



Tikrit Journal of Pure Science

ISSN: 1813 – 1662 (Print) --- E-ISSN: 2415 – 1726 (Online)

Journal Homepage: <http://tjps.tu.edu.iq/index.php/j>



Modified new conjugate gradient method for Unconstrained Optimization

Zeyad M. Abdullah¹, Hameed M. Sadeq², Hisham M. Azzam³, Mundher A. Khaleel⁴

¹ Department of Computer Science, College of Computer Science and Mathematics, University of Tikrit, Tikrit, Iraq

² Directorate of Education Nineveh, Mosul, Iraq

³ Mathematics Department, College of Computer Science and Mathematics, Mosul University, Mosul, Iraq

⁴ Department of Computer Science, College of Computer Science and Mathematics, University of Tikrit, Tikrit, Iraq

DOI: <http://dx.doi.org/10.25130/tjps.24.2019.095>

ARTICLE INFO.

Article history:

-Received: 29 / 5 / 2019

-Accepted: 24 / 6 / 2019

-Available online: / / 2019

Keywords: Search Direction, Conjugate Gradient, Optimization.

Corresponding Author:

Name: Zeyad M. Abdullah

E-mail: zeyamoh1978@tu.edu.iq

Tel:

1. Introduction

Minimizing a given objective function is considered as a problem arising in many practical situations, where a real n -dimensional vector exists. The unconstrained case represents a significant class of practical problems. This is due to some reasons, including: firstly, many constrained problems can be easily converted to and solved by methods of unconstrained optimization. Secondly, solving unconstrained sub problems of many problems of optimization is necessary. The generic unconstrained optimization problem is defined by this equation:

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

Classes of the so called GD methods are among the most natural and widely used class of algorithms of iterative descent.

Supposing that $f: R^n \rightarrow R$ represents “a continuously differentiable function”, then examine the problem of unconstrained optimization assumed in the first equation. Overall, finding a universal Min of f could be too ambitious [Bannas et al., 2006].

This paper only seeks the stationary points of f , i.e., the $x^* \in R^n$ point satisfying $g(x^*)=0$ to start. Suppose that $x_1 \in R^n$ is an initial estimation with $g(x_1) \neq 0$. For achieving progress, proceeding in a search direction $d_k \in R^n$ is required. For example, the iterates can be updated in accordance with:

ABSTRACT

The current paper modified method of conjugate gradient for solving problems of unconstrained optimization. The modified method convergence is achieved by assuming some hypotheses. The statistical results demonstrate that the modified method is efficient for solving problems of Unconstrained Nonlinear Optimization in comparison with methods FR and HS.

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

It is known that how far proceeding is achieved in the d_k direction is controlled by “ $\alpha_k > 0$ ”.

To solve the problem of non-linear unconstrained optimization given in equation (1), where “ $f: R^n \rightarrow R$ ” is “a continuously differentiable function” constrained from below beginning with “an initial guess $x_1 \in R^n$ ”, a sequence is generated by a CG method in accordance with (2) as well as generating the d_k directions, in this way:

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

There are several selections of β_k scalar (called “conjugacy parameter”), giving diverse implementation on “non-quadratic functions”; however, they are equal to “quadratic functions”. The CG algorithms contain a line search, which often depends on the conditions of strong or standard Wolfe [Andrei, 2007a]. These algorithms are required for ensuring convergence and for enhancing stability. They are defined by equations (2) and (3), where the β_k parameter is calculated using one of these ways:

$$\beta^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \quad (4)$$

$$\beta^{DY} = \frac{g_{k+1}^T g_{k+1}}{d_k^T y_k} \quad (5)$$

$$\beta^{CD} = \frac{-g_{k+1}^T g_{k+1}}{g_k^T d_k} \quad (6)$$

$$\beta^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \quad (7)$$

$$\beta^{PR} = \frac{g_{k+1}^T y_k}{g_k^T g_k} \quad (8)$$

$$\beta^{LS} = -\frac{g_{k+1}^T y_k}{g_k^T d_k} \quad (9)$$

Notice that these above algorithms could be categorized as algorithms with “ $g_{k+1}^T g_{k+1}$ ” in “the β_k numerator” and “algorithms with $g_{k+1}^T y_k$ in the β_k parameter numerator”. Fletcher and Reeves [1964] introduced the first algorithm of CG [(4) FR] for non-linear function. Dai and Yuan [1999] proposed the method of DYCG, which is defined in (5); while Fletcher [1987] introduced the method of CD conjugate descent as defined in (6). The algorithm $g_{k+1}^T g_{k+1}$ in the β_k numerator has strong convergence theory, but all these methods are susceptible to jamming. Few steps are made by these methods without any significant advancement to the minimum [Kinsella, 2009]. However, new methods were proposed. For example, Hestenes and Stiefel [1952] suggested the (HSCG) method as defined in (7); Polak and Ribiere [1969] developed the (PRCG) method as described in (8); and the (LSCG) method derived by Liu and Story [1991] as described in (9). A feature of integrated restart is found in methods with “ $g_{k+1}^T y_k$ ” in “the parameter β_k numerator”. This feature addresses the jamming phenomenon. If “ S_k step” is small, “ y_k factor” in “the β_k numerator” will tend to (0). So, “ β_k ” will be small and “the new d_{k+1} direction” in (3) is basically “SD direction – g_{k+1} ”. This means that “ β_k ” will be automatically adjusted by methods of “HS, PR and LS” to avoid jamming; adding to that the performance of these methods will be better than that of method with “ $g_{k+1}^T g_{k+1}$ ” in “the β_k numerator” [Dai and Yuan, 2003].

In the analysis of convergence and implementation of method of conjugate gradient, “the exact and inexact line search” like conditions of Wolfe or “strong Wolfe conditions” is often required. The α_k parameter is found by “the Wolfe line search”, so that $f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \dots (10)$
 $d_k^T g(x_k + \alpha_k d_k) \geq \sigma d_k^T g_k \dots (11)$
 with $0 < \delta < \sigma$. “The strong Wolfe line search” is to find “ α_k ”, so that

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \dots (12)$$

$$|d_k^T g(x_k + \alpha_k d_k)| \leq -\sigma d_k^T g_k \dots (13)$$

where $0 < \delta < \sigma < 1$ are constants [Li, Z. and Weijun, Z., 2008].

2. New conjugate gradient method:

The Dai and Liao’s algorithm of “conjugate gradient” is among the best methods to solve the problem of “large scale nonlinear optimization”.

As known, Perry in 1978 suggested this condition:

$$“d_{k+1} = -g_{k+1}^T s_k” \quad (14)$$

After that, Dai and Lia [2001] introduced the following condition of conjugacy:

$$“d_{k+1}^T y_k = -t g_{k+1}^T s_k” \quad (15)$$

Where $t \geq 0$ is a scalar.

For ensuring that the condition of conjugacy is satisfied by “the search direction d_{k+1} ” in (14), it is necessary to multiply (15) with y_k and utilize this new conjugacy condition (15). Hence, Dai and Liao obtained the following new formula for β_k

$$\beta_k = \frac{g_{k+1}^T y_k}{d_k^T y_k} - t \frac{g_{k+1}^T s_k}{d_k^T y_k} \dots (16)$$

In this case, in the vein of Dai and Liao’s [2001] formulas in (16), we also construct the following conjugate gradient formula.

$$\beta_k^{ZHH} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} - t \frac{(g_{k+1}^T y_k)^2}{2g_k^T y_k - \|g_{k+1}\|^2} \frac{s_k^T g_{k+1}}{g_k^T g_k} \dots (17)$$

Where $t \in (0,1)$, Assuming

$$\lambda = \frac{(g_{k+1}^T y_k)^2}{2g_k^T y_k - \|g_{k+1}\|^2} \frac{s_k^T g_{k+1}}{g_k^T g_k}$$

Then, we have

$$\beta_k^{ZHH} = \beta^{FR} - t \lambda$$

Therefore, the search direction for the new formula is

$$d_{k+1} = -g_{k+1} + \beta_k^{ZHH} d_k \dots (18)$$

Where β_k^{ZHH} is defined in the equation (17).

Now, the new algorithms of “conjugate gradient” can be obtained, in this way:

New Algorithm

The first step: Initialization: choose $x_1 \in R^n$ and the $0 < \delta_1 < \delta_2 < 1$

parameters. Then, calculate “ $f(x_1)$ ” and “ g_1 ”; next, consider “ $d_1 = -g_1$ ” and develop “the initial guess $\alpha_1 = 1/\|g_1\|$ ”.

The second step: Examine the iterations continuity. If “ $\|g_{k+1}\| \leq 10^{-6}$ ”, after that stop.

The third step: “Line search”: calculate “ $\alpha_{k+1} > 0$ ” to satisfy “the Wolfe line search” conditions in (12) & (13), then update the “ $x_{k+1} = x_k + \alpha_k d_k$ ” variables.

The fourth step: “ β_k ” conjugate gradient parameter, which is defined in (17).

The fifth step: computation of direction: calculate $d_{k+1} = -g_{k+1} + \beta_k d_k$. When satisfying “the restart

criterion of Powell $|g_{k+1}^T g_k| \geq 0.2 \|g_{k+1}\|^2$, then develop “ $d_{k+1} = -g_{k+1}$ ”; if not, then define “ $d_{k+1} = d$ ”. Calculate “the initial guess $\alpha_k = \alpha_{k-1} \|d_{k-1}\| / \|d_k\|$,” then develop “ $k = k + 1$ ” and continue with second step.

3. The Descent Property of the new formula

To show that “the search directions” of (18) are “descent directions”:

3.1. Theorem

Assume that the “line search” is satisfying “the conditions of Wolfe” in (12) and (13), then “ d_{k+1} ” given by (18) is “a descent direction”.

Proof

The proof is by induction.

1- If $k=1$ then $g_1^T d_1 < 0$ $d_1 = -g_1 \rightarrow < 0$.

2- Let the relation $g_k^T d_k < 0$ for all k .

3- The relation is proved to be true when $k = k + 1$; by multiplying the equation (18) in g_{k+1} , we get:

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -g_{k+1}^T g_{k+1} + \left(\frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} - t \frac{(g_{k+1}^T y_k)^2}{2g_k^T y_k - \|g_{k+1}\|^2} \frac{S_k^T g_{k+1}}{g_k^T g_k} \right) g_{k+1}^T S_k \\ g_{k+1}^T y_k &\leq L S_k^T g_{k+1} \\ g_{k+1}^T d_{k+1} &= -g_{k+1}^T g_{k+1} + \left(\frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} - t \frac{(L S_k^T g_{k+1})^2}{2L S_k^T g_{k+1} - \|g_{k+1}\|^2} \frac{S_k^T g_{k+1}}{g_k^T g_k} \right) g_{k+1}^T S_k \\ g_{k+1}^T d_{k+1} &= -g_{k+1}^T g_{k+1} + \left(\frac{g_{k+1}^T g_{k+1} g_{k+1}^T S_k}{g_k^T g_k} - t \frac{(L S_k^T g_{k+1})^4}{(2L S_k^T g_{k+1} - \|g_{k+1}\|^2) g_k^T g_k} \right) g_{k+1}^T S_k \\ g_{k+1}^T d_{k+1} &= -g_{k+1}^T g_{k+1} + \left(\frac{g_{k+1}^T S_k}{g_k^T g_k} - t \frac{(L S_k^T g_{k+1})^4}{(2L S_k^T g_{k+1} - \|g_{k+1}\|^2) g_k^T g_k} \right) g_{k+1}^T g_{k+1} \\ g_{k+1}^T d_{k+1} &= - \left(1 - \left(\frac{g_{k+1}^T S_k}{g_k^T g_k} - t \frac{(L S_k^T g_{k+1})^4}{(2L S_k^T g_{k+1} - \|g_{k+1}\|^2) g_k^T g_k} \right) \right) g_{k+1}^T g_{k+1} \\ c &= \left(1 - \left(\frac{g_{k+1}^T S_k}{g_k^T g_k} - t \frac{(L S_k^T g_{k+1})^4}{(2L S_k^T g_{k+1} - \|g_{k+1}\|^2) g_k^T g_k} \right) \right) g_{k+1}^T g_{k+1} \\ g_{k+1}^T d_{k+1} &< -c \|g_{k+1}\|^2 \dots \dots \dots (19) \end{aligned}$$

The First Assumption

Assume f is bounded below in the level set “

$S = \{x \in R^n : f(x) \leq f(x_0)\}$ ”. In some part N of “ S , f ” is “continuously differentiable” and “its gradient” is Lipschitz continuous, there exist $L > 0$, so that:

$$\|g(x) - g(y)\| \leq L \|x - y\| \quad \forall x, y \in N \dots (20)$$

4. Global Convergence Property

The convergence of suggested methods was studied by using “a uniformly convex function”. Then, there was a constant $\mu > 0$, so that

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \mu \|x - y\|^2, \text{ for any } x, y \in S$$

or equally

$$y_k^T S_k \geq \mu \|S_k\|^2 \text{ and } \mu \|S_k\|^2 \leq y_k^T S_k \leq L \|S_k\|^2 \dots (22)$$

On the other hand, under Assumption (1), it is obvious that there are “positive constants B ”, like

$$\|x\| \leq B, \quad \forall x \in S \quad (23)$$

4.1 Proposition

Under Assumption 1 and equation (23) on “ f ”, there is “a constant $\bar{\gamma} > 0$ ”, so that:

$$\|\nabla f(x)\| \leq \bar{\gamma}, \quad \forall x \in S \quad (24)$$

4.2 Lemma (1)

Assume that assumption (1) and equation (23) were conducted. Examine any method of conjugate gradient in forms (2) & (3), where “ d_k ” is “a descent direction” and “ α_k ” is obtained by “the strong Wolfe line search conditions”. If

$$\sum_{k=1}^{\infty} \frac{1}{\|d_{k+1}\|^2} = \infty \quad (25)$$

then we have

$$\lim_{k \rightarrow \infty} (\inf \|g_k\|) = 0. \quad (26)$$

More details can be found in [Dai.Y and Liao.L, 2001] and [Tomizuka.H and Yabe. H, 2004].

4.3 Theorem

Assume that Assumption (1) and equation (23) and the descent condition were conducted, consider the new algorithm (New), where α_k is calculated by “the conditions of Wolfe line search” (12) and (13). If the objective function is uniformly convex on S , then $\liminf_{k \rightarrow \infty} \|g_k\| = 0$.

Proof

$$\begin{aligned} d_{k+1} &= -g_{k+1} + \beta_k^{ZHH} S_k \\ \|d_{k+1}\| &= \|-g_{k+1} + \beta_k^{ZHH} S_k\| \\ \|d_{k+1}\| &= \left\| -g_{k+1} + \left(\frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} - t \frac{(g_{k+1}^T y_k)^2}{2g_k^T y_k - \|g_{k+1}\|^2} \frac{S_k^T g_{k+1}}{g_k^T g_k} \right) S_k \right\| \\ \|d_{k+1}\| &\leq \|g_{k+1}\| + \left(\frac{\|g_{k+1}\|^2}{\|g_k\|^2} - t \frac{\|g_{k+1}^T y_k\|^2}{2\|g_{k+1}\| \|g_k\|^2 + \|g_k\|^2 \|g_{k+1}\|^2} \right) \|S_k\| \\ \|d_{k+1}\| &\leq \|g_{k+1}\| + \left(\frac{\|g_{k+1}\|^2 \|S_k\|}{\|g_k\|^2} - t \frac{\|g_{k+1}\|^2 \|y_k\| \|g_k\| + \|g_k\|^2 \|g_{k+1}\| \|S_k\|}{2\|g_{k+1}\| \|y_k\| \|g_k\| + \|g_k\|^2 \|g_{k+1}\|^2} \right) \\ \|d_{k+1}\| &\leq \|g_{k+1}\| + \left(\frac{\|g_{k+1}\|^2 \|S_k\|}{\|g_k\|^2} - t \frac{\|g_{k+1}\|^2 \|y_k\|^2 \|S_k\|^2}{2\|y_k\| \|g_k\|^2 + \|g_k\|^2 \|g_{k+1}\|^2} \right) \\ \|d_{k+1}\| &\leq \left(1 + \left(\frac{\|g_{k+1}\|^2 \|S_k\|}{\|g_k\|^2} - t \frac{\|g_{k+1}\|^2 \|y_k\|^2 \|S_k\|^2}{2\|y_k\| \|g_k\|^2 + \|g_k\|^2 \|g_{k+1}\|^2} \right) \right) \|g_{k+1}\| \\ M &= \left(1 + \left(\frac{\|g_{k+1}\|^2 \|S_k\|}{\|g_k\|^2} - t \frac{\|g_{k+1}\|^2 \|y_k\|^2 \|S_k\|^2}{2\|y_k\| \|g_k\|^2 + \|g_k\|^2 \|g_{k+1}\|^2} \right) \right) \\ \|d_{k+1}\| &\leq M \|g_{k+1}\| \\ \|d_{k+1}\| &\leq M \bar{\gamma}^{-2} \\ \|d_{k+1}\| &\leq \frac{1}{\bar{\gamma}^2} M \left(\bar{\gamma}^{-2} \right)^2 \\ L &= M \left(\bar{\gamma}^{-2} \right)^2 \\ \|d_{k+1}\| &\leq L \frac{1}{\bar{\gamma}^2} \\ \sum_{k=1}^{\infty} \frac{1}{\|d_k\|} &\geq \frac{1}{L} \bar{\gamma}^{-2} \sum_{k=1}^{\infty} 1 = \infty \dots (27) \end{aligned}$$

Therefore, we have $\lim_{k \rightarrow \infty} \|g_{k+1}\| = 0$

5. Numerical results and comparisons

This section compares the new CG method performance to another classical conjugate gradient method (FR and HS methods). The (75) large scale unconstrained optimization problem was selected for each test problem taken from [Andrei, 2008]. There are statistical results for each test function with the variables number $n = 100, \dots, 1000$. The new algorithm is compared with the well-known FR and HS algorithms of conjugate gradient. All these algorithms were implemented with “strong Wolfe line search conditions” (12) & (13). The stopping criteria in all of these cases is $\|g_k\| = 10^{-6}$. Writing of all codes was done using “double precision FORTRAN Language with F77 default compiler settings”. The test functions usually begin a point standard. Initially, a summary of numerical results was recorded in the figures (1), (2), (3). The performance profile presented by Dolan. E. D and J. J. Moré [2002] is used to display the performance of the new developed CG method of conjugate gradient algorithm in comparison to FR and HS algorithms. $p = 750$ is defined as “the whole set of n_p test problems”; while $S = 3$ refers to “the set of the interested solvers”. Suppose that “ $l_{p,s}$ ” is “the evaluations number of objective function”, which “solver S ” requires them for “problem p ”, then define the ratio of performance as

$$r_{p,s} = \frac{l_{p,s}}{l_p^*} \quad (28)$$

Where “ $l_p^* = \min\{l_{p,s} : s \in S\}$ ”. Obviously, $r_{p,s} \geq 1$ for all p, s . When a problem is not solved by a solver, the $r_{p,s}$ ratio is considered as a large number M .

Each solver S has a performance profile, which is defined below in the function of cumulative distribution for the ratio $r_{p,s}$ of performance,

$$\rho_s(\tau) = \frac{\text{size}\{p \in P : r_{p,s} \leq \tau\}}{n_p} \quad (29)$$

Obviously, $\rho_s(1)$ refers to the problems percentage for which solver S is the best. The study of Dolan. E. D and J. J. Moré [2002] illustrates the performance profile in more details. It is possible to

utilize this performance profile in analyzing the iterations number, the gradient evaluations number and the cpu time. In addition, for clear observation, “the horizontal coordinate” is given “a log-scale” as shown in these figures:

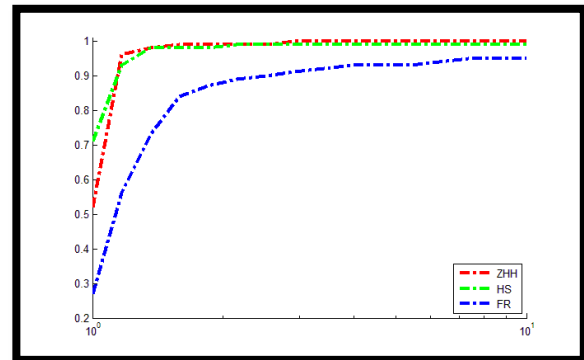


Figure (1): Performance based on iteration

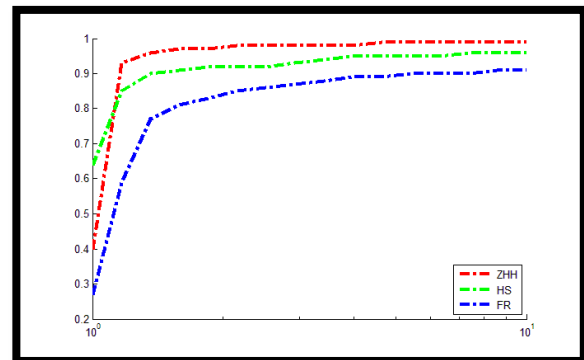


Figure (2): Performance based on Function

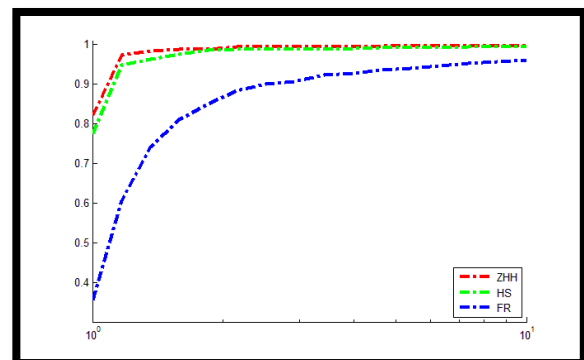


Figure (3): Performance based on Time

References

- [1] Andrei, N. (2008). "An unconstrained optimization test functions collection". Adv. Model. Optim, 10(1), 147-161.
- [2] Andrei, N. (2007a), "Numerical comparison of conjugate gradient algorithms for unconstrained optimization", Studies in Informatics and Control, 16.
- [3] Bannas, J., Gilbert, J., Lemarechal, C. and Sagastizable, G. (2006), "Numerical Optimization, Second Eddition. Springer-Verlag Brlin".
- [4] Dai, Y. and Yuan, Y. (1999), "A Nonlinear conjugate gradient method with a strong global convergence property", SIAM J, Optim, 10.
- [5] Dai, Y. and Yuan, Y. (2003), "A class of globally convergent CG methods", Sci. china Ser. A, 46.
- [6] Dai. Y. H and Liao. L. Z, (2001), "New conjugacy condition and related nonlinear conjugate gradient methods", Applied Mathematics and Optimization 43, pp. 87-101.
- [7] Dolan. E. D and J. J. Mor'e, (2001), "Benchmarking optimization software with performance profiles", Math. Programming, 91 pp. 201-213.
- [8] Fletcher, R. (1987), "Practical Methods For Optimization", John Wiley & Sons Lth.
- [9] Fletcher, R. and Reeves, G. (1964), "Function minimization by gradients computer", Journal, Vol(7).
- [10] Hestenes, M and Stiefel, E. (1952), "Methods of CG for solving linear Systems", J. Research Nat. Bar . standards Sec . B . 49.
- [11] Kinsella, J . (2009), "Course Notes for MS4327 optimization" <http://jkcray.Maths.ul.ie/ms4327/slides.pdf> 2009.
- [12] Li, Z. and Weijun, Z. (2008), "Tow descent hybrid conjugate gradient methods for optimization", Journal of computational and Appl. Math. 216, pp251-264.
- [13] Liu, D. and Story, C. (1991), "Efficient generalized CG algorithms", part I : Theory, Journal of Optimization Theory and Applications 69.
- [14] Perry, A. (1978). "A modified conjugate gradient algorithm. Operations Research", 26(6), 1073-1078..
- [15] Polak, E. and Ribiere, G. (1969), "Not sur la convergence de directions conjugate". Rev. Franaisse Informants. Research operational, 3e Anne., 16.
- [16] Tomizuka, H and Yabe, H, (2004), "A Hybrid Conjugate Gradient method for unconstrained Optimization", February, 2004, revised July, 2004.

تطوير طريقة جديدة للتدرج المترافق في الامثلية اللامقيدة

زياد محمد عبدالله¹ ، حميد محمد صادق² ، هشام محمد عزام³ ، منذر عبدالله خليل⁴

¹ قسم علوم الحاسوب ، كلية علوم الحاسوب والرياضيات ، جامعة تكريت ، تكريت ، العراق

² مديرية تربية نينوى ، الموصل ، العراق

³ قسم الرياضيات ، كلية علوم الحاسوب والرياضيات ، جامعة الموصل ، الموصل ، العراق

⁴ قسم علوم الحاسوب ، كلية علوم الحاسوب والرياضيات ، جامعة تكريت ، تكريت ، العراق

الملخص

في البحث الحالي تم تطوير طريقة جديدة للتدرج المترافق لحل مشاكل الامثلية الغير المقيد. تم تحقيق التقارب للطريقة المطورة الجديد من خلال افتراض بعض الفرضيات. توضح النتائج العددية أن الطريقة المطورة الجديدة فعالة في حل مسائل الامثلية الغير مقيدة الخطية مقارنة بالطرق ل HS و FR.