EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
2025, Vol. 18, Issue 3, Article Number 6377
ISSN 1307-5543 – ejpam.com
Published by New York Business Global

An Enhanced Conjugate Gradient Method for Nonlinear Minimization Problems

Ahmed Anwer Mustafa[1,*], Hussein Ageel Khatab[2]

[1] *Department of Mathematics, College of Education, University of Zakho, Kurdistan Region, Iraq*
[2] *Department of Mathematics, College of Science, University of Zakho, Kurdistan Region, Iraq*

Abstract. Because of their computing efficiency and minimal memory requirements, conjugate gradient techniques are a fundamental family of algorithms for handling large-scale unconstrained nonlinear optimization problems. A new version of the Hestenes-Stiefel (HS) technique is presented in this study with the goal of improving convergence properties without compromising ease of use. We rigorously prove the global convergence qualities of the proposed approach under standard assumptions and show that it meets the conjugacy, descent, and adequate descent constraints. Numerous numerical tests, covering a wide range of benchmark issues, show that the suggested strategy routinely performs better than the traditional HS approach in terms of function evaluations and iteration count.
2020 Mathematics Subject Classifications: 65K05, 46N10, 90C26, 90C30
Key Words and Phrases: Nonlinear Optimization, Unconstrained Optimization, Conjugate Gradient Method, Descent Condition, Global Convergence

1. Introduction

With applications in scientific computing, machine learning, and engineering design, nonlinear unconstrained optimization remains a fundamental topic of study in numerical optimization [1]. Iterative techniques like conjugate gradient (CG) algorithms work well because they can strike a compromise between computing economy and convergence guarantees. Numerous alterations to traditional CG techniques have been put forth recently in an effort to boost efficiency and guarantee convergence under lax circumstances [2, 3]. Notably, three-term formulations have demonstrated potential for reducing storage needs and speeding up convergence [4].

Global convergence of these CG variations has been made easier by improvements in the line search algorithms, especially those that meet the strong Wolfe criteria [5]. Spectral

*Corresponding author.
DOI: https://doi.org/10.29020/nybg.ejpam.v18i3.6377

Email addresses: ahmed.mustafa@uoz.edu.krd (A. A. Mustafa), hussein.khatab@uoz.edu.krd (H. A. Khatab)

conjugate gradient techniques, which dynamically integrate curvature information into parameter updates, have also received fresh attention [6, 7]. Large-scale optimization situations, such as deep learning training procedures and sparse regression, have benefited from such advancements [8].

Furthermore, by carefully examining the adequate descent and conjugacy aspects of CG approaches, current research has shown their theoretical soundness [9, 10]. These advancements highlight the necessity of more research into hybrid or parameter-adaptive CG techniques that maintain desired convergence behavior in a variety of issue scenarios. This paper contributes to this ongoing research by introducing a modified CG method with demonstrable performance gains on classical benchmark problems.

Consider the unconstrained optimization problem:

$$\min f(x), \quad x \in \mathbb{R}^n \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a real-valued, continuously differentiable function.

A nonlinear conjugate gradient method generates a sequence $\{x_k\}$, $k \geq 0$, starting from an initial guess $x_0 \in \mathbb{R}^n$, using the recurrence

$$x_{k+1} = x_k + \alpha_k d_k \tag{1.2}$$

where $v_k = x_{k+1} - x_k$, α_k is positive step size and is obtained by some line searches, and d_k is a search direction. The conjugate gradient method's simplicity and extremely low memory requirements make it a potent line search technique for resolving large-scale optimization problems.

The search direction d_k of the conjugate gradient methods is given by:

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \quad k \geq 0, \quad d_0 = -g_0 \tag{1.3}$$

where $g_{k+1} = \nabla f(x_{k+1})$ and β_k is a scalar. Many formulas for β_k are suggested, like Fletcher-Reeves (FR) [11], Dai and Yuan (DY) [12], Hestenes-Stiefel (HS) [13], Polak-Ribiere-Polyak (PRP) [14], conjugate descent (CD) [15], and Liu-Storey (LS) method [16], these formulas are as follows:

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \tag{1.4}$$

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T (g_{k+1} - g_k)} \tag{1.5}$$

$$\beta_k^{HS} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{d_k^T (g_{k+1} - g_k)} \tag{1.6}$$

$$\beta_k^{PRP} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{\|g_k\|^2} \tag{1.7}$$

$$\beta_k^{CD} = \frac{\|g_{k+1}\|^2}{-d_k^T g_k} \tag{1.8}$$

$$\beta_k^{LS} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{-d_k^T g_k} \tag{1.9}$$

where $y_k = g_{k+1} - g_k$, and the symbol $\| \cdot \|$ denotes the Euclidean norm of vectors.

Also, many parameters are proposed: Hussein Ageel Khatab and Salah Gazi Shareef suggested a new conjugate gradient method for solving nonlinear unconstrained optimization problems by using three terms conjugate gradient method [4]. Ahmed Anwer Mustafa [6] proposed a novel algorithm to perform spectral conjugate gradient descent for an unconstrained, nonlinear optimization problem. The same researcher suggested another algorithm, see [7].

The properties of conjugate gradient methods have been explored greatly through their global convergence properties. Global convergence results for the FR method without regular restarts and in exact line search have been obtained by Zoutendijk [17], and by Al-Baali [18] in connection with inexact line searches. Dai and Yuan [12] shown that the DY method is descent and globally convergent if the Wolfe line search

$$f(x_k + \alpha_k d_k) - f(x_k) \leq c_1 \alpha_k g_k^T d_k, \quad g_{k+1}^T d_k \geq c_2 g_k^T d_k$$

is used.

The structure of this paper is as follows: In Section 2, a new conjugate gradient method will be proposed. Its conjugacy condition, descent condition, and sufficient descent condition will be proved in Section 3. In Section 4, its global convergence will be examined. Some numerical experiments about this new conjugate gradient method will be presented in Section 5, and the conclusion will be provided in Section 6.

2. New Conjugate Gradient Method and Its Algorithm

In this section, we develop a modified nonlinear HS method for solving nonlinear unconstrained optimization problems.

To obtain the generated sufficient descent direction, Hager and Zhang [19] showed a new conjugate gradient method (CG-C) obtained by modifying the HS method, which generates sufficient descent directions at each step, without relying on any line search. The scalar β_k in CG-C method is determined by

$$\beta_k^{NHS} = \max\left\{\beta_k^N,\ \mu_k\right\} \tag{2.1}$$

where

$$\beta_k^N = \frac{g_{k+1}^T y_k}{d_k^T y_k} - \frac{2\|y_k\|^2 g_{k+1}^T d_k}{(d_k^T y_k)^2},$$

$$\mu_k = \frac{1}{\|d_k\| \min\{\|g_k\|, \mu\}},$$

and $\mu > 0$ is a constant. With the weak Wolfe-Powell line search, Hager and Zhang [19] established a global convergence result for (2.1) when the objective function $f(x)$ is a general nonlinear function.

Also, Hussein Ageel Khatab and Salah Gazi Shareef [20] suggested the following conjugate gradient method

$$\beta_k^{NEW} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - \frac{g_{k+1}^T v_k}{d_k^T y_k} - \mu \frac{g_{k+1}^T d_k}{d_k^T y_k \, \|g_k\|^2} g_k^T y_k \tag{2.2}$$

where $\mu \in (0,1)$.

We suggest a new parameter as follows:

$$\beta_k^{New} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - \mu \frac{\|g_{k+1}\|^2 \|v_k\|^2 \|x_{k+1}\|}{(d_k^T y_k)^2} g_{k+1}^T d_k \tag{2.3}$$

where $\mu > 0$.

Algorithm of The New Conjugate Gradient Method (β_k^{NEW})

Step 1: Select x_0 and $\varepsilon = 10^{-5}$.

Step 2: Set

$$d_0 = -g_0, \quad g_k = \nabla f(x_k), \quad \text{Set } k = 0.$$

Step 3: Compute the step length $\alpha_k > 0$ satisfying the Wolfe line search conditions

$$f(x_k + \alpha_k d_k) - f(x_k) \le c_1 \alpha_k g_k^T d_k,$$

$$|g_{k+1}^T d_k| \le c_2 |g_k^T d_k|,$$

where $0 < c_1 < c_2 < 1$.

Step 4: Compute

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

$$g_{k+1} = \nabla f(x_{k+1}), \quad \text{If } \|g_{k+1}\| \le \varepsilon, \text{ then stop.}$$

Step 5: Compute β_k^{New} by equation (2.3).

Step 6: Compute

$$d_{k+1} = -g_{k+1} + \beta_k^{New} d_k.$$

Step 7: If

$$|g_{k+1}^T g_k| > 0.2 \|g_{k+1}\|^2,$$

then go to Step 2. Otherwise,

$$k = k + 1,$$

and go to Step 3.

3. Conjugacy Condition, Descent Condition, and Sufficient Condition

In this section, we will prove the Conjugacy condition, Descent Condition, and Sufficient of the new method.

Theorem 1: - Assume that the sequence $\{x_k\}$ is generated by (1.2), then the new method (2.3) satisfies the conjugacy condition, i.e. $d_{k+1}^T y_k = t\, g_{k+1}^T v_k$, where $t > 0$.

Proof: - By putting (2.3) in equation (1.3) to get

$$d_{k+1} = -g_{k+1} + \beta_k^{new} d_k$$

or,

$$d_{k+1} = -g_{k+1} + \left(\frac{g_{k+1}^T y_k}{d_k^T y_k} - \mu \frac{\|g_{k+1}\|^2 \|v_k\|^2 \|x_{k+1}\|}{(d_k^T y_k)^2} g_{k+1}^T d_k \right) d_k \tag{2.4}$$

Multiply both sides of equation (2.4) by y_k from right-hand side, we get

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

or,

$$d_{k+1}^T y_k = -\mu \frac{\|g_{k+1}\|^2 \|v_k\|^2 \|x_{k+1}\|}{d_k^T y_k} g_{k+1}^T d_k \tag{2.5}$$

implies that

$$d_{k+1}^T y_k = -\left(\mu \frac{1}{\alpha_k} \frac{\|g_{k+1}\|^2 \|v_k\|^2 \|x_{k+1}\|}{d_k^T y_k} \right) g_{k+1}^T v_k \tag{2.6}$$

Suppose that $t = \left(\mu \frac{1}{\alpha_k} \frac{\|g_{k+1}\|^2 \|v_k\|^2 \|x_{k+1}\|}{d_k^T y_k} \right)$
then,

$$d_{k+1}^T y_k = -t\, g_{k+1}^T v_k \qquad \square$$

Theorem 2: - Assume that the sequence $\{x_k\}$ is generated by (1.2), then the search direction (1.3) of the new method (2.3) satisfies the descent condition, i.e. $d_{k+1}^T g_{k+1} \leq 0$ with exact and inexact line search.

Proof: - We will prove by mathematical induction. If $k = 0$, then,

$$d_0 = -g_0, \quad d_0^T g_0 = -g_0^T g_0 = -\|g_0\|^2 \leq 0,$$

We suppose that

$$d_k^T g_k \leq 0.$$

Now, we prove the case $k+1$. From (1.3) and (2.3), we have (2.4), multiply both sides of equation (2.4) by g_{k+1} from the right-hand side, we get

$$d_{k+1}^T g_{k+1} = -g_{k+1}^T g_{k+1} + \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k^T g_{k+1} - \mu \frac{\|g_{k+1}\|^2 \|v_k\|^2 \|x_{k+1}\|}{(d_k^T y_k)^2} g_{k+1}^T d_k d_k^T g_{k+1} \tag{2.7}$$

If the step length α_k is chosen by an exact line search which requires $d_k^T g_{k+1} = 0$, then the proof is complete. If the step length α_k is chosen by inexact line search which requires $d_k^T g_{k+1} \neq 0$, then the first two terms on the right-hand side of equation (2.7) are less than or equal to zero since HS satisfies the descent condition. The third term is less than or equal to zero, so we get

$$d_{k+1}^T g_{k+1} \leq -\mu \frac{\|g_{k+1}\|^2 \|v_k\|^2 \|x_{k+1}\|}{(d_k^T y_k)^2} (g_{k+1}^T d_k)^2 \tag{2.8}$$

Then, $d_{k+1}^T g_{k+1} \leq 0$. □

Theorem 3: - Assume that the sequence $\{x_k\}$ is generated by (1.2), then the search direction (1.3) with the new method (2.3) satisfies the sufficient descent condition, i.e.

$$d_{k+1}^T g_{k+1} \leq -C \|g_{k+1}\|^2.$$

Proof: - We multiply the right-hand side of inequality (2.8) by $\frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2}$, we get

$$d_{k+1}^T g_{k+1} \leq -\mu \frac{\|g_{k+1}\|^2 \|v_k\|^2 \|x_{k+1}\|}{(d_k^T y_k)^2} (g_{k+1}^T d_k)^2 \left(\frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^2} \right) \tag{2.9}$$

Suppose that

$$C = \mu \frac{\|g_{k+1}\|^2 \|v_k\|^2 \|x_{k+1}\|}{\|g_{k+1}\|^2 (d_k^T y_k)^2} (g_{k+1}^T d_k)^2$$

Then,

$$d_{k+1}^T g_{k+1} \leq -C \|g_{k+1}\|^2 \qquad \square$$

4. Convergence Analysis

Assume that:

- The level set $S = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ is bounded.
- In a neighborhood N of S, the function f is continuously differentiable and its gradient is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L \|x - y\| \quad \text{for all } x, y \in N.$$

Under these assumptions on f, there exists a constant $\rho \geq 0$ such that

$$\|g(x)\| \leq \rho, \quad \text{for all } x \in S.$$

In [5], it is proved that for any conjugate gradient method with a strong Wolfe line search the following general result holds:

Lemma 1: Let assumption (i) and (ii) hold and consider any conjugate gradient method (1.3) and (2.3), where d_k is a descent direction and α_k is obtained by the strong Wolfe line search. If

$$\sum_{k \geq 1} \frac{1}{\|d_{k+1}\|^2} = \infty \tag{4.1}$$

then,

$$\lim_{k \to \infty} \|g_{k+1}\| = 0. \tag{4.2}$$

For a uniformly convex function which satisfies the above assumptions, we can prove that the norm of d_{k+1} given by (1.3) and (2.3) is bounded above. Assume that the function f is uniformly convex, i.e., there exists a constant $\delta \geq 0$, such that for all $x, x_k \in S$

$$(g(x) - g(x_k))^T (x - x_k) \geq \delta \|x - x_k\|^2. \tag{4.3}$$

and the step length α_k is given by a strong Wolfe line search:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k g_k^T d_k \tag{4.4}$$

$$\left|\nabla f(x_k + \alpha_k d_k)^T d_k\right| \leq c_2 |g_k^T d_k| \tag{4.5}$$

Using Lemma 1 the following result can be proved.

Theorem 4: Suppose that the assumptions (i) and (ii) hold. Consider the algorithm (2.3) and where $\mu > 0$ and α_k is obtained by a strong Wolfe line search. If d_k tends to zero and there exist nonnegative constants η_1 and η_2 such that

$$\|g_k\|^2 \geq \eta_1 \|v_k\|^2 \quad \text{and} \quad \|g_{k+1}\|^2 \leq \eta_2 \|v_k\| \tag{4.6}$$

and f is a uniformly convex function, then

$$\lim_{k \to \infty} g_{k+1} = 0 \tag{4.7}$$

Proof: Consider the new method

$$\beta_k^{New} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - \mu \frac{\|g_{k+1}\|^2 \|v_k\|^2 \|x_{k+1}\|}{(d_k^T y_k)^2} g_{k+1}^T d_k$$

or,

$$|\beta_k^{New}| = |\frac{g_{k+1}^T y_k}{d_k^T y_k} - \mu \frac{\|g_{k+1}\|^2 \|v_k\|^2 \|x_{k+1}\|}{(d_k^T y_k)^2} g_{k+1}^T d_k|$$

This implies that

$$|\beta_k^{new}| \leq \frac{\|g_{k+1}\|\|y_k\|}{d_k^T y_k} + \mu \frac{\|g_{k+1}\|^2\|v_k\|^2\|x_{k+1}\|}{\alpha_k (d_k^T y_k)^2}\|g_{k+1}\|\|v_k\|$$

Here,

$$|\beta_k^{new}| \leq \frac{\rho L\|v_k\|\|v_k\|}{\delta\|v_k\|^2} + \mu \frac{\eta_2\|v_k\|\|v_k\|^2\|x_{k+1}\|}{\alpha_k(\delta\|v_k\|^2)^2}\rho\|v_k\|$$

or,

$$|\beta_k^{new}| \leq \frac{\rho L}{\delta} + \mu\eta_2 \frac{\rho\|x_{k+1}\|}{\alpha_k \delta}$$

Now,

$$\|d_{k+1}\| \leq \|g_{k+1}\| + \frac{1}{\alpha_k}\left(\frac{\rho L}{\delta} + \mu\eta_2 \frac{\rho\|x_{k+1}\|}{\alpha_k \delta}\right)\|v_k\|$$

Implies that

$$\|d_{k+1}\| \leq \rho + \frac{1}{\alpha_k}\left(\frac{\rho L}{\delta} + \mu\eta_2 \frac{\rho\|x_{k+1}\|}{\alpha_k \delta}\right) D_1$$

Where $D_1 = \{y - z\},\ y, z \in S$ is the diameter of the level set S. There exists a constant D_2 such that $\|x_{k+1}\| \leq D_2$, then

$$\|d_{k+1}\| \leq \rho + \frac{1}{\alpha_k}\left(\frac{\rho L}{\delta} + \mu\eta_2 \frac{\rho D_2}{\alpha_k \delta}\right) D_1$$

Showing that (4.1) is true. By Lemma 1, it follows that (4.2) is true, which for uniformly convex functions is equivalent to (4.7).

5. Numerical Results

We compared the new method with the conjugate gradient (HS) method in this section, the comparative tests involve Well-known nonlinear problems (standard test function) with different dimensions $n = 10,\ 300,\ 1000,$ and 4000, all programs are written in FORTRAN90 language and for all cases the stopping condition is $\|g_{k+1}\| \leq 10^{-5}$. The results given in Table 1 specifically quote the number of function NOF and the number of iteration NOI. More experimental results in Table 1 and the figures confirm that the new conjugate gradient method is superior to the standard conjugate gradient (HS) method concerning the NOI and NOF. The performance results are shown in Figures 1 and 2, using the performance profile introduced by Dolan and More (2002)[21].

Table 1: Comparative Performance of the Two Algorithms (HS and New Method)

No	Test Function	N	NOI (HS)	NOF (HS)	NOI (New)	NOF (New)
1	Wood	10	30	68	30	68
		300	30	68	26	60
		1000	30	68	31	71
		4000	30	68	24	57
2	OSP	10	13	58	13	58
		300	88	290	79	251
		1000	156	473	140	419
		4000	230	700	228	678
3	Powell	10	38	108	39	119
		300	40	122	34	109
		1000	41	124	30	92
		4000	41	124	44	115
4	Miele	10	31	102	31	102
		300	40	146	40	147
		1000	46	176	39	139
		4000	54	211	35	122
5	G-Central	10	22	159	22	159
		300	23	171	23	170
		1000	23	171	22	164
		4000	28	248	31	269
6	Shallow	10	8	21	8	21
		300	8	21	8	21
		1000	9	24	8	21
		4000	9	24	8	21
7	Beal	10	11	28	11	28
		300	12	30	11	28
		1000	12	30	9	25
		4000	12	30	13	33
8	Sum	10	6	34	6	34
		300	19	105	19	105
		1000	23	128	23	128
		4000	32	142	31	128
9	Wolfe	10	32	65	32	65
		300	48	97	44	89
		1000	70	141	45	93
		4000	169	355	55	115
10	Cubic	10	13	37	13	37
		300	13	37	13	37
		1000	13	37	13	37
		4000	13	37	13	37
11	Non-Diagonal	10	26	72	26	72
		300	29	79	30	81
		1000	29	79	27	77
		4000	F	F	F	F
12	Rosen	10	30	83	30	83
		300	30	83	30	83
		1000	30	83	30	83
		4000	30	83	30	83
13	TRI	10	F	F	9	19
		300	148	297	147	295
		1000	288	577	288	577
		4000	600	1201	600	1201
14	G-Dixon	10	21	45	21	45
		300	508	1136	536	1206
		1000	4005	8015	534	1163
		4000	560	1217	505	1107
15	Fred	10	8	23	8	23
		300	8	23	8	23
		1000	8	23	8	23
		4000	8	23	8	23

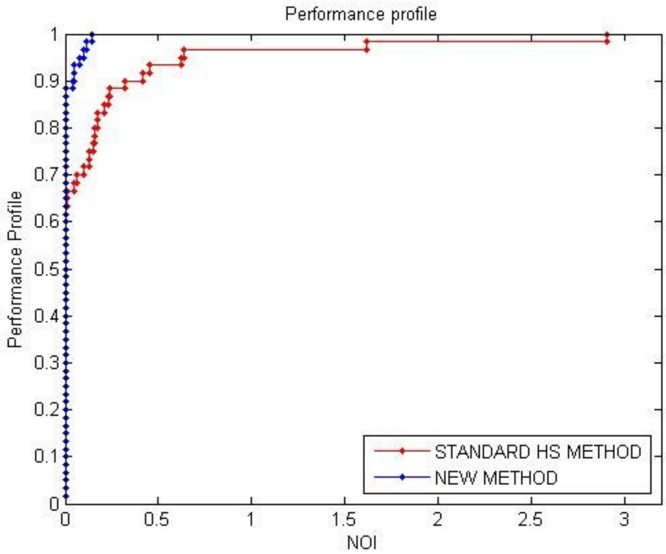

Figure 1: Performance profile on the number of iterations

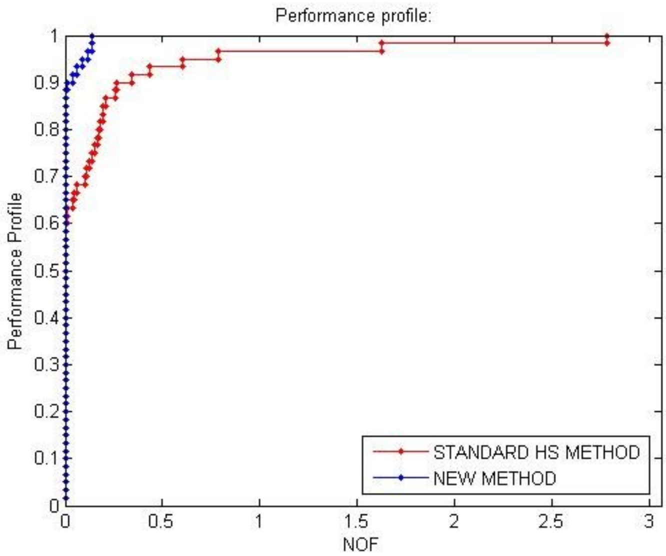

Figure 2: Performance profile on the number of functions

6. Conclusion

In this paper, we suggested a new conjugate gradient method for unconstrained optimization problems. We proved the conjugacy condition, descent condition, sufficient descent condition, and the global convergence of the new method. Implemented and tested to some extent. The new method was compared with the standard conjugate gradient (HS) method. The numerical tests were carried out on low- and high-dimensionality problems and comparisons were made amongst different test functions with inexact line search. We used fifteen function with different dimensions (10, 300, 1000, and 4000). Two important measures are used to assess the approaches' performance: the number of iterations (NOI) and the number of function evaluations (NOF). The numerical results in Table 1 and the figures 1 and 2 demonstrate that the proposed method performs effectively.

References

[1] I. Ahmed and B. Mohanty. Recent advances in nonlinear optimization methods for machine learning models. *Applied Intelligence*, 53:1587–1603, 2023.

[2] R. Bakar and Z. J. Weng. An adaptive conjugate gradient method with descent property. *Numerical Algorithms*, 94:107–129, 2023.

[3] H. Chen and J. Zhang. Three-term conjugate gradient algorithms with inexact line search for large-scale optimization. *Optimization Letters*, 16(8):2315–2333, 2022.

[4] A. K. Hussein and S. G. Shareef. A new parameter conjugate gradient method based on three terms unconstrained optimization. *General Letters in Mathematics*, 7(1):39–44, 2019.

[5] Y. H. Dai, J. Y. Han, G. H. Liu, D. F. Sun, H. X. Yin, and Y. X. Yuan. Convergence properties of nonlinear conjugate gradient methods. *SIAM Journal on Optimization*, 10:348–358, 1999.

[6] A. A. Mustafa. New spectral ls conjugate gradient method for nonlinear unconstrained optimization. *International Journal of Computer Mathematics*, 100(4):838–846, 2023.

[7] A. A. Mustafa. A new algorithm for spectral conjugate gradient in nonlinear optimization. *Mathematics and Statistics*, 10(2):293–300, 2022.

[8] Y. Liu and H. Wang. Spectral-type cg methods for deep learning optimization. *Journal of Computational Mathematics*, 43(1):45–67, 2024.

[9] A. K. Mohammed and S. R. Al-Taie. Global convergence of parameterized conjugate gradient algorithms. *Mathematics*, 12(3):389–404, 2024.

[10] F. Zhang and C. Li. Sufficient descent and global convergence in nonlinear cg methods with modified β-updates. *Journal of Optimization Theory and Applications*, 195:623–647, 2025.

[11] R. Fletcher and C. M. Reeves. Function minimization by conjugate gradients. *The Computer Journal*, 7(2):149–154, 1964.

[12] Y. H. Dai and Y. Yuan. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization*, 10:177–182, 1999.

[13] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6):409–436, 1952.

[14] E. Polak and G. Ribière. Note sur la convergence de méthodes de directions conjuguées. *Revue Française d'Informatique et de Recherche Opérationnelle*, 16:35–43, 1969.

[15] R. Fletcher. *Practical Methods of Optimization*. Wiley, New York, 2 edition, 1987.

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

[17] G. Zoutendijk. Nonlinear programming, computational methods. In J. Abadie, editor, *Integer and Nonlinear Programming*, pages 37–86. North-Holland, 1970.

[18] A. Al-Baali. Descent property and global convergence of the fletcher-reeves method with inexact line search. *IMA Journal of Numerical Analysis*, 5:121–124, 1985.

[19] W. W. Hager and H. Zhang. A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM Journal on Optimization*, 16(1):170–192, 2005.

[20] S. G. Hussein and A. K. Hussein. *General Letters in Mathematics*, 9(2), 2020.

[21] E. D. Dolan and J. J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2):201–213, 2002.