Indonesian Journal of Electrical Engineering and Computer Science
Vol. 16, No. 2, November 2019, pp. 661~668
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v16.i2.pp661-668

A descent extension of the Dai - Yuan conjugate gradient technique

Basim A. Hassan[1], Zeyad M. Abdullah[2], Hawraz N. Jabbar[3]
[1]Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, Iraq
[2]Faculty of Computers Sciences and Mathematics, University of Tikrit, Iraq
[3]Department of Mathematics, College of Sciences, University of Kirkuk, Iraq

Article Info

Article history:

Received Jan 21, 2019
Revised May 1, 2019
Accepted May 22, 2019

Keywords:

Conjugate coefficient
Conjugate gradient method
Global convergence property

ABSTRACT

The conjugate coefficient is the very establishment of an variety of the conjugate gradient methods for solving unconstrained optimization problems. Based on Dai-Yuan technique, a new extension method was proposed. In contrary to the CG classical method, this proposed method employs a coefficient that has an important role in constructing the new method with less computational efforts. This proposed method was found to be efficient on the basis of theory analysis and also numerical results to show the efficiency of the proposed method.

Corresponding Author:

Basim A. Hassan,
Department of Mathematics,
College of Computers Sciences and Mathematics,
University of Mosul, Iraq.
Email: basimabas39@gmail.com

1. INTRODUCTION

The technique of conjugate gradient is useful in finding the problem's minimum value. Consider this formula:

$$Min \quad f(x) \ , \ \mathrm{x} \in \mathrm{R}^{\mathrm{n}} \tag{1}$$

where $f : R^n \to R$ is smooth [1]. The iteration technique is often used to solve Equation (1) and it is written as:

$$x_{k+1} = x_k + \alpha_k d_k \ , \tag{2}$$

where α_k is the step length. In any complete conjugate gradient method, the step length α_k is regularly chosen to satisfy the firm line search conditions. As the most eminent line search conditions, Wolfe line search necessitates that:

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k \tag{3}$$

$$d_k^T g(x_k + \alpha_k d_k) \geq \sigma\, d_k^T g_k \tag{4}$$

where $0 < \delta < \sigma$. More details are found in [2]. Therefore, this study will focus on the CG method whose direction of search is written as:

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \tag{5}$$

β_k is a scalar known as the CG update and g_{k+1} is gradient. Different CG techniques chiefly convert into varied selections for the CG update, as reviewed in [3].

In convergence properties, the conjugate gradient method proposed by Dai and Yuan [4] is generally believed to be one of the best CG techniques. The CG update of the Dai and Yuan (DY) technique is given by:

$$\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{d_k^T y_k} \tag{6}$$

where $y_k = g_{k+1} - g_k$.

We utter that the search directions d_{k+1} satisfy the sufficient descent condition if:

$$g_{k+1}^T d_{k+1} \leq -c \left\| g_{k+1} \right\|^2 \tag{7}$$

where c is a positive constant. More details on performance profile are given in [5]. Now we make extension of the Dai - Yuan conjugate gradient technique and analyze its convergence.

2. EXTENSION OF THE DAI - YUAN CONJUGATE GRADIENT TECHNIQUE

Attempting to make an extension on the CG update suggested by Dai and Yuan through multiplying (5) with g_{k+1} and using (6), this equation is obtained:

$$g_{k+1}^T d_{k+1} = \frac{g_{k+1}^T g_{k+1}}{d_k^T y_k} g_k^T d_k \tag{8}$$

which with (6) gives an equivalent formula to (6):

$$\beta_k = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k} \tag{9}$$

For more details can be found in [6]. Obviously, if the Hessian matrix of f is non negative definite, the majority efficient search direction at x_k will be the Newton direction:

$$d_{k+1} = -\nabla^2 f(x_{k+1})^{-1}\ g_{k+1} \tag{10}$$

The Hessian matrices satisfy the secant equation:

$$\nabla^2 f(x_{k+1})\ s_k = y_k \tag{11}$$

where $s_k = x_{k+1} - x_k$.

Suppose that I denotes the identical matrix and assume that a matrix $\nabla^2 f(x_{k+1})$ is desired with a simplified construction satisfying the secant equation. Specifically, this formula is desired:

After some algebraic manipulations one obtains :

$$\nabla^2 f(x_{k+1}) = \eta_{k+1} I_{k+1} \tag{12}$$

So, the equation of secant is formed as follows:

$$\eta_{k+1}\ s_k = y_k \tag{13}$$

This equation describes the most popular spectral gradient method for optimization with the direction of search:

$$d_{k+1} = -\frac{1}{\eta_{k+1}}\ g_{k+1} \tag{14}$$

where d_{k+1}, which is defined in (14), would be replaced by d_{k+1} in (8), the following equation is obtained:

$$\beta_k = -\frac{1}{\eta_{k+1}} \frac{g_{k+1}^T g_{k+1}}{g_k^T d_k} \tag{15}$$

If a step - length α_k is inexact, then:

$$\beta_k = \frac{1}{\eta_{k+1}} \frac{g_{k+1}^T g_{k+1}}{y_k^T d_k} \tag{16}$$

Denoted this formula procedure as β_k^{BDY}. Definitely, more sufficient test would be required η_{k+1}. The coefficient η_{k+1} plays an important role in the current method. It can be computed by the formula:

$$1/\eta_{k+1} = 2\left[f_k - f_{k+1} + g_{k+1}^T s_k\right]/\ y_k^T s_k \tag{17}$$

This coefficient defines the most popular Yuan method [7]. The Extension algorithm is presented below and called Algorithm BDY.

Algorithm BDY

0 : Given constants $\varepsilon > 0,\ \delta \in (0,1),\ \sigma \in (\delta,1)$ and $x_0 \in R^n$. Let $d_0 = -g_0,\ k = 0$.

1 : Examine a stopping criterion. If yes, then stop; if not, continue with the next step.

2 : Define a step length α_k using a proper line search.

3 : Let $x_{k+1} = x_k + \alpha_k d_k$, and compute β_k^{BDY} by (16).

4 : Compute the search direction $d_{k+1} = -g_{k+1} + \beta_k d_k$.

5 : Set $k = k+1$ and continue with step 1.

3. CONVERGENCE ANALYSIS

In this paper, $f(x)$ is assumed as that :

1. For a given $x_0 \in R^n$, the function $f(x)$ has lower bound on $\Psi = \{x \in R^n : f(x) \le f(x_0)\}$.

2. A positive constant L exists; so:

$$\|g(x)-g(y)\| \le L\,\|x-y\|\,,\ \forall\ x,\ y\in\Phi \tag{18}$$

where $\Phi \in \Psi$ is an open convex set. That is to say, the gradient $g(x)$ is Lipschitz continuous in Φ. More details are found in [8, 9].

Theorem 1

Suppose that the search direction d_{k+1} is created by (5), and $\eta_{k+1} \le 1$. Then the descent condition holds, i.e.,

$$g_{k+1}^T d_{k+1} \le 0 \tag{19}$$

Proof:

This theorem is proved by induction. As $d_0 = -g_0$, there is $g_0^T d_0 = -\|g_0\|^2 < 0$. Assume that $g_k^T d_k < 0$ for all $k \in n$. Through multiplying (5) by g_{k+1}, the following equation is obtained:

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \eta_{k+1}\frac{\|g_{k+1}\|^2}{y_k^T d_k} g_{k+1}^T d_k \\ &= \|g_{k+1}\|^2\left[-1+\eta_{k+1}\frac{g_{k+1}^T d_k}{y_k^T d_k}\right] \\ &= \|g_{k+1}\|^2\left[\frac{\eta_{k+1} g_{k+1}^T d_k - y_k^T d_k}{y_k^T d_k}\right] \end{aligned} \tag{20}$$

Since $y_k^T d_k = g_{k+1}^T d_k - g_k^T d_k$, then :

$$\begin{aligned} g_{k+1}^T d_{k+1} &= \|g_{k+1}\|^2\left[\frac{\eta_{k+1} g_{k+1}^T d_k - g_{k+1}^T d_k + g_k^T d_k}{y_k^T d_k}\right] \\ &= \|g_{k+1}\|^2\left[\frac{(\eta_{k+1}-1)\, g_{k+1}^T d_k + g_k^T d_k}{y_k^T d_k}\right] \\ &= \frac{\|g_{k+1}\|^2}{y_k^T d_k}\left[(\eta_{k+1}-1)\, g_{k+1}^T d_k + g_k^T d_k\right] \end{aligned} \tag{21}$$

From above equation we get:

$$\frac{\|g_{k+1}\|^2}{y_k^T d_k} = \frac{g_{k+1}^T d_{k+1}}{\left[(\eta_{k+1}-1)\, g_{k+1}^T d_k + g_k^T d_k\right]} = \beta_k^{DY} \tag{22}$$

Now,

$$\begin{aligned} \beta_k^{\mathrm{BDY}} &= \eta_{k+1}\frac{\|g_{k+1}\|^2}{y_k^T d_k} = \eta_{k+1}\,\beta_k^{DY} \\ &= \frac{\eta_{k+1} g_{k+1}^T d_{k+1}}{\left[(\eta_{k+1}-1)\, g_{k+1}^T d_k + g_k^T d_k\right]} \end{aligned} \tag{23}$$

From above equation we get :

$$\beta_k^{\mathrm{BDY}} = \omega_{k+1} \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k} \tag{24}$$

where $\omega_{k+1} = \eta_{k+1} /(1+(\eta_{k+1}-1)l_k)$ and $l_k = g_{k+1}^T d_k / g_k^T d_k$. Assume that $g_k^T d_k < 0$. From l_k and Wolfe state $g_{k+1}^T d_k \geq \sigma g_k^T d_k$ we get :

$$l_k \leq \sigma \quad \rightarrow \quad l_k \leq \sigma < 1 \tag{25}$$

and

$$\omega_{k+1} \leq \eta_{k+1} /(1+(\eta_{k+1}-1)\sigma) \tag{26}$$

where $\eta_{k+1} /(1+(\eta_{k+1}-1)\sigma > 0$. The above equation to get :

$$g_{k+1}^T d_{k+1} = \frac{1}{\omega_{k+1}} \beta_k^{\mathrm{BDY}} g_k^T d_k \leq 0 \tag{27}$$

Therefore, the proof is complete.

Due to playing an important role in analyzing the convergence property for conjugate gradient, Zoutendijk's condition [10] will be proved to be a part of the proposed Algorithm in this study.

Lemma 1

Let that d_{k+1} is generated by (5) and step size α_k fulfills (3) and (4), if $f(x)$ satisfies the Assumptions, then :

$$\sum_{k=1}^{\infty} \frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} < \infty \tag{28}$$

holds.

Theorem 2

Let that Assumptions 1 and 2 are held, and that the search direction d_{k+1} is calculated by the formulation in the second step, then:

$$\lim_{k\to\infty} \inf \|g_{k+1}\| = 0 \tag{29}$$

Proof:

Now, equation (29) can be proved by contradiction, assuming that there is a constant $\gamma > 0$ such that :

$$\|g_{k+1}\| \geq \gamma \,, \ \forall \ k \geq 0 \tag{30}$$

Since $d_{k+1} = -g_{k+1} + \beta_k d_k$, then the result is the following equation:

$$\|d_{k+1}\|^2 = (\beta_k)^2 \|d_k\|^2 - 2d_{k+1}^T g_{k+1} - \|g_{k+1}\|^2 \tag{31}$$

Dividing both sides of (31) by $(d_{k+1}^T g_{k+1})^2$ and using (24), the following formula is obtained:

$$\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} \le \omega_{k+1} \frac{\|d_k\|^2}{(d_k^T g_k)^2} - \frac{2}{(d_{k+1}^T g_{k+1})} - \frac{\|g_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2}$$

$$\le \omega_{k+1} \frac{\|d_k\|^2}{(d_k^T g_k)^2} - \left(\frac{\|g_{k+1}\|}{d_{k+1}^T g_{k+1}} + \frac{1}{\|g_{k+1}\|} \right)^2 + \frac{1}{\|g_{k+1}\|^2} \tag{32}$$

$$\le \omega_{k+1} \frac{\|d_k\|^2}{(d_k^T g_k)^2} + \frac{1}{\|g_{k+1}\|^2}$$

which, along with (26), can yield:

$$|\ \omega_{k+1}| \le 1 \tag{33}$$

Using (32) and (33), the following equation can be obtained:

$$\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} \le \frac{\|d_k\|^2}{(d_k^T g_k)^2} + \frac{1}{\|g_{k+1}\|^2} \tag{34}$$

Using (34) recursively and noting that $\|d_1\|^2 = -d_1^T g_1 = \|g_1\|^2$,

$$\frac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} \le \sum_{i=1}^{k+1} \frac{1}{\|g_{i+1}\|^2} \le \frac{k}{\gamma} \tag{35}$$

Then, from this equation and (30), the following formula can be derived:

$$\frac{(d_{k+1}^T g_{k+1})^2}{\|d_{k+1}\|^2} \ge \frac{\gamma}{k}, \tag{36}$$

and this denotes that:

$$\sum_{k=1}^{\infty} \frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} = \infty \tag{37}$$

This is in contrary to the Zoutendijk condition (28). Thus, (29) holds.

4. NUMERICAL RESULTS AND DISCUSSION

In this paper, the algorithm of extension was tested and compared to the method of Dai-Yuan [4]. For comparing methods employed in this study with other classical ones, the gradient errors will be checked for measuring the criteria of stop for algorithms. In general, the iteration will be forced to stop when the gradient norm satisfies this inequality:

$$\|g_{k+1}\| < 10^{-6} \tag{38}$$

By employing the process of Wolfe line search, the parameters will be selected as follows:

$$\delta_1 = 0.001 \text{ and } \delta_2 = 0.9. \quad (39)$$

In this study, experiments were conducted on 30 problems of unconstrained optimization test of Andrei's collection with the dimensions 100 and 1000; for details see Andrei [11]. Optimization problems is an important tool in all papers optimization [12-15]. The comparing data contain number of iterations (IN), the number of restart (NR) and the number of function evaluations (NF) (Table 1). Using Fortran 90 to code this methods.

Table 1. Comparing different conjugate gradient methods with different test functions

P. No.	n	DY algorithm			BDY algorithm		
		NI	NR	NF	NI	NR	NF
1	100	27	8	14	26	7	13
	1000	26	8	13	25	7	13
2	100	28	7	11	25	5	9
	1000	49	15	23	48	14	22
3	100	125	29	83	125	29	83
	1000	629	106	393	550	90	342
4	100	21	5	10	13	3	6
	1000	29	7	15	26	7	13
5	100	61	15	40	61	11	37
	1000	101	27	64	79	22	49
6	100	19	6	10	19	6	10
	1000	35	12	22	19	6	10
7	100	151	26	79	137	24	73
	1000	156	28	85	153	25	81
8	100	28	8	10	28	8	11
	1000	33	10	15	29	11	13
9	100	59	12	31	47	10	25
	1000	51	10	26	54	11	28
10	100	173	37	111	163	33	105
	1000	767	160	493	583	117	373
11	100	133	26	85	126	25	83
	1000	381	71	243	347	69	229
12	100	136	24	87	136	24	87
	1000	421	58	250	400	63	241
13	100	47	11	30	53	4	29
	1000	62	15	37	62	14	36
14	100	33	9	17	40	11	20
	1000	48	12	22	35	9	16
15	100	60	11	35	56	12	36
	1000	55	14	35	54	8	31
Total		**3944**	**787**	**2389**	**3519**	**684**	**2077**

Problems numbers indicant for : "1. is the Extended Beale, 2. is the Penalty, 3. is the Perturbed Quadratic, 4. is the Extended Tridiagonal 1, 5. is the Generalized Tridiagonal 2, 6. is the Extended Himmelblau, 7. is the Extended Powell, 8. is Extended Cliff, 9. is the Extended Wood, 10. is the Quadratic QF2, 11. is the DIXMAANE (CUTE), 12. is the Partial Perturbed Quadratic, 13. is the Broyden Tridiagonal, 14. is the LIARWHD (CUTE), 15. is the Generalized quartic GQ2.".

The summary of our results in Table 2. The result presented in Table 2 imply that BDY method improved over the performance of DY method. The improvement of method BDY over DY is 13 %, in average, of the iterations number, and 13 %, in average, of the restart number and 10 %, in average, of function evaluations number.

Table 2. Ratio of algorithm BDY cost to DY cost

	NI	NR	NF
DY algorithm	100 %	100 %	100 %
BDY algorithm	86. 94 %	86. 91 %	89. 22 %

5. CONCLUSION

An extended method of nonlinear DY conjugate gradient was proposed in this paper. The extended direction refers to the descent direction. Analysis of convergence analysis was carried out and the statistics showed that the proposed technique is efficient for the given problems of nonlinear optimization test.

REFERENCES

[1] Ahmad A. and Zabidin S., “Modification of Nonlinear Conjugate Gradient Method with Weak Wolfe-Powell Line Search,” *Hindawi Abstract and Applied Analysis*, pp. 1-6, 2017.

[2] Nocedal J. and Yuan Y., “Analysis of a self-scaling quasi-Newton method,” *J. Mathematical programming*, vol. 61, pp. 19-37, 1993.

[3] Hager W. W. and Zhang H., “A survey of nonlinear conjugate gradient methods,” *Pac. J. Optim.*, vol. 2, pp. 35-58, 2006.

[4] Y. H. Dai and Yuan Y., “A Nonlinear Conjugate Gradient with a Strong Global Convergence Property,” *SIAM Journal of Optimization*, vol. 10, pp. 177-182, 2000.

[5] Babaie S. K., “An eigen value study on the sufficient descent property of a modified Polak-Ribi- Polak conjugate gradient method,” *Bull. Iranian Math. Soc*., vol. 40, pp. 235-242, 2014.

[6] Y. H. Dai and Yuan Y., “An Efficient Hybrid Conjugate Gradient Method for Unconstrained Optimization,” *Annals of Operations Research*, vol. 103, pp. 33-47, 2001.

[7] Y. Yuan, “A modified BFGS algorithm for unconstrained optimization,” *IMA Journal Numerical Analysis*, vol. 11, pp. 325-332, 1991.

[8] Guangming Z., “A descent algorithm without line search for unconstrained optimization,” *Applied Mathematics and Computation*, vol. 215, pp. 2528-2533, 2009.

[9] Basim A. H., “A new formula for conjugate parameter computation based on the quadratic model,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 3, pp. 954-961, 2019.

[10] Zoutendijk G., “Nonlinear programming computational methods,” *Integer and Nonlinear Programming, J. Abadie, ed, North-Holland, Amsterdam*, pp. 37-86, 1970.

[11] Andrie N., “An Unconstrained Optimization Test functions collection,” *Advanced Modeling and optimization*, vol. 10, pp. 147-161, 2008.

[12] Li C., “A modified conjugate gradent method for unconstrained optimization,” *TELKOMNIKA Telecommunication, Computing, Electronics and Control*”, vol. 11, pp. 6373-6380, 2013.

[13] Basim A. H. and Hussein K. K., “A new class of BFGS updating formula based on the new quasi-newton equation,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 3, pp. 945-953, 2019.

[14] Ghani N., Kamaruddin S., Ramli M., Musirin I. And Hashim H. “Enhanced BFGS quasi-Newton backpropagation models on MCCI data”. *Indonesian Journal of Electrical Engineering and Computer Science.* 2017;8:101-106.

[15] Li C. Fang L. and Cao X. “Global convergenceof a kind of conjugate gradient method”. *TELKOMNIKA Telecommunication, Computing, Electronics and Control*. 2013; 11: 544-549,2013.