BAREKENG: Journal of Mathematics and Its Applications
September 2025 Volume 19 Issue 3 Page 1973-1988
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss3pp1973-1988

A THREE-TERM CONJUGATE GRADIENT METHOD FOR LARGE-SCALE MINIMIZATION IN ARTIFICIAL NEURAL NETWORKS

Umar A Omesa[1], Muhammad Y. Waziri[2], Issam A.R. Moghrabi[3], Sulaiman M. Ibrahim[4*], Gudu E B[5], Fakai S L[6], Rabiu Bashir Yunus[7], Elissa Nadia Madi[8]

[1,5,6]*Department of Mathematics, College of Science, Federal University of Agriculture Zuru, Kebbi State, 872211, Nigeria*

[1]*Numerical Optimization Research Group, Bayero University*
[2]*Department of Mathematics, Faculty of Physical Sciences, Bayero University Kano, Gwarzo Road Kano, PMB 3011, Nigeria*

[3]*Department of Information Systems and Technology, Kuwait Technical College Street 5, Abu-Halifa, P.O.Box: 232, Kuwait*

[4]*School of Quantitative Sciences, College of Arts and Sciences, Universiti Utara Malaysia, Sintok, 06010, Kedah, Malaysia.*

[4]*Faculty of Arts and Education, Sohar University Sohar, Sultanate of Oman, P.O Box: 44, P.C 311, Oman*

[7]*Department of Mathematics, Aliko Dangote University of Science and Technology Wudil 711011, Gaya North, Kano State, Nigeria,*

[8]*Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia. Besut Campus, Besut Terengganu Darul Iman, 22200, Malaysia.*

*Corresponding author's e-mail:** ***sulaimancga@gmail.com***

Article History:

Received: 18th February 2025
Revised: 5th April 2025
Accepted: 8th April 2025
Published: 1st July 2025

Keywords:

Conjugate Gradient;
Global Convergence;
Neural Networks;
Sufficient Descent Condition;
Unconstrained Optimization.

ABSTRACT

Conjugate Gradient (CG) methods are widely used for solving unconstrained optimization problems due to their efficiency and low memory requirements. However, standard CG methods may not always guarantee sufficient descent condition, which can impact their robustness and convergence behavior. Additionally, their effectiveness in training artificial neural networks (ANNs) remains an area of interest. In response, this paper presents a three-term conjugate gradient (CG) method for unconstrained optimization problems. The new parameter is formulated so that the search direction satisfies the sufficient descent condition. The global convergence result of the new algorithm is discussed under suitable assumptions. To evaluate the performance of the new method we considered some standard test problems for unconstrained optimization and applied the proposed method to train different ANNs on some benchmark data sets contained in the NN toolbox. The experimental results show that performance is encouraging for both unconstrained minimization test problems and in training neural networks.

How to cite this article:

U. A. Omesa, M. Y. Waziri[3], I. A.R. Moghrabi, S. M. Ibrahim, Gudu E. B, Fakai S. L, R. B. Yunus and E. N. Madi., "A THREE-TERM CONJUGATE GRADIENT METHOD FOR LARGE-SCALE MINIMIZATION IN ARTIFICIAL NEURAL NETWORKS," *BAREKENG: J. Math. & App.*, vol. 19, no. 3, pp. 1973-1988, September, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Unconstrained optimization problems (UOP) play a fundamental role in various scientific and engineering applications, including pattern recognition [1], machine learning [2], [3], security [4], signal processing [5], robotics [6], self-driving cars [7], and meteorology [8]. Over the years, numerous numerical algorithms have been developed to solve these problems efficiently. Early approaches include first-order methods such as the steepest descent method, which, despite its simplicity, suffers from slow convergence. More advanced second-order methods, including Newton's and quasi-Newton's algorithms, leverage second derivative (Hessian) information to achieve faster convergence [9]. However, these methods often require significant computational and storage resources, making them impractical for large-scale problems.

To address these challenges, conjugate gradient (CG) methods emerged as a preferred choice due to their low memory requirements, strong convergence properties, and ability to handle large-scale optimization problems effectively [10]. Unlike steepest descent, CG methods iteratively construct search directions using gradient information and a conjugacy condition to ensure efficient exploration of the search space [11]. Among the classical CG algorithms, the Fletcher-Reeves [12], Polak-Ribière-Polyak [13], [14], Hestenes-Stiefel [15], Dai-Yuan [16], and Liu-Storey [17] methods have been widely studied and applied in various fields. Despite their theoretical advantages, classical CG methods exhibit certain limitations, such as numerical instability and slow convergence, particularly when dealing with ill-conditioned functions or small gradient norms [18]. To enhance performance, recent research has focused on developing three-term conjugate gradient (TTCG) methods, which incorporate additional information from previous iterations to improve stability and descent properties. While TTCG methods have demonstrated improved robustness, they still face challenges such as parameter tuning complexities and occasional inefficiency in high-dimensional optimization problems. For recent advances on this topic see [19], [20], [21]

Motivated by these advancements, this study develops a new TTCG method called NEWTT based on modification of the Barzilai-Borwein [22] step size and DY method. The proposed method incorporates both gradient information and the search direction in determining the step size. The proposed approach aims to achieve superior performance in large-scale unconstrained optimization problems, including applications in training feed-forward neural networks. Theoretical analysis will be conducted to establish convergence guarantees, and extensive numerical experiments will be performed with application to evaluate its efficiency on standard benchmark problems and real-life application problems.

The contributions of the paper are highlighted as follows

1. Proposing a simple three-term CG method based on the methods of DY and BB step size that is effective for training artificial neural networks and solving unconstrained optimization problems.
2. The new method presented is such that it satisfies the descent condition which is a crucial element in proving global convergence
3. Showing that the global convergence of the new method under the Wolfe condition and the assumption of a Lipschitz condition
4. Showing the efficiency of the proposed method through numerical experiments conducted with recent methods.

The remaining sections of the paper are structured as follows: the formulation process of the proposed method is given in the next section, while Section 3 presents the global convergence of the new method under appropriate assumptions. Section 4 provides numerical experiments; section 5 contains the application of the proposed method in training neural networks and the last section includes concluding remarks.

2. RESEARCH METHOD

Generally, the CG method is employed to solve the following unconstrained optimization problems [23]:

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

by generating a sequence of points $\{x_k\}, k \geq 1$ via:

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

where $f: R^n \to R$ from (1) is a continuously differentiable function, $\alpha_k > 0$ from **Equation (2)** is the step length computed along the direction d_k. The direction is generally required to satisfy the descent condition

$$g_k^T d_k < 0 \tag{3}$$

to guarantees that d_k is a descent direction of $f(x)$ at x_k. The direction is generated using

$$d_k = -g_k + \beta_k d_k, \quad d_0 = -g_0 \tag{4}$$

for $k \geq 1$. From the iteration **Equation (2)**, α_k is the line search parameter in CG formulas and is often based on the general Wolfe condition

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \rho \sigma_k g_k^T d_k \tag{5}$$

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k \tag{6}$$

where d_k is a descent direction and $0 < \rho \leq \sigma < 1$. The stronger version of the Wolfe line search condition is given by **Equation (5)** and

$$\left| g_{k+1}^T d_k \right| \leq -\sigma g_k^T d_k \tag{7}$$

to enhance stability and guarantee convergence. If the direction is exact, then **Equation (4)** satisfies the descent condition given as:

$$g_k^T d_k = -g_k^T g_k \leq 0$$

The parameter β_k in **Equation (4)** is the CG parameter and $g_k = \nabla f(x_k)$ is the gradient of f at x_k. The selection of the parameter β_k varies when considering the minimization of a strongly convex function and the general nonlinear functions. For the minimization of a strongly quadratic convex function, the parameter β_k is selected so that the directions of d_k and d_{k+1} are subject to the Hessian of the quadratic function **[24]**. The effectiveness of these algorithms relies on the precision of the line search. However, when applied to general nonlinear functions, the parameter β_k is often determined using alternative formulas that do not satisfy the conjugacy condition **[24]**.

2.1 Related Work

Several well-known CG methods include those proposed by Polak, Ribiere, and Polyak (PRP) **[13]**, **[14]**, Fletcher and Reeves (FR) **[12]**, Hestenes and Stiefel (HS) method **[15]**, Liu and Storey (LS) **[17]**, Dai and Yuan (DY) **[16]**, Dai and Liao (DL) **[25]**, Fletcher's conjugate descent (CD) method **[26]**. The corresponding CG parameters are as follows:

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}, \quad \beta_k^{PRP} = \frac{y_k^T g_{k+1}}{g_k^T g_k}, \quad \beta_k^{LS} = \frac{-y_k^T g_{k+1}}{g_k^T g_k}, \quad \beta_k^{HS} = \frac{g_k^T y_k}{d_{k-1}^T y_{k-1}}$$

$$\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k}, \quad \beta_k^{DL} = \frac{g_{k+1}^T (y_k - t s_k)}{y_k^T s_k}, \text{where } t > 0, \quad \beta_k^{CD} = \frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}$$

where $y_k = g_{k+1} - g_k$, and $s_k = x_{k+1} - x_k$. By extension, we refer to all these as CG algorithms, even though some parameters, such as β_k^{FR}, and β_k^{DY} do not satisfy the conjugacy condition **[24]**. Assuming that the function f is a strongly convex function, and that the exact line search is employed, then in theory, all the choices for the parameter β_k in **Equation (4)** are equivalent. However, in the case of the non-quadratic objective function, the choice of the parameter β_k performs differently depending on the algorithm used.

In this study, the authors are more interested in three-term CG methods. These classes of CG techniques are designed to improve the convergence rate of CG methods. Among the early three-term CG is that proposed by Beale **[27]** as

$$d_{k+1} = -g_{k+1} + \beta_k d_k + \gamma_k d_t \tag{8}$$

where the choice of β_k can be $\beta_k^{HS} (or\ \beta_k^{FR}, \beta_k^{DY}$ etc) and

$$d_{k+1} = \begin{cases} 0, & k = t+1 \\ \dfrac{g_{k+1}^T y_t}{d_t^T y_t} & k > t+1 \end{cases}$$

the d_t in **Equation (9)** is the restart direction.

Another three-term CG method was proposed by Nazareth **[28]** to achieve finite convergence for an arbitrary initial search direction, based on a three-term recurrence

$$d_{k+1} = -y_k + \frac{y_k^T y_k}{d_k^T y_k} d_k + \frac{y_{k-1}^T y_k}{d_{k-1}^T y_{k-1}} d_{k-1} \tag{9}$$

where $d_{-1} = 0$ and d_0 an arbitrary descent direction. Although both methods **Equation (8)** and **Equation (9)** exhibit finite termination properties, they are not practically efficient **[29]**.

Furthermore, these methods do not always ensure the generation of descent directions, meaning they are not classified as descent methods **[30]**.

Dai and Yuan **[16]** and Deng and Li **[31]** studied the general three-term CG formula given as

$$d_{k+1} = -g_{k+1} + \beta_k d_k + \gamma_k d_{t(p)} \tag{10}$$

and $t(p)$ in **Equation (10)** is the number of the p^{th} restart iteration satisfying $t(p) < k \leq t(p+1) \leq +\infty$.

Demonstrating that, under certain mild conditions, the algorithm achieves global convergence.

Hager and Zhang **[32]** developed a three-term CG method given as

$$d_{k+1} = -g_{k+1} + \beta_k^N d_k, \qquad \beta_k^N = \frac{1}{d_k^T y_k}\left(y_k - 2d_k \frac{\|y_k\|^2}{d_k^T y_k}\right)^T g_{k+1} \tag{11}$$

They established the global convergence for strongly convex functions and general nonlinear functions by utilizing the Goldstein conditions **[33]** and the Wolfe condition, respectively.

Another three-term CG method was presented by Zhang, Zhou, and Li **[30]** by considering the (HS) CG method as presented below

$$\beta_k^{ZHS} = \begin{cases} -g_k & if\ k = 0 \\ -g_k + \beta_k^{HS} d_{k-1} - \theta y_{k-1} & if\ k \geq 1 \end{cases}$$

where $\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}$, and $\theta = \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}}$.

Another efficient three-term parameter was given by Lukman and Muhammad **[34]** as

$$\beta_k^{NTT} = \begin{cases} -g_k & \text{if } k = 0 \\ -g_k + \beta_k^{HS} d_k - (1-\theta)\dfrac{y_{k-1}^T y_k g_k^T d_{k-1}}{y_{k-1}^T v_{k-1} d_{k-1}^T y_{k-1}} y_{k-1} & \text{otherwise} \end{cases} \tag{12}$$

which is a modification of the (HS) conjugate gradient method. When $\theta = 1$ or the exact line search is satisfied then the three-term parameter becomes the HS method.

Despite the extensive research on conjugate gradient (CG) methods, there remains a gap in developing a three-term CG scheme that integrates both the Dai-Yuan (DY) **[16]** parameter and a modified Barzilai-Borwein (BB) step size **[22]** while ensuring descent properties independent of the line search. The DY method is known for its strong theoretical convergence properties; however, its performance can be inconsistent due to its sensitivity to line search strategies, often leading to slow convergence in practical applications. Additionally, most existing three-term CG methods do not fully exploit step-size modifications that can enhance numerical stability and efficiency. To address these issues, this study proposes a novel three-term CG method (NEWT) that incorporates an additional memory term while leveraging a modified BB step size. This approach improves robustness in large-scale optimization and mitigates the DY method's sensitivity to line search, leading to better numerical performance across different problem classes.

2.2 New Three-Term Conjugate (NEWTT) Gradient Method

The proposed three-term conjugate gradient method satisfying the descent condition independent of the line search is formulated as follows. Consider the Barzilai-Borwein (BB) [22] step size defined as:

$$\gamma_k^1 = \frac{s_k^T s_k}{y_k^T s_k} \text{ and } \gamma_k^2 = \frac{s_k^T y_k}{y_k^T y_k}$$

Inverting γ_k^2 and subtracting from γ_k^1 we get

$$\frac{\left(s_k^T s_k - \|y_k\|^2\right)}{y_k^T s_k} g_{k-1} \tag{13}$$

Now, considering the Dai and Yuan conjugate gradient parameter $\left(\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k}\right)$ and **Equation (13)**, we have our three-term scheme as

$$d_{k+1} = -g_{k+1} + \beta_k d_k - \theta_k \frac{g_{k-1}}{y_k^T s_k} \tag{14}$$

where $\beta_k = \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k}$, and $\theta_k = s_k^T s_k - \|y_k\|^2$.

The following algorithm illustrates the computational process of the proposed NEWTT algorithm.

Algorithm 2.1: NEWTT Algorithm

Step 1: Input $x_0 \in R^n, \theta \in (0,1)$ Set $k = 0$ and $d_0 = -g_0$

Step 2: If $\|g_k\| < \epsilon$, stop; otherwise, continue with step 3

Step 3: Compute d_k by **Equation (14)**

Step 4: Find the step length α_k by **Equation (4)** and **Equation (5)**

Step 5: Set $x_{k+1} = x_k + \alpha_k d_k$.

Step 6: Proceed to Step 2 after setting $k = k + 1$.

3. RESULTS AND DISCUSSION

This section will discuss the convergence results of the proposed method and further demonstrate its computational efficiency and real-life applications.

3.1 Convergence Analysis

This subsection presents the global convergence result of the NEWTT algorithm. We consider the line search that satisfies **Equation (5)** and **Equation (6)**.

Assumption 3.1

(A_1): The level set $L = \{x \in R^n : f(x) \leq f(x_0)\}$ is bounded below. Thus, for some constant $L > 0$ it follows that $\|x\| \leq L$ for all $x \in L_0$.

(A_2): The function f is smooth in some neighborhood N of the level set L_0,with Lipschitz continuous gradient, implying that, for some $L > 0$, the following holds

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \forall x, y \in N. \tag{15}$$

Considering the assumptions on f, it is obvious that there exists a constant $\gamma \geq 0$ such that $\|g(x)\| \leq \gamma$ for all $x \in N$. However, we shall consider the assumption that the level set is bounded which is a stronger assumption than the assumption that states the function is bounded below.

Note that the search direction in **Equation (2)** always ensures a descent direction. To establish convergence, we restrict the selection of α_k by proposing that the Wolfe line search consistently provides a lower bound for the step length α_k. Consider the following proposition.

Proposition 1: Let d_k be a descent direction and assume that ∇f satisfies the Lipschitz condition

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

for all x on the line segment between x_k and x_{k+1}, where L is a positive constant. If the line search adheres to the Wolfe conditions **Equation (5)** and **Equation (6)**, then

$$\alpha_k \geq \frac{(1-\sigma)|g_k^T d_k|}{L\|d_k\|^2} \tag{16}$$

Proof: from **Equation (6)** subtract $g_k^T d_k$ from both sides

$$\begin{aligned} g_{k+1}^T d_k - g_k^T d_k &\geq \sigma g_k^T d_k - g_k^T d_k \\ (g_{k+1} - g_k)^T d_k &\geq (\sigma - 1) g_k^T d_k \\ {y_k}^T d_k &\geq (\sigma - 1) g_k^T d_k \end{aligned}$$

Now using the Lipschitz continuity, we get

$$(\sigma - 1) g_k^T d_k \leq (g_{k+1} - g_k)^T d_k = {y_k}^T d_k \leq \|{y_k}^T\| \|d_k\| \leq \alpha_k L \|d_k\|^2$$

Since d_k is a descent direction and $\sigma < 1$, **Equation (16)** follows immediately. ■

For the proof of the global convergence, the Zoutendijk condition is applied to nonlinear conjugate gradient algorithms. The following proposition proves that our proposed method satisfies the Zoutendijk condition under the general Wolfe line search **Equation (5)** and **Equation (6)**.

Proposition 2: Assume that conditions(A_1) and (A_2) hold. Given **Equation (2)** and **Equation (14)**, where d_k is a descent direction and α_k is determined using the standard Wolfe line search, then

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

From **Equation (4)** and proposition 1

$$f(x_k) - f(x_{k+1}) \geq -\rho\sigma_k g_k^T d_k \geq \frac{\rho(1-\sigma)(g_k^T d_k)^2}{L\|d_k\|^2}$$

Hence, from assumption 1 we get the Zoutendijk condition **Equation (16)**. ■

Proposition 3: Assume that conditions(A_1) and (A_2) hold. Given **Equation (2)** and **Equation (14)**, where d_k is a descent direction and α_k is determined using the standard Wolfe line search, if

$$\sum_{k\geq 1}^{\infty} \frac{1}{\|d_k\|^2} = \infty \tag{18}$$

Then

$$lim \inf_{k\to\infty} \|g_k\| = 0 \tag{19}$$

Theorem 1. Suppose that assumption (A_1) and (A_2) holds. Consider **Equation (2)** and **Equation (14)**, and α_k satisfies the standard Wolfe condition, then

$$lim \inf_{k\to\infty} \|g_k\| = 0$$

Proof. From Lipschitz continuity, we have $\|y_k\| \leq L\|s_k\|$. On the other hand, from uniform convexity $y_k^T s_k \geq \mu\|s_k\|^2$. Now from **Equation (14)**

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

Multiplying by g_{k+1}^T, we have

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

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

for exact line search $g_k^T d_k = -g_k^T g_k \leq 0$, we have

$$g_{k+1}^T d_{k+1} = \frac{g_{k+1}^T d_k}{y_k^T s_k} - \frac{(s_k^T s_k - \|y_k\|^2)}{y_k^T s_k} \tag{20}$$

But from Lipschitz's continuity, we can write as

$$g_{k+1}^T d_{k+1} = \frac{1}{y_k^T s_k} - \frac{(1-L)\|s_k\|}{y_k^T s_k} \tag{21}$$

Also, from uniform convexity $y_k^T s_k \geq \mu \|s_k\|^2$

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

$$\|d_{k+1}\| \leq \frac{(1-L)}{\mu \|s_k\|} + \frac{1}{\mu \|s_k\|^2} \tag{23}$$

Showing that **Equation (18)** is true. ■

3.2. Numerical Efficiency

This section presents the report on the numerical performance of the proposed conjugate gradient algorithm (NEWTT) by considering some standard test problems on UOP sourced from Andrei **[35]**. For each test problem, arbitrary initial points are considered together with the range of dimensions from small-scale problems (2 variables) and large-scale problems (1,000 variables). To compare the accuracy of the computed results from the proposed method, we considered results obtained from the following conjugate gradient methods:

1. The Dai-Yuan (DY) **[16]** algorithm with conjugate gradient parameter given as $\beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k}$.
2. The three-term conjugate gradient algorithm by Zhang, Zhou, and Li **[30]** with CG parameter as

$$\beta_k^{ZHANGTT} = -g_k + \beta_k^{HS} d_{k-1} - \theta y_{k-1}, \beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \qquad \theta = \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}}$$

The algorithms were implemented in MATLAB® R2023b environment, utilizing double-precision arithmetic for numerical computation. The experiments were carried out on a personal computer with Windows 10 pro, equipped with Intel (R) Pentium (R) Dual CPU T3400 @2.16GHz 2.17GHz. We considered the Wolfe line search method throughout the numerical computation with the parameter values defined as $\rho = 0.01$ and $\sigma = 0.5$ for all the algorithms.

The termination criteria for all algorithms were $\|g(x_k)\| \leq 10^{-6}$ or any of the following

1. The iteration number reaches 1000
2. The code fails to execute as a result of low memory.

The results from numerical experiment are illustrated in **Figure 1** – **Figure 2** using performance profile tool introduced by Dolan and More **[36]**. The plots represent the CPU time to execute the problem (**Figure 1**) and the number of iterations across problems of different dimensions (**Figure 2**).

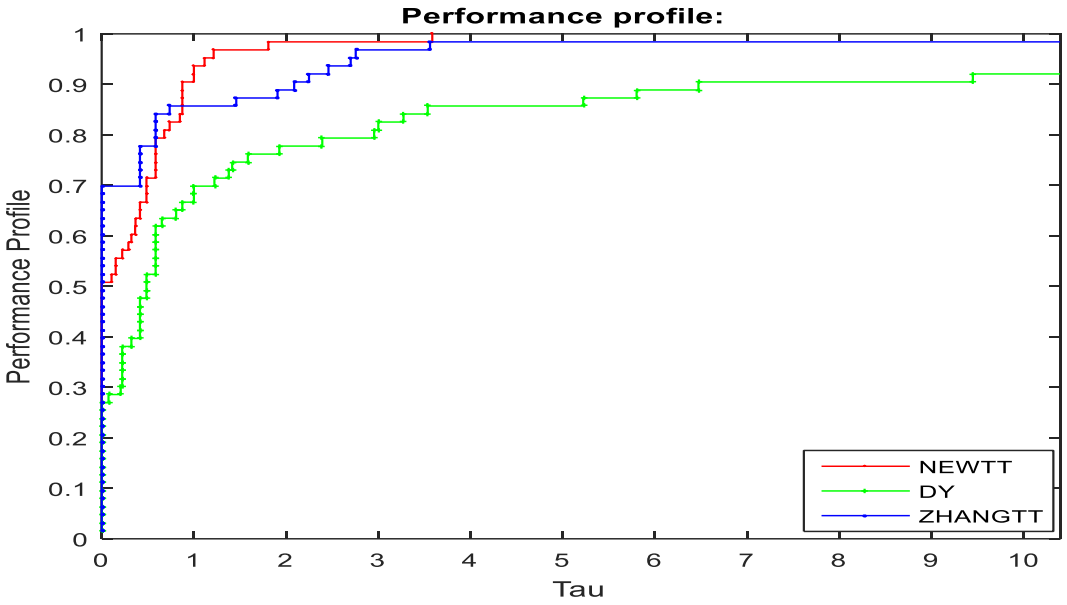

Figure 1. Performance Profile Based on CPU time

Figure 1 illustrates the performance profile of three unconstrained optimization methods: NEWTT, DY, and ZHANGTT based on CPU time. The proposed NEWTT method (red) demonstrates the best overall performance, solving the majority of problems with the least computational time, followed closely by ZHANGTT (blue). Both methods exhibit rapid convergence, reaching a high probability of success at lower τ values. The DY method (green) is less efficient, requiring significantly more computational time, as indicated by its slower rise in profile. Overall, NEWTT and ZHANGTT are more computationally efficient, making preferable choices for solving UOP when minimizing CPU time is crucial. This shows that the numerical result for the new method is encouraging.

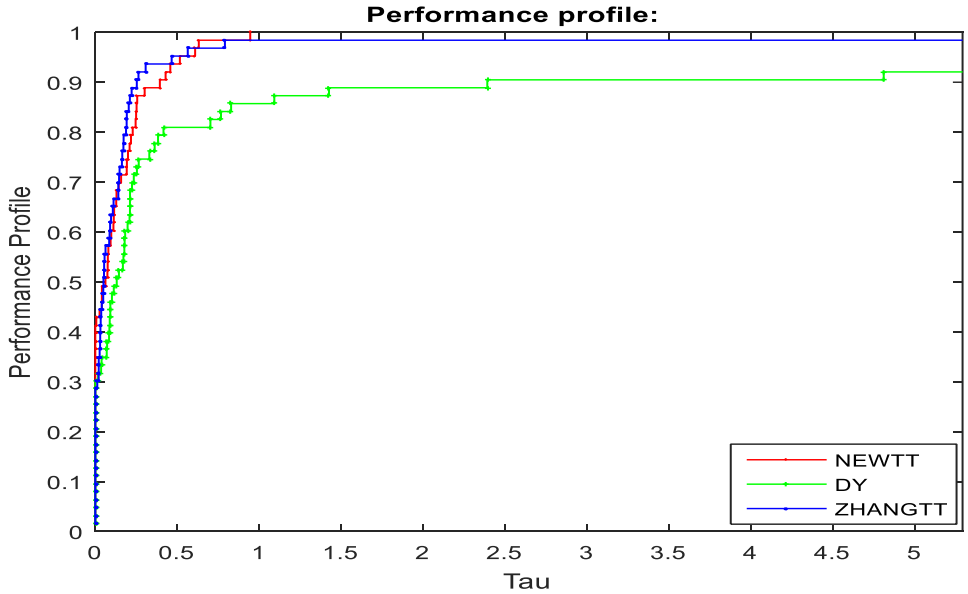

Figure 2. Performance Profile Based on the Number of Iterations

Also, the results based on the number of iterations as illustrated in **Figure 2** demonstrate superiority performance of the proposed method in terms of iteration efficiency. The performance profile, based on the number of iterations, evaluates the effectiveness of three unconstrained optimization methods: NEWTT, DY, and ZHANGTT. The proposed NEWTT method (red) achieves the highest percentage of problem solutions with fewer iterations, followed closely by ZHANGTT (blue). Both algorithms exhibit rapid convergence, attaining high performance probabilities at lower τ values. On the other hand, the DY algorithm (green) requires more iterations on average, as evidenced by its slower rise in the profile. This suggests that the

proposed NEWTT method is the most efficient in terms of iteration count, making it the preferred choice for solving unconstrained optimization problems.

3.3. Application in Neural Networks

Artificial neural networks (ANN) are widely known for their versatility, accuracy, and powerful predictive capabilities, and as such they are developed and applied in various areas like computer science, and autonomous driving [7], [37], [38] digital twin [39], prognostics and health management [40], reliability analysis [41], topology optimization [42], security [4], and multi-objective optimization [43]. However, training a neural network is an UOP where the weight function is minimized [44], [45], [46]. For prediction and minimization of the cost functions the weights W of an artificial neural network (ANN) are optimized by

$$w = arg\min_{w} f(w) \tag{24}$$

and the weights are updated as follows

$$w_{k+1} = w_k + \alpha_k d_k \;\; d_k = \nabla E(w_k)$$

where d_k and the α_k denotes the search direction (first-moment vector) and the step size (learning rate). w_k is a vector representing the weight at the iteration (epoch) step k.

As earlier stated, the technique of the gradient method is to minimize the cost function by determining the search direction as the first-order gradient $\nabla E(w_k)$, whereas the step size is calculated by adopting any of the available line search methods. Though no single best optimization method has been adopted, many existing adaptive methods use the first-order gradient for the search direction, including the conjugate gradient method. In this paper, we propose a simple three-term conjugate gradient algorithm. To illustrate the performance of the proposed method MATLAB program was used to code the algorithm and its implementation.

3.3.1. Experiments and Results

The numerical experiment was carried out using the standard data set contained in the Neural Network (NN) Toolbox version 4.0.2 (R13).

Problem 1. A chemical sensor data set was used to train the neural network applying our algorithm. The network architecture contains one hidden layer with 10 neurons and a single output layer. A maximum of 1000 iterations is set as the termination criterion, and all the parameters are as specified in the NN toolbox.

Table 1. CPU Time and Number of Iterations

CG Parameter	CPU Time	Min Epoch (Number of Iterations)	Best Validation Performance
Zhang three-term	0.00.05	358	5.7016 at epoch 350
FR	0.00.02	141	3.9003 at epoch 135
New method	0.00.02	68	7.6443 at epoch 62

Table 1 shows the simulation performance of our proposed method compared with FR and ZhangTT conjugate gradient methods. The table shows the number of iterations (Epoch), CPU time, and the best validation performance. It is clear from the table that our new method is more efficient compared to the other methods.

Table 2. Regression Analysis

CG parameter	Training	Validation	Test
Zhang three-term	0.95393	0.95868	0.95616
FR	0.94917	0.96056	0.94769
New method	0.94649	0.92198	0.95864

Table 2 shows the performance of the regression analysis of our proposed method against FR and Zhang three-term conjugate methods. From the table, it is obvious that our method is promising and can be appropriate for training and testing neural networks.

3.3.2. Mean Square Error

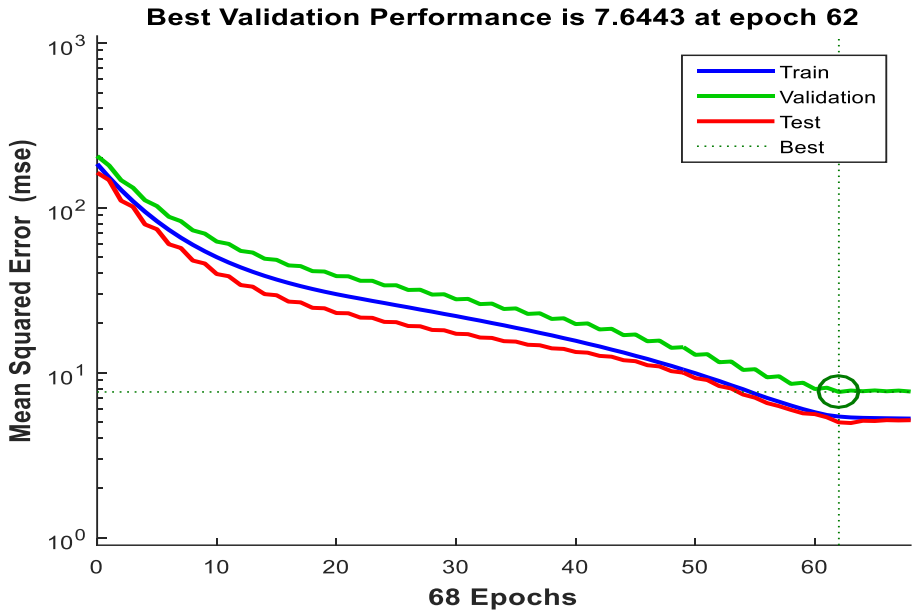

Figure 3. Mean Square Error Performance.

Figure 3 shows the best validation performance of our proposed method as 9.5811 at epoch 35. However, the total number of iterations (epoch) for the testing and training of the data set for the chemical sensor is 41 iterations.

3.3.3 Regression Analysis

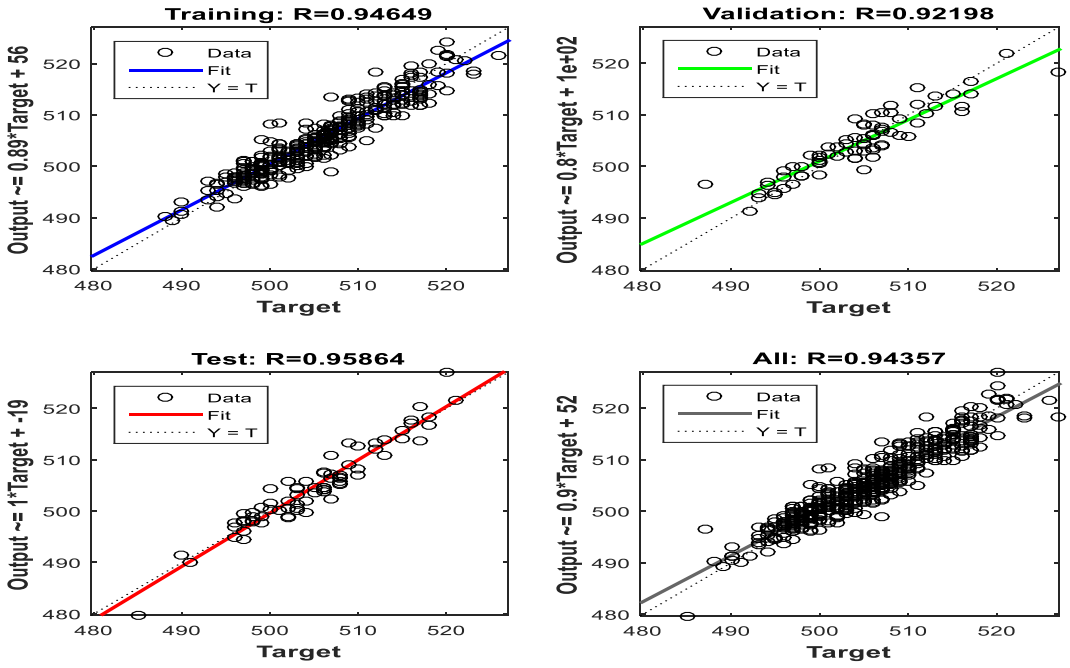

Figure 4. R-Values

Figure 4 shows the performance of regression analysis of our new method. The R-values for training output, validation output, and test output show that our proposed method is promising and encouraging.

3.3.4. Training State

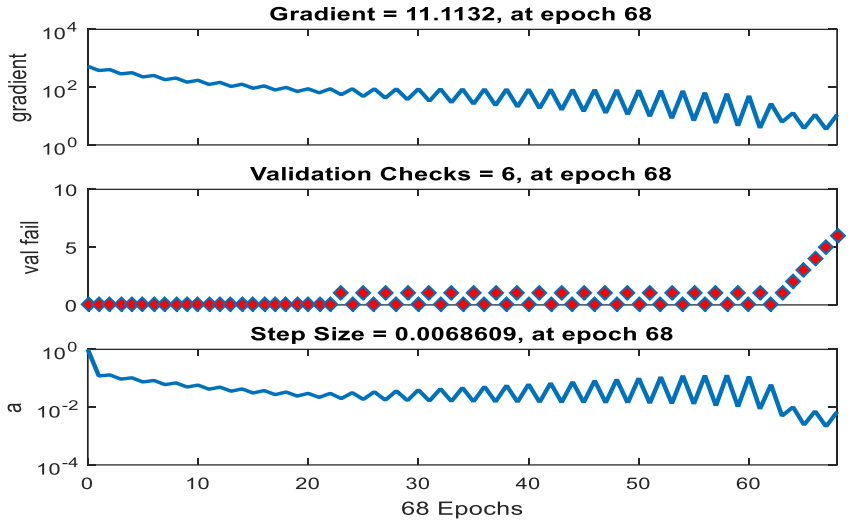

Figure 5. Training state

Figure 5 shows that the gradient value of our method is 34.7 at iteration (epoch) 41. At epoch 41, the validation checks equal 6 and the value of the step size is 0.03.

Problem 2: The second problem analyzed is the Cholesterol dataset from the neural network toolbox. The network structure consists of 21 input nodes, 10 hidden layers, and three output nodes. All parameters are set to their default values as specified in the NN toolbox, with a maximum of 1000 iterations as the termination criterion. Performance is evaluated by over 100 simulations based on the number of iterations (epochs), CPU time, and best validation epoch performance. **Table 3** presents the training and validation performance of the training functions for the FR method, Zhang's three-term method, and the proposed approach. The results clearly indicate that the proposed method is efficient and has potential applications in other fields.

Table 3. CPU Time and Number of Iterations

CG parameter	CPU Time	Min Epoch (Number of Iterations)	Best Validation Performance
Zhang three-term	0.00.02	118	698.2119 at epoch 112
FR	0.00.02	117	4.667 at epoch 111
New method	0.00.01	44	5.3113 at epoch 38

Table 3 shows the simulation performance of our proposed method compared with FR and ZhangTT conjugate gradient methods. The table shows the number of iterations (Epoch), CPU time, and the best validation performance. It is clear from the table that our new method is more efficient compared to the other methods.

Table 4. Regression Analysis

CG parameter	Training	Validation	Test
Zhang three-term	0.92452	0.89466	0.88531
FR	0.95804	0.94916	0.94656
New method	0.94317	0.94206	0.91462

Table 4 shows the performance of the regression analysis of our proposed method against FR and Zhang's three-term conjugate methods. From the table, it is clear that our method is promising and can be appropriate for training and testing neural networks.

3.3.5. Mean Square Error

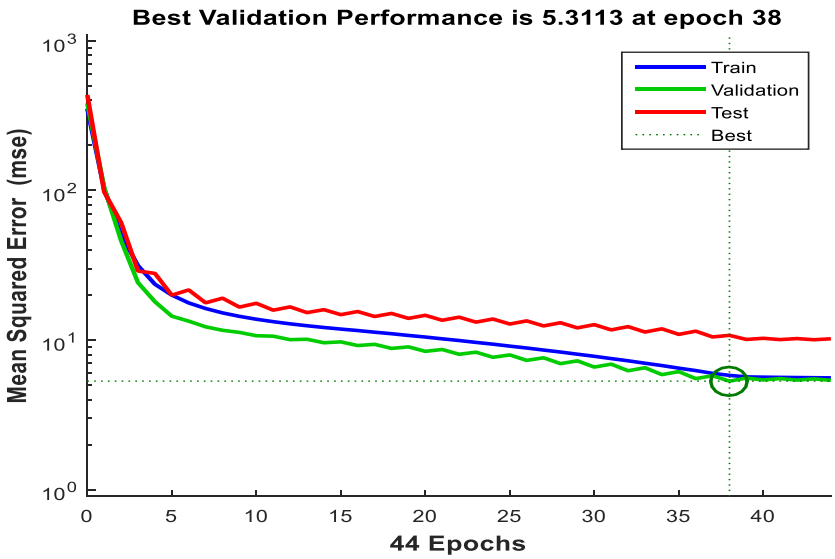

Figure 6. Mean Square Error Performance.

Figure 6 shows the best validation performance of our proposed method as 9.5811 at epoch 35. However, the total number of iterations (epoch) for the testing and training of the data set for the chemical sensor is 41 iterations.

3.3.6. Regression Analysis

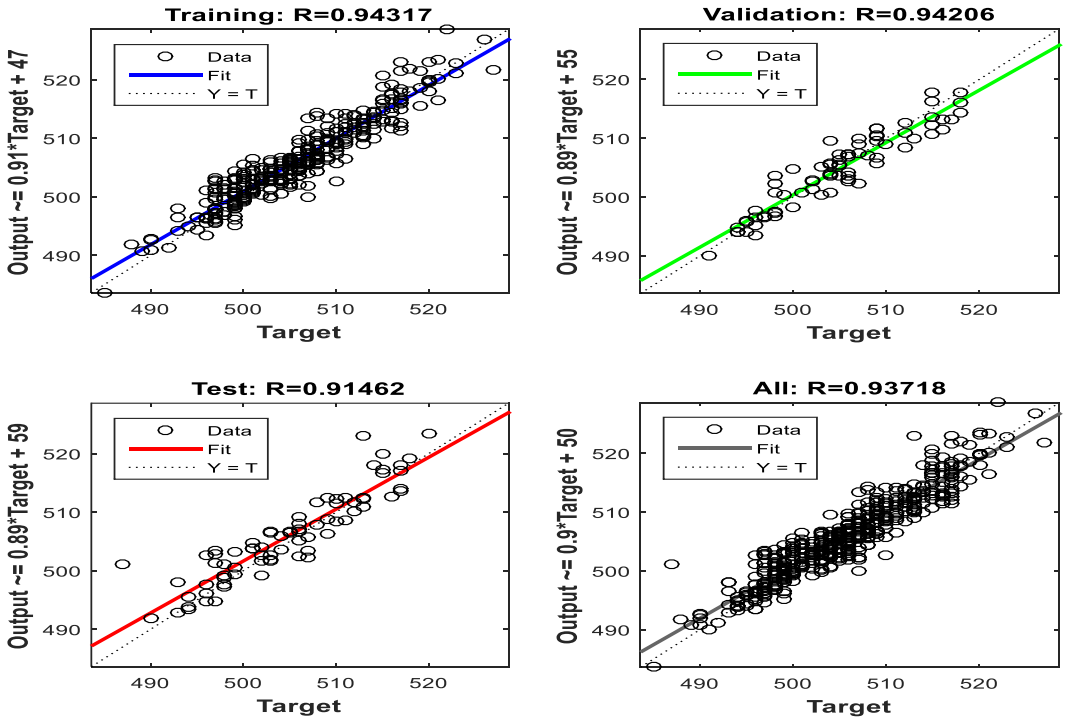

Figure 7. R-values

Figure 7 shows the performance of the regression analysis of our new method. The R-values for training output, validation output, and test output show that our proposed method is promising and encouraging.

3.3.7 Training State

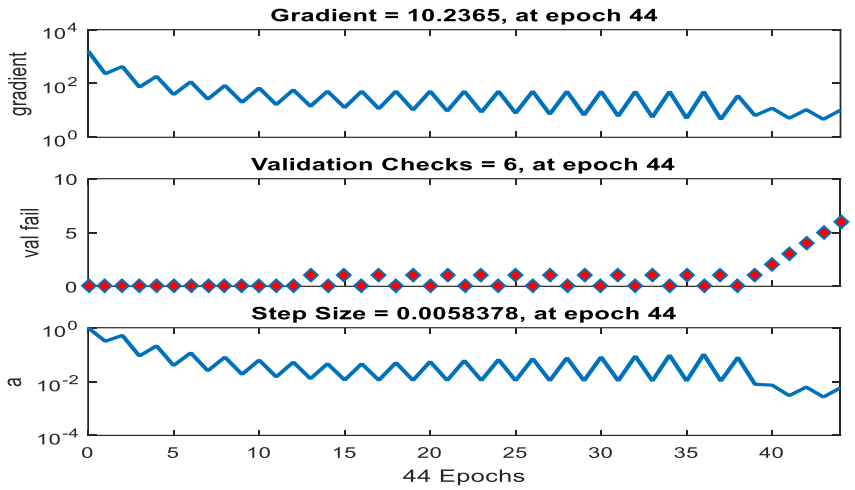

Figure 8. Training state

Figure 8 shows that the gradient value of our method is 777.6379 at iteration (epoch) 37. At epoch 37, the validation checks equal 6 and the value of the step size is 0.0092474.

4. CONCLUSIONS

In this paper, we propose a new conjugate gradient method for UOP. This method is a simple three-term conjugate gradient method based on the Dai-Yuan conjugate gradient method and Barzilai-Borwein method. An interesting feature of the proposed method is that the method possesses sufficient descent property independent of the line search strategy. From the rigorous numerical results and analysis, the new method demonstrated overwhelming performance as compared to other methods for UOP and neural networks.

ACKNOWLEDGMENT

This study was supported by the Tertiary Education Trust Fund research grant of Nigeria number (TETF/DR&D/CE/UNI/ZURU/IBR/2024/VOL.1).

REFERENCES

[1] V. Dhaka and M. P. Singh, "HANDWRITTEN CHARACTER RECOGNITION USING MODIFIED GRADIENT DESCENT TECHNIQUE OF NEURAL NETWORKS AND REPRESENTATION OF CONJUGATE DESCENT FOR TRAINING PATTERNS," *Int. J. Eng.*, vol. 22, no. 2, pp. 145–158, Jun. 2009.

[2] N. Sato, K. Izumi, and H. Iiduka, "SCALED CONJUGATE GRADIENT METHOD FOR NONCONVEX OPTIMIZATION IN DEEP NEURAL NETWORKS," *J. Mach. Learn. Res.*, vol. 25, no. 395, pp. 1–37, 2024.

[3] Z. Yang, "ADAPTIVE STOCHASTIC CONJUGATE GRADIENT FOR MACHINE LEARNING," *Expert Syst. Appl.*, vol. 206, p. 117719, Nov. 2022, doi: https://doi.org/10.1016/j.eswa.2022.117719.

[4] W. Huang and L. Ju, "A NOVEL INTRUSION DETECTION METHOD BASED ON CONJUGATE GRADIENT NEURAL NETWORK," in *2010 International Conference On Computer Design and Applications*, Jun. 2010, pp. V2-470-V2-472. doi: https://doi.org/10.1109/ICCDA.2010.5541362.

[5] H. Zhu, Y. Xiao, and S.-Y. Wu, "LARGE SPARSE SIGNAL RECOVERY BY CONJUGATE GRADIENT ALGORITHM BASED ON SMOOTHING TECHNIQUE," *Comput. Math. Appl.*, vol. 66, no. 1, pp. 24–32, Aug. 2013, doi: https://doi.org/10.1016/j.camwa.2013.04.022.

[6] I. M. Sulaiman, M. Malik, W. Giyarti, M. Mamat, M. A. H. Ibrahim, and M. Z. Ahmad, "The Application of Conjugate Gradient Method to Motion Control of Robotic Manipulators," in *Enabling Industry 4.0 through Advances in Mechatronics*, Eds., Singapore: Springer Nature, 2022, pp. 435–445. doi: https://doi.org/10.1007/978-981-19-2095-0_37.

[7] B. Paden, M. Čáp, S. Z. Yong, D. Yershov, and E. Frazzoli, "A SURVEY OF MOTION PLANNING AND CONTROL TECHNIQUES FOR SELF-DRIVING URBAN VEHICLES," *IEEE Trans. Intell. Veh.*, vol. 1, no. 1, pp. 33–55, Mar. 2016, doi: https://doi.org/10.1109/TIV.2016.2578706.

[8] I. M. Navon and D. M. Legler, "CONJUGATE-GRADIENT METHODS FOR LARGE-SCALE MINIMIZATION IN METEOROLOGY," Aug. 1987, doi: https://doi.org/10.1175/1520-0493(1987)115<1479:CGMFLS>2.0.CO;2
[9] A. Muhammed Awwal, P. Kumam, K. Sitthithakerngkiet, A. Muhammad, A. Halilu, and S. Ibrahim, "DERIVATIVE-FREE METHOD BASED ON DFP UPDATING FORMULA FOR SOLVING CONVEX CONSTRAINED NONLINEAR MONOTONE EQUATIONS AND APPLICATION," *AIMS Math.*, vol. 6, pp. 8792–8814, Jun. 2021, doi: 10.3934/math.2021510.
[10] S. Ibrahim, M. Mamat, A. Abashar, M. Rivaie, and Z. Salleh, "A MODIFIED NONLINEAR CONJUGATE GRADIENT METHOD FOR UNCONSTRAINED OPTIMIZATION," *Appl. Math. Sci.*, vol. 9, pp. 2671–2682, Jan. 2015, doi: https://doi.org/10.12988/ams.2015.5141.
[11] M. Malik, M. Mamat, S. Abas, S. Ibrahim, and S. Firman, "PERFORMANCE ANALYSIS OF NEW SPECTRAL AND HYBRID CONJUGATE GRADIENT METHODS FOR SOLVING UNCONSTRAINED OPTIMIZATION PROBLEMS," *IAENG Int. J. Comput. Sci.*, vol. 48, pp. 66–79, Feb. 2021.
[12] R. Fletcher and C. M. Reeves, "FUNCTION MINIMIZATION BY CONJUGATE GRADIENTS," *Comput. J.*, vol. 7, no. 2, pp. 149–154, Jan. 1964, doi: https://doi.org/10.1093/comjnl/7.2.149.
[13] E. Polak and G. Ribiere, "NOTE SUR LA CONVERGENCE DE MÉTHODES DE DIRECTIONS CONJUGUÉES," *Rev. Fr. Inform. Rech. Opérationnelle Sér. Rouge*, vol. 3, no. 16, Art. no. 16, 1969, doi: https://doi.org/10.1051/m2an/196903R100351.
[14] B. T. Polyak, "THE CONJUGATE GRADIENT METHOD IN EXTREMAL PROBLEMS," *USSR Comput. Math. Math. Phys.*, vol. 9, no. 4, pp. 94–112, Jan. 1969, doi: https://doi.org/10.1016/0041-5553(69)90035-4.
[15] M. R. Hestenes and E. Stiefel, "METHODS OF CONJUGATE GRADIENTS FOR SOLVING LINEAR SYSTEMS," *J. Res. Natl. Bur. Stand.*, vol. 49, no. 6, p. 409, Dec. 1952, doi: https://doi.org/10.6028/jres.049.044.
[16] Y. H. Dai and Y. Yuan, "A NONLINEAR CONJUGATE GRADIENT METHOD WITH A STRONG GLOBAL CONVERGENCE PROPERTY," *SIAM J. Optim.*, vol. 10, no. 1, pp. 177–182, Jan. 1999, doi: https://doi.org/10.1137/S1052623497318992.
[17] Y. Liu and C. Storey, "EFFICIENT GENERALIZED CONJUGATE GRADIENT ALGORITHMS, PART 1: THEORY," *J. Optim. Theory Appl.*, vol. 69, no. 1, pp. 129–137, Apr. 1991, doi: https://doi.org/10.1007/BF00940464.
[18] M. Malik, S. S. Abas, M. Mamat, Sukono, and I. S. Mohammed, "A NEW HYBRID CONJUGATE GRADIENT METHOD WITH GLOBAL CONVERGENCE PROPERTIES," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 05, Art. no. 05, Apr. 2020.
[19] S. M. Ibrahim, N. Idalisa, N. Zullpakkal, P. Kaelo, and N. Shapiee, "A DESCENT THREE-TERM CONJUGATE GRADIENT DIRECTION FOR PROBLEMS OF UNCONSTRAINED OPTIMIZATION WITH APPLICATION," *Int. J. Comput. Methods*, Feb. 2025, doi: https://doi.org/10.1142/S0219876225500100.
[20] A. Mousavi, M. Esmaeilpour, and A. Sheikhahmadi, "TWO EFFICIENT THREE-TERM CONJUGATE GRADIENT METHODS FOR IMPULSE NOISE REMOVAL FROM MEDICAL IMAGES," *Multimed. Tools Appl.*, vol. 83, no. 15, pp. 43685–43703, May 2024, doi: https://doi.org/10.1007/s11042-023-17352-z.
[21] A. Alhawarat, Z. Salleh, H. Alolaiyan, H. El Hor, and S. Ismail, "A THREE-TERM CONJUGATE GRADIENT DESCENT METHOD WITH SOME APPLICATIONS," *J. Inequalities Appl.*, vol. 2024, no. 1, p. 73, May 2024, doi: https://doi.org/10.1186/s13660-024-03142-0.
[22] J. BARZILAI and J. M. BORWEIN, "TWO-POINT STEP SIZE GRADIENT METHODS," *IMA J. Numer. Anal.*, vol. 8, no. 1, pp. 141–148, Jan. 1988, doi: https://doi.org/10.1093/imanum/8.1.141.
[23] I. M. Sulaiman, Sukono, S. Supian, and M. Mamat, "NEW CLASS OF HYBRID CONJUGATE GRADIENT COEFFICIENTS WITH GUARANTEED DESCENT AND EFFICIENT LINE SEARCH," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 621, no. 1, p. 012021, Oct. 2019, doi: https://doi.org/10.1088/1757-899X/621/1/012021.
[24] N. Andrei, "A SIMPLE THREE-TERM CONJUGATE GRADIENT ALGORITHM FOR UNCONSTRAINED OPTIMIZATION," *J. Comput. Appl. Math.*, vol. 241, pp. 19–29, Mar. 2013, doi: https://doi.org/10.1016/j.cam.2012.10.002.
[25] Y. -H. Dai and L. -Z. Liao, "NEW CONJUGACY CONDITIONS AND RELATED NONLINEAR CONJUGATE GRADIENT METHODS," *Appl. Math. Optim.*, vol. 43, no. 1, pp. 87–101, Jan. 2001, doi: https://doi.org/10.1007/s002450010019.
[26] R. Fletcher, "A NEW APPROACH TO VARIABLE METRIC ALGORITHMS," *Comput. J.*, vol. 13, no. 3, pp. 317–322, Jan. 1970, doi: https://doi.org/10.1093/comjnl/13.3.317.
[27] B. E. M, "A DERIVATION OF CONJUGATE-GRADIENTS," *Numer. Methods Non-Linear Optim.*, 1972, Accessed: Feb. 26, 2025. [Online]. Available: https://cir.nii.ac.jp/crid/1573950399913297920
[28] L. Nazareth, "A Conjugate Direction Algorithm Without Line Searches," *J. Optim. Theory Appl.*, vol. 23, no. 3, pp. 373–387, Nov. 1977, doi: https://doi.org/10.1007/BF00933447.
[29] W. W. Hager and H. Zhang, "A SURVEY OF NONLINEAR CONJUGATE GRADIENT METHODS".
[30] L. Zhang, W. Zhou, and D. Li, "SOME DESCENT THREE-TERM CONJUGATE GRADIENT METHODS AND THEIR GLOBAL CONVERGENCE," *Optim. Methods Softw.*, vol. 22, no. 4, pp. 697–711, Aug. 2007, doi: https://doi.org/10.1080/10556780701223293.
[31] N. Deng and Z. Li, "GLOBAL CONVERGENCE OF THREE TERMS CONJUGATE GRADIENT METHODS ∗," *Optim. Methods Softw.*, Jan. 1994, doi: https://doi.org/10.1080/10556789508805593.
[32] W. W. Hager and H. Zhang, "A NEW CONJUGATE GRADIENT METHOD WITH GUARANTEED DESCENT AND AN EFFICIENT LINE SEARCH," *SIAM J. Optim.*, vol. 16, no. 1, pp. 170–192, Jan. 2005, doi: https://doi.org/10.1137/030601880.
[33] A. A. Goldstein, "ON STEEPEST DESCENT," *J. Soc. Ind. Appl. Math. Ser. Control*, vol. 3, no. 1, pp. 147–151, Jan. 1965, doi: https://doi.org/10.1137/0303013.
[34] A. L. Ibrahim and M. G. Mohammed, "A NEW THREE-TERM CONJUGATE GRADIENT METHOD FOR TRAINING NEURAL NETWORKS WITH GLOBAL CONVERGENCE," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 28, no. 1, Art. no. 1, Oct. 2022, doi: https://doi.org/10.11591/ijeecs.v28.i1.pp551-558.
[35] N. Andrei, "AN UNCONSTRAINED OPTIMIZATION TEST FUNCTIONS COLLECTION," 2008. Accessed: Feb. 26, 2025. [Online]. Available: https://www.semanticscholar.org/paper/An-Unconstrained-Optimization-Test-Functions-Andrei/3bcfd42ec838188f3366d75d3568c0638a79c8c4
[36] E. D. Dolan and J. J. Moré, "BENCHMARKING OPTIMIZATION SOFTWARE WITH PERFORMANCE PROFILES," *Math. Program.*, vol. 91, no. 2, pp. 201–213, Jan. 2002, doi: https://doi.org/10.1007/s101070100263.

[37] K. Tu, S. Yang, H. Zhang, and Z. Wang, "HYBRID A∗ BASED MOTION PLANNING FOR AUTONOMOUS VEHICLES IN UNSTRUCTURED ENVIRONMENT," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2019, pp. 1–4. doi: https://doi.org/10.1109/ISCAS.2019.8702779.
[38] Y. Zhang *et al.*, "HYBRID TRAJECTORY PLANNING FOR AUTONOMOUS DRIVING IN HIGHLY CONSTRAINED ENVIRONMENTS," *IEEE Access*, vol. 6, pp. 32800–32819, 2018, doi: 10 https://doi.org/10.1109/ACCESS.2018.2845448.
[39] H. Zhang, Q. Qi, W. Ji, and F. Tao, "AN UPDATE METHOD FOR DIGITAL TWIN MULTI-DIMENSION MODELS," *Robot. Comput.-Integr. Manuf.*, vol. 80, p. 102481, Apr. 2023, doi: https://doi.org/10.1016/j.rcim.2022.102481.
[40] T. Sutharssan, S. Stoyanov, C. Bailey, and C. Yin, "PROGNOSTIC AND HEALTH MANAGEMENT FOR ENGINEERING SYSTEMS: A REVIEW OF THE DATA-DRIVEN APPROACH AND ALGORITHMS," *J. Eng.*, vol. 2015, no. 7, pp. 215–222, 2015, doi: https://doi.org/10.1049/joe.2014.0303.
[41] G. Ezzati, M. Mammadov, and S. Kulkarni, "A New Reliability Analysis Method Based On The Conjugate Gradient Direction," *Struct. Multidiscip. Optim.*, vol. 51, no. 1, pp. 89–98, Jan. 2015, doi: https://doi.org/10.1007/s00158-014-1113-z.
[42] H. K. Seong, H. Shin, J. Yoo, T. Yamada, and S. Nishiwaki, "REACTION-DIFFUSION EQUATION BASED TOPOLOGY OPTIMIZATION COMBINED WITH THE MODIFIED CONJUGATE GRADIENT METHOD," *Finite Elem. Anal. Des.*, vol. 140, pp. 84–95, Feb. 2018, doi: https://doi.org/10.1016/j.finel.2017.11.009.
[43] L. Montoli, "MULTI-OBJECTIVE TOPOLOGY OPTIMIZATION FOR CONJUGATE HEAT TRANSFER WITH THE ADJOINT METHOD," Apr. 2021, Accessed: Feb. 26, 2025. [Online]. Available: https://www.politesi.polimi.it/handle/10589/174045
[44] K. Uba kamfa, S. Ibrahim, M. Awang Nawi, M. Waziri, and M. Mamat, "PERFORMANCE EVALUATION OF A NOVEL CONJUGATE GRADIENT METHOD FOR TRAINING FEED FORWARD NEURAL NETWORK," *Math. Model. Comput.*, vol. 10, Apr. 2023, doi: 10.23939/mmc2023.02.326.
[45] H. Kim *et al.*, "VARIABLE THREE-TERM CONJUGATE GRADIENT METHOD FOR TRAINING ARTIFICIAL NEURAL NETWORKS," *Neural Netw.*, vol. 159, pp. 125–136, Feb. 2023, doi: https://doi.org/10.1016/j.neunet.2022.12.001.
[46] I. Livieris and P. Pintelas, "Performance Evaluation of Descent CG Methods for Neural Network Training," in *HERMIS-μπ. Hellenic European Research on Mathematics and Informatics Science*, 2009, pp. 40–46.