



Kirkuk University Journal /Scientific Studies (KUJSS)

Volume 12, Issue 2, March 2017

ISSN 1992 – 0849

طريقة جديدة للمتجهات المترافقة غير الخطية أستناداً الى طريقتي Dai- Liao و Kafaki-Ghanbari

أ.د. خليل خضر عبو اينور جودت نامق

قسم الرياضيات

كلية علوم الحاسوب والرياضيات

جامعة الموصل

kh_196538@yahoo.com aynoor.Javdet@gmail.com

الملخص

استناداً الى الخوارزميات Dai-Liao و Kafaki-Ghanbari تم اقتراح الطريقة الجديدة في التدرج المترافق الغير الخطي. بفرض بعض الشروط أثبت خاصية الأنحدار الكافي وكذلك خاصية التقارب تم أثبات التقارب المطلق للخوارزمية المقترحة وذلك بأستخدام خط البحث Wolfe الأعتيادي. النتائج العددية أظهرت بأن الخوارزمية المقترحة تمتلك كفاءة جيدة في حل مسائل الأختبار.

كلمات مفتاحية: طريقة التدرج المترافق، اتجاه الأنحدار، الامثلية الغير المقيدة.



Kirkuk University Journal /Scientific Studies (KUJSS)

Volume 12, Issue 2, March 2017

ISSN 1992 – 0849



Kirkuk University Journal /Scientific Studies (KUJSS)

Volume 12, Issue 2, March 2017

ISSN 1992 – 0849

A new non-linear conjugate gradient method based on the Dai-Liao and Kafaki-Ghanbari methods

Prof. Khalil K. Abbo

Aynur J. Namik

Department of Mathematics

College of computers sciences and mathematics

Mosul University

kh_196538@yahoo.com

Aynoor.Javdet@gmail.com

Abstract

Based on the Dai-Liao and Kafaki-Ghanbari methods, a new non-linear conjugate gradient method is proposed. Under proper conditions, it is briefly shown that our proposed method possess the descent property and generates conjugate directions. We also show that the suggested method with Wolfe line search conditions is globally convergent. Numerical results illustrates that our suggested method can efficiently solve the test problems and therefore is promising.

Keywords: Conjugate gradient method, descent direction, unconstrained optimization.



Kirkuk University Journal /Scientific Studies (KUJSS)

Volume 12, Issue 2, March 2017

ISSN 1992 – 0849



1. Introduction

Recently, due to the features of strong global convergence properties and low memory requirement, conjugate gradient (CG) methods constitute an active choice for efficiently solving the large- scale unconstrained optimization problems [4]. We refer to an excellent survey [8] for a review on recent advances in this area.

Conjugacy condition is an important factor in CG methods. The searching directions in CG methods are often selected in such a way that, when applied to minimize a strongly quadratic convex function, two successive directions are conjugate if no round-off error exists, subject to the Hessian of the quadratic function. That is to say, minimizing a convex quadratic function in a subspace spanned by a set of mutually conjugate directions is equivalent in the sense that one minimizes this function along each conjugate direction in turn. But for the general nonlinear function, the searching directions in most methods fail to satisfy the conjugacy condition. This feature motivates us to solve unconstrained problems by seeking efficient conjugacy conditions [8].

Consider the following unconstrained optimization problem:

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



where $f : R^n \rightarrow R$ is a smooth nonlinear function and its gradient $g(x)$ is available. The iterative formula of a CG method is given by

$$x_{k+1} = x_k + s_k, \quad s_k = \alpha_k d_k \quad k = 1, 2, \dots \quad (2)$$

where d_k is a search direction updated by $d_1 = -g_1$ and

$$d_{k+1} = -g_{k+1} + \beta_k d_k \quad (3)$$

and the step-length $\alpha_k > 0$ is commonly chosen to satisfy certain line search conditions [11]. Among them, the so-called Wolfe conditions have attracted special attention in the convergence analyses and the implementations of CG methods, requiring that

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \rho \alpha_k d_k^T g_k \quad (4)$$

$$g(x_k + \alpha_k d_k)^T d_{k+1} \geq \sigma d_k^T g_k \quad (5)$$

The stronger version of the Wolfe line search conditions are (4) and

$$|g_{k+1}^T d_k| \leq -\sigma g_k^T d_k \quad (6)$$

where $0 < \rho < \sigma < 1$ are often imposed on the line search.

With a distinct choice of the parameter β_k in (3), the obtained method has different theoretical property and numerical performance. The leading parameters formulate for β_k are called the Fletcher-Reeves (FR) [6], Hestenes-Stiefel (HS) [7] and Polak–Ribie`re (PR) [9], More recent reviews on nonlinear conjugate gradient methods can be found in Hager and Zhang [8].

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \quad \beta_k^{HS} = \frac{y_k^T g_{k+1}}{y_k^T d_k}, \quad \beta_k^{PR} = \frac{y_k^T g_{k+1}}{\|g_k\|^2}$$

Here and throughout the paper, we always use $\|\cdot\|$ to stand for the Euclidian norm of vectors and $y_k = g_{k+1} - g_k$.

More recently, many researchers highlighted two properties in designing new CG methods, the first is the conjugacy condition and the second is the descent property, which play a crucial role in obtaining global convergence and nice actual performance.

In order to accelerate the CG method, the conjugacy condition is often utilized to obtain the order accuracy in the approximation of the curvature of the function as high as possible. By modifying the HS method, Dai and Liao [3] proposed the following Dai–Liao (DL) conjugacy condition

$$d_{k+1}^T y_k = -t s_k^T g_{k+1} \tag{7}$$

Based on the above conjugacy condition Dai and Liao in [3] suggested the following conjugacy parameter

$$\beta^{DL} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - t \frac{s_k^T g_{k+1}}{d_k^T y_k} \tag{8}$$

Very recently Kafaki and Ghanbari [2] discussed the optimal value for the parameter t , see [2], and suggested the following choices for t

$$t_{k1}^* = \frac{s_k^T y_k}{\|s_k\|^2} + \frac{\|y_k\|}{\|s_k\|}, \quad \text{and} \quad t_{k2}^* = \frac{\|y_k\|}{\|s_k\|} \tag{9}$$

Hence the search directions for Kafaki and Ghanbari methods are

$$d_{k+1}^{KF1} = -g_{k+1} + \frac{y_k^T g_{k+1}}{s_k^T y_k} s_k - \left(\frac{s_k^T y_k}{\|s_k\|^2} + \frac{\|y_k\|}{\|s_k\|} \right) \frac{s_k^T g_{k+1}}{s_k^T y_k} s_k \tag{10}$$

$$d_{k+1}^{KF2} = -g_{k+1} + \frac{y_k^T g_{k+1}}{s_k^T y_k} s_k - \left(\frac{\|y_k\|}{\|s_k\|} \right) \frac{s_k^T g_{k+1}}{s_k^T y_k} s_k \tag{11}$$

The latter property is an indispensable factor in the convergence analysis of CG methods. Exactly, a direction d_k satisfies the so-called sufficient descent condition, if there exists a constant $c > 0$ such that

$$d_k^T g_k \leq -c \|g_k\|^2, \quad k \geq 1 \tag{12}$$

The paper is organized as follows. Section 2 describes the suggested method and their properties. In section 3 the global convergence analysis for the proposed method is discussed. Section 4 is devoted to providing numerical results.

2. Derivation of the new (AK1 say) CG method

The aim of this section is to derive a new conjugate gradient method Aynur and Khalil (AK1 say) by using Dai-Liao and Kafaki-Ghanbari CG methods. consider the search direction given by Dai – Liao

$$d_{k+1}^{DL} = -g_{k+1} + \frac{y_k^T g_{k+1}}{s_k^T y_k} s_k - t \frac{s_k^T g_{k+1}}{s_k^T y_k} s_k \tag{13}$$

It is remarkable that numerical performance of the DL method is very dependent on the parameter t for which there is no any optimal choice [1]. It has been attempts to find an ideal value for t . We suggest the following value for t . let

$$t = t_{k1}^* - t_{k2}^* = \frac{s_k^T y_k}{\|s_k\|^2} + \frac{\|y_k\|}{\|s_k\|} - \frac{\|y_k\|}{\|s_k\|} = \frac{s_k^T y_k}{\|s_k\|^2} \tag{14}$$

Therefore if we substitute the above value for t in the DL method we get the new search direction (AK1) can be defined as follows:

$$d_{k+1}^{AK1} = -g_{k+1} + \frac{y_k^T g_{k+1}}{s_k^T y_k} s_k - \frac{s_k^T g_{k+1}}{s_k^T s_k} s_k \tag{15}$$

We can define the suggested (AK1) algorithm as follows:

Algorithm (AK1)

Step (1): Select a starting point $x_1 \in \text{dom } f$ and $\varepsilon > 0$, compute $f_1 = f(x_1)$ and $g_1 = \nabla f(x_1)$. Select some positive values for ρ and σ . Set $d_1 = -g_1$ and $k = 1$.

Step (2): Test for convergence .If $\|g_k\|_\infty \leq \varepsilon$, then stop x_k is optimal ; otherwise go to step (3).

Step (3): Determine the step length α_k , by using the Wolfe line search conditions (4)-(5).

Step (4): Update the variables as : $x_{k+1} = x_k + \alpha_k d_k$. Compute f_{k+1} and g_{k+1} . Compute $y_k = g_{k+1} - g_k$ and $s_k = x_{k+1} - x_k$.

Step (5): Compute the search direction as: d_{k+1}^{AK1} in (15) .

Step (6): Set $k = k + 1$ and go to step 2.

In the following theorems we will prove that our method generates conjugate directions and sufficient descent directions.

Theorem (1): Suppose that the step-size α_k satisfies the standard Wolfe conditions (SDWC), consider the search directions d_k generated from (15) then the search directions d_{k+1} are conjugate , for all k . i.e.

$$y_k^T d_{k+1}^{AK1} = -t s_k^T g_{k+1} \tag{16}$$

Proof:

By multiplying both sides of equation (15) to y_k^T we get

$$y_k^T d_{k+1}^{AK1} = -\left(\frac{y_k^T s_k}{s_k^T s_k}\right) s_k^T g_{k+1}$$

As with Wolfe condition $s_k^T y_k > 0$

$$\therefore t = \left(\frac{y_k^T s_k}{s_k^T s_k}\right) > 0 \tag{17}$$

Theorem (2): Suppose that the objective function is uniformly convex and step-size α_k satisfies the standard Wolfe conditions (SDWC), consider the search directions d_k generated from (15) then the search directions d_{k+1} satisfies the sufficient descent condition

$$d_k^T g_k \leq -c\|g_k\|^2, \quad k \geq 1 \tag{18}$$

Proof:

The proof is by induction.

$$d_1 = -g_1 \rightarrow d_1^T g_1 = -\|g_1\|^2$$

Know let $s_k^T g_k < -c\|g_k\|^2$ to proof for $k + 1$, multiply (15) by g_{k+1}^T to get

$$d_{k+1}^T g_{k+1} = -g_{k+1}^T g_{k+1} + \left[\frac{y_k^T g_{k+1}}{s_k^T y_k} - \frac{s_k^T g_{k+1}}{\|s_k\|^2} \right] s_k^T g_{k+1}$$

Now we simplify the equation to get the following

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

Using Lipschitz condition $g_{k+1}^T y_k \leq L s_k^T g_{k+1}$ in the second -term of the above equation to get

$$d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 + \frac{L}{s_k^T y_k} (s_k^T g_{k+1})^2 - \frac{(s_k^T g_{k+1})^2}{\|s_k\|^2}$$

On the other hand, since the objective function is uniformly convex that is

$s_k^T y_k \geq \eta s_k^T s_k$ satisfies the following inequality:

$$\left(\frac{L}{s_k^T y_k} - \frac{1}{\eta \|s_k\|^2} \right) = -c_1 \quad \text{where} \quad c_1 > 0$$

$$d_{k+1}^T g_{k+1} \leq -\|g_{k+1}\|^2 - c_1 (s_k^T g_{k+1})^2$$

The proof is complete.

3. Convergence analysis

Assume the following.

(1) The level set $S = \{x \in R^n : f(x) \leq f(x_0)\}$ is bounded, i.e. there exists

positive constant $B > 0$ such that, for all $x \in S$, $\|x\| \leq B$.

(2) In a neighborhood N of S the function f is continuously differentiable and its gradient is Lipschitz continuous for all $x, y \in N$.

Under these assumptions on f , there exists a constant $\Gamma \geq 0$ such that $\|\nabla f(x)\| \leq \Gamma$, for all $x \in S$. Observe that the assumption that the function f is bounded below is weaker than the usual assumption that the level set is bounded. Although the search directions generated by (15) are always descent directions, to ensure convergence of the algorithm we need to constrain the choice of the step-length α_k . The following proposition shows that the Wolfe line search always gives a lower bound for the step-length α_k .

Proposition 1 [14]. Suppose that d_k is a descent direction and that the gradient ∇f satisfies the Lipschitz condition for all x on the line segment connecting x_k and x_{k+1} . If the line search satisfies the Wolfe conditions (4) and (5), then

$$\alpha_k \geq \frac{(1-\sigma)|g_k^T d_k|}{L\|d_k\|^2} \tag{19}$$

Proposition 2 [10]. Suppose that assumptions (1) and (2) hold. Consider the algorithm (2) and (15), where d_k is a descent direction and α_k is computed by the general Wolfe line search (4) and (5). Then

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \tag{20}$$

Proposition 3 [12,13]. Suppose that assumptions (1) and (2) hold, and consider any conjugate gradient algorithm (2), where d_k is a descent direction and α_k is obtained by the strong Wolfe line search (4) and (6). If

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} = \infty \tag{21}$$

Then $\lim_{k \rightarrow \infty} \inf \|g_k\| = 0$

For uniformly convex functions, we can prove that our suggested AK1 algorithm is globally convergent (theorem4).

Theorem (4): Suppose that assumptions (1) and (2) hold, and consider the algorithm (2) and (15), where d_k is a descent direction and α_k is computed by the strong Wolfe line search (4) and (6). Suppose that f is a uniformly convex function on S , i.e. there exists a constant $\mu > 0$ such that

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \mu \|x - y\|^2, \quad \forall x, y \in N \tag{22}$$

Then $\lim_{k \rightarrow \infty} \inf \|g_k\| = 0$

Proof: The prove is by Contradiction.

$$\begin{aligned} \|d_{k+1}\| &= \left\| -g_{k+1} + \left(\frac{y_k^T g_{k+1}}{s_k^T y_k} - \frac{s_k^T g_{k+1}}{\|s_k\|^2} \right) s_k \right\| \\ &\leq \|g_{k+1}\| + \frac{L\|s_k\|^2 \|g_{k+1}\|}{\mu\|s_k\|^2} + \frac{\|s_k\|^2 \|g_{k+1}\|}{\|s_k\|^2} \\ &\leq \|g_{k+1}\| \left(1 + \frac{L}{\mu} + 1 \right) \\ \|d_{k+1}\| &\leq \left(\frac{2\mu + L}{\mu} \right) \Gamma \end{aligned}$$

From the above relation we get:

$$\sum_{k \geq 1} \frac{1}{\|d_k\|^2} \geq \left(\frac{\mu}{2\mu + L} \right)^2 \frac{1}{\Gamma^2} \sum_{k \geq 1} 1 = \infty$$

Which is contradiction therefore $\lim_{k \rightarrow \infty} \|g_k\| = 0$

4. Numerical results and comparisons

In this section, we report some numerical results on 75 nonlinear unconstrained test problems. For each test problem, the dimension $n=100, \dots, 1000$. The Fortran77 expression of its function and gradient can be downloaded from N. Andrei's website: <http://www.ici.ro/camo/neculai/SCALCG/evalfg.for>.

The following CG methods in the form of (2) and (3), only different in the choice of the CG parameter, are test:

The Dai- Liao (DL) method [3]:

$$\beta^{DL} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - t \frac{s_k^T g_{k+1}}{d_k^T y_k}, \quad t = 1$$

1. The Kafaki- Ghanbari (KF1) method [2]:

$$\beta^{KF1} = \frac{y_k^T g_{k+1}}{s_k^T y_k} - \left(\frac{s_k^T y_k}{\|s_k\|^2} + \frac{\|y_k\|}{\|s_k\|} \right) \frac{s_k^T g_{k+1}}{s_k^T y_k}$$

2. The Kafaki- Ghanbari (KF2) method [2]:

$$\beta^{KF2} = \frac{y_k^T g_{k+1}}{s_k^T y_k} - \frac{\|y_k\|}{\|s_k\|} \frac{s_k^T g_{k+1}}{s_k^T y_k}$$

3. The Aynur and Abbo (AK1) method :

$$\beta^{AK1} = \frac{y_k^T g_{k+1}}{s_k^T y_k} - \frac{s_k^T g_{k+1}}{s_k^T s_k}$$

Here we utilize the source code Fortran 77 on N. Andrei’s website. All the parameters, including the parameters $\rho = 0.0001$, $\sigma = 0.9$, are set as default. The implementations are run on PC with 1.3 GHz CPU

processor and 760 MB RAM memory. We stop the iterations if the inequality $\|g_k\|_\infty \leq 10^{-6}$ is satisfied.

We adopt the performance profiles by Dolan and Moré [5] to compare the performance among the tested methods. For n_s and n_p problems, the performance profile $P : \mathfrak{R} \rightarrow [0,1]$ is defined as follows:

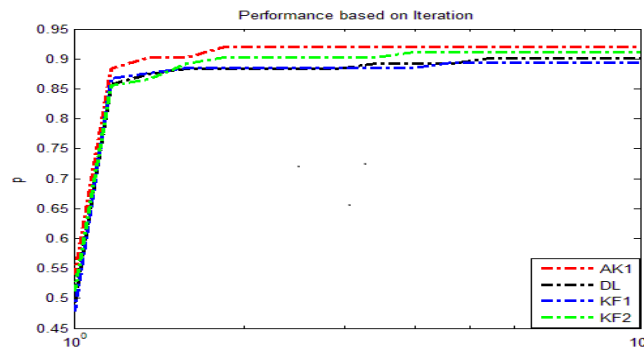
Let \mathbf{P} and \mathbf{S} be the set of problems and the set of solvers, respectively. For each problem $p \in \mathbf{P}$ and for each solver $s \in \mathbf{S}$, we define $t_{p,s} :=$ (computing time or (number of iterations, etc) required to solve problem p by solver s). The performance ratio is given by $r_{p,s} := t_{p,s} / \min_{s \in \mathbf{S}} t_{p,s}$. Then the performance profile is defined by:

$$P(\tau) = \frac{1}{n_p} \text{size} \{p \in \mathbf{P} : r_{p,s} \leq \tau\}, \quad \forall \tau \in \mathfrak{R}$$

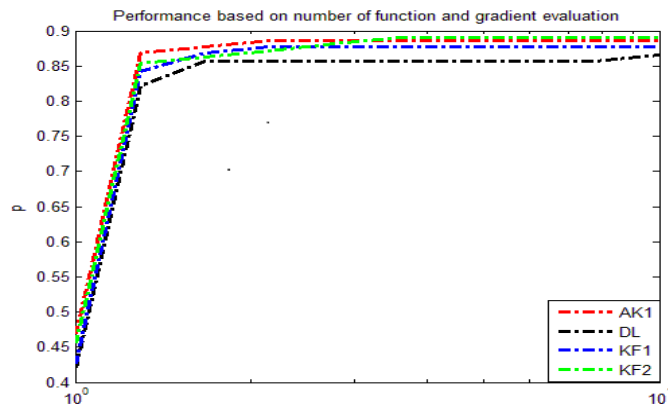
where size $\{p \in \mathbf{P} : r_{p,s} \leq \tau\}$ stands

For the number of elements of the set $\{p \in \mathbf{P} : r_{p,s} \leq \tau\}$. Note that if the performance profile of a method is over the performance profiles of the other methods, then this method performed better than the other methods.

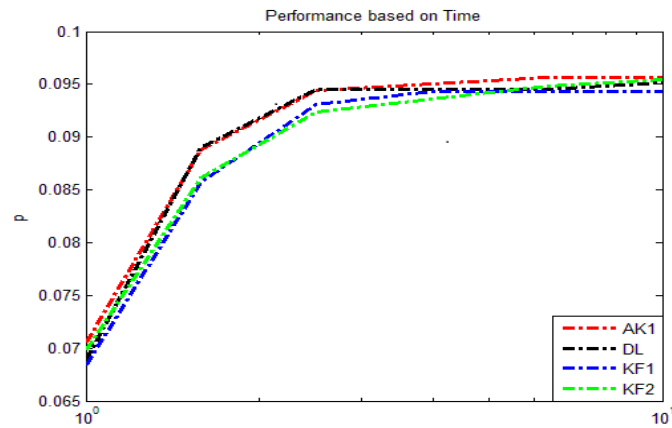
Figures 1-3 are the performance profiles measured by the number of iterations, the number of function and gradient evaluations, and CPU time respectively. From Figures1-3, we can observe that our proposed method (AK1) numerically outperforms with slight superiority to the other methods, since the figures graphically illustrate that the curves of AK1 are always the top performer for almost all values of τ . The possible reason is that our method suggests optimal value for the parameter t which is an open question.



Fig(1) Performance profile with respect number of iterations.



Fig(2) Performance profile with respect number of function gradient evaluations.



Fig(3) Performance profile with respect to the total number of CPU time.

Table(1) shows the comparison of the algorithms AK1, DL, KF1 and KF2 with respect to the total number of iteration(iter), total number of function and gradient evaluations (fg) and total required for solving 750 test problems.

Table(1) comparison of the algorithms

Algorithm	Total iter	Total fg	Total Time
AK1	136404	224772	2981
DL	137538	231532	3346
KF1	137694	231885	3369
KF2	136494	224853	2992

5.Conclusions

In this paper we have developed a new conjugate gradient method which is based on Dai-Liao and Kafaki-Ghanbari CG methods and generates sufficient descent search direction. Under suitable assumptions our method have been shown to converge globally.



Kirkuk University Journal /Scientific Studies (KUJSS)

Volume 12, Issue 2, March 2017

ISSN 1992 – 0849

References



Kirkuk University Journal /Scientific Studies (KUJSS)

Volume 12, Issue 2, March 2017

ISSN 1992 – 0849



- [1] N. Andrei, (2011). Open Problems in Nonlinear Conjugate Gradient Algorithms for Unconstrained Optimization. BULLETIN of the Malaysian Mathematical Sciences Society. 34, 319–330.
- [2] S. Babaie-Kafaki and R. Ghanbari, (2014). The Dai-Liao nonlinear conjugate gradient methods with optimal parameter choices, European Journal of Operational Research, 234, 625-630.
- [3] Y. H. Dai and L. Z. Liao, (2001). New Conjugacy Conditions and Related Nonlinear Conjugate Gradient Methods, Applied Mathematics and Optimization, Springer-Verlag, New York, USA, 43, PP. 87-101.
- [4] X. Dong, H. Liu and Y. He, (2015): A Self-Adjusting Conjugate Gradient Method with Sufficient Descent Condition and Conjugacy Condition. J. Optim Theory Appl 165.
- [5] E.D. Dolan and J.J. Mor'e ,(2002). "Benchmarking optimization software with performance profiles", Math. Programming, 91, pp. 201-213.
- [6] R. Fletcher and C.M. Reeves, (1964). Function Minimization by Conjugate Gradients. Computer Journal, 7, PP. 149-154.
- [7] M.R. Hestenes and E. Stiefel, (1952). Methods of Conjugate Gradients for Solving Linear Systems, Journal of Research of the National Bureau of Standards, Vol.(5), No.(49), pp. 409-436.
- [8] W. Hager and H. Zhang, (2006): A survey of nonlinear conjugate gradient methods. Pac. J. Optim. 2(1).
- [9] E. Polak and G. Ribière, (1969). Note Sur la Convergence de Directions Conjuguée. Revue Francaise Information, Recherche. Operationnelle, (16), pp.35-43.
- [10] M.J.D. Powell, (1984). Nonconvex Minimization Calculations and the Conjugate Gradient Method.in : Numerical Analysis (Dundee,1983),In Lecture Notes in mathematics, Springer-Verlag, Berlin, 1066, PP. 122-141.





- [11] W. Sun and Y. Yuan, (2006). Optimization Theory and Methods, Nonlinear programming, Springer Science, Business Media, LLC.,New York.
- [12] P. Wolfe, (1968). Convergence conditions for ascent methods , SIAM Review, 11, 226-235.
- [13] P. Wolfe, (1971). Convergence conditions for ascent methods, (II): some corrections, SIAM Review, 13, 185-188.
- [14] G. Zoutendijk, (1970). Nonlinear Programming, Computational Methods. Integer and Nonlinear Programming (J. Abadie ED.). North-Holland, Amsterdam, PP. 37-86.





Kirkuk University Journal /Scientific Studies (KUJSS)

Volume 12, Issue 2, March 2017

ISSN 1992 – 0849



Kirkuk University Journal /Scientific Studies (KUJSS)

Volume 12, Issue 2, March 2017

ISSN 1992 – 0849