Inverse Problems

PAPER

A new nonlocal low-rank regularization method with applications to magnetic resonance image denoising

IOP Publishing

Inverse Problems

Inverse Problems **38** (2022) 065012 (24pp)

https://doi.org/10.1088/1361-6420/ac65ac

A new nonlocal low-rank regularization method with applications to magnetic resonance image denoising

Jian Lu[1,5], Chen Xu[1,5], Zhenwei Hu[1], Xiaoxia Liu[1,*], Qingtang Jiang[2], Deyu Meng[3] and Zhouchen Lin[4]

[1] Shenzhen Key Laboratory of Advanced Machine Learning and Applications, College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, People's Republic of China

[2] Department of Mathematics and Statistics, University of Missouri-St. Louis, St. Louis, MO 63121, United States of America

[3] School of Mathematics and Statistics and the Ministry of Education, Key Laboratory of Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China

[4] Key Laboratory of Machine Perception (MoE), School of Artificial Intelligence, Peking University, Beijing 100871, People's Republic of China

E-mail: xiaoxia_liu_math@outlook.com

Received 22 August 2021, revised 8 February 2022
Accepted for publication 8 April 2022
Published 6 May 2022

CrossMark

Abstract

Magnetic resonance (MR) images are frequently corrupted by Rician noise during image acquisition and transmission. And it is very challenging to restore MR data because Rician noise is signal-dependent. By exploring the nonlocal self-similarity of natural images and further using the low-rank prior of the matrices formed by nonlocal similar patches for 2D data or cubes for 3D data, we propose in this paper a new nonlocal low-rank regularization (NLRR) method including an optimization model and an efficient iterative algorithm to remove Rician noise. The proposed mathematical model consists of a data fidelity term derived from a maximum *a posteriori* estimation and a NLRR term using the log-det function. The resulting model in terms of approximated patch/cube matrices is non-convex and non-smooth. To solve this model, we propose an alternating reweighted minimization (ARM) algorithm using the Lipschitz-continuity of the gradient of the fidelity term and the concavity of the logarithmic function in the log-det function. The subproblems of the ARM algorithm have closed-form

[*] Author to whom any correspondence should be addressed.

[5] These authors contributed equally to this work.

solutions and its limit points are first-order critical points of the problem. The ARM algorithm is further integrated with a two-stage scheme to enhance the denoising performance of the proposed NLRR method. Experimental results tested on 2D and 3D MR data, including simulated and real data, show that the NLRR method outperforms existing state-of-the-art methods for removing Rician noise.

Keywords: MR images, Rician noise, denoising, nonlocal low-rank regularization, alternating reweighted minimization

(Some figures may appear in colour only in the online journal)

1. Introduction

Magnetic resonance (MR) imaging is an essential technique for the non-invasive observation of internal organs of the human body and plays an increasingly important role in disease diagnosis. However, in the acquisition process, due to the instability of the internal hardware of MR scanners, MR images are disturbed by the noise of Rician distribution [1]. Noise reduces the resolution of MR data, affects the accuracy of medical examination, and increases the difficulty of further computer-aided analysis of MR data, including recognition, segmentation and classification.

To remove Rician noise from MR images, filtering methods such as the anisotropic diffusion filters [2, 3] and wavelet-based filters [4] have been proposed to retain image information as much as possible. Total variation (TV) based methods have also been proposed to preserve edge features. Getreuer *et al* [5] proposed a maximum *a posteriori* (MAP) estimation model with a TV regularization. This model is complicated to solve because the objective function is non-convex. To address this issue, Getreuer *et al* [5] proposed a convex model, referred to as the GTV model, to approximate the non-convex MAP model and solved the convex model by the split Bregman iteration (SBI) algorithm. By adding a data fidelity term to the non-convex MAP model, Chen and Zeng [6] proposed a model that is strictly convex under some mild conditions. Chen *et al* [7] further improved the model using sparse representation and dictionary learning, referred to as the C-KSVD model, and solved the model using the SBI algorithm. However, the aforementioned convex MAP models may require bias corrections. In addition to the methods for 2D MR images denoising, the multi-channel denoising convolutional neural network (MCDnCNN) method [8] for 3D MR images denoising was proposed based on DnCNN [9] and achieved competitive denoising performance.

Recently, many works have been developed by exploiting the prior knowledge of nonlocal self-similarity (NSS) [10] to reduce noise in 2D and 3D MR data. According to NSS, there are many repeated local patterns in natural images and each pattern in a local patch or cube has many similar patterns across the entire image. Using the spatial pattern redundancy, Manjón *et al* improved the nonlocal means filter [10] for Gaussian noise removal to address the characteristics of Rician noise. They proposed a 2D denoising method [11], referred to as the NLM2D method, and a prefiltered rotationally invariant 3D denoising method [12], referred to as the PRI-NLM3D method, which uses a prefiltered image obtained by the oracle DCT3D method [12]. Foi [13] developed forward and inverse variance-stabilizing transformations (VSTs) to transform between Rician distribution and Gaussian distribution. Noisy data were first transformed via the forward VST, then restored by denoising methods designed for Gaussian noise removal, and later transformed back via the inverse VST. Then the benchmark NSS-based algorithms such as the block matching 3D (BM3D) algorithm [14] for 2D data and the block matching 4D (BM4D) algorithm [15] for 3D data can work with VSTs to remove Rician noise.

NSS implies that the matrix or tensor formed by nonlocal similar patches or cubes has a low-rank structure. Using this low-rank prior that seeks for sparsity in flexible bases, Zhang *et al* [16] adopted the high order singular value decomposition for 3D MR data denoising and Kong *et al* [17] adopted the tensor singular value decomposition. However, both methods like the BM3D and the BM4D methods required the forward and inverse VSTs, and the denoising process and VSTs were performed separately in sequential order. As a result, any slight perturbation in the forward VST may mislead the following denoising methods, and the errors in the denoising process may be persisted or be enlarged via the inverse VST. This unavoidable limitation of VST-based methods significantly affects the performance of those NSS-based methods in MR data denoising.

Inspired by the nonlocal low-rank regularization (NLRR) discussed in [18, 19], we develop in this paper a novel NLRR method that consists of a new optimization model for Rician noise removal and a new alternating-minimization-like algorithm to solve this model. The main contributions of this paper are summarized as follows.

- We propose a new NLRR based model to reduce Rician noise. The objective function of the proposed optimization model is composed of a NLRR term and a data fidelity term over approximated patch/cube matrices. The NLRR term is characterized using the log-det function as a non-convex non-smooth surrogate of the matrix rank, while the data fidelity term is derived from the non-convex MAP model based on the Rician distribution. The proposed model does not require any pre-processing or post-processing such as forward and inverse VSTs or bias corrections, in contrast to some existing methods for Rician noise removal such as VST-based methods and convex MAP models.
- We propose a new alternating reweighted minimization (ARM) algorithm by taking advantage of the smoothness of the fidelity term and the concavity of the logarithmic function in the NLRR term. The subproblems of the proposed ARM algorithm can be computed explicitly, in contrast to some existing NLRR based methods [18–20] that require either the ADMM algorithm or Newton's method to iteratively solve the subproblems.
- We demonstrate that the proposed NLRR method is theoretically and practically reliable. On one hand, it can be proven that any limit point of the sequence generated by the ARM algorithm is a first-order critical point of the objective function. On the other hand, we further integrate the ARM algorithm with a two-stage scheme with criteria on noise estimations can guarantee the quality of the denoised image is improved gradually.
- We show the superiority of the proposed NLRR method over several existing methods for Rician noise removal by state-of-the-art denoising results tested on 2D and 3D MR data, including simulated and real data.

The rest of this paper is organized as follows. In section 2, we review Rician noise in MR images and the MAP model. Then we propose in section 3 a new NLRR model for Rician noise removal and develop in section 4 an ARM algorithm to solve the proposed model with some convergence results. Later, in section 5, we introduce the NLRR method with a two-stage scheme and in section 6 we present the numerical results. Finally, we conclude in section 7.

2. Background

2.1. Rician noise in MRI

Rician noise naturally occurs in MR magnitude images due to the thermal noise during the acquisition process. In low signal-to-noise ratio regimes, Rician noise not only causes random fluctuations in pixel values but also introduces a signal-dependent bias that reduces image contrast [4]. Thus, it is difficult to separate the image from noise.

Suppose that $x \in \mathbb{R}^N$ is the original image. Then the observed magnitude image $y \in \mathbb{R}^N$ can be mathematically expressed as

$$y = \sqrt{(x + \eta_1)^2 + \eta_2^2},$$

where $\eta_1, \eta_2 \sim N(0, \delta^2)$. The noise in each pixel follows a Rician distribution [5] with the probability density function (PDF) defined as

$$p(y_i|x_i) = \frac{y_i}{\delta^2} e^{-\frac{y_i^2 + x_i^2}{2\delta^2}} B_0\left(\frac{y_i x_i}{\delta^2}\right),$$

where $B_0(\cdot)$ is the modified Bessel function of the first kind with order zero [21] defined as

$$B_0(t) := \frac{1}{\pi}\int_0^{\pi} e^{t\ \cos\theta}\,\mathrm{d}\theta. \tag{1}$$

2.2. The MAP model

To restore MR images corrupted by Rician noise, Getreuer *et al* [5] proposed a MAP estimation model. The image x is estimated by MAP probability and applying Bayes' theorem and the negative logarithm as follows

$$\max_x P(x|y) \iff \min_x\{-\log(P(y|x)) - \log(P(x))\}.$$

The image x is assumed to follow a TV prior with the PDF as $P(x) = \exp(-\gamma\|x\|_{\mathrm{TV}})$, where γ is a parameter, $\|x\|_{\mathrm{TV}} := \sum_i \sqrt{(\nabla_1 x)_i^2 + (\nabla_2 x)_i^2}$ is the TV of x, and ∇_1 and ∇_2 are discrete first order difference operators in the horizontal and vertical directions, respectively. Hence, the MAP model with a TV regularization term is formulated as

$$\min_x \frac{1}{2\delta^2}\|x\|_2^2 - \left\langle \log\, B_0\left(\frac{xy}{\delta^2}\right), \mathbf{1}\right\rangle + \gamma\|x\|_{\mathrm{TV}}, \tag{2}$$

where $B_0\left(\frac{xy}{\delta^2}\right)$ is performed componentwise, $B_0(\cdot)$ is defined as in (1) and $\mathbf{1}$ denotes an $N \times 1$ vector of all ones. Throughout the paper, we refer to model (2) as the MAP model.

Let $f : \mathbb{R}^N \to \mathbb{R}$ be the data fidelity term of the MAP model in (2) defined as

$$f(x) := \frac{1}{2\delta^2}\|x\|_2^2 - \left\langle \log\, B_0\left(\frac{xy}{\delta^2}\right), \mathbf{1}\right\rangle. \tag{3}$$

The function f is differentiable and its gradient ∇f can be computed as follows

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_N}\right]^{\mathrm{T}}, \tag{4}$$

where

$$\frac{\partial f}{\partial x_i} = \frac{1}{\delta^2} x_i - \frac{y_i}{\delta^2} \frac{B_0'\left(\frac{x_i y_i}{\delta^2}\right)}{B_0\left(\frac{x_i y_i}{\delta^2}\right)}.$$

It is verified in [22] that f is lower bounded and smooth with a Lipschitz-continuous gradient ∇f as follows.

Definition 1 (Lipschitz continuous). A function $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ is Lipschitz continuous with a Lipschitz constant $L > 0$ (or called L-Lipschitz continuous) if for all $x_1, x_2 \in \mathbb{R}^d$,

$$\|f(x_1) - f(x_2)\|_2 \leqslant L\|x_1 - x_2\|_2.$$

Proposition 2. *Given $y \in \mathbb{R}^N$, let f be defined as in (3). Then the following statements hold:*

(*a*) *f is lower bounded, that is, $\inf_{x\in\mathbb{R}^N} f(x) > -\infty$;*
(*b*) *f has a Lipschitz-continuous gradient ∇f with the Lipschitz constant $L = \frac{1}{\delta^2}$.*

The MAP-based fidelity term defined as in (3) is not widely used in the literature due to its non-convexity that may make the resulting model difficult to solve. The existing studies for Rician noise removal focus on its convex approximation, e.g., GTV model and CZ model, but they may require bias corrections, e.g., mean shifting, leading to unexpected artifacts. As shown in proposition 2, the lower boundedness of f leads to the existence of minimizers of the MAP model and the smoothness of f helps develop efficient algorithms to solve the MAP model. Therefore, we are in favor of the non-convex MAP-based fidelity term (3) over convex fidelity terms and we will overcome the difficulty in solving a non-convex model by analysing its Lipschitz-continuous gradient.

3. A new NLRR based model for MRI denoising

In this section, we propose a new model based on NLRR to remove Rician noise in 2D and 3D MR images.

3.1. Nonlocal low-rank regularization

NSS refers to the assumption that for each patch in a natural image, similar patches can be found across the image. Intuitively, the matrix formed by nonlocal similar patches should be low-rank. To regularize the low-rank prior of patch matrices, there are two key components in NLRR based models [18–20]: patch grouping and low-rank matrix approximation.

First, we group together nonlocal patches with similar patterns by block matching [14] and formulate the extraction of a nonlocal similar patch matrix. Suppose that $\hat{x} \in \mathbb{R}^N$ is an estimated clean image. For an exemplary patch of size $\sqrt{m} \times \sqrt{m}$ at position j denoted by $\hat{x}_j \in \mathbb{R}^m$, we search within a wide neighborhood (e.g., a 30×30 window) for a total of n patches that are most similar to the exemplary patch in terms of the Euclidean distance. In the patch group G_j containing those similar patches, a patch matrix $R_j(\hat{x})$ formed by all the patches in G_j can be formulated using an extraction operator $R_j : \mathbb{R}^N \to \mathbb{R}^{m\times n}$ defined as follows

$$R_j(x) := \begin{bmatrix} \hat{R}_{j1}x & \hat{R}_{j2}x & \dots & \hat{R}_{jn}x \end{bmatrix}, \tag{5}$$

where $\hat{R}_{ji} \in \mathbb{R}^{m\times N}$ is a binary matrix such that $\hat{R}_{ji}\hat{x}$ is the ith patch in group G_j. According to NSS, the patch matrix $R_j(x)$ with similar structures should be a low-rank matrix if x is close to the clean image $\hat{x}$.

Second, after grouping nonlocal similar patches, we regularize the patch matrices using low-rank constraints. In this paper, we use the log-det function, which is a non-convex surrogate of the rank function, to characterize the low-rank property of nonlocal similar patch matrices. For a matrix $X \in \mathbb{R}^{m\times n}$, we consider log $\det((XX^{\mathrm{T}})^{1/2} + \varepsilon I)$ if $m \leqslant n$ and log $\det((X^{\mathrm{T}}X)^{1/2} + \varepsilon I)$ if $m > n$. The log-det function of X can be further written as

$$\Psi(X) := \sum_{i=1}^{\ell} \log\left(\sigma_i\left(X\right) + \varepsilon\right), \tag{6}$$

where $\sigma_i(X)$ is the ith largest singular value of X, $\ell = \min\{m, n\}$ and ε represents a small positive constant. The log-det function can be interpreted as a sum of the logarithms of the singular values. The logarithmic function is a concave but smooth function.

The NLRR discussed above for 2D data can be extended to 3D data based on the additional spatial similarity (the depth) in volumetric data or the temporal similarity in videos. We can group nonlocal similar cubes by cube matching and form nonlocal similar cube matrices by stacking the vectorized cubes together. The function Ψ defined in (6) can also work efficiently in regularizing the low-rank property of these nonlocal similar cube matrices.

3.2. Proposed model

We propose to develop a new model by combining the NLRR with the MAP-based data fidelity to remove Rician noise. The resulting model is a global optimization problem in terms of the entire unknown image and approximated nonlocal similar patch/cube matrices.

Let $y \in \mathbb{R}^N$ denote the noisy image and let $x \in \mathbb{R}^N$ denote the unknown clean image. We consider the patch/cube matrices $R_j(x)$ and $R_j(y)$ extracted from x and y, respectively. Then $R_j(y)$ corrupted with Rician noise can be expressed as

$$R_j\left(y\right) = \sqrt{(R_j(x) + \eta_1)^2 + \eta_2^2},$$

where $R_j(x)$ is the clean patch/cube matrix and $\eta_1, \eta_2 \sim N(0, \delta^2)$.

First, to measure the closeness between $R_j(x)$ and $R_j(y)$, we extend the MAP-based fidelity term defined as in (3) to be a fidelity term in terms of a patch/cube matrix $X_j \in \mathbb{R}^{m\times n}$ as follows

$$f_j\left(X_j\right) := \frac{1}{2\delta^2}\|X_j\|_F^2 - \langle \log\, B_0\left(\frac{X_jY_j}{\delta^2}\right), \mathbf{1}\rangle, \tag{7}$$

where $Y_j = R_j(y)$, $B_0(\cdot)$ is defined as in (1), log $B_0\left(\frac{X_jY_j}{\delta^2}\right)$ is performed componentwise, $\mathbf{1}$ denotes an $m \times n$ matrix of all ones and $\|\cdot\|_F$ is the Frobenius norm for matrices.

Second, to characterize the low-rank property on $R_j(x)$, we consider the NLRR defined as in (6). Then a NLRR based model for Rician noise removal can be written as follows

$$\min_x \sum_{j=1}^{J} \left[f_j\left(R_j\left(x\right)\right) + \lambda_j\Psi(R_j\left(x\right))\right], \tag{8}$$

where $\lambda_j > 0$ is a regularization parameter. Due to the non-convexity of f_j and the non-convexity and non-smoothness of Ψ, model (8) is non-convex and non-smooth and is difficult to solve. Furthermore, the model is a composite optimization problem with a linear but not invertible operator R_j, which significantly increases the complexity of the model.

Lastly, to overcome the difficulty in solving the composite optimization problem (8), we apply the variable splitting method to relax the model. By introducing auxiliary variables $X_j \in \mathbb{R}^{m\times n}$ such that $X_j = R_j(x)$, $j = 1, 2, \ldots, J$, and then relaxing these equalities, we obtain the following model

$$\min_{x,X_1,\ldots,X_J} \sum_{j=1}^{J} \left[\frac{1}{2}\|X_j - R_j(x)\|_F^2 + \mu_j f_j(X_j) + \lambda_j \Psi(X_j)\right], \tag{9}$$

where f_j is the MAP-based defined as in (6), R_j is the patch/cube matrix extraction operator defined as in (5), $\mu_j > 0$ is a fidelity parameter and $\lambda_j > 0$ is a regularization parameter.

This new model enforces the MAP-based fidelity and the low-rank regularization over approximated nonlocal similar patch/cube matrices X_j and aims to restore the entire image x by minimizing the Euclidean distance between the desired patch/cube matrices $R_j(x)$ and the approximated patch/cube matrices X_j. Compared to existing NLRR based models [18–20], the proposed model in terms of approximated patch/cube matrices has low complexity. This helps develop efficient algorithms, especially using the smoothness of the fidelity term. Moreover, the proposed model of a general form can be extended to other applications in image processing, especially in the case that the data fidelity term has a Lipschitz-continuous gradient.

In section 4, we will show that model (9) can be efficiently solved by a reweighted version of alternating minimization (AM) algorithm whose limit points are first-order critical points of the problem; and in section 6, the experimental results will demonstrate that our proposed model can outperform other existing methods for Rician noise removal.

4. ARM algorithm and convergence analysis

In this section, we propose a new optimization algorithm to solve model (9) and provide a convergence analysis of the proposed algorithm.

To begin with, we denote the objective function of model (9) as

$$\Phi(X_1, \ldots, X_J, x) := \sum_{j=1}^{J} F_j(X_j, x), \tag{10}$$

where the function $F_j : \mathbb{R}^{m\times n} \times \mathbb{R}^N \to \mathbb{R}$ is defined as

$$F_j(X_j, x) := \frac{1}{2}\|X_j - R_j(x)\|_F^2 + \mu_j f_j(X_j) + \lambda_j \Psi(X_j), \tag{11}$$

f_j is defined as in (7), Ψ is defined as in (6) and R_j is defined as in (5).

The objective function Φ is non-convex and non-smooth because f_j is non-convex and Ψ is non-convex and non-smooth. Fortunately, f_j is lower bounded and smooth according to proposition 2. The lower boundedness of f_j implies that the solutions of model (9) exist. The smoothness of f_j helps develop efficient algorithms for solving model (9).

In the following, we first introduce a class of first-order critical points of Φ, then develop an ARM algorithm, and later establish a convergence analysis of the proposed algorithm.

4.1. Critical points of the proposed model

Before characterizing the first-order critical points of the objective function Φ, we first introduce some notations.

First, we introduce the transpose of R_j, which is linear. Define $R_j^{\mathrm{T}} : \mathbb{R}^{m\times n} \to \mathbb{R}^N$ as

$$R_j^{\mathrm{T}}(X) := \sum_{i=1}^{n} \hat{R}_{ji}^{\mathrm{T}} x_i, \tag{12}$$

where $x_i \in \mathbb{R}^m$ is the ith vector of X. Note that $\langle R_j(x), X\rangle_F = \langle x, R_j^{\mathrm{T}}(X)\rangle$, for all $x \in \mathbb{R}^N$ and $X \in \mathbb{R}^{m\times n}$. Let

$$W := \sum_{j=1}^{J} R_j^{\mathrm{T}} \circ R_j = \sum_{j=1}^{J}\sum_{i=1}^{n} \hat{R}_{ji}^{\mathrm{T}} \hat{R}_{ji}. \tag{13}$$

Since $\hat{R}_{ji}$ has linearly independent rows, $\hat{R}_{ji}^{\mathrm{T}}\hat{R}_{ji}$ is a diagonal matrix. Then W is a diagonal matrix whose diagonal entries are the counts for each pixel. As we assume each pixel belongs to at least one nonlocal similar patch/cube group, W is invertible.

Then we recall some notations for the SVD. For a vector $a \in \mathbb{R}^\ell$, let $\mathrm{Diag}(a)$ denote the $\ell \times \ell$ diagonal matrix whose ith diagonal element is a_i. For a matrix $X \in \mathbb{R}^{m\times n}$, the SVD of X is expressed as $X = U\Sigma V^{\mathrm{T}}$, where $U \in \mathbb{R}^{m\times \ell}$ and $V \in \mathbb{R}^{n\times \ell}$ are orthogonal matrices with $U^{\mathrm{T}}U = V^{\mathrm{T}}V = I$, and $\Sigma \in \mathbb{R}^{\ell\times\ell}$ is a diagonal matrix with $\Sigma = \mathrm{Diag}(\sigma(X))$, $\sigma(X) := [\sigma_1(X), \dots, \sigma_\ell(X)]^{\mathrm{T}}$, where $\sigma_i(X)$ is the ith largest singular value of X and $\ell = \min\{m, n\}$.

Next, motivated by the class of first-order critical points for ℓ_p regularized low-rank approximation problems introduced in [23], we define a class of first-order critical points for model (9) using

$$\bar{\mathcal{M}}(X) := \left\{(\bar{U}, \bar{V}) \in \mathbb{R}^{m\times r} \times \mathbb{R}^{n\times r} : \bar{U}^{\mathrm{T}}\bar{U} = \bar{V}^{\mathrm{T}}\bar{V} = I \text{ and } X = \bar{U}\mathrm{Diag}(\bar{\sigma}(X))\bar{V}^{\mathrm{T}}\right\},$$

where $\bar{\sigma}(X) := [\sigma_1(X), \dots, \sigma_r(X)]^{\mathrm{T}}$ and $r = \mathrm{rank}(X)$. Note that $\bar{\mathcal{M}}(X)$ is the set of all such pairs $(\bar{U}, \bar{V})$ of the rank reduced SVD of X.

Definition 3. A point $\left(X_1^*, \dots, X_J^*, x^*\right)$ is a first-order critical point of problem (9) if

$$x^* = W^{-1}\sum_{j=1}^{J} R_j^{\mathrm{T}}\left(X_j^*\right), \tag{14}$$

and for all $j = 1, 2, \dots, J$,

$$0 \in \left\{\bar{U}_j^{\mathrm{T}}\left(X_j^* - R_j(x^*) + \mu_j \nabla f_j(X_j^*)\right)\bar{V}_j + \lambda_j\, Diag((\bar{\sigma}(X^*) + \varepsilon)^{-1}) : (\bar{U}_j, \bar{V}_j) \in \bar{\mathcal{M}}\left(X_j^*\right)\right\}, \tag{15}$$

where R_j^{T} is defined as in (12) and W is defined as in (13).

The next theorem shows that a local minimizer of problem (9) is a first-order critical point.

Theorem 4. *Suppose that* $\left(X_1^*, \dots, X_J^*, x^*\right)$ *is a local minimizer of problem* (9)*. Then* $\left(X_1^*, \dots, X_J^*, x^*\right)$ *is a first-order critical point of problem* (9)*, that is,* (14) *and* (15) *hold at* $\left(X_1^*, \dots, X_J^*, x^*\right)$.

Proof. If $\left(X_1^*, \dots, X_J^*, x^*\right)$ is a local minimizer of problem (9), then (a) x^* is a local minimizer of $\Phi(X_1^*, \dots, X_J^*, x)$ with respect to x; (b) X_j^* is a local minimizer of $F_j(X_j, x^*)$ with respect to X_j, for all $j = 1, 2, \dots, J$.

Since $\Phi(X_1^*,\ldots,X_J^*,x)$ is differentiable with respect to x, then (a) implies $0=\nabla_x\Phi(X_1^*,\ldots,X_J^*,x^*)=Wx^*-\sum_{j=1}^J R_j^{\mathrm{T}}(X_j^*)$. Hence, $x^*=W^{-1}\sum_{j=1}^J R_j^{\mathrm{T}}\left(X_j^*\right)$.

Let $X_j^*=U_j\,\mathrm{Diag}(\bar{\sigma}(X_j^*))V_j^{\mathrm{T}}$ for some $(U_j,V_j)\in\bar{\mathcal{M}}(X_j^*)$ and $r_j=\mathrm{rank}(X_j^*)$. Then (b) implies that 0 is a local minimizer of the following problem

$$\min_{Z\in\mathbb{R}^{r_j\times r_j}}\left\{\frac{1}{2}\left\|(X_j^*+U_jZV_j^{\mathrm{T}})-R_j(x^*)\right\|_F^2+\mu_jf_j(X_j^*+U_jZV_j^{\mathrm{T}})+\lambda_j\Psi(X_j^*+U_jZV_j^{\mathrm{T}})\right\}.$$

This together with the fact

$$\Psi\left(X_j^*+U_jZV_j^{\mathrm{T}}\right)=\Psi\left(U_j\,\mathrm{Diag}(\bar{\sigma}(X_j^*))V_j^{\mathrm{T}}+U_jZV_j^{\mathrm{T}}\right)=\Psi\left(\mathrm{Diag}(\bar{\sigma}(X_j^*))+Z\right),$$

implies that 0 is a local minimizer of the problem

$$\min_{Z\in\mathbb{R}^{r_j\times r_j}}\left\{\frac{1}{2}\left\|(X_j^*+U_jZV_j^{\mathrm{T}})-R_j(x^*)\right\|_F^2+\mu_jf_j(X_j^*+U_jZV_j^{\mathrm{T}})+\lambda_j\Psi(\mathrm{Diag}(\bar{\sigma}(X_j^*))+Z)\right\}. \tag{16}$$

Let $S(Z)$ denote the objective function of problem (16). By theorem 7.1 in [24] and the definition of $\bar{\mathcal{M}}(\cdot)$, the subdifferential of $S(Z)$ at $Z=0$ is given by

$$\begin{aligned}\partial S(0)=\big\{&U_j^{\mathrm{T}}\left(X_j^*-R_j(x^*)+\mu_j\nabla f_j(X_j^*)\right)V_j\\&+\lambda_j\tilde{U}_j\,\mathrm{Diag}((\bar{\sigma}(X_j^*)+\varepsilon)^{-1})\tilde{V}_j^{\mathrm{T}}:(\tilde{U}_j,\tilde{V}_j)\in\bar{\mathcal{M}}\left(\mathrm{Diag}(\bar{\sigma}(X_j^*))\right)\big\}.\end{aligned}$$

Since 0 is a local minimizer of problem (16), the first-order optimality condition of (16) yields $0\in\partial S(0)$. Hence, there exists some $(\tilde{U}_j,\tilde{V}_j)\in\bar{\mathcal{M}}\left(\mathrm{Diag}(\bar{\sigma}(X_j^*))\right)$ such that

$$U_j^{\mathrm{T}}\left(X_j^*-R_j(x^*)+\mu_j\nabla f_j(X_j^*)\right)V_j+\lambda_j\tilde{U}_j\,\mathrm{Diag}((\bar{\sigma}(X_j^*)+\varepsilon)^{-1})\tilde{V}_j^{\mathrm{T}}=0. \tag{17}$$

Upon pre- and post-multiplying (17) by $\tilde{U}_j^{\mathrm{T}}$ and $\tilde{V}_j$, and using $\tilde{U}_j^{\mathrm{T}}\tilde{U}_j=\tilde{V}_j^{\mathrm{T}}\tilde{V}_j=I$, we obtain

$$\bar{U}_j^{\mathrm{T}}\left(X_j^*-R_j(x^*)+\mu_j\nabla f_j(X_j^*)\right)\bar{V}_j+\lambda_j\,\mathrm{Diag}((\bar{\sigma}(X_j^*)+\varepsilon)^{-1})=0,$$

where $\bar{U}_j=U_j\tilde{U}_j$ and $\bar{V}_j=V_j\tilde{V}_j$. Since $(U_j,V_j)\in\bar{\mathcal{M}}(X_j^*)$ and $(\tilde{U}_j,\tilde{V}_j)\in\bar{\mathcal{M}}\left(\mathrm{Diag}(\bar{\sigma}(X_j^*))\right)$, then we have

$$\bar{U}_j\,\mathrm{Diag}(\bar{\sigma}(X_j^*))\bar{V}_j^{\mathrm{T}}=U_j\left(\tilde{U}_j\,\mathrm{Diag}(\bar{\sigma}(X_j^*))\tilde{V}_j^{\mathrm{T}}\right)V_j^{\mathrm{T}}=U_j\,\mathrm{Diag}(\bar{\sigma}(X_j^*))V_j^{\mathrm{T}}=X_j^*.$$

□

Hence, $(\bar{U}_j,\bar{V}_j)\in\bar{\mathcal{M}}(X_j^*)$ and (15) holds.

4.2. Proposed algorithm

We aim to develop a new algorithm to solve model (9). The AM algorithm and the well-known ADMM algorithm are not applicable to this problem because it is challenging to minimize $F_j(X_j,x)$ with respect to X_j due to the non-convexity of f_j and the non-convexity and non-smoothness of Ψ. To overcome this challenge, we take advantage of the Lipschitz continuity of ∇f_j and the concavity of the logarithmic function in Ψ and further propose an efficient algorithm whose subproblems have closed-form solutions and which has convergence results.

We first recall a descent lemma [25] for functions with a Lipschitz continuous gradient and a lemma derived from the concavity of a smooth logarithmic function.

Lemma 5 (Descent lemma). *Let $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ be a continuously differentiable function with gradient ∇f assumed L-Lipschitz continuous. Then, for all $X_1, X_2 \in \mathbb{R}^{m\times n}$,*

$$f(X_1) \leqslant f(X_2) + \langle X_1 - X_2, \nabla f(X_2)\rangle + \frac{L}{2}\|X_1 - X_2\|_F^2.$$

Lemma 6 (Property of log-det). *Let $\Psi : \mathbb{R}^{m\times n} \to \mathbb{R}$ be defined as in (6). Then, for all $X_1, X_2 \in \mathbb{R}^{m\times n}$,*

$$\Psi(X_1) \leqslant \Psi(X_2) + \sum_{i=1}^{\ell} w_i(\sigma_i(X_1) - \sigma_i(X_2)),$$

where $w_i = \frac{1}{\sigma_i(X_2)+\varepsilon}$, $i = 1, 2, \ldots, \ell$.

Proof. Since the logarithmic function $\log(t + \varepsilon)$ is concave and continuously differentiable on $[0, \infty)$, by the definition of supergradients, we have

$$\log(\sigma_i(X_1) + \varepsilon) \leqslant \log(\sigma_i(X_2) + \varepsilon) + w_i(\sigma_i(X_1) - \sigma_i(X_2)),$$

where $w_i = \frac{1}{\sigma_i(X_2)+\varepsilon}$, $i = 1, 2, \ldots, \ell$. By summing the inequality from $i = 1$ to $i = \ell$, the result holds. □

To tackle the challenge in minimizing $F_j(X_j, x)$ with respect to X_j, we apply the above lemmas to approximate $F_j(X_j, x)$. In particular, we apply lemma 5 to $\frac{1}{2}\|X_j - R_j(x)\|_F^2 + \mu_j f_j(X_j)$ at points $X_1 = X_j$ and $X_2 = X_j^k$ and achieve a quadratic upper approximation; and we use lemma 6 at points $X_1 = X_j$ and $X_2 = X_j^k$ and achieve a linear upper approximation of $\Psi(X_j)$ with a reweighted scheme.

Now we propose to develop an *ARM* algorithm. That is, starting with some given initial point $(X_1^0, \ldots, X_J^0, x^0)$, we generate a sequence $\{(X_1^k, \ldots, X_J^k, x^k)\}_{k\in\mathbb{N}}$ by alternately minimizing an approximation of the objective function Φ with respect to the approximated low-rank patch/cube matrix X_j and the entire image x as follows

$$X_j^{k+1} = \underset{X_j}{\operatorname{argmin}} \left\{ \langle X_j - X_j^k, X_j^k - R_j(x^k) + \mu_j \nabla f_j(X_j^k)\rangle + \frac{c_j}{2}\|X_j - X_j^k\|_F^2 + \lambda_j \sum_{i=1}^{\ell} (w_j^k)_i \sigma_i(X_j) \right\}, \quad (18)$$

$$x^{k+1} = W^{-1} \sum_{j=1}^{J} R_j^{\mathrm{T}}(X_j^{k+1}), \quad (19)$$

where $(w_j^k)_i = \frac{1}{\sigma_i(X_j^k)+\varepsilon}$ and $c_j > 0$, $j = 1, 2, \ldots, J$.

We observe that the update of X_j^{k+1} in (18) at the $(k + 1)$th step is equivalent to

$$X_j^{k+1} = \underset{X_j}{\operatorname{argmin}} \left\{ \lambda_j \sum_{i=1}^{\ell} (w_j^k)_i \sigma_i(X_j) + \frac{c_j}{2} \left\| X_j - \left(X_j^k - \left(X_j^k - R_j(x^k) + \mu_j \nabla f_j(X_j^k) \right) / c_j \right) \right\|_F^2 \right\}. \quad (20)$$

Algorithm 1. ARM.

Input: Noisy image y, initial image x^0
Initialize: $\mu_j > 0$, $\lambda_j > 0$, $\varepsilon > 0$, $c_j > 1 + \mu_j L$,
$X_j^0 = R_j(x^0)$, $(w_j^0)_i = \frac{1}{\sigma_i(X_j^0)+\varepsilon}$, $k = 0$

1: Set R_j via block/cube matching based on x^0
2: **repeat**
3: **for** j from 1 to J **do**
4: $Z_j^k = X_j^k - \left(X_j^k - R_j(x^k) + \mu_j \nabla f_j(X_j^k)\right)/c_j$
5: $[U_j^k, \Lambda_j^k, V_j^k] = \mathrm{SVD}\left(Z_j^k\right)$
6: $\Sigma_j^{k+1} = S_{\lambda_j/c_j, w_j^k}(\Lambda_j^k)$
7: $X_j^{k+1} = U_j^k \Sigma_j^{k+1} (V_j^k)^{\mathrm{T}}$
8: $(w_j^{k+1})_i = 1/((\Sigma_j^{k+1})_{ii} + \varepsilon)$
9: **end for**
10: $x^{k+1} = W^{-1} \sum_{j=1}^{J} R_j^{\mathrm{T}} \left(X_j^{k+1}\right)$
11: $k \leftarrow k+1$
12: **until** stopping criterion is satisfied
Output: x^{k+1}

Note that the weight vector $w_j^k = [(w_j^k)_1, (w_j^k)_2, \ldots, (w_j^k)_\ell]^{\mathrm{T}}$ is in an ascending order, that is, $0 \leqslant (w_j^k)_1 \leqslant (w_j^k)_2 \leqslant \cdots \leqslant (w_j^k)_\ell$, since $\sigma_1(X_j^k) \geqslant \sigma_2(X_j^k) \geqslant \cdots \geqslant \sigma_\ell(X_j^k) \geqslant 0$. Then the solution of (20) can be uniquely obtained by the weighted singular value thresholding operator in the following theorem from [26].

Theorem 7. *For $Y \in \mathbb{R}^{m\times n}$, $\lambda > 0$ and $w = [w_1, w_2, \ldots, w_\ell]^{\mathrm{T}}$ such that $0 \leqslant w_1 \leqslant w_2 \leqslant \cdots \leqslant w_\ell$, $\ell = \min\{m, n\}$, the following problem*

$$\hat{X} = \operatorname*{argmin}_X \left\{ \frac{1}{2}\|X - Y\|_F^2 + \lambda \sum_{i=1}^{\ell} w_i \sigma_i(X) \right\},$$

has a global optimal solution which is given by

$$\hat{X} = U S_{\lambda,w}(\Sigma) V^{\mathrm{T}},$$

where $Y = U\Sigma V^{\mathrm{T}}$ is the SVD of Y, and $S_{\lambda,w}(\Sigma)$ is the weighted singular value thresholding operator of diagonal matrix Σ with its (i, i)-entry given by

$$S_{\lambda,w}(\Sigma)_{ii} = \max\left(\Sigma_{ii} - \lambda w_i, 0\right).$$

In algorithm 1, we present the proposed ARM algorithm for solving model (9). In particular, the solution of (20) is computed in line 4–7 according to theorem 7.

4.3. Convergence analysis

We prove that any limit point of the sequence $\{(X_1^k, \ldots, X_J^k, x^k)\}_{k\in\mathbb{N}}$ generated by the ARM algorithm in algorithm 1 is a first-order critical point of Φ.

First, we show that $\Phi(X_1^k, \ldots, X_J^k, x^k)$ is decreasing as k goes to ∞ and the distance between two successive iterates tends to 0.

Theorem 8. *Let Φ be the objective function of model (9) and let $\{(X_1^k, \ldots, X_J^k, x^k)\}_{k\in\mathbb{N}}$ be a sequence generated by the ARM algorithm in algorithm 1. Suppose $c_j > 1 + \mu_j L$, $j = 1, 2, \ldots, J$. Then the following statements hold:*

(*a*) *For all* $k \in \mathbb{N}$,

$$\Phi(X_1^k, \ldots, X_J^k, x^k) - \Phi(X_1^{k+1}, \ldots, X_J^{k+1}, x^{k+1})$$
$$\geqslant \sum_{j=1}^{J} \frac{c_j - (1 + \mu_j L)}{2} \|X_j^{k+1} - X_j^k\|_F^2 > 0. \tag{21}$$

(*b*) $\lim_{k\to\infty} \|X_j^{k+1} - X_j^k\|_F^2 = 0$, $j = 1, 2, \ldots, J$, *and* $\lim_{k\to\infty} \|x^{k+1} - x^k\|_2^2 = 0$.

Proof.

(a) Let F_j be defined as in (11). By first applying lemmas 5 and 6 to $F_j(X_j, x^k)$ and then substituting $X_j = X_j^{k+1}$, we have

$$F_j(X_j^{k+1}, x^k) \leqslant F_j(X_j^k, x^k) + \frac{1 + \mu_j L}{2} \|X_j^{k+1} - X_j^k\|_F^2$$
$$+ \langle X_j^{k+1} - X_j^k, X_j^k - R_j(x^k) + \mu_j \nabla f_j(X_j^k) \rangle$$
$$+ \lambda_j \sum_{i=1}^{\ell} (w_j^k)_i \left(\sigma_i(X_j^{k+1}) - \sigma_i(X_j^k) \right)$$
$$\leqslant F_j(X_j^k, x^k) - \frac{c_j - (1 + \mu_j L)}{2} \|X_j^{k+1} - X_j^k\|_F^2,$$

where $(w_j^k)_i = \frac{1}{\sigma_i(X_j^k) + \varepsilon}$, $i = 1, 2, \ldots, \ell$, $\ell = \min\{m, n\}$, and the last inequality is derived from (18). Summing this inequality from $j = 1$ to J, we have

$$\Phi(X_1^{k+1}, \ldots, X_J^{k+1}, x^k) \leqslant \Phi(X_1^k, \ldots, X_J^k, x^k)$$
$$- \sum_{j=1}^{J} \frac{c_j - (1 + \mu_j L)}{2} \|X_j^{k+1} - X_j^k\|_F^2.$$

Also, the update of x^{k+1} given in (19) implies that $\Phi(X_1^{k+1}, \ldots, X_J^{k+1}, x^{k+1}) \leqslant \Phi(X_1^{k+1}, \ldots, X_J^{k+1}, x^k)$ and then statement (a) holds.

(b) Summing two sides of (21) from $k = 0$ to ∞, we have

$$\sum_{j=1}^{J} \frac{c_j - (1 + \mu_j L)}{2} \sum_{k=0}^{\infty} \|X_j^{k+1} - X_j^k\|_F^2 \leqslant \Phi(X_1^0, \ldots, X_J^0, x^0)$$
$$- \Phi(X_1^\infty, \ldots, X_J^\infty, x^\infty) < \infty,$$

since $\inf \Phi > -\infty$.

As we assume $c_j > 1 + \mu_j L$, $j = 1, 2, \ldots, J$, we have $\lim_{k\to\infty} \|X_j^{k+1} - X_j^k\|_F^2 = 0$, $j = 1, 2, \ldots, J$. Due to (19), we also have $\lim_{k\to\infty} \|x^{k+1} - x^k\|_2^2 = 0$.

Second, we show that the sequence $\{(X_1^k, \ldots, X_J^k, x^k)\}_{k\in\mathbb{N}}$ is bounded, which implies the existence of limit points. Then we further prove that any limit point is a first-order critical point of problem (9).

Theorem 9. *Let $\{(X_1^k, \ldots, X_J^k, x^k)\}_{k\in\mathbb{N}}$ be a sequence generated by algorithm 1. Then the following statements hold:*

(*a*) *The sequence $\{(X_1^k, \ldots, X_J^k, x^k)\}_{k\in\mathbb{N}}$ is bounded;*

(*b*) *Let $(X_1^*, \ldots, X_J^*, x^*)$ be any limit point of $\{(X_1^k, \ldots, X_J^k, x^k)\}_{k\in\mathbb{N}}$. Then $(X_1^*, \ldots, X_J^*, x^*)$ is a first-order critical point of problem (9), i.e., (14) and (15) hold at $(X_1^*, \ldots, X_J^*, x^*)$.*

Proof.

(a) We know that $\Phi(X_1^k, \ldots, X_J^k, x^k) \leqslant \Phi(X_1^0, \ldots, X_J^0, x^0)$ for all $k \in \mathbb{N}$. This together with the definition of Φ and $\underline{F} = \sum_{j=1}^J \inf_{X_j\in\mathbb{R}^{m\times n}} f_j(X_j)$ implies that

$$\begin{aligned}\underline{F} + \sum_{j=1}^J \lambda_j \Psi(X_j^k) &\leqslant \sum_{j=1}^J \left(\frac{1}{2}\left\|X_j^k - R_j(x^k)\right\|_F^2 + \mu_j f_j(X_j^k) + \lambda_j\Psi(X_j^k)\right)\\ &\leqslant \Phi(X_1^0, \ldots, X_J^0, x^0) < \infty.\end{aligned}$$

Since $\lim_{\|X\|_F\to\infty} \Psi(X) = \infty$, $\{X_j^k\}_{k\in\mathbb{N}}$ is bounded for $j = 1, 2, \ldots, J$. Due to (19), we know $\{x^k\}_{k\in\mathbb{N}}$ is also bounded. Hence, $\{(X_1^k, \ldots, X_J^k, x^k)\}_{k\in\mathbb{N}}$ is bounded.

(b) Since $(X_1^*, \ldots, X_J^*, x^*)$ is a limit point of $\{(X_1^k, \ldots, X_J^k, x^k)\}_{k\in\mathbb{N}}$, there exists a subsequence $\{(X_1^k, \ldots, X_J^k, x^k)\}_{k\in\mathcal{K}}$ such that $\{(X_1^k, \ldots, X_J^k, x^k)\}_{k\in\mathcal{K}} \to (X_1^*, \ldots, X_J^*, x^*)$.

Since $X_j^{k+1} = U_j^k \Sigma_j^{k+1} (V_j^k)^{\mathrm{T}}$ is the SVD of X_j^{k+1} and $\{X_j^{k+1}\}_{k\in\mathcal{K}} \to X_j^*$, we have

$$\left\{(\Sigma_j^{k+1})_{ii}\right\}_{k\in\mathcal{K}} \to \sigma_i(X_j^*), \quad \forall\ i = 1, 2, \ldots, \ell, \tag{22}$$

where $\ell = \min\{m, n\}$.

Let $r_j = \operatorname{rank}\left(X_j^*\right)$. Then it follows from (22) that there exists some $k_0 > 0$ such that $(\Sigma_j^{k+1})_{ii} > 0$ for all $1 \leqslant i \leqslant r_j$ and $k \in \mathcal{K}_0 = \{k \in \mathcal{K} : k > k_0\}$. This together with $\Sigma_j^{k+1} = S_{\lambda/c_j, w_j^k}(\Lambda_j^k)$ leads to

$$(\Sigma_j^{k+1})_{ii} = (\Lambda_j^k)_{ii} - \lambda_j (w_j^k)_i / c_j, \quad \forall\ 1 \leqslant i \leqslant r_j, \quad k \in \mathcal{K}_0.$$

Since $X_j^{k+1} = U_j^k \Sigma_j^{k+1} (V_j^k)^{\mathrm{T}}$ and $Z_j^k = U_j^k \Lambda_j^k (V_j^k)^{\mathrm{T}}$, then for all $k \in \mathcal{K}_0$

$$\begin{aligned}&X_j^{k+1} - \sum_{i=r_j+1}^{\ell} (\Sigma_j^{k+1})_{ii} (U_j^k)_i (V_j^k)_i^{\mathrm{T}}\\ &\quad = Z_j^k - \sum_{i=r_j+1}^{\ell} (\Lambda_j^k)_{ii} (U_j^k)_i (V_j^k)_i^{\mathrm{T}} - \frac{\lambda_j}{c_j}\sum_{i=1}^{r_j} (w_j^k)_i (U_j^k)_i (V_j^k)_i^{\mathrm{T}}.\end{aligned}$$

Using $Z_j^k = X_j^k - \frac{1}{c_j}\left(X_j^k - R_j(x^k) + \mu_j \nabla f_j(X_j^k)\right)$, we obtain

$$\begin{aligned}&X_j^{k+1} - X_j^k + \frac{1}{c_j}\left(X_j^k - R_j(x^k) + \mu_j\nabla f_j(X_j^k)\right) - \sum_{i=r_j+1}^{\ell}\left((\Sigma_j^{k+1})_{ii} - (\Lambda_j^k)_{ii}\right)(U_j^k)_i (V_j^k)_i^{\mathrm{T}}\\ &\quad + \frac{\lambda_j}{c_j}\sum_{i=1}^{r_j} (w_j^k)_i (U_j^k)_i (V_j^k)_i^{\mathrm{T}} = 0.\end{aligned}\tag{23}$$

Let $\bar{U}_j^k = [(U_j^k)_1 \dots (U_j^k)_{r_j}]$ and $\bar{V}_j^k = [(V_j^k)_1 \dots (V_j^k)_{r_j}]$. Upon pre- and post-multiplying (23) by $(\bar{U}_j^k)^{\mathrm{T}}$ and $\bar{V}_j^k$, and using $(\bar{U}_j^k)^{\mathrm{T}}\bar{U}_j^k = (\bar{V}_j^k)^{\mathrm{T}}\bar{V}_j^k = I$, we can see that for all $k \in \mathcal{K}_0$,

$$c_j(\bar{U}_j^k)^{\mathrm{T}}(X_j^{k+1} - X_j^k)\bar{V}_j^k + \lambda_j Diag((w_j^k)_1, \dots, (w_j^k)_{r_j}) \\ + (\bar{U}_j^k)^{\mathrm{T}} \left(X_j^k - R_j(x^k) + \mu_j \nabla f_j(X_j^k)\right) \bar{V}_j^k = 0. \tag{24}$$

Notice that $\{\bar{U}_j^k\}_{k\in\mathcal{K}}$ and $\{\bar{V}_j^k\}_{k\in\mathcal{K}}$ are bounded. Considering a convergent subsequence if necessary, assume without loss of generality that $\{\bar{U}_j^k\}_{k\in\mathcal{K}} \to \bar{U}_j^*$ and $\{\bar{V}_j^k\}_{k\in\mathcal{K}} \to \bar{V}_j^*$. Since $\{X_j^{k+1}\}_{k\in\mathcal{K}} \to X_j^*$, we have $\{(w_j^k)_i\}_{k\in\mathcal{K}} \to (w_j^*)_i$, where $(w_j^*)_i = (\sigma_i(X_j^*) + \varepsilon)^{-1}$. Taking limits on both sides of (24) as $k \in \mathcal{K}_0 \to \infty$ and using $\{\|X_j^{k+1} - X_j^k\|_F\}_{k\in\mathcal{K}} \to 0$, $\{X_j^{k+1}\}_{k\in\mathcal{K}} \to X_j^*$, $\{x^k\}_{k\in\mathcal{K}} \to x^*$, we can get

$$\lambda_j Diag((\sigma(X_j^*) + \varepsilon)^{-1}) + (\bar{U}_j^*)^{\mathrm{T}} \left(X_j^* - R_j(x^*) + \mu_j \nabla f_j(X_j^*)\right) \bar{V}_j^* = 0. \tag{25}$$

Taking limits as $k \in \mathcal{K}_0 \to \infty$ to $(\bar{U}_j^k)^{\mathrm{T}}\bar{U}_j^k = (\bar{V}_j^k)^{\mathrm{T}}\bar{V}_j^k = I$, we have $(\bar{U}_j^*)^{\mathrm{T}}\bar{U}_j^* = (\bar{V}_j^*)^{\mathrm{T}}\bar{V}_j^* = I$. Using $X_j^{k+1} = U_j^k\Sigma_j^{k+1}(V_j^k)^{\mathrm{T}}$, (22), $r_j = \mathrm{rank}(X_j^*)$, $\{X_j^{k+1}\}_{k\in\mathcal{K}} \to X_j^*$, $\{\bar{U}_j^k\}_{k\in\mathcal{K}} \to \bar{U}_j^*$ and $\{\bar{V}_j^k\}_{k\in\mathcal{K}} \to \bar{V}_j^*$, one can get that $X_j^* = \bar{U}_j^* Diag(\sigma(X_j^*))(\bar{V}_j^*)^{\mathrm{T}}$. Hence, $(\bar{U}_j^*, \bar{V}_j^*) \in \bar{\mathcal{M}}(X_j^*)$. Using those relations and (25), we can conclude that (15) holds at $(X_1^*, \dots, X_J^*, x^*)$ with $U_j = \bar{U}_j^*$, $V_j = \bar{V}_j^*$.

On the other hand, due to $\{x^k\}_{k\in\mathcal{K}} \to x^*$, $\{X_j^{k+1}\}_{k\in\mathcal{K}} \to X_j^*$ and (19), we have that (14) holds. □

Note that the function $f_j(X_j)$ in Φ may not have a Kurdyka–Łojasiewicz property at $X_j = 0$ and hence the convergence analysis in [27, 28] is not applicable to algorithm 1.

5. Rician noise removal via the NLRR method

In this section, we propose the NLRR method with a two-stage method to efficiently remove Rician noise in MR images.

The proposed method for Rician noise removal has two stages as shown in algorithm 2. The first stage yields a preliminary estimation of the image by iteratively applying the ARM algorithm to remove most of the noise, and the second stage further refines the output of the first stage. The ARM algorithms in these two stages use the same noisy image as an input, but they have different initial images and stopping criteria.

In the first stage, the iterative scheme updating the initial image helps boost the denoising performance because the patch/cube matrix extraction operator R_j becomes more and more accurate in collecting patches/cubes with similar textures as the initial image is updated at each iteration. And the stopping criterion of the ARM algorithm based on the standard deviation of Rician noise in the denoised image guarantees that the noise is reduced gradually and that the ARM algorithm is efficiently terminated. In the second stage, the quality of the denoised image is further improved using the estimated image in the first stage as the initial image and the final estimated image is achieved in a reliable manner using the relative error in the stopping criterion of the ARM algorithm.

Algorithm 2. The NLRR method for Rician noise removal.

Input: Noisy image y
Initialize: $0 < \rho < 1$, $\delta_{\text{tol}} > 0$, $\epsilon_{\text{tol}} > 0$,
$x_0 = y$, $\delta_0 = \delta$, $t = 0$
- *Stage* 1*:*
  repeat
  $x_{t+1} = ARM(y, x_t)$
  with stopping criterion: $\delta_{\text{est}}(x_{t+1}) < \rho \cdot \delta_t$
  $\delta_{t+1} = \delta_{\text{est}}(x_{t+1})$
  $t \leftarrow t + 1$
  until $\delta_{t+1} < \delta_{\text{tol}}$
- *Stage* 2*:*
  $\hat{x} = \text{ARM}(y, x_{t+1})$
  with stopping criterion: relative error $<\epsilon_{\text{tol}}$

Output: $\hat{x}$

6. Numerical experiments

In this section, we perform numerical experiments on 2D medical images, and 3D volumetric medical data to demonstrate the superiority of our proposed NLRR method in removing Rician noise. All the experiments were performed in Matlab R2020a running a 64 bit Windows 10 system and executed on an eight-core Intel Core i9-9900K 64 GB CPU at 3.60 GHz with one NVIDIA GeForce GTX 1660 Ti 32 GB GPU.

All of the denoising methods were tested on both simulated and real data. The simulated data, including T1 weighted (T1w), T2 weighted (T2w) and proton density weighted (PDw) images, is degraded by Rician noise with noise level ranging from 1% to 19%, where the noise level denotes the percentage of noise standard deviation δ relative to the maximum value of the original data. And the quality of restored data were evaluated using the peak-signal-to-noise ratio (PSNR) [29] and the structural similarity index measure (SSIM) [12, 30]. PSNR is based on the root mean squared error (RMSE) and is defined as

$$\text{PSNR} = 20 \log_{10} \left(\frac{\text{MAX}_I}{\text{RMSE}} \right),$$

where MAX_I is the maximum possible pixel value. SSIM is a measure more consistent with the human visual system, and is defined as in [30] for 2D images and defined as in [12] for 3D data. Both PSNR and SSIM values were computed only in the region of interest (tissues) obtained with the background removed. The larger values of PSNR and SSIM, the better quality of the restored data. The real test data, including T1w and diffusion tensor imaging (DTI) data, is naturally contaminated by noise and the noise level is approximately 1.5%–4% according to the noise estimation [13] on the region of interest. And the residual image is used to evaluate the quality of the restored T1w images, which measures pointwise difference between the noisy and estimated data; the fractional anistropy (FA) map and color FA map computed via DSI Studio[6] are used to assess the DTI data. In particular, FA measures the directionality of molecular displacement by diffusion and varies between 0 (isotropic diffusion) and 1 (anisotropic diffusion); color FA is color-coded by the orientation of the principal eigenvector of the diffusion tensor.

[6] http://dsi-studio.labsolver.org/.

Table 1. Parameter settings.

Data	Parameters	Noise level (%)					
		$\leqslant 1$	$\leqslant 3$	$\leqslant 7$	$\leqslant 11$	$\leqslant 15$	>15
2D	Patch size	6	6	6	7	8	9
	Group size (n)	45	50	50	80	90	120
	$\lambda_j/(\delta^2\sqrt{n})$	7	6.4	5.8	5.8	5.8	5.4
3D	Cube size	4	4	4	5	5	6
	Group size (n)	60	60	60	100	100	170
	$\lambda_j/(\delta^2\sqrt{n})$	15.2	13.6	10.4	11.6	11.2	10.8

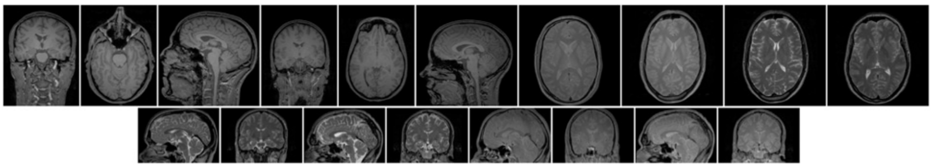


Figure 1. Original images.

Next, we give the parameter settings of the proposed method. In stage 1 of algorithm 2, we use the noise estimation approach introduced in [13] as $\delta_{\mathrm{est}}(\cdot)$ to estimate the standard deviation δ of Rician noise in the estimated image, and set $\delta_{\mathrm{tol}} = 1$ and $\rho = 0.2$; in stage 2, we set the stopping condition based on the following relative improvement inequality:

$$\frac{\|x^{k+1} - x^k\|_2}{\|x^k\|_2} < \epsilon_{\mathrm{tol}},$$

where ϵ_{tol} (e.g., 10^{-3}) is the error tolerance. Moreover, the parameter settings of the ARM algorithm in these two stages are the same. We set $c_j = 1.01 \cdot (1 + \frac{\mu_j}{\delta^2})$, $\varepsilon = 10^{-16}$ and

$$\mu_j = \frac{\delta^2}{0.003\,2 \cdot \mathrm{mean}(R_j(x^0)) + 1.8}.$$

And as shown in table 1, the settings for block/cube matching and parameter λ_j are chosen dependent upon the noise level.

6.1. Denoising of 2D MR images

To demonstrate the superior performance of the proposed method in removing Rician noise in 2D MR images, we compare the proposed NLRR method with the GTV [5], NLM2D [11], C-KSVD [7], and VST-BM3D [13, 14] methods. And we test all the denoising methods on 18 simulated 2D T1w, T2w and PDw images generated from selected slices of IXI database[7], see figure 1, and we test on real 2D T1w image 'OAS1_0092' (256×256) from a selected slice of open access series of imaging studies database [31], see figure 5(a).

[7] http://brain-development.org/ixi-dataset/.

Table 2. Average PSNR and SSIM values tested on 2D data by different denoising methods.

Noise	GTV		NLM2D		C-KSVD		VST-BM3D		NLRR (ours)	
level (%)	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	40.82	0.977	38.21	0.973	15.96	0.285	42.55	0.986	**42.81**	**0.987**
3	33.33	0.893	33.87	0.914	32.73	0.931	35.75	0.941	**36.06**	**0.944**
5	30.10	0.807	30.58	0.839	30.16	0.861	32.77	0.893	**33.03**	**0.899**
7	29.02	0.775	28.19	0.765	27.43	0.772	30.96	0.848	**31.32**	**0.855**
9	27.33	0.705	26.28	0.693	25.46	0.686	29.60	0.805	**29.94**	**0.814**
11	26.65	0.680	24.76	0.626	24.10	0.615	28.53	0.764	**28.87**	**0.770**
13	25.40	0.626	23.40	0.563	23.18	0.566	27.63	0.724	**27.93**	**0.732**
15	24.53	0.594	22.13	0.501	22.56	0.533	26.81	0.686	**27.22**	**0.694**
17	22.49	0.480	21.01	0.441	22.20	0.510	26.10	0.649	**26.50**	**0.652**
19	21.91	0.460	19.97	0.381	22.02	0.496	25.40	0.618	**25.90**	**0.629**

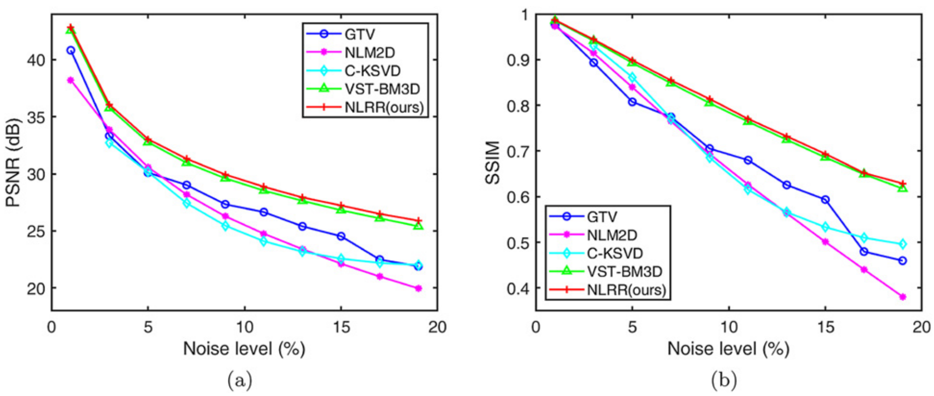


Figure 2. Average numerical results tested on 2D data. (a) The average PSNR (dB) values; (b) the average SSIM values.

Table 2 presents the average PSNR and SSIM values of the restored images tested on simulated data and figure 2 plots the average PSNR and SSIM values versus the noise level (%). In terms of PSNR values, the proposed NLRR method outperforms other methods for every test image at every noise level. In particular, at noise level 19%, the average PSNR value of the image obtained by the proposed method is 5.9 dB larger than the NLM2D method, 3.8 dB larger than the GTV and C-KSVD method, and 0.5 dB larger than the benchmark VST-BM3D method.

Figures 3 and 4 present the restored images with zoomed-in views tested on simulated data at noise level 5%. And figure 5 presents the restored images and the residual image tested on real data at noise level 4%. In terms of visual quality, the restored images of the proposed NLRR method have less noise and fewer artifacts and retain more details than other methods. For example, the GTV method in figures 3(c), 4(c) and 5(b) has some noise and artifacts. The NLM2D method in figures 3(d), 4(d) and 5(c) and the C-KSVD method in figures 3(e), 4(e) and 5(d) are too smooth after denoising. As for the loss of details, the proposed method has a better trade-off than the other four methods between retaining fine details while removing noise. For example, the proposed method preserves more neck textures in the left lower box

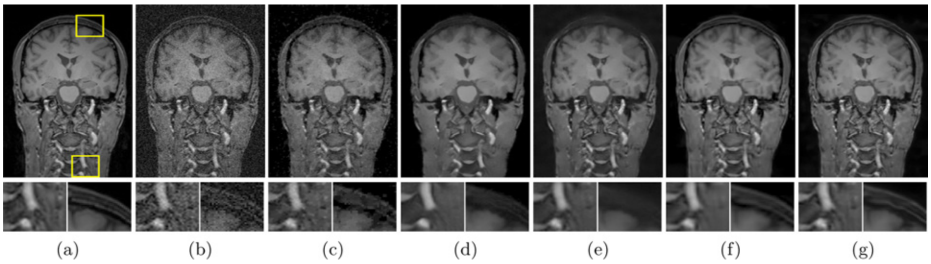


Figure 3. Denoised results tested on 'Brain 1' at noise level 5%. From top to bottom, the images are the original and zoomed-in views of denoised images. From left to right, the denoising methods with (PSNR, SSIM) values are (a) ground truth; (b) noisy image; (c) GTV (29.87 dB, 0.819); (d) NLM2D (30.18 dB, 0.836); (e) C-KSVD (30.19 dB, 0.860); (f) VST-BM3D (31.93 dB, 0.888); (g) NLRR (ours) (32.23 dB, 0.898).

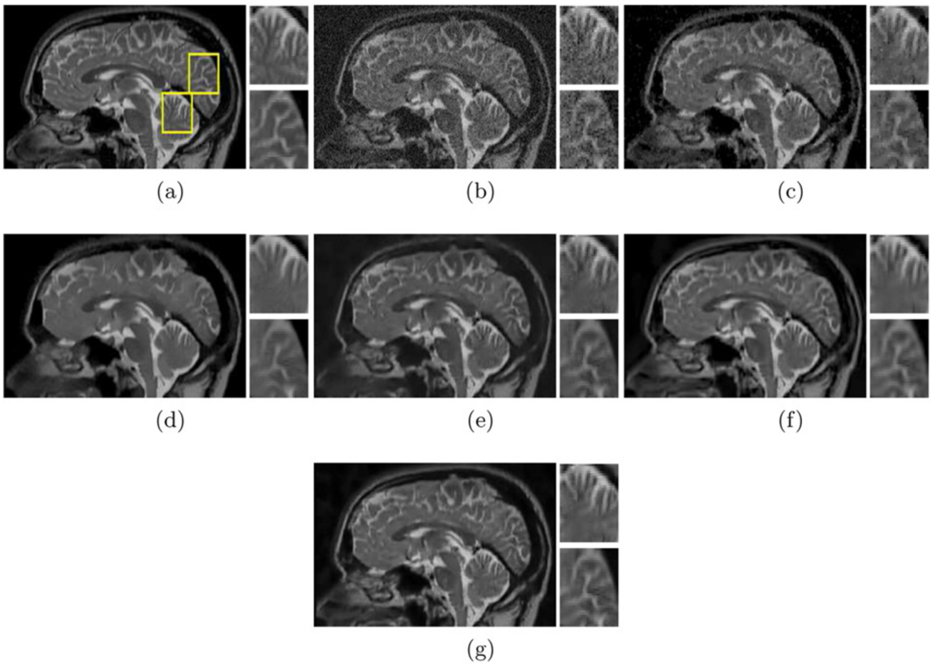


Figure 4. Denoised results tested on 'Brain 2' at noise level 5%. From left to right, the images are the original and zoomed-in views of denoised images. From left upper to right lower, the denoising methods with (PSNR, SSIM) values are (a) ground truth; (b) noisy image; (c) GTV (29.56 dB, 0.839); (d) NLM2D (29.53 dB, 0.834); (e) C-KSVD (30.01 dB, 0.864); (f) VST-BM3D (31.84 dB, 0.896); (g) NLRR (ours) (32.10 dB, 0.904).

of figure 3(g) and more brain details in the right upper box of figure 4(g) than the VST-BM3D method in figures 3(f) and 4(f), respectively. Also, the residual images in figure 5 show that the noise removed by the proposed method is relatively random with less geometric structures than other methods.

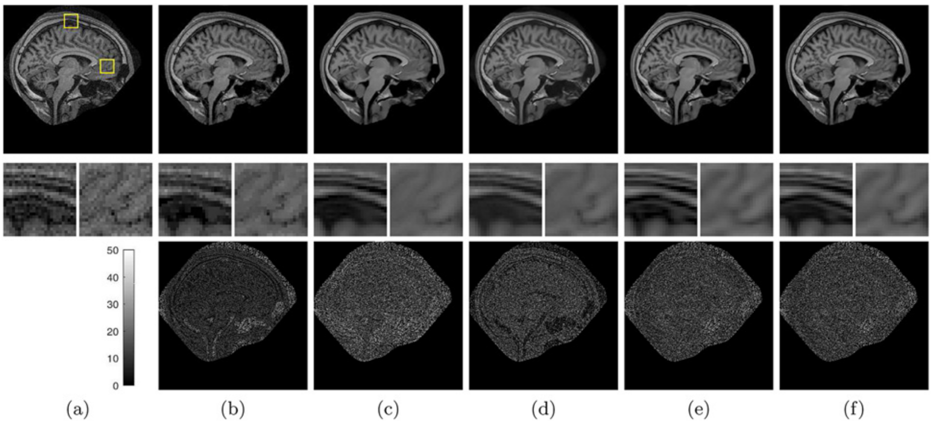


Figure 5. Denoised results tested on 'OAS1_0092' at approximated noise level 4.5%. From top to bottom, the images are the original, zoomed-in views, and residual images of denoised images. From left to right, the denoising methods are (a) noisy image; (b) GTV; (c) NLM2D; (d) C-KSVD; (e) VST-BM3D; (f) NLRR (ours).

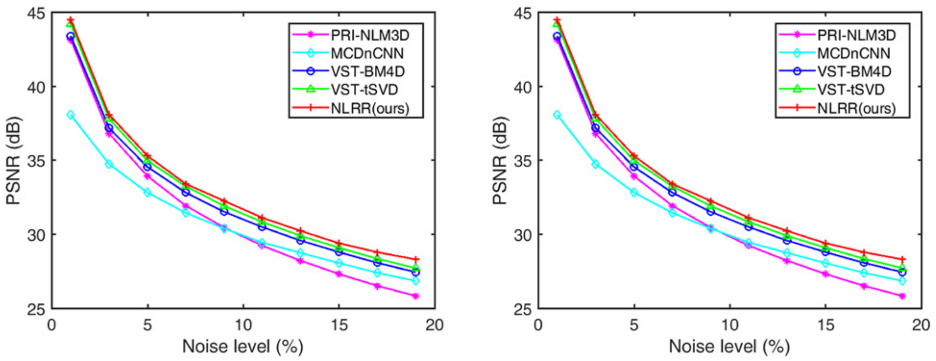
Table 3. Average PSNR and SSIM values tested on 3D data by different denoising methods.

Noise	PRI-NLM3D		MCDnCNN		VST-BM4D		VST-tSVD		NLRR (ours)	
level (%)	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
1	43.17	0.995	38.09	0.989	43.43	0.995	44.23	**0.996**	**44.52**	**0.996**
3	36.83	0.983	34.77	0.974	37.21	0.983	37.84	0.985	**38.11**	**0.986**
5	33.96	0.968	32.85	0.960	34.58	0.971	35.01	0.972	**35.33**	**0.974**
7	31.96	0.951	31.48	0.946	32.84	0.958	33.27	**0.961**	**33.43**	**0.961**
9	30.49	0.932	30.42	0.932	31.55	0.946	31.94	0.948	**32.28**	**0.951**
11	29.26	0.911	29.47	0.918	30.51	0.932	30.86	0.935	**31.14**	**0.937**
13	28.24	0.889	28.77	0.903	29.60	0.917	29.92	0.920	**30.25**	**0.924**
15	27.34	0.868	28.08	0.888	28.82	0.903	29.12	0.906	**29.42**	**0.909**
17	26.54	0.846	27.43	0.871	28.10	0.887	28.38	0.892	**28.82**	**0.902**
19	25.87	0.826	26.88	0.855	27.47	0.872	27.73	0.876	**28.33**	**0.892**

In summary, the proposed NLRR method achieves the best quantitative assessments among all the methods for denoising 2D MR images and provides smoother areas, better shape retention, and better detail preservation than other methods in the denoised images. Other competing methods suffer from some limitations. The GTV method uses the TV regularization, which may cause staircase artifacts; the C-KSVD and NLM2D method require a bias correction to estimate the mean value of the clean image, which may cause issues with image contrast; and the VST-BM3D method requires forward and inverse VSTs, which may introduce unexpected errors.

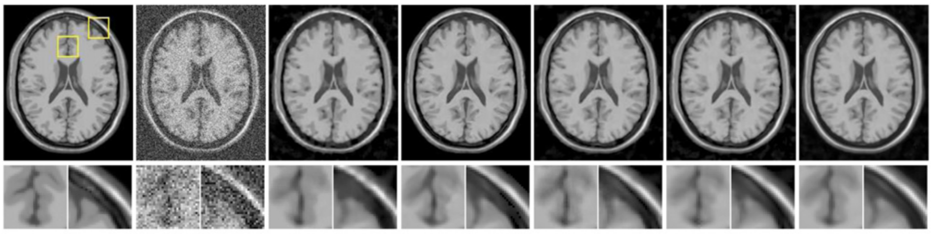
6.2. Denoising of 3D MR data

To show the great performance of the proposed NLRR method in 3D MR data denoising, we compare the proposed method with the PRI-NLM3D [12], VST-BM4D [15], VST-tSVD [17] and MCDnCNN [8] methods. The MCDnCNN method is trained with unknown noise level on



(a) (b)

Figure 6. Average numerical results tested on 3D data. (a) The average PSNR (dB) values; (b) the average SSIM values.



(a) (b) (c) (d) (e) (f) (g)

Figure 7. Denoised results tested on T1w data at noise level 13%. From top to bottom, the images are the original and zoomed-in views of selected slides of denoised images. From left to right, the denoising methods with (PSNR, SSIM) values are (a) ground truth; (b) noisy image; (c) PRI-NLM3D (29.76 dB, 0.893); (d) MCDnCNN (30.34 dB, 0.909); (e) VST-BM4D (30.89 dB, 0.916); (f) VST-tSVD (31.09 dB, 0.917); (g) NLRR (ours) (31.49 dB, 0.923).

healthy subjects randomly selected from IXI database, including T1w, T2w and PDw images. Ten images for each type of MR images were used for training. The simulated 3D MR data are T1w, T2w, and PDw volumes from BrainWeb phantoms [32] of size $181 \times 217 \times 181$ with $1 \times 1 \times 1$ mm^3 resolution. The real 3D MR data includes T1w volume from 'OAS1_0105' of size $256 \times 256 \times 128$ with $1 \times 1 \times 1.25$ mm^3 resolution and DTI data[8] (30 directions) of size $112 \times 112 \times 50$ with $2 \times 2 \times 2$ mm^3.

Table 3 presents the average PSNR and SSIM values of the restored volumetric data tested on simulated data, and figure 6 plots the average PSNR values versus the noise level (%). It is shown that the proposed NLRR method achieves the best average PSNR and SSIM values. For example, at noise level 19%, the image obtained by the proposed method has 2.5 dB larger average PSNR value and 0.07 larger average SSIM value than the PRI-NLM3D methods.

Figures 7 and 8 present with zoomed-in views the selected slides of the MR data tested on simulated data at noise level 13%. Figure 9 presents the denoised MR data and its residual

[8] http://cmic.cs.ucl.ac.uk/camino//uploads/Tutorials/example_dwi.zip.

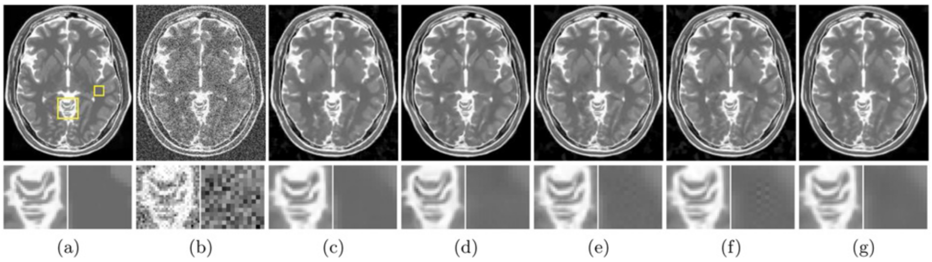


Figure 8. Denoised results tested on T2w data at noise level 13%. From top to bottom, the images are the original and zoomed-in views of selected slides of denoised images. From left to right, the denoising methods with (PSNR, SSIM) values are (a) ground truth; (b) noisy image; (c) PRI-NLM3D (26.62 dB, 0.909); (d) MCDnCNN (27.02 dB, 0.921); (e) VST-BM4D (27.85 dB, 0.929); (f) VST-tSVD (28.33 dB, 0.935); (g) NLRR (ours) (28.75 dB, 0.940).

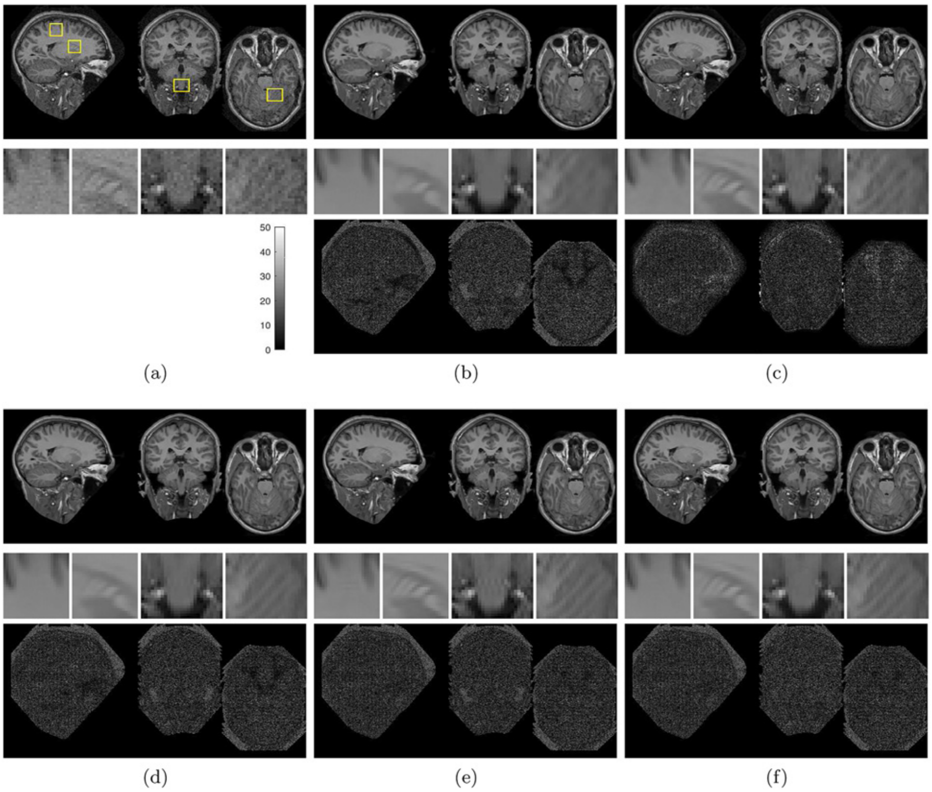


Figure 9. Denoised results tested on 'OAS1_0105' at approximated noise level 3%. From top to bottom, the images are the original, zoomed-in views, and residual images of denoised images. The denoising methods are (a) noisy image; (b) PRI-NLM3D; (c) MCDnCNN; (d) VST-BM4D; (e) VST-tSVD; (f) NLRR (ours).

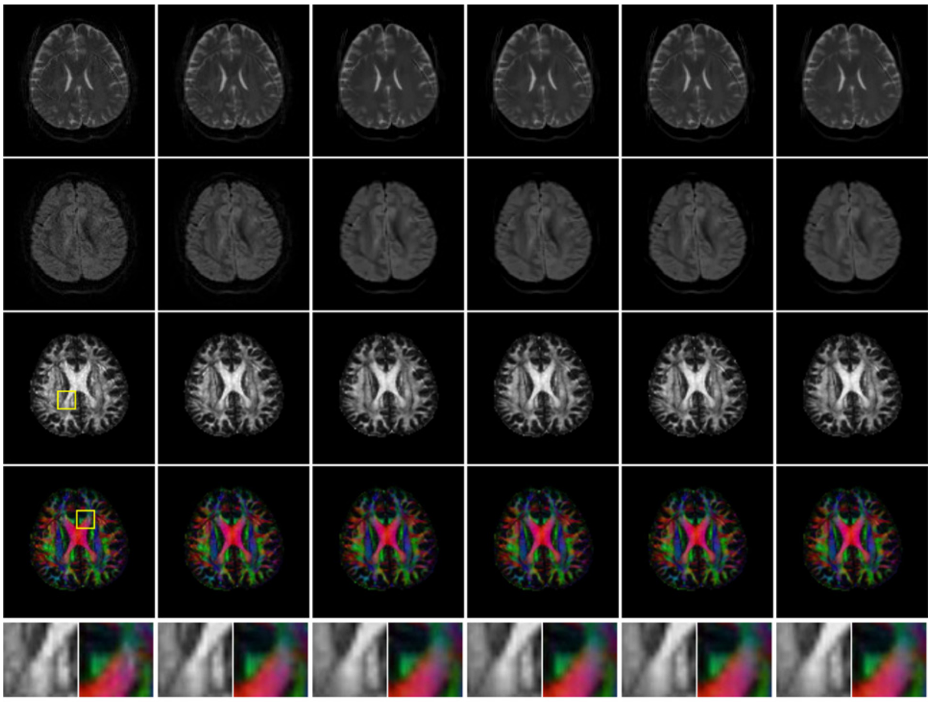


Figure 10. Denoised results tested on DTI data at approximated noise level between 1.5% and 2.5%. From top to bottom, the images are the noisy/denoised image ($b = 0$), the noisy/denoised image ($b = 1000$), FA map, color FA map and zoomed-in views. From left to right, the denoising methods are (a) noisy image; (b) PRI-NLM3D; (c) MCDnCNN; (d) VST-BM4D; (e) VST-tSVD; (f) NLRR (ours).

images tested on real T1w data. The restored MR data of the proposed NLRR method generate smoother tissue textures and fewer artifacts, and preserve more fine details as shown in figures 7, 8(g) and 9(f). On the contrary, the PRI-NLM3D method tends to over-smooth the edges of brain layers in figure 7(c) and remove some fine details of brain tissues in figure 9(b). And due to the limitation of VSTs, the stated-of-the-art 3D denoising methods such as the VST-BM4D and VST-tSVD methods suffer from unexpected artifacts. For example, there are mosaic patterns in the second zoomed-in view of figures 8(e) and (f) and the third zoomed-in view of figures 9(d) and (e). Also, the MCDnCNN method tends to remove more noise in brighter regions and less noise in darker regions as shown in its residual image in figure 9(c). Moreover, the denoising results in figure 10 tested on real DTI data also demonstrate the great capability of the proposed method in removing noise. The FA map and color FA map derived from the proposed method in figure 10(f) have fewer artifacts and more natural transition.

In summary, quantitative results and denoised images demonstrate the superiority of the proposed method in denoising 3D MR data. The MCDnCNN method relies on the training dataset and its performance seems not stable. And NSS-based methods with flexible bases, like the VST-tSVD and the proposed method, outperform the NSS-based methods with fixed bases, like the PRI-NLM3D and VST-BM4D method. Also, the proposed method introduces fewer artifacts than the methods required VSTs.

7. Conclusion

In this paper, we propose a new optimization model that employs a NLRR and a MAP-based fidelity term for denoising of MR images with Rician noise. Specifically, we utilize the log-det function to regularize the low-rank prior of nonlocal similar patch/cube matrices and adopt the data fidelity term of the MAP model to reflect the statistics of Rician noise. Using the facts that the logarithmic function in the log-det function is concave and that the MAP-based fidelity term has a Lipschitz continuous gradient, we develop the ARM algorithm to solve the proposed model and prove that any limit point of the sequence generated by the ARM algorithm is a first-order critical point of the objective function. Then a two-stage method enhances the ARM algorithm for removing Rician noise. Furthermore, numerical experiments on 2D and 3D denoising demonstrate the effectiveness of the proposed NLRR method for removing Rician noise and the superiority over the state-of-the-art VST-based and DnCNN-based methods in terms of average PSNR and SSIM values and of the visual quality of restored images.

Acknowledgments

The authors would like to thank the authors of [7, 8, 11–15, 17] for sharing the source codes. Also, the authors would like to thank the anonymous reviewers for providing us valuable comments and suggestions which lead to improvements of our article. This work was supported by the National Natural Science Foundation of China (NSFC) under Grants 61972265, 11871348, 61872429 and 11701383, by the Natural Science Foundation of Guangdong Province of China under Grant 2020B1515310008, by the Educational Commission of Guangdong Province of China under Grant 2019KZDZX1007, and by the Shenzhen Key Laboratory of Media Security. Q Jiang was supported by the Simons Foundation under Grant 353185. Z Lin was supported by the NSFC under Grant 61731018 and by the NSFC Tianyuan Fund for Mathematics under Grant 12026606.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Jian Lu https://orcid.org/0000-0003-4599-7281
Xiaoxia Liu https://orcid.org/0000-0002-2289-7520

References

[1] Wright G A 1997 Magnetic resonance imaging *IEEE Signal Process. Mag.* **14** 56–66
[2] Guido G, Kubler O, Kikinis R and Jolesz F A 1992 Nonlinear anisotropic filtering of MRI data *IEEE Trans. Med. Imag.* **11** 221–32
[3] Krissian K and Aja-Fernández S 2009 Noise-driven anisotropic diffusion filtering of MRI *IEEE Trans. Image Process.* **18** 2265–74
[4] Nowak R D 1999 Wavelet-based Rician noise removal for magnetic resonance imaging *IEEE Trans. Image Process.* **8** 1408–19
[5] Getreuer P, Tong M and Vese L A 2011 A variational model for the restoration of MR images corrupted by blur and Rician noise *Advances in Visual Computing* vol 6938 (Berlin: Springer) pp 686–98

[6] Chen L and Zeng T 2015 A convex variational model for restoring blurred images with large Rician noise *J. Math. Imaging Vis.* **53** 92–111
[7] Chen W, You J, Chen B, Pan B, Li L, Pomeroy M and Liang Z 2018 A sparse representation and dictionary learning based algorithm for image restoration in the presence of Rician noise *Neurocomputing* **286** 130–40
[8] Jiang D, Dou W, Vosters L, Xu X, Sun Y and Tan T 2018 Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network *Japan. J. Radiol.* **36** 566–74
[9] Zhang K, Zuo W, Chen Y, Meng D and Zhang L 2017 Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising *IEEE Trans. Image Process.* **26** 3142–55
[10] Buades A, Coll B and Morel J-M 2005 A non-local algorithm for image denoising *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition* vol 2 (Piscataway, NJ: IEEE) pp 60–5
[11] Manjon J, Carbonellcaballero J, Lull J, Garciamarti G, Martibonmati L and Robles M 2008 MRI denoising using non-local means *Med. Image Anal.* **12** 514–23
[12] Manjón J V, Coupé P, Buades A, Louis Collins D and Robles M 2012 New methods for MRI denoising based on sparseness and self-similarity *Med. Image Anal.* **16** 18–27
[13] Foi A 2011 Noise estimation and removal in MR imaging: the variance-stabilization approach *IEEE Int. Symp. Biomedical Imaging: From Nano to Macro* (Piscataway, NJ: IEEE) pp 1809–14
[14] Dabov K, Foi A, Katkovnik V and Egiazarian K 2007 Image denoising by sparse 3D transform-domain collaborative filtering *IEEE Trans. Image Process.* **16** 2080–95
[15] Maggioni M, Katkovnik V, Egiazarian K and Foi A 2013 Nonlocal transform-domain filter for volumetric data denoising and reconstruction *IEEE Trans. Image Process.* **22** 119–33
[16] Zhang X, Xu Z, Jia N, Yang W, Feng Q, Chen W and Feng Y 2015 Denoising of 3D magnetic resonance images by using higher-order singular value decomposition *Med. Image Anal.* **19** 75–86
[17] Kong Z, Han L, Liu X and Yang X 2018 A new 4D nonlocal transform-domain filter for 3D magnetic resonance images denoising *IEEE Trans. Med. Imaging* **37** 941–54
[18] Dong W, Shi G, Li X, Ma Y and Huang F 2014 Compressive sensing via nonlocal low-rank regularization *IEEE Trans. Image Process.* **23** 3618–32
[19] Liu X, Lu J, Shen L, Xu C and Xu Y 2020 Multiplicative noise removal: nonlocal low-rank model and its proximal alternating reweighted minimization algorithm *SIAM J. Imaging Sci.* **13** 1595–629
[20] Huang T, Dong W, Xie X, Shi G and Bai X 2017 Mixed noise removal via Laplacian scale mixture modeling and nonlocal low-rank approximation *IEEE Trans. Image Process.* **26** 3171–86
[21] Bowman F 1958 *Introduction to Bessel Functions* (New York: Dover)
[22] Lu J, Tian J, Jiang Q, Liu X, Hu Z and Zou Y 2021 Rician noise removal via weighted nuclear norm penalization *Appl. Comput. Harmon. Anal.* **53** 180–98
[23] Lu Z, Zhang Y and Lu J 2017 ℓ_p regularized low-rank approximation via iterative reweighted singular value minimization *Comput. Optim. Appl.* **68** 619–42
[24] Lewis A S and Sendov H S 2005 Nonsmooth analysis of singular values: I. Theory *Set-Valued Anal.* **13** 213–41
[25] Bertsekas D P 1999 *Nonlinear Programming* (Belmont, MA: Athena Scientific)
[26] Gu S, Xie Q, Meng D, Zuo W, Feng X and Zhang L 2017 Weighted nuclear norm minimization and its applications to low level vision *Int. J. Comput. Vis.* **121** 183–208
[27] Bolte J, Sabach S and Teboulle M 2014 Proximal alternating linearized minimization for nonconvex and nonsmooth problems *Math. Program.* **146** 459–94
[28] Sun T, Jiang H and Cheng L 2017 Convergence of proximal iteratively reweighted nuclear norm algorithm for image processing *IEEE Trans. Image Process.* **26** 5632–44
[29] Gibson J D and Bovik A C 2000 *Handbook of Image and Video Processing* (New York: Academic)
[30] Wang Z, Bovik A C, Sheikh H R and Simoncelli E P 2004 Image quality assessment: from error visibility to structural similarity *IEEE Trans. Image Process.* **13** 600–12
[31] Marcus D S, Wang T H, Parker J, Csernansky J G, Morris J C and Buckner R L 2007 Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults *J. Cogn. Neurosci.* **19** 1498–507
[32] Cocosco C A, Kollokian V, Kwan R K-S and Evans A C 1997 Brainweb: online interface to a 3D MRI simulated brain database *NeuroImage* **5** S425