

A Novel Class of Hybrid Conjugate Gradient Methods for Unconstrained Optimization with Applications to Image Denoising



Rega Jalal¹, Hawraz N. Jabbar²

¹College of Science, Department of Mathematics, University of Sulaimani, Sulaymaniyah, 46001, Kurdistan Region of Iraq, ² College of Science, Department of Mathematics, University of Kirkuk, Kirkuk, 36013, Iraq.

ABSTRACT

This paper presents a new hybrid conjugate gradient (CG) method for solving large-scale unconstrained optimization problems where classical CG algorithms may not perform well. This new method integrates four classical algorithms, namely, Liu and Storey, Fletcher and Reeves, Dai and Yuan, and Polak, Ribiere, and Polyak, using a convex combination of CGs and an inexact line search based on strong Wolfe conditions, and adheres to the Dai-Liao conjugate condition to enhance convergence properties. The theoretical study established the conditions for sufficient descent and global convergence of the hybrid CG method. Numerical studies demonstrated that the proposed method significantly reduced the number of iterations and computational time, achieving superior evaluation efficiency compared with other CG methods. In particular, the effectiveness of the proposed algorithm is validated for image impulse denoising, where it can recover high-quality images while preserving significant features, implying its practical applicability to real-world signal and image processing problems.

Index Terms: Conjugate Gradient Method, Strong Wolfe Condition, Global Convergence, Unconstrained Optimization, Convex Combination, Hybrid Method, Image Processing

1. INTRODUCTION

Conjugate gradient (CG) methods are a category of unconstrained optimization algorithms that are distinguished by their low memory requirements and robust local and global convergence characteristics [1]. The main objective of the CG method was to solve a linear system, Hestenes and Stiefel [2]. Later, the method was modified to address an unconstrained optimization problem because solving a linear system is equivalent to minimizing a positive definite quadratic function [3].

CG methods have significantly contributed to solving large-scale non-linear optimization problems. The CG method is often used to solve the following unconstrained optimization problem:

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

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a smooth non-linear function with an available gradient.

The iterative formula of the CG method is expressed as follows,

$$x_0 \in \mathbb{R}^n$$

$$x_{k+1} = x_k + \alpha_k d_k, k = 0, 1, \dots, n \quad (2)$$

and

Access this article online

DOI:10.21928/uhdjst.v9n2y2025.pp166-183

E-ISSN: 2521-4217

P-ISSN: 2521-4209

Copyright © 2025 Jalal R, Jabbar HN. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Rega Jalal, College of Science, Department of Mathematics, University of Sulaimani, Sulaymaniyah, 46001, Kurdistan Region of Iraq. E-mail: rega.awrahman@univsul.edu.iq

Received: 03-07-2025

Accepted: 16-08-2025

Published: 27-09-2025

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geq 1. \end{cases} \quad (3)$$

In this context, x_0 is the initial point, and x_k represents the present iteration. The gradient of f at point x_k is given by $g_k = \nabla f(x_k)$. The search direction is denoted by d_k , and the conjugate (update) parameter $\beta_k \in \mathbb{R}$ determines various iterations of the CG method. In addition, the step size, which can be achieved using numerous line search methods, is represented by a positive value $\alpha_k > 0$.

In this study, we use an inexact line search called the strong Wolfe line search, which has garnered significant focus in convergence analysis and the execution of CG methods and is characterized by the following requirements [4,5].

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k \quad (4)$$

$$|g_{k+1}^T d_k| \leq -\sigma g_k^T d_k \quad (5)$$

where δ, σ are scalars and satisfy $0 < \delta \leq \sigma < 0.5$.

Assume that $\|\cdot\|$ is *Euclidean norm* and $y_k = g_{k+1} - g_k$. If f is a strongly convex quadratic, then theoretically, all six options for the update parameter in Table 1 are equal to an exact line search [12]. However, they exhibit distinct behavior for

TABLE 1: Choices of β_k in standard conjugate gradient methods.

Formula	Description
$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}$	(1952) Hestenes and Stiefel linear (CG) [2]
$\beta_k^{FR} = \frac{\ g_{k+1}\ ^2}{\ g_k\ ^2}$	(1964) Fletcher and Reeves non-linear (CG) [6]
$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\ g_k\ ^2}$	(1969) Polak and Ribiere [7], Polyak [8]
$\beta_k^{CD} = \frac{\ g_{k+1}\ ^2}{-g_k^T d_k}$	(1987) Conjugate Descent (CD) by Fletcher [9]
$\beta_k^{LS} = \frac{g_{k+1}^T y_k}{-d_k^T g_k}$	(1991) Liu and Storey [10]
$\beta_k^{DY} = \frac{\ g_{k+1}\ ^2}{y_k^T d_k}$	(1999) Dai and Yuan [11]

general non-quadratic functions with inexact line search. Various CG methods correspond to different values based on the scalar parameter β_k . The methods can be classified into two groups.

In the first group, all CG parameters have the same numerator $g_{k+1}^T y_k$. The methods have been proposed by Hestenes and Stiefel (HS), Polak, Ribiere, and Polyak (PRP), and Liu and Storey (LS) using the following CG parameters [1].

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{y_k^T d_k}, \beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}, \beta_k^{LS} = \frac{g_{k+1}^T y_k}{-d_k^T g_k}.$$

In the second group, $\|g_{k+1}\|^2$ is the common numerator of all the CG parameters. The following CG parameters were proposed by Fletcher and Reeves (FR), Fletcher (conjugate descent or CD), and Dai and Yuan (DY), and include the following CG parameters [1].

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \beta_k^{CD} = \frac{\|g_{k+1}\|^2}{-g_k^T d_k}, \beta_k^{DY} = \frac{\|g_{k+1}\|^2}{y_k^T d_k}$$

Each CG method has distinct advantages and disadvantages. The second group of FR, CD, and DY methods with $\|g_{k+1}\|^2$ in the numerator has significantly worse numerical performance than the first group of HS, PRP, and LS algorithms [13]. The poor numerical performance of the second group can be attributed to the jamming phenomenon, whereby the algorithm may execute several brief iterations without achieving much advancement toward the minimum [13]. Nonetheless, the first group HS, PRP, and LS with $g_{k+1}^T y_k$ in the numerator inherently has an automated approximation restart mechanism that prevents jamming. Specifically, when the step α_k is small, the component y_k in the numerator approaches zero. Consequently, β_k decreases, and the new search direction d_{k+1} approximates the steepest descent direction. However, Powell [14] developed a three-dimensional counterexample demonstrating that the PRP and HS methods may cycle indefinitely without a solution, indicating their lack of global convergence under specific conditions.

The FR, CD, and DY methods generally exhibit strong global convergence features [15], although their computational performance may be limited [13]. Simultaneously, the HS, PRP, and LS methods can fail to consistently converge, even though they often exhibit superior computing efficiency [13].

In Hager and Zhang [1] asserts that a comprehensive review of several CG approaches was presented.

The most important category of CG algorithms is hybrid CG algorithms. The goal of this study was to improve the computing speed and maintain robust global convergence by combining four types of CG methods. Hybrid CG methods include switching a parameter in the second group to the first group when iterations are stuck [16]. Recently, some hybrid CG methods have been proposed. Andrei proposed the following hybrid CG method: $\beta_k^{hyb} = (1 - \theta_k) \beta_k^{HS} + \theta_k \beta_k^{DY}$ [17]. Djordjević proposed following hybrid CG methods: $\beta_k^{hyb} = (1 - \theta_k) \beta_k^{PRP} + \theta_k \beta_k^{FR}$, $\beta_k^{hyb} = (1 - \theta_k) \beta_k^{LS} + \theta_k \beta_k^{FR}$ [18,19]. Babando *et al.* proposed this hybrid method: $\beta_k^{hyb} = (1 - \theta_k) \beta_k^{PRP} + \theta_k \beta_k^{DY}$ [3]. Liu and Li introduced a new hybrid approach that combines the two methods: $\beta_k^{hyb} = (1 - \theta_k) \beta_k^{LS} + \theta_k \beta_k^{DY}$ [20]. Hanachi *et al.* presented a hybrid approach that is a combination of three CG methods, as described in [21] and [22].

$$\begin{aligned}\beta_k^{New} &= \gamma_k \beta_k^{FR} + \eta_k \beta_k^{DY} + (1 - \gamma_k - \eta_k) \beta_k^{PRP}, \\ \beta_k^{New} &= \delta_k \beta_k^{LS} + \gamma_k \beta_k^{FR} + (1 - \delta_k - \gamma_k) \beta_k^{PRP}.\end{aligned}$$

Inspired by the research of Hanachi *et al.* [21] and [22], we propose a new hybrid CG method that combines the LS, FR, DY, and PRP methods to solve problem (1). This study represents the first research effort to propose and analyze a hybrid CG method that utilizes a convex combination of four classical CG methods. To the best of our knowledge, no previous study has explored such a combination for developing of CG algorithms. The remainder of this paper is organized as follows: In section 2, we introduce the hybrid CG approach and demonstrate how we obtained the parameters ζ_k, ξ_k and ω_k using various methods. Under moderate circumstances, we show that the chosen method, when used with a strong Wolfe line search, provides directions that satisfy the sufficient descent condition. The algorithm is described in section 3. In section 4, we examine the convergence properties of the proposed method. Through comprehensive numerical comparisons with several existing hybrid methods in the literature using 170 distinct test problems, demonstrating the improved efficiency of the proposed method in section 5, was also validated for an application to image impulse denoising in section 6. Section 7 concludes with a brief summary.

2. HYBRID CG ALGORITHM

In this section, to enhance the performance of the CG updating parameter proposed by Hestenes and Stiefel [2], Fletcher and Reeves [6], Dai and Yuan [11], Polak and Ribiere [7] and Polyak [8], CG methods, we employed their convex combination, in which the parameter β_k in the proposed method, referred to as β_k^{hRH} , i.e.,

$$\beta_k^{hRH} = \zeta_k \beta_k^{LS} + \xi_k \beta_k^{FR} + \omega_k \beta_k^{DY} + (1 - \zeta_k - \xi_k - \omega_k) \beta_k^{PRP} \quad (6)$$

The parameters ζ_k, ξ_k, ω_k in equation (6) that meet the condition $0 \leq \zeta_k, \xi_k, \omega_k \leq 1$ are referred to as hybridization parameters. These parameters are obtained in a specific manner, which will be discussed in detail later and the search direction d_k is computed from

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \beta_k^{hRH} d_{k-1}, & \text{if } k \geq 1. \end{cases} \quad (7)$$

To determine the step length a_k , we utilize the strong Wolfe line search (4), (5).

There are 15 separate cases depending on the parameter values where $0 \leq \zeta_k, \xi_k, \omega_k \leq 1$ and $\zeta_k + \xi_k + \omega_k \leq 1$.

Case 1: If $\zeta_k = 1, \xi_k = 0, \omega_k = 0$, then $\beta_k^{hRH} = \beta_k^{LS}$

Case 2: If $\zeta_k = 0, \xi_k = 0, \omega_k = 1$, then $\beta_k^{hRH} = \beta_k^{FR}$

Case 3: If $\zeta_k = 0, \xi_k = 0, \omega_k = 1$, then $\beta_k^{hRH} = \beta_k^{DY}$

Case 4: If $\zeta_k = 0, \xi_k = 0, \omega_k = 0$, then $\beta_k^{hRH} = \beta_k^{PRP}$

Case 5: If $0 < \zeta_k < 1, \xi_k = 0, \omega_k = 0$, then $\beta_k^{hRH} = \zeta_k \beta_k^{LS} + (1 - \zeta_k) \beta_k^{PRP}$

Case 6: If $\zeta_k = 0, 0 < \xi_k < 1, \omega_k = 0$, then $\beta_k^{hRH} = \xi_k \beta_k^{FR} + (1 - \xi_k) \beta_k^{PRP}$

Case 7: If $\zeta_k = 0, \xi_k = 0, 0 < \omega_k < 1$, then $\beta_k^{hRH} = \omega_k \beta_k^{DY} + (1 - \omega_k) \beta_k^{PRP}$

Case 8: If $0 < \zeta_k < 1, 0 < \xi_k < 1, \omega_k = 0$ and $\zeta_k + \xi_k = 1$, then $\beta_k^{hRH} = \zeta_k \beta_k^{LS} + (1 - \zeta_k) \beta_k^{FR}$

Case 9: If $0 < \zeta_k < 1$, $0 < \xi_k < 1$, $\omega_k = 0$ and $0 < \zeta_k + \xi_k < 1$, then

$$\beta_k^{hRH} = \zeta_k \beta_k^{LS} + \xi_k \beta_k^{FR} + (1 - \zeta_k - \xi_k) \beta_k^{PRP}$$

Case 10: If $0 < \zeta_k < 1$, $\xi_k = 0$, $0 < \omega_k < 1$ and $\zeta_k + \omega_k = 1$, then

$$\beta_k^{hRH} = \zeta_k \beta_k^{LS} + (1 - \zeta_k) \beta_k^{DY}$$

Case 11: If $0 < \zeta_k < 1$, $\xi_k = 0$, $0 < \omega_k < 1$ and $0 < \zeta_k + \omega_k < 1$, then

$$\beta_k^{hRH} = \zeta_k \beta_k^{LS} + \omega_k \beta_k^{DY} + (1 - \zeta_k - \omega_k) \beta_k^{PRP}$$

Case 12: If $\zeta_k = 0$, $0 < \xi_k < 1$, $0 < \omega_k < 1$ and $\xi_k + \omega_k = 1$, then

$$\beta_k^{hRH} = \xi_k \beta_k^{FR} + (1 - \xi_k) \beta_k^{DY}$$

Case 13: If $\zeta_k = 0$, $0 < \xi_k < 1$, $0 < \omega_k < 1$ and $0 < \xi_k + \omega_k < 1$, then

$$\beta_k^{hRH} = \xi_k \beta_k^{FR} + \omega_k \beta_k^{DY} + (1 - \xi_k - \omega_k) \beta_k^{PRP}$$

Case 14: If $0 < \zeta_k < 1$, $0 < \xi_k < 1$, $0 < \omega_k < 1$ and $\zeta_k + \xi_k + \omega_k = 1$, then

$$\beta_k^{hRH} = \zeta_k \beta_k^{LS} + \xi_k \beta_k^{FR} + \omega_k \beta_k^{DY}$$

Case 15: If $0 < \zeta_k < 1$, $0 < \xi_k < 1$, $0 < \omega_k < 1$ and $0 < \zeta_k + \xi_k + \omega_k < 1$,

$$\beta_k^{hRH} = \zeta_k \beta_k^{LS} + \xi_k \beta_k^{FR} + \omega_k \beta_k^{DY} + (1 - \zeta_k - \xi_k - \omega_k) \beta_k^{PRP}$$

It is clear from equations (6) and (7) that

$$d_k = \begin{cases} -g_k, & \text{if } k = 0 \\ -g_k + \zeta_k \frac{g_k^T y_{k-1}}{-d_{k-1}^T g_{k-1}} d_{k-1} \\ + \xi_k \frac{\|g_k\|^2}{\|g_{k-1}\|^2} d_{k-1} + \omega_k \frac{\|g_k\|^2}{y_{k-1}^T d_{k-1}} d_{k-1} \\ + (1 - \zeta_k - \xi_k - \omega_k) \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} d_{k-1}, & \text{if } k \geq 1 \end{cases} \quad (8)$$

For both numerical computation and convergence analysis, the traditional conjugacy condition is crucial in the CG method. This condition is based on an exact line search to obtain the step size a_k ; however, in practical computations, where an inexact line search is commonly utilized, the condition may not be strictly satisfied.

$$d_{k+1}^T y_k = 0 \quad (9)$$

In 1978, Perry [25] generalized the conjugacy condition and incorporated ideas from quasi-Newton methods. In particular, Perry supplemented the conjugacy condition by adding modifications to the standard conjugacy condition. The secant equation was incorporated into this formulation. Perry's extension enables the CG method to be brought in line with the quasi-Newton direction, is flexible, and is more robust under an exact line search.

$$d_{k+1}^T y_k = -g_{k+1}^T s_k \quad (10)$$

However, it is still contingent on the exact line search and is therefore not particularly effective in certain numerical optimization procedures. In addition, inexact line searches rather than exact line searches are used to determine the step size a_k .

Therefore, in 2001, Dai and Liao [26] suggested a less stringent conjugacy condition called the D-L conjugacy condition:

$$d_{k+1}^T y_k = -t g_{k+1}^T s_k, \quad t \geq 0 \quad (11)$$

Where $s_k = x_{k+1} - x_k$, and t is a non-negative modulating parameter. The selection of an optimal value for t remains an active area of research, with several potential choices proposed in [27-35]. Notably, for $t=0$, equation (11) simplifies to equation (9), while for $t=1$, it corresponds to equation

$$(10). \text{ To select parameter } t, \text{ this study employs } t = \frac{s_k^T y_k}{s_k^T s_k} + \frac{y_k}{s_k}$$

as specified in [30]. The parameters ζ_k , ξ_k , and ω_k are chosen by applying the D-L conjugacy condition, as given in equation (11). Taking the inner product of equation (3) with the vector y_k^T and calculating β_k using equation (6) yields

$$d_{k+1}^T y_k = -g_{k+1}^T y_k + \beta_k^{hRH} (d_k^T y_k)$$

By enforcing the D-L condition $d_{k+1}^T y_k = -t g_{k+1}^T s_k$, we obtain

$$-t g_{k+1}^T s_k = -g_{k+1}^T y_k + \left(\zeta_k \beta_k^{LS} + \xi_k \beta_k^{FR} + \omega_k \beta_k^{DY} + (1 - \zeta_k - \xi_k - \omega_k) \beta_k^{PRP} \right) (d_k^T y_k)$$

To simplify the presentation, we define the following terms:

$$A_{LS} = \beta_k^{LS} (d_k^T y_k) = \frac{g_{k+1}^T y_k}{-g_k^T d_k} (d_k^T y_k) \quad (12)$$

$$A_{FR} = \beta_k^{FR} (d_k^T y_k) = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} (d_k^T y_k) \quad (13)$$

$$A_{DY} = \beta_k^{DY} (d_k^T y_k) = \frac{\|g_{k+1}\|^2}{d_k^T y_k} (d_k^T y_k) = \|g_{k+1}\|^2 \quad (14)$$

$$A_{PRP} = \beta_k^{PRP} (d_k^T y_k) = \frac{g_{k+1}^T y_k}{\|g_k\|^2} (d_k^T y_k) \quad (15)$$

Rearranging the equation to solve for the hybridization parameters yields

$$-g_{k+1}^T s_k + g_{k+1}^T y_k - A_{PRP} = \zeta_k (A_{LS} - A_{PRP}) + \xi_k (A_{FR} - A_{PRP}) + \omega_k (A_{DY} - A_{PRP})$$

This is an underdetermined linear equation in ζ_k, ξ_k, ω_k . For our algorithm, we propose a strategy in which we compute one parameter based on the values of the other two (which can be set to defaults or previous values). This yields the following update rule:

$$\zeta_k = \frac{(-g_{k+1}^T s_k + g_{k+1}^T y_k - A_{PRP}) - \xi_k (A_{FR} - A_{PRP}) - \omega_k (A_{DY} - A_{PRP})}{A_{LS} - A_{PRP}} \quad (16)$$

$$\xi_k = \frac{(-g_{k+1}^T s_k + g_{k+1}^T y_k - A_{PRP}) - \zeta_k (A_{LS} - A_{PRP}) - \omega_k (A_{DY} - A_{PRP})}{A_{FR} - A_{PRP}} \quad (17)$$

$$\omega_k = \frac{(-g_{k+1}^T s_k + g_{k+1}^T y_k - A_{PRP}) - \zeta_k (A_{LS} - A_{PRP}) - \xi_k (A_{FR} - A_{PRP})}{A_{DY} - A_{PRP}} \quad (18)$$

Note that while these formulas appear simpler, substituting the full expressions for A_{LS} , A_{FR} , A_{DY} , and A_{PRP} would recover more detailed forms of the equations. However, this formulation makes the underlying structure of the linear system explicit. In our algorithm, care is taken to handle cases in which any denominator is close to zero.

The parameters ζ_k, ξ_k , and ω_k may lie beyond the interval $[0, 1]$. The following rule must be applied to obtain a valid convex combination in equation (6): If $\zeta_k, \xi_k, \omega_k \leq 0$, then set

$\zeta_k, \xi_k, \omega_k = 0$ in (6). If $\zeta_k, \xi_k, \omega_k \geq 1$, then we assign $\zeta_k, \xi_k, \omega_k = 1$ in (6). If $\zeta_k + \xi_k + \omega_k \geq 1$, then we change $\zeta_k, \xi_k, \omega_k = 1$ in (6).

Algorithm 1: hybrid Rega and Hawraz Algorithm

Require: Initial point $x_0 \in R^n$, parameters δ, σ with $0 < \delta \leq \sigma < 0.5$, tolerance: $\epsilon = 10^{-6}$

Initialization: Set $k = 0$, compute $f(x_0)$, $g_0 = \nabla f(x_0)$. If $\|g_0\| < \epsilon$, stop.

Set: $d_0 = -g_0$.

while $\|g_k\| \geq \epsilon$ do

Line Search: Find step size $a_k > 0$ satisfying strong Wolfe conditions (4) and (5).

Update variables:

$$x_{k+1} = x_k + a_k d_k; g_{k+1} = \nabla f(x_{k+1})$$

$$s_k = x_{k+1} - x_k; y_k = g_{k+1} - g_k$$

Compute hybridization parameters:

$$\text{Compute } t: \text{ Set } t = \frac{s_k^T y_k}{\|s_k\|^2} + \frac{\|y_k\|}{\|s_k\|}$$

Compute $A_{LS}, A_{FR}, A_{DY}, A_{PRP}$ using equations (12)-(15).

Compute initial values for ζ_k, ξ_k, ω_k using equations (16)-(18).

If any denominator is near zero, the corresponding parameter is set to zero.

Enforce convex combination constraints: Clamp each parameter to the interval $[0, 1]$:

$$\zeta_k \leftarrow \max(0, \min(1, \zeta_k))$$

$$\xi_k \leftarrow \max(0, \min(1, \xi_k))$$

$$\omega_k \leftarrow \max(0, \min(1, \omega_k))$$

if $\zeta_k + \xi_k + \omega_k > 1$ then

Normalize the sum to 1:

$$Sum \leftarrow \zeta_k + \xi_k + \omega_k; \zeta_k \leftarrow \frac{\zeta_k}{Sum}; \xi_k \leftarrow \frac{\xi_k}{Sum}; \omega_k \leftarrow \frac{\omega_k}{Sum}.$$

end if

Compute conjugate parameter β_k^{hRH}

Calculate
 $\beta_k^{hRH} = \zeta_k \beta_k^{LS} + \xi_k \beta_k^{FR} + \omega_k \beta_k^{DY} + (1 - \zeta_k - \xi_k - \omega_k) \beta_k^{PRP}$

Update search direction: Set $d_{k+1} = -g_{k+1} + \beta_k^{hRH} d_k$

Restart condition (Powell):

if $|g_{k+1}^T g_k| \geq 0.2 \|g_{k+1}\|^2$ then

$d_{k+1} \leftarrow -g_{k+1}$

end if

Iteration Update: Set $k \leftarrow k+1$

end while

3. SUFFICIENT DESCENT CONDITION

In this section, we demonstrate the sufficient descent of the proposed method, which is crucial for examining global convergence. These fundamental assumptions regarding the objective function are required to demonstrate the global convergence of the hRH method.

Assumption 1. The level set $\mathcal{L} = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}$ is bounded, i.e., \exists a constant $B > 0$ such that

$$\|x\| \leq B, \forall x \in \mathcal{L} \quad (19)$$

Assumption 2. In a neighborhood N of L , the function f is continuously differentiable, and its gradient $\nabla f(x)$ is Lipschitz continuous, that is, \exists a constant $0 < L < \infty$ such that:

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|, \forall x, y \in N \quad (20)$$

Under Assumptions (1) and (2) on f , \exists a constant c such that

$$\|\nabla f(x)\| \leq \Gamma, \forall x \in L \quad (21)$$

Theorem 3.1. If the relations (6) and (7) are valid, then

$$d_{k+1}^{hRH} = \zeta_k d_{k+1}^{LS} + \xi_k d_{k+1}^{FR} + \omega_k d_{k+1}^{DY} + (1 - \zeta_k - \xi_k - \omega_k) d_{k+1}^{PRP}.$$

Proof. Now, from (7), we have

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

Given in the equation (6), the final form is given by

$$d_{k+1}^{hRH} = -g_{k+1} + \left(\zeta_k \beta_k^{LS} + \xi_k \beta_k^{FR} + \omega_k \beta_k^{DY} + (1 - \zeta_k - \xi_k - \omega_k) \beta_k^{PRP} \right) d_k$$

can write

$$\begin{aligned} d_{k+1}^{hRH} &= -\left(1 - \zeta_k + \zeta_k - \xi_k + \xi_k - \omega_k + \omega_k\right) g_{k+1} \\ &\quad + \left(\zeta_k \beta_k^{LS} + \xi_k \beta_k^{FR} + \omega_k \beta_k^{DY} + (1 - \zeta_k - \xi_k - \omega_k) \beta_k^{PRP}\right) d_k \\ d_{k+1}^{hRH} &= -\left(\zeta_k g_{k+1} + \xi_k g_{k+1} + \omega_k g_{k+1} + (1 - \zeta_k - \xi_k - \omega_k) g_{k+1}\right) \\ &\quad + \left(\zeta_k \beta_k^{LS} + \xi_k \beta_k^{FR} + \omega_k \beta_k^{DY} + (1 - \zeta_k - \xi_k - \omega_k) \beta_k^{PRP}\right) d_k \end{aligned}$$

It follows that

$$\begin{aligned} d_{k+1}^{hRH} &= \zeta_k \left(-g_{k+1} + \beta_k^{LS} d_k\right) + \xi_k \left(-g_{k+1} + \beta_k^{FR} d_k\right) + \\ &\quad \omega_k \left(-g_{k+1} + \beta_k^{DY} d_k\right) + (1 - \zeta_k - \xi_k - \omega_k) \left(-g_{k+1} + \beta_k^{PRP} d_k\right) \end{aligned}$$

Here, we obtain

$$d_{k+1}^{hRH} = \zeta_k d_{k+1}^{LS} + \xi_k d_{k+1}^{FR} + \omega_k d_{k+1}^{DY} + (1 - \zeta_k - \xi_k - \omega_k) d_{k+1}^{PRP} \quad (22)$$

Thus, the theorem has been proven.

Theorem 3.2. Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by the hRH method. The search direction d_k satisfies the sufficient descent condition.

$$g_{k+1}^T d_{k+1} \leq -c \|g_{k+1}\|^2, \forall k \geq 0, \text{ where } c > 0 \quad (23)$$

Proof. According to the hRH method, when the Powell restart criterion is satisfied, that is, $|g_{k+1}^T g_k| \geq 0.2 \|g_{k+1}\|^2$ it follows that $d_{k+1} = -g_{k+1}$ and (23) is also valid. We proceeded with the assumption that the Powell restart criterion was invalid. Next, we obtain

$$|g_{k+1}^T g_k| \leq 0.2 \|g_{k+1}\|^2 \quad (24)$$

We demonstrate the proof by mathematical induction for $k=0$, $d_0 = -g_0$ so $g_0^T d_0 = -\|g_0\|^2$. We found that the sufficient descent condition is satisfied for $k=0$. We now assume that (23) is hold for k , thus

$$\mathbf{g}_k^T \mathbf{d}_k \leq -c \|\mathbf{g}_k\|^2 \quad (25)$$

We now show that this is true for $k+1$. From the strong Wolfe condition (5),

$$\leq \frac{\Gamma LD}{c r^2} + \frac{\Gamma^2}{r^2} + \frac{\Gamma^2}{c(1-\sigma)r^2} + \frac{LD}{r^2} = M. \quad (26)$$

Multiplying (22) by \mathbf{g}_{k+1}^T from the left yields

$$\begin{aligned} \mathbf{g}_{k+1}^T \mathbf{d}_{k+1} &= \zeta_k \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{LS}} + \xi_k \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{FR}} + \omega_k \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{DY}} \\ &+ (1 - \zeta_k - \xi_k - \omega_k) \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{PRP}} \end{aligned} \quad (27)$$

It is necessary to prove 15 cases depending on the parameter values:

Case 1: If $\zeta_k=1$, $\xi_k=0$ and $\omega_k=0$, then the relation (27) becomes

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1} = \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{LS}}.$$

The search direction for the LS method is given by

$$\mathbf{d}_{k+1}^{\text{LS}} = -\mathbf{g}_{k+1} + \beta_k^{\text{LS}} \mathbf{d}_k \quad (28)$$

Taking the inner product of (28) with \mathbf{g}_{k+1}^T

$$\begin{aligned} \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{LS}} &= \mathbf{g}_{k+1}^T (-\mathbf{g}_{k+1}) + \beta_k^{\text{LS}} \mathbf{g}_{k+1}^T \mathbf{d}_k \\ &= -\|\mathbf{g}_{k+1}\|^2 + \left(\frac{\mathbf{g}_{k+1}^T \mathbf{y}_k}{-\mathbf{d}_k^T \mathbf{g}_k} \right) \mathbf{g}_{k+1}^T \mathbf{d}_k \end{aligned}$$

Using the definition $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$, we can expand the term $\mathbf{g}_{k+1}^T \mathbf{y}_k$:

$$\begin{aligned} \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{LS}} &= -\|\mathbf{g}_{k+1}\|^2 + \left(\frac{\mathbf{g}_{k+1}^T (\mathbf{g}_{k+1} - \mathbf{g}_k)}{-\mathbf{d}_k^T \mathbf{g}_k} \right) \mathbf{g}_{k+1}^T \mathbf{d}_k \\ &= -\|\mathbf{g}_{k+1}\|^2 + \left(\frac{\|\mathbf{g}_{k+1}\|^2 - \mathbf{g}_{k+1}^T \mathbf{g}_k}{-\mathbf{d}_k^T \mathbf{g}_k} \right) \mathbf{g}_{k+1}^T \mathbf{d}_k \\ &= -\|\mathbf{g}_{k+1}\|^2 + \frac{\|\mathbf{g}_{k+1}\|^2 \mathbf{g}_{k+1}^T \mathbf{d}_k}{-\mathbf{d}_k^T \mathbf{g}_k} - \frac{(\mathbf{g}_{k+1}^T \mathbf{g}_k)(\mathbf{g}_{k+1}^T \mathbf{d}_k)}{-\mathbf{d}_k^T \mathbf{g}_k} \end{aligned}$$

By taking the absolute value and applying the triangle inequality, we obtain

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{LS}} \leq -\|\mathbf{g}_{k+1}\|^2 + \left| \frac{\mathbf{g}_{k+1}^T \mathbf{d}_k}{-\mathbf{d}_k^T \mathbf{g}_k} \right| \|\mathbf{g}_{k+1}\|^2 + \left| \frac{\mathbf{g}_{k+1}^T \mathbf{d}_k}{-\mathbf{d}_k^T \mathbf{g}_k} \right| |\mathbf{g}_{k+1}^T \mathbf{g}_k|.$$

From the second strong Wolfe condition (5) and \mathbf{d}_k is a descent direction (25), which gives

$$\left| \frac{\mathbf{g}_{k+1}^T \mathbf{d}_k}{-\mathbf{d}_k^T \mathbf{g}_k} \right| \leq \sigma \quad (29)$$

Substituting (29) into our main expression, we obtain

$$\begin{aligned} \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{LS}} &\leq -\|\mathbf{g}_{k+1}\|^2 + \sigma \|\mathbf{g}_{k+1}\|^2 + \sigma |\mathbf{g}_{k+1}^T \mathbf{g}_k| = \\ &= -(1-\sigma) \|\mathbf{g}_{k+1}\|^2 + \sigma |\mathbf{g}_{k+1}^T \mathbf{g}_k| \end{aligned}$$

Since we are assumed that the Powell restart condition is not met, we have from (24) that $|\mathbf{g}_{k+1}^T \mathbf{g}_k| \leq 0.2 \|\mathbf{g}_{k+1}\|^2$. Substituting into (30):

$$\begin{aligned} \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{LS}} &\leq -(1-\sigma) \|\mathbf{g}_{k+1}\|^2 + \sigma (0.2 \|\mathbf{g}_{k+1}\|^2) \\ &= -(1-\sigma-0.2\sigma) \|\mathbf{g}_{k+1}\|^2 \\ &= -(1-1.2\sigma) \|\mathbf{g}_{k+1}\|^2 \end{aligned}$$

Let $c_1 = (1-1.2\sigma)$. Since $\sigma < 0.5$, we have $c_1 > 0$. Therefore:

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{LS}} \leq -c_1 \|\mathbf{g}_{k+1}\|^2, \text{ where } c_1 > 0$$

Therefore, the search direction $\mathbf{d}_{k+1}^{\text{LS}}$ fulfills the sufficient descent condition.

Case 2: If $\zeta_k=0$, $\xi_k=1$ and $\omega_k=0$, then the relation (27) becomes

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1} = \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{FR}}.$$

A sufficient descent condition for the FR method under strong Wolfe conditions was established in [1]. Thus, \exists a constant $c_2 > 0$ such that

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{FR}} \leq -c_2 \|\mathbf{g}_{k+1}\|^2$$

Case 3: If $\zeta_k=0$, $\xi_k=0$ and $\omega_k=1$, then the relation (27) becomes

$$\begin{aligned}
 \mathbf{g}_{k+1}^T \mathbf{d}_{k+1} &= \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{DY}} \\
 &= \mathbf{g}_{k+1}^T \left(-\mathbf{g}_{k+1} + \beta_{k+1}^{\text{DY}} \mathbf{d}_k \right) \\
 &= -\|\mathbf{g}_{k+1}\|^2 + \beta_{k+1}^{\text{DY}} \mathbf{g}_{k+1}^T \mathbf{d}_k \\
 &= -\|\mathbf{g}_{k+1}\|^2 + \frac{\|\mathbf{g}_{k+1}\|^2}{\mathbf{y}_k^T \mathbf{d}_k} \mathbf{g}_{k+1}^T \mathbf{d}_k \\
 &\leq -\|\mathbf{g}_{k+1}\|^2 + \frac{\|\mathbf{g}_{k+1}\|^2}{\mathbf{d}_k^T \mathbf{y}_k} |\mathbf{d}_k^T \mathbf{g}_{k+1}| \\
 &\leq -\|\mathbf{g}_{k+1}\|^2 + \frac{\|\mathbf{g}_{k+1}\|^2}{-(1-\sigma)\mathbf{d}_k^T \mathbf{g}_k} (-\sigma \mathbf{g}_k^T \mathbf{d}_k) \text{ (by using (5) and (26))} \\
 &\leq -\frac{1-2\sigma}{1-\sigma} \mathbf{g}_{k+1}^2
 \end{aligned} \tag{31}$$

Let $\epsilon_3 = \frac{1-2\sigma}{1-\sigma}$. Since $\sigma < 0.5$, we have $\epsilon_3 > 0$. Therefore:

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{DY}} \leq -\epsilon_3 \|\mathbf{g}_{k+1}\|^2$$

Thus, it has been shown that $\mathbf{d}_{k+1}^{\text{DY}}$ fulfills the sufficient descent condition.

Case 4: If $\zeta_k=0$, $\xi_k=0$ and $\omega_k=0$ then the relation (27) becomes

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1} = \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{PRP}}.$$

We demonstrate the sufficient descent condition of the PRP method by the following theorem.

Theorem 3.3. [19] Under Assumptions (1) and (2), suppose that the step sizes a_k satisfy conditions (4) and (5) with $\sigma < \frac{1}{2}$. If $\|\mathbf{s}_k\| \rightarrow 0$ and there exist non-negative constants τ_1 and τ_2 such that

$$\|\mathbf{g}_k\|^2 \geq \tau_1 \|\mathbf{s}_k\|^2 \tag{32}$$

$$\|\mathbf{g}_{k+1}\|^2 \leq \tau_2 \|\mathbf{s}_k\| \tag{33}$$

where $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k = a \mathbf{d}_k$. Subsequently, $(\mathbf{d}_{k+1}^{\text{PRP}})$ satisfies the sufficient descent condition $\forall k$.

Proof.

$$\mathbf{d}_{k+1} = -\mathbf{g}_{k+1}^T + \beta_{k+1}^{\text{PRP}} \mathbf{s}_k \tag{34}$$

Now, multiplying (34) by \mathbf{g}_{k+1}^T from the left, we get

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{PRP}} = -\|\mathbf{g}_{k+1}\|^2 + \beta_{k+1}^{\text{PRP}} \mathbf{g}_{k+1}^T \mathbf{s}_k \tag{35}$$

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{PRP}} = -\|\mathbf{g}_{k+1}\|^2 + \frac{\mathbf{g}_{k+1}^T \mathbf{y}_k}{\|\mathbf{g}_k\|^2} \mathbf{g}_{k+1}^T \mathbf{s}_k \tag{36}$$

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{PRP}} \leq -\|\mathbf{g}_{k+1}\|^2 + \frac{|\mathbf{g}_{k+1}^T \mathbf{y}_k| \mathbf{g}_{k+1}^T \mathbf{s}_k}{\|\mathbf{g}_k\|^2} \tag{37}$$

By the Cauchy-Schwarz inequality, we have

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{PRP}} \leq -\|\mathbf{g}_{k+1}\|^2 + \frac{\|\mathbf{g}_{k+1}\| \|\mathbf{y}_k\| \|\mathbf{g}_{k+1}\| \|\mathbf{s}_k\|}{\|\mathbf{g}_k\|^2} \tag{38}$$

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{PRP}} \leq -\|\mathbf{g}_{k+1}\|^2 + \frac{\|\mathbf{g}_{k+1}\|^2 \|\mathbf{y}_k\| \|\mathbf{s}_k\|}{\|\mathbf{g}_k\|^2} \tag{39}$$

From (20), $\|\mathbf{y}_k\| \leq L \|\mathbf{s}_k\|$, so

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{PRP}} \leq -\|\mathbf{g}_{k+1}\|^2 + \frac{L \|\mathbf{g}_{k+1}\|^2 \|\mathbf{s}_k\|^2}{\|\mathbf{g}_k\|^2} \tag{40}$$

By using (33), we get

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{PRP}} \leq -\|\mathbf{g}_{k+1}\|^2 + \frac{L \tau_2 \|\mathbf{s}_k\|^3}{\|\mathbf{g}_k\|^2} \tag{41}$$

Using (32), it further becomes

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{PRP}} \leq -\|\mathbf{g}_{k+1}\|^2 + \frac{1}{\tau_1} L \tau_2 \|\mathbf{s}_k\| \tag{42}$$

However, because assumption $\|\mathbf{s}_k\| \rightarrow 0$, the second summand in (42) tends to zero, so there exists a number $0 < \gamma < 1$, such that

$$\frac{1}{\tau_1} L \tau_2 \|\mathbf{s}_k\| \leq \gamma \|\mathbf{g}_{k+1}\|^2 \tag{43}$$

Now, (42) becomes

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{PRP}} \leq -\|\mathbf{g}_{k+1}\|^2 + \gamma \|\mathbf{g}_{k+1}\|^2 \tag{44}$$

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{PRP}} \leq (1-\gamma) \|\mathbf{g}_{k+1}\|^2 \tag{45}$$

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{PRP}} \leq -\epsilon_4 \|\mathbf{g}_{k+1}\|^2 \tag{46}$$

$$\text{where } \epsilon_4 = 1-\gamma > 0 \tag{47}$$

Therefore, a sufficient descent condition holds for all k .

Case 5: If $\zeta_k=0$, $\omega_k=0$ and $0<\zeta_k<1$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \zeta_k g_{k+1}^T d_{k+1}^{LS} + (1 - \zeta_k) g_{k+1}^T d_{k+1}^{PRP} = g_{k+1}^T d_{k+1}^{LSPRP}$$

The sufficient descent condition is satisfied and discussed in [36], such that

$$g_{k+1}^T d_{k+1}^{LSPRP} \leq -c_5 \|g_{k+1}\|^2, \text{ where } c_5 > 0$$

Case 6: If $\zeta_k=0$, $0<\zeta_k<1$ and $\omega_k=0$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \xi_k g_{k+1}^T d_{k+1}^{FR} + (1 - \xi_k) g_{k+1}^T d_{k+1}^{PRP} = g_{k+1}^T d_{k+1}^{FRPRP}$$

Djordjević established in [18] that the sufficient descent condition is satisfied for d_{k+1}^{FRPRP}

$$g_{k+1}^T d_{k+1}^{FRPRP} \leq -c_6 \|g_{k+1}\|^2 \text{ where } c_6 > 0$$

Case 7: If $\zeta_k=0$, $\xi_k=0$ and $0<\omega_k<1$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \omega_k g_{k+1}^T d_{k+1}^{DY} + (1 - \omega_k) g_{k+1}^T d_{k+1}^{PRP} = g_{k+1}^T d_{k+1}^{DYPRP}$$

In Babando *et al.* [3], proved that the sufficient descent condition holds for d_{k+1}^{DYPRP}

$$g_{k+1}^T d_{k+1}^{DYPRP} \leq -c_7 \|g_{k+1}\|^2, \text{ where } c_7 > 0$$

Case 8: If $\zeta_k + \xi_k = 1$, $0<\zeta_k<1$, $0<\xi_k<1$ and $\omega_k=0$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \zeta_k g_{k+1}^T d_{k+1}^{LS} + (1 - \zeta_k) g_{k+1}^T d_{k+1}^{FR} = g_{k+1}^T d_{k+1}^{LSFR}$$

In Djordjević [19], Djordjević shown that the sufficient descent condition is satisfied for d_{k+1}^{LSFR} .

$$g_{k+1}^T d_{k+1}^{LSFR} \leq -c_8 \|g_{k+1}\|^2$$

Case 9: If $0<\zeta_k<1$, $0<\xi_k<1$, $\omega_k=0$ and $0<\zeta_k+\xi_k<1$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \zeta_k g_{k+1}^T d_{k+1}^{LS} + \xi_k g_{k+1}^T d_{k+1}^{FR} + (1 - \zeta_k - \xi_k) g_{k+1}^T d_{k+1}^{PRP} = g_{k+1}^T d_{k+1}^{LSFRPRP}$$

Hanachi *et al.* proved in [22] that $d_{k+1}^{LSFRPRP}$ satisfies the sufficient descent condition $\forall k$, i.e.,

$$g_{k+1}^T d_{k+1}^{LSFRPRP} \leq -c_9 \|g_{k+1}\|^2, \text{ where } c_9 > 0$$

Case 10: If $0<\zeta_k<1$, $0<\omega_k<1$, $\xi_k=0$ and $\zeta_k+\omega_k=1$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \zeta_k g_{k+1}^T d_{k+1}^{LS} + (1 - \zeta_k) g_{k+1}^T d_{k+1}^{DY} = g_{k+1}^T d_{k+1}^{LSDY}$$

Liu and Li in [20] shown that d_{k+1}^{LSDY} satisfies the sufficient descent condition $\forall k$, i.e.,

$$g_{k+1}^T d_{k+1}^{LSDY} \leq -c_{10} \|g_{k+1}\|^2, \text{ where } c_{10} > 0$$

Case 11: If $0<\zeta_k<1$, $0<\omega_k<1$, $\xi_k=0$ and $0<\zeta_k+\omega_k<1$, then there exist constants $\varphi_1, \varphi_2, \varphi_3, \varphi_4$ such that $0<\varphi_1 \leq \zeta_k \leq \varphi_3 < 1$ and $0<\varphi_2 \leq \omega_k \leq \varphi_4 < 1$. Now, the relation (27) becomes

$$\begin{aligned} g_{k+1}^T d_{k+1}^{LSDYPRP} &\leq \varphi_1 g_{k+1}^T d_{k+1}^{LS} + \varphi_2 g_{k+1}^T d_{k+1}^{DY} \\ &\quad + (1 - \varphi_3 - \varphi_4) g_{k+1}^T d_{k+1}^{PRP} \\ &\leq -\varphi_1 c_1 \|g_{k+1}\|^2 - \varphi_2 c_3 \|g_{k+1}\|^2 \\ &\quad - (1 - \varphi_3 - \varphi_4) c_4 \|g_{k+1}\|^2 \end{aligned}$$

Hence

$$g_{k+1}^T d_{k+1}^{LSDYPRP} \leq -\Omega \|g_{k+1}\|^2, \text{ with } \Omega = \begin{bmatrix} \varphi_1 c_1 + \varphi_2 c_2 \\ + (1 - \varphi_3 - \varphi_4) c_4 \end{bmatrix} > 0$$

Thus, it has been shown that $d_{k+1}^{LSDYPRP}$ fulfills the sufficient descent condition.

Case 12: If $0<\zeta_k<1$, $0<\omega_k<1$, $\xi_k=0$ and $\zeta_k+\omega_k=1$, then the relation (27) becomes

$$g_{k+1}^T d_{k+1} = \xi_k g_{k+1}^T d_{k+1}^{FR} + (1 - \xi_k) g_{k+1}^T d_{k+1}^{DY} = g_{k+1}^T d_{k+1}^{FRDY}$$

Abubakar *et al.* in [37] shown that d_{k+1}^{FRDY} satisfies the sufficient descent condition $\forall k$, i.e.,

$$g_{k+1}^T d_{k+1}^{FRDY} \leq -c_{12} \|g_{k+1}\|^2, \text{ where } c_{12} > 0$$

Case 13: If $0<\xi_k<1$, $0<\omega_k<1$, $\zeta_k=0$ and $0<\xi_k+\omega_k<1$, then the relation (27) becomes

$$\begin{aligned} g_{k+1}^T d_{k+1} &= \xi_k g_{k+1}^T d_{k+1}^{FR} + \omega_k g_{k+1}^T d_{k+1}^{DY} + (1 - \xi_k - \omega_k) g_{k+1}^T d_{k+1}^{PRP} \\ &= g_{k+1}^T d_{k+1}^{FRDYPRP} \end{aligned}$$

Hanachi *et al.* proved in [21] that $d_{k+1}^{FRDYPRP}$ satisfies the sufficient descent condition $\forall k$, i.e.,

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{FRDYPRP}} \leq -\epsilon_{13} \|\mathbf{g}_{k+1}\|^2, \text{ where } \epsilon_{13} > 0$$

Case 14: If $0 < \zeta_k < 1$, $0 < \omega_k < 1$ and $0 < \xi_k < 1$ with $\zeta_k + \xi_k + \omega_k = 1$, then there exist constants $\gamma_1, \gamma_2, \gamma_3$ such that $0 < \gamma_1 \leq \zeta_k \leq 1$, $0 < \gamma_2 \leq \xi_k \leq 1$, and $0 < \gamma_3 \leq \omega_k \leq 1$. Now, the relation (27) becomes

$$\begin{aligned} \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{LSFRDY}} &\leq \gamma_1 \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{LS}} + \gamma_2 \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{FR}} + \gamma_3 \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{DY}} \\ &\leq -\gamma_1 \epsilon_1 \|\mathbf{g}_{k+1}\|^2 - \gamma_2 \epsilon_2 \|\mathbf{g}_{k+1}\|^2 - \gamma_3 \|\mathbf{g}_{k+1}\|^2 \end{aligned}$$

Hence

$$\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{LSFRDY}} \leq -\Phi \mathbf{g}_{k+1}^2, \text{ with } \Phi = [\gamma_1 \epsilon_1 + \gamma_2 \epsilon_2 + \gamma_3 \epsilon_3] > 0$$

Thus, it has been shown that $\mathbf{d}_{k+1}^{\text{LSFRDY}}$ fulfills the sufficient descent condition.

Case 15: If $0 < \zeta_k < 1$, $0 < \omega_k < 1$ and $0 < \xi_k < 1$ with $0 < \zeta_k + \xi_k + \omega_k < 1$, then there exist constants $\varpi_1, \varpi_2, \varpi_3, \varpi_4, \varpi_5, \varpi_6$ such that $0 < \varpi_1 \leq \zeta_k \leq \varpi_4 < 1$, $0 < \varpi_2 \leq \xi_k \leq \varpi_5 < 1$, and $0 < \varpi_3 \leq \omega_k \leq \varpi_6 < 1$. Now, the relation (27) becomes

$$\begin{aligned} \mathbf{g}_{k+1}^T \mathbf{d}_{k+1} &\leq \varpi_1 \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{LS}} + \varpi_2 \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{FR}} + \varpi_3 \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{DY}} \\ &\quad + (1 - \varpi_4 - \varpi_5 - \varpi_6) \mathbf{g}_{k+1}^T \mathbf{d}_{k+1}^{\text{PRP}} \\ &\leq -\varpi_1 \epsilon_1 \|\mathbf{g}_{k+1}\|^2 - \varpi_2 \epsilon_2 \|\mathbf{g}_{k+1}\|^2 \\ &\quad - \varpi_3 \epsilon_3 \|\mathbf{g}_{k+1}\|^2 - (1 - \varpi_4 - \varpi_5 - \varpi_6) \epsilon_4 \|\mathbf{g}_{k+1}\|^2 \\ \mathbf{g}_{k+1}^T \mathbf{d}_{k+1} &\leq -\Psi \|\mathbf{g}_{k+1}\|^2 \\ &\text{with } \Psi = [\varpi_1 \epsilon_1 + \varpi_2 \epsilon_2 + \varpi_3 \epsilon_3 + (1 - \varpi_4 - \varpi_5 - \varpi_6) \epsilon_4] > 0. \end{aligned}$$

Thus, it has been shown that $\mathbf{d}_{k+1}^{\text{LSFRDYPRP}}$ fulfills the sufficient descent condition.

4. CONVERGENCE ANALYSIS

In this section, we study the global convergence properties of the proposed CG method. For further consideration, we need Assumptions (1) and (2), and the following important result, established by Zoutendijk [38] and Wolfe [4].

Theorem 4.1. Suppose that Assumptions (1) and (2) hold. Consider the iterative method (2), in which, $\forall k \geq 0$, the search direction \mathbf{d}_k is the descent direction, and the step size a_k satisfies the strong Wolfe conditions (4) and (5). Then

$$\sum_{k=0}^{\infty} \cos^2 \phi_k \|\mathbf{g}_k\|^2 < \infty \quad (48)$$

Where

$$\cos \phi_k = -\frac{\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{g}_k\| \|\mathbf{d}_k\|} \quad (49)$$

Inequality (48) is called the Zoutendijk condition.

Lemma 1. Assume that Assumptions (1) and (2) hold. Let \mathbf{d}_k be a descent direction (23) and suppose the step size a_k satisfies (5). Then, the step size a_k satisfies the following lower bound,

$$\alpha_k \geq \frac{(1-\sigma) \|\mathbf{d}_k^T \mathbf{g}_k\|}{L \|\mathbf{d}_k\|^2}, \forall k \geq 0 \quad (49)$$

Proof. From the (26), we have

$$-(1-\sigma) \mathbf{d}_k^T \mathbf{g}_k \leq \sigma \mathbf{d}_k^T \mathbf{g}_k - \mathbf{d}_k^T \mathbf{g}_k = \mathbf{d}_k^T (\mathbf{g}_{k+1} - \mathbf{g}_k)$$

By applying the Cauchy-Schwarz inequality and Lipschitz condition (20), we obtain

$$\begin{aligned} \mathbf{d}_k^T (\mathbf{g}_{k+1} - \mathbf{g}_k) &\leq \|\mathbf{d}_k\| \cdot \|\mathbf{g}_{k+1} - \mathbf{g}_k\| \leq \|\mathbf{d}_k\| \cdot L \alpha_k \|\mathbf{d}_k\| \\ &= L \alpha_k \|\mathbf{d}_k\|^2 \end{aligned}$$

Combining the above inequalities yields

$$-(1-\sigma) \mathbf{d}_k^T \mathbf{g}_k \leq L \alpha_k \|\mathbf{d}_k\|^2$$

Rearranging gives the desired result (49).

According to Lemma (1) and Assumption (1), (2), the strong Wolfe condition (4), (5), and decent direction (23), we conclude that a_k , obtained from our innovative method, is non-zero, i.e., \exists a constant $\eta > 0$ such that

$$a_k \geq \eta, \forall k \geq 0 \quad (50)$$

The global convergence of the hRH method is demonstrated by the following theorem:

Theorem 4.2. Assume that Assumptions (1) and (2) holds. The sequence $\{\mathbf{x}_k\}$ and $\{\mathbf{d}_k\}$ generated by the hRH algorithm, and the search direction \mathbf{d}_k is a descent direction, and the stepsize a_k satisfies the strong Wolfe conditions (4) and (5), then

$$\lim_{k \rightarrow \infty} \inf \mathbf{g}_k = 0 \quad (51)$$

Proof. A contradiction is used to prove this theorem. Assume that formula (51) does not hold. Therefore, \exists a constant $r > 0$ such that

$$\|g_k\| \geq r, \forall k \geq 1 \quad (52)$$

According to Theorem (3.2), it can be deduced that

$$g_k^T d_k \leq -c \|g_k\|^2, \forall c > 0$$

We can write it as

$$-g_k^T d_k \geq c \|g_k\|^2, \forall c > 0$$

Then, by (26) and (53), we get

$$d_k^T y_k = d_k^T g_{k+1} - d_k^T g_k \geq -(1-\sigma) d_k^T g_k \geq c(1-\sigma) \|g_k\|^2 \quad (54)$$

From (20), we obtain

$$\|y_k\| = \|g_{k+1} - g_k\| \leq L \|s_k\| \leq LD \quad (55)$$

where $D = \max\{\|x - y\|, x, y \in N\}$ is the diameter of N and $s_k = x_{k+1} - x_k$.

We have

$$\begin{aligned} d_{k+1} &= -g_{k+1} + \beta_K^{\text{hRH}} d_k \\ \|d_{k+1}\| &\leq \|g_{k+1}\| + |\beta_K^{\text{hRH}}| \|d_k\| \end{aligned}$$

From (6), we obtain

$$\begin{aligned} |\beta_K^{\text{hRH}}| &= \zeta_K \beta_K^{\text{LS}} + \xi_K \beta_K^{\text{FR}} + \omega_K \beta_K^{\text{DY}} + (1 - \zeta_K - \xi_K - \omega_K) \beta_K^{\text{PRP}} \\ |\beta_K^{\text{hRH}}| &= \frac{|g_{k+1}^T y_k|}{|d_k^T g_k|} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} + \frac{|g_{k+1}^T y_k|}{|d_k^T y_k|} + \frac{|g_{k+1}^T y_k|}{\|g_k\|^2} \\ &\leq \frac{\|g_{k+1}\| \|y_k\|}{c \|g_k\|^2} + \frac{\|g_{k+1}\|^2}{\|g_k\|^2} + \frac{\|g_{k+1}\|^2}{c(1-\sigma) \|g_k\|^2} + \frac{\|g_{k+1}\| \|y_k\|}{\|g_k\|^2} \end{aligned}$$

By using (52) and (54), we get

$$\leq \frac{\Gamma LD}{c r^2} + \frac{\Gamma^2}{r^2} + \frac{\Gamma^2}{c(1-\sigma) r^2} + \frac{LD}{r^2} = M.$$

The initial inequality is derived from the conditions $0 < \zeta_K, \xi_K, \omega_K < 1$ and $0 < 1 - \zeta_K - \xi_K - \omega_K < 1$. The second inequality uses the Cauchy-Schwarz inequality with (23). The final

inequality utilizes (21), (52), and (55). Therefore, (3) and (50) imply that

$$\begin{aligned} \|d_{k+1}\| &\leq \|g_{k+1}\| + |\beta_K^{\text{hRH}}| \|d_k\| \leq \|g_{k+1}\| + \frac{|\beta_K^{\text{hRH}}| \|s_k\|}{\alpha_k} \leq \Gamma \\ &+ \frac{MD}{\eta} = W, \end{aligned}$$

imply that

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} = \infty.$$

However, from (48), (52), and (23), we obtain

$$c^2 r^4 \sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} \leq \sum_{k=0}^{\infty} \frac{c^2 \|g_k\|^4}{\|d_k\|^2} \leq \sum_{k=0}^{\infty} \frac{(d_k^T g_k)^2}{\|d_k\|^2} < \infty.$$

This contradicts Theorem (4.1). Therefore, equation (52) is invalid, and the claim in (51) is proven.

5. THE NUMERICAL RESULTS

This section presents a numerical evaluation of the proposed hRH method compared to the established methods: DHF [23], HLSY [20], HLSFR [19], and HDYCDHS [24]. The evaluations used 170 benchmark test problems obtained from [39-41] with different dimensions (n) ranging from 2 to 200,000. The measures adopted included the time required to solve the problem, the number of iterations, and the number of function and gradient evaluations. The implementation was performed using MATLAB (R2024b) on a personal computer with an Intel(R) Core (TM) i5-8250U CPU @ 1.60 GHz, 8.00 GB RAM, and a 64-bit Windows 11 Education operating system. The convergence criterion was set to $\|g_k\| \leq 10^{-6}$ or when the maximum number of iterations reaches 2000. A comprehensive list of these test problems is provided in Table 2. All of these algorithms implement the strong Wolfe parameters and are fixed at $\delta = 0.0001$ and $\sigma = 0.1$. The numerical comparisons were based on the performance profile framework proposed by Dolan and Moré [42]. In the analysis of the performance profiles, the upper curve indicates that the method is better. The effectiveness of the method is determined according to the performance metric $P_s(\tau)$, which measures the performance of the solver on a set of benchmark problems. Here, $P_s(\tau)$ denotes the proportion of problems solved by the solver with a performance ratio less than or equal to τ . Using this metric, we obtained the ratio of the solver's capacity

to properly assess how this optimization problems can be solved. Thus, efficient procedures must be considered and

TABLE 2: List of test problems, dimensions, and initial points

Function name	Dimensions tested (n)	Initial point (s)
ARGLINB	2, 500, 1000	
BV	500, 1000, 10000, 20000, 30000	(0,...,0)
COSINE	10, 100, 1000, 10000, 100000	(1,...,1)
DIXMAANA	300, 30000, 60000, 90000, 120000	(2,...,2)
DIXMAANL	3000, 9000, 30000, 60000	(2,...,2)
DIXMAANK	3000, 9000, 30000, 60000, 120000	(2,...,2)
DIAGONAL 4	100, 1000, 10000, 100000, 200000	(1,...,1)
DIAGONAL 5	100, 1000, 10000, 50000, 200000	(1.1,...,1.1)
DIAGONAL 6	100, 1000, 10000, 50000, 100000	(-1.3,...,-1.3)
DIAGONAL 8	1000, 5000, 10000	(1,...,1)
DQRTIC	5000, 10000, 50000, 100000, 150000	(2,...,2)
EDENSCH	2, 100, 1000, 10000, 100000	(0,...,0)
DENSCHNF	10, 100, 1000, 10000, 100000	(0.02,...,0.02)
EDENSCHNB	2, 100, 10000, 100000	(0.3,...,0.3)
ETRIG1	2, 1000, 10000	(0.5,...,0.5)
ENGVAL1	10, 1000, 10000	(-1,...,-1)
EVF	2	(11,...,11)
EXPENALTY	100, 1000, 25000, 50000	(0.5,...,0.5)
EXTROSNB	2, 500, 1000, 10000	(-2,...,-2)
EXROSEN	10, 100, 1000, 5000, 10000, 50000, 100000	(2,...,2)
EXHIMMELBLAU	10, 1000, 10000	(-0.5,...,0.5)
GENHUMPS	2, 100, 500, 1500	(1,...,-1)
GENQUARTIC	1000, 25000, 75000	(1,...,1)
GQUARTIC	50, 100, 500, 1000	(1,...,1)
HARKERP	100, 1000, 10000, 50000	(-1,...,-1)
HIMMELBH	100, 10000, 100000	(0.2,...,0.2)
HIMMELBG	100, 1000, 10000, 100000	(1.5,...,1.5)
IE	100, 500, 1000	(0,...,0)
LIN	100, 250, 500	(1,...,1)
LIARWHD	1000, 50000, 100000	(4,...,4)
NONSCOMP	100, 5000, 25000, 50000, 100000, 200000	(-1,...,-1)
PEN1	100, 250, 500	(1. or 0.5.)
QP1	1000, 10000, 25000, 50000	(-1,...,-1)
QP2	100, 1000, 10000, 100000	(10,...,10)
QP3	100, 1000, 10000, 100000	(-1,...,-1)
QUARTC	100, 1000, 10000, 25000, 50000	(0.23,...,0.23)
QUARTICM	1000, 10000, 100000	(0.03,...,0.03)
RAYDAN1	50, 500, 50000	(1,...,1)
SINE	100, 1000, 5000, 50000, 100000	(1,...,1)
SHALLOW	25000, 50000, 100000	(10,...,10)
TRID	10, 100, 1000, 2000, 5000	(-1,...,-1)
WOODS	1000, 100000, 150000, 200000	(0.5,...,-2, 0,...,0)

simplified. Taking $P_j(\tau) = P(\tau)$, it is also taken that $S = \tau$. The numerical results were compared graphically. Figs. 1-3 show the comparison results.

In Fig. 1, we present the performance curves of the computational time. The vertical axis represents the fraction solved within a given ratio τ , while the horizontal axis represents the time ratio τ . When $0 < \tau \leq 0.5$, hRH shows better performance and solves 77% of the problems in less time than the other methods, while HLSDY and HLSFR 70% and HDYCDHS and DHF 67% and 51%, respectively. With τ rising to $\tau \geq 0.5$, hRH remains the most efficient at 95%,

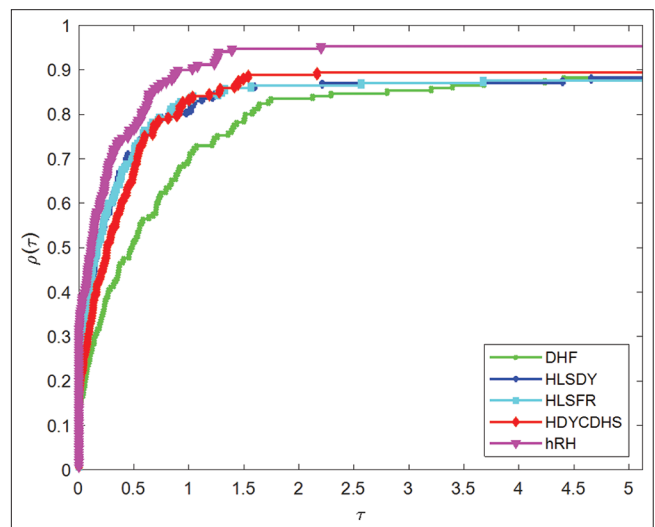


Fig. 1. Time performance for DHF, HLSDY, HLSFR, HDYCDHS, and hRH methods.

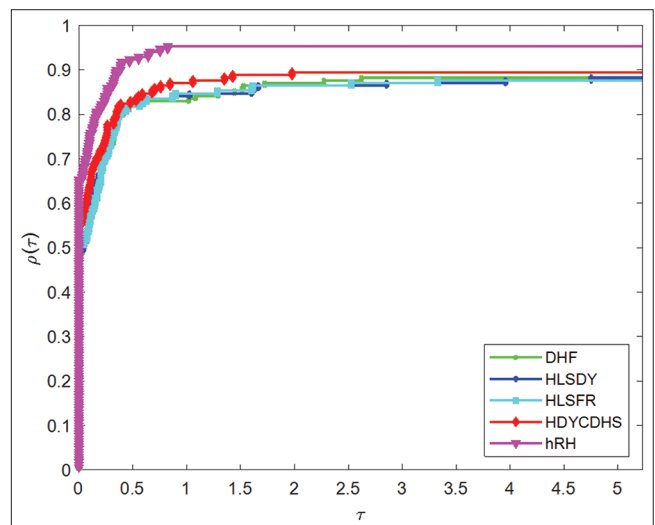


Fig. 2. Iteration count performance for DHF, HLSDY, HLSFR, HDYCDHS, and hRH methods.

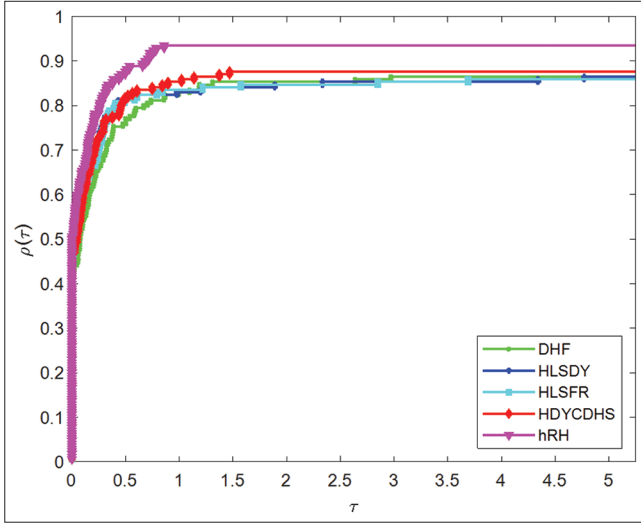


Fig. 3. Objective function/gradient calculation performance for DHF, HLSDY, HLSFR, HDYCDHS, and hRH methods.

While HLSDY takes the second place producing 89%, third place is taken by DHF and HLSDY 88%, and HLSFR in last place 87%. As expected, hRH is faster in terms of computational time than the other algorithms. Concurrently, if τ is between $0.5 < \tau \leq 1$, our method solved 90% problems, and that of the second-best method, HDYCDHS solved 82% problems.

Fig. 2 shows the relative measures, which are the iteration counts of each algorithm for convergence. In $0 < \tau \leq 0.5$, the hRH method demonstrated the highest resolution rate compared to the other methods, solving 92% of the problems with a smaller number of iterations than all the other methods, approximately 82%. Again, as $\tau \geq 0.5$, the efficiency of the hRH method is 95% while for HDYCDHS, it is 89%, and for the other methods, it is 88%. Meanwhile, if τ is between $0.5 < \tau \leq 1$, the proposed method solved 95%, the HDYCDHS method 87% and the other methods 84%. These results demonstrate that the proposed hRH algorithm achieves a better solution with fewer iterations, particularly in high-dimensional or challenging problem domains.

The performance profile of the number of function/gradient evaluations is depicted in Fig. 3. The relative performance of hRH within the range $0 < \tau \leq 0.5$ is as follows: hRH solves 88% of the problems with fewer evaluations compared to HLSDY, HLSFR, and HDYCDHS solve 81% and DHF solves 76%. Up to $\tau \geq 1$, the efficiency of hRH is 95%, that of the second-best method HDYCDHS is 89%, and that of the others is 87%. The results show that the proposed

hRH algorithm requires fewer function/gradient evaluations than other algorithms, while significantly reducing the overall computational cost.

For the EXROSEN function, the proposed (hRH) method uniquely succeeded in solving the problem across the seven tested dimensions, whereas the comparative methods (DHF, HLSDY, HLSFR, and HDYCDHS) failed in all dimensions because of jamming, that is, generating many short steps without making significant progress toward the solution. The numerical simulation also confirmed the effectiveness of the hRH method for the investigated performance parameters. It can be observed that it always performs better than the other methods in terms of computational time, number of iterations, and number of function/gradient evaluations. These results demonstrate that the proposed hRH algorithm is stable and fast for optimizing unconstrained optimization problems.

6. IMAGE RESTORATION PROBLEMS

In this section, we evaluate the performance of the proposed hRH algorithm in restoring color images corrupted by impulse noise.

Following the spirit of [43], we treat high-density noise in digital images using a dual-phase framework. In the first phase, we used an adaptive median-based method to identify the corrupted pixels. Let us assume that Z is a color image of dimensions $M \times N$ and $\Omega = \{1, \dots, M\} \times \{1, \dots, N\}$ is the pixel index domain. Therefore, we assume that the subset of noisy pixels, denoted by D , is a subset of and choose $|D|$ to be the total number of noisy pixels. Every noisy pixel at location (i, j) has four immediate neighbors in the neighborhood \mathcal{W}_{ij} around it: $\{(i-1, j), (i+1, j), (i, j-1), (i, j+1)\}$.

We observe the noisy pixel value, denoted as z_{ij} , and the recovered value, denoted as v_{ij} . The denoising process is formulated as a non-smooth optimization.

$$\min_{\mathbf{v}} \sum_{(i,j) \in \mathcal{D}} \left[|v_{i,j} - z_{i,j}| + \frac{\lambda}{2} (2 \cdot R_{i,j}^{(1)} + R_{i,j}^{(2)}) \right] \quad (56)$$

where the regularization terms are given by

$$R_{i,j}^{(1)} = \sum_{(m,n) \in \mathcal{W}_{i,j} \setminus \mathcal{D}} \psi_{\delta} (v_{i,j} - z_{m,n}), \quad R_{i,j}^{(2)} = \sum_{(m,n) \in \mathcal{W}_{i,j} \cap \mathcal{D}} \psi_{\delta} (v_{i,j} - v_{m,n}).$$

The function $\psi_\delta(t) = \sqrt{t^2 + \delta}$ acts as a smooth penalty term that preserves the edges and is characterized by a positive parameter δ . The vector $v = [v_{ij}]_{(ij) \in \mathcal{D}}$ is a lexicographically ordered list of unknowns. Directly minimizing (56) can be computationally expensive, particularly for high-resolution images with dense noise.

To resolve this issue, a smooth surrogate model was presented in [44], eliminating the non-smooth term and resulting in the following equivalent formulation:

$$\min_v G_\delta(v) := \sum_{(i,j) \in \mathcal{D}} \left(2 \sum_{(m,n) \in \mathcal{W}_{i,j} \setminus \mathcal{D}} \psi_\delta(v_{i,j} - z_{m,n}) + \sum_{(m,n) \in \mathcal{W}_{i,j} \cap \mathcal{D}} \psi_\delta(v_{i,j} - v_{m,n}) \right) \quad (57)$$

With an increasing noise ratio, the number of corrupted pixels also increases, and the size and complexity of the problem (57) increase drastically. Nevertheless, as demonstrated in [44], CGMs remain efficient and scalable for this problem, even under severe corruption.

Nevertheless, color images require more attention than grayscale images because they are multichannel. Each of the three-color images contains three channels: Red, Green, and Blue (RGB), and there may be noise present in each of these channels differently. Therefore, the restoration process is much more complex. Therefore, we adopt a channel-wise restoration strategy where each RGB component is selectively restored independently using the same optimization procedure. To preserve reality in the final output, care is

taken to ensure that the overall structure consists of what is present across channels. We show that this strategy can be used effectively to handle independently corrupted color channels without introducing artifacts in the reconstructed image.

We then applied this model to tackle salt-and-pepper noise, which is the most prevalent form of impulse corruption in color images. It was first operated using an adaptive median filter [45] to detect the noisy pixels across all color channels. The solution of (57) using our proposed hRH algorithm is compared in the reconstruction step to several classical CGMs: PRP, LS, and FR methods. We evaluated the performance on several widely used test images of size 768×512 : Girl, Lighthouse, Hats, and Macaws. The termination condition is defined as:

$$\text{Iterations} > 300 \text{ or } \frac{|G_\delta(v_k) - G_\delta(v_{k-1})|}{|G_\delta(v_k)|} \leq 10^{-4}$$

The experimental platform matches the one described in section 5. Restoration quality is quantified using the Peak Signal-to-Noise Ratio (PSNR), computed as

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{255^2}{\frac{1}{MN} \sum_{i,j} (x_{i,j}^r - x_{i,j}^o)^2} \right)$$

where $x_{i,j}^r$ and $x_{i,j}^o$ denote the restored and original pixel values, respectively, and

TABLE 3: Image denoising results for different methods (hRH, PRP, LS, and FR)

Image	Noise	Methods											
		hRH			PRP			LS			FR		
		ltr	Time	PSNR	ltr	Time	PSNR	ltr	Time	PSNR	ltr	Time	PSNR
Girl	0.30	17	28.59	83.14	77	81.96	18.67	27	83.28	77.27	82	21.67	51.10
	0.50	21	54.45	80.79	74	79.97	17.67	28	81.01	73.68	80	20.00	43.03
	0.70	22	66.74	78.48	71	77.98	18.33	64	78.42	70.89	77	22.00	71.70
	0.90	31	139.79	75.78	69	75.33	23.00	72	69.19	68.67	75	22.33	75.34
House	0.30	21	32.75	81.46	81	83.06	20.33	28	81.57	81.69	83	15.33	16.08
	0.50	20	67.59	79.57	80	81.15	20.67	72	79.04	79.27	82	17.33	61.51
	0.70	26	51.05	76.89	77	79.10	13.33	41	75.46	75.60	79	21.00	41.35
	0.90	16	71.86	72.66	72	74.53	24.67	86	73.10	74.55	76	20.67	89.77
Macaws	0.30	22	24.31	84.30	85	84.66	20.33	18	83.98	84.97	83	24.33	18.99
	0.50	19	23.10	82.08	83	81.88	18.67	25	81.85	83.28	82	23.67	24.46
	0.70	23	48.26	79.33	80	80.45	20.00	33	80.52	79.53	78	14.00	36.27
	0.90	28	51.33	76.07	77	76.84	26.33	58	74.07	75.24	77	19.67	29.64
Hats	0.30	21	15.01	84.77	85	86.34	21.33	23	84.34	84.93	86	18.67	20.42
	0.50	20	18.97	82.81	84	84.27	19.00	28	81.67	83.88	85	16.33	21.44
	0.70	22	45.77	80.73	81	82.79	16.33	32	79.19	79.90	82	20.33	33.79
	0.90	25	25.49	77.37	78	73.91	22.00	42	77.20	75.88	76	24.67	62.55

The key metrics, including the iteration count, calculation time, and PSNR for all tested methods, are summarized in Table 3. We only provide visual results for noise levels of 70% and 90% in Figs. 4 and 5 to illustrate the original, noisy, and restored images, to conserve space.

The proposed hRH method demonstrates a marked advantage over the classical CG methods (PRP, LS, and FR) for image denoising. While the LS algorithm provides competitive PSNR values, our hRH method excels in high-noise scenarios, delivering superior or comparable image quality where it

matters most. The most significant strength of hRH lies in its convergence efficiency, as it consistently requires a fraction of the iterations needed by the other methods. Although the FR method is frequently faster in terms of computation time, it suffers from poor accuracy. Conversely, PRP and LS are generally slower without consistently outperforming hRH in quality. Overall, the hRH method provides the best trade-off between accuracy and efficiency, proving to be the most robust and reliable algorithm tested, particularly in challenging denoising tasks.



Fig. 4. (a-x) Original images (first row), noisy images with 70% salt-and-pepper noise (second row), and restored results with hRH (row 3), PRP, (row 4), LS (row 5) and FR (row 6).

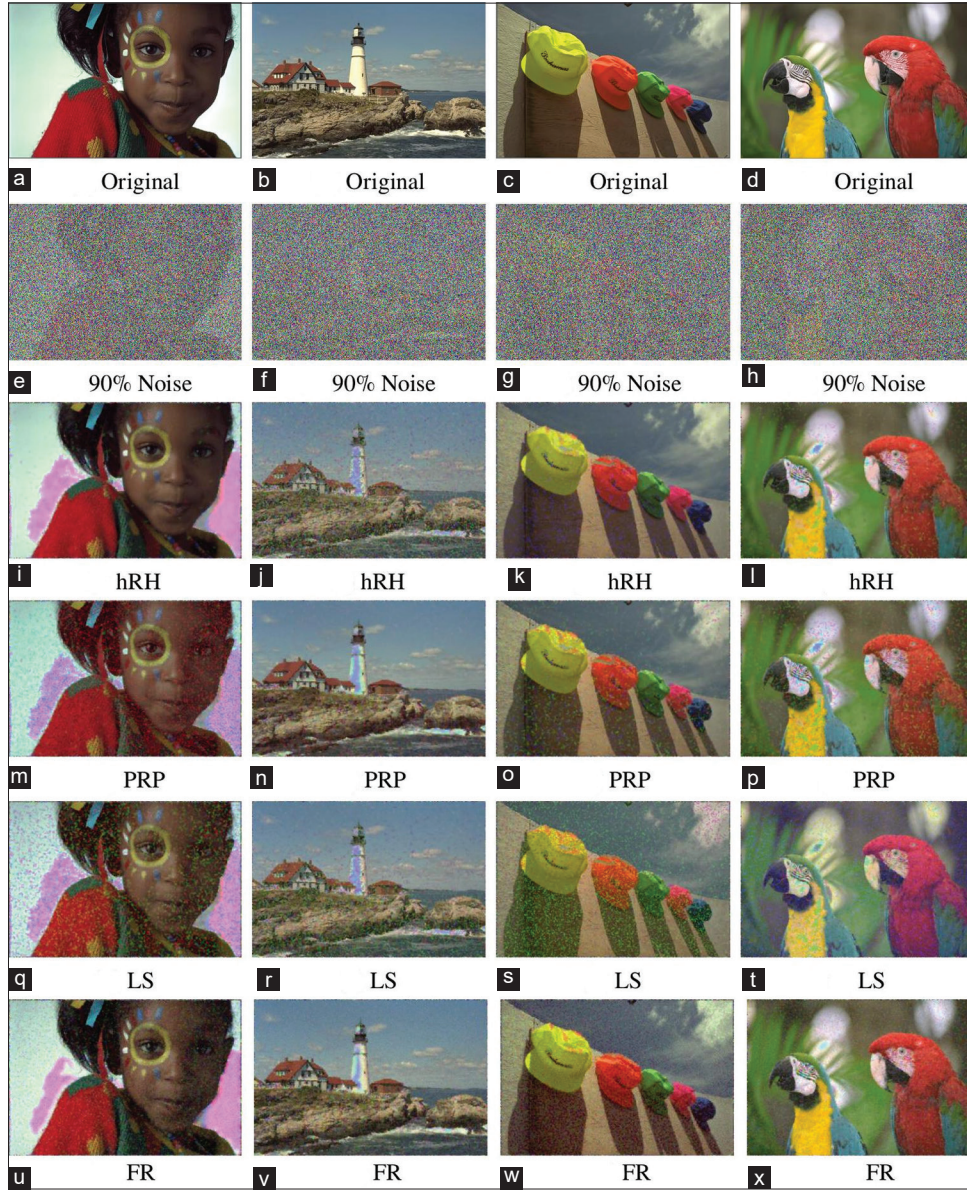


Fig. 5. (a-x) Original images (first row), noisy images with 90% salt-and-pepper noise (second row), and restored results with hRH (row 3), PRP, (row 4), LS (row 5) and FR (row 6).

7. CONCLUSION

In this study, a novel hybrid CG algorithm for unconstrained optimization is proposed and called β_k^{hRH} . The update parameter is a convex combination of the established CG formulas:

$$\beta_k^{\text{hRH}} = \zeta_k \beta_k^{\text{LS}} + \xi_k \beta_k^{\text{FR}} + \omega_k \beta_k^{\text{DY}} + (1 - \zeta_k - \xi_k - \omega_k) \beta_k^{\text{PRP}},$$

where the coefficients are selected such that they satisfy the Dai-Liao conjugacy condition. The proposed method guarantees descent directions by an inexact line search and

demonstrates global convergence under certain general conditions. Computational testing of the hRH method on a range of benchmark problems showed good performance, which was superior to that of conventional CG methods. In addition, the practical feasibility of the proposed method was demonstrated by its successful application to image impulse denoising, where the images were restored clearly despite the preservation of the intrinsic structural information. The above results thus show the versatility and applicability of the proposed method to standard optimization problems and real-world image processing problems.

ACKNOWLEDGMENT

We sincerely thank the anonymous reviewers for their thoughtful comments and insightful suggestions, which helped us improve the clarity and quality of this study. We are also grateful to the editor for their kind support and careful handling of the review process.

REFERENCES

- [1] W. W. Hager and H. Zhang. "A survey of nonlinear conjugate gradient methods". *Pacific Journal of Optimization*, vol. 2, no. 1, pp. 35-58, 2006.
- [2] M. R. Hestenes and E. Stiefel. "Methods of Conjugate Gradients for Solving Linear Systems". vol. 49. NBS, Washington, DC, 1952.
- [3] H. A. Babando, M. Barma, S. Nasiru and S. Suraj. "A dai-liao hybrid prp and dy schemes for unconstrained optimization". *International Journal of Development Mathematics*, vol. 1, no. 1, pp. 41-53, 2024.
- [4] P. Wolfe. "Convergence conditions for ascent methods". *SIAM Review*, vol. 11, no. 2, pp. 226-235, 1969.
- [5] P. Wolfe. "Convergence conditions for ascent methods. II: Some corrections". *SIAM Review*, vol. 13, no. 2, pp. 185-188, 1971.
- [6] R. Fletcher and C. M. Reeves. "Function minimization by conjugate gradients". *The Computer Journal*, vol. 7, no. 2, pp. 149-154, 1964.
- [7] E. Polak and G. Ribiere. "Note sur la convergence de méthodes de directions conjuguées. Revue française d'informatique et de recherche opérationnelle". *Série Rouge*, vol. 3, no. 16, pp. 35-43, 1969.
- [8] B. T. Polyak. "The conjugate gradient method in extremal problems". *USSR Computational Mathematics and Mathematical Physics*, vol. 9, no. 4, pp. 94-112, 1969.
- [9] R. Fletcher. "Practical Methods of Optimization". John Wiley and Sons, Hoboken, 2000.
- [10] Y. Liu and C. Storey. "Efficient Generalized conjugate gradient algorithms, part 1: Theory". *Journal of Optimization Theory and Applications*, vol. 69, pp. 129-137, 1991.
- [11] Y. H. Dai and Y. Yuan. "A nonlinear conjugate gradient method with a strong global convergence property". *SIAM Journal on Optimization*, vol. 10, no. 1, pp. 177-182, 1999.
- [12] J. Nocedal and S. J. Wright. "Numerical Optimization". Springer, Berlin, 2006.
- [13] N. Andrei. "Numerical comparison of conjugate gradient algorithms for unconstrained optimization". *Studies in Informatics and Control*, vol. 16, no. 4, pp. 333-352, 2007.
- [14] M. J. D. Powell. Nonconvex Minimization Calculations and the Conjugate Gradient Method. In: "Numerical Analysis: Proceedings of the 10th Biennial Conference Held at Dundee, Scotland, June 28-July 1, 1983", Springer, pp. 122-141, 1984.
- [15] Y. Dai, J. Han, G. Liu, D. Sun, H. Yin and Y. X. Yuan. "Convergence properties of nonlinear conjugate gradient methods". *SIAM Journal on Optimization*, vol. 10, no. 2, pp. 345-358, 2000.
- [16] S. Babaie-Kafaki. "A hybrid conjugate gradient method based on a quadratic relaxation of the dai-yuan hybrid conjugate gradient parameter". *Optimization*, vol. 62, no. 7, pp. 929-941, 2013.
- [17] N. Andrei. "Another nonlinear conjugate gradient algorithm for unconstrained optimization". *Optimization Methods and Software*, vol. 24, no. 1, pp. 89-104, 2009.
- [18] S. S. Djordjević. "New hybrid conjugate gradient method as a convex combination of FR and PRP methods". *Filomat*, vol. 30, no. 11, pp. 3083-3100, 2016.
- [19] S. S. Djordjević. "New hybrid conjugate gradient method as a convex combination of LS and FR methods". *Acta Mathematica Scientia*, vol. 39, pp. 214-228, 2019.
- [20] J. K. Liu and S. J. Li. "New hybrid conjugate gradient method for unconstrained optimization". *Applied Mathematics and Computation*, vol. 245, pp. 36-43, 2014.
- [21] S. B. Hanachi, B. Sellami and M. Belloufi. "New iterative conjugate gradient method for nonlinear unconstrained optimization". *RAIRO-Operations Research*, vol. 56, no. 4, pp. 2315-2327, 2022.
- [22] S. B. Hanachi, B. Sellami and M. Belloufi. "A new family of hybrid conjugate gradient method for unconstrained optimization and its application to regression analysis". *RAIRO-Operations Research*, vol. 58, no. 1, pp. 613-627, 2024.
- [23] N. Salihu, M. R. Odekunle, A. M. Saleh, S. Salihu. "A dai-liao hybrid hestenes-stiefel and fletcher-reeves methods for unconstrained optimization". *International Journal of Industrial Optimization*, vol. 2, no. 1, pp. 33-50, 2021.
- [24] A. Hallal, M. Belloufi and B. Sellami. "An efficient new hybrid cg-method as convex combination of DY and CD and HS algorithms". *RAIRO-Operations Research*, vol. 56, no. 6, pp. 4047-4056, 2022.
- [25] A. Perry. "A modified conjugate gradient algorithm". *Operations Research*, vol. 26, no. 6, pp. 1073-1078, 1978.
- [26] Y. H. Dai and L. Z. Liao. "New conjugacy conditions and related nonlinear conjugate gradient methods". *Applied Mathematics and Optimization*, vol. 43, pp. 87-101, 2001.
- [27] A. B. Abubakar and P. Kumam. "A descent dai-liao conjugate gradient method for nonlinear equations". *Numerical Algorithms*, vol. 81, pp. 197-210, 2019.
- [28] N. Andrei. "A dai-liao conjugate gradient algorithm with clustering of eigenvalues". *Numerical Algorithms*, vol. 77, no. 4, pp. 1273-1282, 2018.
- [29] S. Babaie-Kafaki. "On optimality of two adaptive choices for the parameter of dai-liao method". *Optimization Letters*, vol. 10, pp. 1789-1797, 2016.
- [30] S. Babaie-Kafaki and R. Ghanbari. "The dai-liao nonlinear conjugate gradient method with optimal parameter choices". *European Journal of Operational Research*, vol. 234, no. 3, pp. 625-630, 2014.
- [31] S. Babaie-Kafaki and R. Ghanbari. "A descent family of dai-liao conjugate gradient methods". *Optimization Methods and Software*, vol. 29, no. 3, pp. 583-591, 2014.
- [32] S. Babaie-Kafaki and R. Ghanbari. "Two optimal dai-liao conjugate gradient methods". *Optimization*, vol. 64, no. 11, pp. 2277-2287, 2015.
- [33] S. Babaie-Kafaki and R. Ghanbari. "Two adaptive dai-liao nonlinear conjugate gradient methods". *Iranian Journal of Science and Technology, Transactions A: Science*, 42, pp. 1505-1509, 2018.
- [34] B. Ivanov, G. V. Milovanović and P. S. Stanimirović. "Accelerated dai-liao projection method for solving systems of monotone nonlinear equations with application to image deblurring". *Journal of Global Optimization*, vol. 85, no. 2, pp. 377-420, 2023.
- [35] M. Y. Waziri, K. Ahmed and J. Sabi'u. "A dai-liao conjugate gradient method via modified secant equation for system of nonlinear equations". *Arabian Journal of Mathematics*, vol. 9, pp. 443-457, 2020.

- [36] A. B. Abubakar, P. Kumam, M. Malik and A. H. Ibrahim. "A hybrid conjugate gradient based approach for solving unconstrained optimization and motion control problems". *Mathematics and Computers in Simulation*, vol. 201, pp. 640-657, 2022.
- [37] A. B. Abubakar, P. Kumam, M. Malik, P. Chaipunya and A. H. Ibrahim. "A hybrid FR-DY conjugate gradient algorithm for unconstrained optimization with application in portfolio selection". *AIMS Mathematics*, vol. 6, no. 6, pp. 6506-6527, 2021.
- [38] G. Zoutendijk. "Nonlinear programming, computational methods". *Integer and Nonlinear Programming*, pp. 37-86, 1970.
- [39] N. Andrei. "An unconstrained optimization test functions collection". *Advanced Modeling and Optimization*, vol. 10, no. 1, pp. 147-161, 2008.
- [40] N. I. M, Gould, D. Orban and P. L. Toint. "Cuter and sifdec: A constrained and unconstrained testing environment, revisited". *ACM Transactions on Mathematical Software*, vol. 29, no. 4, pp. 373-394, 2003.
- [41] J. J. Moré, B. S. Garbow and K. E. Hillstom. "Testing unconstrained optimization software". *ACM Transactions on Mathematical Software*, vol. 7, no. 1, pp. 17-41, 1981.
- [42] E. D. Dolan and J. J. Moré. "Benchmarking optimization software with performance profiles". *Mathematical Programming*, vol. 91, pp. 201-213, 2002.
- [43] R. H. Chan, C. W. Ho and M. Nikolova. "Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization". *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1479-1485, 2005.
- [44] J. F. Cai, R. Chan and B. Morini. Minimization of an Edge-Preserving Regularization Functional by Conjugate Gradient Type Methods. In: *Image Processing Based on Partial Differential Equations: Proceedings of the International Conference on PDEBased Image Processing and Related Inverse Problems*, CMA, Oslo, Springer, pp. 109-122, 2007.
- [45] H. Hwang and R. A. Haddad. "Adaptive median filters: New algorithms and results". *IEEE Transactions on Image Processing*, vol. 4, no. 4, pp. 499-502, 1995.