MATHEMATICAL MODELING AND COMPUTING, Vol. 10, No. 2, pp. 326–337 (2023)

Performance evaluation of a novel Conjugate Gradient Method for training feed forward neural network

Kamilu K.[1], Sulaiman M. I.[2,3], Muhammad A. L.[1], Mohamad A. W.[4], Mamat M.[5]

[1]*Department of Mathematical Science, Faculty of Computing and Mathematics, Kano University of Science and Technology, 713101, Wudil, Nigeria*
[2]*School of Quantitative Sciences, Universiti Utara Malaysia, 06010, Kedah, Malaysia*
[3]*Institute of Strategic Industrial Decision Modelling (ISIDM), SQS, Universiti Utara Malaysia, Sintok, 06010, Kedah, Malaysia*
[4]*School of Dental Sciences, Universiti Sains Malaysia, Health Campus (USM), 16150, Kubang Kerian, Kelantan, Malaysia*
[5]*Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia*

(Received 29 June 2022; Revised 28 January 2023; Accepted 5 February 2023)

In this paper, we construct a new conjugate gradient method for solving unconstrained optimization problems. The proposed method satisfies the sufficient decent property irrespective of the line search and the global convergence was established under some suitable. Further, the new method was used to train different sets of data via a feed forward neural network. Results obtained show that the proposed algorithm significantly reduces the computational time by speeding up the directional minimization with a faster convergence rate.

Keywords: *conjugate gradient method; neural network; line search; convergence analysis.*
2010 MSC: 90C30, 90C26, 90C06, 90C90, 90C47 **DOI:** 10.23939/mmc2023.02.326

1. Introduction

This study will consider the following model

$$\min\{f(x)\colon x \in \mathbb{R}^n\}, \tag{1}$$

where f is a smooth function defined by $f\colon \mathbb{R}^n \to \mathbb{R}$ whose gradient $g(x) = \nabla f(x)$ is always available [1].

Problem in the form (1) can be found in various specialized disciplines such as computer science, machine learning, neural network, engineering, statistics and many more (see [2–7]). For simplicity, the following abbreviations $\nabla(f(x_k))$ and $f(x_k)$ would be represented by ∇_k and f_k throughout this study and $\|\ \|$ denote the Euclidean norm of vectors.

In recent years, different types of numerical algorithms have been developed for solving (1), however, the conjugate gradient (CG) method has been considered as the most preferred, because of its low memory needs, good convergence features and simple implementation, especially when solving large-scale problems [8, 9]. Like many optimization algorithms, CG method builds the sequence of iterate using the following recursive computational scheme:

$$w_{k+1} = w_k + \sigma_k d_k, \tag{2}$$

where w_k and w_{k+1} are the present and the next iteration points, respectively, σ_k is the learning rate obtained using either exact or inexact line search approaches [10]. The exact line search requires computing σ_k such that the cost function is minimized along the search direction d_k. This procedure is costly and time consuming, thus, many studies considered use the inexact line search approaches such as Armijo line search (WP), backtracking and the Standard Wolfe line search (SWP) [11]. The SWP is computed such as σ_k satisfies

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) + \mu\alpha_k \nabla f(x_k)^T d_k,$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geqslant \sigma \nabla f(x_k)^T d_k, \tag{3}$$

with $0 < \mu < \sigma < 1$. In some cases, a constant value, like $\sigma \in [0, 1]$, is set as the learning rate [12]. Recently, Sun and Zhang [13] proposed a new approach to replace the line search procedure discussed above. The outcome from the study was very encouraging. For detail discussion of this see Sun et al. [13] and Wu [14]).

Effective line search procedure and search direction d_k play important role in the convergence analysis of the CG methods. For this reason, [15] highlighted the importance of starting the process with the steepest descent direction, i.e. $d_0 = -g_0$ [16], otherwise, the convergence rate will be linear even for a strongly convex quadratic function [17]. The successive directions are obtained in a predefined sequence, defined as

$$d_k = -\nabla_k + \beta_k d_{k-1} \quad \text{for} \quad k \geqslant 1. \tag{4}$$

The choice of β_k lead to four classes of CG method, namely, scaled CG method [3], three term CG method [18, 19], classical CG method (two term) [11, 20] and hybrid CG method [17]. All the classes were developed to improve either the convergence or computational efficiency of the classical CG algorithm [5].

Discussion on the convergence of the well known classical CG method like Hestenes–Stiefel (HS) [21], Polak–Ribiere–Polyak (PRP) [22, 23], Fletcher–Reeves (FR) [24], Liu–Storey (LS) [25], conjugate descent (CD) [26], and Dai–Yuan (DY) [27] have been provided by many researchers [11, 28]. The PRP formula is considered to be the most effective in terms of numerical computation but its convergence under the several line search is not guaranteed [5]. This drawback led to various modifications of the PRP parameter. See Yuan et al. [29], Andrei [28], and Zhang et al. [30] for more detailed discussions about the PRP method.

Recently, Rivaie et al. [5] defined a new denominator for the PRP method named the RMIL method and discussed the convergence analysis of the method under both exact line search and inexact (Strong Wolfe) line searches. However, Dai [31] raised concern about its convergence and suggested that the convergence can only be valid if the CG parameter is restricted as follows:

$$\beta^{RMIL+} = \begin{cases} \frac{g_{k+1}^T (g_{k+1} - g_k)}{\|d_k\|^2} & \text{if } g_{k+1}^T g_k \leqslant \|g_{k+1}\|^2, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Yousif [32] investigated this method under the Wolfe line search while Sulaiman et al., [33] presented a new three-term direction using the above suggestion. The restriction above follows from the work of Gilbert and Nocedal [34] on PRP method which states that if PRP method is restricted to be positive and the learning rate is obtained by the line search strategy satisfying the sufficient descent condition:

$$g_k^T d_k \leqslant \mu \|g_k\|^2 \quad \mu > 0, \tag{6}$$

then the method would be globally convergent [23]. Based on this result, Wei et al. [35] proposed the WYL parameter with formula given as

$$\beta^{WYL} = \frac{g_k^T \left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1} \right)}{\|g_{k-1}\|^2}. \tag{7}$$

The study shows that the parameter is always positive in addition to satisfying the important property mentioned by Gilbert and Nocedel [34]. Various modifications of (7) satisfying the descent property based on inexact line search have been provided (see [36]). One of the recent modifications is the work of Zabidin et al. [37] with the parameter defined as

$$\beta^{A1} = \begin{cases} \frac{\|g_k\|^2 - \mu |g_k^T g_{k-1}|}{m |g_k^T d_{k-1}| + \|g_{k-1}\|^2} & \text{if } \|g_k\|^2 > \mu_k |g_k^T g_{k-1}|, \\ \mu_k \frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}} & \text{otherwise,} \end{cases} \tag{8}$$

where $\mu = \frac{\|s_{k-1}\|}{\|y_{k-1}\|}$.

Based on the above trend, it is obvious that the modifications of the classical PRP CG method are generally centred around changing the denominator or numerator. As indicated by Andrei [11, 28],

most of the modifications were defined by authors that a new denominator performs well but the methods often consist of the previous search direction in the conjugate parameter. Kamilu et al. [12] also construct a new denominator for the PRP coefficient which does not contain the search direction as

$$\beta_k^{KMAR} = \frac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T(g_k + g_{k-1})}. \tag{9}$$

This new parameter has been found to be efficient under exact line search [12, 19, 38]. Another three-term CG method based on the parameter has also been developed for unconstrained optimization and image restoration [39].

Motivated by the above trend, this paper presents an improvement of (9) and applies the modification to solving various data set in feed forward neural network. The remaining part of this paper is as follows: Section 2 contains the derivation process of the new conjugate parameter with detailed description of it algorithm. A sufficient descent condition and global convergence properties of the method are discussed in Section 3. Section 4 consists of the numerical result generated by testing the new method on some benchmark test problems. Lastly, the method was extended to solve real-life application problems in feed forward neural network.

2. New method and algorithm

Consider the HS [21] parameter defined as

$$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})}. \tag{10}$$

Let the denominator for $k = 1$ be defined as

$$M = d_{k-1}^T(g_k - g_{k-1}).$$

Substituting the initial search direction will give

$$\begin{aligned} M &= -g_{k-1}^T(g_k - g_{k-1}) \\ &= g_{k-1}^T g_{k-1} - g_{k-1}^T g_k \\ &= -g_{k-1}^T(g_k + g_{k-1}). \end{aligned}$$

Algorithm 1 KMAR + and KMAR++ methods.

Require: Initial point $x_0 \in \mathbb{R}^n$ for $k = 0$;
Ensure: $\varepsilon_0 > 0$;
1: Compute $g(x_0)$ set $d_0 = -g_0$, $\varepsilon = \varepsilon_0$ and $k = 0$
2: Check if $\|g_k\| \leqslant \varepsilon$ then stop.
3: Compute the learning rate using (3)
4: Update the new point using (2)
5: Compute β_k using (11) or (12) and update d_k using (4)
6: Set $k = k + 1$, Go to Step 2

To modify the KMAR parameter, so as there is sufficient decrease in the function value, a number of iteration and CPU time, and, above all, to make the parameter useful in the neural network, we present the following two methods based on KMAR,

$$\beta^{KMAR^+} = \begin{cases} \frac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T(g_k + g_{k-1})} & \text{if } \|g_k\|^2 > \mu_k |g_k^T g_{k-1}|, \\ 0 & \text{otherwise,} \end{cases} \tag{11}$$

$$\beta^{KMAR^{++}} = \begin{cases} \frac{g_k^T\left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}\right)}{g_{k-1}^T(g_k + g_{k-1})} & \text{if } \|g_k\|^2 > \mu_k |g_k^T g_{k-1}|, \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

The new denominator employed in the new method is expected to play a significant role in the convergence analysis. Note, this alteration differentiates the new method with classical PRP and WYL method. Next, the algorithm of the new method called Algorithm 1 is presented below. The order of the Algorithm was adapted from the work of Andrei [28] and Riviea et al. [5].

3. Convergence result

We begin this section by showing that the new parameter satisfies the sufficient descent property. From [7], it follows that KMAR parameter will reduce to the following:

$$0 \leqslant \beta_k^{KMAR} \leqslant \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \tag{13}$$

which would play an important role in the convergence analysis of the proposed method.

For a new CG method to be considered efficient, it must be able to possess the sufficient descent property (6) and converge under certain conditions [15]. The proof presented in this section will be supported by numerical results generated using different test functions in subsequent sections. The convergence discussion will begin with the following assumption on the objective function.

Lemma 1. *Let the sequence g_n and d_n be generated by the KMAR+ and step length by the Strong Wolfe line search, then*

$$d_k^T g_k < -\theta \|g_k\|^2, \quad \forall\, k \geqslant 0. \tag{14}$$

Proof. For $k = 0$, we have

$$\begin{aligned} d_0 &= -g_0, \\ g_0^T d_0 &= -g_0^T g_0 \\ &= -\|g_0^2\|. \end{aligned}$$

For $k \geqslant 1$, we have from (4), (11), and (13)

$$\begin{aligned} d_k &= -g_k + \beta_k^{KMAR} d_{k-1} \\ &= -g_k + \frac{\|g_k\|^2}{\|g_{k-1}\|^2} d_{k-1}. \end{aligned}$$

Multiply both sides by g_k^T

$$d_k^T g_k = -\|g_k\|^2 + \frac{\|g_k\|^2}{\|g_{k-1}\|^2} g_k^T d_{k-1}. \tag{15}$$

Factor $\|g_k\|^2$ in RHS

$$d_k^T g_k = -\left[1 - \frac{\|g_k\| \|d_{k-1}\|}{\|g_{k-1}\|^2}\right] \|g_k\|^2, \tag{16}$$

hence (14) holds where $\theta = \left[1 - \frac{\delta w}{\delta}\right]$. ∎

Assumptions A

1. The function f is bounded below on the level set $Y = \{x \in \mathbb{R}^n \colon f(x) \leqslant f(x_0)\}$, where $x - 0$ is the starting guess.
2. f is smooth in some neighbourhood N of Y and its gradient $g(x) = \nabla f(x)$ is Lipschitz continuous, i.e., there exists a constant $H > 0$ such that $\|g(x) - g(y)\| \leqslant H\|x - y\|\ \forall x, y \in N$.

Lemma 2. *Let Assumption A be true, using the new parameter, define in Algorithm A, where d_k is a descent direction and the step length α_k satisfying the standard Wolfe condition, then*

$$\sum_{k=0}^{\infty} \frac{\nabla f(x_k)^T d_k}{\|d_k\|^2} < +\infty. \tag{17}$$

Lemma 3. *Let Assumption A hold and the parameter KMAR+ generate the sequence* $\{x_k, d_k, \alpha_k, g_k\}$. *Then there exists a constant* $\theta > 0$ *such that*

$$\alpha_k \geqslant \theta \quad \forall k \geqslant 1. \tag{18}$$

Proof. The Lipschitz condition and first inequalities of standard Wolfe line search provide us with the following

$$\begin{aligned} \alpha_k H &\geqslant (g_k - g_{k+1})^T d_{k-1} \\ &\geqslant -(1-\sigma) g_k^T d_k \\ &\geqslant (1-\sigma)\|g_k\|^2, \end{aligned}$$

$$\alpha_k \geqslant \frac{1-\sigma}{H}\frac{\|g_k\|^2}{\|d_k\|^2} \geqslant \frac{1-\sigma}{H\gamma}.$$

Let $\lambda \in \left(0, \frac{1-\sigma}{H\gamma}\right)$. ∎

Theorem 1. *Suppose Assumption A is true, consider Algorithm A with the step length satisfying the standard Wolfe condition, then*

$$\lim_{k\to\infty} \inf \|g_k\|^2 = 0. \tag{19}$$

Proof. If $k = 0$ then the statement holds. Suppose that (18) is not true, then there exists a constant $\varepsilon > 0$ such that

$$\|g_k\| \geqslant \varepsilon \quad \forall k. \tag{20}$$

From (8), we have

$$\begin{aligned} d_k &= -g_k + \beta_k d_{k-1} \\ &\leqslant \|g_k\| + \frac{\|g_k\|^2}{\|g_{k-1}\|^2} d_{k-1}, \\ \|d_k\|^2 &\leqslant \|g_k\|^2 + \frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_{k-1}\|^2, \\ \frac{\|d_k\|^2}{\|g_k\|^4} &= \frac{1}{\|g_k\|^2} + \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} \\ &\leqslant \sum_{n=0}^{k-1} \frac{1}{\|g_k\|^2} \\ &\leqslant \frac{k}{\varepsilon^2}, \end{aligned} \tag{21}$$

therefore, (18) implies

$$\sum_{k\geqslant 1} \frac{\|g_k\|^4}{\|d_k\|^2} \geqslant \varepsilon^2 \sum_{k\geqslant 1} \frac{1}{k} = +\infty,$$

which contradicts Lemma 2, hence

$$\lim_{k\to\infty} \inf \|g_k\| = 0,$$

which completes the convergence proof. ∎

4. Numerical results

The performance of KMAR+ and KMAR++ was investigated using the standard test function from Andrei [11], with various initial points ranging from 2 to 10000. A comparison is made with β^{RMIL+} and β^{AMRI} based on a number of iterations and CPU time.

A personal computer Intel Core i3-3217u 4GB DDR3 Memory 500 GB HDD was used to run each algorithm after being coded on Matlab R2015b software. $\|g_k\| \leqslant \varepsilon$ is set as the stopping condition or the iteration would be terminated when the number of iteration exceeds 1000. A summary of the

performance based on a number of iterations and CPU time is presented in Tables 1 and 2 respectively. To further analyse the results, we employed a well-known performance profile tool introduced by Dolan and More [40]. Details of the profile can be obtained in [5]. Figure 1 shows the performance in terms of the number of iterations and CPU time.

Table 1. Performance Analysis Based On the Number of Iterations.

CG Method	RMIL+	KMAR+	KMAR++	AMRI+
Number of Success	17	79	51	15
Percentage of Success	0.1604	0.7453	0.4811	0.1415
Number of Failure	24	0	0	24
Percentage of Failure	0.7736	1	1	0.7736

Table 2. Performance Analysis Based On the CPU Time.

CG Method	RMIL+	KMAR+	KMAR++	AMRI+
Number of Success	0	32	59	17
Percentage of Success	0	0.3019	0.5566	0.1604
Number of Failure	24	0	0	24
Percentage of Failure	0.7736	1	1	0.7736

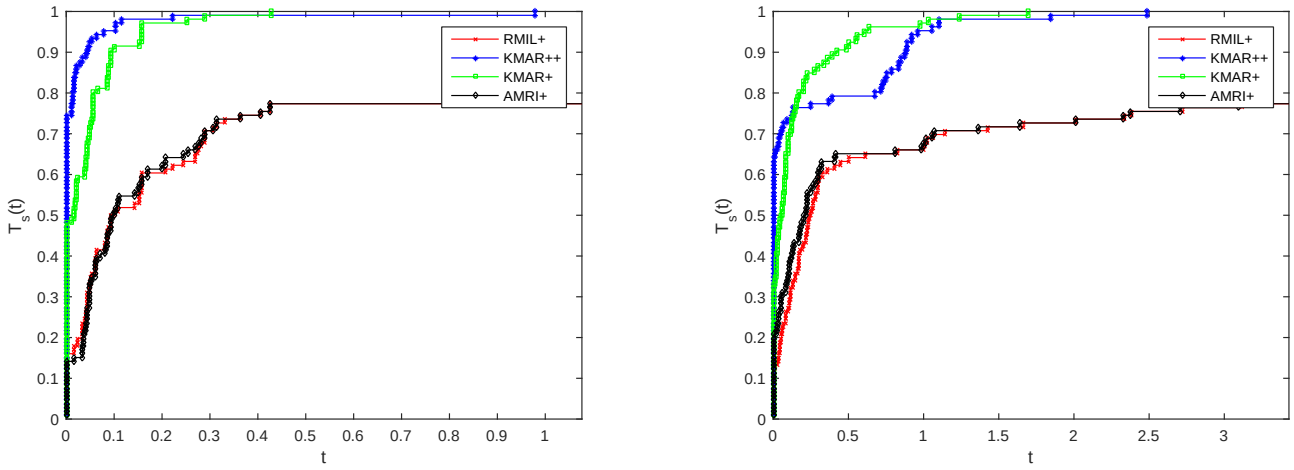

Fig. 1. Performance profile outputs for RMIL+, KMAR+, KMAR++ and AMRI+ based on NOI (left) and CPU time (right).

5. Application in neural network

Artificial Intelligence (AI) was designed to imitate the human brain [41]. There are various branches of AI, popular among them is artificial neural network (ANN) [15]. This class is used in classifying, optimizing, or prediction of a given set of data/information to give an appropriate results or output. Training and testing of the data are the main stages in ANN. It is carried out in order to give the room for the ANN to understand the pattern of data even when data set is incomplete [42, 43]

In the literature, various learning methods have been designed for training set of data, function minimization, and pattern recognition. A lot of these methods are based on the gradient descent method [42, 43]. One of the shortfall of this method is the bad convergence rate and poor computation results. To address this shortfall different Quasi-Newton methods have been employed and the outcomes is remarkable [41, 44]. However, this procedure requires the use of the inverse of the Hessian matrix or its approximation, which requires a lot of storage and, therefore, cannot handle large-scale data set.

To address this shortfall, this work applied a new CG method to train the set of data. The choice of CG method due to the fact that it does not require any Hessian evaluation or its approximation

during the iteration process. Also, the CG method can handle large-scale of data set. In what follows, we present a CG based NN algorithm based on the ideas of [41–43].

1. The given vector u_0 is transformed into the output vector u_k by solving the following equation
$$u_i^l(k) = f(x_i^l) = f\left(\sum_{j=1}^{n_{l-1}} v_{ij}^l u_j^{l-1} + b_i^l\right).$$
2. The error that is the difference between the desired output and actual output is obtained using
$$\delta_i^L(k) = f'(x_i^L)(d_i - u_i^L).$$
3. Based on the following formula, we propagate the error signal at the output units backwards through the whole network
$$\delta_j^{l-1}(k) = f'(x_j^{l-1}) \sum_{i=1}^{n_l} \delta_i^l v_{ij}^l.$$
4. Learning update using CG search direction based on Algorithm 1.

To illustrate the performance of the proposed methods, their algorithms were coded in Matlab program. The performance was compared with the classical training function like traincgf (FR) and traincgp (PR) using the default parameters throughout.

6. Experiment and results

1. For the first problem, the study considers the Chemical sensor data set from the neural tool box. The network architecture for this problem contains the one hidden layer with 10 neurons and a single output layer. All the parameters use the default values as mentioned in the NN tool and a maximum of 1000 iterations is set as the termination condition. For 100 simulation, the performance based on the min number of iterations (epochs), maximum number of iteration (epochs) and the percentage of the success of the algorithms are reported in Table 3. Figure 2 presents the regression analysis of each of the training function for KMAR+ and FR methods while Figure 3 illustrates the training performance. Also, Figure 4 presents the regression analysis and training performance of each of the training function for PR method.

Table 3. Simulation Performance for Chemical Sensor Data Set.

Training Function	Min Epoch	Max Epoch	Succ
FR	17	97	100 %
PR	16	76	100 %
New	11	65	100 %

2. The second problem we consider is the Body fat percentage data set from a neural network tool box. The network architecture for this problem contains the one hidden layer with 50 neurons and a single output layer. All the parameters use the default values as mentioned in the NN tool and a maximum of 1000 iterations is set as the termination condition. For 100 simulations, the performance based on the min number of iteration (epochs), maximum number of iteration (epochs) and the percentage of the success of the algorithms are reported in Table 4. Figures 5 - 7 demonstrate the training and validation performance of each of the training function for KMAR+, FR, and PR methods, respectively. Based on these results, it is obvious that the proposed method is efficient and can further find applications in other fields.

Table 4. Simulation Performance for Body Fat Percentage Data Set.

Training Function	Min Epoch	Max Epoch	Succ
FR	19	87	100 %
PR	18	46	100 %
New	10	41	100 %

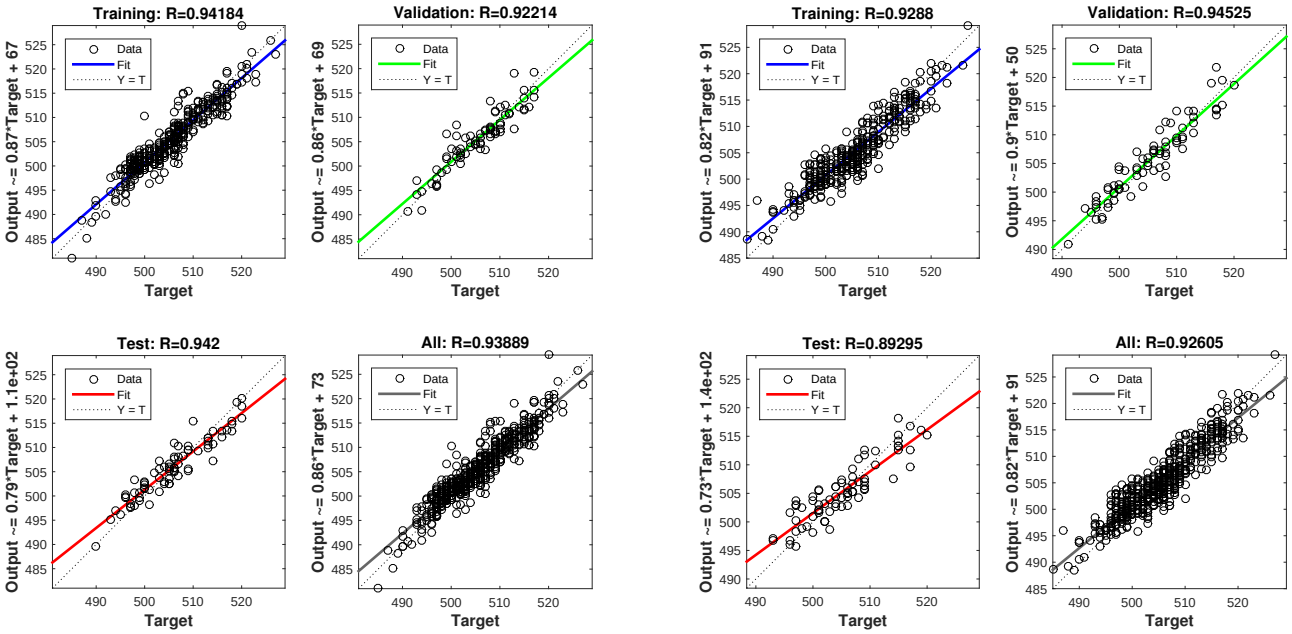

Fig. 2. Regression Analysis for KMAR+ method (left) and FR method (right).

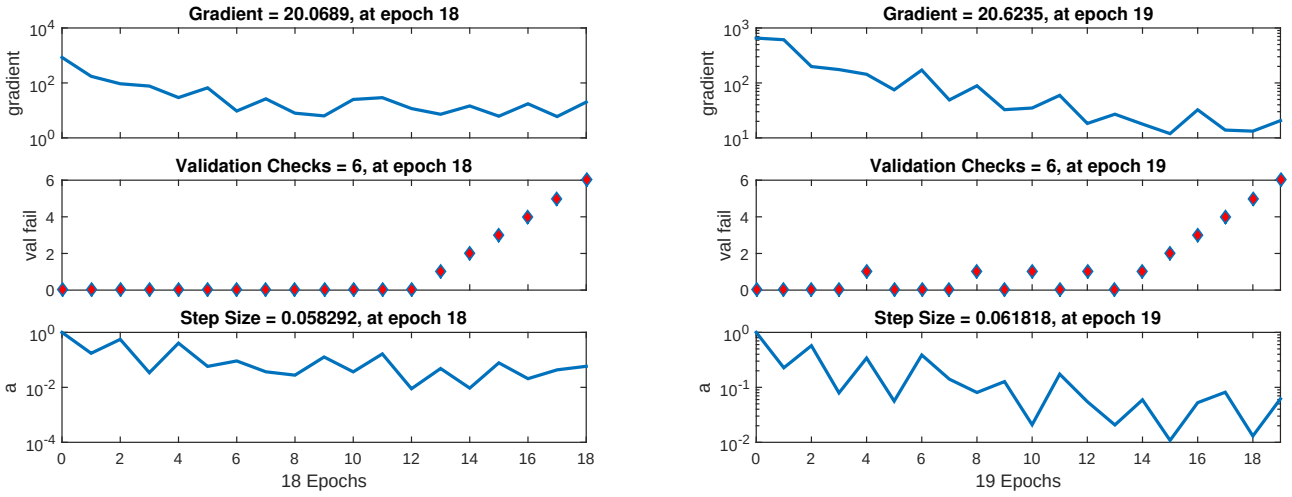

Fig. 3. Training Performance KMAR+ method (left) and FR method (right).

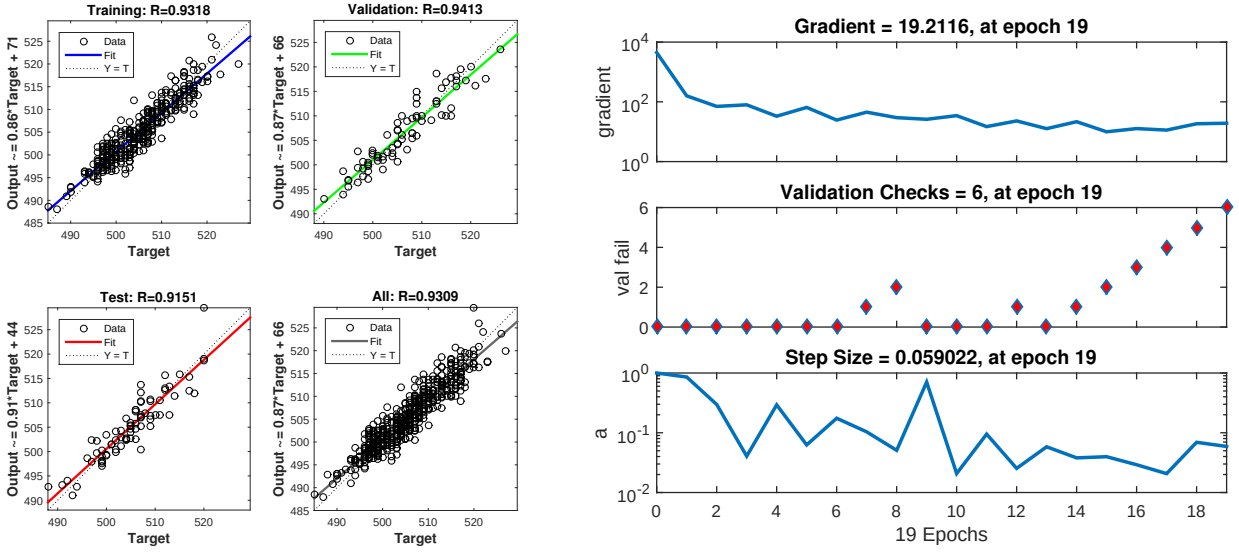

Fig. 4. Regression Analysis (left) and Training Performance (right) of PR method.

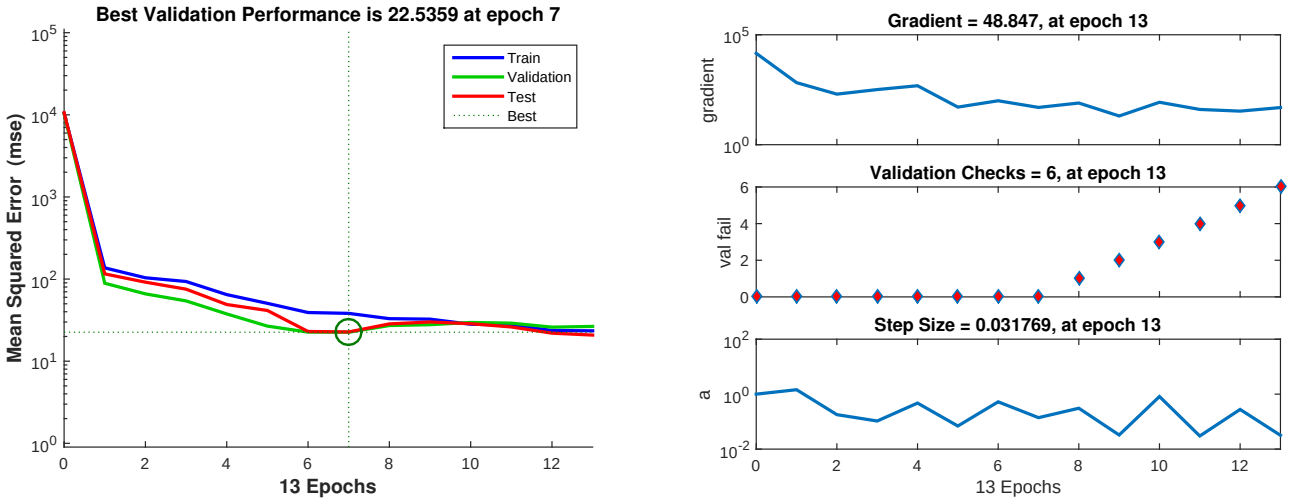

Fig. 5. Validation Performance (left) and Training Performance (right) of KMAR+ method.

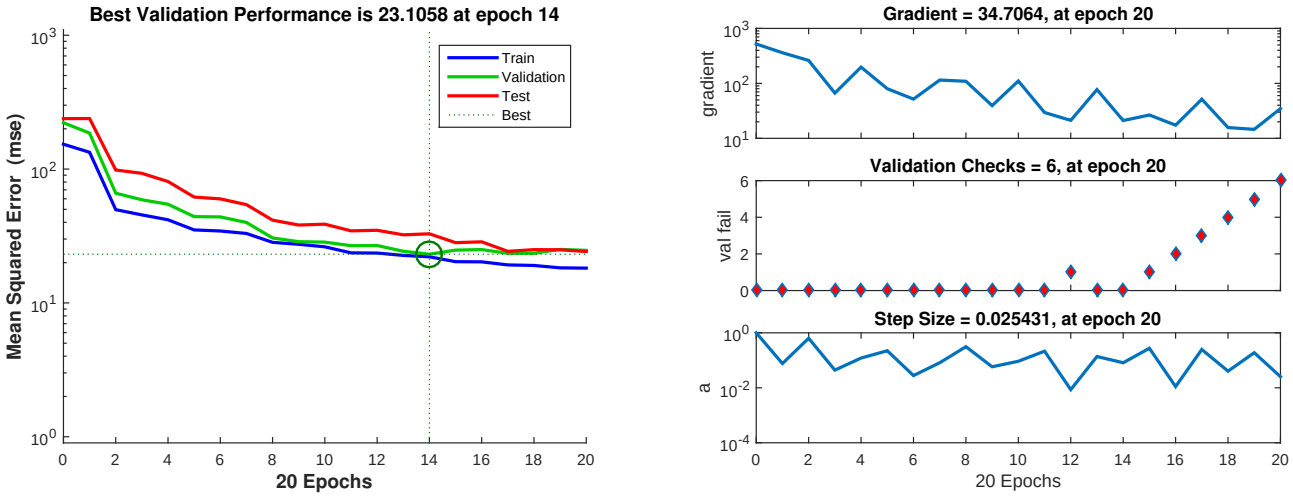

Fig. 6. Validation Performance (left) and Training Performance (right) of FR method.

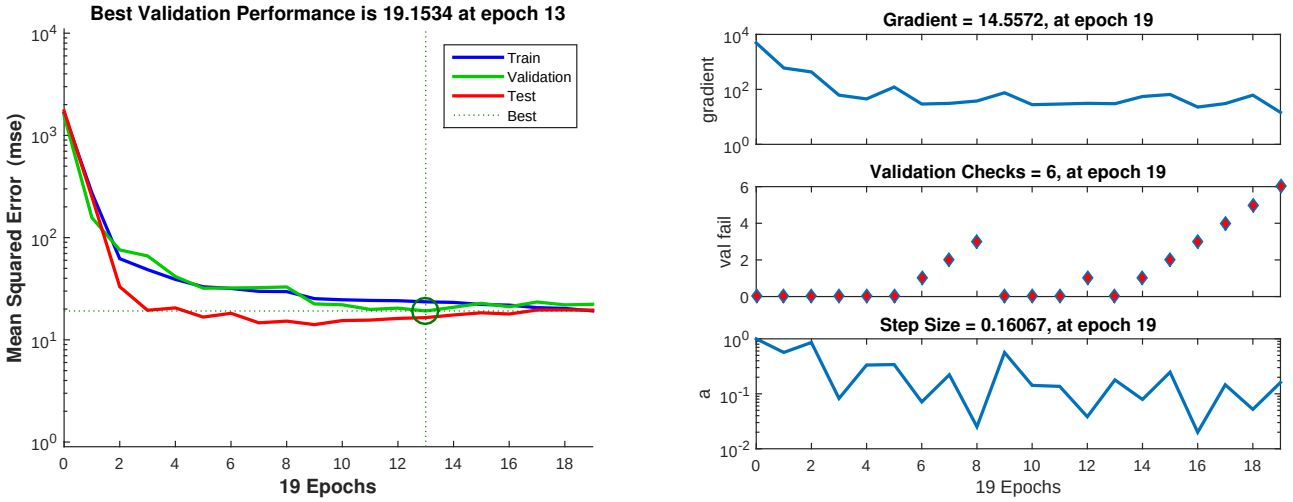

Fig. 7. Validation Performance (left) and Training Performance (right) of PR method.

7. Conclusion

In this paper, a modified KMAR conjugate parameter is proposed for solving unconstrained optimization problems. The new method is an improvement of the classical CG method and possess the sufficient descent property irrespective of the line search. Further, the global convergence of the method is discussed under suitable conditions. Results from numerical computation show that the new method is promising. In addition, the method is extended on feed forward neural network and the outcome has shown a reduction in a number of iterations and CPU time when compared to Classical training function of FR and PR.

[1] Sulaiman I. M., Mamat M. A new conjugate gradient method with descent properties and its application to regression analysis. Journal of Numerical Analysis, Industrial and Applied Mathematics. **14** (1–2), 25–39 (2020).

[2] Dennis J. E., Schnable R. B. Numerical Methods for Unconstrained Optimization and Nonlinear Equations. SIAM, Philadelphia (1993).

[3] Abashar A., Mamat M., Rivaie M., Ismail M. Global convergence properties of a new class of conjugate gradient method for unconstrained optimization. Applied Mathematics and Computation. **8** (67), 3307–3319 (2014).

[4] Rivaie M., Mamat M., Mohd I., Fauzi M. A comparative study of conjugate gradient coefficient for unconstrained optimization. Australian Journal of Basic and Applied Sciences. **5** (9), 947–951 (2011).

[5] Rivaie M., Mamat M., Leong W. J., Mohd I. A new conjugate gradient coefficient for large scale nonlinear unconstrained optimization. International Journal of Mathematical Analysis. **6** (23), 1131–1146 (2012).

[6] Yakubu U. A., Sulaiman I. M., Mamat M., Ghazali P., Khalid K. The global convergence properties of a descent conjugate gradient method. Journal of Advanced Research in Dynamical and Control Systems. **12** (2), 1011–1016 (2020).

[7] Malik M., Mamat M., Abas S. S., Sulaiman I. M., Sukono. A new spectral conjugate gradient method with descent condition and global convergence property for unconstrained optimization. Journal of Mathematical and Computational Science. **10** (5), 2053–2069 (2020).

[8] Awwal A. M., Sulaiman I. M., Malik M., Mamat M., Kumam P., Sitthithakerngkiet K. A Spectral RMIL+ Conjugate Gradient Method for Unconstrained Optimization With Applications in Portfolio Selection and Motion Control. IEEE Access. **9**, 75398–75414 (2021).

[9] Ishak M. I., Marjugi S. M., June W. A new modified conjugate gradient method under the strong Wolfe line search for solving unconstrained optimization problems. Mathematical Modeling and Computing. **9** (1), 111–118 (2022).

[10] Kamfa K., Waziri M. Y., Mamat M., Mohamed M. A., Puspa L. G. A New Modified Three Term CG Search Direction for Solving Unconstrained Optimization Problems. Journal of Advanced Research in Modeling and Simulation. **1** (1), 23–30 (2018).

[11] Andrei N. An unconstrained optimization test functions collection. Advanced Modelling and Optimization. **10** (1), 147–161 (2008).

[12] Kamfa K., Mamat M., Abashar A., Rivaie M., Ghazali P. L., Salleh Z. Another modified conjugate gradient coefficient with global convergence properties. Applied Mathematical Sciences. **9** (37), 1833–1844 (2015).

[13] Sun J., Zhang L. Global convergence of conjugate gradient methods without line search. Annals of Operation Research. **103**, 161–173 (2001).

[14] Wu Q.-j. A Nonlinear Conjugate Gradient Method without Line Search and Its Global Convergence. 2011 International Conference on Computational and Information Sciences. 1148–1152 (2011).

[15] Hager W. W., Zhang H. A survey of nonlinear conjugate gradient methods. Pacific Journal of Optimization. **2** (1), 35–58 (2006).

[16] Kamilu K., Waziri M. Y., Mamat M., Mohamad A. M. A derivative free Newton-like method with improved rational approximation model for solving nonlinear equations. Far East Journal of Mathematical Sciences. **105** (1), 119–129 (2018).

[17] Kamilu K., Waziri M. Y., Ibrahim S. M., Mamat M., Abas S. S. An Efficient Hybrid BFGS-CG Search Direction for Solving Unconstrained Optimization Problems. Journal of Advanced Research in Dynamical and Control Systems. **12** (2), 1035–1041 (2020).

[18] Kamfa K., Sulaiman I. M., Waziri M. Y., Abashar A. Another improved three term PRP-CG method with global convergent properties for solving unconstrained optimization problems. Malaysian Journal of Computing and Applied Mathematics. **1** (1), 1–10 (2018).

[19] Kamfa K., Waziri M. Y., Sulaiman I. M., Ibrahim M. A. H., Mamat M. An Efficient Three Term CG Method using a Modified FR Formula for Solving Unconstrained Optimization Problems. Journal of Advanced Research in Dynamical and Control System. **12** (2), 1027–1034 (2020).

[20] Kamfa K. U., Mamat M., Abashar A., Rivaie M., Ghazali P. L. B., Salleh Z. Another Modified DPRP Conjugate Gradient Method with Global Convergent Properties. Far East Journal of Mathematical Sciences. **9** (37), 1833–1844 (2015).

[21] Hestenes M. R., Stiefel E. Methods of conjugate gradients for solving linear systems. Journal of Research of the National Bureau of Standards. **49** (6), 409–435 (1952).

[22] Polak E., Ribiere G. Note sur la convergence de méthodes de directions conjuguées. ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique. **3** (16), 35–43 (1969).

[23] Polyak B. T. The conjugate gradient method in extremal problems. USSR Computational Mathematics and Mathematical Physics. **9** (4), 94–112 (1969).

[24] Fletcher R., Powell M. J. D. A rapidly convergent descent method for minimization. The Computer Journal. **6** (2), 163–168 (1963).

[25] Liu Y., Storey C. Efficient generalized conjugate gradient algorithms, part 1: Theory. Journal of Optimization Theory and Applications. **69** (1), 129–137 (1991).

[26] Fletcher R. Practical Methods of Optimization. John Wiley & Sons (2020).

[27] Dai Y., Han J., Liu G., Sun D., Yin H., Yuan Y. X. Convergence properties of nonlinear conjugate gradient methods. SIAM Journal on Optimization. **10** (2), 345–358 (2000).

[28] Andrei N. Nonlinear Conjugate Gradient Methods for Unconstrained Optimization. Springer Optimization and its application (2020).

[29] Yuan G., Wei Z., Lu X. Global convergence of BFGS and PRP methods under a modified weak Wolfe–Powell line search. Applied Mathematical Modelling. **47**, 811–825 (2017).

[30] Zhang L., Zhou W., Li D.-H. A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence. IMA Journal of Numerical Analysis. **26** (4), 629–640 (2006).

[31] Dai Z. Comments on a new class of nonlinear conjugate gradient coefficients with global convergence properties. Applied Mathematics and Computation. **276**, 297–300 (2016).

[32] Yousif O. O. O. The convergence properties of RMIL+ conjugate gradient method under the strong Wolfe line search. Applied Mathematics and Computation. **367**, 124777 (2020).

[33] Sulaiman I. M., Malik M., Awwal A. M., Kumam P., Mamat M., Al-Ahmad S. On three-term conjugate gradient method for optimization problems with applications on COVID-19 model and robotic motion control. Advances in Continuous and Discrete Models. **2022**, 1 (2022).

[34] Gilbert J. C., Nocedal J. Global convergence properties of conjugate gradient methods for optimization. SIAM Journal on Optimization. **2** (1), 21–42 (1992).

[35] Wei Z., Yao S., Liu L. The Convergence Properties of some New Conjugate Gradient Methods. Applied Mathematics and Computation. **183** (2), 1341–1350 (2006).

[36] Dai Z., Wen F. Another improved Wei–Yao–Liu non-linear conjugate gradient method with sufficient descent property. Applied Mathematics and Computation. **218** (14), 7421–7430 (2012).

[37] Zabidin S., Adel A., Ahmad A. Two efficient modifications of AZPRP conjugate gradient method with sufficient descent property. Journal of Inequalities and Applications. **2022**, 14 (2022).

[38] Kamfa K., Ibrahim S. M., Sufahani S. F., Yunus R. Y., Mamat M. A modified BFGS method via new rational approximation model for solving unconstrained optimization problems and its application. Advances in Mathematics: Scientific Journal. **5**, 10771–10786 (2020).

[39] Ma G., Lin H., Han D. Two modified conjugate gradient methods for unconstrained optimization with applications in image restoration problems. Journal of Applied Mathematics and Computing. **68**, 4733–4758 (2022).

[40] Dolan E., Moré J. J. Benchmarking optimization software with performance profile. Mathematical Programming. **91**, 201–213 (2002).

[41] Yoksal A. L., Abbo K. K., Hisham M. K. Training feed forward neural network with modified Fletcher–Reeves method. Journal of Multidisciplinary Modelling and Optimization. **1** (1), 14–22 (2018).

[42] Livieris I., Pintelas P. Performance evaluation of descent CG methods, for neural net-works training. Proceedings of the 9th Hellenic European Research on Computer Mathematics and its Applications Conference (HERCMA '09). 40–46 (2009).

[43] Rumelhart D. E., Hinton G. E., Williams R. J. Learning representations by back-propagating errors. Nature. **323**, 533–536 (1986).

[44] Battiti R. First-and second-order methods for learning: between steepest descent and Newton's method. Neural Computation. **4** (2), 141–166 (1992).

Оцінка продуктивності нового методу спряженого градієнта для навчання нейронної мережі з прямим зв'язком

Камілу К.[1], Сулейман М. І.[2,3], Мухаммад А. Л.[1], Мохамад А. В.[4], Мамат М.[5]

[1] *Кафедра математичних наук, Факультет обчислювальної техніки та математики, Науково-технічний університет Кано, 713101, Вуділ, Нігерія*
[2] *Школа кількісних наук, Університет Утара Малайзія, 06010, Кедах, Малайзія*
[3] *Інститут моделювання стратегічних промислових рішень (ISIDM), SQS, Університет Північної Малайзії, Сінток, 06010, Кедах, Малайзія*
[4] *Школа стоматологічних наук, Університет Святих Малайзії, Кампус здоров'я (USM), 16150, Кубанг Керіан, Келантан, Малайзія*
[5] *Факультет інформатики та обчислювальної техніки, Університет Султана Зайнала Абідіна, Теренгану, Малайзія*

У цій статті створено новий метод спряженого градієнта для розв'язання задач необмеженої оптимізації. Запропонований метод задовольняє властивість достатнього спуску незалежно від лінійного пошуку, і глобальна збіжність була встановлена за деяких умов. Крім того, новий метод використовувався для навчання різного набору даних через нейронну мережу з прямим зв'язком. Отримані результати показують, що запропонований алгоритм значно скорочує час обчислення за рахунок прискорення спрямованої мінімізації з вищою швидкістю збіжності.

Ключові слова: *метод спряженого градієнта; нейронна мережа; лінійний пошук; аналіз збіжності.*