International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 7 Issue 6, June-2018, Impact Factor: 4.059

A New Scale in Conjugate Gradient Methods for Solving Unconstrained Optimization

Eman T. Hamed[1], Huda I. Ahmed[2], Eilaf S. Khaleel[3]

[1,2]Department of Operation Research and Intelligent Techniques, College of Computer Sciences and Mathematics
[3]Department of mathematic, College of Computer Sciences and Mathematic. University of Mosul, Mosul, Iraq

ABSTRACT

Conjugate gradient methods arean important class methods for unconstrained optimization, especially for large-scale problems. The descent property for the suggested method is proved under some conditions. Also, we prove that for strongly convex functions the modified method is global convergent. Finally, the numerical results showed that the new method is very efficient for general problems.

HOW TO CITE THIS ARTICLE

Eman T. Hamed Huda I. Ahmed, Eilaf S. Khaleel, "A New Scale in Conjugate Gradient Methods for Solving Unconstrained Optimization", International Journal of Enhanced Research in Science, Technology & Engineering, ISSN: 2319-7463, Vol. 7 Issue 6, June -2018.

1. INTRODUCTION

Conjugate gradient methods are a class of important method for solving:

$$\min\{f(x): x \in R^n\}^{[5]} \qquad (1)$$

where $f: R^n \to R$ is a continuously differentiable. This method generate a sequence x_i of the function f . In general the method has the following form:

$$x_{i+1} = x_i + \alpha_i d_i \qquad i = 0,1,2,\ldots.. \qquad (2)$$

$$d_{i+1} = \begin{cases} -g_{i+1} & \text{if } i = 0 \\ -g_{i+1} + \beta_i d_i & \text{if } i \geq 1 \end{cases} \qquad (3)$$

where $g_i = \nabla f(x_i)$, α_i is a step length obtained by a line search, and β_i is a scalar parameter. There are many ways to select β , and some well-known formulas are given in [7, 8, 9, 13].

The step length α_i in (2) is computed by carrying out line search. The Wolfe line search consist of finding a positive step-length α_i such that:

$$f(x_i + \alpha_i d_i) - f(x_i) \leq \delta\alpha_i g_i^T d_i \qquad (4)$$

$$\left|g(x_i + \alpha_i d_i)^T d_i\right| \leq -\sigma\, g_i^T d_i \qquad (5)$$

where $0 < \delta < \sigma < 1$ [15],[16].

2. A NEWSCALED CONJUGATE GRADIENT METHOD θ_{i+1}

The algorithm in this class of nonlinear conjugate gradient algorithms generate the sequence x_i in (2) and defined the search direction d_{i+1} by:

$$d_{i+1} = -\theta_{i+1} g_{i+1} + \beta_i d_i \qquad (6)$$

where θ_{i+1} is a positive scalar , observe that if $\theta_i = 1$, then we get the classical conjugate gradient algorithm according to the value of formulas ofβ_i .

Eilaf [6] proposed a parameter β_i which is defined by :

$$\beta_i = \frac{g_{i+1}^T y_i - g_{i+1}^T s_i}{d_i^T y_i} - \frac{g_{i+1}^T d_i}{d_i^T g_i} \qquad (7)$$

where $s_i = x_{i+1} - x_i$, $y_i = g_{i+1} - g_i$ .We used this β_i in (6) and in order to get the new scalarθ_{i+1} in our method , multiply both sides of (6) by y_i .

$$d_{i+1}^T y_i = -\theta_{i+1} g_{i+1}^T y_i + \beta_i d_i^T y_i \quad (8)$$

since $d_{i+1}^T y_i = -s_i^T g_{i+1}$ (9)
which is called Perry 's condition [10].
substituting (7) and (9) in (8) :

$$-s_i^T g_{i+1} = -\theta_{i+1} g_{i+1}^T y_i + \left(\frac{g_{i+1}^T y_i - g_{i+1}^T s_i}{d_i^T y_i} - \frac{g_{i+1}^T d_i}{d_i^T g_i} \right) d_i^T y_i$$

$$= -\theta_{i+1} g_{i+1}^T y_i + g_{i+1}^T y_i - s_i^T g_{i+1} - \frac{g_{i+1}^T d_i}{d_i^T g_i} d_i^T y_i$$

$$\theta_{i+1} = 1 - \frac{g_{i+1}^T d_i}{d_i^T g_i} \cdot \frac{d_i^T y_i}{g_{i+1}^T y_i} \quad (10)$$

then the new direction is defined by :

$$d_{i+1} = -\theta_{i+1} g_{i+1} + \beta_i d_i$$

$$= -\left(1 - \frac{g_{i+1}^T d_i}{d_i^T g_i} \cdot \frac{d_i^T y_i}{g_{i+1}^T y_i}\right) g_{i+1} + \left(\frac{g_{i+1}^T y_i - g_{i+1}^T s_i}{d_i^T y_i} - \frac{g_{i+1}^T d_i}{d_i^T g_i} \right) d_i \quad (11)$$

If we used exact line search (ELS) then the new scalar (θ_{i+1}) is equal to 1 .

Algorithm 2.1
Step 1: given $x_1 \in R^n$,$\in > 0$ set k=1
Step 2 : set $d_i = -g_i$, $if \ \|g_i\| < \in$ then stop
Step 3 : find $\alpha_k > 0$ satisfying (4) , (5)
Step 4 : let $x_{i+1} = x_i + \alpha_i d_i$ if $\|g_k\| < \in$ then stop , otherwise continue
Step 5 : compute β_i , θ_{i+1} by the formula (7) , (10) respectively and generate the new search direction d_{i+1} by (11)
Step 6 : if k=n or $|g_{k+1}^T g_k| \geq 0.2\|g_{k+1}\|^2$ (powell restarting [13]) is satisfy , go to step (2) else set k=k+1 and go to step 3

3- CONVERGENCE ANALYSIS

Assume the following basic on the objective function .

Assumption 3.1 [17]

i) The function $f(x)$is bounded in the level set S i.e. there exists positive constant $z > 0$, such that $\|x\| \leq z$,$\forall x \in S$ (12)

ii) If there is a neighborhood W of S ,and it is gradient is Lipschitz continuous , i.e. there exists a constant $L > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \ , \forall \ x, y \ \in W \quad (13)$$

Under these assumptions on f , there exists a constant $\Gamma \geq 0$ such that
$\|\nabla f(x)\| \leq \Gamma$, $\forall \ x \in S$ (14)

Theorem 3.2[4 and 17]

Suppose that the function $f(x)$ is a uniformly convex then there exists a constant $M > 0$ such that

$$\left(g(x) - g(y)\right)^T (x - y) \geq M\|x - y\|^2, \forall \ x, y \ \in S \quad (15)$$

we can rewrite (15) in the following manner

$$y_i^T s_i \geq M\|s_i\|^2 \quad (16)$$

Then (16) with (13) , implies that :

$$M\|s_i\|^2 \leq y_i^T s_i \leq L\|s_i\|^2 \quad (17)$$

Theorem 3.3

Suppose that α_i satisfies the Wolfe condition (4) and (5) , then the direction d_{i+1} given by (11) is a sufficient descent direction .
proof :
since $d_0 = -g_0$, we have $g_0^T d_0 = -\|g_0\|^2 < 0$
now , multiplying (11) by g_{i+1} we have :

$$d_{i+1}^T g_{i+1} = -\left(1 - \frac{g_{i+1}^T d_i}{d_i^T g_i}\frac{d_i^T y_i}{g_{i+1}^T y_i}\right)\|g_{i+1}\|^2 + \left(\frac{g_{i+1}^T y_i - s_i^T g_{i+1}}{d_i^T y_i} - \frac{g_{i+1}^T d_i}{d_i^T g_i}\right) d_i^T g_{i+1}$$

Divided both said by $\|g_{i+1}\|^2$

$$\frac{d_{i+1}^T g_{i+1}}{\|g_{i+1}\|^2} = -\left(1 - \frac{g_{i+1}^T d_i}{d_i^T g_i}\frac{d_i^T y_i}{g_{i+1}^T y_i}\right) + \left(\frac{g_{i+1}^T y_i - s_i^T g_{i+1}}{d_i^T y_i} - \frac{g_{i+1}^T d_i}{d_i^T g_i}\right)\frac{d_i^T g_{i+1}}{\|g_{i+1}\|^2}$$

$$\frac{d_{i+1}^T g_{i+1}}{\|g_{i+1}\|^2} + 1 = \left(\frac{g_{i+1}^T d_i}{d_i^T g_i}\frac{d_i^T y_i}{g_{i+1}^T y_i}\right) + \left(\frac{g_{i+1}^T y_i - s_i^T g_{i+1}}{d_i^T y_i} - \frac{g_{i+1}^T d_i}{d_i^T g_i}\right)\frac{d_i^T g_{i+1}}{\|g_{i+1}\|^2}$$

since $g_{i+1}^T d_i = d_i^T y_i + g_i^T d_i$

$\therefore g_{i+1}^T d_i < d_i^T y_i$ also $s_i^T g_{i+1} \le s_i^T y_i$ and from (5) we have

$$\frac{d_{i+1}^T g_{i+1}}{\|g_{i+1}\|^2} + 1 \le \left(\frac{-\sigma g_i^T d_i}{g_i^T d_i}\frac{d_i^T y_i}{g_{i+1}^T y_i}\right) + \left(\frac{g_{i+1}^T y_i - s_i^T y_i}{d_i^T y_i} + \frac{\sigma g_i^T d_i}{g_i^T d_i}\right)\frac{d_i^T y_i}{\|g_{i+1}\|^2}$$

$$\le \frac{-\sigma d_i^T y_i}{g_{i+1}^T y_i} + \frac{g_{i+1}^T y_i}{\|g_{i+1}\|^2} - \frac{s_i^T y_i}{\|g_{i+1}\|^2} + \sigma\frac{d_i^T y_i}{\|g_{i+1}\|^2}$$

since $g_{i+1}^T y_i \le \|g_{i+1}\|.\|y_i\|$

and from (13) we have $s_i^T y_i \le L\|s_i\|^2$, $d_i^T y_i \le \alpha L\|d_i\|^2$

then we have

$$\le \frac{\sigma\alpha L\|d_i\|^2}{-\|g_{i+1}\|.\|y_i\|} + \frac{\|g_{i+1}\|.\|y_i\|}{\|g_{i+1}\|^2} - \frac{L\|s_i\|^2}{\|g_{i+1}\|^2} + \frac{\sigma\alpha L\|d_i\|^2}{\|g_{i+1}\|^2}$$

$$\le \frac{\|y_i\|}{\|g_{i+1}\|} + \frac{\sigma\alpha L\|d_i\|^2}{\|g_{i+1}\|^2}$$

Let $c_1 = \frac{\|y_i\|}{\|g_{i+1}\|} + \frac{\sigma\alpha L\|d_i\|^2}{\|g_{i+1}\|^2}$

$$\frac{d_{i+1}^T g_{i+1}}{\|g_{i+1}\|^2} + 1 \le c_1$$

$$\frac{d_{i+1}^T g_{i+1}}{\|g_{i+1}\|^2} \le c_1 - 1$$

$$d_{i+1}^T g_{i+1} \le -(1 - c_1)\|g_{i+1}\|^2$$

$$d_{i+1}^T g_{i+1} \le -c_2\|g_{i+1}\|^2$$

Where $c_2 = 1 - c_1$ the proof is complete.

Lemma 3.4 (see [5])

If the Assumption 3.1 holds, and suppose that any conjugate gradient method of the form (2) , (3) where d_i is satisfying the descent condition and α_i is satisfied the strong Wolfe line search .

If $\sum_{i=1}^{\infty} \frac{1}{\|d_i\|^2} = \infty$ (18)

we have that

$\lim_{i\to\infty} \inf\|g_i\| = 0$ (19)

Theorem 3.5

Suppose that Assumption 3.1 hold and the function is a uniformly convex . Consider the conjugate gradient method in the form (2) , (11) where d_{i+1} satisfy the sufficient descent condition then the new method satisfy the global convergence (i.e.$\lim_{i\to\infty}\|g_i\| = 0$) .

Proof:

$d_{i+1} = -\theta_{i+1} g_{i+1} + \beta_i d_i$

$\|d_{i+1}\| = |\theta_{i+1}|\|g_{i+1}\| + |\beta_i|\|d_i\|$ (20)

$$|\beta_i| \le \left|\frac{g_{i+1}^T y_i}{d_i^T y_i}\right| + \left|\frac{s_i^T g_{i+1}}{d_i^T y_i}\right| + \left|\frac{g_{i+1}^T d_i}{g_i^T d_i}\right|$$

since $s_i^T g_{i+1} \le s_i^T y_i$

$$\le \frac{\|g_{i+1}\|\|y_i\|}{\alpha M\|d_i\|^2} + \frac{s_i^T y_i}{\alpha M\|d_i\|^2} + \frac{\sigma g_i^T d_i}{g_i^T d_i}$$

$\le \frac{\|g_{i+1}\|\|y_i\|}{\alpha M\|d_i\|^2} + \frac{L\|s_i\|^2}{\alpha M\|d_i\|^2} + \sigma = c_3$ (21)

$\theta_{i+1} = 1 - \frac{g_{i+1}{}^T d_i}{d^T{}_i g_i} \cdot \frac{d_i{}^T y_i}{g_{i+1}^T y_i}$

$\therefore |\theta_{i+1}| \leq 1 + \left| \frac{\sigma g_i{}^T d_i}{g^T{}_i d_i} \cdot \frac{\alpha L \|d_i\|^2}{g_{i+1}^T y_i} \right|$

$$\leq 1 + \frac{\sigma \alpha L \|d_i\|^2}{|g_{i+1}^T y_i|}$$

since $g_{i+1}^T y_i = \|g_{i+1}\|^2 - g_{i+1}^T g_i$

and since $|g_{i+1}^T g_i| \geq 0.2\|g_{i+1}\|^2$ (Powell restart [13])

$g_{i+1}^T y_i \geq \|g_{i+1}\|^2 - 0.2\|g_{i+1}\|^2$

$g_{i+1}^T y_i \geq 0.8\|g_{i+1}\|^2$

i.e. $\frac{1}{g_{i+1}^T y_i} \leq \frac{1}{0.8\|g_{i+1}\|^2}$

$\therefore |\theta_{i+1}| \leq 1 + \frac{\sigma \alpha L \|d_i\|^2}{0.8\|g_{i+1}\|^2} = c_4$ (22)

use (21) and (22) in (20) then we get :

$$\|d_{i+1}\| = c_4\|g_{i+1}\| + c_3\|d_i\| = c_5$$

$$\frac{1}{\|d_{i+1}\|^2} \geq \left(\frac{1}{c_5}\right)^2 \sum_{i\geq 1} 1 = \infty$$

$\lim_{i\to\infty}\|g_i\| = 0$ (23)

by lemma (3.4), it follows that (19) is true, which for uniformly convex function is equivalent to (23).

5. NUMERICAL RESULTS

Inorder to performance demonstrate the analytical results and effectiveness of such computational algorithm each of these method reached the optimal results. All program arewritten in FORTRAN language by using a set of well known unconstrained optimization test functions these test function are contributed in CUTE[2]. In Table(1) we have compared our new algorithm with Perry method, take the dimension(n=1000,10000) where Table(2)using the total dimension (n=1000,2000,…,10000) for every test function. The comparative performance of the algorithm is evaluated by considering both the (NOF) which is number of function evaluations and the (NOI) which is the number of iterations and all these methods terminate when the following stooping criterion when $\|g_i\| \leq \epsilon$, $\epsilon = 10^{-7}$

Table(1): Numerical Comparisons between the Perry method and New method.

N	Test fu.	Dim	Perry	Newmethod
1	Extended freudentein & roth	1000 10000	13(8) 12(8)	68(60) 12(7)
2	Trigonometric	1000 10000	29(28) 16(15)	29(18) 16(14)
3	Rosenbrock	1000 10000	39(22) 37(21)	34(18) 35(19)
4	White & holst	1000 10000	32(17) 33(19)	35(20) 35(20)
5	Beale	1000 10000	12(7) 12(7)	14(8) 14(8)
6	Penalty	1000 10000	2(2) 2(2)	2(2) 2(2)
7	perturbed quadratic	1000 10000	320(84) 1089(307)	314(89) 1002(274)
8	Raydon 2	1000 10000	4(4) 4(4)	4(4) 4(4)
9	Diagonal 2	1000 10000	183(56) 708(211)	209(74) 535(173)
10	Generalized tridiagonal1	1000 10000	24(5) 24(8)	27(14) 89(75)
11	Extended three exponential terms	1000 10000	10(6) 12(7)	8(4) 8(4)
12	Extended Himmeblau	1000 10000	19(10) 19(10)	10(6) 19(10)

13	Extended maratos	1000 10000	69(34) 66(33)	70(32) 65(30)
14	Extended psc1	1000 10000	6(5) 7(5)	6(5) 7(5)
15	quadratic Diagonal perturbed	1000 10000	132(25) 569(98)	124(21) 508(87)
16	Quadratic function qf1	1000 10000	342(92) 1134(305)	310(86) 1032(280)
17	Extended quadratic penalty QP2	1000 10000	37(20) 42(21)	35(20) 40(20)
18	Nondquar	1000 10000	1647(281) 1344(272)	1685(270) 1120(229)
19	Dixmaan E cute	1000 10000	189(88) 988(616)	232(89) 855(523)
20	Fletcher	1000 10000	124(110) 364(354)	44(31) 276(268)

Table(2): The total Numerical Comparisons between the Perry method and New method.

N	Test fu.	Perry	Newmethod
1	Extended freudentein & roth	629(563)	194(133)
2	Trigonometric	204(168)	203(168)
3	Rosenbrock	386(216)	349(189)
4	White & holst	326(177)	350(200)
5	Beale	120(79)	139(79)
6	Penalty	42(42)	42(42)
7	perturbed quadratic	8042(2204)	7713(2102)
8	Raydon 2	40(40)	40(40)
9	Diagonal 2	4113(1194)	4188(1314)
10	Generalized tridiagonal1	306(149)	341(189)
11	Extended three exponential terms	109(65)	81(42)
12	Extended Himmeblau	190(100)	172(92)
13	Extended maratos	67(50)	67(50)
14	Extended psc1	668(333)	664(303)
15	quadratic Diagonal perturbed	3444(615)	3439(607)
16	Quadratic function qf1	8000(2186)	7705(2157)
17	Extended quadratic penalty QP2	415(209)	408(208)
18	Nondquar	13896(2609)	13866(2436)
19	Dixmaan E cute	5842(3076)	5599(2899)
20	Fletcher	1771(1655)	1347(1240)

CONCLUSION

The research concludes that the proposed new scalar θ_{i+1} of conjugate gradient methods is very effective and increase ability when applied on problem of high dimensional. Under some conditions, we establish that the new proposed methodis globally convergent for uniformly convex.

Table (1) and Table (2), showing that the new algorithm is superior to reach of the optimal solution depend on NOF and NOI.

APPENDIX

The specifics the test functions that used in search can be find in [2]

1- Extended Freudentein and Roth Function .
2- Trigonometric Function .
3- Rosenbrock Function :
4- White &Holst Function .
5- Beale Function .
6- Penalty Function .
7- Perturbed quadratic Function .
8- Raydan2 Function .
9- Diagonal 2Function .
10- Generalization tridiagonal function .
11- Extended three exponential terms .
12- Extended Himmelblau Function .
13- Extended maratos Function .
14- Extended Psc1 Function.
15- Quadrati Diagonal Perturbed Function .
16- Quadrati function QF.
17- Extended quadratic penalty QP2 function .
18- Nondquar Function.
19- Dixmaan E cute Function.
20- Fletcher Function.

REFERENCES

[1]. Andrei, A.(2013), A simple three-term conjugate gradient algorithm for unconstrained optimization , Journal of Computational and Applied Mathematics,Vol.21, PP.419-29.
[2]. Andrei,A. (2008), An Unconstrained Optimization Test Functions Collection ,Advanced Modeling and Optimization , Vol.10 , pp. 147-164.
[3]. Bongratz, K.E. , Conn, A.R. , Gould , N.I and Toint , P.L. , (1995), CUTE: conatrained and unconstrained testing envirouments. ACM Trans- Math. Software , Vol.21, pp.123-160.
[4]. Chen , X. ,and Sun , J. (2002), Global convergence of two parameter family of conjugate gradient methods with out line search, J.Cop. and Applied Math., Vol.146, pp.37-45.
[5]. Dai, H. ,Han, Y. , liu. H., Sun, F., Yin, X., and yuan , Y. (1999), convergence properties of nonlinear conjugate gradient methods, SIAM J. Optim., Vol.10, pp. 348-358.
[6]. Eilaf, S. K.(2018), Global convergence properties of a moodified HS conjugate gradient methods for unconstrained opttimization, Inernational Journal of Enhanced Researchin Science, Technology & Engineering, ISSN:2319-7463, Vol.7,Issue 3.
[7]. Fletcher, R., Revees, C.M. (1964), Function Minimization byconjugate gradients, Computer Journal, Vol.7, pp. 149-154.
[8]. Hestense, M.R., Stiefel, E.L. (1952), Methods of conjugate gradients for solving linear system, Journal of Research of the National Bureau of Standards, Vol. 49, pp.409-436.
[9]. Liu, Y. and Storey, C.(1999),Efficient generalized conjugate gradients algorithms, Part 1: theory Journal of Optimization Theory Applications, Vol. 69, pp.129-137.
[10]. Perry, A.(1978), Amodified conjugate gradient algorithm ,Operation Research. Vol.26, pp.1073-1078
[11]. Perry, J.M. (1977), A class of conjugate gradient algorithms with a two step variable metric memory , discussion paper 269 , Center for Mathematical studies in Economics and Management Science, North Western Universty .
[12]. Polak, E. and Ribiere, G.(1969), Not sur la convergence de direction conjugue'e, Rev. Françoise Informant Research Operational, 3e Anne'e, Vol.16, pp.35-43.
[13]. Powell , M. (1977), Restart procedure for the conjugate gradient methods, mathematics programming , vol. 12 , pp. 241-254.
[14]. Powell , M.J.D. (1976),Reatart procedure for the conjugate gradient method, Math. Prog., Vol.1 , pp. 241-254.
[15]. Wolfe , P. (1971),Convergence conditions for a scent methods (II); some corrections , SIAM Review, Vol.13, pp.185-188.
[16]. Wolfe , P.(1968), Convergence conditions for a scent methods , SIAM Review, Vol. 11, pp.226-235.
[17]. Yabe ,Y. and Sakaiwa, N. (2005), A new nonlinear conjugate gradient methods for unconstrained optimization, Journal Operations Research, Soc. of Japan, Vol.48, , pp.284-296.