Journal of Mathematical Modeling
Vol. -, No. -, 2026, pp. -. Research Article

JMM

An improved nonlinear conjugate gradient method and its application to satellite image restoration

Atefe Bay[†*], **Zohreh Akbari**

[†]*Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Mazandaran, Babolsar, Iran*

Email(s): a.bay01@umail.umz.ac.ir, z.akbari@umz.ac.ir

Abstract. In this paper, an improved nonlinear conjugate gradient method is proposed for solving unconstrained optimization problems. Due to the high computational cost of Newton-type methods, conjugate gradient methods have emerged as efficient alternatives for large-scale problems. However, their performance heavily depends on the choice of search directions and algorithmic parameters. In the proposed method, a novel parameter and a modified search direction are introduced to ensure sufficient descent at each iteration. Global convergence of the method is established under standard assumptions. Numerical experiments on satellite image restoration demonstrate the superiority of the proposed method over the Polak–Ribiere–Polyak method in terms of noise reduction and image quality enhancement.

Keywords: Nonlinear conjugate gradient method, unconstrained optimization, global convergence, image restoration.

AMS Subject Classification 2010: 34A34, 65L05.

1 Introduction

Consider the following unconstrained optimization problem

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

where the objective function $f:\mathbb{R}^n\to\mathbb{R}$ is continuously differentiable. In recent years, nonlinear optimization has become a central topic in computational mathematics, particularly due to its essential role in solving high-dimensional problems arising in diverse domains such as medicine, engineering, economics, and astronomy [5, 13, 14].

*Corresponding author
Received: 10 May 2025/ Revised: 05 February 2026/ Accepted: 15 February 2026
DOI: 10.22124/jmm.2026.30611.2744

A key example is image restoration, where the objective is to reconstruct high-quality images from degraded or noisy observations. This problem is inherently ill-posed and highly sensitive to noise and initial conditions, requiring optimization algorithms that are both numerically stable and computationally efficient.

One of the most widely used techniques for large-scale unconstrained optimization is the nonlinear conjugate gradient (CG) method. Its appeal lies in low memory requirements and fast convergence, making it suitable for high-dimensional applications such as image and signal processing. In general, the iterations of a CG method are given by

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

where α_k is the step size and d_k is the search direction. The search direction is defined recursively as

$$d_k = -g_k + \beta_k d_{k-1}, \quad (d_0 = -g_0),$$

where $g_k = \nabla f(x_k)$ is the gradient of the objective function and β_k is known as the conjugate gradient parameter [4, 11].

Different choices for computing β_k lead to different CG algorithms. Some well-known formulas for β_k include the Fletcher-Reeves (FR) [9], Hestenes-Stiefel (HS) [10], Polak–Ribière–Polyak (PRP) [12], and Dai–Yuan (DY) [8] methods, which are defined as follows:

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

where $y_k = g_{k+1} - g_k$, $\|\cdot\|$ denotes the Euclidean norm, and $(\cdot)^T$ stands for the matrix (or vector) transpose.

Recent research has extensively studied these methods and proposed various improvements to the standard CG framework. Arazm et al. [2] introduced a Dai–Liao-type extension for nonconvex functions with global convergence guarantees. Chan-Renous and Royer [7] developed a CG method with complexity bounds for regression problems. Zhang and Yang [16] proposed a hybrid method inspired by quasi-Newton techniques, while Yuan et al. [15] and Ali and Mahdi [1] focused on enhanced CG strategies for image restoration and large-scale problems. These studies reflect an ongoing effort to design CG algorithms that are both theoretically sound and practically applicable.

Proposed methods in the recent literature have introduced diverse modifications to the CG framework; however, there remains room for developing variants that specifically target instability and convergence issues, especially in ill-conditioned problems. The PRP method has attracted significant attention due to its favorable performance in practice. However, it may suffer from slow convergence or oscillatory behavior in complex or noisy settings, such as satellite image restoration.

Motivated by these developments and the limitations of classical PRP, this paper proposes a new nonlinear CG method that introduces a modified β_k and combines it with a stable search direction inspired by the work of Cao et al. [6]. The new algorithm guarantees global convergence and improves stability.

Cao's search direction is defined as

$$d_k = -g_k + \frac{\beta_k}{\gamma_k} d_{k-1} - \frac{\beta_k}{\gamma_k} \frac{d_{k-1}^T g_k}{\|g_k\|^2} g_k, \quad (d_0 = -g_0), \tag{2}$$

where $\gamma_k = \frac{|\beta_k|\|d_{k-1}\|}{\|g_k\|}$. This modified direction (2) is designed to ensure sufficient descent while enhancing convergence speed.

The organization of the paper is as follows. In Section 2, we present the proposed algorithm, its properties, and a proof of global convergence. Section 3 evaluates the algorithm's performance in solving an image restoration problem. Finally, Section 4 presents the conclusions.

2 Improved CG algorithm and its properties

In this section, inspired by the favorable features of the PRP method [12], we introduce a new CG parameter β_k. Furthermore, the sufficient descent condition and global convergence of the proposed method are discussed.

The new formula for β_k^{New} is defined as

$$\beta_k^{New} = \frac{g_{k+1}^T y_k}{\max\{\mu\|d_k\|\|y_k\|, \|g_k\|^2\}}, \qquad \mu = \tau\frac{\|g_{k+1}\|}{\|g_k\|}, \tag{3}$$

where $\tau > 0$ is a positive constant.

In equation (3), the value of μ is selected to improve numerical stability and better adapt the algorithm to challenging problems. The main idea is to update μ in proportion to the variation in the gradient. When the gradient changes sharply, μ increases, thereby stabilizing β_k and preventing instability. Moreover, the parameter τ allows adjusting the influence of μ in the definition of β_k^{New}, making the algorithm more adaptable to different optimization conditions. This approach enhances the performance of the algorithm in solving complex problems.

The proposed nonlinear CG algorithm under the Armijo line search rule is presented below.

Algorithm 1: Proposed nonlinear CG algorithm

1: Choose an initial point $x_0 \in \mathbb{R}^n$ and positive constants $\varepsilon \in (0,1)$, $\delta_0 > 0$, $\delta \in (0,1)$, and $\rho \in (0,1)$. Set $k = 0$ and $d_0 = -g_0$.

2: If $\|g_k\| \leq \varepsilon$, then stop. Otherwise, compute a step size $\alpha_k = \delta_0\rho^{i_k}$ such that the following condition holds

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta\alpha_k g_k^T d_k, \tag{4}$$

where $i_k = \min\{0,1,2,\ldots\}$ is the smallest nonnegative integer satisfying (4).

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

4: If $\|g_{k+1}\| \leq \varepsilon$, then stop. Otherwise, update the search direction d_{k+1} using equation (2) with β_k^{New} from (3).

5: Set $k := k+1$ and go to Step 2.

The following lemma establishes that the search direction d_k satisfies the sufficient descent condition.

Lemma 1. *Consider the search direction d_k defined by equation (2), with β_k given by equation (3). Then, the following properties hold*

$$d_k^T g_k = -\|g_k\|^2, \tag{5}$$

$$\|d_k\| \leq c\|g_k\|, \tag{6}$$

for some constant $c > 0$.

Proof. From (2), for $k = 0$, we get $d_0^T g_0 = -\|g_0\|^2$. For $k \geq 1$, we obtain

$$d_k = -g_k + \frac{\beta_k}{\gamma_k}\left(d_{k-1} - \frac{d_{k-1}^T g_k}{\|g_k\|^2} g_k\right),$$

$$d_k^T g_k = (-g_k)^T g_k + \frac{\beta_k}{\gamma_k}\left(d_{k-1}^T g_k - \frac{d_{k-1}^T g_k}{\|g_k\|^2} g_k^T g_k\right),$$

$$d_k^T g_k = -\|g_k\|^2.$$

Now we prove that (6) holds, For $k = 0$, we directly have

$$\|d_0\| = \|g_0\|,$$

which satisfies (6) with $c = 1$.

For $k \geq 1$, using (2) and the triangle inequality, we get

$$\|d_k\| \leq \|g_k\| + \frac{1}{\gamma_k}\left(|\beta_k|\|d_{k-1}\| + |\beta_k|\left|\frac{d_{k-1}^T g_k}{\|g_k\|^2}\right|\|g_k\|\right).$$

Since in CG methods the vectors d_{k-1} and g_k are nearly orthogonal, we have

$$|d_{k-1}^T g_k| \ll \|d_{k-1}\|\|g_k\|,$$

and hence

$$\|d_k\| \leq \|g_k\| + \frac{2|\beta_k|\|d_{k-1}\|}{\gamma_k}.$$

Substituting $\gamma_k = \frac{|\beta_k|\|d_{k-1}\|}{\|g_k\|}$, we obtain

$$\|d_k\| \leq \|g_k\| + 2\|g_k\| = 3\|g_k\|.$$

Thus, (6) holds with $c = 3$. □

2.1 Global convergence of the algorithm

To establish the global convergence of Algorithm 1, we make the following standard assumptions:

(i) The level set $S_0 = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$ is bounded.

(ii) The objective function $f \in C^2$ is bounded from below. Moreover, its gradient is Lipschitz continuous; that is, there exists a constant $L > 0$ such that

$$\|\nabla f(x_k) - \nabla f(x_y)\| \leq L\|x - y\| \quad \forall x, y \in \mathbb{R}^n.$$

Although the search directions from (2) are always descent directions, the convergence of the algorithm also depends on the step length α_k. The next lemma establishes that the Armijo line search provides a guaranteed lower bound for α_k.

Lemma 2. *Suppose that Assumptions (i) and (ii) hold, and let the sequences* $\{x_k\}$ *and* $\{\alpha_k\}$ *be generated by Algorithm 1, Then*

$$\alpha_k \geq \frac{2(1-\delta)\lambda}{L\|d_k\|^2}, \quad \forall k.$$

Proof. From Assumption (ii) and Taylor's theorem, it follows that for any $\alpha > 0$

$$f(x_k + \alpha d_k) \leq f(x_k) + \alpha g_k^T d_k + \frac{L}{2}\alpha^2 \|d_k\|^2.$$

Combining this with the Armijo condition (4) yields

$$f(x_k) + \alpha g_k^T d_k + \frac{L}{2}\alpha^2 \|d_k\|^2 \leq f(x_k) + \delta \alpha g_k^T d_k.$$

Rearranging terms and defining $\lambda := |g_k^T d_k| = -g_k^T d_k$ gives

$$(1-\delta)\lambda \geq \frac{L}{2}\alpha \|d_k\|^2,$$

hence a sufficient α satisfying the Armijo condition is

$$\alpha \leq \frac{2(1-\delta)\lambda}{L\|d_k\|^2}.$$

In the backtracking procedure of the Armijo line search, if α becomes smaller than $\bar{\alpha} = \frac{2(1-\delta)\lambda}{L\|d_k\|^2}$, the previous step must have had $\alpha > \bar{\alpha}$. Therefore, the selected α_k satisfying the Armijo condition is bounded below

$$\alpha_k \geq \bar{\alpha}.$$

□

Based on the above assumptions and Lemma 2, we now prove the global convergence of the proposed algorithm.

Theorem 1. *Suppose that Assumptions 1 and 2 hold, and the sequence* $\{x_k\}$ *is generated by Algorithm 1. Then, we have*

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

Proof. At each iteration k, the Armijo condition (4) ensures

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k.$$

Substituting (5) and (6) into the above inequality yields

$$f(x_{k+1}) \leq f(x_k) - \delta \alpha_k c \|g_k\|^2.$$

By Lemma 2, since $\alpha_k \geq \bar{\alpha} > 0$, we obtain

$$f(x_{k+1}) \leq f(x_k) - \delta c \bar{\alpha} \|g_k\|^2.$$

Defining $\eta := \delta c \bar{\alpha} > 0$, the inequality can be rewritten as

$$f(x_k) - f(x_{k+1}) \geq \eta \|g_k\|^2.$$

Summing both sides over $k = 0, 1, \ldots, N-1$ yields

$$\sum_{k=0}^{N-1} [f(x_k) - f(x_{k+1})] \geq \eta \sum_{k=0}^{N-1} \|g_k\|^2.$$

Since the left-hand side telescopes, this simplifies to

$$f(x_0) - f(x_N) \geq \eta \sum_{k=0}^{N-1} \|g_k\|^2.$$

Taking the limit as $N \to \infty$ and using that $f(x_N) \geq f^*$ (since f is bounded below), we arrive at

$$f(x_0) - f^* \geq \eta \sum_{k=0}^{\infty} \|g_k\|^2.$$

Hence

$$\sum_{k=0}^{\infty} \|g_k\|^2 \leq \frac{f(x_0) - f^*}{\eta} < \infty.$$

Since the sum of squared norms is finite, it follows that

$$\lim_{k \to \infty} \|g_k\|^2 = 0,$$

and therefore

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

□

3 Numerical results

In this section, we present a comprehensive numerical evaluation of the proposed nonlinear conjugate gradient method for satellite image restoration. The method is evaluated in terms of restoration quality, convergence speed, and computational efficiency, and compared with two other algorithms: the classical PRP algorithm [12] and an improved PRP variant introduced by Barrouk et al. [3], which incorporates a novel β-update strategy based on a tunable parameter. The comparison is carried out using three distinct satellite images of size 512×512. All simulations were conducted using MATLAB R2022b on a personal computer equipped with an 11th Gen Intel® Core™ i7-1165G7 @ 2.80GHz processor, 16.0 GB of RAM, running Windows 11 Pro (Version 24H2, 64-bit), and integrated Intel® Iris® Xe Graphics (128 MB).

The image restoration problem is modeled as

$$\min_{x \in \mathbb{R}^{m \times n}} \frac{1}{2} \|b - Ax\|^2 + \lambda R(x),$$

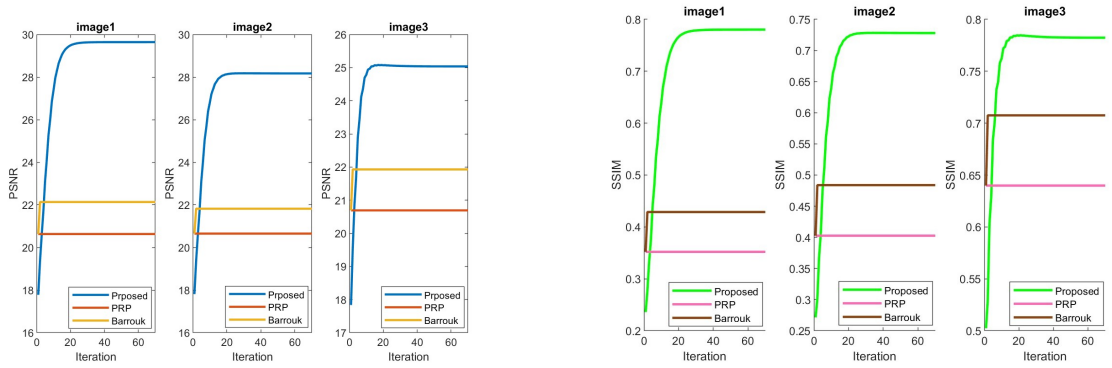

Figure 1: PSNR and SSIM variation per iteration

Table 1: Comparison of final SSIM and PSNR values for each method

	Proposed		**PRP**		**Barrouk**	
Images	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
Image1	0.79	30.00	0.35	20.43	0.42	22.14
Image2	0.72	28.10	0.40	20.66	0.48	21.82
Image3	0.76	25.59	0.63	20.66	0.70	21.91

Table 2: Comparison of final CPU time, objective function, gradient norm, and number of iterations

	Proposed				**PRP**				**Barrouk**			
Images	n	t	f	$\|g\|$	n	t	f	$\|g\|$	n	t	f	$\|g\|$
Image1	70	20.66	974.86	0.45	70	29.40	3474.9	20.604	70	24.84	2896.3	0.00
Image2	70	20.20	1033.2	0.71	70	29.40	6189.3	25.841	70	24.78	5419.9	0.00
Image3	70	20.90	1527.9	6.94	70	33.88	5394.4	27.203	70	24.97	4680.3	1.00

where λ is the regularization parameter and $R(x)$ is the regularization term. The term $\|b-Ax\|^2$ measures the discrepancy between the restored image and the noisy data, while $R(x)$ reduces sensitivity to noise and prevents excessive oscillations in the solution. In this paper, to ensure the smoothness of the problem, the regularizer is chosen as $R(x) = \|\nabla x\|_2^2$. The parameter λ controls the influence of the regularization on the final result.

In this experiment, each satellite image was corrupted with Gaussian noise (signal-to-noise ratio of 30 dB) and salt-and-pepper noise. The algorithms were executed until a stopping criterion was satisfied or the maximum number of 70 iterations was reached. The parameters used in the experiments were initialized as follows

$$\varepsilon = 1 \times 10^{-6}, \quad \lambda = 0.05, \quad \delta = 0.05, \quad \rho = 0.8, \quad \delta_0 = 1.$$

To evaluate the restoration quality, we use two widely adopted image quality metrics: Peak Signal-

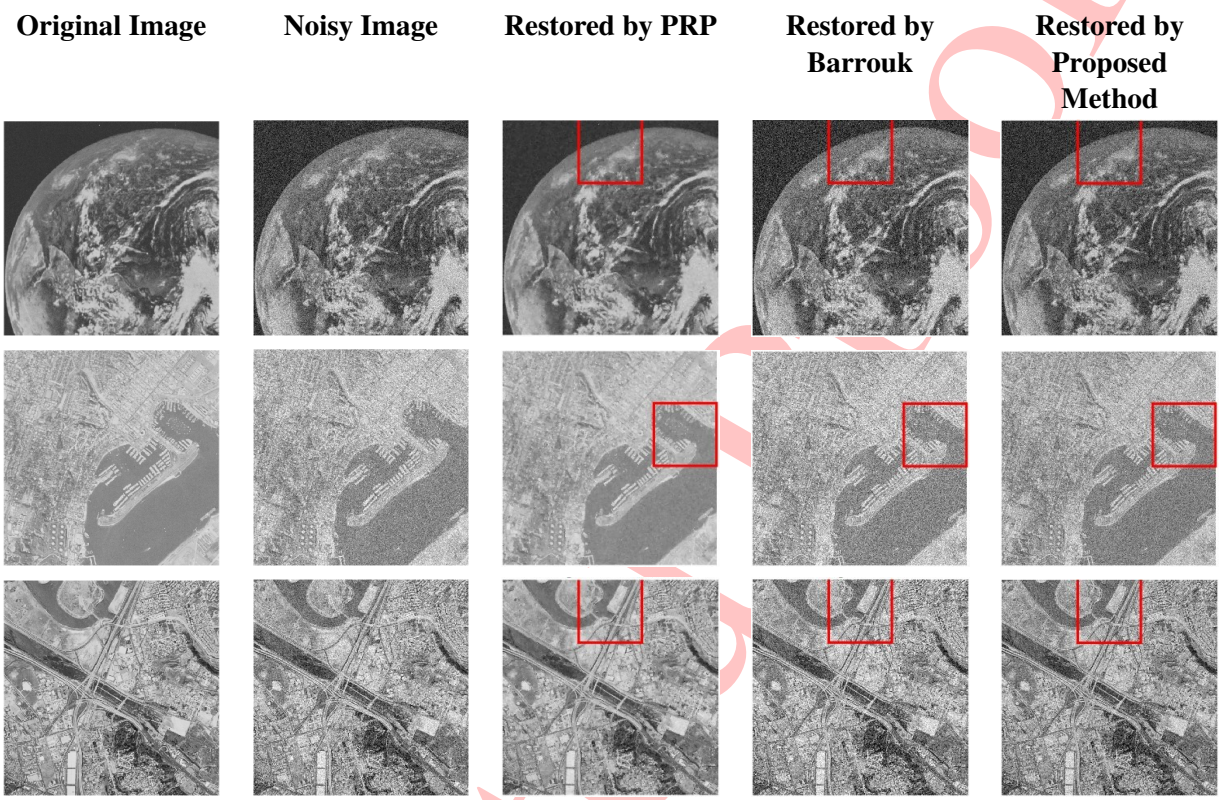

Figure 2: Original, noisy, and restored images by different methods

to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The PSNR is defined as

$$\text{PSNR} = -10\log_{10}\left(\frac{\|x_k - x_{\text{orig}}\|^2}{mn}\right),$$

where x_k is the restored image and x_{orig} is the original image. The SSIM is used to measure structural similarity and is defined as

$$\text{SSIM}(x, x_{\text{orig}}) = \frac{(2\mu_x\mu_{x_{\text{orig}}} + C_1)(2\sigma_{xx_{\text{orig}}} + C_2)}{(\mu_x^2 + \mu_{x_{\text{orig}}}^2 + C_1)(\sigma_x^2 + \sigma_{x_{\text{orig}}}^2 + C_2)},$$

where

- μ_x and $\mu_{x_{\text{orig}}}$ are the local mean intensities,
- σ_x^2 and $\sigma_{x_{\text{orig}}}^2$ are the local variances,
- $\sigma_{xx_{\text{orig}}}$ is the local covariance between x and x_{orig},
- $C_1 = (k_1 L)^2$ and $C_2 = (k_2 L)^2$ are small constants to stabilize the division when the denominator is close to zero,

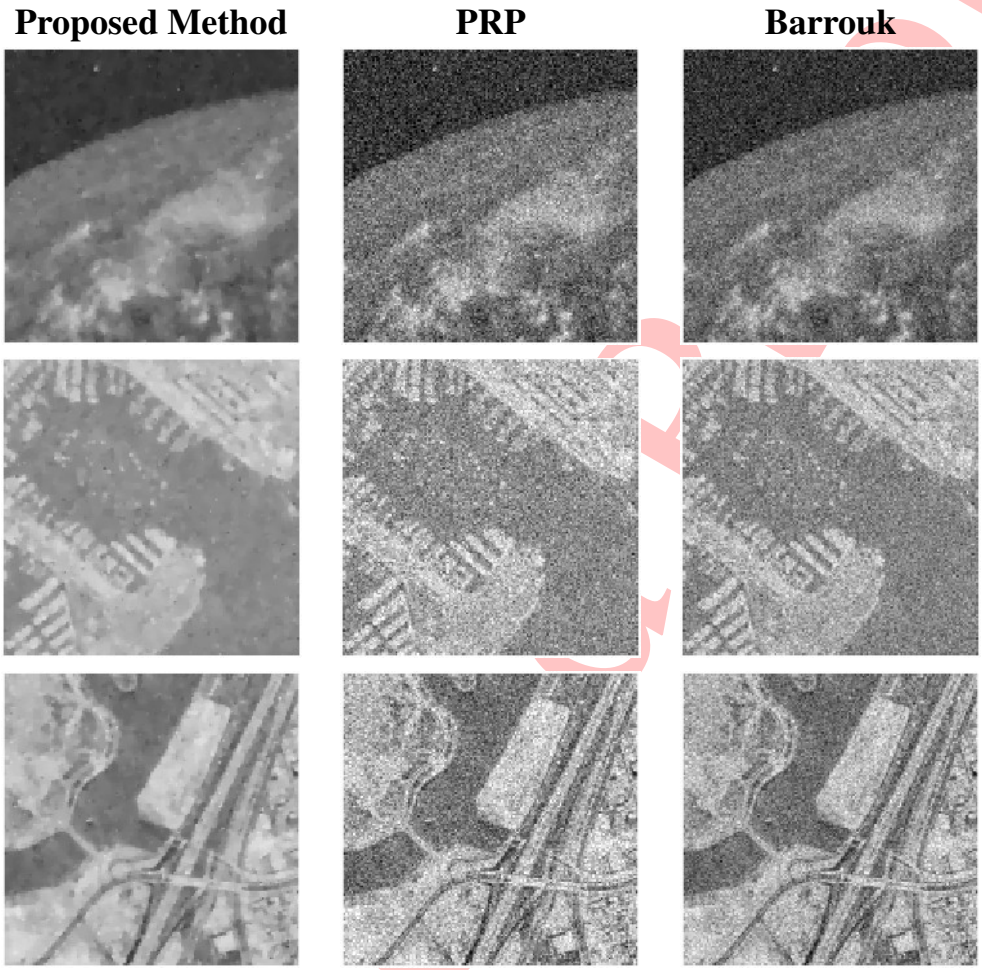

Figure 3: Zoomed-in regions showing detailed visual comparison

- L is the dynamic range of pixel values (e.g., $L = 255$ for 8-bit images),
- $k_1 = 0.01$ and $k_2 = 0.03$ are default constants.

As reported in Table 1, the proposed method achieves higher PSNR and SSIM values for all test images, demonstrating superior restoration quality in terms of noise reduction and structural preservation. Figures 2 and 3 further illustrate the visual quality of the restored images, including zoomed-in views for clearer comparison. Table 2 presents the numerical results for the proposed method, PRP, and the method of Barrouk et al. for the considered satellite images, where n denotes the total number of iterations, t represents the CPU time (in seconds), f is the objective function value at termination, and $\|g\|$ denotes the Euclidean norm of the gradient at the final iteration. These results collectively confirm that the proposed method offers an efficient and robust optimization approach for satellite image restoration problems.

4 Conclusion

In this study, an improved version of the PRP conjugate gradient algorithm was proposed by introducing a modified parameter β_k and combining it with Cao's efficient search direction. The proposed method not only guarantees sufficient descent but also achieves faster convergence and greater stability in image restoration problems. Numerical results indicate that the proposed algorithm outperforms the classical PRP method by producing clearer images and achieving higher PSNR values. These outcomes demonstrate the algorithm's effectiveness in handling complex and sensitive image processing tasks. In future work, we intend to extend the algorithm to nonsmooth optimization problems and evaluate its performance in broader areas of optimization and image processing.

Acknowledgements

The author would like to sincerely thank her supervisor for generous support and valuable guidance throughout this research.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] E. Ali, S. Mahdi, *A family of developed hybrid four-term conjugate gradient algorithms for unconstrained optimization with applications in image restoration*, Symmetry **15** (2023) 1203.

[2] M.R. Arazm, S. Babaie-Kafaki, R. Ghanbari, *An extended Dai–Liao conjugate gradient method with global convergence for nonconvex functions*, Glas. Mat. **52** (2017) 361–375.

[3] B. Barrouk, M. Belloufi, R. Benzine, T. Bechouat, *An improved PRP conjugate gradient method for optimization computation*, Int. J. Nonlinear Anal. Appl. **15** (2023) 139–147.

[4] D.P. Bertsekas, *Nonlinear Programming*, 2nd ed., Athena Scientific, Belmont, MA, 2003.

[5] A. Bertuzzi, F. Conte, F. Papa, C. Sinisgalli, *Applications of nonlinear programming to the optimization of fractionated protocols in cancer radiotherapy*, Information **11** (2020) 313.

[6] J. Cao, J. Wu, *A conjugate gradient algorithm and its applications in image restoration*, Appl. Numer. Math. **152** (2020) 243–252.

[7] R. Chan-Renous-Legoubin, C.W. Royer, *A nonlinear conjugate gradient method with complexity guarantees and its application to nonconvex regression*, arXiv:2201.08568. (2022).

[8] Y.H. Dai, Y. Yuan, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM J. Optim. **10** (1999) 177–182.

[9] R. Fletcher, C.M. Reeves, *Function minimization by conjugate gradients*, Comput. J. **7** (1964) 149–154.

[10] M.R. Hestenes, E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Natl. Bur. Stand. **49** (1952) 409–436.

[11] J. Nocedal, S.J. Wright, *Numerical Optimization*, 2nd ed., Springer, New York, 2006.

[12] B.T. Polyak, *The conjugate gradient method in extreme problems*, USSR Comput. Math. Math. Phys. **9** (1969) 94–112.

[13] E. Rubtsov, I. Chilingarian, I. Katkov, K. Grishin, V. Goradzhanov, S. Borisov, *Hybrid minimization algorithm for computationally expensive multi-dimensional fitting*, arXiv:2112.03413 (2021).

[14] F. You, I. Grossmann, *Mixed-integer nonlinear programming models and algorithms for large-scale supply chain design with stochastic inventory management*, Ind. Eng. Chem. Res. **47** (2008) 7070–7083.

[15] G. Yuan, T. Li, W. Hu, *A conjugate gradient algorithm for large-scale nonlinear equations and image restoration problems*, Appl. Numer. Math. **147** (2020) 129–141.

[16] X. Zhang, Y. Yang, *A new hybrid conjugate gradient method close to the memoryless BFGS quasi-Newton method and its application in image restoration and machine learning*, AIMS Math. **9** (2024) 27535–27556.