The Convergence Properties of Conjugate Gradient Method Using AMRI Parameter with Exact Line Search

LAILY DWI RETNO WAHYUNINGTIAS[1], SALMAH[1*]
[1]Department of Mathematics, Universitas Gadjah Mada
Yogyakarta 55281, INDONESIA

Abstract: The conjugate gradient method is a simple way to get a solution of optimization problems without constraints. In this work, we offer a conjugate gradient method algorithm using AMRI parameter where the step of length is decided by an exact line search. The proposed algorithm accomplishes the condition of sufficiently descent and globally convergence with several assumptions. Computation results prove that the modified method is effective.

Key-Words:-conjugate gradient method, conjugate parameter, globally convergent, sufficient descent condition.

Received: September 8, 2022. Revised: March 11, 2023. Accepted: April 22, 2023. Published: May 31, 2023.

1 Introduction

Solving various optimization problems is often an important topic for engineers and scientists. One easy-to-use optimization technique is the conjugate gradient method. For example, at Yuan et al. [1] discuss the image processing problem which can be formulated into an optimization problem. Abubakar et al. [2] discuss the signal improvement problems that can be solved by constructing optimization problem and Helmig et al. [3], create an optimization problem to estimate the distance and the number of sensors in the inverse calculation of temperature boundary conditions, and others (see [4], [5], [6], [7] and [8]). Its application include optimization problems, with and without constraints.

The method that we discuss in this paper is the conjugate gradient method. This method is a helpful and convenient way to get solution of unconstrained optimization problems because it takes less memory and is easier to compute. This method determines the iteration solution direction through the objective function's gradient, the conjugate parameter, and the search direction of the previous iteration. The development of conjugate gradient methods is very diverse, especially in modifying conjugate gradient parameters.

The first conjugate gradient parameter introduced is FR parameter [9]. Powell [10] further investigated FR parameter and found that this parameter with exact line search can generate small step length without giving significant results to the optimal solution. Then, Polak and Ribiere in [11] introduce PR parameters. Numerically, the conjugate gradient method with PR parameter has better performance than FR parameter. Polak and Ribiere [11] showed that the PR parameter under exact line search for convex objective functions yield global convergent properties. Nevertheless, Powell [12] has also shown that this is not necessarily the case for non-convex function. In addition, Powell [12] found that PR parameter by exact line search, can rotate indefinitely and do not approach the solution point. This behavior can occur when the conjugate gradient parameter is negative so Powell [12] suggests that the conjugate gradient parameter non-negative. Therefore, Gilbert and Nocedal [13] modified the PR parameter to the maximum value between the number zero and the PR parameter and obtained global convergence results with inexact line search.

Rivaie et. al. [14] modified the PR parameter by changing the numerator from the norm of gradient of the objective function to the norm of search direction. These are called RMIL parameters hereafter. In general, RMIL parameter is better than PR parameter because RMIL parameter can solves more test functions in optimization problems. However, the RMIL parameter is not necessarily negative, so Dai [15] modified the parameter so that if the RMIL parameter is negative, the RMIL parameter value is set to zero. With this modification, a globally convergent method is obtained.

Another parameter that is a modification of RMIL is the AMRI parameter [16]. This parameter always has a positive or zero value and it is clear that this parameter accomplishes the condition of sufficiently descent with inexact line search. Based on the paper's numerical results, the AMRI parameter is better than the RMIL parameters.

In this work, the AMRI parameter with exact line search are also proved to satisfy sufficient descent conditions. Moreover, this AMRI parameter-based conjugate gradient method meets the property of global convergence with an exact line search step length. Several numerical computations and comparisons between methods are also presented in this paper.

The systematics of writing this work is given as follows: Some definitions and assumptions related to this study are given in section 2. AMRI parameter and its algorithm are described in section 3. We will provide a convergence analysis in section 4. It includes condition of sufficiently descent and globally convergent. Numerical results and comparisons 1between methods with AMRI parameter and existing parameters are showed in section 5. As a finale, given the conclusions presented in section 6.

2 Preliminaries

We consider optimization problem without constraints below:

$$\min_{\mathrm{x}\in\mathbb{R}^n} f(\mathrm{x}) \qquad (1)$$

with $f:\mathbb{R}^n \to \mathbb{R}$ is continuous and differentiable function and $\mathbb{R}$ is real numbers. The method discussed here is the modified conjugate gradient where this method is an algorithm for numerical solutions that is often implemented as an iterative algorithm of the form

$$\mathrm{x}_{k+1} = \mathrm{x}_k + \alpha_k \mathrm{d}_k, k = 0,1,2,... \qquad (2)$$

where x_k is the solution point for the *k*th iteration, $\alpha_k > 0$ is length of step and notation d_k is the search direction. Length of step is determined by a one-dimensional search called a line search. The most commonly used is the exact line search, or

$$f(\mathrm{x}_k + \alpha_k \mathrm{d}_k) = \min_{\alpha\geq 0} f(\mathrm{x}_k + \alpha \mathrm{d}_k) \qquad (3)$$

The most frequently used length of step is the exact line search. This is due to its ability to produce optimal step length [17]. In 2015, research showed that modern technology with faster processors and better tools solves the speed issues often encountered by equation (3), as indicated in Rivaie et al. [18].

The formulae for search direction d_k is

$$\mathrm{d}_k = \begin{cases} -g_k, & k = 0 \\ -g_k + \beta_k \mathrm{d}_{k-1}, & k \geq 1, \end{cases} \qquad (4)$$

where g_k is the gradient of the function f at x_k and β_k is the parameter of the conjugate gradient. As mentioned in section 1, there are several well-known conjugation parameter formulas such as FR parameter, PR parameter and RMIL parameter below.

$$\beta_k^{FR} = \frac{g_k^T g_k}{\|g_{k-1}\|^2} \qquad (5)$$

$$\beta_k^{PR} = \frac{g_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2} \qquad (6)$$

$$\beta_k^{RMIL} = \frac{g_k^T (g_k - g_{k-1})}{\|\mathrm{d}_{k-1}\|^2} \qquad (7)$$

where $\|.\|$ is the norm of euclid. The equation **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.-Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.** became known as Fletcher and Reeves (FR) in [9], Polak and Ribiere (PR) in [11], and Rivaie, Mustafa, Ismail and Leong (RMIL) in [14] respectively. There are two important properties for analyzing the convergence properties of the conjugate gradient method, namely sufficiently descend and globally convergent [19].

Definition 1: An algorithm is said to sufficiently descent if there is $C > 0$ for any d_k

$$g_k^T d_k \leq -C\|g_k\|^2, \qquad \forall k \geq 0. \qquad (8)$$

Definition 2: An algorithm of conjugate gradient method is globally convergent if

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \qquad (9)$$

The following assumptions are required to get the prove of equation (9) of this method.

Assumption 1: Function f on the level set $\mathbb{R}^n$ is bounded below and is continuous and differentiable at the starting point in the neighborhood $\mathcal{B}$ of the level set $\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$.

Assumption 2: The gradient g is a Lipschitz continuous function in $\mathcal{B}$, in other words there exist $K > 0$, such that $\|g(x) - g(y)\| \leq K\|x - y\|$ for any $x, y \in \mathcal{B}$.

DOI: 10.37394/232030.2023.2.5

A lemma is obtained by Assumption 1 and Assumption 2, hereinafter referred to as the Zoutendijk condition. The lemma also applies when the step length is specified with inexact line search.

Lemma 1: For any conjugate gradient method with (2)-(4), where α_k as formulated by exact line search (3). Under Assumption 1 and Assumption 2 holds will satisfy

$$\sum_{k=0}^{\infty} \frac{\left(g_k^T \mathrm{d}_k\right)^2}{\|\mathrm{d}_k\|^2} < \infty. \qquad (10)$$

Zoutendijk has proved this lemma in [20].

3 Conjugate Gradient Method With AMRI Parameter

In this section, the conjugate gradient method algorithm is formed with the AMRI parameter and equation (3). The AMRI parameter used is the conjugation parameter proposed by Abashar, Mamat, Rivaie, Ismail, and Omer in the form

$$\beta_k^{AMRI} = \frac{g_k^T \left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}\right)}{\|\mathrm{d}_{k-1}\|^2}. \qquad (11)$$

The algorithm is given as follows based on (2), (3), (4), and (11).

1) Initialization. Given $\mathrm{x}_0 \in \mathbb{R}^n, \varepsilon > 0$ and set $k = 0$.
2) Calculate $\|g_k\|$, if $\|g_k\| \leq \varepsilon$, then x_k is solution point. If $\|g_k\| > \varepsilon$, go to step (3).
3) Calculate β_k based on (11).
4) Calculate d_k based on (4).
5) Calculate step length α_k by equation (3).
6) Set $k = k + 1$ and calculate next step by equation (2) and move to step (2).

4 Convergent Analysis

In this section, we will prove that the conjugate gradient method with the AMRI conjugation parameter accomplishes condition of sufficiently descent and global convergence. To discuss these two matters, we give a property that shows that the AMRI parameter is always non-negative.

Lemma 2 : Given the conjugate gradient method algorithm with the AMRI parameter. For every $k \geq 0$ applies $\beta_k^{AMRI} \geq 0$

Proof : Based on the conjugation gradient method algorithm with the AMRI parameter, if $k = 0$, then $\beta_k^{AMRI} = 0 \geq 0$ applies. If $k \geq 0$, then using the Cauchy-Schwartz inequality, we have

$$\begin{aligned}
\beta_k^{AMRI} &= \frac{g_{k+1}\left(g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k\right)}{\|d_k\|^2}. \\
&= \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} g_{k+1}^T g_k}{\|d_k\|^2} \\
&\geq \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} \|g_{k+1}\| \|g_k\|}{\|d_k\|^2} \\
&= 0
\end{aligned} \qquad (12)$$

■

Theorem 1 below will show that the method with the AMRI parameter accomplishes the condition of sufficiently descent.

Theorem 1: For a conjugate gradient method with d_k formulated by (4) and parameter β_k^{AMRI} formulated by (11), then

$$g_k^T \mathrm{d}_k \leq -C\|g_k\|^2, \qquad (13)$$

where $C > 0$.

Proof : We prove that the conjugate gradient method with the AMRI parameter accomplishes (13). Note that

$$\mathrm{d}_k = \begin{cases} -g_k, & k = 0 \\ -g_k + \beta_k^{AMRI} \mathrm{d}_{k-1}, & k \geq 1, \end{cases} \qquad (14)$$

If $k = 0$, then we get

$$g_o^T \mathrm{d}_0 = -g_0^T g_0 = -\|g_0\|^2. \qquad (15)$$

In other words, the inequality (13) is satisfied for $k = 0$. Furthermore, if $k \geq 1$, then multiply (14) with g_k^T, we get

$$\begin{aligned}
g_k^T \mathrm{d}_k &= -g_k^T(-g_k + \beta_k^{AMRI} \mathrm{d}_{k-1}) \\
&= -\|g_k\|^2 + \beta_k^{AMRI} g_k^T \mathrm{d}_{k-1}
\end{aligned}$$

Apply exact line search method, we get $g_k^T d_{k-1} = 0$. Consequently,

$$g_k^T \mathrm{d}_k = -\|g_k\|^2.$$

This means that d_{k+1} is sufficiently descen. Therefore inequation (13) applies. We have finished the proof.■

Next, we will prove that the conjugate gradient method with AMRI parameter accomplishes the global convergence properties, in other words it is proved to satisfy equation (9). To analyze the global convergence of this modified method, it is necessary to use the Assumption 1, Assumption 2, and Lemma 1.

Theorem 2 : Assume that Assumption 1, Assumption 2, and Theorem 1 satisfied. Let the conjugate gradient method with equation (2)-(4), where β_k is computed by (11) and α_k is formulated by equation (3). If $\|\alpha_k \mathrm{d}_k\| \to 0$ for $k \to \infty$, then

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \quad (16)$$

Proof : Suppose θ_k be the angle between $-g_k$ and d_k with $\cos\theta_k = \frac{-g_k^T \mathrm{d}_k}{\|g_k\|\|\mathrm{d}_k\|}$. With equation (3) and equation (4), we have

$$\|\mathrm{d}_k\| = \sec\theta_k \|g_k\|. \quad (17)$$

Using the fact that $g_k^T \mathrm{d}_{k-1} = 0$, we get

$$\beta_{k+1}\|\mathrm{d}_k\| = \tan\theta_{k+1}\|g_{k+1}\|. \quad (18)$$

Furthermore, by combining (17) and (18), we get

$$\tan\theta_{k+1} = \beta_{k+1}^{AMRI} \sec \frac{\|g_k\|}{\|g_{k+1}\|}$$

$$= \sec\theta_k \frac{\|g_k\|}{\|g_{k+1}\|}\left(\frac{g_{k+1}\left(g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k\right)}{\|d_k\|^2}\right)$$

$$\leq \sec\theta_k \frac{\|g_k\|}{\|g_{k+1}\|} \frac{\|g_{k+1}\| \left\|g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k\right\|}{\|d_k\|^2}$$

$$= \sec\ \theta_k \frac{\|g_k\| \left\|g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k\right\|}{\|d_k\|^2}$$

With exact line search and (4), we get

$$\tan\theta_{k+1} \leq \sec\theta_k \frac{\|g_k\| \left\|g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k\right\|}{\|g_k\|^2 + \beta_k^{AMRI}\|\mathrm{d}_{k-1}\|^2}.$$

Based on Lemma 2, we have

$$\tan\theta_{k+1} \leq \sec\theta_k \frac{\|g_k\| \left\|g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k\right\|}{\|g_k\|^2}$$

$$= \sec\theta_k \frac{\left\|g_{k+1} - g_k + g_k \frac{\|g_{k+1}\|}{\|g_k\|} g_k\right\|}{\|g_k\|}$$

$$\leq \sec\theta_k \frac{\|g_{(k+1)} - g_k\| + |\|g_k\| - \|g_{k+1}\||}{\|g_k\|}$$

$$\leq \sec\theta_k \frac{2\|g_{k+1} - g_k\|}{\|g_k\|} \quad (19)$$

If (16) does not hold, then for every $k \geq 0$ there is $\gamma > 0$ such that

$$\|g_k\| \geq \gamma. \quad (20)$$

Because $\|\alpha_k \mathrm{d}_k\| \to 0$ and from Assumption (2) g is a Lipschitz function, there exist a non-negative integer M such that for every $k \geq M$, holds

$$\|g_{k+1} - g_k\| \leq \frac{1}{4}\gamma. \quad (21)$$

By (19)-(21) we have

$$\tan\theta_{k+1} \leq \frac{1}{2}\sec\theta_k. \quad (22)$$

Note that for every $\theta \in [0, \pi/2)$, the following inequation holds

$$\sec\theta \leq 1 + \tan\theta. \quad (23)$$

Inequation (22) and (23), induce

$$\tan\theta_{k+1} \leq \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots + \left(\frac{1}{2}\right)^{k+1-m} (1 + \tan\theta_m)$$

$$\leq 1 + \tan\theta_m \quad (24)$$

This show that the angle θ_k is always less than angle fixed $\bar{\theta}$ which is less than $\pi/2$. By Lemma 1, we get

$$\sum_{k=0}^{\infty} \frac{\left(g_k^T \mathrm{d}_k\right)^2}{\|\mathrm{d}_k\|^2} = \sum_{k=0}^{\infty} \|g_k\|^2 (\cos\theta_k)^2 < \infty. \quad (25)$$

This implicitly means that $\lim_{k\to\infty} \inf\|g_k\| = 0$, which contradicts (20). We have finished the proof. ■

5 Numerical Implementation and Discussions

In this section, we provide numerical implementation of the conjugate gradient method using FR, PR, RMIL, and AMRI parameters to show each method's efficiency. Some of the test functions used are taken from Andrei [21] starting from low, medium, to high dimensions as in Rivaie et al. [14], namely 2, 4, 10, 50, 100, 500. Criteria for stopping iteration based on $\|g_k\| \leq \varepsilon$ where $\varepsilon = 10^{-6}$. From the starting point that is closest to the optimal point to the starting point that is farthest from the optimal point, each test function utilizes a different set of starting points. Table 1 give the list of test functions and starting points.

Table 1. Test function's list.

No.	Function	n	Starting point
1	Six-hump	2	$(8,8), (-8,-8), (10,10)$
2	Three-hump	2	$(-1,1), (1,-1), (-2,2)$
3	Booth	2	$(4,4), (8,8), (16,16)$
4	Goldstein-Price	2	$(2,-2), (9,9), (15,15)$
6	Zettl	2	$(3,3), (5,5), (7,7)$
7	Cube	2	$(5,5), (10,10), (20,20)$
8	Rosenbrock	2	$(-2,-2), (2,2), (11,11)$
9	Quartic	4	$(2,\ldots,2), (5,\ldots,5), (10,\ldots,10)$
10	Ext. Maratos	2,4	$(8,\ldots,8), (22,\ldots,22), (44,\ldots,44)$
11	Ext. White and Holst	4,10	$(2,\ldots,2), (3,\ldots,3), (-2,\ldots,-2)$
12	Ext. Frudenstein and Roth	4,100	$(3,\ldots,3), (5,\ldots,5), (10,\ldots,10)$
13	Beale	2,4,10	$(2,\ldots,2), (4,\ldots,4), (6,\ldots,6)$
14	Raydan1	2,4,10	$(-1,\ldots,-1), (1,\ldots,1), (2,\ldots,2)$
15	Liarwhd	2,4,10	$(3,\ldots,3), (5,\ldots,5), (7,\ldots,7)$
16	Fletcher	2,4,10	$(5,\ldots,5), (10,\ldots,10), (40,\ldots,40)$
17	Edencsch	2,4,10	$(3,\ldots,3), (23,\ldots,23), (43,\ldots,43)$
18	Gen.Quartic	2,4,100	$(1,\ldots,1), (10,\ldots,10), (20,\ldots,20)$
19	Ext. Denschnf	2,4,100	$(2,\ldots,2), (13,\ldots,13), (50,\ldots,50)$
20	Ext. Denschnb	2,4,100	$(4,\ldots,4), (8,\ldots,8), (15,\ldots,15)$
21	Himmelblau	2,10,100	$(15,\ldots,15), (25,\ldots,25), (35,\ldots,35)$
22	Ext. Penalty	2,10,100	$(2,\ldots,2), (5,\ldots,5), (10,\ldots,10)$
23	Tridiagonal1	2,10,500	$(5,\ldots,5), (7,\ldots,7), (15,\ldots,15)$

Table 2. Numerical experiments's summary.

Method	Total of NI	Total of CPU time	Successful
FR	14,344	2944.653350	98%
PRP	1,647	1396.406475	93%
RMIL	2,720	1710.859500	98%
AMRI	2,182	1461.500225	100%

The test functions that is given in Table 1 are completed by Matlab R2020a and done on a laptop with specifications; Intel(R) Celeron(R) processor, 4.00 GB RAM, 64-bit Windows 10 Operating System Home Single Language. If the test function solved by a method produces a negative step length, then that method's function is included in the failed category. Iteration number (NI) and the time taken by the CPU are compared. Table 2 will give summary of numerical results. From these numerical results, a performance representation is obtained based on NI and CPU time shown respectively in Fig. 1 and Fig. 2. The result is a performance representation curve established by Dolan and More [22].

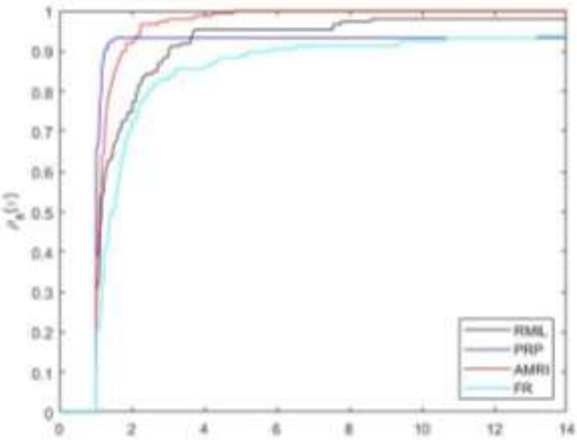

Fig. 1. Performance representation dependent on iteration count.

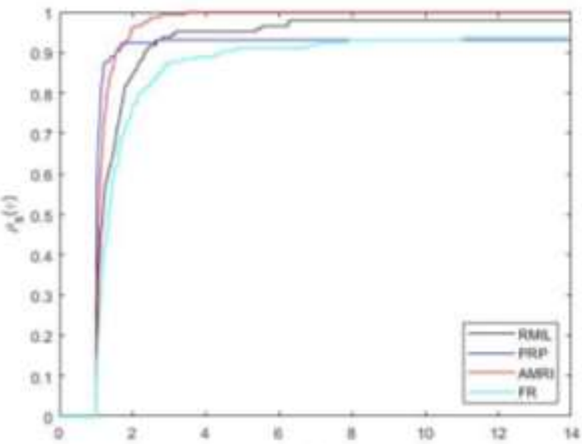

Fig. 2. Performance representation dependent on the time taken by CPU

Dolan and More [22] established how evaluation and comparation of the performance of each method with various test functions. Let S is a collection of n_S solver and P is a collection of n_P test functions. For each $s \in S$ solver and $p \in P$, is defined $r_{p,s}$, which is the iteration's number or time the CPU needs to complete function p with solver s. Performance solver s in the test function p is compared with the best performance of the existing method with the same test function, or it can be written as follows

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s}: s \in S\}}.$$

Then $r_{p,s}$ is called performance rate. Choose a constant r_M with $r_M \geq r_{p,s}$ for each $p \in P$ and $s \in S$ and $r_{p,s} = r_M$ if and only if solver s failed to solve function p.

It is interesting to compare the performance of each solver against a test function, but the author wants to obtain an overall comparison. If defined

$$\rho_s(\tau) = \frac{1}{n_p} \text{size}\{p \in P : r_{p,s} \leq \tau\},$$

then $\rho_s(\tau)$ is the probability of the solver $s \in S$ where the performance rate $r_{p,s}$ is in the $\tau \in \mathbb{R}$ factor of the best possible rate. In general, the method with a high $\rho_s(\tau)$ score or the curve position on the top right is the best solver.

Fig. 1 and Fig. 2 respectively represent the execution of method with FR, PR, RMIL and AMRI parameter based on the iteration number and the entire time spent by the CPU in solving each test function. The curve of the method with the AMRI parameter is at the top left to right compared to other parameters. Although the PRP method previously achieved a higher probability value than the AMRI method, there are several functions that the PRP method has been unable to complete. The AMRI method can complete the entire test function, the RMIL method can complete the 98% test function, and the PRP method can complete the 93% test function. The FR method can solve the 98% test function, but the iteration is high enough, so its performance is not better than the other methods. Because of these results, we can say that the method with AMRI parameter is a method capable of achieving better performance than FR, PR and RMIL parameter.

6 Conclusion

This paper aims to initiate an algorithm of conjugate gradient method with AMRI parameter using exact line search step length. The AMRI method accomplishes two important properties, namely sufficiently descent and globally convergence with exact line search. Based on numerical experiments on the test functions, it has shown that the conjugate gradient method with AMRI parameters is an efficient method.

Acknowledgments:
The author would like to thank the Direktorat Riset, Teknologi, dan Pengabdian Kepada Masyarakat, Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia for funding and support under (089/E5/PG.02.00.PT/2022;1957/UN1/DITLIT/Dit-Lit/PT.01.03/2022)

References:

[1] G. Yuan, T. Li and W. Hu, A conjugate gradient algorithm for large scale nonlinear equation and image restoration problems, *Applied Numerical Mathematics,* vol. 147, 2020, pp. 129-141.

[2] A. B. Abubakar, P. Kumam and A. M. Awwal, Global convergence via descent modified three-term conjugate gradient projection algorithm with applications to signal recovery, *Results in Applied Mathematics, vol. 4,* 2019, p. 100069.

[3] T. Helmig, F. Al-Sibai and R. Kneer, Estimating sensor number and spacing for inverse calculation of thermal boundary conditions using the conjugate gradient method, *International Journal of Heat and Mass Transfer,* vol. 153, 2020, p. 119638.

[4] J. Cao and J. Wu, A conjugate gradient algorithm and its applications in image restoration, *Applied Numerical Mathematics,* vol. 152, 2020, pp. 243-252.

[5] A. H. Ibrahim, P. Kumam, A. B. Abubakar, W. Jirakitpuwapat and J. Abubakar, A hybrid conjugate gradient algorithm for constrained monotone equations with application in compressive sensing, *Heliyon,* vol. 6, no. 3, 2020, p. e03466.

[6] I. A. R. Moghrabi, A New Preconditioned Conjugate Gradient Method for Optimization, *IAENG International Journal of Applied Mathematics,* vol. 49, no. 1, 2019, pp. 29-36.

[7] O. Kardani, A. V. Lyamin and K. Krabbenhoft, A Comparative Study of Preconditioning Techniques for Large Sparse Systems Arising in Finite Element Limit Analysis, *IAENG International Journal of Applied Mathematics,* vol. 43, no. 4, 2013, pp. 195-203.

[8] D. Kumar, S. Gupta and P. Sehgal, Improved Training of Predictive ANN with Gradient Techniques, in *Lecture Notes in Engineering and Computer Science : Proceedings of The International MultiConference of Engineers and Computer Scientists 2014,* Hong Kong.

[9] R. Fletcher and C. M. Reeves, Function minimization by conjugate gradients, *The Computer Journal,* vol. 7, no. 2, 1964, pp. 149-154.

[10] M. J. D. Powell, Restart Procedures for the Conjugate Gradient, *Mathematical Programming,* vol. 12, 1977, pp. 241-254.

[11] E. Polak and G. Ribiere, Note sur la convergence de m´ethodes de directions conjugu´ees, *ESAIM: Mathematical Modelling and Numerical Analysis-Mod´elisation Math´ematique et Analyse Num´erique,* vol. 3, no. R1, 1969, pp. 35-43.
[12] M. J. D. Powell, Nonconvex minimization calculations and the conjugate gradient method, *Lecture Notes in Mathematics*, vol. 1066, 1984, pp. 122-141.
[13] J. C. Gilbert and J. Nocedal, Global Convergence Properties of Conjugate Gradient Methods for Optimization, *SIAM J. Optimizat*, vol. 2, no. 1, 1992, pp. 21-42.
[14] M. Rivaie, M. Mamat, L. W. June and I. Mohd, A new class of nonlinear conjugate gradient coefficient with global convergence properties, *Appl. Math. comput.*, vol. 218, 2012, pp. 11323-11332.
[15] Z. Dai, Comments on a new class of nonlinear conjugate gradient coefficients with global convergence properties, *Applied Mathematics and Computation*, vol. 276, 2016, pp. 297-300.
[16] A. Abashar, M. Mamat, M. Rivaie, I. Mohd and O. Omer, The Proof of Sufficient Descent Condition for a New Type of Conjugate Gradient Methods, *AIP Conf. Proc.*, 1602, 2014, pp. 296-303.
[17] M. A. Hery, M. Mamat and L. W. June, BFGS Method : A New Search Direction, *Sains Malaysiana*, vol. 40, no. 10, 2014, pp. 1591-1597.
[18] M. Rivaie, M. Mamat and A. Abashar, new class of nonlinear conjugate gradient coefficients with exact and inexact line searches, *Appl. Math. Comput.*, vol. 268, 2015, pp. 1152-1163.
[19] M. Malik, M. Mamat, S. S. Abas, I. M. Sulaiman and Sukono, A New Modification of NPRP Conjugate Gradient Method for Unconstrained Optimization, *Advances in Mathematics : Scientific Journal*, vol. 9, no. 7, 2020, pp. 4955-4970.
[20] G. Zoutendijk, Nonlinear programming, computational methods, *Integer and Nonlinear Programming*, 1970, pp. 37-86.
[21] N. Andrei, An Unconstrained Optimization Test Functions Collection, *Advanced Modeling and Optimization, Volume 10, Number 1*, vol. 10, no. 1, 2008, pp. 147-161.
[22] E. D. Dolan and J. J. More, Benchmarking optimization software with performance profiles, *Math. Prog.*, vol. 91, 2002, pp. 201-213.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflicts of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.