New Proposed Conjugate Gradient Method for Nonlinear Unconstrained Optimization

Salah **G**azi **Sh**areef and **D**lovan **H**aji **O**mar
Department of Mathematics, Faculty of Science, University of Zakho, Kurdistan Region – Iraq.
(Accepted for publication: December 25, 2016)

Abstract:

In this paper, we suggest a new conjugate gradient method for unconstrained optimization by using homotopy theory. Our suggestion algorithm satisfies the conjugacy and descent conditions. Numerical result shows that our new algorithm is better than the standard CG algorithm with respect to the NOI and NOF.
Keywords*: Unconstrained Optimization, Conjugate Gradient Method, and Homotopy Theory.*

1. Introduction

The conjugate gradient (CG) method is one of the most popular and well known iterative techniques for solving sparse symmetric positive definite (SPD) system of linear equations. It was originally developed as a direct method, but became popular for its properties as an iterative method especially following the development of sophisticated precondition techniques.

Method of linear conjugate gradient is iterative method to solve minimization problem,

$$\min \boldsymbol{f}(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{G}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + \boldsymbol{c} \quad (1.1)$$

where b is an nx1 vector, c is a constant and G is an nxn positive symmetric definite matrix, we can show that (1.1) is equivalent to a system of linear equations,

$$\boldsymbol{Gx} = \boldsymbol{b} \quad (1.2)$$

Then the unique solution of (1.1) is the same as the solution of (1.2).

Consider the unconstrained minimization problem

$$\min f(x) \quad (1.3)$$

and the conjugate gradient method of the form:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{\alpha}_k\boldsymbol{d}_k \quad (1.4)$$

$$\boldsymbol{d}_{k+1} = \begin{Bmatrix} -\boldsymbol{g}_k & \boldsymbol{for\ k = 0} \\ -\boldsymbol{g}_{k+1} + \boldsymbol{\beta}_k\boldsymbol{d}_k & \boldsymbol{for\ k \geq 0} \end{Bmatrix} \quad (1.5)$$

where $x_k \in R^n$ is the current iterative, d_k is a decent direction of $f(x) at\ x_k$

$g_k = \nabla f(x_k)$ and β_k is a scalar. The scalar chosen so that the methods (1.4) and (1.5) reduce to the linear conjugate gradient method when f is a strictly convex quadratic and when α_k is the exact one-dimensional minimizer. Various conjugate gradient methods have been proposed, and they mainly differ in the choice of the parameter β_k .Some formulas for β_k, called the Fletcher-Reeves (FR)[3], Polak-Ribiere-Polyak (PRP)[7], Hestenes-Stiefel (HS)[5] , Conjugate Descent(CD)[4] and β_k [2] respectively, are given below:

$$\boldsymbol{\beta}_k^{FR} = \|\boldsymbol{g}_{k+1}\|^2/\|\boldsymbol{g}_k\|^2 \quad (1.6)$$

$$\boldsymbol{\beta}_k^{PRP} = \frac{\boldsymbol{g}_{k+1}^T\boldsymbol{y}_k}{\|\boldsymbol{g}_{k-1}\|^2} \quad (1.7)$$

$$\boldsymbol{\beta}_k^{HS} = \frac{\boldsymbol{g}_{k+1}^T\boldsymbol{y}_k}{\boldsymbol{d}_k^T\boldsymbol{y}_k} \quad (1.8)$$

$$\boldsymbol{\beta}_k^{CD} = \frac{-\|\boldsymbol{g}_{k+1}\|^2}{\boldsymbol{d}_k^T\boldsymbol{g}_k} \quad (1.9)$$

$$\boldsymbol{\beta}_k = \boldsymbol{\mu}\frac{\boldsymbol{g}_{k+1}^T\ \boldsymbol{y}_k}{\boldsymbol{d}_k^T\ \boldsymbol{y}_k}(\boldsymbol{1} - \frac{\boldsymbol{2}\ \boldsymbol{d}_k^T\ \boldsymbol{g}_{k+1}}{\|\boldsymbol{y}_k\|^2})(\frac{\boldsymbol{g}_{k+1}^T\ \boldsymbol{y}_k}{\boldsymbol{d}_k^T\ \boldsymbol{y}_k}), \quad (1.10)$$

where $\|.\|$ denotes the Euclidean norm, μ is control parameter ($\mu \in (0,4)$), and $y_k = g_{k+1} - g_k$. The linear conjugate gradient methods generate a sequence of search directions (1.5) such that the following conjugacy condition holds:

$$\boldsymbol{d}_{k+1}^T\boldsymbol{G}\boldsymbol{d}_k = \boldsymbol{0}, \quad (1.11)$$

where G is the Hessian of the objective function. There is many conjugacy condition are suggested, for example:

$$\boldsymbol{d}_{k+1}^T\boldsymbol{y}_k = \boldsymbol{0} \quad (1.12)$$

Extension of (1.12) has been studied by Dai Y.H. and Yuan Y. in [1], introduce the following conjugacy condition:

$$\boldsymbol{d}_{k+1}^T\boldsymbol{y}_k = -\boldsymbol{\tau}\boldsymbol{g}_{k+1}^T\boldsymbol{v}_k\ ,\ \boldsymbol{\tau} \geq \boldsymbol{0} \quad (1.13)$$

The conjugate gradient method is a very efficient line search method for solving large unconstrained problems, due to its lower storage and simple computation. The conjugate gradient method is still the best choice for solving (1. 3).

In practical computations, it is generally believed that the conjugate gradient method is preferred to the relatively exact line searches. As a result, in the conjugate gradient method, the standard Wolfe conditions [8], namely,

$$\boldsymbol{f}(\boldsymbol{x}_k + \boldsymbol{\alpha}_k\boldsymbol{d}_k) - \boldsymbol{f}(\boldsymbol{x}_k) \leq \boldsymbol{\delta}\boldsymbol{\alpha}_k\boldsymbol{g}_k^T\boldsymbol{d}_k \quad (1.14)$$

$$\boldsymbol{g}(\boldsymbol{x}_k + \boldsymbol{\alpha}_k\boldsymbol{d}_k)^T \geq \boldsymbol{\sigma}\boldsymbol{g}_k^T\boldsymbol{d}_k\ , \quad (1.15)$$

where $0 < \delta < \sigma < 1$ are often on the line search, and the strong Wolfe conditions, namely (2.9) and

$|\boldsymbol{x_k} + \boldsymbol{\alpha_k d_k}| \leq -\boldsymbol{\sigma g_k^T d_k}$ (1.16)

Algorithm 1.1: (Classical Conjugate Gradient Algorithm [5])

Set $k = 0$, select

Step (1): the initial point x_0 ,

Step (2) : $\boldsymbol{g_k = \nabla f(x_k)}, \boldsymbol{if\ g_k = 0}$, Then stop

else

set $\boldsymbol{d_k = -g_k}$

Step (3): Compute α_k to minimize $f(x_{k+1})$

Step (4): $\boldsymbol{x_{k+1} = x_k + \alpha_k d_k}$,

Step (5) : $\boldsymbol{g_{k+1} = \nabla f(x_{k+1})}, \boldsymbol{if\ g_{k+1} = 0}$,Then stop .

Step (6) : Compute $\boldsymbol{\beta_k}$, $\boldsymbol{\beta_k^{HS}} = \frac{g_{k+1}^T y_k}{d_k^T y_k}$

Step (7): $\boldsymbol{d_{k+1} = -g_{k+1} + \beta_k d_k}$.

Step (8): If k = n, then go to step 2.

else

$\boldsymbol{k = k+1}$ and go to step 3.

This paper is organized as follows: In Sect.2, we suggested the new (CG) algorithm and we show that the our suggestion is satisfy the conjugacy condition. In section 3, we prove the descent condition of our method. In Section 4, we show the numerical results of our new algorithm. In section 5, we give the conclusions.

2. New Conjugate Gradient method ($\boldsymbol{\beta_k^{new}}$)

In this section, we proposed a new conjugate gradient method by using homotopy theory with parameters (2.8) and (2.10),we can denote the parameters (2.8) and (2.10) as follows β_k^1 and β_k^2 respectivly. The iterates $x_0, x_1, x_2, \ldots$ of our suggestion (β_K^{NEW}) computed by (1.4), where step size α_k>0 is determined according to the Wolfe line search condition and the direction d_k are computed by the rule

$\boldsymbol{d_{k+1} = -g_{k+1} + \beta_K^{New} d_k}$ (2.1)

We will combine two parameters (β_k^1 and β_k^2) by homotopy theory [6],

$$\boldsymbol{\beta_K^{NEW} = (1-\theta_k)\beta_k^1 + \theta_k \beta_k^2}$$

Or

$$\boldsymbol{\beta_K^{NEW}} = (\boldsymbol{1-\theta_k})\frac{g_{k+1}^T y_k}{d_k^T y_k} + \boldsymbol{\theta_k}[\boldsymbol{\mu}\frac{g_{k+1}^T\ y_k}{d_k^T\ y_k}(\boldsymbol{1} - \frac{2\,d_k^T\ g_{k+1}}{\|y_k\|^2})(\frac{g_{k+1}^T\ y_k}{d_k^T\ y_k})] \quad (2.2)$$

Where θ_k is scalar parameter satisfying $0 \leq \theta_k \leq 1$ and we take $\mu = 1$.

We note that if $\theta_k = 0$, then $\beta_K^{NEW} = \beta_k^1$ which is the parameter of HS and if $\theta_k = 1$, then $\beta_K^{NEW} = \beta_k^2$ which is the parameter (1.10).

To show that the our suggestion is satisfy the conjugacy condition, by substituting (2.2) in (2.1) to obtain

$$\mathbf{d_{k+1}} = -\mathbf{g_{k+1}} + \left[(\mathbf{1}-\boldsymbol{\theta}_\mathbf{k})\frac{\mathbf{g_{k+1}^T y_k}}{\mathbf{d_k^T y_k}} + \boldsymbol{\theta}_\mathbf{k}[\boldsymbol{\mu}\frac{\mathbf{g_{k+1}^T\ y_k}}{\mathbf{d_k^T\ y_k}}(\mathbf{1} - \frac{\mathbf{2\,d_k^T\ g_{k+1}}}{\|\mathbf{y_k}\|^2})(\frac{\mathbf{g_{k+1}^T\ y_k}}{\mathbf{d_k^T\ y_k}})]\right]\mathbf{d_k} \quad (2.3)$$

Multiply both sides of above equation by y_k , to get

$$\boldsymbol{d_{k+1}^T\ y_k} = -\boldsymbol{g_{k+1}^T\ y_k} + \left[(\boldsymbol{1}-\boldsymbol{\theta_k})\frac{\boldsymbol{g_{k+1}^T y_k}}{\boldsymbol{d_k^T y_k}} + \boldsymbol{\theta_k}[\boldsymbol{\mu}\frac{\boldsymbol{g_{k+1}^T\ y_k}}{\boldsymbol{d_k^T\ y_k}}(\boldsymbol{1} - \frac{\boldsymbol{2\,d_k^T\ g_{k+1}}}{\|\boldsymbol{y_k}\|^2})(\frac{\boldsymbol{g_{k+1}^T\ y_k}}{\boldsymbol{d_k^T\ y_k}})]\right]\boldsymbol{d_k y_k}$$

This implies that

$$\mathbf{d_{k+1}^T\ y_k} = -\mathbf{g_{k+1}^T\ y_k} + \left[(\mathbf{1}-\boldsymbol{\theta}_\mathbf{k})\mathbf{g_{k+1}^T y_k} + \boldsymbol{\theta}_\mathbf{k}[\boldsymbol{\mu}\mathbf{g_{k+1}^T\ y_k}(\mathbf{1} - \frac{\mathbf{2\,d_k^T\ g_{k+1}}}{\|\mathbf{y_k}\|^2})(\frac{\mathbf{g_{k+1}^T\ y_k}}{\mathbf{d_k^T\ y_k}})]\right]$$

Since $\mu = 1$, so, we have

$$\mathbf{d_{k+1}^T\ y_k} = -\boldsymbol{\theta}_\mathbf{k}(\frac{\mathbf{2\,d_k^T\ g_{k+1}}}{\|\mathbf{y_k}\|^2})(\frac{(\mathbf{g_{k+1}^T\ y_k})^2}{\mathbf{d_k^T\ y_k}}) \quad (2.4)$$

Now, we know that $v_k = \alpha_k d_k$, then, $d_k = \frac{1}{\alpha_k} v_k$

So, (2.4) becomes

$$\mathbf{d_{k+1}^T\ y_k} = -\boldsymbol{\theta}_\mathbf{k}\frac{\mathbf{1}}{\boldsymbol{\alpha}_\mathbf{k}}(\frac{\mathbf{2\,v_k^T\ g_{k+1}}}{\|\mathbf{y_k}\|^2})(\frac{(\mathbf{g_{k+1}^T\ y_k})^2}{\mathbf{d_k^T\ y_k}})$$

Or

$$\mathbf{d_{k+1}^T\ y_k} = -\mathbf{v_k^T\ g_{k+1}}\frac{\mathbf{1}}{\boldsymbol{\alpha}_\mathbf{k}}(\frac{\mathbf{2}\,\boldsymbol{\theta}_\mathbf{k}}{\|\mathbf{y_k}\|^2})(\frac{(\mathbf{g_{k+1}^T\ y_k})^2}{\mathbf{d_k^T\ y_k}}) \quad (2.5)$$

Since $\frac{\mathbf{1}}{\boldsymbol{\alpha}_\mathbf{k}}(\frac{\mathbf{2}\,\boldsymbol{\theta}_\mathbf{k}}{\|\mathbf{y_k}\|^2})(\frac{(\mathbf{g_{k+1}^T\ y_k})^2}{\mathbf{d_k^T\ y_k}}) > \mathbf{0}$

Here , we suppose that

$$\boldsymbol{\tau} = \frac{\mathbf{1}}{\boldsymbol{\alpha}_\mathbf{k}}(\frac{\mathbf{2}\,\boldsymbol{\theta}_\mathbf{k}}{\|\mathbf{y_k}\|^2})(\frac{(\mathbf{g_{k+1}^T\ y_k})^2}{\mathbf{d_k^T\ y_k}})$$

Then, (2.5) becomes

$$\mathbf{d_{k+1}^T\ y_k} = -\boldsymbol{\tau}\mathbf{v_k^T\ g_{k+1}}$$

Theorem2.1: Assume that the sequence $\{x_k\}$ is generated by (1.4), then the modified CG-method in from (2.2) is satisfied the descent condition, i.e. $d_{k+1}^T g_{k+1} \leq 0$ in both cases: exact and inexact line search.

Proof: From (2.1) and (2.2) , we get

$$\boldsymbol{d_{k+1}} = -\boldsymbol{g_{k+1}} + \left[(\boldsymbol{1}-\boldsymbol{\theta_k})\frac{g_{k+1}^T y_k}{d_k^T y_k} + \boldsymbol{\theta_k}[\mu\frac{g_{k+1}^T\, y_k}{d_k^T\, y_k}(\boldsymbol{1}-\frac{2\, d_k^T\, g_{k+1}}{\|y_k\|^2})(\frac{g_{k+1}^T\, y_k}{d_k^T\, y_k})]\right]\boldsymbol{d_k}$$

Multiply both sides of above equation by g_{k+1} , we have

$$\boldsymbol{d_{k+1}^T g_{k+1}} = -\|\boldsymbol{g_{k+1}}\|^2 + \left[(\boldsymbol{1}-\boldsymbol{\theta_k})\frac{g_{k+1}^T y_k}{d_k^T y_k} + \boldsymbol{\theta_k}[\mu\frac{g_{k+1}^T\, y_k}{d_k^T\, y_k}(\boldsymbol{1}-\frac{2\, d_k^T\, g_{k+1}}{\|y_k\|^2})(\frac{g_{k+1}^T\, y_k}{d_k^T\, y_k})]\right]\boldsymbol{d_k^T\, g_{k+1}}$$

Implies that

$$\boldsymbol{d_{k+1}^T g_{k+1}} = -\|\boldsymbol{g_{k+1}}\|^2 + \left[(\boldsymbol{1}-\boldsymbol{\theta_k})\frac{g_{k+1}^T y_k}{d_k^T y_k}(\boldsymbol{d_k^T\, g_{k+1}}) + [(\boldsymbol{\theta_k}\mu\frac{g_{k+1}^T\, y_k}{d_k^T\, y_k}(\boldsymbol{d_k^T\, g_{k+1}}) - (\boldsymbol{\theta_k}\mu)(\frac{2\,(d_k^T\, g_{k+1})^2}{\|y_k\|^2})(\frac{g_{k+1}^T\, y_k}{d_k^T\, y_k})^2]\right]$$

Since $\mu = 1$, then the above equation becomes

$$\boldsymbol{d_{k+1}^T g_{k+1}} = -\|\boldsymbol{g_{k+1}}\|^2 + \frac{g_{k+1}^T y_k}{d_k^T y_k}(\boldsymbol{d_k^T\, g_{k+1}}) - \boldsymbol{\theta_k}(\frac{2\,(d_k^T\, g_{k+1})^2}{\|y_k\|^2})(\frac{g_{k+1}^T\, y_k}{d_k^T\, y_k})^2 \quad (2.6)$$

The proof is complete if the step length α_k is chosen by an exact line search which requires $d_k^T g_{k+1} = 0$.

Now, if the step length $\alpha_k$is chosen by inexact line search which requires $d_k^T g_{k+1} \neq 0$.

We know that, the first two terms of equation (2.6) are less than or equal to zero because the parameter of (HS) is satisfies the descent condition and the third term clearly is less than or equal to zero since $0 < \theta_k < 1$, then,

$$\boldsymbol{d_{k+1}^T g_{k+1}} = -\|\boldsymbol{g_{k+1}}\|^2 + \frac{g_{k+1}^T y_k}{d_k^T y_k}(\boldsymbol{d_k^T\, g_{k+1}}) - \boldsymbol{\theta_k}(\frac{2\,(d_k^T\, g_{k+1})^2}{\|y_k\|^2})(\frac{g_{k+1}^T\, y_k}{d_k^T\, y_k})^2 \leq \boldsymbol{0}$$

Algorithm *2*.1: (New Conjugate Gradient Method)

Step (1) : Set k =0, select initial point x_k.

Step (2) : $g_k = \boldsymbol{\nabla f}\,(\boldsymbol{x_k})$, $\boldsymbol{If\ g_k = 0}$, *then stop*.

Else

Set $\boldsymbol{d_k = -g_k}$.

Step (3): Compute α_k , to minimize $\boldsymbol{f(x_{k+1})}$.

Step (4): $\boldsymbol{x_{k+1} = x_k + \alpha_k d_k}$.,

Step (5): $\boldsymbol{g_{k+1} = \nabla f(x_k + 1)}$, $\boldsymbol{If}\ \|\boldsymbol{g_{k+1}}\|_\infty \leq \boldsymbol{10^{-5}}$, *then stop*.

Step (6): Compute β_k^{new}

Where $\boldsymbol{\beta_K^{NEW}} = (\boldsymbol{1}-\boldsymbol{\theta_k})\frac{g_{k+1}^T y_k}{d_k^T y_k} + \boldsymbol{\theta_k}[\mu\frac{g_{k+1}^T\, y_k}{d_k^T\, y_k}(\boldsymbol{1}-\frac{2\, d_k^T\, g_{k+1}}{\|y_k\|^2})(\frac{g_{k+1}^T\, y_k}{d_k^T\, y_k})]$

$\boldsymbol{0 < \theta_k < 1}$ and $\boldsymbol{\mu = 1}$.

Step (7): $\boldsymbol{d_{k+1} = -g_{k+1} + \beta_k^{new}\ d_k}$

Step (8): If k=n then go to step 2 , where n is natural number

else

k=k+1 and go to step 3.

3. Numerical Results

This section is devoted to test the implementation of new method. We compare the our method with Conjugate Gradient (HS), the comparative tests involve well-known nonlinear problems (standard test function) with different dimensions $4 \leq n \leq 5000$, all programs are written in FORTRAN95 language and for all cases the stopping condition is

$\|g_{k+1}\|_\infty \leq 10^{-5}$ the results given in Table 1 specifically quote the number of function NOF and the nuber of iteration NOI Experimental results in Table 1 confirm that the new CG is superior to standard CG method with respect to the NOI and NOF.

Table (1): Comparative Performance of the Two Algorithms (New Conjugate Gradient Method and HS)

Test functions	N	Algorithm of HS		New algorithm	
		NOI	NOF	NOI	NOF
Powell	4	38	108	32	85
	10	38	108	32	85
	50	38	108	35	102
	100	40	122	35	102
	500	41	124	38	121
	1000	41	124	38	121
	5000	41	124	38	121

Powell 3	4	16	36	14	31
	10	16	36	14	31
	50	16	36	15	33
	100	16	36	15	33
	500	16	36	15	33
	1000	16	36	15	33
	5000	16	36	15	33
Wood	4	30	68	26	60
	10	30	68	26	60
	50	30	68	26	60
	100	30	68	26	60
	500	30	68	26	60
	1000	30	68	27	62
	5000	30	68	27	62
Non diagonal	4	24	64	24	63
	10	26	72	24	69
	50	29	79	29	79
	100	29	79	30	81
	500	-	-	30	80
	1000	29	79	30	80
	5000	30	81	30	80
Mile	4	28	85	28	87
	10	31	102	28	87
	50	31	102	28	87
	100	33	114	31	104
	500	40	146	31	104
	1000	46	176	31	104
	5000	54	211	31	104
G-Central	100	32	184	32	182
	500	32	184	32	182
	1000	32	184	32	182
	5000	32	184	32	182
Total		1217	3832	1068	3325

Note: The fail result in standard CG is considered a twice value of new CG results.

Table (2): Percentage of Improving the New Formula

	HS	New Algorithm
NOI	100%	87.7567789647 %
NOF	100%	86.7693110647 %

3. Conclusion

Our method has been analyzed, implemented and tested to some extent, while numerical tests were carried out, on low and high dimensionality problems, and comparisons were made amongst different non-quadratic and quadratic models with inexact line search.

4. References

Dai Y.H. and Yuan Y, (2001), New conjugacy conditions and related nonlinear conjugate gradient methods, Appl. Math. Optima. 43, 87–101.

Fletcher, R.,(1987). Practical Methods of Optimization, Vol I: Unconstrained Optimization. New York: Wiley.

Fletcher, R. and Reeves, C. (1964), Function minimization by conjugate gradients. J. Comput., 7, 149–154.

Hestenes, M.R. and Stiefel, E. L. ,(1952), Method of conjugate gradient for solving linear systems. J. Res. Natl.Bur. Stand., 49, 409–432.

Omar D. H., (2013), Numerical Methods for Unconstrained Optimization Algorithms with Chaos Theory , University of Zakho, Kurdistan Region – Iraq.

Polak,E. and Ribiere, G., (1969), Note surla convergence des methods' de directions conjugu´ ees. Rev. Fr.Imform. Rech. Op-er., 16, 35–43.

Watson L. T. and Haftka, (1988), Modern Homotopy Methods in Optimization, TR 88-51, Virginia Polytechnic Institute and state University , Blacksburg, November 14, VA 24061.

Wolfe P., (1969), Convergence conditions for ascent method, SIAM Rev. 11, pp.226-235.

کورتیا لێکولینێ:

دڤێ ڤەکولینێدا، مە پێشنیارا رێکەکا نوی کر بوٚ پێنگاڤ پێنگاڤکرنا ئاوەلێ بوٚ ئوپتیمیکا نە گرێدایی بکارئینانا (homotopy theory).ئەو ئەلگوریزما هاتیە پێشنیارکرن مەرجێن ئاوەلی بجھ دئینت . ئەنجامێن ژمارەی یێن ئەلگوریزما نوی باشترە ژ ئەلگوریزما ئاسایی CG بو هەردوو NOI و NOF.

الملخص:

تم في هذا البحث، اقتراح طريقة جديدة للانحدار المرافق لحالات التفضيل الغير المقيدة باستخدام (homotopy theory) .الاشتقاق المقترح يحقق شرط الترافق وشرط الانحدار. اثبتت النتائج الرقمية بان الاشتقاق المقترح هو افضل من النموذج القياسي CG بالنسبة لكل في NOI و NOF.