Large-Scale Optimization Using Modified Memoryless SR1 Algorithm

Radhwan Basim Thanoon and Ghada Moayid Al-Naemi
Department of Mathematics, College of Computer and Mathematics, University of Mosul, 41002 Mosul, Iraq
rathwanbasim@gmail.com, radhwan_81@nan.epedu.gov.iq, radhwan.23csp88@student.uomosul.edu.iq, drghadaalnaemi@uomosul.edu.iq

Keywords: Quasi-Newton SR1, Spectral Parameter, Memoryless Algorithms, Strong Wolfe Criteria Line Search, Dolan and Moré Performance Profiles.

Abstract: In this paper, we present a modified spectral memoryless quasi-Newton method for solving unconstrained optimization problems. The proposed method is based on the symmetric rank-one (SR1) update, which introduces a rank-one correction to the inverse Hessian approximation. To enhance efficiency and robustness, the method employs a non-quadratic spectral parameter derived from gradient information. This parameter approximates curvature information without the need to store or update full matrices, significantly reducing computational cost and making the method suitable for large-scale problems. A key feature of the proposed algorithm is its ability to preserve the descent property of the search direction at each iteration. This is ensured by incorporating a line search procedure that satisfies the strong Wolfe conditions, thereby improving stability and convergence reliability. Theoretical analysis demonstrates that, under standard assumptions, the algorithm converges globally to a stationary point. Extensive numerical experiments conducted on benchmark test functions show that the proposed method is competitive with, and often superior to, several state-of-the-art optimization methods in terms of convergence speed, robustness, and accuracy.

1 INTRODUCTION

Unconstrained optimization is common in scientific and technical professions where the goal is to minimize a continuously differentiable function without using explicit constraints. This paper considers the general unconstrained optimization problem:

$$\text{Minimizer} f(x), \quad x \in \Re^n.\backslash, \tag{1}$$

where $f: \Re^n \to \Re$ is a twice continuously differentiable function that is enclosed below. The gradient of f is denoted as g(x) = ∇f(x). Solving the optimization problem in (1) frequently includes an iterative strategy, that is, starting with x_1 additional formatting to meet the specified criteria outlined below. as the initial position, the next approximation is x_n, To calculate the$x$ solution of (1), apply the formula:

$$x_n = x_{n-1} + \alpha_{n-1} d_{n-1}, \forall n \geq 1. \tag{2}$$

The step size, $\alpha_{n-1} > 0$, is chosen using line search, and the search direction, d_(n)∈R^n, seeks to reduce the objective function [1]. The step size α_{n-1} can be computed using both exact and inexact line search techniques.

The Conjugate Gradient (CG) framework is typically used to define the search direction:

$$d_n = \begin{cases} -g_n\,, & n = 1 \\ -g_n + \beta_{n-1} d_{n-1}, & n \geq 2 \end{cases}. \tag{3}$$

The conjugate gradient (CG) coefficient, represented by the scalar β_{n-1},, is crucial in establishing the search direction and changes according to the CG formulation used [2]. CG approaches are often categorized into two groups. The first group comprises the classical approaches, such as the Hestenes and Stiefel (1952), the Fletcher-Reeves (FR) method (1964), the Polak-Ribiere (PR) method (1969), the Dai-Yuan (DY) method (1999), and the Dai-Liao (DL) method (2001):

$$\left.\begin{matrix} \beta_{n-1}^{HS} = \frac{g_n^T y_{n-1}}{y_{n-1}^T d_{n-1}}; \; \beta_{n-1}^{FR} = \frac{g_n^T g_{n-1}}{g_n^T g_{n-1}}; \; \beta_{n-1}^{PR} = \frac{g_n^T y_{n-1}}{\|g_{n-1}\|^2} \\ \beta_{n-1}^{DY} = \frac{g_n^T g_n}{y_{n-1}^T d_{n-1}}; \; \beta_{n-1}^{DL} = -\frac{g_n^T (y_{n-1} - t\, s_{n-1})}{d_{n-1}^T y_{n-1}} \end{matrix}\right\}, \tag{4}$$

where $y_{n-1} = g_n - g_{n-1}$ and $s_{n-1} = x_n - x_{n-1} = \alpha_{n-1}d_{n-1}$ and $\|.\|$ is the Euclidean norm of vectors.

The second category comprises a number of modified CG methods that aim to address some of the drawbacks of traditional approaches, such as global convergence, numerical stability, and practical performance. [1]-[4].Accurate line searches are frequently impractical for ensuring convergence and maintaining the effectiveness of the conjugate gradient (CG) search direction, particularly in large-scale issues. To find a minimum, the search direction d_n nmust be downward. Therefore, the step size α_{n-1} that meets the strong Wolfe criteria (SWC) is regarded appropriate [4].

$$f(x_{n-1} + \alpha_{n-1}d_{n-1}) \leq f(x_{n-1}) + \delta_1\alpha_{n-1}g_{n-1}^T d_{n-1} \quad (5)$$

$$|g(x_{n-1} + \alpha_{n-1}d_{n-1})^T d_{k-1}| \leq \delta_2|g_{n-1}^T d_{n-1}|, \quad (6)$$

where $0 < \delta_1 < \delta_2 < 1$ are user-defined parameters that control the accuracy of the line search.

Several solutions have been devised to address the unconstrained optimization problem (1). The Newton technique has a high second-order convergence rate, but only when the $\nabla^2 f(x_n)$ Hessian matrix is positive definite. However, the method does not always ensure that the calculated search direction is a descending direction. Furthermore, each iteration necessitates the accurate computation of the Hessian matrix, which can be computationally intensive, especially for large-scale issues. To address these restrictions, quasi-Newton methods have been developed. These approaches estimate the Hessian matrix to identify the search direction, removing the complexity and, in some situations, can handle non-differentiable objective functions [5].

The BFGS, DFP, and SR1 algorithms are some of the most well-known quasi-Newton approaches. The BFGS approach is well-known for its robustness and efficiency, as it updates an approximation of the inverse Hessian matrix with gradient evaluations while maintaining positive definiteness under moderate conditions. The DFP approach, an earlier formulation, likewise updates the inverse Hessian approximation, but with a somewhat different updating mechanism. Although it is less widely utilized than BFGS, it lay the groundwork for future improvements. Unlike the inverse, the SR1 technique updates the Hessian approximation immediately. It provides more flexibility by not requiring positive definiteness, which can be useful in some situations, particularly where curvature information is difficult to obtain. Each of these methods balances computing efficiency and convergence features, making them useful for large-scale unconstrained optimization [6].

In quasi-Newton methods, the search direction d_n is usually obtained by solving the following linear system:

$$d_n B_n = -g_n. \quad (7)$$

Where B_n is an approximation of the Hessian matrix $\nabla^2 f(x_n)$, and $g_n = \nabla f(x_n)$ is the gradient of the objective function at the point x_n. Alternatively, the search direction can be calculated directly using the inverse Hessian approximation.

$$d_n = -H_n g_n, \quad (8)$$

where $H_n$is an approximation of the inverse Hessian matrix, such that

$$H_n = B_n^{-1}. \quad (9)$$

Novelty of the study: This work introduces a non-quadratic spectral parameter within the scaled memoryless SR1 framework. The proposed parameter enhances the approximation of the curvature while eliminating the need to store or update full Hessian matrices, thus significantly reducing computational cost. Furthermore, it improves the convergence speed and robustness of the algorithm when applied to large-scale unconstrained optimization problems.

1.1 Research Problem

Despite the effectiveness of traditional quasi-Newton methods, their computational demands and storage requirements increase significantly with problem size. In large-scale settings, the explicit storage and updating of Hessian approximations become infeasible. Furthermore, classical models often rely on quadratic assumptions that may not accurately capture the curvature of non-quadratic functions. Hence, there is a need for a memoryless and computationally light quasi-Newton approach that maintains convergence stability without storing full matrices.

1.2 Importance and Practical Relevance

In modern computational environments, large-scale unconstrained optimization plays a vital role in numerous real-world applications, such as image reconstruction, signal processing, and machine learning model training. Developing optimization algorithms that are both memory-efficient and robust can significantly improve the speed and scalability of these applications, making the research highly relevant and impactful for current scientific and engineering challenges.

1.3 Objective of the Research

The primary objective of this study is to approach a modified spectral memoryless quasi-Newton method based on the SR1 update that integrates a non-quadratic spectral parameter derived from gradient information. The aim is to improve convergence stability and computational efficiency for large-scale unconstrained optimization problems. Additionally, this study evaluates the numerical performance of the proposed approach across standard benchmark functions and compares it against established state-of-the-art algorithms [7].

2 THE SCALED MEMORYLESS SR1 (SMSR1) MATRIX

In quasi-Newton methods, the enforcement of the secant equation at each iteration is some critical criteria for updating the Hessian approximation. This condition is written as:

$$B_n s_{n-1} = y_{n-1} \text{ or equivalently. } H_n y_{n-1} = s_{n-1} \quad (10)$$

Memoryless quasi-Newton algorithms have been approach to meet the computational and memory requirements of large-scale unconstrained optimization problems. These methods avoid storing entire matrix approximations by depending just on data from the current and previous iterations. Within this framework, we use both direct and inverse Hessian approximations to create memoryless implementations of the Symmetric Rank-One (SR1) update. These memoryless SR1 techniques are the foundation of the algorithm provided in this paper. Spectral scaling techniques are also used to improve the robustness and numerical stability. This research focuses on the SR1 update, which is obtained by addressing the following problem. Given a symmetric matrix B_{n-1} with vectors s_{n-1} and y_{n-1}, construct a symmetric matrix B_n that has rank one and satisfies the secant equation $B_n s_{n-1} = y_{n-1}$. When $(y_{n-1} - B_{n-1}s_{n-1})^T s_{n-1} \neq 0$, the unique solution is given by:

$$B_n = B_{n-1} + \frac{(y_{n-1}-B_{n-1}s_{n-1})(y_{n-1}-B_{n-1}s_{n-1})^T}{(y_{n-1}-B_{n-1}s_{n-1})^T s_{n-1}}. \quad (11)$$

If $B_{n-1}s_{n-1} = y_{n-1}$, then $B_n = B_{n-1}$and there is no need to update. If $(y_{n-1} - B_n s_{n-1})^T s_{n-1} = 0$ and $y_{n-1} \neq B_{n-1}s_{n-1}$, the SR1 update is undefined.
The Sherman-Morrison-Woodbury formula gives the following SR1 update for the inverse Hessian approximation H_n:

$$H_n = H_{n-1} + \frac{(s_{n-1}-H_{n-1}y_{n-1})(s_{n-1}-H_{n-1}y_{n-1})^T}{(s_{n-1}-H_{n-1}y_{n-1})^T y_{n-1}}. \quad (12)$$

This expression is valid provided the denominator is non-zero and H_n remains invertible. To derive a memoryless SR1 method with a direct Hessian approximation, we assume $B_{n-1} = I$ (the identity matrix) in (11), yielding:

$$B_n = I + \frac{(y_{n-1}-s_{n-1})(y_{n-1}-s_{n-1})^T}{(y_{n-1}-s_{n-1})^T s_{n-1}}. \quad (13)$$

This formulation is simple and does not retain information across iterations. Consequently, the memoryless SR1 method inherits the same limitation as the classical SR1 update: it becomes undefined when the denominator$(y_{n-1} - s_{n-1})^T s_{n-1}$ is zero or near zero.

Similarly, by setting $H_{n-1} = I$ in equation (12), we obtain the memoryless SR1 inverse update:

$$H_n = I + \frac{(s_{n-1}-y_{n-1})(s_{n-1}-y_{n-1})^T}{(s_{n-1}-y_{n-1})^T y_{n-1}}. \quad (14)$$

From this expression, the search direction for the memoryless SR1 method is given by:

$$d_n = -H_n g_n, \quad (15)$$

where g_n is the gradient at iteration n. This again illustrates the simplicity of the approach, as no cumulative Hessian information is stored between iterations. However, as with the classical SR1 method, this memoryless variant is undefined when the denominator in the update formula is zero or nearly zero[8].

While the memoryless SR1 method offers computational simplicity and reduced storage demands, its performance can be sensitive to poor scaling, particularly when the search directions or gradients exhibit large variation in magnitude. To address this issue, spectral scaling can be introduced into the memoryless SR1 framework, resulting in the scaled memoryless SR1 (SMSR1) method.

The key idea is to enhance the update formula by incorporating a scaling factor η_{n-1} that reflects curvature information derived from the iterates. This factor adjusts the base matrix (typically the identity matrix) used in the memoryless formulation. Specifically, the scaled SR1 update is constructed using the modified secant equation (10) with H_n now centered around a scaled identity matrix η_{n-1}, rather than the unscaled identity. The SMSR1 matrix is thus defined as:

$$H_n = \eta_{n-1}\left[I + \frac{(s_{n-1}-y_{n-1})(s_{n-1}-y_{n-1})^T}{(s_{n-1}-y_{n-1})^T y_{n-1}}\right], \quad (16)$$

provided the denominator is non-zero. The choice of the scaling parameter η_{n-1} plays a crucial role in the performance of the algorithm. A common and effective strategy is to use a spectral coefficient, such as the Barzilai–Borwein (BB) [9], type scaling:

$$\eta_{n-1} = \frac{y_{n-1}^T y_{n-1}}{y_{n-1}^T s_{n-1}} \ or\ \eta_{n-1} = \frac{s_{n-1}^T y_{n-1}}{s_{n-1}^T s_{n-1}}.$$

These spectral choices are motivated by quasi-Newton and gradient method literature and are known to incorporate second-order information implicitly while maintaining low computational complexity.

3 MOTIVATION AND NEW NON QUADRATIC PARAMETER USING SMSR1 MATRIX

The search direction using the scaled inverse approximation is then given by (15). which respects the curvature of the objective function more effectively than the unscaled version. In summary, the SMSR1 method retains the simplicity and low memory footprint of its memoryless counterpart, while significantly improving its robustness and numerical behavior through spectral scaling.

This section introduces a new parameter obtained from a non-quadratic model and based on the SMSR1 approach.

By applying (16) to (15), we obtain

$$d_n = -\eta_{n-1}[\, g_n + \frac{(s_{n-1}-y_{n-1})^T g_n}{(s_{n-1}-y_{n-1})^T y_{n-1}}\,(s_{n-1} - y_{n-1})].$$

Multiplying both sides of the above equation by y_{n-1}, yields:

$$d_n^T y_{n-1} = -\eta_{n-1}[\, g_n^T y_{n-1} + (s_{n-1} - y_{n-1})^T y_{n-1}\,]. \quad (17)$$

Multiplying both sides of (3) by y_{n-1}, substituting (17) into it, and performing some algebraic manipulations, we obtain:

$$\beta_{n-1}^{RG} = \frac{\alpha_{n-1} d_{n-1}^T g_n - \eta_{n-1}\, g_n^T y_{n-1}}{d_{n-1}^T y_{n-1}}. \quad (18)$$

In this study, we assess η_{n-1} in (18) using a more generic model than the usual quadratic one, which is offered as a basis for the conjugate gradient technique. Instead of constraining $\eta_{n-1}$to the binary set $[0,1]$ [6]. If $q(x)$ is a quadratic function, then $F(q(x)\,)$ is defined as a nonlinear scaling of $q(x)$, provided the following conditions are satisfied:

$$f(x) = F\big(q(x)\big), \frac{df}{dq} = f' \ > 0 \ \&\ q(x) > 0. \ (19)$$

The above conditions directly determine the following scaling properties:

- The contour lines of $q(x)$ are also contour lines of $f(x)$.
- If x^* is a global minimizer of $q(x)$, then x^* is likewise a global minimizer of $f(x)$.

If x^* is a local minimizer of $q(x)$, it is also a local minimizer of $f(x)$.

Boland (1979) discovered that $q(x)$, and $F(q(x)\,)$ yield identical search paths, preserving the algorithm's limited termination property. A conjugate gradient approach for minimizing the function

$$f(x) = q(q(x))^p \ , \qquad p > 0 \,\&\, x \in R^n.$$

Fried (1971) described a step-by-step approach. We summarize numerous published papers that illustrate special examples of this fundamental structure.

- [7], A polynomial model which has been investigated which is defined by:

$F(q(x)) = \varepsilon_1 q(x) + \frac{1}{2}\varepsilon_2 q^2(x)$, where $\varepsilon_1 \,\&\, \varepsilon_2$ are positive scaler.

- [10] two rational models which have been investigated which is defined by:

$$F\big(q(x)\big) = \frac{(\varepsilon_1 q(x)+1)}{\varepsilon_2 q(x)}, \varepsilon_1 > 0\,, \varepsilon_2 < 0\,,$$

where $\varepsilon_1 \,\&\, \varepsilon_2$ are scaler.

The second one was

$$F\big(q(x)\big) = \frac{\varepsilon q(x)}{q(x)+1}\,, \varepsilon > 0,$$

- [11], [12] Another rational models which have been investigated, the first one was:

$$F\big(q(x)\big) = \frac{\varepsilon_1 q(x)}{(1-\varepsilon_2 q(x))}, \varepsilon_1 > 0\,, \varepsilon_2 < 0\,.$$

The second one was:

$$F\big(q(x)\big) = \frac{1}{\left(e^{q^2(x)} - 1\right)}, q(x) > 0 \ and\ \frac{df}{dq} > 0.$$

In this study, we propose a novel rational model defined as:

$$F\big(q(x)\big) = \frac{1}{(1-e^{q^2(x)})}, q(x) > 0 \ \&\ \frac{df}{dq} > 0. \quad (20)$$

The unknown quantities η_{n-1} are expressed in terms of known algorithmic quantities, specifically the gradient values of the objective function, using the expression:

$$\eta_{n-1} = \frac{f'_{n-1}}{f'_n}, \tag{21}$$

wheref'_n must be evaluated as a function of known quantities from the relationship:

$$g_n = f'_n H(x_n - x^*) \Rightarrow f'_n = \frac{g_n}{H(x_n - x^*)}, \tag{22}$$

$$g_{n-1} = f'_{n-1} H(x_n - x^*) \Rightarrow f'_{n-1} = \frac{g_{n-1}}{H(x_n - x^*)}. \tag{23}$$

Here, G is the Hessian matrix and x^* denotes the minimizer of the function. Substituting (22) and (23) into (21), we obtain:

$$\begin{aligned} g^T_{n-1}(x_n - x^*) &= g^T_{n-1}(x_{n-1} + \alpha_{n-1} d_{n-1} - x^*) \\ &= g^T_{n-1}(x_n - x^*) \\ &+ \alpha_{n-1} g^T_{n-1} d_{n-1}, \end{aligned}$$

and,

$$\begin{aligned} g^T_n(x_{n-1} - x^*) &= g^T_n(x_n - \alpha_{n-1} d_{n-1} - x^*) \\ &= g^T_n(x_n - x^*). \end{aligned}$$

Since $g^T_n d_{n-1} = 0$, therefore, we can express η_{n-1}

$$\eta_{n-1} = \frac{g_{n-1}{}^T(x_n - x^*) + \alpha_{n-1} g_{n-1}{}^T d_{n-1}}{g_n{}^T(x_{n-1} - x^*)}. \tag{24}$$

From (22), (23) and (24), we get:

$$\eta_{n-1} = \frac{f'_{n-1}(x_{n-1} - x^*) H(x_n - x^*) + \alpha_{n-1} g_{n-1}{}^T d_{n-1}}{f'_n(x_{n-1} - x^*) H(x_n - x^*)}.$$

Therefore

$$\eta_{n-1} = \frac{2 f'_{n-1} q_{n-1} + \alpha_{n-1} g_{n-1}{}^T d_{n-1}}{2 f'_n q_n}. \tag{25}$$

If we calculate f'_{n-1} and f'_n from (18), we get:

$$f'_{n-1} = 2 q_{n-1} f_{n-1}(1 + f_{n-1}) \;\; \& \;\; f'_n = 2 q_n f_n (1 + f_n). \tag{26}$$

Now, to find $q(x)$ we can solving (20), its yield:

$$q(x) = \left[ln\left(\frac{f-1}{f}\right) \right]^{\frac{1}{2}}. \tag{27}$$

By substituting (26), (27) in (25), we get:

$$\eta_{n-1} = \frac{\psi_1 + \mathcal{B}}{\psi_2}. \tag{28}$$

Where:

$$\psi_1 = \left(ln \frac{(f_{n-1} - 1)}{f_{n-1}} \right) * f_{n-1} * (f_{n-1} - 1), \tag{29}$$

$$\psi_2 = \left(ln \frac{(f_n - 1)}{f_n}\right) * f_n * (f_n - 1), \tag{30}$$

and,

$$\mathcal{B} = \frac{\omega_{n-1} g^T_{n-1} d_{n-1}}{4}. \tag{31}$$

3.1 Algorithm (I)

The proposed optimization procedure is summarized in Algorithm I, which describes the iterative steps used to update the search direction and convergence process:

1) Consider an initial point $x_0, \varepsilon > 0, d_n = -g_n, set\ n = 0$.
2) If $\|g_0\| \le \varepsilon$, stop; otherwise, go to the next.
3) Determine a step size α_{n-1}as in (5) & (6).
4) Update the variable $x_n = x_{n-1} + \alpha_{n-1} d_{n-1}$, and compute f_n, g_n.
5) Compute the direction d_n as in (3), (18) & (28).
6) If $\|g_n\| \le \varepsilon$, stop; otherwise, go to the next.
7) Set $n = n + 1$ and go to 3.
8) End.

3.2 Presumptions (I)

A1-The level set $\Upsilon = \{x \in \Re^n, f(x) \le f(x_1)\}$, is bounded from below where x_0 is the starting point of conjugate gradient (2) and (3). That is, there exists some positive constants such that

$$\left.\begin{aligned} \bar{\sigma} \le \|d_{n-1}\| \le \sigma \\ \bar{\gamma_1} \le \|g_{n-1}\| \le \gamma_1 \\ \bar{\gamma_2} \le \|g_n\| \le \gamma_2 \\ \|y_{n-1}\| \le \gamma_3 \end{aligned}\right\} . \tag{32}$$

A2-The function f is smooth, and its gradient is Lipschitz continuous in a specific neighborhood $\mathbb{N}$ of Υ; specifically, there is a constant $\check{\mathcal{L}}$ greater than zero, so that:

$$\|\nabla f(x_1) - \nabla f(x_2)\| \le \check{\mathcal{L}} ||x_1 - x_2||, \forall x_1, x_2 \in \mathbb{N} \tag{33}$$

The relationship will be satisfied when the function is uniformly convex [13]:

$$d^T_{n-1} y_{n-1} \ge \mu \|d_{n-1}\|. \|y_{n-1}\|. \tag{34}$$

3.3 Theorem (1)

Let the search direction d_n be defined as in (3), (18), and (28), and suppose that the parameter ω_n satisfies conditions (5) and (6). Then the inequality

$$g^T_n d_n \le -\tau \|g_n\|^2, \tag{35}$$

holds for all $n \ge 0$.

Proof: For the initial iteration $n, n = 0$, we have

$$g^T_0 d_0 = -\|g_0\|^2.$$

Assume by induction that condition (3) holds for all values up to $n - 1$, i.e.

$$g^T_{n-1} d_{n-1} = -\|g_{n-1}\|^2. \tag{36}$$

We now show that this leads to the condition holding for n as well.

$$g_n^T d_n = -\|g_n\|^2 + \beta_{n-1}^{RG} g_n^T d_{n-1}. \quad (37)$$

For η_{n-1} as defined in equation (28), the presence of a logarithmic function requires that the argument inside the logarithm be positive. This leads to two possible cases:

Case 1: If $(f_{n-1} - 1) < 0$, then it must be that $f_{n-1} < 0$. Consequently, $\frac{f_{n-1}-1}{f_{n-1}} > 0$, which implies that $\psi_1 > 0$.

There is quantity:

$$0 < \psi_1 < \rho_1. \quad (38)$$

Case 2: If $(f_{n-1} - 1) > 0$, then $f_{n-1} > 0$. Again, we have $\frac{f_{n-1}-1}{f_{n-1}} > 0$, which implies that $\psi_2 > 0$.

i.e., the quantity

$$0 < \rho_2 < \psi_2 < \rho_3. \quad (39)$$

Substitutes (36) in (31), we have:

$$\mathcal{B} \le -\frac{\omega_{n-1}}{4}\tau\|g_{n-1}\|^2 \le \rho_4 . \quad (40)$$

Now, Substitute (38)-(40) in (28), we get:

$$\eta_{n-1} \le \frac{\rho_1+\rho_4}{\rho_2} = \rho_5. \quad (41)$$

Now for (37), we have:

$$g_n^T d_n = -\|g_n\|^2 + \frac{\omega_{n-1} g_n^T d_{n-1} - \eta_{n-1} g_n^T y_{n-1}}{d_{n-1}^T y_{n-1}} g_n^T d_{n-1}, \quad (42)$$

We have:

$$(g_n^T d_{n-1})^2 \le \|g_n\|^2 . \|d_{n-1}\|^2, \quad (43)$$

$$g_n^T d_{n-1} \le d_{n-1}^T y_{n-1}, so \ \frac{g_n^T d_{n-1}}{d_{n-1}^T y_{n-1}} \le 1, \quad (44)$$

$$g_n^T y_{n-1} = \|g_n\|^2 - g_n^T g_{n-1} = \|g_n\|^2 \left(1 - \frac{\|g_{n-1}\|}{\|g_n\|} cos\theta\right), |\, cos\theta| \le 1. \quad (45)$$

By using SWC and (36), we get:

$$d_{n-1}^T y_{n-1} \ge (1-\sigma)\,\tau\|g_{n-1}\|^2. \quad (46)$$

Substitute (41) and (43)-(46) in to (42), we obtain:

$$g_n^T d_n \le -[1 + \frac{\omega_{n-1}}{(1-\sigma)\tau}\frac{\|d_{n-1}\|^2}{\|g_{n-1}\|^2} - \rho_5\left(\frac{\|y_{n-1}\|}{\|g_n\|}\right)]\|g_n\|^2. \quad (47)$$

Now using (32) in (47), its yield:

$$g_n^T d_n \le -[1 + \frac{\omega_{n-1}}{(1-\sigma)}\frac{\sigma^2}{\Upsilon_1^2} - \rho_5\left(\frac{\Upsilon_3}{\Upsilon_2}\right)]\|g_n\|^2 = -\tau\|g_n\|^2.$$

This inequality implies (35) is satisfies for n.

4 CONVERGENCE ANALYSIS

4.1 Lemma (1)

Let f be a continuously differentiable function on $\Re^n$, and let d_n be the descent direction at the point x_n. If the function f is bounded below along the path $\{x_{n-1} + \alpha_{n-1} d_{n-1}, \alpha_{n-1} > 0\}$, for any two constants δ_1 *and* $\delta_2$such that $0 < \delta_1 < \delta_2 < 1$, there exists a range of step sizes where the SWC are satisfied [14].

4.2 Lemma (2)

Suppose that Presumptions (I) holds, and each conjugate gradient method has descending direction d_n , and the step size ω_{n-1} is achieved by SWC. if

$$\sum_{n\ge 0}\frac{1}{\|d_n\|^2} = \infty, \quad (48)$$

then the next equation is holds.

$$\lim_{n\to\infty} \inf\|g_n\| = 0. \quad (49)$$

4.3 Theorem (2)

Suppose that Presumptions (I) holds, search direction d_n is given by (3), (18) and (28) satisfied descent condition and α_{n-1} has SWC. Then (49) is holds.

Proof: If $\lim_{n\to\infty} \inf\|g_n\| \ne 0$. Then, there exists a constant $\gamma_2 > 0$ such that $\|g_n\| \ge \overline{\gamma_2}, \forall n.$

By taking the norm on both sides of (3), we obtain:

$$\|d_n\| = \|-g_n + \beta_{n-1} d_{n-1}\| \le \|g_n\| + |\beta_{n-1}|\|d_{n-1}\|,$$

$$|\beta_{n-1}^{RG}| = \left|\omega_{n-1}\frac{d_{n-1}^T g_n}{d_{n-1}^T y_{n-1}} - \eta_{n-1}\frac{g_n^T y_{n-1}}{d_{n-1}^T y_{n-1}}\right|.$$

From (44) the above equation become:

$$|\beta_{n-1}^{RG}| \le |\alpha_{n-1}| + |\eta_{n-1}| \left|\frac{g_n^T y_{n-1}}{d_{n-1}^T y_{n-1}}\right|.$$

Now by the definition of η_{n-1} in (28), we have:

$$|\eta_{n-1}| = \left|\frac{\psi_1 + \mathcal{B}}{\psi_2}\right|.$$

Taking the absolute value to both sides of (40), we get:

$$|\mathcal{B}| \le \left|-\frac{\alpha_{n-1}}{4}\tau\|g_{n-1}\|^2\right| \le \frac{\alpha_{n-1}}{4}\tau\gamma_1^2 \le \rho_5 \quad (50).$$

Using (38), (39) and (50), its yield:

$$|\eta_{n-1}| \le \frac{\rho_1 + \rho_5}{\rho_2} = \rho_{6,}$$

$$|\beta_{n-1}^{RG}| \leq |\omega_{n-1}| + \rho_6 \frac{\|g_n\|\|y_{n-1}\|}{\mu\|d_{n-1}\|\|y_{n-1}\|}$$
$$\leq w + \rho_6 \frac{\gamma_2}{\mu\bar{\sigma}} = u,$$

$$\|d_n\| = \|g_n\| + u_3\|d_{n-1}\| \leq \gamma_3 + u = U,$$

So, we get $\sum_{n\geq1} \frac{1}{\|d_n\|^2} \geq \frac{1}{U} \sum_{n\geq1} = +\infty$.

This inequality contradicts the result established in Lemma 2. Therefore, the assumption that $\lim_{n\to\infty} \inf\|g_n\| \neq 0$ must be false, and thus, the conclusion in (49) holds.

5 NUMERICALLY RESULTS

In this section, the numerical performance of various conjugate gradient (CG) methods is evaluated. A total of 46 test functions from the CUTEr library were selected to assess the efficiency of the approach method incorporating the new parameter β_(n-1)^RG. Each test function was implemented at three different levels of dimensionality: N=1000, 5000 and 10000, depending on the problem's requirements.

The performance of the approach method was compared against two classical CG methods: the Dai–Yuan (DY) method, using β_{n-1}^{DY}, and the Hestenes–Stiefel (HS) method, using β_(n-1)^HS. All methods employed a line search strategy satisfying the Sufficient Wolfe Conditions (SWC). The comparison metrics include the number of iterations (NI) and the number of function evaluations (NF).

The parameters for the line search were set asδ_1=0.001, δ_2=0.9. The stopping criterion was defined by the gradient norm condition$\|g_n\| \leq 1 * 10^{-6}$. If a method exceeded 2000 iterations for any test function, the run was considered a failure.

All algorithms were implemented in Fortran (Microsoft Developer Studio Fortran Power Station v4.0) with double precision. The results were visualized using MATLAB (R2020a), following the benchmarking approach described by Dolan and Moré [15]. Key findings include Figures 1-6.

The findings indicate that the approach technique delivers enhanced numerical efficiency when compared with classical conjugate gradient (CG) methods. Furthermore, the consistent performance observed across all tested problem dimensions (N=1000,5000,10000) underscores the robustness and computational effectiveness of the approach RG method. Its superior convergence properties position it as a strong candidate for tackling large-scale unconstrained optimization problems, surpassing traditional CG approaches such as DY and HS.

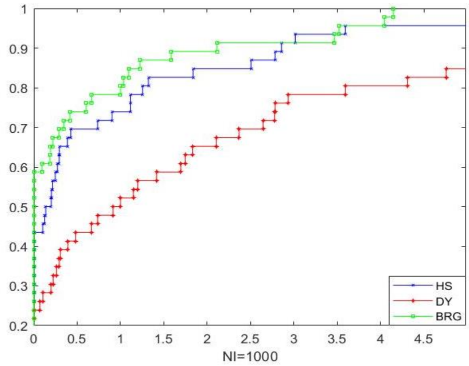

Figure 1: At the dimensional scale N=1000, the approach RG method exhibits superior performance with respect to the number of iterations (NI), surpassing both the DY and HS methods.

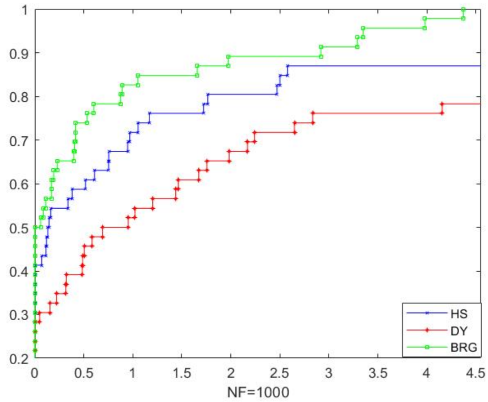

Figure 2: Likewise, in terms of the number of function evaluations (NF), the RG method attains the most favorable results at N=1000, as indicated by its placement on the upper curve in the performance profile graph.

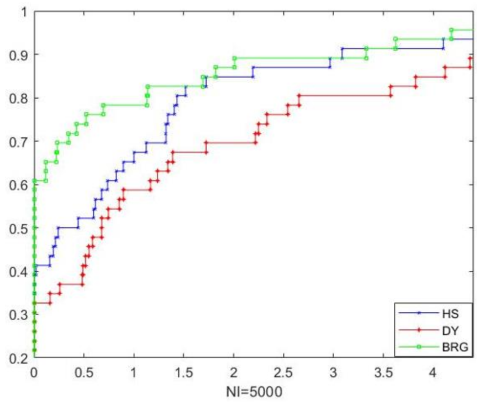

Figure 3: At the dimensional scale N=5000, the approach RG method exhibits superior performance with respo the number of iterations (NI), surpassing both the DY and HS methods.

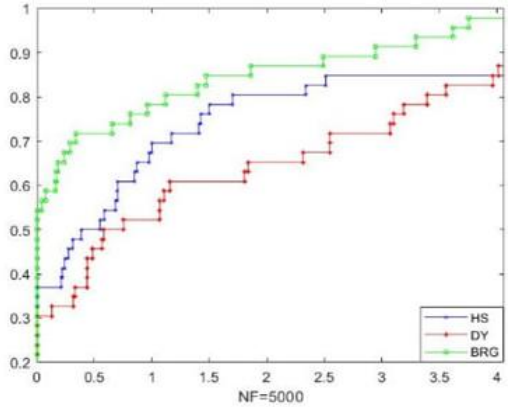

Figure 4: Likewise, in terms of the number of function evaluations (NF), the RG method attains the most favorable results at N=5000, as indicated by its placement on the upper curve in the performance profile graph.

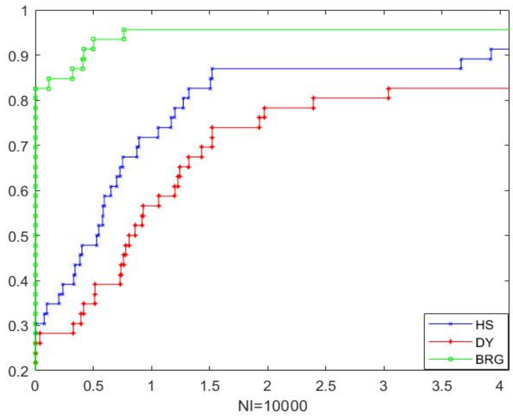

Figure 5: Likewise, in terms of the number of function evaluations (NF), the RG method attains the most favorable.

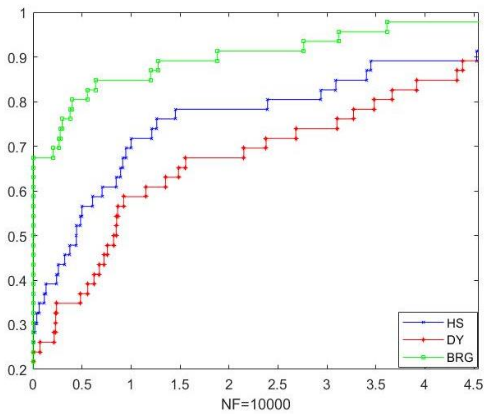

Figure 6: Additionally, in terms of the number of function evaluations (NF), the RG method maintains its advantage at N=10000, as confirmed by its dominance in the upper curve of the performance profile graph.

6 CONCLUSIONS

This paper presents a novel algorithmic framework based on a conjugate gradient method of type RG, inspired by the Dai–Yuan (DY) and Hestenes–Stiefel (HS) methods. The proposed method addresses challenges associated with large-scale unconstrained optimization problems, where traditional algorithms often suffer from reduced efficiency and scalability.

A key feature of the proposed algorithm is its ability to guarantee a sufficient descent condition independently of the line search technique used. This enhances stability and prevents inefficient search directions.

Unlike classical approaches that rely on quadratic models, this work introduces a more flexible non-quadratic model. This allows better approximation of real-world objective functions, which are often non-quadratic.

Theoretical analysis confirms the global convergence of the proposed method under standard assumptions. Numerical experiments demonstrate that the proposed algorithm provides competitive and often superior performance in terms of convergence speed, computational efficiency, and robustness compared to classical CG methods.

Overall, the results indicate that incorporating a non-quadratic model within the conjugate gradient framework is a promising direction for solving large-scale optimization problems.

7 FUTURE WORK

In light of the results and insights obtained in this study, several directions for future research are approach:

- Enhancing numerical performance through adaptive parameter tuning, preconditioning techniques, or hybrid strategies.
- Extending the algorithm to constrained optimization problems, including those with nonlinear or inequality constraints.
- Incorporating the approach methods into machine learning and artificial intelligence frameworks, particularly in large-scale training problems and model optimization tasks.

REFERENCES

[1] J. Lu, Y. Li, and H. Pham, “A Modified Dai-Liao Conjugate Gradient Method with a New Parameter for Solving Image Restoration Problems,” Mathematical Problems in Engineering, vol. 2020, 2020.

[2] B. Ivanov, G. V. Milovanović, P. S. Stanimirović, A. M. Awwal, L. A. Kazakovtsev, and V. N. Krutikov, "A Modified Dai-Liao Conjugate Gradient Method Based on a Scalar Matrix Approximation of Hessian and Its Application," Journal of Mathematics, vol. 2023, 2023.

[3] Yusuf, N. H. Manjak, H. Mohammad, A. I. Kiri, and A. B. Abubakar, "A Solution Method for Nonlinear Monotone Equations via Hybrid Spectral Conjugate Gradient and Signal Recovery Problems," Operations Research Forum, vol. 5, no. 2, pp. 0-14, 2024.

[4] F. N. Jardow and G. M. Al-Naemi, "A new parameter to enhance three-term conjugate gradient method with inexact line search," 2025.

[5] I. M. Sulaiman, P. Kaelo, R. Khalid, and M. K. M. Nawawi, "A Descent Generalized RMIL Spectral Gradient Algorithm for Optimization Problems," International Journal of Applied Mathematics and Computer Science, vol. 34, no. 2, pp. 225-233, 2024.

[6] J. Frédéric Bonnans, J. Charles Gilbert, C. Lemaréchal, and C. A. Sagastizábal, "Numerical optimization: Theoretical and practical aspects," Numerical Optimization: Theoretical and Practical Aspects, pp. 1-494, 2006.

[7] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," IMA Journal of Numerical Analysis, vol. 8, no. 1, pp. 141-148, 1988.

[8] G. M. Al-Naemi, "Modules With Chain Conditions On δ-Small Submodules," Iraqi Journal of Science, vol. 55, no. 1, pp. 202-217, 2014.

[9] W. R. Boland, E. R. Kamgnia, and J. S. Kowalik, "A conjugate-gradient optimization method invariant to nonlinear scaling," Journal of Optimization Theory and Applications, vol. 27, no. 2, pp. 221-230, Feb. 1979.

[10] A. Tassopoulos and C. Storey, "A conjugate-direction method based on a nonquadratic model," Journal of Optimization Theory and Applications, vol. 43, no. 3, pp. 371-381, Jul. 1984.

[11] G. M. Al-Naemi, "A Modified Hestenes-Stiefel Conjugate Gradient Method and its Global convergence for unconstrained optimization," Iraqi Journal of Science, vol. 55, no. 1, pp. 202-217, 2014.

[12] G. M. R. Al-Naemi, "New multi-version extended conjugate gradient methods for non-linear optimization," 1993.

[13] J. C. Gilbert, "Global convergence properties of conjugate gradient methods for optimization," SIAM Journal on Optimization, vol. 2, no. 1, pp. 21-42, Feb. 1992.

[14] S. S. Djordjević, "New hybrid conjugate gradient method as a convex combination of LS and CD methods," Filomat, vol. 31, no. 6, pp. 1813-1825, 2017.

[15] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles," Mathematical Programming, vol. 91, no. 2, pp. 201-213, 2002.