribière (PR))}$$

$$\beta_k^{BA} = \frac{-y_k^T y_k}{d_k^T g_k} \qquad \text{(Al-Bayati \& Al-Assady,1986)}$$

$$\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{d_k^T y_k} \qquad \text{(Dai-Yuan (DY),1999)}$$

The stepsize λ_k is usually chosen to satisfy certain line search conditions. Among them, the so-called strong Wolfe line search conditions require that,The weak Wolfe-conditions:

$$f(x_k + \lambda_k d_k) - f(x_k) \le \delta \lambda_k g_k^T d_k \quad \text{....(4)}$$

$$g(x_k + \lambda_k d_k)^T d_k \ge \sigma g_k^T d_k \quad \text{.......(5)}$$

the strong Wolfe-conditions:

$$f(x_k + \lambda_k d_k) - f(x_k) \le \delta \lambda_k g_k^T d_k \quad \text{.................(6)}$$

$$\left| g(x_k + \lambda_k d_k)^T d_k \right| \le -\sigma g_k^T d_k \quad \text{..........(7)}$$

where $\delta \in (0, \frac{1}{2})$ and $\sigma \in (0,1)$

Beale's three-term restart conjugate gradient method is a well-known three-term conjugate gradient method in which

$$d_{k+1} = -g_{k+1} + \beta_k d_k + \gamma_k d_l \text{ ....(8)}$$

where 1 ≤ $l < k$, $\gamma_k = g_k^T \; y_l / d_l^T y_l$ Another well-known method is

Nazareth's three-term recurrence , where

$$d_{k+1} = -y_k + \frac{y_{k-1}^T y_{k-1}}{y_{k-1}^T d_{k-1}} d_{k-1} + \frac{y_k^T y_k}{y_k^T d_k} \; d_k \text{ ....(9)}$$

Zhang, Li , Zhou ,Weijun,(2007) and Zhang, L.,Weijun Zhou, (2007) proposed a three-term PRP conjugate gradient method (TTPRP) and a three-term FR conjugate gradient method (TTFR), respectively, that is,

$$d_{k+1}^{TTPRP} = -g_{k+1} + \beta_k^{PRP} d_k - \theta_k^1 y_k \text{ ....(10)}$$

$$d_{k+1}^{TTFR} = -g_{k+1} + \beta_k^{FR} d_k - \theta_k^2 g_{k+1} \text{ ....(11)}$$

$\theta_k^1 = \dfrac{g_{k+1}^T d_k}{\|g_k\|^2}$ and $\theta_k^2 = \dfrac{d_k^T y_k}{\|g_k\|^2}$

2.New proposed method

in this paper we will get new projection matrix from three-term CG-algorithm as follows:

$$d_{k+1} = -g_{k+1} + \beta_k^{FR} d_k - \beta_k^{FR} \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k \qquad \ldots\ldots(12)$$

If we use the exact line search and, then our method (12) becomes the nonlinear conjugate gradient method (3)

$$d_{k+1} = -g_{k+1} + \beta_k^{FR} (I - \frac{y_k g_{k+1}^T}{d_k^T y_k}) d_k \qquad \ldots\ldots(13)$$

Where the matrix $\left(I - \frac{y_k g_{k+1}^T}{d_k^T y_k} \right)$ is also a projection matrix into the orthogonal complement of Span $\{g_{k+1}\}$

3. The New Algorithm :

Step 1 : For the initial point τ_0 , $x_1 \in R^n, \varepsilon$, Set $d_1 = -g_1$,k = 1 , if $\|g_1\| \le \varepsilon$, then
stop.

Step 2: Set $d_k = -g_k$

Step 3 : Find $\lambda_k > 0$ satisfying the strong wolf conditions.

Step 4: Let $x_{k+1} = x_k + \lambda_k d_k$ and If $\|g_{k+1}\| \le \varepsilon$ then stop .

Step 5 : compute the search direction d_{k+1} by (13)
.

Step 6 : If $k = n$ or $\frac{|g_k^T g_{k+1}|}{\|g_{k+1}\|^2} \ge 0.2$,then go to step 2.

Step 7: Set k: = $k+1$, go to Step 3.

4. Global Convergence Properties for the new Suggestion algorithm:

In this section we will study the convergence of the new proposed method depending by the following assumption

Assumption(A) :

(i) The level set $S = [x \in R^n ; f(x) \le f(x_0)\}$ is bounded.

(ii) In a neighborhood N of S, the function f is continuously differentiable and its gradient is Lipschitz continuous, i.e there exists a constant $L > 0$ such that

$$\|\nabla f(x) - f(y)\| \le L\|x - y\| \qquad \text{for all } x, y \in N \quad ..(14)$$

We can get from assumption (A) that there exists positive constant $\psi > 0$, such that:

$$\|g(x)\| \le \psi \qquad \forall\, x \in S \qquad ….(15)$$

Lemma (1). Suppose that the assumption (A) hold and consider any conjugate gradient method (2) and (3), where is a descent direction d_k and λ_k is obtained by the strong Wolfe line search

If

$$\sum_{k\ge 1} \frac{1}{\|d_k\|^2} = \infty \qquad ……………….…(16)$$

Then

$$\liminf_{k\to\infty} \|g_k\| = 0 \qquad ………………………...(17)$$

(Dai, Y.H., et al, 1999)

Lemma:

Suppose the assumption (A) hold , let the sequence $\{ x_k \}$ generated by (2) and the step length λ_k satisfies wolf conditions, then the direction which is define in (13)is satisfied sufficient condition

Proof :

By multiply both side of (13) by g_{k+1}^T and dividing by $\|g_{k+1}^T\|^2$ we get :

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} = \beta_k^{FR} \frac{g_{k+1}^T d_k}{\|g_{k+1}^T\|^2} - \beta_k^{FR} \frac{(g_{k+1}^T y_k)(g_{k+1}^T d_k)}{d_k^T y_k \|g_{k+1}\|^2} \quad ………………………..(18)$$

By using strong wolf condition we get:

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \le \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \left(\frac{-\sigma d_k^T g_k}{\|g_{k+1}^T\|^2} \right) + \frac{(g_{k+1}^T y_k)(\sigma d_k^T g_k)}{\|g_{k+1}\|^2 \, d_k^T y_k} \quad ……………………(19)$$

Since $g_{k+1}^T y_k \le \|g_{k+1}\| \|y_k\|$

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \le \frac{-\sigma d_k^T g_k}{\|g_k\|^2} + \frac{\|g_{k+1}\| \|y_k\| \sigma d_k^T g_k}{d_k^T y_k \|g_{k+1}\|^2} \quad \ldots\ldots(20)$$

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \le \frac{\sigma d_k^T y_k}{(\sigma+1)\|g_k\|^2} - \frac{\|y_k\| \sigma d_k^T y_k}{d_k^T y_k \|g_{k+1}\| \ (\sigma+1)} \quad \ldots\ldots(21)$$

$$\frac{g_{k+1}^T d_{k+1} + \|g_{k+1}\|^2}{\|g_{k+1}\|^2} \le \frac{\sigma d_k^T y_k}{(\sigma+1)\|g_k\|^2} = c \quad \ldots\ldots(22)$$

where c is small positive constant

$\therefore g_{k+1}^T d_{k+1} \le -(1-c)\|g_{k+1}\|^2$

the proof is complete.

Theorem (2):

(Global convergence for new proposed method):

Consider the iteration method which is define in (2) where d_k defined by,(13) and suppose the assumption A holds. Then the new algorithm either stops at stationary point i.e. $\|g_k\| = 0$ or $\lim\inf_{k\to\infty} \|g_k\| = 0$

Proof:

Form (12),we get

$$d_{k+1} = -g_{k+1} + \beta_k^{FR} d_k - \beta_k^{FR} \frac{g_{k+1}^T y_k}{d_k^T y_k} y_k$$

$$\|d_{k+1}\| \le \|g_{k+1}\| + |\beta_k^{FR}| \|d_k\| + |\beta_k^{FR}| \frac{g_{k+1}^T d_k}{d_k^T y_k} \|y_k\| \quad \ldots\ldots(23)$$

$$\|d_{k+1}\| \le \|g_{k+1}\| + |\beta_k^{FR}| \|d_k\| + |\beta_k^{FR}| \frac{\sigma \|g_k\|^2}{d_k^T y_k} \|y_k\| \quad \ldots\ldots(24)$$

$$\|d_{k+1}\| \le \psi$$

$$\sum_{k\ge1} \frac{1}{\|d_{k+1}\|^2} \ge \sum_{K\ge1} \frac{1}{\psi} = \sum_{k\ge1} 1 = \infty$$

$$\sum_{k\ge1} \frac{1}{\|d_{k+1}\|} = \infty$$

by using lemma(1) we get :

$$\lim_{k\to\infty} \|g_k\| = 0$$

5:Numerical experiments

In this section, we will test the feasibility and effectiveness of the Algorithm 2.1. The algorithm is implemented in Fortran77 code using double precision arithmetic and Comparison our new algorithm with standard three term CG-algorithm in Table (1)

overall the calculation and for different dimension for ($100 \le n \le 5000$), all the algorithms in this paper use the same ILS strategy.

.

All the results are obtained using (Pentium 4 computer). All programs are written in FORTRAN 90 language and for all cases the stopping criterion taken to be:

$$\|g_{k+1}\| \le 10^{-5}$$

The comparative performance for all of these algorithms is evaluated by considering number of function Evaluations NOF and number of iterations NOI.

Table (1) Comparison of our new algorithm with standard FR CG-algorithm.

Test fun.	**Dim.**	**FR. Algorithm**		**New Algorithm**	
		NOI	**NOF**	**NOI**	**NOF**
Powell	100	50	136	40	109
Central	100	33	222	30	234
Edger	100	6	15	6	15
Cubic	100	16	44	16	46
Wolfe	100	49	99	46	93
Sum	100	12	64	14	69
Wood	100	29	67	30	69
Miele	100	46	146	45	144
Rosen	100	22	55	22	55
Recip	100	5	16	5	16
Powell	1000	54	164	40	109
Central	1000	40	312	33	278
Edger	1000	6	15	6	15
Cubic	1000	16	44	16	46
Wolfe	1000	64	129	53	107
Sum	1000	21	110	23	114
Wood	1000	29	67	30	69

Miele	1000	53	180	45	144
Rosen	1000	22	55	22	55
Recip	1000	5	16	5	16
Powell	10000	56	168	43	128
Central	10000	45	384	34	292
Edger	10000	6	15	6	15
Cubic	10000	16	44	16	46
Wolfe	10000	118	238	131	266
Sum	10000	32	161	35	165
Wood	10000	29	67	30	69
Miele	10000	53	180	53	182
Rosen	10000	22	55	22	55
Recip	10000	6	18	6	18
Total		961	3286	903	3039

References

[1] AL - Bayati, A.Y. and AL-Assady, N.H. (1986). “Conjugate gradient method” Technical Research report, No (1), school of computer studies, Leeds University.

[2] Dai,Y. and Yuan, Y. (1999), "A Nonlinear conjugate gradient method with a strong global convergence property", SIAM J. Optim.,10. ,177– 182.

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

[4] Hestenes, M.R and Stiefel, E., (1952)," Method of conjugate gradient for solving linear equations", J. Res. Nat. Bur. Stand. 49, 409–436.

Wanyou Cheng *y* Dong-Hui Li (2010)*z* "A Two Term PRP Based Descent Method " College of Mathematics and Econometrics, Hunan University,

[5] Liu,Y. and Storey,C.,(1991) ," Efficient generalized Conjugate Gradient Algorithms", part1:theory,J. Optimizat .Theory conjugate Appl.69,129–137.

[6] Zhang,L., Zhou,Li, , D.H. ,(2006),” Global convergence of a modified Fletcher-Reeves conjugate Method with Armijo-type line search” Numer. Math. 104 ,561–572.

[7] Zhang, Li , Zhou ,Weijun,(2007) ," A descent nonlinear conjugate gradient method for large-scale unconstrained optimization Applied Mathematics and Computation 187 , 636–643.

[8] Zhang,L.,Weijun Zhou, (2007) , "Some descent three term conjugate gradient methods and their global convergence", College of Mathematics and Computational Science, Journal of Computational and Applied Mathematics 216 ,251 –264.

[9] Wanyou Cheng *y* Dong-Hui Li (2010) "A Two Term PRP Based Descent Method " College of Mathematics and Econometrics, Hunan University,

Appendix

1. Generalized Central Function:

$$f(x) = \sum_{i=1}^{n/4} \begin{bmatrix} (\exp(x_{4i-3}) - x_{4i-2})^4 + 100(x_{4i-2} - x_{4i-1})^6 + \\ \arctan(x_{4i-1} - x_{4i}))^4 + x_{4i-3} \end{bmatrix},$$

x_0=(1,2,2,2;...)^T.

2. Extended Wood Function

$$f(x) = \sum_{i=1}^{n/4} (100(x_{4i-3}^2 - x_{4i-2})^2 + (x_{4i-3} - 1)^2 + 90(x_{4i-1}^2 - x_{4i})^2 +$$
$$(1 - x_{4i-1})^2 + 10.1((x_{4i-2} - 1)^2 + (x_{4i} - 1)^2) + 19.8(x_{4i-2} - 1)(x_{4i} - 1)),$$

$$x_0 = (-3,-1,-3,-1,...,-3,-1,-3,-1)^T.$$

3. Generalized Powell Function:

$$f(x) = \sum_{i=1}^{n/4} \begin{bmatrix} (x_{4i-3} + 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + \\ (x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-3} - x_{4i})^4 \end{bmatrix}$$

$$x_0 = (3,-1,0,1,...,3,-1,0,1)^T.$$

4. Sum of Quadritics (SUM) function:

$$f(x) = \sum_{i=1}^{n} (x_i - i)^4 \quad .$$

x_0=(1 ,…)^T.

5. Wolfe Function:

$$f(x) = [-x_1(3-x_1/2)+2x_2-1]^2 + \sum_{i=1}^{n-1}\begin{bmatrix}[x_{i-1}-x_i(3-x_i/2+2x_{i+1}-1)]^2 \\ +[x_{n-1}-x_n(3-x_n/2)-1]^2\end{bmatrix}$$

$x_0=(-1,\ldots)^T$

6. Rosenborck Function:

$$f(x) = \sum_{i=1}^{n/2} 100(x_{2i}-x_{2i-1}^2)^2 + (1-x_{2i-1})^2$$

$x_0=(-1.2,1,\ldots)^T$.

7. Generalized Recip Function:

$$f(x) = \sum_{i=1}^{n/3}\left[(x_{3i-1}-5)^2 + x_{9i-1}^2 + \frac{x_{3i}^2}{(x_{3i-1}-x_{3i-2})^2}\right],$$

$$X_0 = (2\,,5,1,\ldots,2\,,5,1)^T$$

8. Miele Function:

$$f(x) = \sum_{i=1}^{n/4}\left(\exp(x_{4i-3})+10x_{4i-2}\right)^2 + 100(x_{4i-2}+x_{4i-1})^6 + \left(\tan(x_{4i-1}-x_{4i})\right)^4 + (x_{4i-3})^8 + (x_{4i}-1)^2,$$

$$x_0 = (1,2,2,2,\ldots,1,2\,,2,2)^T$$

9. Generalized Edger Function:

$$f(x) = \sum_{i=1}^{n/2}(x_{2i-1}-2)^4 + (x_{2i-1}-2)^2 x_{2i}^2 + (x_{2i}+1)^2$$

$$x_0 = (1.,0.,\ldots,1.,0.)^T$$

10. Generalized Cubic Function:

$$f(x) = \sum_{i=1}^{n/2}\left[100(x_{2i}-x_{2i-1}^3)^2 + (1-x_{2i-1})^2\right],$$

$$x_0 = (-1.2,1,\ldots,-1.2,1)^T$$