Kirkuk University Journal /Scientific Studies (KUJSS)
Volume 13, Issue 2, June 2018, pp. (101 - 113)
ISSN 1992 – 0849 (Print), 2616 – 6801 (Online)

Four–Term Conjugate Gradient (CG) Method Based on Pure Conjugacy Condition for Unconstrained Optimization

Hawraz N. Jabbar[1], Khalil K. Abo[2], Hisham. M. Azzam[3]
[1] Department of mathematics, College of Science, Kirkuk University, Kirkuk, Iraq.
[2] Department of mathematics, College of Education, University of Talafer, Talafer, Iraq.
[3] Department of mathematics, College of Science, University of Mosul, Mosul, Iraq.
[1]hawraz_na@yahoo.com, [2]kh_196538@yahoo.com, [3]hisham892012@gmail.com

Abstract

A four-term CG-method based on pure conjugacy condition are proposed, Research activities on extending three-term CG-method to the four-term conjugate gradient method. The new method shown that the suggested CG-methods owns the sufficient descent property. The global convergence of the proposed scheme with the general Wolfe conditions under a suitable assumption was verified. Finally, the computational experiment show that the new method is efficient and robust.

Keywords: Unconstrained optimization, conjugacy condition, Conjugate gradient methods.

Kirkuk University Journal /Scientific Studies (KUJSS)
Volume 13, Issue 2, June 2018, pp. (101 - 113)
ISSN 1992 – 0849 (Print), 2616 – 6801 (Online)

طرائق التدرج المترافق ذات الحدود الاربعة باستخدام شرط الترفق الصرف في الامثلية غير المقيدة

هوراز ناظم جبار[1]، خليل خضر عبو[2]، هشام محمد عزام[3]
[1] قسم الرياضيات، كلية العلوم، جامعة كركوك، كركوك، العراق.
[2] قسم الرياضيات، كلية التربية الاساسية ، جامعة تلعفر، تلعفر، العراق.
[3] قسم الرياضيات، كلية علوم الحاسوب والرياضيات، جامعة الموصل، الموصل، العراق.
[1]hawraz_na@yahoo.com, [2]kh_196538@yahoo.com, [3]hisham892012@gmail.com

الملخص

في هذا البحث تم تقديم طرائق التدرج المترافق ذات الحدود الاربعة بالاعتماد على شرط الترافق الصرف, وقد التركيز على توسيع طرائق التدرج المترافق ذات الثلاثة حدود الى ذات الاربعة حدود. اظهرت الطريقة المقترحة الجديدة بانها تحقق خاصية الانحدار. ان خاصية التقارب الشامل للطريقة المقترحة باستخدام شرط وولف العام قد تحققت. اظهرت النتائج العددية كفاءة الطريقة الجديدة مقارنة بالطرق ذات الحدود الثلاثة

الكلمات المفتاحية : الامثلية غير المقيدة, شرط الترافق, طرائق التدرج المترافق

1. Introduction

A problem, which arises in many practical situations, is to minimize a given objective function, where   is a real n-dimensional vector, which may be subject to a number of constraints. The unconstrained case represents a significant class of practical problems, firstly because many constrained problems can be easily converted to and solved by methods of unconstrained optimization and secondly because many optimization problems require the solution of unconstrained sub problems. The generic unconstrained optimization problem is defined by,

$$Minimize\ f(x), x \epsilon R^n$$

where $f(x)$ is assumed to be at least twice continuously differentiable and where the function and first derivatives can be evaluated at any point x [1].

Let $f(x): R^n \to R$ be a function with continuous derivatives consider the unconstrained optimization problem given in equation (1). In general, it may be too ambitious to find a global Min of f [2]. The gradient of $f(x)$ is defined as the vector containing the first order partial derivatives of $f(x)$ given by [3].

$$g(x)=\nabla f(x)=\left[\frac{\partial f(x)}{\partial x_1}\ \frac{\partial f(x)}{\partial x_2}\ \ldots\ldots\ldots\ldots\ldots\frac{\partial f(x)}{\partial x_n}\right]^T \tag{1}$$

Hence, we will just look for stationary points of f i.e. a point $x^* \in R^n$ that satisfies $g(x^*)=0$ to begin. Let $x_1 \in R^n$ be an initial estimate with $g(x_1)\neq 0$. In order to achieve progress we need to proceed in some search direction $d_k \in R^n$. For instance, we can update the iterates according to

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

We know that α_k is a positive scalar and called the step length which is determined by some line search, controls how far we proceed in the direction d_k. and we can calculate search direction as in next equation:

$$\begin{aligned} d_1 &= -g_1 & k&=1 \\ d_{k+1} &= -g_{k+1} + \beta_k\ d_k & k&\geq 1 \end{aligned} \tag{3}$$

where the parameter β_k with d_{k+1} is computed as in one of the following manners (4 - 9):

$d_1 = -g_1,\ y_k = g_{k+1} - g_k$ and

$$d_{k+1} = -g_{k+1} + \frac{y_k^T g_{k+1}}{y_k^T d_k} d_k\ ,\ [4] \tag{4}$$

$$d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} d_k\ ,\ [5] \tag{5}$$

$$d_{k+1} = -g_{k+1} + \frac{y_k^T g_{k+1}}{g_k^T g_k} d_k\ ,\ [6] \tag{6}$$

$$d_{k+1} = -g_{k+1} - \frac{g_{k+1}^T g_{k+1}}{d_k^T g_k} d_k \text{, [3]} \tag{7}$$

$$d_{k+1} = -g_{k+1} - \frac{g_{k+1}^T y_{k+1}}{d_k^T g_k} d_k \text{, [7]} \tag{8}$$

$$d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T \, g_{k+1}}{d_k^T \, y_k} d_k \text{, [8]} \tag{9}$$

In the convergence analysis and implementation of conjugate gradient method, one often requires the exact and inexact line search such as the Wolfe conditions or the strong Wolfe conditions. The Wolfe line search is to find α_k such that

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k \tag{10}$$

$$d_k^T g(x_k + \alpha_k d_k) \ge \sigma \, d_k^T g_k \tag{11}$$

with $0 < \delta < \sigma$. The strong Wolf line search is to find α_k such that

$$f\,(x_k + \alpha_k d_k) \le f\,(x) + \delta \alpha_k g_k^T d_k \tag{12}$$

$$\left| d_k^T g(x_k + \alpha_k d_k) \right| \le -\sigma \, d_k^T g_k \tag{13}$$

where $0 < \delta < \sigma < 1$ are constants, by Li and Weijun, **[**9**]**.

There are other types of conjugate gradient method is called three-term conjugate gradient methods

$$d_{k+1} = -g_{k+1} + \left(\frac{y_k^T g_{k+1}}{s_k^T y_k} - \left(1 + \frac{y_k^T y_k}{s_k^T y_k}\right) \frac{s_k^T g_{k+1}}{s_k^T y_k} \right) s_k + \frac{s_k^T g_{k+1}}{s_k^T y_k} y_k \text{, [10]} \tag{14}$$

$$d_{k+1} = -y_k + \frac{y_k^T y_k}{y_k^T d_k} d_k + \frac{y_{k-1}^T y_k}{y_{k-1}^T d_k} d_{k-1} \text{, [11]} \tag{15}$$

Kirkuk University Journal /Scientific Studies (KUJSS)
Volume 13, Issue 2, June 2018, pp. (101 - 113)
ISSN 1992 – 0849 (Print), 2616 – 6801 (Online)

$$d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T y_k}{s_k^T y_k} s_k - \frac{g_{k+1}^T s_k}{s_k^T y_k} y_k \quad (16)$$
, [12]

$$d_{k+1} = -g_{k+1} + \beta^{HS} d_k - t \frac{g_{k+1}^T s_k}{|d_k^T y_k|} d_k - \frac{g_{k+1}^T s_k}{d_k^T y_k} y_k \quad (17)$$

$t_k = \max\left\{\xi, 1 - \frac{\|y_k\|^2}{s_k^T y_k}\right\}$ where ξ is a nonnegative constant. [13].

2. A New Four-Term CG-Method:

The search direction (17) can be viewed as a Four-term extension of the DL method. If t=0 then the method reduces to the three-term CG method ZZL which satisfies the sufficient descent condition

$$d_{k+1} = -g_{k+1} + \beta^{HS} d_k - t \frac{g_{k+1}^T s_k}{|d_k^T y_k|} d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k \quad (18)$$

Then from pure conjugacy condition

$$d_{k+1}^T y_k = 0 \quad (19)$$

$$d_{k+1}^T y_k = -y_k^T g_{k+1} + \beta^{HS} s_k^T y_k - t_k \frac{g_{k+1}^T s_k}{|s_k^T y_k|} s_k^T y_k - \frac{g_{k+1}^T s_k}{s_k^T y_k} y_k^T y_k = 0 \quad (20)$$

$$t_k = -\frac{y_k^T y_k}{s_k^T y_k} \quad (21)$$

Algorithm

Step1. Given $x_1 \in R^n, \varepsilon > 0$, $d_1 = -g_1$; $k = 1$

Step2. If $\|g_{k+1}\| \le \varepsilon$, stop ,else go to Step 3

Step3. Find α_k satisfying Wolfe condition (12) and (13).

Step4.Compute new iterative x_{k+1} by $x_{k+1} = x_k + \alpha_k d_k$.

Step5. Compute β^{HS} , and d_{k+1} from (18) and t_k from (21) and set k=k+1 go to step 2.

3. Descent Property and Global Convergence Analysis

Next we will show that our three-term CG (20)-method satisfies the descent property and global converges.

Assumption (1)

Assume f is bound below in the level set $S=\{x\in R^n : f(x)\le f(x_\circ)\}$; In some neighborhood N of S, f is continuously differentiable and its gradient is Lipchitz continuous, there exist $L>0$ such that:

$$\|g(x)-g(y)\|\le L\|x-y\| \quad \forall \text{x, y} \in \text{N} \ . \tag{22}$$

or equivalently

$$y_k^T s_k \ge \mu\|s_k\|^2. \quad \text{and} \quad \mu\|s_k\|^2 \le y_k^T s_k \le L\|s_k\|^2. \tag{23}$$

From (20) we get

$$y_k^T y_k \le L y_k^T s_k . \tag{24}$$

On the other hand, under Assumption (1) , It is clear that there exist positive constant B such

$$\|x\|\le B \ ,\forall x \in \Omega \tag{25}$$

$$\underline{\gamma} \le \|g(x)\| \le \overline{\gamma} \ ,\forall x \in \Omega \tag{26}$$

Suppose that Assumption (1) holds and if the line search satisfies the Wolfe condition. It follows from [6] that

$$s_k^T y_k = s_k^T (g_{k+1}-g_k) \ge (\sigma-1) s_k^T g_k \tag{27}$$

Lemma (1)

Suppose that Assumption (1) and equation (25) hold .consider any conjugate gradient method in from (2) and (3), where d_k is a descent direction and α_k is obtained by the strong Wolfe (SW) .If

$$\sum_{k>1} \frac{1}{\|d_{k+1}\|^2} = \infty \tag{28}$$

Then we have

Kirkuk University Journal /Scientific Studies (KUJSS)
Volume 13, Issue 2, June 2018, pp. (101 - 113)
ISSN 1992 – 0849 (Print), 2616 – 6801 (Online)

$$\lim_{k\to\infty}\left(\inf\|g_k\|\right)=0 . \tag{29}$$

More details can be found in [14], [15]

Theorem (1)

Let $\{x_k\}$ and $\{d_k\}$ be generated by the equation (2), (3) and α_k satisfies Wolfe line search conditions (SWLSC) (12) and (13), then $d_k^T g_k < 0$ hold for all $k \geq 1$.

Proof. The conclusion can be proved by induction. When *K*=1, we have $d_1^T g_1 \geq -\|g_1\|^2 < 0$

Suppose that $d_k^T g_k < 0$ hold for all k. From (20) we have

$$d_{k+1} = -g_{k+1} + \beta^{HS} s_k - t\frac{g_{k+1}^T s_k}{|s_k^T y_k|} d_k - \frac{g_{k+1}^T s_k}{s_k^T y_k} y_k$$

$$t_k = -\frac{y_k^T y_k}{s_k^T y_k}$$

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1} + \frac{y_k^T g_{k+1}}{y_k^T s_k} g_{k+1}^T s_k + \frac{y_k^T y_k}{s_k^T y_k}\frac{g_{k+1}^T s_k}{|s_k^T y_k|} g_{k+1}^T s_k - \frac{g_{k+1}^T s_k}{s_k^T y_k} g_{k+1}^T y_k$$

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1} + \frac{y_k^T y_k}{s_k^T y_k}\frac{g_{k+1}^T s_k}{|s_k^T y_k|} g_{k+1}^T s_k$$

$$y_k^T y_k \leq L y_k^T s_k$$

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1} + \frac{L\left(g_{k+1}^T s_k\right)^2}{s_k^T y_k}$$

$$s_k^T y_k = s_k^T\left(g_{k+1} - g_k\right) \geq (\sigma - 1) s_k^T g_k$$

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1} + \frac{L\max\{g_{k+1}^T s_k, 0\}}{(\sigma - 1)}$$

$0 < \sigma < 1$, 0<L<1

$$g_{k+1}^T d_{k+1} \leq -g_{k+1}^T g_{k+1} + \frac{L\max\{g_{k+1}^T s_k, 0\}}{(\sigma - 1)}$$

Kirkuk University Journal /Scientific Studies (KUJSS)
Volume 13, Issue 2, June 2018, pp. (101 - 113)
ISSN 1992 – 0849 (Print), 2616 – 6801 (Online)

Theorem (2)

Suppose that assumption (1) holds, and consider the new algorithm (New), where α_k is computed by the Wolfe line search conditions (12) and (13) then:

Proof:-

$$\|d_{k+1}\|^2 = \left\| -g_{k+1} + \beta^{HS} s_k - t\frac{g_{k+1}^T s_k}{|s_k^T y_k|} d_k - \frac{g_{k+1}^T s_k}{s_k^T y_k} y_k \right\|^2$$

$$\|d_{k+1}\|^2 \le \|g_{k+1}\|^2 + \|g_{k+1}\|^2 \left(\frac{L\alpha\|s_k\|^2}{\|y_k^T s_k\|^2}\|s_k\|^2 + t\frac{\alpha\|s_k\|^2}{\|y_k^T s_k\|^2}\|s_k\|^2 + \frac{\alpha\|s_k\|^2}{\|y_k^T s_k\|^2}\|y_k\|^2 \right)$$

$$\|d_{k+1}\|^2 \le \|g_{k+1}\|^2 \left(1 + \frac{L\alpha\|s_k\|^2}{\mu\|s_k\|^2}\|s_k\|^2 + t\frac{\alpha\|s_k\|^2}{\mu\|s_k\|^2}\|s_k\|^2 + \frac{\alpha\|s_k\|^2}{\mu\|s_k\|^2}\|y_k\|^2 \right)$$

$$\|d_{k+1}\|^2 \le \|g_{k+1}\|^2 \left(1 + \frac{L\alpha}{\mu}\|s_k\|^2 + t\frac{\alpha}{\mu}\|s_k\|^2 + \frac{\alpha}{\mu}\|y_k\|^2 \right)$$

Let B= $\|s_k\|^2$, C= $\|y_k\|^2$

$$\|d_{k+1}\|^2 \le \|g_{k+1}\|^2 \left(1 + \frac{L\alpha}{\mu}B + t\frac{\alpha}{\mu}B + \frac{\alpha}{\mu}C \right)$$

Let M= $\left(1 + \frac{L\alpha}{\mu}B + t\frac{\alpha}{\mu}B + \frac{\alpha}{\mu}C \right)$

$$\|d_{k+1}\|^2 \le \|g_{k+1}\|^2 M \le \bar{\gamma}^2 M$$

$$\sum_{k\ge 1}\frac{1}{\|d_{k+1}\|^2} \ge \frac{1}{\bar{\gamma}^2 M}\sum_{k\ge 1} 1 = \infty$$

$$\lim_{k\to\infty}\left(\inf \|g_k\|\right) = 0$$

4. Numerical Results and Comparisons

In this section, we compare the performance of new formal developed a New method of conjugate gradient method to Three term conjugate gradient method (SBK). we have selected (75) large scale unconstrained optimization problem, for each test problems taken from [16]. For each test function we have considered numerical experiments with the

Kirkuk University Journal /Scientific Studies (KUJSS)
Volume 13, Issue 2, June 2018, pp. (101 - 113)
ISSN 1992 – 0849 (Print), 2616 – 6801 (Online)

number of variables n=100,…,1000. These new version is compared with well-known conjugate gradient algorithm (SBK). All these algorithms are implemented with standard Wolfe line search conditions (12) and (13) with. In all these cases, the stopping criteria is the $\|g_k\| = 10^{-6}$. All codes are written in double precision FORTRAN Language with F77 default compiler settings. The test functions usually start point standard initially summary numerical results recorded in the Figs. 1, 2, 3. The performance profile by [17] is used to display the performance of the developed a New method of conjugate gradient algorithm with TTCG (SBK) algorithms. Define p=750 as the whole set of n_p test problems and S=2 the set of the interested solvers. Let $l_{p,s}$ be the number of objective function evaluations required by solver s for problem p. Define the performance ratio as

$$r_{p,s} = \frac{l_{p,s}}{l_p^*} \tag{30}$$

Where $l_p^* = \min\{l_{p,s} : s \in S\}$. It is obvious that $r_{p,s} \geq 1$ for all p,s. If a solver fails to solve a problem, the ratio $r_{p,s}$ is assigned to be a large number M. The performance profile for each solver s is defined as the following cumulative distribution function for performance ratio $r_{p,s}$,

$$\rho_s(\tau) = \frac{size\{p \in P : r_{p,s} \leq \tau\}}{n_p} \tag{31}$$

Obviously, $p_s(1)$ represents the percentage of problems for which solver s is the best. See [17] for more details about the performance profile. The performance profile can also be used to analyze the number of iterations, the number of gradient evaluations and the CPU time. Besides, to get a clear observation, we give the horizontal coordinate a log-scale in the Figures (1-3).

Kirkuk University Journal /Scientific Studies (KUJSS)
Volume 13, Issue 2, June 2018, pp. (101 - 113)
ISSN 1992 – 0849 (Print), 2616 – 6801 (Online)

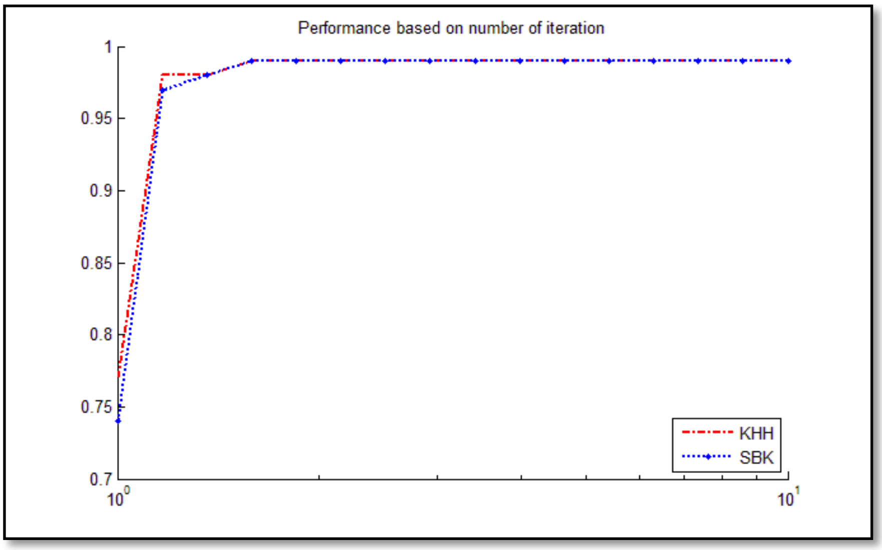

Fig. 1: Performance based on iteration

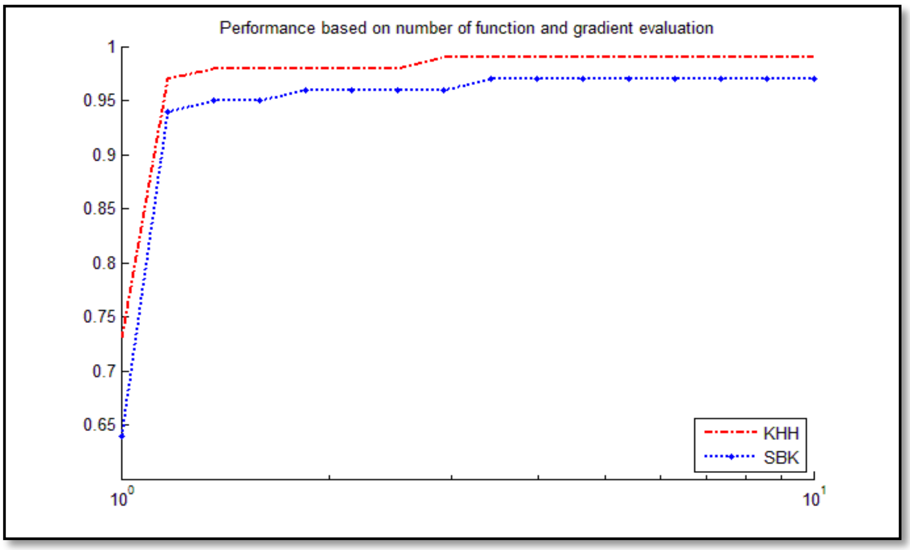

Fig. 2: Performance based on Function

Kirkuk University Journal /Scientific Studies (KUJSS)
Volume 13, Issue 2, June 2018, pp. (101 - 113)
ISSN 1992 – 0849 (Print), 2616 – 6801 (Online)

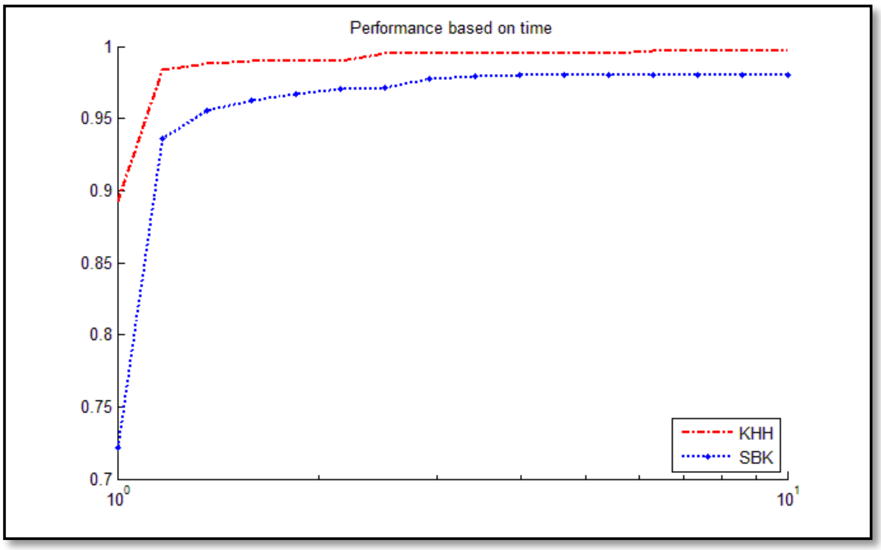

Fig. 3: Performance based on Time

References

[1] Lenir, A. "***Variable Storage CG-Methods***", M.Sc. Thesis, University of leed, Dep. Of Computer Studies, (1981).

[2] Bonanza, J., Gilbert, J., Lemarechal, C. and Sagastizable, G. "***Numerical Optimization***" Second Edition. Springer-Verlag Berlin Heidelberg, (2006).

[3] Fletcher, R., "***Practical Methods of Optimization (second edition)***", John Wiley and Sons, New York (1987).

[4] Hestenes, M. and Stiefel, E., "***Method of conjugate gradients for solving linear systems***", J. Research Nat. Standards, 49, 409(1952).

[5] Fletcher, R. and Reeves, C.M. "***Function Minimization by Conjugate Gradients***", Computer Journal 7,149 (1964).

[6] Polak E., and Ribiere, G., "***Not sur la convergence de directions conjugate***". Rev. Franaise Informant Researche operationelle, 3e Anne'e., 16, 35 (1969).

Kirkuk University Journal /Scientific Studies (KUJSS)
Volume 13, Issue 2, June 2018, pp. (101 - 113)
ISSN 1992 – 0849 (Print), 2616 – 6801 (Online)

[7] Dixon, L. C., "***Conjugate Gradient Algorithms for Quadratic Termination without Linear Searches***", Journal of the Institute of Mathematics and its Applications, 15, 317 (1975).

[8] Dai Y.H. and Yuan Y., "***A nonlinear conjugate gradient method with a strong global convergence property***", SIAM Journal on Optimization, 10, 177 (1999).

[9] Li, Z. and Weijun, Z., "***Tow descent hybrid conjugate gradient methods for optimization***" Journal of computational and Appl. Math, 216, 251 (2008).

[10] Shanno, D., "***Conjugate gradient methods with inexact searches***", Math. of operation Research, 3, 244 (1978).

[11] Nazareth, L "***A conjugate direction algorithms without line search***", Journal of Optimition Theory and Applications 23, 373 (1977).

[12] Zhang, L. Zhou, W. and Li, D., "***Some descent three-term conjugate gradient methods and their global convergence"*** Optimization Methods and Software, 22, 697 (2007).

[13] Kafaki B, Saman, "***A Class of Descent Four–Term Extension of the Dai–Liao Conjugate Gradient Method Based On the Scaled Memoryless Bfgs Update***". Journal of Industrial and Management Optimization", Department of Mathematics, Faculty of Mathematics, Statistics and Computer Science Semnan University, P.O. Box: 35195–363, Semnan, Iran (2017).

[14] Tomizuka. H and Yabe. H., "***A Hybrid Conjugate Gradient method for unconstrained Optimization***", Journal of Operation Research Society of Japan, 47, 1 (2004).

[15] Dai , Y. and Liao, L., "***New conjugacy conditions and related non-linear CG methods***" Appl. Math optim.,(43). Spring-verlag, New York (2001).

Volume 13, Issue 2, June 2018, pp. (101 - 113)
ISSN 1992 – 0849 (Print), 2616 – 6801 (Online)

[16] Andrei N., "***An Unconstrained Optimization test function collection***". Adv. Model. Optimization. 10, 147 (2008).

[17] Dolan. E. D and Mor´e. J. J, "***Benchmarking optimization software with performance profiles***", Math. Programming, 91, 201 (2002).