VARIATIONAL RICIAN NOISE REMOVAL VIA SPLITTING ON SPHERES

ZHIFANG LIU*, HUIBIN CHANG†, AND YUPING DUAN‡

Abstract. We propose a novel variational method to restore magnitude images corrupted by Rician noises in magnetic resonance (MR) imaging via signal-noise splitting. By the link between the Gaussian noise removal of complex MR images and the Rician noise removal of magnitude MR images, the proposed nonlinear optimization model consists of a total variation regularizer, two quadratic terms, and a constraint on the field of spheres. Specifically, this constraint represents the forward model of calculating the magnitude MR images from complex MR images degraded by Gaussian noises. Namely, the proposed model is completely different from the existing variational methods, which were usually derived by maximum a posterior of Rician distribution such that they inevitably involved the Bessel function causing high computation cost. We further adopt the alternating direction method of multipliers for solving the proposed model efficiently and briefly analyze its convergence. Numerical comparisons with existing variational methods show the proposed method produces comparable results in terms of image quality but saves about 50% of overall computational cost on average.

Key words. MR image denoising, signal-noise splitting, the field of spheres, Rician noise, total variation, alternating direction method of multipliers

1. Introduction.

Magnitude magnetic resonance (MR) images are widely used in medical image processing. Compared with the complex-valued MR images, such images discard the phase information, thus avoiding phase artifacts [20]. The MR images are reconstructed from the MR scan data with inevitable measurable noises. One of the major sources of these noises is the thermal noise caused by patients during the MR scan [29, 2]. Consequently, the MR images are always noisy. Significantly, the noises in the magnitude MR images negatively affect different medical image processing and analysis tasks, such as visualization, segmentation, registration, classification, and diffusion tensor estimation [2]. Therefore, it is a fundamental problem to remove noises in magnitude MR images to obtain high-quality ones.

The noisy magnitude MR data is commonly modeled by the Rician distribution [20, 2]. Thus, we usually refer to estimating clean magnitude MR image from a noisy one as Rician noise removal. Because of the signal dependence of the Rician noise, it is a great challenge to extract the clean magnitude MR image directly. Next, we present some previous works devoted to address this challenge problem. We first review some statistical-based noise-removal methods. In [20], Henkelman used the first moment of the Rician distribution to estimate the MR signal and proposed a lookup table to correct the Rician basis. Furthermore, the second moment of the Rician distribution was employed in [5, 28]. In [32], Sijbers and den Dekker implemented the maximum likelihood method for the estimation of MR signal amplitude. In [1], Aja-Fernández *et al.* derived the linear minimum mean square error (LMMSE) estimator based on the local sample statistics. Other statistical-based methods, such as the non-local means (NLM) method and the variance-stabilizing transformation (VST) method, can be found in [24, 37, 14]. Recently, learning-based methods have become popular in image processing and have been applied to Rician noise reduction. In [25], Manjón

*School of Mathematical Sciences, Tianjin Normal University, Tianjin 300387, China (matlzhf@tjnu.edu.cn).

†Corresponding author. School of Mathematical Sciences, Tianjin Normal University, Tianjin 300387, China (changhuibin@gmail.com).

‡Center for Applied Mathematics, Tianjin University, Tianjin, China (yuping.duan@tju.edu.cn).

and Coupe presented a two-stage strategy method that combines convolutional neural network and non-local means filter and can automatically deal with both stationary and spatially varying noises. In [40], You *et al.* studied the wider denoising neural network for denoising of MR images with Rician noise. In [39], Yang *et al.* proposed a noise adaptive trainable non-linear reaction-diffusion method for Rician noise removal that is robust against noise level changes. In [22], by fitting the distribution of pixel-level and feature-level intensities, Li *et al.* developed a RicianNet for MR image denoising.

Besides the abovementioned methods, the variational regularization method is also an important mathematical tool for Rician noise removal. The variational method is interpretable and stable for various image processing tasks. Thus, it has attracted much research during the last several decades. The variational model usually contains two terms: regularization term and data fidelity term. One of the most common regularizations is the total variation (TV), which is firstly proposed for removing Gaussian noise in [31]. Due to its edge preservation properties, TV regularization has also been introduced for Rician noise removal. Based on maximum a posterior (MAP) estimates, Getreuer *et al.* [16] and Martin [26] independently proposed the variational MAP model using TV regularization with the Rician likelihood fidelity term. The existence theory of the MAP model was roughly mentioned by Getreuer *et al.* in [16] and further rigorously analysed by Martín *et al.* in [27]. Getreuer *et al.* [16] solved the MAP model with the ℓ^2 and Sobolev H^1 gradient descents. However, those algorithms converged slowly and could get stuck in a local minimum without proper initialization and numerical discretization. Therefore, Getreuer *et al.* [16] replaced the fidelity term in the MAP model with one of its convex approximations and then solved the convexified model with the split Bregman method. In [10], Chen and Zeng added a quadratic term based on the statistics property of Rician distribution to the MAP model and then obtained a strictly convex model, solved with a primal-dual algorithm. One can find some other variants based on the MAP model in [41, 23].

The existing variational methods were mainly derived from the MAP of the Rician noise [16] for the magnitude MR image, and they involved the sophisticated Bessel function causing relatively high computational complexities, compared with the purely Gaussian noise denoising. This paper will explore an alternative method based on the link of the Gaussian denoising of complex images and Rician noise removal of magnitude MR images. To this end, we first examine the original variational model of the complex image with complex Gaussian noise, consisting of TV regularization for the phase and magnitude separately, a ℓ^2 fitting term, and a constraint for the forward noise model. Separating the magnitude and phase of the complex image leads to an equivalent form. Then, a variational model to optimize the magnitude of the complex image is obtained by dropping off the regularization term for the phase part of the complex image and taking the absolute part of the constraint for the complex variable. Further combined with a quadratic correction term, one finally deduces a new variational model, essentially a non-convex optimization model with the constraint on the field of spheres. Due to the closed-form projection onto spheres, an efficient alternating direction method of multipliers (ADMM) can be applied to the proposed model. A rigorous convergence guarantee could be further provided following the convergence theory for the non-convex ADMM [33].

The main contributions of this paper are listed below:

- By exploring the link between the complex denoising model and the Rician noise model of magnitude MR image, we propose a novel variational model by directly seeking a piecewise-constant solution from a field of spheres, instead of using Bessel functions. Namely, completely different with existing variational models coping with sophisticated Bessel functions, the proposed model consists of the standard TV norm of the underlying image, two quadratic terms and a constraint on a field of spheres, such that one can solve it as efficiently as standard Gaussian denoising.

- Based on the closed-form expression of the projection onto spheres, we develop an efficient ADMM algorithm. We further eliminate the redundant update for the multiplier to reduce the computation cost. The corresponding convergence to the stationary point of the proposed model are proved with a sufficiently large stepsize.
- We conduct numerous experiments to evaluate the performance of the proposed method. Numerically, the proposed method can produce comparable quality results in Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), among all the compared variational methods. Namely, it reduces about 50% of overall computational cost on average compared to the main existing variational methods, since the cost per iteration dramatically decreases without calculating the sophisticated Bessel functions.

The rest of this paper is organized as follows. In Section 2, we give some preliminaries and review some related works. In Section 3, we present the new Rician noise removal model. In Section 4, we present an efficient iterative algorithm for solving the proposed model. The numerical experiments for both synthetic and real MR images are shown in Section 5. We conclude the paper in Section 6.

2. Preliminary and related works.

2.1. Preliminary. The raw MR data is measured through a quadrature detector that generates two- or three-dimensional complex arrays in k-space and always corrupted by Gaussian-distributed noise. After the inverse Fourier transformation of this data, the MR images in x-space are obtained and remain Gaussian distribution. The magnitude images are computed from these complex MR images, and mathematically, the measured magnitude image $f \in \mathbb{R}^{p\times q}$ degraded by noises can be expressed [2] by

$$f_{i,j} = \sqrt{\left(u_{i,j} + (n_1)_{i,j}\right)^2 + \left(n_2\right)_{i,j}^2},\ 1 \le i \le p, 1 \le j \le q, \tag{2.1}$$

where $u \in \mathbb{R}^{p\times q}$ is the true amplitude of noise-free image, and $n_1 \in \mathbb{R}^{p\times q}$ and $n_2 \in \mathbb{R}^{p\times q}$ are two independent white Gaussian noise variables both with zero mean and standard deviation σ. Due to the non-linear transformation used to obtain magnitude images, the distribution of overall noises for the measured data defined in (2.1) is no longer Gaussian. The probability density function for each pixel of the magnitude MR image f is a Rician distribution [20, 2], i.e., for any $1 \le i \le p, 1 \le j \le q$,

$$\mathbb{P}(f_{i,j} \mid u_{i,j}, \sigma) = \frac{f_{i,j}}{\sigma^2} \exp\Big(-\frac{f_{i,j}^2 + u_{i,j}^2}{2\sigma^2}\Big) I_0\Big(\frac{f_{i,j}u_{i,j}}{\sigma^2}\Big). \tag{2.2}$$

where $I_0(\cdot)$ is the zero-order modified Bessel function of the first kind. Series forms of modified Bessel function of the first kind with real order ν are given by [35]

$$I_\nu(x) = \sum_{p=0}^{\infty} \frac{\left(\frac{1}{2}x\right)^{\nu+2p}}{p!\Gamma(\nu+p+1)}, \nu \in \mathbb{R}.$$

2.2. Related works of variational approaches. Numerous variational methods have been proposed to estimate the noise-free image u from the noisy observed image f. We briefly review some of these methods based on the Rician distribution (2.2) in the following.

In the MAP approach [16, 26], the image u is estimated by maximizing a posterior given f, that is $\tilde{u} = \arg\max_u \mathbb{P}(u \mid f)$. Applying the Bayes's rule and using the TV prior, the MAP model is given by

$$\min_u \mathrm{TV}(u) + \alpha\left(\frac{1}{2\sigma^2}\|u\|^2 - \sum_{i,j} \log I_0\left(\frac{f_{i,j}u_{i,j}}{\sigma^2}\right)\right) \tag{2.3}$$

with

$$\mathrm{TV}(u) := \sum_{i,j} \|(\nabla u)_{i,j}\| = \sum_{i,j} \sqrt{(\nabla_x u)_{i,j}^2 + (\nabla_y u)_{i,j}^2}. \tag{2.4}$$

Here, $\nabla u = (\nabla_x u, \nabla_y u)$ and ∇_x and ∇_y are first-order finite difference operators with proper boundary conditions in x and y directions, respectively. The MAP model (2.3) is non-smooth and non-convex, that poses great challenge of fast algorithms designing. Alternatively, Getreuer *et al.* [16] considered a smoothed version of MAP model

$$\min_u \sum_{i,j} \sqrt{\|(\nabla u)_{i,j}\|^2 + \epsilon^2} + \alpha\left(\frac{1}{2\sigma^2}\|u\|^2 - \sum_{i,j} \log I_0\left(\frac{f_{i,j}u_{i,j}}{\sigma^2}\right)\right) \tag{2.5}$$

and coped with this smoothed energy successfully with the ℓ^2 gradient descent algorithm. In [16], Getreuer *et al.* also proposed an elegant way to avoid the non-convexity, where the basic idea was to replace the non-convex fidelity function with one of its convex approximations. Hence the convex model [16] was derived as below:

$$\min_u \mathrm{TV}(u) + \alpha \sum_{i,j} G_\sigma(u_{i,j}), \tag{2.6}$$

where

$$G_\sigma(u_{i,j}) = \begin{cases} H_\sigma(u_{i,j}), & \text{if } u_{i,j} \geq c\sigma, \\ H_\sigma(c\sigma) + H'_\sigma(c\sigma)(u_{i,j} - c\sigma), & \text{if } u_{i,j} \leq c\sigma, \end{cases}$$
$$H'_\sigma(u_{i,j}) = \frac{u_{i,j}}{\sigma^2} - \frac{f_{i,j}}{\sigma^2} A\left(\frac{f_{i,j}u_{i,j}}{\sigma^2}\right), A(t) = \frac{t^3 + 0.950037t^2 + 2.38944t}{t^3 + 1.48937t^2 + 2.57541t + 4.65314},$$

with $c = 0.8426$ and $H_\sigma(\cdot)$ as the primitive function of $H'_\sigma(\cdot)$.

Following a similar idea for convexification [12] of the variational multiplicative noise removal, Chen and Zeng [10] added a quadratic term $\frac{1}{\sigma}\sum_{i,j}(\sqrt{u_{i,j}} - \sqrt{f_{i,j}})^2$ to the MAP model (2.3) based on the boundedness of $\mathbb{E}\big((\sqrt{u_{i,j}} - \sqrt{f_{i,j}})^2\big)/\sigma$ ($\mathbb{E}$ denoting the mathematical expectation), that led to the following model

$$\min_{u \in S} \mathrm{TV}(u) + \alpha\left(\frac{1}{2\sigma^2}\|u\|^2 - \sum_{i,j} \log I_0\left(\frac{f_{i,j}u_{i,j}}{\sigma^2}\right)\right) + \frac{\alpha}{\sigma}\sum_{i,j}(\sqrt{u_{i,j}} - \sqrt{f_{i,j}})^2. \tag{2.7}$$

Note that this model is strictly convex [10, 9] in the domain $S = \big\{w \mid 0 \leq w_{i,j} \leq 255,\ 1 \leq i \leq p, 1 \leq j \leq q\big\}$.

3. Proposed model.

3.1. Complex Gaussian noise removal. First, go back to the original model for complex-valued MR images. Let $U \in \mathbb{C}^{p\times q}$ be the complex image, and $F \in \mathbb{C}^{p\times q}$ be the noisy degraded image after contamination by complex white Gaussian noise $(N_1, N_2) \in \mathbb{R}^{p\times q} \times \mathbb{R}^{p\times q}$ as

$$F = U + N_1 + \boldsymbol{i} N_2, \tag{3.1}$$

with $\boldsymbol{i}^2 = -1$, and $(N_1)_{i,j}, (N_2)_{i,j} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma)$ $\forall 1 \le i \le p, 1 \le j \le q$. To recover U from F, directly applying the variational approach to (3.1) can yield the following model

$$\min_{U,N_1,N_2} \mathrm{TV}_{\mathbb{C}}(U) + \frac{\alpha}{2}\|N_1 + \boldsymbol{i}N_2\|^2, \quad \text{s.t. } F = U + N_1 + \boldsymbol{i}N_2, \tag{3.2}$$

with the parameter $\alpha > 0$ to balance the regularization (first term of the minimization problems in (3.2)) and data fitting terms (the second term of the minimization problems in (3.2)). The regularization term $\mathrm{TV}_{\mathbb{C}}(U)$ denotes the summation of TV of the magnitude and TV of the phase of U [42], which is given by

$$\mathrm{TV}_{\mathbb{C}}(U) := \mathrm{TV}(u) + \mathrm{TV}(\theta), \tag{3.3}$$

with $u \in \mathbb{R}^{p\times q}$ and $\theta \in \mathbb{R}^{p\times q}$ being the amplitude and phase of the complex-valued image U respectively, i.e. $U := u \circ \exp(\boldsymbol{i}\theta)$ ($\circ$ denotes the Hadamard product).

Remark 3.1. By naturally extending the definition of TV for the real-valued images to the complex-valued ones, one has

$$\widehat{\mathrm{TV}}(U) := \sum_{i,j} \|(\nabla U)_{i,j}\| = \sum_{i,j} \sqrt{\|(\nabla u)_{i,j}\|^2 + u_{i,j}^2 \|(\nabla \theta)_{i,j}\|^2}. \tag{3.4}$$

Readily one sees that the magnitude and the phase are coupled together. That poses potential difficulty for further modeling, since one cannot simply discard one of the them. Thus, we adopt the definition of TV for complex-valued images as (3.3).

3.2. Proposed model. Letting the complex variables be expressed in polar form $F := f \circ \exp(\boldsymbol{i}\phi)$ with the amplitude $f \in \mathbb{R}^{p\times q}$ and phase $\phi \in \mathbb{R}^{p\times q}$, the forward model (3.1) can be rewritten as

$$f = \exp\big(\boldsymbol{i}(\theta - \phi)\big) \circ (u + n_1 + \boldsymbol{i}n_2), \tag{3.5}$$

with n_1, n_2 denoting the real and imaginary parts of $(N_1 + \boldsymbol{i}N_2) \circ \exp(-\boldsymbol{i}\theta)$ respectively. The TV based regularization model (3.2) can also be rewritten as

$$\begin{aligned} &\min_{u,\theta,n_1,n_2} \mathrm{TV}(u) + \mathrm{TV}(\theta) + \frac{\alpha}{2}\|n_1 + \boldsymbol{i}n_2\|^2, \\ &\text{s.t. } f = \exp\big(\boldsymbol{i}(\theta - \phi)\big) \circ (u + n_1 + \boldsymbol{i}n_2). \end{aligned} \tag{3.6}$$

Since it is common to consider the magnitude image in practice [1], we focus on the recovery of the magnitude of the true image. Further, by the neglect of the phase θ and simplifying the constraint for (3.6), one has

$$\begin{aligned} &\min_{u,n_1,n_2} \mathrm{TV}(u) + \frac{\alpha}{2}\big(\|n_1\|^2 + \|n_2\|^2\big), \\ &\text{s.t. } f_{i,j} = \sqrt{\big(u_{i,j} + (n_1)_{i,j}\big)^2 + (n_2)^2_{i,j}},\ 1 \le i \le p, 1 \le j \le q, \end{aligned} \tag{3.7}$$

where the constraint is derived by taking the absolute part of both sides of (3.5). Note that this constraint is exactly the same as (2.1), referred as the spherical constraint, which essentially makes the vector $(u+n_1, n_2)$ (generated from the image u and noises (n_1, n_2)) onto a field of spheres.

The current model essentially decomposes the noisy measurement f into three variables u, n_1, n_2. Compared to the model (3.6), it employs one half (only the magnitude) of the original complex data F while discarding only one unknown variable θ among all the four unknown variables u, θ, n_1, n_2. Hence, the prior containing only the TV regularization for the image and the L^2 norm penalty for the noises seems insufficient to determine this decomposition. Other conditions should be further discussed in order to optimize (3.7).

This candidate could be $\|u - f\|^2$ by directly using an ℓ^2 fitting. However, the Rician noise introduces signal-dependent bias to the data that reduces the image contrast [29], thus the mean of the true image u is not equal to that of the observed image f. Heckelmann [20] and Chen-Zeng [10] showed the mean of each pixel of noise-free images could be approximated below:

$$\mathbb{E}(u_{i,j}) \approx \begin{cases} \mathbb{E}\left(\sqrt{\max\left\{0, f_{i,j}^2 - \sigma^2\right\}}\right), & \text{if } u_{i,j} > 2\sigma; \\ \mathbb{E}\left(\sqrt{\max\left\{0, f_{i,j}^2 - 2\sigma^2\right\}}\right), & \text{otherwise.} \end{cases}$$

Hence consider an estimate g as

$$g_{i,j} := \sqrt{\max(f_{i,j}^2 - c\sigma^2, 0)},\ 1 \le i \le p, 1 \le j \le q, \tag{3.8}$$

with a tunable scaling factor $c \in [1, 2]$. One may notice that this estimate is adaptive, i.e. when $f_{i,j}$ is small (not greater than $\sqrt{c}\sigma$) the estimate $g_{i,j}$ become zeros, while when $f_{i,j}$ gets larger (greater than $\sqrt{c}\sigma$), it becomes $\sqrt{f_{i,j}^2 - c\sigma^2}$. As a result, we consider the more efficient way by adding the quadratic term $\|u - g\|^2$ to (3.7) based on this estimate, in order to restrict the target not far away from a rough guess g.

Therefore, we propose the following variational model on the field of spheres as

$$\begin{cases} \min\limits_{u,n_1,n_2} \mathrm{TV}(u) + \dfrac{\alpha}{2}\left(\|n_1\|^2 + \|n_2\|^2\right) + \dfrac{\beta}{2}\|u - g\|^2, \\ \text{s.t. } f_{i,j} = \sqrt{\left(u_{i,j} + (n_1)_{i,j}\right)^2 + \left(n_2\right)_{i,j}^2},\ 1 \le i \le p, 1 \le j \le q, \end{cases} \tag{3.9}$$

with g as the estimate derived by (3.8).

We remark that the proposed model is entirely different from the existing variational models [16, 26, 10]. It is derived initially from the complex-valued image domain and uses the spherical constraint to express the relationship between images and noises. The spherical constraint reproduces the imaging process of the magnitude MR image. Compared with the sophisticated Bessel function, the spherical constraint is easy to handle by the operator splitting method [21], by which we can develop a more efficient algorithm for the Rician noise removal.

4. Proposed algorithm with convergence guarantee. In this section, we will propose an efficient operator splitting method for the non-convex model (3.9) with the spherical constraint and present its convergence analysis briefly.

We introduce a new auxiliary variable $\boldsymbol{v} := (v_1, v_2) \in \mathbb{R}^{p\times q} \times \mathbb{R}^{p\times q}$, with

$$\boldsymbol{v} = (u + n_1, n_2),$$

to split the spherical constraint such that the constrained problem (3.9) is equivalent to the following optimization problem

$$\begin{cases} \min\limits_{u,n_1,n_2,\boldsymbol{v}} \mathrm{TV}(u) + \frac{\alpha}{2}\left(\|n_1\|^2 + \|n_2\|^2\right) + \frac{\beta}{2}\|u-g\|^2, \\ \text{s.t. } \boldsymbol{v} = (u+n_1, n_2), \sqrt{(v_1)_{i,j}^2 + (v_2)_{i,j}^2} = f_{i,j},\ 1 \le i \le p, 1 \le j \le q. \end{cases} \tag{4.1}$$

For simplicity of presentation, we introduce some notations to reformulate the problem (4.1). Let $\mathcal{K}$ denotes the field of 2D-spheres given by

$$\mathcal{K} := \left\{ \boldsymbol{v} = (v_1, v_2) \mid \sqrt{(v_1)_{i,j}^2 + (v_2)_{i,j}^2} = f_{i,j},\ 1 \le i \le p, 1 \le j \le q \right\}. \tag{4.2}$$

Then define $\mathbb{I}_{\mathcal{K}}$ as the indicator function of $\mathcal{K}$ by

$$\mathbb{I}_{\mathcal{K}}(\boldsymbol{v}) = \begin{cases} 0, & \text{if } \boldsymbol{v} \in \mathcal{K}, \\ +\infty, & \text{if } \boldsymbol{v} \notin \mathcal{K}. \end{cases}$$

Let $\boldsymbol{Q} : \mathbb{R}^{p\times q} \to \mathbb{R}^{p\times q} \times \mathbb{R}^{p\times q}$ denotes the linear operator defined as

$$\boldsymbol{Q}u := (u, \boldsymbol{0}).$$

Finally, the TV model with spherical constraint (4.1) can be rewritten as

$$\begin{cases} \min\limits_{u,\boldsymbol{n},\boldsymbol{v}} \mathrm{TV}(u) + \frac{\alpha}{2}\|\boldsymbol{n}\|^2 + \frac{\beta}{2}\|u-g\|^2 + \mathbb{I}_{\mathcal{K}}(\boldsymbol{v}), \\ \text{s.t.} \quad \boldsymbol{Q}u + \boldsymbol{n} = \boldsymbol{v}, \end{cases} \tag{4.3}$$

with $\boldsymbol{n} = (n_1, n_2) \in \mathbb{R}^{p\times q} \times \mathbb{R}^{p\times q}$. The above problem has the non-smooth TV term and the non-convex indicator for the spherical constraint. One of the effective ways to cope with this problem is the operator-splitting algorithm, in which we can solve the subproblem involving TV term efficiently via many well-designed methods and obtain an analytic projection solution of the subproblem related to the field of spheres set $\mathcal{K}$. The spherical constrained problem is a special case of orthogonality constrained problems. There exist many first-order operator-splitting algorithms for the spherical constraint optimization problems, such as the curvilinear search method [36], the method of splitting orthogonality constraints [21], the coordinate descent-based method [15], and ADMM [18, 17]. Here, we adopt ADMM, which has been successfully applied in various convex and non-convex optimization problems and received considerable attention in recent years [6, 38, 33].

For the constrained optimization problem (4.3), we define the augmented Lagrangian function as follows

$$\begin{aligned} \mathcal{L}(\boldsymbol{v}, u, \boldsymbol{n}; \boldsymbol{\lambda}) = \mathbb{I}_{\mathcal{K}}(\boldsymbol{v}) + \mathrm{TV}(u) + \frac{\beta}{2}\|u-g\|^2 + \frac{\alpha}{2}\|\boldsymbol{n}\|^2 \\ + \langle \boldsymbol{\lambda}, \boldsymbol{Q}u + \boldsymbol{n} - \boldsymbol{v} \rangle + \frac{r}{2}\|\boldsymbol{Q}u + \boldsymbol{n} - \boldsymbol{v}\|^2, \end{aligned}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2) \in \mathbb{R}^{p\times q} \times \mathbb{R}^{p\times q}$ is a Lagrange multiplier with the penalization parameter $r > 0$. The ADMM is an iterative method to seek a saddle point of this

Lagrangian function. It minimizes the augmented Lagrangian function with respect to the primal variables $u, \boldsymbol{n}, \boldsymbol{v}$ alternately and updates the dual variable $\boldsymbol{\lambda}$. Here, we employ an ADMM scheme given below:

$$\boldsymbol{v}^{k+1} = \arg\min_{\boldsymbol{v}} \mathbb{I}_{\mathcal{K}}(\boldsymbol{v}) + \frac{r}{2}\left\|\boldsymbol{v} - (\boldsymbol{Q}u^k + \boldsymbol{n}^k + \boldsymbol{\lambda}^k/r)\right\|^2; \tag{4.4}$$

$$u^{k+1} = \arg\min_{u} \mathrm{TV}(u) + \frac{\beta}{2}\|u-g\|^2 + \frac{r}{2}\left\|u - (v_1^{k+1} - n_1^k - \lambda_1^k/r)\right\|^2; \tag{4.5}$$

$$\boldsymbol{n}^{k+1} = \arg\min_{\boldsymbol{n}} \frac{\alpha}{2}\|\boldsymbol{n}\|^2 + \frac{r}{2}\left\|\boldsymbol{n} - (\boldsymbol{v}^{k+1} - \boldsymbol{Q}u^{k+1} - \boldsymbol{\lambda}^k/r)\right\|^2; \tag{4.6}$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + r(\boldsymbol{Q}u^{k+1} + \boldsymbol{n}^{k+1} - \boldsymbol{v}^{k+1}). \tag{4.7}$$

In the following, we show how to solve these subproblems. It is interesting to show that we can further eliminate the multiplier updating step (4.7) and finally cope with three subproblems. Indeed, from the first-order optimization condition of the $\boldsymbol{n}$-subproblem (4.6), which is a quadratic optimization problem, we have

$$\alpha\boldsymbol{n}^{k+1} + \boldsymbol{\lambda}^k + r(\boldsymbol{n}^{k+1} + \boldsymbol{Q}u^{k+1} - \boldsymbol{v}^{k+1}) = 0. \tag{4.8}$$

Combined with (4.7), it follows that

$$\boldsymbol{\lambda}^{k+1} = -\alpha\boldsymbol{n}^{k+1}. \tag{4.9}$$

Replace therefore $\boldsymbol{\lambda}^k$ by $-\alpha\boldsymbol{n}^k$ in (4.4), (4.5) and (4.6). Thus, the solution of the $\boldsymbol{n}$-sub problem is

$$\boldsymbol{n}^{k+1} = \frac{\alpha\boldsymbol{n}^k + r(\boldsymbol{v}^{k+1} - \boldsymbol{Q}u^{k+1})}{\alpha + r}. \tag{4.10}$$

The $\boldsymbol{v}$-sub problem (4.4) is equivalent to pq projection problems of the independent component in the form

$$\min_{\boldsymbol{v}_{i,j}\in\mathbb{R}^2} \|\boldsymbol{v}_{i,j} - \hat{\boldsymbol{v}}_{i,j}\|^2 \text{ s.t. } \|\boldsymbol{v}_{i,j}\| = f_{i,j},\ 1\le i\le p, 1\le j\le q,$$

with

$$\hat{\boldsymbol{v}}_{i,j} = (\boldsymbol{Q}u^k + \boldsymbol{n}^k)_{i,j} + \boldsymbol{\lambda}_{i,j}^k/r = (\boldsymbol{Q}u^k)_{i,j} + (1-\alpha/r)\boldsymbol{n}_{i,j}^k \in \mathbb{R}^2.$$

Each of the above problems has a closed form solution

$$\boldsymbol{v}_{i,j} = f_{i,j}\frac{\hat{\boldsymbol{v}}_{i,j}}{\|\hat{\boldsymbol{v}}_{i,j}\|},\ 1\le i\le p, 1\le j\le q. \tag{4.11}$$

The u-sub problem (4.5) is a TV-ℓ_2 optimization problem

$$\min_u \left\{\mathrm{TV}(u) + \frac{\beta + r}{2}\|u - \hat{u}\|^2\right\}, \tag{4.12}$$

where

$$\hat{u} = \frac{\beta g + r\left(v_1^{k+1} - n_1^k\right) - \lambda_1^k}{\beta + r} = \frac{\beta g + r v_1^{k+1} + (\alpha - r)n_1^k}{\beta + r}.$$

The problem (4.12) is known as the Rudin-Osher-Fatemi (ROF) model [31] and has a unique solution due to the strict convexity of the objective function. It has been well-studied during the last three decades and has many efficient solvers nowadays,

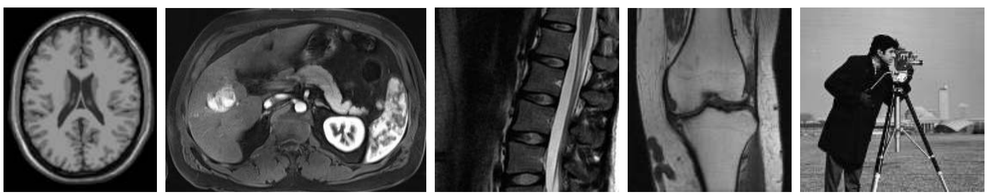
Algorithm 4.1 Simplified ADMM for solving the model (4.3)

Initialization:
$k = 0$, $g = \sqrt{\max(f^2 - c\sigma^2, 0)}$, $c \in [1, 2]$, $u^0 = g$, $n_1^0 = 0$, $n_2^0 \sim \mathcal{N}(0, \sigma)$, $tol = 1e-4$, $K = 500$.

Iteration:

1. For given $(u^k, \boldsymbol{n}^k)$, compute $\boldsymbol{v}^{k+1}$ through (4.11);
2. For given $(\boldsymbol{n}^k, \boldsymbol{v}^{k+1})$, find u^{k+1} through solving (4.12) by Chambolle's projection method;
3. For given $(u^{k+1}, \boldsymbol{v}^{k+1})$, compute $\boldsymbol{n}^{k+1}$ from (4.10);

Until: $\|u^k - u^{k+1}\|/\|u^k\| < tol$ or $k \geq K$.



(a) (b) (c) (d) (e)

FIG. 1. *Clean magnitude MR images. (a) Brain (*181×217*), (b) Liver (*304×214*), (c) Spine (*200×200*), (d) Knee (*194×218*), (e) Cameraman (*256×256*).*

such as Chambolle's projection method [7], the fast iterative shrinkage thresholding algorithm (FISTA) [4], the split Bregman method [19], the augmented Lagrangian method [38], and primal-dual algorithms [13, 8].

In summary, we describe a simplified ADMM for solving the model (4.3) in the Algorithm 4.1.

Next, we present a brief convergence analysis of the ADMM algorithm (4.4)-(4.7). Our algorithm falls into the algorithmic framework of the non-smooth and non-convex ADMM in [33]. The global convergence results have been established in [33], and we refer to Appendix A for details. The convergence of the proposed algorithm is given as follows.

THEOREM 4.1. *For any sufficiently large stepsize r, the ADMM algorithm* (4.4)-(4.7) *generates a sequence $(u^k, \boldsymbol{v}^k, \boldsymbol{n}^k, \boldsymbol{\lambda}^k)$ that converges to a stationary point of $\mathcal{L}$.*

Finally, we give some comments on solving the subproblem (4.12). As is known to all, a proper approximate solution to the subproblem is sufficient for the numerical convergence of ADMM. Please also see the numerical examples in Section 5.2. Moreover, the overall cost will be significantly reduced by very few iterations of Chambolle's projection. More rigorous analysis for the inexact version algorithm will be explored in the future.

5. Numerical experiments. In this section, we present some numerical experiments to evaluate the performance of the proposed model for Rician noise removal. We conduct the experiments with MATLAB R2020a on a desktop computer with a 4-cores 3.4GHz Intel Core i7-6700 processor and 16GB RAM. Firstly, we apply our method to synthetic MR images to verify its effectiveness. Then, we discuss the choices of the inner iteration numbers, the role of the correction term, and the influences of parameters in our method. Next, we compare our approach with other methods, including variational and deep learning methods. Finally, we apply our method to some

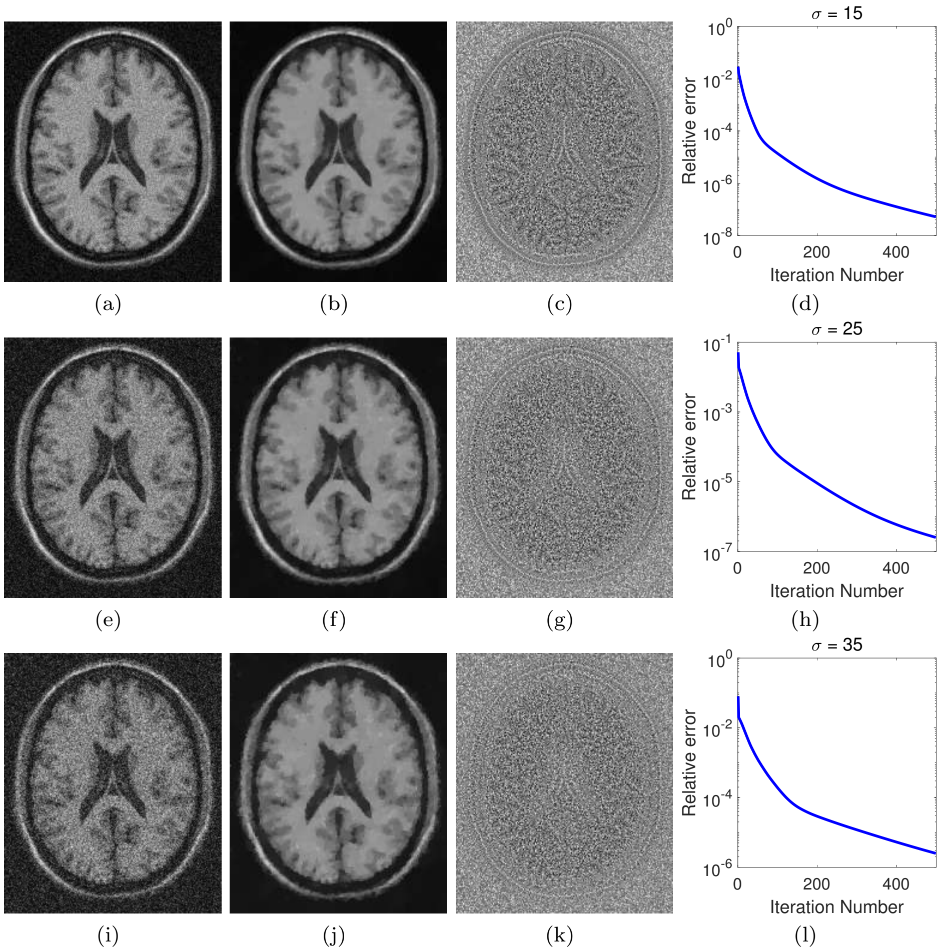


Fig. 2. *The denoised results of the proposed method for the synthetic MR "Brain" image. Here, (a), (e), and (i) are images degraded by Rician noises with standard deviation* $\sigma = 15$, 25, *and* 35. *(b), (f) and (j) are restored images, and (c), (g) and (k) are residual images. The parameters are: (b)* $\alpha = 0.015$ *and* $\beta = 0.08$*; (f)* $\alpha = 0.01$ *and* $\beta = 0.045$*; and (j)* $\alpha = 0.005$ *and* $\beta = 0.03$*. Here, (b) PSNR: 31.46dB, (f) PSNR: 28.34dB, and (j) PSNR: 26.16dB.*

raw MR images.

We carry out the trials on five images including "Brain" (size of 181×217), "Liver" (size of 304×214), "Lumbar spine" (size of 200×200), "Knee" (size of 194×218), and "Cameraman" (size of 256×256), as shown in Fig. 1. The range of gray-scale image is $[0, 255]$. In our experiments, all these images are corrupted with Rician noises with standard deviations $\sigma = 15$, 25, and 35.

We use two quantitative metrics to measure the qualities of the restoration results. The first metric is the PSNR, which is defined as

$$\mathrm{PSNR}(u, \tilde{u}) = 10 \log_{10} \frac{255^2 \times m^2}{\sum_{i,j} (u_{i,j} - \tilde{u}_{i,j})^2},$$

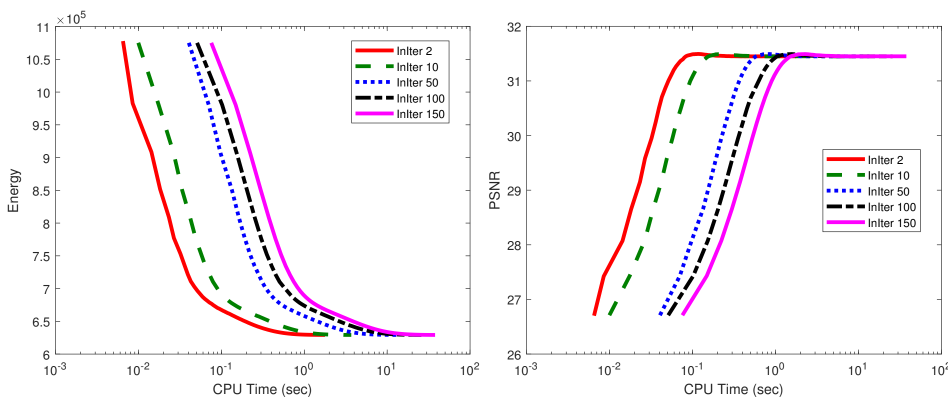
where u and $\tilde{u}$ are the noiseless image and the recovered image, respectively. The second one is the SSIM [34], which is computed via

$$\mathrm{SSIM}(u,\tilde{u}) = \frac{(2\mu_u\mu_{\tilde{u}} + C_1)(2\sigma_{u\tilde{u}} + C_2)}{(\mu_u^2 + \mu_{\tilde{u}}^2 + C_1)(\sigma_u^2 + \sigma_{\tilde{u}}^2 + C_2)},$$

with two constants C_1 and C_2, where μ_u, $\mu_{\tilde{u}}$, σ_u, $\sigma_{\tilde{u}}$, and $\sigma_{u\tilde{u}}$ are local means, standard deviations and cross-covariance for images u and $\tilde{u}$ respectively.

5.1. Application to synthetic MR images. We conduct the first experiment on the synthetic MR "Brain" image; see Fig. 1 (a), which are taken from the BrainWeb database[1]. We apply the proposed method to the degraded images of "Brain" with Rician noises at levels $\sigma = 15$, 25, and 35. The restored results are showed in Fig. 2, in which we have choose $\alpha = 0.015$ and $\beta = 0.08$ for Fig. 2 (b), $\alpha = 0.01$ and $\beta = 0.045$ for Fig. 2 (f), and $\alpha = 0.01$ and $\beta = 0.045$ for Fig. 2 (j). Since the "Brain" image in the experiment is synthetic, we can easily get the exact region of the foreground. In Fig. 2, PSNR values are computed for the foreground. We can see that the proposed model can well remove the Rician noises from the degraded images and preserve the main features in the restored images. We also record the evolution relative errors of the proposed ADMM in Fig. 2 (d), (h), and (l), which verify the convergence numerically.

5.2. Extensive tests with various settings. We conduct the second experiment to discuss various aspects of the proposed model by Algorithm 4.1, including the choices of the inner iteration numbers, the role of the correction term, and influences of parameters.



(a) (b)

Fig. 3. *Performance of the proposed ADMM under the same outer iteration number* 500 *and different inner iteration numbers, i.e.* 2, 10, 50, 100, *and* 150. *This example is performed on "Brain" image.*

According to the analysis in Section 4, the proposed ADMM algorithm (4.4)-(4.7) is globally convergent based on the exact solver for the sub-problem (4.5). In practice, the sub-problem (4.5) can be solved inexactly for each outer iteration. In the sequel, we show how the number of iterations in the inexact ROF solver impacts the solution, e.g., setting 2, 10, 50, 100, and 150 iterations for Chambolle's projection method. See Fig. 3 for denoising the degraded "Brain" image (Rician noises with

[1]BrainWeb: Simulated Brain Database: https://brainweb.bic.mni.mcgill.ca/brainweb/

standard deviations $\sigma = 15$). From Fig. 3 (a), we see that the objective function (energy) of the model (3.9) decreases as the evolution of Algorithm 4.1 and stabilizes eventually. It indicates too that the proposed algorithm converges numerically when an inexact solver for the ROF problem (4.5) is used. Combined with Fig. 3 (a) and (b), we also see that using 2 fixed inner iterations can significantly reduce the overall computational cost of the proposed algorithm without affecting the recovery qualities. Thus, using 2 fixed inner iterations are enough to reach a proper estimated solution for the convergence guarantee. In all the experiments, we use two inner loops for the proposed algorithm.

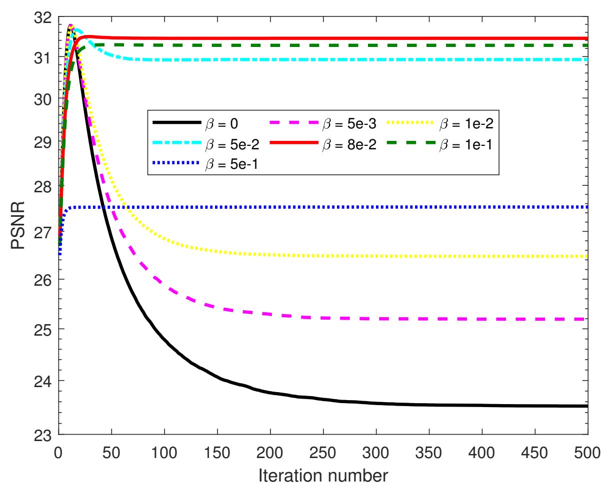


Fig. 4. *Performance of solving the proposed model by ADMM under different β. This example is carried out on "Brain" image.*

As analyzed in Section 3.2, the correction term in the proposed model (3.9) plays an important role in the determination of the variable decomposition. We conduct the tests by gradually increasing the balance parameter β from 0 to 0.5 ($\beta = 0$ means no correction). Please refer to Fig. 4 to see how this term affects the performance. As can be seen, when the model (3.9) does not have the correction term (*i.e.*, $\beta = 0$), the PSNRs of estimated images along the evolution reaches the highest value after few numbers of iteration and then quickly decrease to a steady-state with very low value. Actually, with $\beta = 0$, the final estimate of u loses almost all details, while all of these features move to n_1 and n_2. This coincides with the analysis in Section 3.2, which points out that the original model (3.7) cannot generate ideal decomposition. With correction term by choosing a proper β, the proposed algorithm can effectively prevent the PSNR values of the estimated images from dramatic decrease, as shown in Fig. 4. In other words, the correction term can also be interpreted as the stabilization term of the model (3.9). Therefore, the correction term is essential for a stable recovery. Moreover, the estimate g has a tunable scaling factor $c \in [1, 2]$. To demonstrate the influences of the parameter c, we conduct experiments of denoising the noisy images of "Brain" with Rician noises at level $\sigma = 15$, 25, and 35 with different c in the range $[1, 3]$. The results are put in Fig. 5, which shows that tuning c between 1 and 2 could improve the recovery results compared with $c = 2$, especially when $\sigma = 15$. However, for practical use, $c = 2$ seems to be a good choice for the most cases, and therefore, set $c = 2$ as the default for other experiments unless otherwise specified.

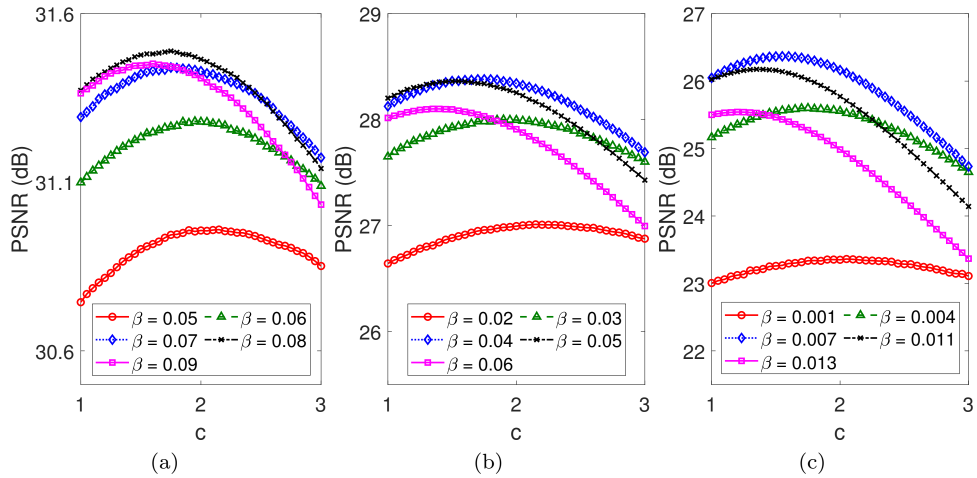


FIG. 5. *The PSNR values under different pairs* (β, c) *for denoising the noisy images of "Brain" with Rician noises at level (a)* $\sigma = 15$*, (b)* $\sigma = 25$*, and (c)* $\sigma = 35$ *by using the proposed method.*

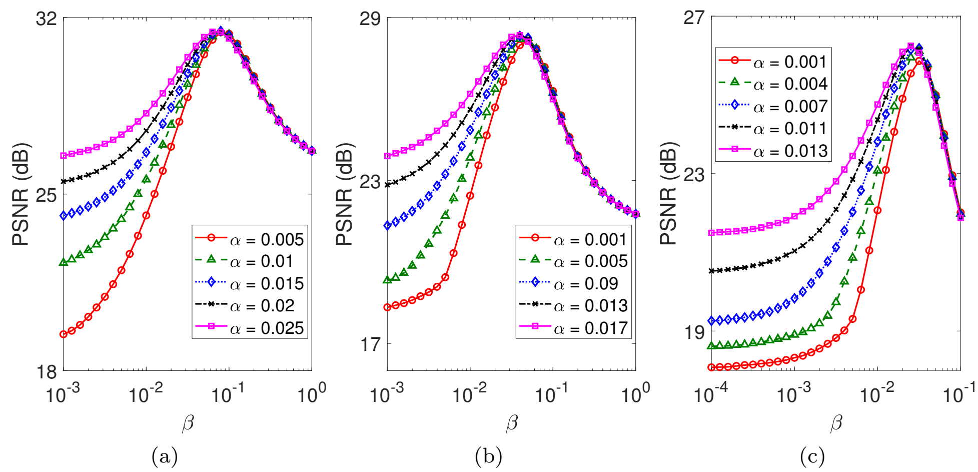


FIG. 6. *The PSNR values under different pairs* (α, β) *for denoising the noisy images of "Brain" with Rician noises at level (a)* $\sigma = 15$*, (b)* $\sigma = 25$*, and (c)* $\sigma = 35$ *by using the proposed method.*

Then we discuss the influence of the proposed model (3.9) by the parameters α and β. Fig. 6 shows the PSNR values of the denoised "Brain" image by our proposed method for different pairs (α, β). The parameter α should be related to the noise level. As the standard deviation of the noise increases, the value of α should decrease to get good recovery results. We suggest that the parameter α is in $[0.001, 0.03]$ for $\sigma = 15$, 25, and 35. As discussed before, the parameter β controls the distance between the final result and the estimate g. Again, in our experiments, we noticed that if β gets too small, the recovered image will be over-smooth; if β gets too large, the recovered image will approach the undesirable noise image g. From Fig. 6, we can observe that the choice of β has a significant impact on the quality of the denoised image. Moreover, we can also see from Fig. 6 that for each noise level, when a good value of β is chosen, the image quality is less sensitive to the parameter α.

The parameter r is the penalty parameter of the augmented Lagrangian function used in the ADMM algorithm (4.4)-(4.7). For the proposed algorithm to converge, we need to set r to be large enough. Hence, we empirically choose $r \in [0.1, 10]$ with $r = 1$ as default.

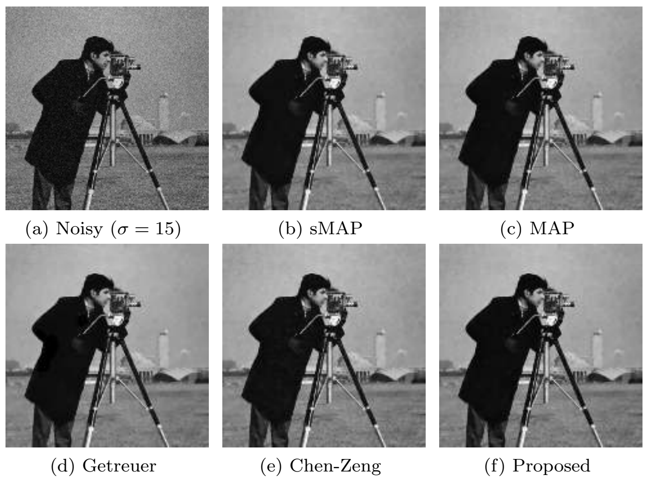


Fig. 7. *Numerical results for the image of "Cameraman". (a) Noisy image with $\sigma = 15$; The denoised images: (b) sMAP model (PSNR = 30.13), (c) MAP model (PSNR = 30.26), (d) Getreuer's model (PSNR = 29.94), (e) Chen-Zeng model (PSNR = 30.15), (f) the proposed model (PSNR = 30.37).*

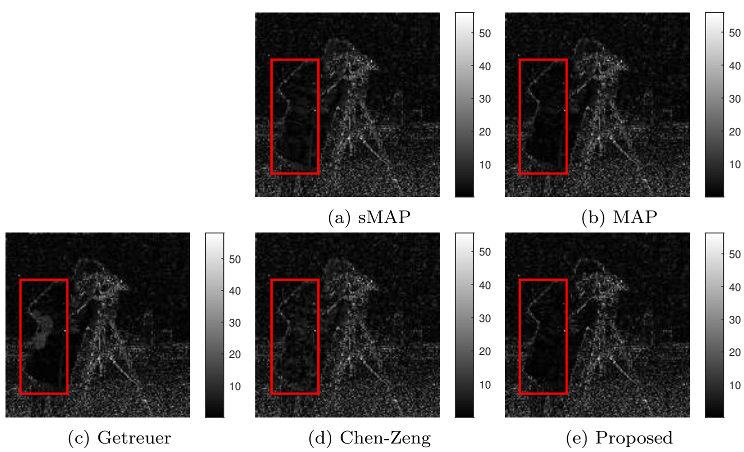


Fig. 8. *The absolute differences between the denoised images in Fig. 7 and the true "Cameraman" image. (a) sMAP model, (b) MAP model, (c) Getreuer's model, (d) Chen-Zeng model, (e) the proposed model.*

5.3. Comparisons with other variational methods. We conduct the third experiment on comparing the proposed method with some existing variational methods listed below:

1) The smoothed MAP (sMAP) model (2.5) solved by the ℓ^2 gradient descent method in [16]. As recommended in [16], we fix the step size $dt = 0.1$.
2) The MAP model (2.3) solved by a proximal point algorithm proposed by Martín *et al.* in [27].

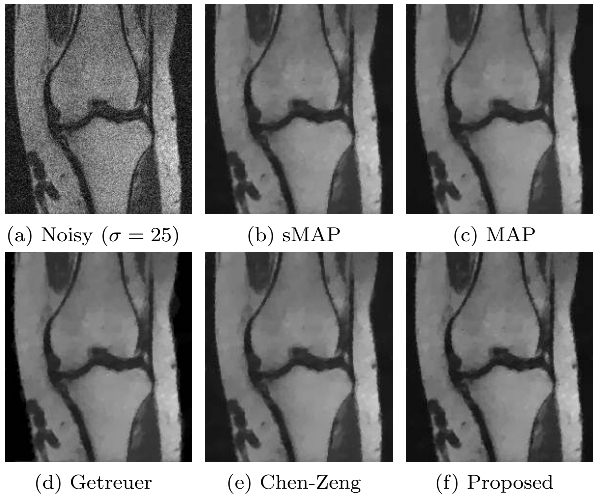


FIG. 9. *Numerical results for the image of "Knee". (a) Noisy image with $\sigma = 25$; The denoised images: (b) sMAP model (PSNR = 28.50), (c) MAP model (PSNR = 28.81), (d) Getreuer's model (PSNR = 27.75), (e) Chen-Zeng model (PSNR = 28.19), (f) the proposed model (PSNR = 28.92).*

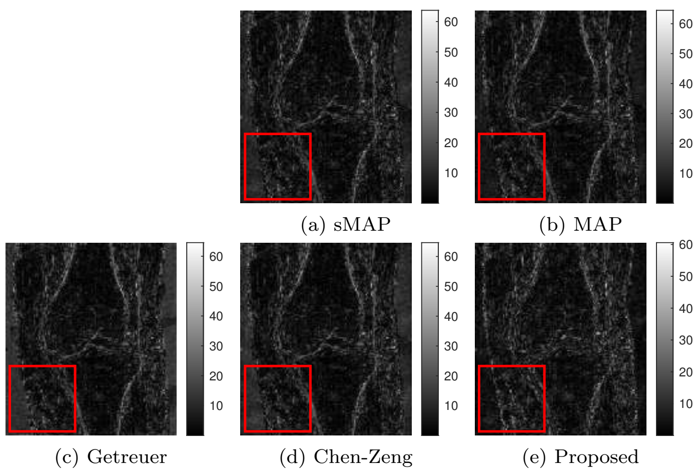


FIG. 10. *The absolute differences between the denoised images in Fig. 9 and the true "Knee" image. (a) sMAP model, (b) MAP model, (c) Getreuer's model, (d) Chen-Zeng model, (e) the proposed model.*

3) The Getreuer's model (2.6) solved by the split Bregman method [16].

4) The Chen-Zeng model (2.7) solved by a primal-dual algorithm [10].

The source codes of all the algorithms are implemented in MATLAB. The code for the proximal point algorithm is kindly provided by the authors of [27], and the codes for the sMAP model, the Getreuer's model, and the CZ model are kindly provided by the authors of [10]. We adopt the following stopping criteria for all the compared algorithms.

$$\frac{\|u^k - u^{k+1}\|}{\|u^k\|} < tol$$

with $tol = 1e-4$. Set the maximum number of iterations as 500 for all the algorithms

Table 1

Comparison of the performance of the proposed model with the sMAP model, MAP model, Getreuer's model, and Chen-Zeng model in terms of PSNR values and SSIM values.

Image	σ		sMAP	MAP	Getreuer	Chen-Zeng	New
Cameraman	15	PSNR	30.13	30.26	29.94	30.15	**30.37**
		SSIM	0.8590	0.8599	0.8513	0.8503	**0.8621**
	25	PSNR	26.88	26.98	26.29	26.43	**27.41**
		SSIM	0.7900	0.7979	0.6858	0.7846	**0.7997**
	35	PSNR	24.23	25.04	24.68	24.09	**25.12**
		SSIM	0.7260	0.7452	0.6409	0.7261	**0.7505**
Brain	15	PSNR	31.41	31.40	31.31	31.39	**31.46**
		SSIM	0.9370	0.9355	0.9372	**0.9373**	0.9372
	25	PSNR	28.34	28.30	28.22	28.28	**28.34**
		SSIM	0.8860	**0.8877**	0.8856	0.8839	0.8867
	35	PSNR	26.33	**26.38**	26.18	26.14	26.16
		SSIM	0.8310	0.8355	0.8282	0.8426	**0.8431**
Liver	15	PSNR	30.82	**30.85**	30.79	30.75	30.75
		SSIM	0.8480	**0.8476**	0.8456	0.8438	0.8429
	25	PSNR	27.93	27.99	27.37	27.80	**28.00**
		SSIM	0.7690	0.7677	0.7071	0.7649	**0.7688**
	35	PSNR	**25.84**	25.76	25.77	25.32	25.81
		SSIM	0.7010	0.6903	0.6387	0.6895	**0.6962**
Spine	15	PSNR	30.23	30.35	29.65	30.09	**30.52**
		SSIM	0.8320	0.8369	0.7535	0.8232	**0.8420**
	25	PSNR	27.21	**27.39**	25.71	26.25	27.29
		SSIM	0.7370	**0.7431**	0.5545	0.7084	0.7347
	35	PSNR	24.54	**25.59**	23.82	23.32	25.25
		SSIM	0.6410	**0.6713**	0.4566	0.6136	0.6625
Knee	15	PSNR	31.51	31.60	31.56	31.44	**31.62**
		SSIM	0.8540	0.8563	0.8544	0.8518	**0.8565**
	25	PSNR	28.50	28.81	27.75	28.19	**28.92**
		SSIM	0.7910	0.7979	0.6949	0.7870	**0.8051**
	35	PSNR	26.03	**26.81**	26.21	25.61	26.51
		SSIM	0.7290	0.7478	0.6435	0.7288	**0.7555**
Average		PSNR	28.00	28.23	27.68	27.68	**28.24**
		SSIM	0.7950	0.8014	0.7319	0.7891	**0.8029**

except for the ℓ^2 gradient descent algorithm, whose maximum iteration number is set as 2000 to obtain high-quality images. To achieve the best performances of these approaches, we manually set σ to the true value. Some noise estimation methods can also be found in [1, 11]. For the sake of fairness, we have tuned the parameters of these methods to achieve the best balance between PSNR values and visual quality.

Table 1 shows the experimental results by these approaches for different images in terms of PSNR and SSIM. The best values of PSNR and SSIM for each case have been highlighted in bold. Some visual results for $\sigma = 15$, 25, and 35 are shown in Figs. 7–12. We give the denoised images of "Cameraman", "Knee", and "Liver" in Figs. 7, 9, and 11, respectively. We also show the absolute differences between the denoised images and the true images in Figs. 8, 10, and 12. Fig. 13 presents the energy evolution via iteration numbers for five different methods applying to denoise the "Brain" image with $\sigma = 15$. Compared with other methods, the proposed method performs better at removing noises (the highest average PSNR) and preserving features (highest average SSIM). It is interesting to see that these non-convex models, including the sMAP model, MAP model, and the proposed model, perform better than Getreuer's model and Chen-Zeng model in most cases. As can be seen from the red boxes in Figs. 8, 10, and 12, these non-convex models can preserve the contrast better than Getreuer's

Table 2

Comparison of the performance of the proposed method with the sMAP method, MAP method, Getreuer's method, and Chen-Zeng method in terms of total CPU time, number of iterations (# of iter), and time of per iteration (per-iter).

Image	σ		sMAP	MAP	Getreuer	Chen-Zeng	New
Cameraman	15	# of iter	283	53	73	**37**	47
		Time(s)	4.54	9.4	2.19	0.5	**0.29**
		per-iter	0.0160	0.1773	0.0300	0.0135	**0.0061**
	25	# of iter	543	113	**49**	50	82
		Time(s)	9.98	20.36	1.3	0.69	**0.37**
		per-iter	0.0184	0.1801	0.0265	0.0137	**0.0045**
	35	# of iter	722	169	**56**	**56**	125
		Time(s)	17.88	31.25	1.48	1.02	**0.6**
		per-iter	0.0248	0.1849	0.0265	0.0183	**0.0048**
Brain	15	# of iter	295	103	**41**	44	49
		Time(s)	2.89	11.99	0.62	0.33	**0.14**
		per-iter	0.0098	0.1164	0.0152	0.0074	**0.0029**
	25	# of iter	529	173	**36**	49	85
		Time(s)	6.1	20.31	0.55	0.42	**0.27**
		per-iter	0.0115	0.1174	0.0154	0.0085	**0.0032**
	35	# of iter	751	237	84	**57**	124
		Time(s)	10.25	28.13	1.27	0.59	**0.37**
		per-iter	0.0136	0.1187	0.0152	0.0103	**0.0030**
Liver	15	# of iter	282	54	**44**	49	51
		Time(s)	6.5	9.56	1.32	0.81	**0.23**
		per-iter	0.0231	0.1771	0.0299	0.0166	**0.0044**
	25	# of iter	510	121	71	**60**	98
		Time(s)	12.6	22.29	2.07	1.07	**0.43**
		per-iter	0.0247	0.1842	0.0291	0.0178	**0.0044**
	35	# of iter	717	202	110	**73**	157
		Time(s)	16.24	37.34	3.15	1.18	**0.66**
		per-iter	0.0226	0.1849	0.0286	0.0162	**0.0042**
Spine	15	# of iter	272	81	95	**55**	59
		Time(s)	3.56	9.66	1.51	0.55	**0.17**
		per-iter	0.0131	0.1193	0.0159	0.0099	**0.0028**
	25	# of iter	819	169	93	**65**	97
		Time(s)	9.38	20.23	1.32	0.56	**0.27**
		per-iter	0.0115	0.1197	0.0142	0.0086	**0.0028**
	35	# of iter	709	260	95	**77**	130
		Time(s)	7.6	30.86	1.26	0.61	**0.43**
		per-iter	0.0107	0.1187	0.0133	0.0079	**0.0033**
Knee	15	# of iter	246	56	54	**46**	49
		Time(s)	2.77	6.64	1.04	0.39	**0.14**
		per-iter	0.0113	0.1186	0.0192	0.0084	**0.0029**
	25	# of iter	505	116	88	**59**	89
		Time(s)	6.9	14.24	1.62	0.59	**0.27**
		per-iter	0.0137	0.1227	0.0184	0.0100	**0.0031**
	35	# of iter	662	176	**65**	66	139
		Time(s)	10.77	22.02	1.19	0.77	**0.43**
		per-iter	0.0163	0.1251	0.0183	0.0117	**0.0031**
Average		# of iter	523	139	70	**56**	92
		Time(s)	8.53	19.62	1.46	0.67	**0.34**
		per-iter	0.0161	0.1443	0.0211	0.0119	**0.0037**

model and Chen-Zeng model.

Finally, we discuss the computational costs of all comparable methods. We record the CPU costs per iteration and overall algorithms and the number of iterations for all the five algorithms and put it in Table 2. As can be seen, the proposed method is the fastest among all these algorithms since it saves about 50% of the overall

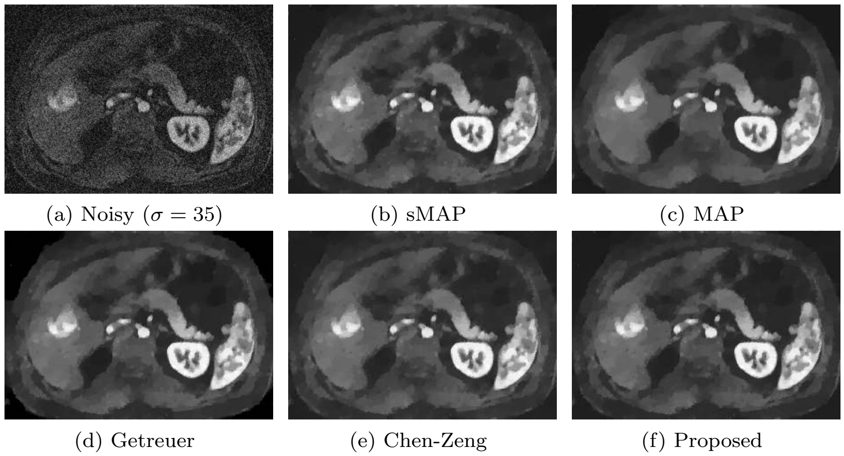


FIG. 11. *Numerical results for the image of "Liver". (a) Noisy image with $\sigma = 35$; The denoised images: (b) sMAP model (PSNR = 25.84), (c) MAP model (PSNR = 25.76), (d) Getreuer's model (PSNR = 25.77), (e) Chen-Zeng model (PSNR = 25.32), (f) the proposed model (PSNR = 25.81).*

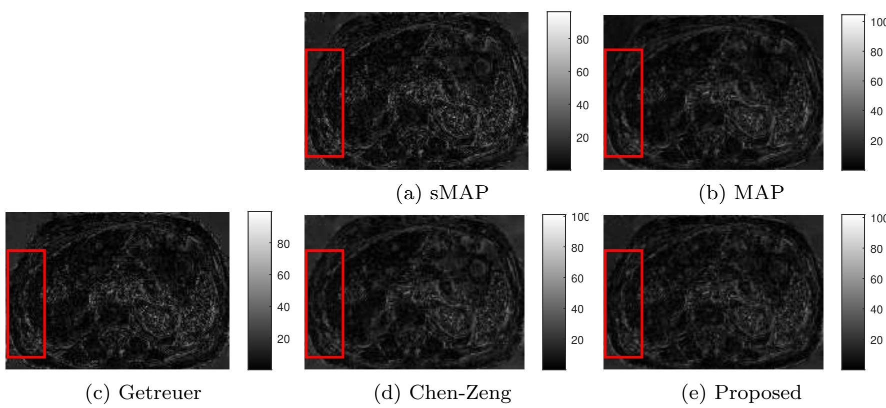
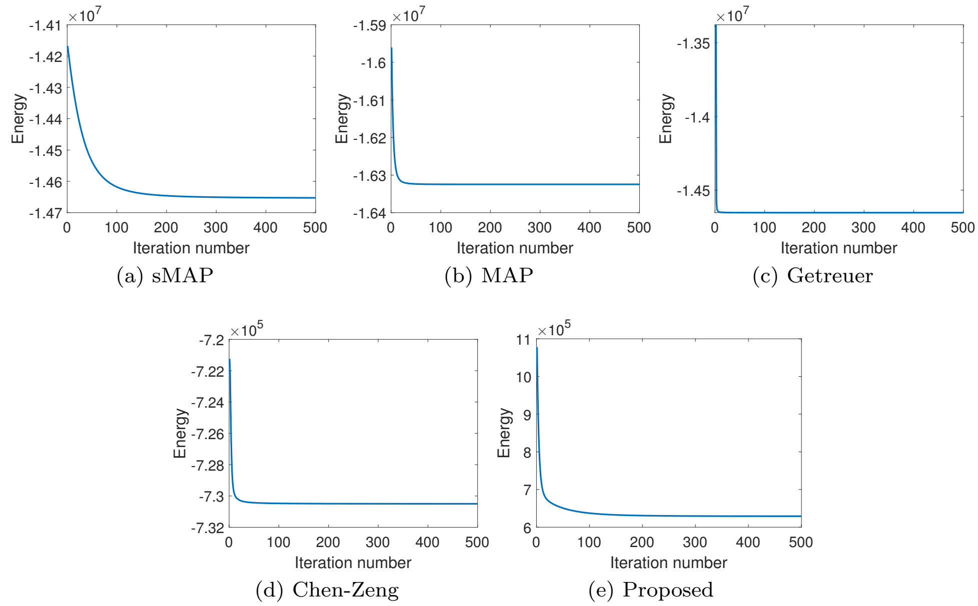


FIG. 12. *The absolute differences between the denoised images in Fig. 11 and the true "Liver" image. (a) sMAP model, (b) MAP model, (c) Getreuer's model, (d) Chen-Zeng model, (e) the proposed model.*

computation cost on average even compared with the convex methods. One may also notice that although the Chen-Zeng method requires the fewest iterations on average, the proposed method is still the fastest in terms of runtime since the cost per iteration is dramatically lower than other methods involving the calculations with the Bessel function. In summary, our method can achieve comparable denoising results with the lowest computational cost.

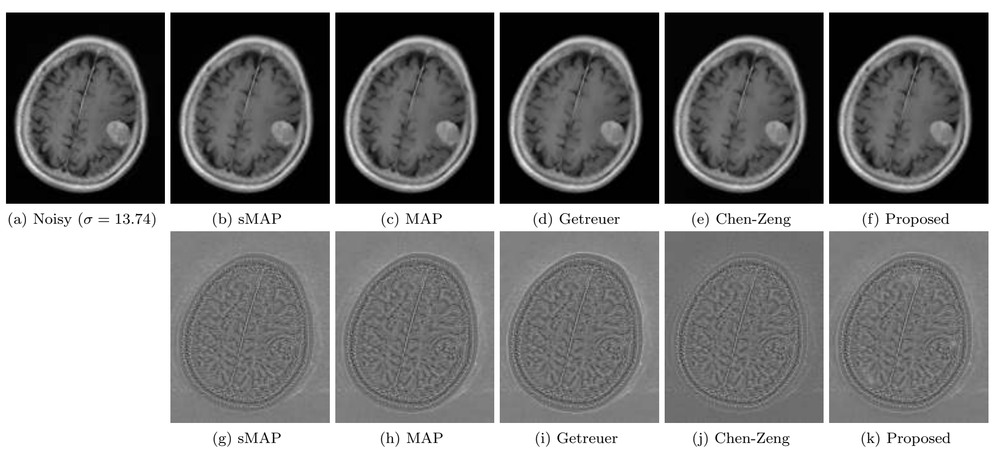
5.4. Comparisons with deep learning methods. We compare the proposed method with deep learning methods, including the trainable non-linear reaction diffusion (TNRD) [39] and the wider denoising neural network (WDNN) [40]. For TNRD, we use the trained model provided by the authors of [39]. For WDNN, we implement the network, train it in a workstation with one NVIDIA Tesla P100 GPU computing processor, and use the trained model for our test. We should mention that compared with deep learning methods, the advantage of our method is that it does not require



(a) sMAP (b) MAP (c) Getreuer (d) Chen-Zeng (e) Proposed

FIG. 13. *Energy evolution via iteration numbers for five different methods. This example is carried out on "Brain" image.*

training data and training processing. We also test these two methods for different noise levels, i.e, $\sigma = 15$, 25, and 35. We find that the average PSNR values rank in the descending order as follows: TNRD (29.18dB), WDNN (28.92dB), and New (28.24dB). However, our method (0.34s) is the fastest, followed by TNRD (0.48s) and WDNN (3.01s).



(a) Noisy ($\sigma = 13.74$) (b) sMAP (c) MAP (d) Getreuer (e) Chen-Zeng (f) Proposed

(g) sMAP (h) MAP (i) Getreuer (j) Chen-Zeng (k) Proposed

FIG. 14. *Numerical results for the raw MR image. (a) Noisy image f with the estimated noise level $\sigma = 13.74$; The denoised images $\tilde{u}$: (b) sMAP model, (c) MAP model, (d) Getreuer's model, (e) Chen-Zeng model, (f) the proposed model; The residual images $f - \tilde{u}$: (g) sMAP model, (h) MAP model, (i) Getreuer's model, (j) Chen-Zeng model, (k) the proposed model.*

5.5. Test on raw MR data. To show the effectiveness of our method, we apply it to raw MR data. We test on two raw MR images shown in Figs. 14(a) and 15(a),

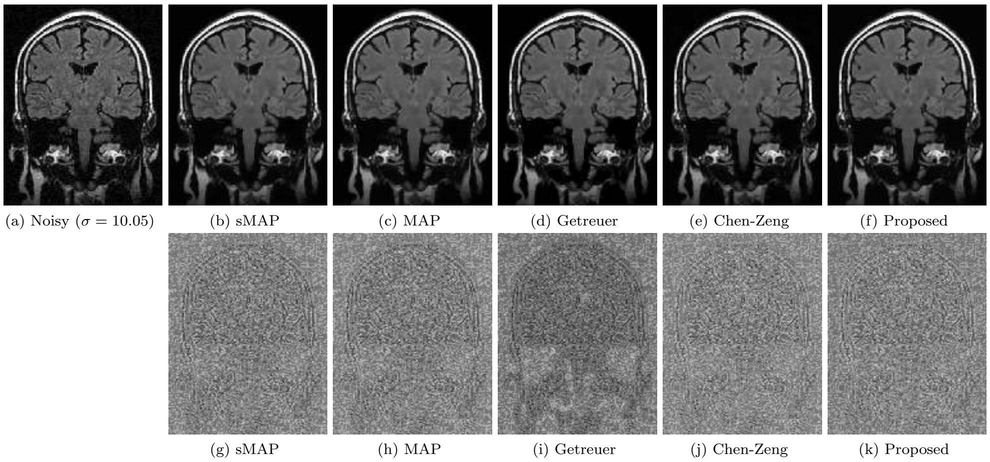


FIG. 15. *Numerical results for the raw MR image. (a) Noisy image f with the estimated noise level $\sigma = 10.05$; The denoised images $\tilde{u}$: (b) sMAP model, (c) MAP model, (d) Getreuer's model, (e) Chen-Zeng model, (f) the proposed model; The residual images $f - \tilde{u}$: (g) sMAP model, (h) MAP model, (i) Getreuer's model, (j) Chen-Zeng model, (k) the proposed model.*

whose noise levels are estimated as $\sigma = 13.74$ and $\sigma = 10.05$. The denoised images and the corresponding residual images are presented in Figs. 14 and 15. As can be seen, our method can remove the Rician noise quite well and the denoised results of our method are comparable to these of other variational methods. Especially from the second row of Fig. 15, the residual image of our method seems more uniform than those of other methods.

6. Conclusions. In this paper, we have proposed a novel variational method for Rician noise removal for magnitude MRI, which avoids using the Bessel functions. It is a nonconvex optimization model with spherical constraints and has been solved by the convergent ADMM efficiently. Numerical experiments have demonstrated the remarkable performance of the proposed method. In future works, we will extend the proposed model to medical image segmentation.

Acknowledgments. The authors would like to thank Dr. Adrián Martín for providing the source code of the MAP model and Dr. Liyuan Chen for providing the source codes of the sMAP model, the Getreuer's model, and the Chen-Zeng model. The authors would like to thank Dr. Huan Yang for providing the source codes of the TNRD. The authors also would like to thank Dr. Yue Lu for helpful discussions.

Appendix A. Convergence of non-convex ADMM.

The convergence of non-smooth and non-convex ADMM, including our algorithm (4.4)-(4.7) as a special case, has been analyzed in [33]. Here, we apply this convergence result to our approach. We will first recall the non-smooth and non-convex optimization problem considered in [33], its ADMM solver, and the five critical assumptions on the problem. We will then show that our model (4.3) can be reformulated into the optimization problem in [33] and verify that our model satisfies the Assumptions A1–A5 in [33], so the convergence of our algorithm will be straightforward.

In [33], Wang et al. consider the following non-convex and non-smooth optimization problem

$$
\text{(A.1)} \qquad \begin{cases} \min\limits_{\boldsymbol{x}=(x_0,x_2,\ldots,x_M),y} \phi(\boldsymbol{x},y) := \xi(\boldsymbol{x}) + \sum\limits_{m=0}^{M} \zeta_m(x_m) + \eta(y), \\ \text{s.t.} \qquad \sum\limits_{m=0}^{M} A_m x_m + By = 0, \end{cases}
$$

where $\xi : \mathbb{R}^{p_0} \times \cdots \times \mathbb{R}^{p_M} \to \mathbb{R} \cup \{+\infty\}$, $\zeta_m : \mathbb{R}^{p_m} \to \mathbb{R} \cup \{+\infty\}, m = 0, 1, \ldots, M$, and $\eta : \mathbb{R}^q \to \mathbb{R} \cup \{+\infty\}$ are lower semi-continuous functions; $A_m \in \mathbb{R}^{s \times p_m}$ and $B \in \mathbb{R}^{s \times q}$ are given matrices.

The augmented Lagrangian function of optimization problem (A.1) is defined as

$$
\mathcal{L}_r(\boldsymbol{x}, y; \lambda) = \phi(\boldsymbol{x}, y) + \left\langle \lambda, \sum_{m=0}^{M} A_m x_m + By \right\rangle + \frac{r}{2} \left\| \sum_{m=0}^{M} A_m x_m + By \right\|^2.
$$

Wang et al. [33] presented the non-convex ADMM for (A.1) as follows:

$$
\text{(A.2)} \qquad \begin{cases} x_m^{k+1} = \arg\min\limits_{x_m} \mathcal{L}_r(x_{<m}^{k+1}, x_m, x_{>m}^k, y^k; \lambda^k), m = 0, 1, \ldots, M; \\ y^{k+1} = \arg\min\limits_{y} \mathcal{L}_r(\boldsymbol{x}^{k+1}, y; \lambda^k); \\ \lambda^{k+1} = \lambda^k + r\left(\sum\limits_{m=0}^{M} A_m x_m^{k+1} + By^{k+1} \right). \end{cases}
$$

Under five critical assumptions on the objective functions and matrices, Wang et al. [33] proved that the iterative scheme (A.2) converges. We recall the five assumptions in the following.

A1 (**coercivity**) The objective function $\phi(\boldsymbol{x}, y)$ is coercive over the feasible set $\{(\boldsymbol{x}, y) : \sum_{m=0}^{M} A_m x_m + By = 0\}$;

A2 (**feasibility**) Denote $\boldsymbol{A} = [A_0, A_1, \ldots, A_M]$. $\mathrm{Im}(\boldsymbol{A}) \subseteq \mathrm{Im}(B)$, where $\mathrm{Im}(\cdot)$ returns the image of a matrix;

A3 (**Lipschitz sub-minimization path**)

(a) For any fixed $\boldsymbol{x}$, $\arg\min_y \{\phi(\boldsymbol{x}, y) : By = z\}$ has a unique minimizer. $H : \mathrm{Im}(B) \to \mathbb{R}^q$ defined by $H(z) \triangleq \arg\min_y \{\phi(\boldsymbol{x}, y) : By = z\}$ is a Lipschitz continuous map.

(b) For $m = 0, 1, \ldots, M$ and any $x_{<m}$, $x_{>m}$ and y,

$$
\arg\min_{x_m} \{\phi(x_{<m}, x_m, x_{>m}, y) : A_m x_m = z\}
$$

has unique minimizer and $\Theta_m : \mathrm{Im}(A_m) \to \mathbb{R}^{p_m}$ defined by

$$
\Theta_m(z) := \arg\min_{x_m} \{\phi(x_{<m}, x_m, x_{>m}, y) : A_m x_m = z\}
$$

is a Lipschitz continuous map.

Moreover, the above Θ_m and H have a universal Lipschitz constant $\bar{L} > 0$.

A4 (**objective-ζ regularity**)

(a) $\xi(\boldsymbol{x})$ is Lipschitz differentiable with constant L_ξ,

(b) ζ_0 is lower semi-continuous, ζ_m is restricted prox-regular [33, Definition 2 and Propostion 1] for $m = 1, 2, \ldots, M$;

A5 (**objective-η regularity**) $\eta(y)$ is Lipschitz differentiable with constant L_η.

Wang et al. [33] stated that A1 holds when the objective function is coercive and A2 and A3 holds when A_m and B have full column rank.

We now give the main convergence theorem in [33] as follows.

THEOREM A.1 ([33], Theorem 1). *Suppose A1-A5 hold. Algorithm (A.2) converges subsequently for any sufficiently large r, that is, starting from any $x_1^0, \ldots, x_M^0$, y^0, λ^0, it generates a sequence that is bounded, has at least one limit point, and that each limit point $(\boldsymbol{x}^*, y^*, \lambda^*)$ is a stationary point of $\mathcal{L}_r$.*

In addition, if $\mathcal{L}_r$ is a Kurdyka-Lojasiewicz function [3], then $(\boldsymbol{x}^k, y^k, \lambda^k)$ converges globally to the unique limit point $(\boldsymbol{x}^*, y^*, \lambda^*)$.

Next, we apply the above convergence results for our method. We denote

$$\Psi_0(\boldsymbol{v}) := \mathbb{I}_{\mathcal{K}}(\boldsymbol{v}), \quad \Psi_1(u) := \mathrm{TV}(u) + \frac{\beta}{2}\|u-g\|^2, \quad \Phi(\boldsymbol{n}) := \frac{\alpha}{2}\|\boldsymbol{n}\|^2.$$

Then the propose model (4.3) can be reformulated as

$$\text{(A.3)} \qquad \begin{cases} \min\limits_{\boldsymbol{v},u,\boldsymbol{n}} \Psi_0(\boldsymbol{v}) + \Psi_1(u) + \Phi(\boldsymbol{n}), \\ \text{s.t. } \boldsymbol{Q}u - \boldsymbol{v} + \boldsymbol{n} = 0. \end{cases}$$

Thus, the problem (4.3) is a special case of problem (A.1) with the following specifications:

1) $x_0 := \boldsymbol{v}$, $x_1 := u$, and $y := \boldsymbol{n}$;
2) $\xi(\boldsymbol{x}) := 0$, $\zeta_0(x_0) := \Psi_0(\boldsymbol{v})$, $\zeta_1(x_1) := \Psi_1(u)$, and $\eta(y) := \Phi(\boldsymbol{n})$;
3) $A_0 := -I$, $A_1 := \boldsymbol{Q}$, and $B := \boldsymbol{I}$.

It can be verified that the model (A.3) satisfies the Assumptions A1–A5 as follows. Note that the set $\mathcal{K}$ is closed and bounded, thus $\Psi_0(\boldsymbol{v}) = \mathbb{I}_{\mathcal{K}}(\boldsymbol{v})$ is coercive and lower semi-continuous [30]. The convexity of $\Psi_1(u)$ implies it a restricted prox-regular function [33, Definition 2 and Propostion 1]. Let us verify Assumptions A1-A5 in [33]. Assumption A1 holds because of the coercivity of $\Psi_0(\boldsymbol{v})$, $\Psi_1(u)$, and $\Phi(\boldsymbol{n})$. Assumptions A2 and A3 hold for $\boldsymbol{Q}$, $-I$, and I being full column rank. Assumption A4 holds for $\Psi_0(\boldsymbol{v})$ is lower semi-continuous and $\Psi_1(u)$ is restricted prox-regular. Assumption A5 holds because $\Phi(\boldsymbol{n})$ is Lipschitz differentiable.

Moreover, one readily sees that $\mathbb{I}_{\mathcal{K}}(\boldsymbol{v})$ is an indicator function of the semi-algebraic set, $\mathrm{TV}(u)$ is semi-algebraic, and $\frac{\beta}{2}\|u-g\|^2 + \frac{\alpha}{2}\|\boldsymbol{n}\|^2 + \langle \boldsymbol{\lambda}, \boldsymbol{Q}u + \boldsymbol{n} - \boldsymbol{v}\rangle + \frac{r}{2}\|\boldsymbol{Q}u + \boldsymbol{n} - \boldsymbol{v}\|^2$ is a polynomial function. Therefore, their sum $\mathcal{L}$ is semi-algebraic [3]. Since semi-algebraic functions satisfy Kurdyka-Łojasiewicz (KL) inequality [3], $\mathcal{L}$ is a KL function.

REFERENCES

[1] S. Aja-Fernandez, C. Alberola-Lopez, and C. Westin, *Noise and signal estimation in magnitude MRI and Rician distributed images: A LMMSE approach*, IEEE Transactions on Image Processing, 17 (2008), pp. 1383–1398. (Cited on pages 1, 5, and 16.)

[2] S. Aja-Fernández and G. Vegas-Sánchez-Ferrero, *Statistical Analysis of Noise in MRI*, Springer International Publishing, 2016. (Cited on pages 1 and 3.)

[3] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the kurdyka-lojasiewicz inequality*, Mathematics of Operations Research, 35 (2010), pp. 438–457. (Cited on page 22.)

[4] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problemss*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 183–202. (Cited on page 9.)

[5] M. A. Bernstein, D. M. Thomasson, and W. H. Perman, *Improved detectability in low signal-to-noise ratio magnetic resonance images by means of a phase-corrected real reconstruction: Technical Reports: Low SNR MR images*, Medical Physics, 16 (1989), pp. 813–817. (Cited on page 1.)

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends® in Machine Learning, 3 (2011), pp. 1–122. (Cited on page 7.)

[7] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 89–97. (Cited on page 9.)

[8] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, 40 (2011), pp. 120–145. (Cited on page 9.)

[9] L. Chen, Y. Li, and T. Zeng, *Variational image restoration and segmentation with Rician noise*, Journal of Scientific Computing, 78 (2019), pp. 1329–1352. (Cited on page 4.)

[10] L. Chen and T. Zeng, *A convex variational model for restoring blurred images with large Rician noise*, Journal of Mathematical Imaging and Vision, 53 (2015), pp. 92–111. (Cited on pages 2, 4, 6, and 15.)

[11] P. Coupé, J. V. Manjón, E. Gedamu, D. Arnold, M. Robles, and D. L. Collins, *Robust Rician noise estimation for MR images*, Medical Image Analysis, 14 (2010), pp. 483–493. (Cited on page 16.)

[12] Y. Dong and T. Zeng, *A convex variational model for restoring blurred images with multiplicative noise*, SIAM Journal on Imaging Sciences, 6 (2013), pp. 1598–1625. (Cited on page 4.)

[13] E. Esser, X. Zhang, and T. F. Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 1015–1046. (Cited on page 9.)

[14] A. Foi, *Noise estimation and removal in MR imaging: The variance-stabilization approach*, in 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE, Mar 2011, pp. 1809–1814. (Cited on page 1.)

[15] B. Gao, X. Liu, X. Chen, and Y.-x. Yuan, *A new first-order algorithmic framework for optimization problems with orthogonality constraints*, SIAM Journal on Optimization, 28 (2018), pp. 302–332. (Cited on page 7.)

[16] P. Getreuer, M. Tong, and L. A. Vese, *A variational model for the restoration of MR images corrupted by blur and Rician noise*, in Advances in Visual Computing, B. G. et al., ed., vol. 6938 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2011, pp. 686–698. (Cited on pages 2, 4, 6, 14, and 15.)

[17] R. Glowinski, *ADMM and Non-convex Variational Problems*, Scientific Computation, Springer International Publishing, 2016, pp. 251–299. (Cited on page 7.)

[18] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*, Society for Industrial and Applied Mathematics, 1989. (Cited on page 7.)

[19] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 323–343. (Cited on page 9.)

[20] R. M. Henkelman, *Measurement of signal intensities in the presence of noise in MR images: Technical reports: Signal intensities in MR image noise*, Medical Physics, 12 (1985), pp. 232–233. (Cited on pages 1, 3, and 6.)

[21] R. Lai and S. Osher, *A splitting method for orthogonality constrained problems*, Journal of Scientific Computing, 58 (2014), pp. 431–449. (Cited on pages 6 and 7.)

[22] S. Li, J. Zhou, D. Liang, and Q. Liu, *MRI denoising using progressively distribution-based neural network*, Magnetic Resonance Imaging, 71 (2020), pp. 55–68. (Cited on page 2.)

[23] J. Lu, J. Tian, L. Shen, Q. Jiang, X. Zeng, and Y. Zou, *Rician noise removal via a learned dictionary*, Mathematical Problems in Engineering, 2019 (2019), pp. 1–13. (Cited on page 2.)

[24] J. V. Manjón, J. Carbonell-Caballero, J. J. Lull, G. García-Martí, L. Martí-Bonmatí, and M. Robles, *MRI denoising using non-local means*, Medical Image Analysis, 12 (2008), pp. 514–523. (Cited on page 1.)

[25] J. V. Manjón and P. Coupe, *MRI denoising using deep learning*, in Patch-Based Techniques in Medical Imaging, W. Bai, G. Sanroma, G. Wu, B. C. Munsell, Y. Zhan, and P. Coupé, eds., Lecture Notes in Computer Science, Springer International Publishing, 2018, pp. 12–19. (Cited on page 1.)

[26] A. Martín, J. F. Garamendi, and E. Schiavi, *Iterated Rician denoising*, in The 2011 International Conference on Image Processing, Computer Vision, and Pattern Recognition

(IPCV'11), Jul 2011, pp. 959–963. (Cited on pages 2, 4, and 6.)

[27] A. Martín, E. Schiavi, and S. Segura de León, *On 1-Laplacian elliptic equations modeling magnetic resonance image Rician denoising*, Journal of Mathematical Imaging and Vision, 57 (2017), pp. 202–224. (Cited on pages 2, 14, and 15.)

[28] G. McGibney and M. R. Smith, *An unbiased signal-to-noise ratio measure for magnetic resonance images*, Medical Physics, 20 (1993), pp. 1077–1078. (Cited on page 1.)

[29] R. Nowak, *Wavelet-based Rician noise removal for magnetic resonance imaging*, IEEE Transactions on Image Processing, 8 (1999), pp. 1408–1419. (Cited on pages 1 and 6.)

[30] R. T. Rockafellar and R. Wets, *Variational Analysis*, vol. 317 of Grundlehren der Mathematischen Wissenschaften, Springer, 1998. (Cited on page 22.)

[31] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena, 60 (1992), pp. 259–268. (Cited on pages 2 and 8.)

[32] J. Sijbers and A. J. den Dekker, *Maximum likelihood estimation of signal amplitude and noise variance from MR data*, Magnetic Resonance in Medicine, 51 (2004), pp. 586–594. (Cited on page 1.)

[33] Y. Wang, W. Yin, and J. Zeng, *Global convergence of ADMM in nonconvex nonsmooth optimization*, Journal of Scientific Computing, 78 (2019), pp. 29–63. (Cited on pages 2, 7, 9, 20, 21, and 22.)

[34] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, *Image quality assessment: from error visibility to structural similarity*, IEEE Transactions on Image Processing, 13 (2004), pp. 600–612. (Cited on page 11.)

[35] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, 1922. (Cited on page 3.)

[36] Z. Wen and W. Yin, *A feasible method for optimization with orthogonality constraints*, Mathematical Programming, 142 (2013), pp. 397–434. (Cited on page 7.)

[37] N. Wiest-Daesslé, S. Prima, P. Coupé, S. P. Morrissey, and C. Barillot, *Rician noise removal by non-local means filtering for low signal-to-noise ratio MRI: Applications to DT-MRI*, in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008, D. Metaxas, L. Axel, G. Fichtinger, and G. Székely, eds., vol. 5242 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2008, pp. 171–179. (Cited on page 1.)

[38] C. Wu and X. Tai, *Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 300–339. (Cited on pages 7 and 9.)

[39] H. Yang, H. Li, and Y. Duan, *Adaptive trainable non-linear reaction diffusion for Rician noise removal*, IET Image Processing, (2020). (Cited on pages 2 and 18.)

[40] X. You, N. Cao, H. Lu, M. Mao, and W. Wanga, *Denoising of MR images with Rician noise using a wider neural network and noise range division*, Magnetic Resonance Imaging, 64 (2019), pp. 154–159. (Cited on pages 2 and 18.)

[41] J. Yuan, *An improved variational model for denoising magnetic resonance images*, Computers & Mathematics with Applications, 76 (2018), pp. 2212–2222. (Cited on page 2.)

[42] F. Zhao, D. C. Noll, J. Nielsen, and J. A. Fessler, *Separate magnitude and phase regularization via compressed sensing*, IEEE Transactions on Medical Imaging, 31 (2012), pp. 1713–1723. (Cited on page 5.)