International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 3 Issue 4, April-2014, pp: (208-216), Impact Factor: 1.252, Available online at: www.erpublications.com

A Globally Convergent of a New Type Nonlinear Conjugate Gradient Methods

Dr. Basim A. Hassan[1], Dr. Abbas Y. AL-Bayati[2]
[1]Asst. Professor, College of Computers Sciences and Math., University of Mosul, Iraq
[2]Professor, College of Computers Sciences and Math., University of Mosul, Iraq

Abstract: In this paper, a new type conjugate gradient method based on the quadratic model is derived. The new algorithm has proved its worth by posses the descent direction in addition to the global convergence property. These new method are tested on some standard test functions and compared with the original DY method showing considerable improvements over all these comparisons.

Keyword: Unconstrained optimization, conjugate gradient method, descent directions, global convergent methods.

Introduction

For solving large - scale unconstrained optimization problems :

$$\min f(x) \ , \ x \in R^n \qquad ....(1)$$

where $f : R^n \to R$ is a continuously differentiable function, bounded from below, of the most elegant and probably the simplest methods are the conjugate gradient methods. For solving this problem, staring from an initial guess $x_0 \in R^n$, a nonlinear conjugate gradient method, generates a sequence $\{x_k\}$ as :

$$x_{k+1} = x_k + \alpha_k d_k \ , \quad k = 0,1,... \qquad ....(2)$$

where α_k is a positive scalar and called the step-length which is determined by some line search, and d_k are generated as :

$$d_{k+1} = \begin{cases} -g_{k+1} & \text{if } k = 0 \\ -g_{k+1} + \beta_k d_k & \text{if } k > 0 \end{cases} \qquad ....(3)$$

In (3) β_k is known as the conjugate gradient parameter, $v_k = x_{k+1} - x_k$ and g_{k+1} is gradient of f at x_{k+1}. The search direction d_k, assumed to be a descent one, plays the main role in these methods. Different conjugate gradient algorithms correspond to different choices for the scalar parameter β_k. Line search in the conjugate gradient algorithms often is based on the standard Wolfe conditions :

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k \qquad ....(4)$$

$$d_k^T g(x_k + \alpha_k d_k) \ge \sigma d_k^T g_k \qquad ....(5)$$

where d_k is supposed to be a descent direction and with $\delta \in (0, 1/2)$, and $\sigma \in (0, \delta)$. A numerical comparison of conjugate gradient algorithms (2) and (3) with Wolfe line search, for different formulae of parameter β_k computation, including the Dolan and More performance profile, is given in [5]. Then the conjugacy condition

$$d_i^T A d_j = 0 \qquad ....(6)$$

Holds for all $i \neq j$. This relation (8) is the original condition used by Hestenes and Stiefel [4] to derive the conjugate gradient algorithms, mainly for solving symmetric positive definite systems of linear equations. More details can be found in [3].

In [2] conjugate gradient algorithms (2) and (3) with exact line search always satisfy the condition

$$g_{k+1}^T d_{k+1} < 0 \,. \qquad \text{..........(7)}$$

For a general nonlinear twice differential function f, by the mean value theorem, there exists some ξ

$$d_{k+1}^T y_k = \alpha_k d_{k+1}^T \nabla^2 f(x_k + \xi \alpha_k d_k) d_k \,. \qquad \text{..........(8)}$$

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

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

This together (3) lead to the equation

$$\beta_k v_k^T y_k - g_{k+1}^T y_k = 0 \qquad \text{..........(10)}$$

or

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{v_k^T y_k} \,. \qquad \text{..........(11)}$$

More details can be found in [7].

Dai and Yuan [6] proposed the conjugate gradient method with

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k}, \qquad \text{..........(12)}$$

which generated a descent search direction at every iteration and they showed that the proposed method converges globally if the Wolfe condition (4) and (5) are satisfied.

The structure of the paper is as follows. In section (2) extension of the Dai-Yuan Method. In section (3) we present a new conjugate gradient method. Section (4) show that the search direction generated by this proposed algorithms at each iteration satisfies the descent condition and new algorithm. Section (5) establishes the global convergence property for the new CG-methods. Section (6) establishes some numerical results to show the effectiveness of the proposed CG-method and Section (7) gives a brief conclusions and discussions.

Extension of the Dai-Yuan Method

We consider a condition that a descent search direction is generated, and we extend the DY method. Suppose that the current search direction d_k is a descent direction, namely, $g_k^T d_k < 0$ at the k^{th} iteration. Now we need to find a β_k that produces a descent search direction d_{k+1}. This requires that

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_k g_{k+1}^T d_k < 0 \,. \qquad \text{..........(13)}$$

Letting γ_{k+1} be a positive parameter, we define

$$\beta_k = \frac{\|g_{k+1}\|^2}{\gamma_{k+1}} \,. \qquad \text{..........(14)}$$

Equation (13) is equivalent to

$$\gamma_{k+1} > g_{k+1}^T d_k \,. \qquad \text{..........(15}a\text{)}$$

Taking the positively of γ_{k+1} in to consideration, we have

$$\gamma_{k+1} > \max\{g_{k+1}^T d_k,\ 0\}. \qquad \text{..........(15}b\text{)}$$

Therefore if condition $(15b)$ is satisfied for all k, the conjugate gradient method with (14) produces a descent search direction at every iteration. From (14) we can get various kinds of conjugate gradient method by choosing various γ_{k+1}. We note that the Wolfe condition (5) guarantees $d_k^T y_k > 0$ and that $d_k^T y_k = d_k^T g_{k+1} - d_k^T g_k > d_k^T g_{k+1}$. This implies that

$$d_k^T y_k > \max\{g_{k+1}^T d_k,\ 0\}. \qquad \text{..........(15}c\text{)}$$

We make use of the modified secant condition again, and it is written as follows :

$$B_{k+1}v_k = \hat{y}_k = y_k + \frac{\theta_k}{v_k^T u_k} u_k \qquad \text{..........(16)}$$

$$\theta_k = 6(f_k - f_{k+1}) + 3(g_{k+1} + g_k)^T v_k \ .$$

where u_k is any vector with $v_k^T u_k \neq 0$. By multiplying (11) by d_k^T, namely,

$$d_k^T \hat{y}_k = d_k^T y_k + \frac{\theta_k}{v_k^T u_k} d_k^T u_k \, . \qquad \text{..........(17)}$$

The above equation is rewritten as $\gamma_{k+1} = d_k^T \hat{y}_k$, More details can be found in [8].

A new type conjugate gradient method

In this section, we present a concrete formula of γ_{k+1} that satisfies condition $(15b)$ and propose a new conjugate gradient method. By we use the Taylor expansion to second-order terms, f can be written as :

$$f(x_k) = f(x_{k+1}) + \nabla f(x_{k+1})^T (x_k - x_{k+1}) + \frac{1}{2}(x_k - x_{k+1})^T \nabla^2 f(x_{k+1})(x_k - x_{k+1}) \, . \qquad \text{..........(18)}$$

Hence

$$f(x_k) = f(x_{k+1}) + g_{k+1}^T v_k + \frac{1}{2} v_k^T \nabla^2 f(x_{k+1}) v_k \ . \qquad \text{..........(19)}$$

Therefore

$$\begin{aligned} v_k^T \nabla^2 f(x_{k+1}) v_k &= 2[f(x_k) - f(x_{k+1})] + 2g_{k+1}^T v_k \\ &= 2[f(x_k) - f(x_{k+1})] + [g_{k+1} + g_k]^T v_k + v_k^T y_k \end{aligned} \qquad \text{..........(20)}$$

From (2) and (20) we get

$$\begin{aligned} \alpha_k d_k^T \nabla^2 f(x_{k+1}) v_k &= \{2[f(x_k) - f(x_{k+1})] + [g_{k+1} + g_k]^T v_k + v_k^T y_k\} \\ d_k^T \nabla^2 f(x_{k+1}) v_k &= \frac{1}{\alpha_k}\{2[f(x_k) - f(x_{k+1})] + [g_{k+1} + g_k]^T v_k + v_k^T y_k\} . \end{aligned} \qquad \text{..........(21)}$$

Therefore, we can define a new denominator nonlinear conjugate gradient methods as follows :

$$\gamma_{k+1} = \frac{1}{\alpha_k}\{2[f(x_k) - f(x_{k+1})] + [g_{k+1} + g_k]^T v_k + v_k^T y_k\} . \qquad \text{..........(22)}$$

This implies that $\alpha_k \gamma_{k+1}$ has a good information of the second order curvature $v_k^T \nabla^2 f(x_{k+1}) v_k$. From this we further define the formula,

$$\beta_k^{New1} = \frac{g_{k+1}^T g_{k+1}}{\gamma_{k+1}} \ . \qquad \text{..........(23)}$$

From (23), we can get several kind of conjugate gradient methods by choosing various γ_{k+1} . The case $\gamma_{k+1} = d_k^T y_k$, formula reduces to the Dai and Yuan (DY) [6].

To create algorithms that has global convergence properties, we modify the above formulas as follow :

$$\beta_k^{New2} = \frac{\|g_{k+1}\|^2}{\max\ (d_k^T y_k \ , \gamma_{k+1})} . \qquad \text{..........(24)}$$

Generate the search direction by

$$d_{k+1} = -g_{k+1} + \frac{\|g_{k+1}\|^2}{\max\ (d_k^T y_k \ , \gamma_{k+1})} d_k \, . \qquad \text{..........(25)}$$

New Algorithm and descent property

In this section, we give the specific form of the proposed conjugate gradient method as follows.

Algorithm 4.1 (The new method)

Step 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 $\alpha_1 = 1/\|g_1\|$.

Step 2. Test for continuation of iterations. If $\|g_{k+1}\| \leq 10^{-6}$, then stop.

Step 3. Line search. Compute $\alpha_{k+1} > 0$ satisfying the Wolfe line search condition (4) and (5) and update the variables $x_{k+1} = x_k + \alpha_k d_k$.

Step 4. β_k conjugate gradient parameter which defined in (23) and (24).

Step 5. Direction computation. Compute $d_{k+1} = -(g_{k+1} + \beta_k d_k)$. If the restart criterion of Powell $|g_{k+1}^T g_k| \geq 0.2\ \|g_{k+1}\|^2$, is satisfied, then set $d_{k+1} = -g_{k+1}$ otherwise define $d_{k+1} = d$. Set $k = k+1$ and continue with step2 .

Now we have the following theorem, which illustrate that the scaled conjugate gradient method can guarantee the descent property with the Wolfe line searches.

Theorem (4.1) :

Suppose that α_k in (2) satisfies the Wolfe conditions (3) and (4), then the direction d_{k+1} given by (27) is a descent direction.

Proof :

Since $d_0 = -g_0$, we have $g_0^T d_0 = -\|g_0\|^2 \leq 0$. Suppose that $g_k^T d_k < 0$ for some $k \in n$. Inequality (4) ensures that

$$\begin{aligned}
\gamma_{k+1} &= \max\ (d_k^T y_k\ \ , \gamma_{k+1}) \\
&\geq \frac{1}{\alpha_k}\left\{2[f(x_k) - f(x_{k+1})] + [g_{k+1} + g_k]^T v_k + v_k^T y_k\right\} \\
&= \frac{1}{\alpha_k}\left\{2[f(x_k) - f(x_{k+1})] + 2g_{k+1}^T v_k\right\} \qquad \text{..........(26)} \\
&\geq \frac{1}{\alpha_k}\left\{-2\delta g_k^T v_k + 2\sigma g_k^T v_k\right\} \\
&\geq \frac{1}{\alpha_k}\left\{-2(\delta - \sigma) g_k^T v_k\right\}\ > 0\ .
\end{aligned}$$

Multiplying (3) by g_{k+1}^T, with (25) we have

$$g_{k+1}^T d_{k+1} = \left[-g_{k+1}^T g_{k+1} + \frac{\|g_{k+1}\|^2}{\max\ (d_k^T y_k\ \ , \gamma_{k+1})}\ g_{k+1}^T d_k \right] \qquad \text{..........(27)}$$

If $g_{k+1}^T d_k \leq 0$, we get :

$$\begin{aligned}
g_{k+1}^T d_{k+1} &= g_{k+1}^T(-g_{k+1} + \beta_k g_{k+1}^T d_k) \qquad \text{..........(28)} \\
&\leq -\|g_{k+1}\|^2\ < 0\ .
\end{aligned}$$

If $g_{k+1}^T d_k > 0$, by $\gamma_{k+1} \geq d_k^T y_k = g_{k+1}^T d_k - g_k^T d_k$ and $g_k^T d_k < 0$, we find that $\gamma_{k+1} \geq g_{k+1}^T d_k$, and hence, $g_{k+1}^T d_k / \gamma_{k+1} < 1$. So, we have :

$$\begin{aligned} g_{k+1}^T d_{k+1} &= g_{k+1}^T(-g_{k+1} + \beta_k g_{k+1}^T d_k) \\ &\leq -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\gamma_{k+1}} g_{k+1}^T d_k \\ &\leq \left[-1 + \frac{g_{k+1}^T d_k}{\gamma_{k+1}}\right] \|g_{k+1}\|^2 \\ &< [-1+1]\|g_{k+1}\|^2 = 0. \end{aligned} \qquad \text{..........(29)}$$

Convergence analysis for strongly convex functions

In [6] throughout this section we assume that f is strongly convex and Lipschitz continuous on the level set

$$L_0 = \{x \in \mathrm{R}^n : f(x) \leq f(x_0)\}. \qquad \text{..........(30)}$$

Assumption A :

This is, there exists constants $\mu > 0$ and L such that

$$(\nabla f(x) - \nabla f(y))^T (x-y) \geq \mu \|x-y\|^2 \qquad \text{..........(31)}$$

and

$$(\nabla f(x) - \nabla f(y)) \leq L \|x-y\|^2 \qquad \text{..........(32)}$$

for all x and y from L_0.

On the other hand, under Assumption, It is clear that there exist positive constants B such

$$\|x\| \leq B \ , \forall x \in S \qquad \text{..........(33)}$$

Proposition :

Under Assumption A and equation (33) on f, there exists a constant $\bar{\gamma} > 0$ such that

$$\|\nabla f(x)\| \leq \bar{\gamma} \ , \ \forall x \in S \qquad \text{..........(34)}$$

Dai et al. [6] proved that for any conjugate gradient method with Wolfe line search the following general result holds :

Lemma (5.1)

Suppose that the assumptions A hold and consider any conjugate gradient method (2) and (3), where d_{k+1} is a descent direction and α_k is obtained by the strong Wolfe line search (4) and (5). If

$$\sum_{k \geq 0} \frac{1}{\|d_{k+1}\|^2} = \infty \ , \qquad \text{..........(35)}$$

then

$$\liminf_{k \to \infty} \|g_{k+1}\| = 0 \ . \qquad \text{..........(36)}$$

As we know, if f is a uniformly convex functions, then there exists a constant $\mu > 0$ such that

$$(\nabla f(x) - \nabla f(y))^T (x-y) \geq \mu \|x-y\|^2 \qquad \text{..........(37)}$$

Equivalently, this can be expressed as

$$f(x) \geq f(y) + \nabla f(y)^T (x-y) + \frac{\mu}{2} \|x-y\|^2 \qquad \text{..........(38)}$$

From (37) and (38) it follows that

$$y_k^T v_k \geq \mu \|v_k\|^2,$$
$$f_k - f_{k+1} \geq -g_{k+1}^T v_k + \frac{\mu}{2}\|v_k\|^2. \quad \text{..........(39)}$$

Obviously, from (35) we get :

$$\mu\|v_k\|^2 \leq y_k^T v_k \leq L\|v_k\|^2, \;\; i.e. \; \mu \leq L. \quad \text{..........(40)}$$

for any $x, y \in S$, More details can be found in [2,3].

Theorem (5.1) :

Suppose that the assumptions A hold and descent condition (29) hold. Consider the CG method in the form of (25), where α_k is computed using the Wolfe line search (4) and (5). If the objective function f is uniformly

$$\lim_{k\to\infty} \inf \|g_{k+1}\| = 0. \quad \text{..........(41)}$$

Proof :

For the of contradiction, we suppose that the conclusion is not true. Suppose that $g_k \neq 0$ for all k.

From (26) and (39) we get :

$$\begin{aligned}
\gamma_{k+1} &= \frac{1}{\alpha_k}\left\{2[f(x_k) - f(x_{k+1})] + [g_{k+1} + g_k]^T v_k + v_k^T y_k\right\} \\
&\geq \frac{1}{\alpha_k}\left\{2(-g_{k+1}^T v_k + \frac{\mu}{2}\|v_k\|^2) + g_{k+1}^T v_k + g_k^T v_k + y_k^T v_k\right\} \\
&= \frac{1}{\alpha_k}\left\{\mu\|v_k\|^2 - [g_{k+1} - g_k]^T v_k + y_k^T v_k\right\} \\
&= \frac{1}{\alpha_k}\left\{\mu\|v_k\|^2\right\}
\end{aligned} \quad \text{..........(42)}$$

On the other hand, suppose that the gradient is Lipschitz continuous. Then we have that

$$\|y_k\| = \|g_{k+1} - g_k\| \leq L\|v_k\|^2. \quad \text{..........(43)}$$

Hence,

$$\left|g_{k+1}^T y_k\right| \leq \|g_{k+1}\|\|y_k\| \leq L\|g_{k+1}\|\|v_k\| \quad \text{..........(44)}$$

From (25) and (44) we get :

$$\begin{aligned}
\|d_{k+1}\| &= \left\|-g_{k+1} + \frac{g_{k+1}^T y_k}{\gamma_{k+1}} d_k\right\| \\
&\leq \|g_{k+1}\| + \frac{\left|g_{k+1}^T y_k\right|}{\left|\gamma_{k+1}\right|}\|d_k\| \\
&\leq \bar{\gamma} + \frac{\bar{\gamma}\, L\|v_k\|}{\frac{1}{\alpha_k}\mu L\|v_k\|^2}\|d_k\| \\
&\leq \bar{\gamma} + \frac{\bar{\gamma}\, L\|v_k\|}{\mu L\|v_k\|^2}\alpha_k\|d_k\| \\
&\leq \bar{\gamma}\left[1 + \frac{1}{\mu}\right]
\end{aligned} \quad \text{..........(45)}$$

This relation implies

$$\sum_{k\geq 1}\frac{1}{\|d_{k+1}\|^2} \geq \left[\frac{\mu}{\mu+L}\right]^2 \frac{1}{\bar{\gamma}^2}\sum_{k\geq 1} 1 = \infty \qquad \text{..........(46)}$$

Therefore, from Lemma (3.1) we have.

$$\lim_{k\to\infty} \inf \|g_{k+1}\| = 0 \qquad \text{..........(47)}$$

Numerical Results

This section provides the result of some numerical experiments. A comparison was made between new conjugate gradient methods and the DY-Algorithm. Both algorithms are implemented in Fortran. Test

problems are from [1]. We have selected (15) large scale unconstrained optimization problems in extended or generalized form, for each test function , we have considered numerical experiments with the number of variable $n = 100,1000$. We use the inequality $\|g_{k+1}\| \leq 10^{-6}$ as the termination condition. The test results for both algorithms with the Wolfe line search $\delta_1 = 0.001$ and $\delta_2 = 0.9$ was obtained and summarized in Tables (6.1) and (6.2), where each column has the following meanings :

Problem : the name of the problem.
Dim : the dimension of the problem.
TNOI : the total number of iterations.
TIRS : the total number of restart.
F* : if NOI exceeded 2000.

Table (6.1) : Comparison of methods for n= 100

Test problems	β_k^{DY}		β_k^{New1}		β_k^{New2}	
	NOI	IRS	NOI	IRS	NOI	IRS
Extended Rosenbrock (CUTE)	40	81	39	79	38	77
Penalty	11	28	9	25	9	25
Perturbed Quadratic	83	125	83	127	83	125
Diagonal 2	61	105	64	107	64	105
Extended Tridiagonal 1	10	21	11	23	10	21
Generalized Tridiagonal 2	40	61	41	64	37	61
Extended Powell	79	151	79	151	64	119
Extended Wood (CUTE)	31	59	22	43	25	48
Quadratic QF2	111	173	108	168	109	175
DIXMAANE (CUTE)	85	133	78	121	85	133
STAIRCASE S1	518	821	481	756	405	639
ENGVAL1 (CUTE)	F	F	27	51	26	52
Extended Block-Diagonal BD2	12	23	10	19	12	23
Generalized quartic GQ2	35	60	34	54	36	60
Hager	27	45	29	48	29	48
Total	1143	1886	1088	1785	1006	1659

Table (6.2) : Comparison of methods for n= 1000

Test problems	β_k^{DY}		β_k^{New1}		β_k^{New2}	
	NOI	IRS	NOI	IRS	NOI	IRS
Extended Rosenbrock (CUTE)	39	85	34	73	30	65
Penalty	23	49	13	39	23	54
Perturbed Quadratic	393	629	297	469	333	527
Diagonal 2	201	329	208	341	198	318
Extended Tridiagonal 1	15	29	15	29	15	29
Generalized Tridiagonal 2	64	101	65	100	56	87
Extended Powell	85	156	83	160	87	163
Extended Wood (CUTE)	26	51	32	61	24	48
Quadratic QF2	493	767	340	538	351	557
DIXMAANE (CUTE)	243	381	220	349	231	357
STAIRCASE S1	F	F	F	F	F	F
ENGVAL1 (CUTE)	F	F	51	955	71	1595
Extended Block-Diagonal BD2	12	23	12	23	12	23
Generalized quartic GQ2	35	55	30	53	36	55
Hager	F	F	F	F	118	1947
Total	1629	2655	1349	2235	1396	2283

Conclusions and Discussions

In this paper, we have proposed a new conjugate gradient- algorithms based on the Taylor expansion to second-order terms, respectively under some assumptions the a new algorithms have been shown to be globally convergent for uniformly convex and satisfies the sufficient descent property. The computational experiments show that the a new CG kinds given in this paper are successful .Table (7.1) gives a comparison between the new-algorithms and the Dai and Yuan (DY) algorithm for convex optimization, this table indicates, see Table (7.1), that the new algorithm saves $(86.61-87.91)\%$ NOI and $(86.80-88.52)\%$ IRS overall against the standard Dai and Yuan (DY) algorithm, especially for our selected group of test problems.

Table (7.1) : Relative efficiency of the new Algorithm $(n=100,1000)$

Tools	NOI	IRS
DY- Algorithm	100 %	100 %
New Algorithm with β_k^{New1}	12.08 %	11.47 %
New Algorithm with β_k^{New2}	13.38 %	13.19 %

secnerefeR

[1]. Andrei N. (2008), An Unconstrained Optimization test function collection. Adv. Model. Optimization . 10. pp.147-161.
[2]. Andrei N. (2007), Numerical comparison of conjugate gradient algorithms for unconstrained optimization. Studies in Informatics and Control, 16, pp.333-352.
[3]. Andrei N. (2008), An accelerated conjugate gradient algorithm with guaranteed descent and conjugacy conditions for unconstrained optimization. Technical Report, pp. 1-18.

[4]. Hestenes, M. R. and Stiefel, E. L. (1952), ' Method of conjugate gradients for Solving linear systems ' Journal National Standards 49, pp. 409-436.
[5]. Dolan E. and More J..(2002),' Benchmarking optimization software with performance profiles ' Math. Programming 91, pp. 201-213.
[6]. Dai, Y.H., Yuan, Y.(1999), A nonlinear conjugate gradient method with a strong global convergence property. SIAM J. Optim. 10, 177–182.
[7]. Matonoha C., Luksan L. and Vlcek J. (2008) 'Computational experience with conjugate gradient methods for unconstrained optimization' Technical Report, No.1038, pp. 1-17.
[8]. Yabe, H., Sakaiwa, N.: A new nonlinear conjugate gradient method for unconstrained optimization. J. Oper. Res. Soc. Japan 48, 284–296 (2005)