METHODS AND APPLICATIONS OF ANALYSIS.
Vol. 21, No. 1, pp. 151–176, March 2014

007

A VECTORIAL TOTAL VARIATION MODEL FOR DENOISING HIGH ANGULAR RESOLUTION DIFFUSION IMAGES CORRUPTED BY RICIAN NOISE*

M. TONG†, Y. KIM‡, L. ZHAN§, G. SAPIRO¶, C. LENGLET‖, B. A. MUELLER**, P. M. THOMPSON§, AND L. A. VESE†

Abstract. The presence of noise in High Angular Resolution Diffusion Imaging (HARDI) data of the brain can limit the accuracy with which fiber pathways of the brain can be extracted. In this work, we present a variational model to denoise HARDI data corrupted by Rician noise. We formulate a minimization model composed of a data fidelity term incorporating the Rician noise assumption and a regularization term given by the vectorial total variation. Although the proposed minimization model is non-convex, we are able to establish existence of minimizers. Numerical experiments are performed on three types of data: 2D synthetic data, 3D diffusion-weighted Magnetic Resonance Imaging (DW-MRI) data of a hardware phantom containing synthetic fibers, and 3D real HARDI brain data. Experiments show that our model is effective for denoising HARDI-type data while preserving important aspects of the fiber pathways such as fractional anisotropy and the orientation distribution functions.

Key words. Total variation, Rician noise, denoising, diffusion imaging.

AMS subject classifications. 35, 49, 65, 68U10.

1. Introduction to HARDI images. High Angular Resolution Diffusion Imaging (HARDI) [24, 25] is a modality of Magnetic Resonance (MR) imaging used in reconstructing fibers in the brain. Water diffusion in the brain causes attenuation of MR signals. Since water diffuses preferentially in the direction of fiber pathways, MR imaging presents a method for tracking fibers in the brain. More specifically, the Stejskal-Tanner equation [22] gives a relationship between water diffusion and corresponding MR signal

$$S(x,\theta,\phi) = S_0(x)\exp\left(-b\cdot d(x,\theta,\phi)\right), \tag{1}$$

where $S(x,\theta,\phi)$ and $d(x,\theta,\phi)$ are the MR signal and corresponding diffusion at position x and direction on the sphere given by

$$(\cos(\theta)\sin(\phi), \sin(\theta)\sin(\phi), \cos(\phi)).$$

$d(x,\theta,\phi)$ is also known as the spherical apparent diffusion coefficient (sADC). $S_0(x)$ is the MR signal at x when no diffusion gradient is applied, and $b \geq 0$ is a parameter used in collecting the data.

*Received February 22, 2013; accepted for publication October 23, 2013.

†Department of Mathematics, University of California, Los Angeles, CA 90095-1555, USA ({meltong; lvese}@math.ucla.edu). Contact author: L. A. Vese

‡Department of Diagnostic Radiology, Yale University, New Haven, CT 06520, USA (yunho.kim @yale.edu).

§Laboratory of Neuro Imaging, University of Southern California, Los Angeles, CA 90032, USA ({liang.zhan; thompson}@loni.usc.edu).

¶Department of Electrical and Computer Engineering, Duke University, Box 90291, Durham, NC 27708, USA (guillermo.sapiro@duke.edu).

‖Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, MN 55455, USA (clenglet@umn.edu).

**Department of Psychiatry, University of Minnesota, Minneapolis, MN 55454, USA (muell093 @umn.edu).

The true MR signal cannot be larger than the MR signal when no diffusion gradient is applied. That is, we have the constraints

$$S(x,\theta,\phi) \leq S_0(x) \quad \text{or} \quad d(x,\theta,\phi) \geq 0. \tag{2}$$

HARDI data is a collection of diffusion weighted MR images (DW-MRI), meaning that for each diffusion direction considered, we have a corresponding MR signal. As the number of diffusion directions increases, the angular resolution increases as well. The number of diffusion directions used in typical HARDI data is ≈ 100 explaining the "High Angular Resolution" part of HARDI. For more details about HARDI, we refer to [6, 11].

HARDI data is often contaminated by noise. The level of noise present varies as a function of the collection parameter b, with larger b values leading to more noise but more accurate diffusivity information [4]. Thus, there is a tradeoff between more accurate information and less noise. However, with any choice of b, the introduction of noise will likely cause violations of the constraint (2) and change important characteristics measured from the HARDI data such as fractional anisotropy (FA) [16, 1, 20] of the fiber pathways. Given this knowledge, it may be beneficial to denoise the HARDI data prior to the extraction of fibers.

During the DW-MR data collection process, a complex valued signal is collected and its magnitude forms the DW-MRI signal. Since the real and imaginary parts of this raw complex signal is corrupted by Gaussian distributed noise, the resulting magnitude signal will have Rician distributed noise [28, 29, 30, 31, 3].

An underlying clean signal u is said to be corrupted with Rician distributed noise if the resulting noisy signal S follows a Rician probability distribution. The Rician probability distribution is given by

$$Pr(S|u) = \frac{S}{\sigma^2} \exp\left(- \frac{S^2+u^2}{2\sigma^2} \right) I_0\left(\frac{Su}{\sigma^2} \right), \tag{3}$$

where σ is the standard deviation of the noise, and I_0 is the zeroth-order modified Bessel function of the first kind [3]. In other words, if at a particular voxel, the underlying clean signal has value u, then $Pr(S|u)$ gives the probability of getting a noisy signal with value S.

As mentioned earlier, a DW-MR signal is formed by taking the magnitude of the raw complex valued signal collected by a DW-MR imaging machine. If the collection process is perfect, then we start with the machine collected complex valued signal

$$z = z_{real} + i z_{imag} = \nu \cos(\theta) + i\nu \sin(\theta),$$

where ν is the magnitude of z and θ is the phase of z. The resulting DW-MR signal is simply the magnitude ν. However, in real applications, the collection process corrupts z with noise. More specifically, the real and imaginary components of z are corrupted with Gaussian distributed noise with some standard deviation σ [28, 29, 30, 31, 3]:

$$z = z_{real} + i z_{imag} \rightarrow (z_{real} + G(\sigma)) + i(z_{imag} + G(\sigma)).$$

The corrupted DW-MR signal is the magnitude of the corrupted complex valued signal:

$$\text{magnitude} = \sqrt{(z_{real} + G(\sigma))^2 + (z_{imag} + G(\sigma))^2}.$$

As a result, a DW-MR signal is the magnitude of two independent normal random variables, with distributions $N(\nu\cos\theta, \sigma^2)$ and $N(\nu\sin\theta, \sigma^2)$. It is easy to see from the discussion above that the resulting DW-MR signal is corrupted by Rician distributed noise.

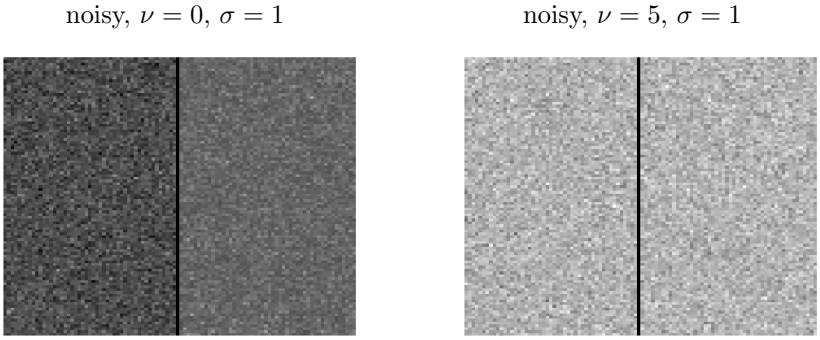

FIG. 1. *Comparison of Gaussian and Rician noise with $\sigma = 1$ on input values $\nu = 0$ (left square) and $\nu = 5$ (right square). Left of each square: additive Gaussian noise; right of each square: Rician noise. For $\nu = 0$, mean of Gaussian: -0.0041, mean of Rician: 1.2550. For $\nu = 5$, mean of Gaussian: 5.0155, mean of Rician: 5.1066.*

In Fig. 1, we start with two squares of constant intensity $\nu = 0$ and $\nu = 5$. We add Gaussian noise to the left half of the squares and Rician noise to the right half of the squares, where the standard deviation of the noise $\sigma = 1$ in all cases. We note that for input values ν that are small compared to the standard deviation σ, we see a difference between Gaussian and Rician noise.

We compare the probability distributions for these two cases in Fig. 2. When the standard deviation σ of the Rician probability distribution is small compared to ν (e.g. $\nu > 2\sigma$), the Rician probability distribution is approximately Gaussian. This case occurs when the signal to noise ratio (SNR) is high (low noise level). However, when σ is larger or on the order of the value of ν (e.g. $\nu \leq \sigma$), the Gaussian probability distribution is a bad approximation to the Rician one. This occurs when the SNR is low and the noise level is high. Therefore, a Gaussian distribution is a good approximation to the Rician distribution in areas corrupted with small amounts of noise but not in areas corrupted with large amounts of noise.

As noted in Fig. 1, for both these examples, the mean of the images corrupted with Rician noise is larger than the mean of the images corrupted with Gaussian noise. In general, let us suppose that the true intensity at a pixel of an image has value ν. If we consider many such identical pixels and independently add Gaussian noise, the mean value of the corrupted intensity values would approach ν. If we independently add Rician noise, the mean value of the corrupted intensity values would approach [30, 33]

$$\text{Rician}_{\text{mean}}(\nu,\sigma) = \sigma\Big(\frac{\pi}{2}\Big)^{1/2}\exp\Big(-\frac{\nu^2}{4\sigma^2}\Big)\times\left(\Big(1+\frac{\nu^2}{2\sigma^2}\Big)I_0\Big(\frac{\nu^2}{4\sigma^2}\Big)+\Big(\frac{\nu^2}{2\sigma^2}\Big)I_1\Big(\frac{\nu^2}{4\sigma^2}\Big)\right).$$

Here, I_1 is the first-order modified Bessel function of the first kind. In Fig. 3, we illustrate the difference $(\text{Rician}_{\text{mean}}(\nu,\sigma) - \text{Gaussian}_{\text{mean}}(\nu,\sigma))$. For any true intensity ν and standard deviation σ, the mean of the corrupted intensity values due to Rician

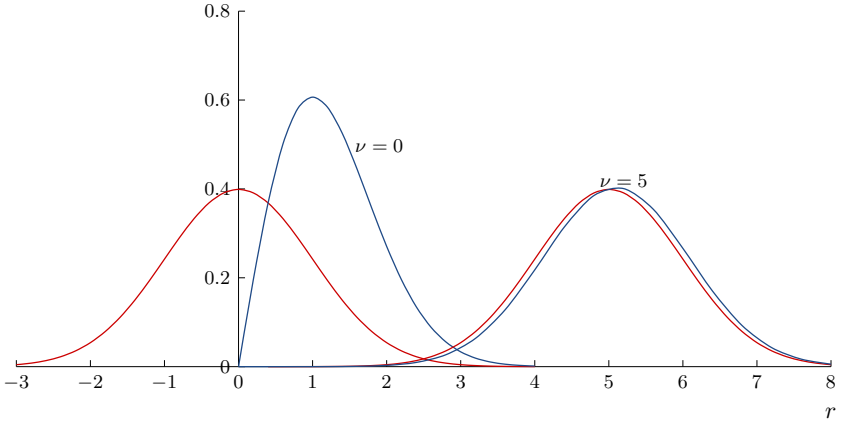

FIG. 2. *Comparison of Rician (blue) and Gaussian (red) probability distribution functions for given ν and standard deviation $\sigma = 1$.*

noise will be larger than the mean of the corrupted intensity values due to Gaussian noise. With an underlying true image, we would have many pixels of varying intensity values. However, if we repeat the argument above for each pixel and its corresponding intensity, hopefully one can see that the mean value of the Rician noise corrupted mean image is larger than the mean value of the Gaussian noise corrupted mean image. This allows us to furthermore conclude that the mean of the means of the Rician noise corrupted images is larger than the mean of the means of the Gaussian noise corrupted images.

In [13], the authors propose a variational restoration model to denoise HARDI data that utilizes a Gaussian probability distribution as an approximation to the Rician probability distribution, assuming high signal to noise ratio. This leads to a simple L^2 fidelity term, where the standard deviation of the noise σ does not appear explicitly. In [9, 10], an L^1 fidelity term is utilized, which is more suitable than an L^2 fidelity term while being simpler to model than Rician noise. Again, σ does not appear explicitly in the variational denoising model.

Recently, [14, 15] look for sparse representations of HARDI signals using spherical ridglets and learned dictionaries, respectively. These approaches are similar to modeling HARDI signals using high-order spherical harmonics series [5] in that the goal is to represent the HARDI signal as a linear combination of some basis functions, but differ in that a *sparse* representation is sought.

In [3], the authors propose a related anisotropic diffusion model to denoise diffusion tensor imaging (DTI) data corrupted by Rician noise. The Rician noise assumption is incorporated in the data fidelity term but without imposing additional necessary constraints on the denoised vector-valued signal.

Of the previous works, the most relevant to ours are [3, 13, 9, 10]. Our proposed model is different from [3, 13] in that we incorporate the data acquisition model (1) and from [9, 10] in that we impose a more accurate noise model, the Rician distributed one. A short and very preliminary version of this work was presented at the 2012 IEEE International Symposium on Biomedical Imaging [23]. The present work can be seen as a continuation of [7, 8], where total variation restoration of scalar-valued

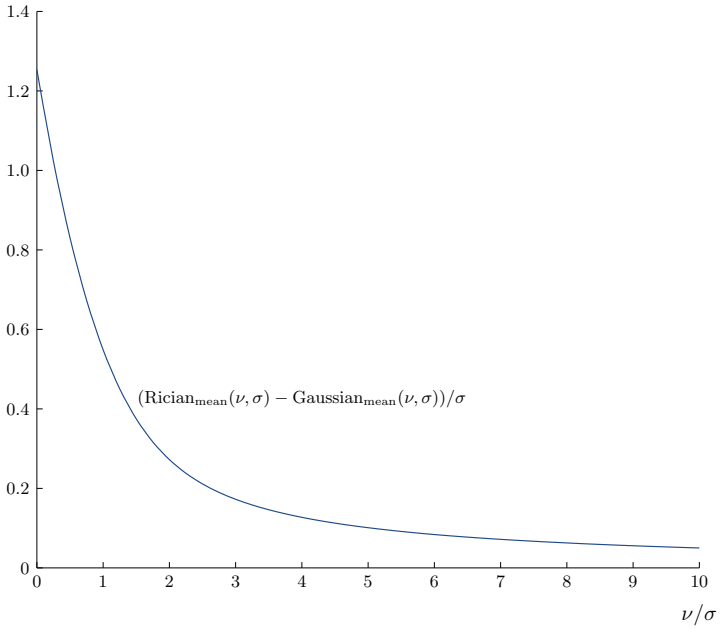

FIG. 3. *Difference between Rician mean and Gaussian mean as a function of input ν and standard deviation σ. This figure shows that at any input ν, the Rician mean is larger than the Gaussian mean, and as $\nu/\sigma \to \infty$, the Rician mean approaches the Gaussian mean.*

MR images corrupted by Rician noise has been proposed and analyzed in theory and practice. However, unlike [7, 8], this work proposes a restoration model that works directly with the sADC instead of the corresponding MR images.

2. Proposed variational HARDI denoising model. We propose a variational denoising model motivated by a *maximum a posteriori* (MAP) estimate which attempts to find the most likely sADC, $d = (d_1, ..., d_n)$ to have generated a given observed noisy signal $S = (S_1, ..., S_n)$. Here, n represents the number of diffusion directions and d_i and S_i represent the sADC and noisy signal in direction i on the sphere. The unknown d is the minimizer of an energy composed of two terms:

(4)
$$\inf_{d:\Omega\to\mathbb{R}^n} \left\{ F(d) = \int_\Omega |D(d)| + \lambda \int_\Omega \sum_{i=1}^{n} \left(-\log I_0\left(\frac{S_i S_0 e^{-P(d_i)}}{\sigma^2}\right) + \frac{(S_0 e^{-P(d_i)})^2}{2\sigma^2}\right) dx \right\},$$

where $\Omega \subset \mathbb{R}^N$ ($N = 2$ for 2D synthetic HARDI data and $N = 3$ for volumetric real HARDI data), $\lambda > 0$ is a tuning parameter, σ is the standard deviation of the noise, S_i is the observed noisy HARDI signal measured in the spherical direction i, and d_i is the denoised unknown sADC in the same direction. We denote by $u_i = S_0 \exp(-P(d_i))$ the recovered denoised HARDI signal (here, we use the data acquisition model (1) with $b = 1$, which rescales d; the tuning parameter λ accounts for this rescaling). P is a projection operator given by

(5)
$$P(z) = \begin{cases} z & \text{if } z \geq 0 \\ 0 & \text{if } z < 0, \end{cases}$$

which is introduced to satisfy the constraint $d_i \geq 0$, and $\int_\Omega |D(d)|$ is the vectorial total variation of d defined by

$$\int_\Omega |D(d)| = \sup\Big\{ \sum_{i=1}^{n} \int_\Omega d_i \mathrm{div}(\phi_i) dx \;\Big|\; \phi = (\phi_1, ..., \phi_n) \text{ and } \phi_i \in C_c^1(\Omega; \mathbb{R}^3),\ ||\phi||_\infty \leq 1 \Big\},$$

where $\|\phi\|_\infty = \operatorname{ess\,sup}_{x\in\Omega} \sqrt{\phi_1^2(x) + \cdots + \phi_n^2(x)}$, which reduces to

$$\int_\Omega |D(d)| = \int_\Omega \sqrt{\sum_{i=1}^{n} |\nabla d_i|^2} dx. \tag{6}$$

when d is differentiable.

The first term of the energy, $\int_\Omega |D(d)|$, is a prior on d which provides edge-preserving regularization, making the minimization problem well-posed. This prior is suitable for denoising as it provides smoothing effects while preserving edges. This is the vectorial extension of the well known total variation regularization used in image processing and introduced in [18].

The second term of the energy is the fidelity term. Up to a constant, it is equal to the sum of $-\log Pr(S_i \mid S_0 \exp(-P(d_i)))$ over Ω and diffusion directions $i = 1, ...n$, where $Pr(\cdot \mid \cdot)$ is the Rician probability density function. This term encourages

$$Pr(S_i \mid S_0 \exp(-P(d_i)))$$

to be large so that the contribution of

$$-\log Pr(S_i \mid S_0 \exp(-P(d_i)))$$

is small. This is equivalent to finding d_i such that the probability of generating the noisy S_i given this d_i is large, assuming that the noisy signal S_i is the result of an underlying clean signal $u_i = S_0 \exp(-P(d_i))$ corrupted with Rician distributed noise.

3. Existence of a minimizer. Although the second (fidelity) term in the proposed energy (4) to be minimized is not convex, we can still establish in this section existence of minimizers d.

Throughout this section, Ω is a bounded open subset of $\mathbb{R}^N$ with Lipschitz boundary $\partial\Omega$ for $N \geq 2$ unless otherwise stated. We will also use only in this section different notations for the input data, f_i for S_i and f for S, for all $1 \leq i \leq n$.

DEFINITION 1. *Let $d = (d_1, \ldots, d_n) \in L^1(\Omega; \mathbb{R}^n)$. We define*

$$\begin{aligned} &|Dd|(\Omega) \\ &= \sup\Big\{ \sum_{i=1}^{n} \int_\Omega d_i \,\mathrm{div}(\psi_i) \;\Big|\; \psi = (\psi_1, \ldots, \psi_n) \text{ and } \psi_i \in C_c^1(\Omega; \mathbb{R}^N),\ \|\psi\|_\infty \leq 1 \Big\}. \end{aligned}$$

The space of functions of bounded variation $BV(\Omega; \mathbb{R}^n)$ is defined by

$$BV(\Omega; \mathbb{R}^n) = \Big\{ d \in L^1(\Omega; \mathbb{R}^n) \Big|\ |Dd|(\Omega) < \infty \Big\}$$

with the norm

$$\|d\|_{BV(\Omega;\mathbb{R}^n)} = |Dd|(\Omega) + \sum_{i=1}^{n} \|d_i\|_{L^1(\Omega)}.$$

We will use $BV(\Omega)$ *when* $n = 1$.

Consider the following minimization problem

$$\inf_{d\in BV(\Omega;\mathbb{R}^n)} \Big\{ F(d) = |Dd|(\Omega) - \lambda \sum_{i=1}^{n} \int_\Omega \Big[\log\Big(\frac{f_i}{\sigma^2}\Big) + \log I_0\Big(\frac{f_i f_0 e^{-P(d_i)}}{\sigma^2}\Big) - \frac{f_i^2 + (f_0 e^{-P(d_i)})^2}{2\sigma^2}\Big] dx \Big\},$$

where log is the natural logarithm and I_0 is the modified Bessel function of the first kind of order 0. Note that if d is differentiable, then

$$|Dd|(\Omega) = \int_\Omega \sqrt{|Dd_1(x)|^2 + \cdots + |Dd_n(x)|^2} dx.$$

The minimization problem above is equivalent to

$$\inf_{d\in BV(\Omega;\mathbb{R}^n)} \Big\{ F(d) = |Dd|(\Omega) - \lambda \sum_{i=1}^{n} \int_\Omega \Big[\log I_0\Big(\frac{f_i f_0 e^{-P(d_i)}}{\sigma^2}\Big) - \frac{f_0^2 e^{-2P(d_i)}}{2\sigma^2}\Big] dx \Big\}. \tag{7}$$

If we define

$$G(x, A) = \log I_0(x) - Ax^2,$$

then (7) can be rewritten as

$$\inf_{d\in BV(\Omega;\mathbb{R}^n)} \Big\{ F(d) = |Dd|(\Omega) - \lambda \sum_{i=1}^{n} \int_\Omega G\Big(\frac{f_i f_0 e^{-P(d_i)}}{\sigma^2}, \frac{\sigma^2}{2f_i^2}\Big) dx \Big\}. \tag{8}$$

We will now investigate the function $G(x, A)$. Note that

$$\frac{\partial G}{\partial x}(x, A) = \frac{I_1(x)}{I_0(x)} - 2Ax \quad \text{and} \quad \frac{\partial^2 G}{\partial x^2}(x, A) = \frac{I_1'(x)I_0(x) - I_1^2(x)}{I_0^2(x)} - 2A.$$

Since $\frac{d}{dx} I_1(x) = \frac{1}{2}(I_0(x) + I_2(x))$, if $0 \le x << 1$ and $0 < A < \frac{1}{4}$, then

$$\frac{\partial^2 G}{\partial x^2}(x, A) = \frac{1}{2} + \frac{I_2(x)}{2I_0(x)} - \Big(\frac{I_1(x)}{I_0(x)}\Big)^2 - 2A > 0.$$

Then, since $\frac{\partial G}{\partial x}(0, A) = 0$, if $0 < A < \frac{1}{4}$, then

$$\frac{\partial G}{\partial x}(x, A) \ge 0 \quad \text{and} \quad \frac{\partial G}{\partial x}(x, A) \quad \text{is strictly increasing for} \quad 0 \le x << 1,$$

which also implies that

$$G(x, A) \ge 0 \quad \text{and} \quad G(x, A) \quad \text{is strictly increasing for} \quad 0 \le x << 1.$$

THEOREM 1. *Given nonnegative functions* $f_0, f_1, \ldots, f_n$ *on* Ω, *we assume that* $f_i \in L^\infty(\Omega)$ *for* $i = 0, 1, \ldots, N$ *and*

$$\inf_{i=1,2,\ldots,n} \operatorname{ess\,inf}_{x\in\Omega} f_i(x)^2 > 2\sigma^2. \tag{9}$$

Then (8) *has a minimizer in* $BV(\Omega; \mathbb{R}^n)$.

Proof. Let

$$M = \sup_{i=0,1,\dots,n} ||f_i||_\infty < \infty.$$

First of all, note that given any $A > 0$, $B > 0$, there exists $K > 0$ such that

$$|G(x, A)| < K \quad \text{for} \quad 0 \le x < B.$$

Hence, the functional F in (8) is bounded below and the infimum exists. Let

$$\{d^k = (d_1^k, d_2^k, \dots, d_n^k)\}_{k\in\mathbb{N}}$$

be a minimizing sequence. Since

$$F(P(d)) \le F(d) \quad \text{for any } d = (d_1, \dots, d_n),$$

we may assume that

$$d_i^k \ge 0 \quad \text{for } i = 1, \dots, n, \text{ and } k \in \mathbb{N}.$$

Note that there exists $0 < A < \frac{1}{4}$ such that for all i,

$$\text{ess sup}_{x\in\Omega} \frac{\sigma^2}{2f_i^2(x)} < A,$$

and, hence, there exists $\delta > 0$ such that

$$-G(y, A) \quad \text{is decreasing on} \quad [0, \delta].$$

As a matter of fact,

$$-G(y, B) \quad \text{is decreasing} \quad \text{for any} \quad B < A \quad \text{on} \quad [0, \delta].$$

Therefore, if we choose $M_1 > 0$ satisfying

$$\frac{M^2 e^{-M_1}}{\sigma^2} < \delta, \tag{10}$$

then it is easy to see that

$$|D(d \wedge M_1)|(\Omega) \le |Dd|(\Omega)$$

and

$$-G\Big(\frac{f_i f_0 e^{-P(d_i\wedge M_1)}}{\sigma^2}, \frac{\sigma^2}{2f_i^2}\Big) \le -G\Big(\frac{f_i f_0 e^{-P(d_i)}}{\sigma^2}, \frac{\sigma^2}{2f_i^2}\Big),$$

where $(d \wedge M_1)(x) = (\min(d_1(x), M_1), \dots, \min(d_n(x), M_1))$. Therefore,

$$F(d \wedge M_1) \le F(d),$$

in other words,

$$\{d^k \wedge M_1\}_{k\in\mathbb{N}}$$

is a minimizing sequence that is bounded in the BV norm, $\|\cdot\|_{BV(\Omega)}$. We now define $\tilde{d}^k = d^k \wedge M_1$ for $k \in \mathbb{N}$. Then, there exists $\tilde{d} \in BV(\Omega;\mathbb{R}^n)$ such that, up to a subsequence,

$$\tilde{d}^k \to \tilde{d} \ \text{ in } \ L^1(\Omega;\mathbb{R}^n) \ \text{ as } \ k \to \infty.$$

Furthermore, up to a subsequence, it is also true that

$$\lim_{k\to\infty} \tilde{d}^k(x) = \tilde{d}(x) \ \text{ a.e. } x \in \Omega.$$

By the lower-semicontinuity property of the total variation, we obtain

$$|D\tilde{d}|(\Omega) \le \liminf_{k\to\infty} |D\tilde{d}^k|(\Omega)$$

and by the dominant convergence theorem, we obtain

$$\lim_{k\to\infty} \int_\Omega G\Big(\frac{f_i f_0 e^{-P(\tilde{d}_i^k)}}{\sigma^2}, \frac{\sigma^2}{2f_i^2}\Big)dx = \int_\Omega G\Big(\frac{f_i f_0 e^{-P(\tilde{d}_i)}}{\sigma^2}, \frac{\sigma^2}{2f_i^2}\Big)dx.$$

Therefore,

$$F(\tilde{d}) \le \liminf_{k\to\infty} F(\tilde{d}^k) = \lim_{k\to\infty} F(d^k),$$

that is,

$$\tilde{d} \in BV(\Omega;\mathbb{R}^n) \ \text{ is a minimizer of (8).}$$

□

REMARK 1. *If*

$$\sup_{i=1,2,\dots,N} \text{ess sup}_{x\in\Omega} f_i(x)^2 < 2\sigma^2,$$

then there may not exist a minimizer of (8) *since the fact that* $G(x,A)$ *for* $A > \frac{1}{4}$ *is strictly decreasing on* $0 \le x << 1$ *and* $G(0,A) = 0$ *implies that if* $P(d_i), i = 1,\dots,n$, *are large enough for some* $d \in BV(\Omega;\mathbb{R}^n)$, *then we could have*

$$F(d+c) < F(d) \text{ for any } c \in \mathbb{R}^n_+.$$

In case that (9) is not satisfied, we have an alternative theorem below.

THEOREM 2. *Given nonnegative functions* $f_0, f_1, \dots, f_n$ *on* Ω, *we assume that* $f_i \in L^\infty(\Omega)$ *for* $i = 0,1,\dots,n$ *and*

$$\min_{i=1,2,\dots,n} \sup_{\alpha\ge 0} \int_\Omega G\Big(\frac{f_i f_0 e^{-\alpha}}{\sigma^2}, \frac{\sigma^2}{2f_i^2}\Big)dx > 0. \tag{11}$$

Then (8) *has a minimizer in* $BV(\Omega;\mathbb{R}^n)$.

REMARK 2. *Note that the supremum in* (11) *for each* $i = 1,2,\dots,n$ *is attained if it is positive since the integrand is continuous in* α *and*

$$G\Big(\frac{f_i f_0 e^{-\alpha}}{\sigma^2}, \frac{\sigma^2}{2f_i^2}\Big) \to 0 \ \text{ as } \ \alpha \to \infty.$$

Proof. The main step is to prove that there exists a bounded minimizing sequence in $BV(\Omega;\mathbb{R}^n)$. Once we have a bounded minimizing sequence, then it follows from the discussion given in Theorem 1 that there exists a minimizer in $BV(\Omega;\mathbb{R}^n)$. First of all, we let

$$M = \sup_{i=0,1,\ldots,n} ||f_i||_\infty < \infty.$$

and define for $i = 1, 2, \ldots, n$,

$$H_i(\alpha) = \int_\Omega G\Big(\frac{f_i f_0 e^{-\alpha}}{\sigma^2}, \frac{\sigma^2}{2f_i^2}\Big)dx.$$

By assumption (11) and Remark 2, for each $i = 1, 2, \ldots, n$, we have

$$\alpha_i = \text{argmax}_{\alpha\geq 0}\, H_i(\alpha) \geq 0.$$

Since $\log I_0(x)$ is nonnegative and strictly increasing for $x \geq 0$, we have, for each $i = 1, 2, \ldots, n$,

$$\begin{aligned}\Big\|G\Big(\frac{f_i f_0 e^{-\alpha}}{\sigma^2}, \frac{\sigma^2}{2f_i^2}\Big)\Big\|_{L^\infty(\Omega)} &= \Big\|\log I_0\Big(\frac{f_i f_0 e^{-\alpha}}{\sigma^2}\Big) - \frac{f_0^2 e^{-2\alpha}}{2\sigma^2}\Big\|_{L^\infty(\Omega)} \\ &\leq \log I_0\Big(\frac{M^2 e^{-\alpha}}{\sigma^2}\Big) + \frac{M^2 e^{-2\alpha}}{2\sigma^2} := \tilde{M}_\alpha.\end{aligned}$$

Note that $\tilde{M}_\alpha \leq \tilde{M}_0$ for all $\alpha \geq 0$. Hence, the functional F in (8) is bounded below and the infimum exists. Moreover,

$$F(d) < 0 \quad \text{for} \quad d = (\alpha_1, \alpha_2, \ldots, \alpha_n)$$

implies that the infimum in (8) is negative. Now let

$$\{d^k = (d_1^k, d_2^k, \ldots, d_n^k)\}_{k\in\mathbb{N}}$$

be a minimizing sequence. We may assume that

$$d_i^k \geq 0 \quad \text{for} \quad i = 1, 2, \ldots, n, \ \text{and} \ \ k \in \mathbb{N}.$$

Suppose that

$$\sup_{i=1,2,\ldots,n} \sup_{k\in\mathbb{N}} \int_\Omega d_i^k(x)dx = \infty. \tag{12}$$

Without loss of generality, we may assume that

$$\int_\Omega d_1^k(x)dx \uparrow \infty \quad \text{as} \quad k \to \infty.$$

By the Poincaré-Wirtinger inequality, we have

$$\|d_1^k - (d_1^k)_\Omega\|_{L^{\frac{N}{N-1}}(\Omega)} \leq C|Dd_1^k|(\Omega),$$

where C is a constant and

$$(d_1^k)_\Omega = \frac{1}{|\Omega|}\int_\Omega d_1^k(x)dx.$$

Let $v_k = d_1^k - (d_1^k)_\Omega$ for $k \in \mathbb{N}$. Then $\{v_k\}_{k\in\mathbb{N}}$ is a bounded sequence in $BV(\Omega)$, implying that there exists $v \in BV(\Omega)$ such that

$$v_k \to v \ \text{ in } \ L^1(\Omega). \tag{13}$$

For a given $T > 0$, we define

$$A_T^k = \{x \in \Omega : d_1^k < T\}.$$

Then, there exists $K_1 \in \mathbb{N}$ such that $(d_1^k)_\Omega > T$ if $k > K_1$ and (13) implies

$$\infty > \sup_{k>K_1} \|v_k\|_{L^1(\Omega)} \geq \int_{A_T^k} |v_k| dx \geq \int_{A_T^k} |(d_1^k)_\Omega - T| dx, \tag{14}$$

which also implies

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

where $|A_T^k|$ is the Lebesgue measure of A_T^k.

For $T >> \alpha_1$, there exits $K_2 > 0$ such that if $k > K_2$, then

$$\int_{\Omega\setminus A_T^k} G\Big(\frac{f_1 f_0 e^{-P(d_1^k)}}{\sigma^2}, \frac{\sigma^2}{2f_1^2}\Big)dx < \frac{1}{2}\int_\Omega G\Big(\frac{f_1 f_0 e^{-\alpha_1}}{\sigma^2}, \frac{\sigma^2}{2f_1^2}\Big)dx = \frac{1}{2}H_1(\alpha_1),$$

which implies that for $k > K_2$,

$$\begin{aligned}
F(d^k) &= |Dd^k|(\Omega) - \lambda\sum_{i=2}^n \int_\Omega G\Big(\frac{f_i f_0 e^{-P(d_i^k)}}{\sigma^2}, \frac{\sigma^2}{2f_i^2}\Big)dx - \lambda\int_\Omega G\Big(\frac{f_1 f_0 e^{-P(d_1^k)}}{\sigma^2}, \frac{\sigma^2}{2f_1^2}\Big)dx \\
&= |Dd^k|(\Omega) - \lambda\sum_{i=2}^n \int_\Omega G\Big(\frac{f_i f_0 e^{-P(d_i^k)}}{\sigma^2}, \frac{\sigma^2}{2f_i^2}\Big)dx \\
&\quad -\lambda\int_{A_T^k} G\Big(\frac{f_1 f_0 e^{-P(d_1^k)}}{\sigma^2}, \frac{\sigma^2}{2f_1^2}\Big)dx - \lambda\int_{\Omega\setminus A_T^k} G\Big(\frac{f_1 f_0 e^{-P(d_1^k)}}{\sigma^2}, \frac{\sigma^2}{2f_1^2}\Big)dx \\
&\geq |D\tilde{d}^k|(\Omega) - \lambda\sum_{i=2}^n \int_\Omega G\Big(\frac{f_i f_0 e^{-P(d_i^k)}}{\sigma^2}, \frac{\sigma^2}{2f_i^2}\Big)dx - \lambda\int_{A_T^k} \tilde{M}_0 dx \\
&\quad -\lambda\int_\Omega G\Big(\frac{f_1 f_0 e^{-\alpha_1}}{\sigma^2}, \frac{\sigma^2}{2f_1^2}\Big)dx + \frac{1}{2}\lambda\int_\Omega G\Big(\frac{f_1 f_0 e^{-\alpha_1}}{\sigma^2}, \frac{\sigma^2}{2f_1^2}\Big)dx \\
&= F(\tilde{d}^k) - \lambda\int_{A_T^k} \tilde{M}_0 dx + \frac{1}{2}\lambda H_1(\alpha_1),
\end{aligned}$$

where

$$\tilde{d}^k = (\alpha_1, d_2^k, \ldots, d_n^k).$$

Therefore, we obtain

$$\begin{aligned}
\lim_{k\to\infty} F(d^k) &\geq \liminf_{k\to\infty} \Big(F(\tilde{d}^k) - \lambda\int_{A_T^k} \tilde{M}_0 dx + \frac{1}{2}\lambda H_1(\alpha_1)\Big) \\
&= \liminf_{k\to\infty} F(\tilde{d}^k) + \frac{1}{2}\lambda H_1(\alpha_1) > \liminf_{k\to\infty} F(\tilde{d}^k).
\end{aligned}$$

This contradicts that

$$\{d^k = (d_1^k, d_2^k, \dots, d_n^k)\}_{k\in\mathbb{N}}$$

is a minimizing sequence. Therefore, we can conclude that

$$\sup_{i=1,2,\dots,n} \sup_{k\in\mathbb{N}} \int_\Omega d_i^k(x)dx < \infty.$$

Since we have now a minimizing sequence bounded in $BV(\Omega;\mathbb{R}^n)$, there exists a minimizer in $BV(\Omega;\mathbb{R}^n)$. □

REMARK 3. *For numerical computations on HARDI brain data, it is common to mask out the region outside the brain without information by setting $f_0, f_1, \dots, f_n$ to be equal to 0 in that region. Note that in this case, where $\Omega_1 \subset \Omega$ is the brain region of interest, then* (11) *becomes*

$$\min_{i=1,2,\dots,n} \sup_{\alpha\geq 0} \int_{\Omega_1} G\Big(\frac{f_i f_0 e^{-\alpha}}{\sigma^2}, \frac{\sigma^2}{2f_i^2}\Big)dx > 0$$

since $G \equiv 0$ in $\Omega \setminus \Omega_1$.

REMARK 4. *Theorem 2 also covers Theorem 1 because if* (9) *is true, then for each $i = 1, 2, \dots, n$,*

$$\begin{aligned}\int_\Omega G\Big(\frac{f_i f_0 e^{-\alpha}}{\sigma^2}, \frac{\sigma^2}{2f_i^2}\Big)dx &= \int_\Omega \Big[\log I_0\Big(\frac{f_i f_0 e^{-\alpha}}{\sigma^2}\Big) - \frac{\sigma^2}{2f_i^2}\Big(\frac{f_i f_0 e^{-\alpha}}{\sigma^2}\Big)\Big]dx \\ &> \int_\Omega \Big[\log I_0\Big(\frac{f_i f_0 e^{-\alpha}}{\sigma^2}\Big) - A\Big(\frac{f_i f_0 e^{-\alpha}}{\sigma^2}\Big)\Big]dx\end{aligned}$$

for some $0 < A < \frac{1}{4}$. In particular, $G(x, A)$ is strictly increasing on $[0, \delta]$ for some $\delta > 0$ with $G(0, A) = 0$. Hence, for M_1 in (10)*, we have*

$$\int_\Omega \Big[\log I_0\Big(\frac{f_i f_0 e^{-M_1}}{\sigma^2}\Big) - A\Big(\frac{f_i f_0 e^{-M_1}}{\sigma^2}\Big)\Big]dx > 0.$$

Therefore, (11) *is satisfied. However, note that* (9) *is easier to check than* (11)*.*

4. Numerical implementation. To solve the minimization problem (4) numerically, we consider the Euler-Lagrange equations in combination with the L^2 gradient descent method. The resulting evolution equations are

$$\frac{\partial d_i}{\partial t} = \lambda P'(d_i)\left(\frac{u_i^2}{\sigma^2} - \frac{I_1(\frac{S_i u_i}{\sigma^2})}{I_0(\frac{S_i u_i}{\sigma^2})}\frac{S_i u_i}{\sigma^2}\right) + \nabla\cdot\left(\frac{\nabla d_i}{\sqrt{\sum_{j=1}^n |\nabla d_j|^2}}\right) \tag{15}$$

for $i = 1, ..., n$. We use forward and backward differencing to discretize the second term and take a fully explicit scheme to obtain an iterative method from (15). As mentioned earlier, the projection operator P is used to help satisfy the constraint $d_i \geq 0$. In the evolution equations (15), we see that the contribution of P appears as a derivative factor, P', multiplying the term in the Euler-Lagrange equations corresponding to the fidelity term of the energy. Since P is the projection operator, then theoretically

P' is the Heaviside function. However, for numerical purposes, we will use a smooth approximation to the Heaviside function

$$P'(z) = \begin{cases} 1 & \text{if } z \geq a \\ 0.5\big(1 + \frac{1}{a}z + \frac{1}{\pi}\sin\big(\frac{\pi}{a}z\big)\big) & \text{if } |z| < a \\ 0 & \text{if } z \leq -a, \end{cases} \tag{16}$$

where a is a small positive parameter.

More explicitly, for three-spatial dimensions $p = 0, ..., N_x$, $q = 0, ..., N_y$, $r = 0, ..., N_z$, iterations $m = 1, 2, ...$ and timestep $\triangle t$, the numerical algorithm is given by

$$\begin{aligned} &\frac{d_i^{m+1}(p,q,r) - d_i^m(p,q,r)}{\triangle t} = \\ &\quad \lambda P'(d_i^m(p,q,r))\left(\frac{(u_i^m(p,q,r))^2}{\sigma^2} - \frac{I_1(\frac{S_i(p,q,r)u_i^m(p,q,r)}{\sigma^2})}{I_0(\frac{S_i(p,q,r)u_i^m(p,q,r)}{\sigma^2})}\frac{S_i(p,q,r)u_i^m(p,q,r)}{\sigma^2}\right) \\ &\quad + c_i^m(p,q,r)(d_i^m(p+1,q,r) - d_i^m(p,q,r)) \\ &\quad\quad - c_i^m(p-1,q,r)(d_i^m(p,q,r) - d_i^m(p-1,q,r)) \\ &\quad + c_i^m(p,q,r)(d_i^m(p,q+1,r) - d_i^m(p,q,r)) \\ &\quad\quad - c_i^m(p,q-1,r)(d_i^m(p,q,r) - d_i^m(p,q-1,r)) \\ &\quad + c_i^m(p,q,r)(d_i^m(p,q,r+1) - d_i^m(p,q,r)) \\ &\quad\quad - c_i^m(p,q,r-1)(d_i^m(p,q,r) - d_i^m(p,q,r-1)), \end{aligned} \tag{17}$$

where

$$\begin{aligned} c_i^m(p,q,r) = &\Big(\sum_{j=1}^{n}(d_j^m(p+1,q,r) - d_j^m(p,q,r))^2 + (d_j^m(p,q+1,r) - d_j^m(p,q,r))^2 \\ &+ (d_j^m(p,q,r+1) - d_j^m(p,q,r))^2 + \epsilon^2\Big)^{-1/2}, \end{aligned} \tag{18}$$

and

$$u_i^m(p,q,r) = S_0(p,q,r)\exp\left(-P(d_i^m(p,q,r))\right). \tag{19}$$

For places where d_i is positive, then $P'(d_i) = 1$, and we get both the contribution of the fidelity and regularization terms in the evolution equations. However, if d_i is negative (violating the constraint that $d_i \geq 0$), then $P'(d_i) = 0$, and only the regularization term appears in the evolution equations. As mentioned, the regularization term provides smoothing; at voxels where d_i is negative, the value of d_i at these voxels will depend on the values of the surrounding d_i's in space. Considering a specific voxel where $d_i < 0$, its new value will be determined by a "weighted averaging" of the surrounding d_i's in a small neighborhood of this voxel. If the values of the surrounding d_i's are positive, then this "weighted averaging" will influence the negative d_i to increase. This is the motivation for the initial guess that we chose:

$$d_i(x) = \begin{cases} 0.005 & \text{if } S_i(x) > S_0(x) \\ -\log\big(\frac{S_i(x)}{S_0(x)}\big) & \text{if } S_i(x) \leq S_0(x). \end{cases} \tag{20}$$

The combination of a positive initial guess and the role of the projection operator P helps satisfy the constraint $d_i \geq 0$. Decreasing the constant 0.005 results in more violations of $d_i \geq 0$ and increasing it may lead to less satisfactory numerical results. As a note, using a sufficiently small timestep in the iterative method will guarantee $d_i \geq 0$ but will lead to a longer computation time.

We implement Neumann boundary conditions $\frac{\partial d_i}{\partial \eta} = 0$ where η is the normal to the boundary $\partial\Omega$. Iterations are stopped when the energy $F(d)$ decreases and stabilizes to an approximate steady state.

Furthermore, for efficiency, we use a cubic rational polynomial approximation to I_1/I_0 introduced in prior work [7]:

$$\frac{I_1(t)}{I_0(t)} \approx \frac{t^3 + 0.950037t^2 + 2.38944t}{t^3 + 1.48937t^2 + 2.57541t + 4.65314}. \tag{21}$$

In some cases, the standard deviation of the noise σ is unknown. Since σ appears in our denoising algorithm, we must find an estimation for this parameter. As in [19], we approximate σ by using the equation

$$E(X^2) = \nu^2 + 2\sigma^2 \tag{22}$$

for the second moment of a Rician distributed variable X. Here, $E(X^2)$ is the expected value of X^2 and ν is the true signal. For regions of a data set where we expect $\nu = 0$ (e.g. background), we can use

$$\sigma \approx \sqrt{E(X^2)/2} \tag{23}$$

where $E(X^2)$ is simply the average value of the square of the noisy signal over the region where the true signal is expected to be zero.

5. Numerical experiments and results. We perform numerical experiments on synthetic, phantom, and real data sets to demonstrate the validity of our proposed variational denoising model. With the first two experiments, we consider the root mean square error (RMSE) for vectorial data as well as a visualization of the data created by the first author of [9] as metrics of evaluation of results. The aims of these initial experiments were to understand how the different components of the model (e.g. projection operator ($P(d)$), initial condition, and tuning parameter λ) affect the denoised result. For the latter experiments, we compute orientation distribution functions (ODFs) and fractional anisotropy (FA) to gain a better idea of the applicability of our denoising model to real applications such as fiber tracking. In the next section, we give a brief summary of the metrics of evaluation that we use.

5.1. Metrics of evaluation. In this section, we briefly explain the metrics of evaluation that we use.

First, we use the root mean square error (RMSE) between two vectorial MR signals u_1 and u_2 given by

$$\mathrm{RMSE}(u_1, u_2) = \sqrt{\frac{1}{n \cdot N} \sum_{i=1}^{N} \sum_{j=1}^{n} (u_1(x_i, s_j) - u_2(x_i, s_j))^2}.$$

$x_i \in \Omega$, $i = 1, ..., N$ are the points of the domain and s_j, $j = 1, ..., n$ represent the diffusion directions.

For our initial visualizations of the vectorial MR signal, we follow [9]. For an example, see Fig. 4 for a visualization of exact two dimensional data provided by McGraw et al. [13]. Every third point in the spatial domain is plotted, and 81 vectors correspond to each point. Each vector points in the direction of diffusion and its length is determined by the magnitude of the MR signal when a diffusion gradient is applied in that direction. The tips of the vectors trace out the surface of a morphed sphere. With the special case where all 81 vectors have equal length, we have a sphere with radius equal to the length of the vectors.

We will also look at fractional anisotropy (FA) of fibers and orientation distribution functions (ODFs) and associated Jensen-Shannon divergence (JSD). Fractional anisotropy (FA) is a scalar between 0 and 1 that represents the amount of anisotropy of water diffusion at a particular voxel. FA value of 0 represents isotropic diffusion, meaning equal diffusion in all directions. FA values near 1 represent a more directed diffusion, so that diffusion occurs mostly in one direction only. FA is computed from the eigenvalues, λ_1, λ_2, and λ_3 corresponding to the eigenvectors of the diffusion tensor (the diffusion tensor data, a simpler representation of the HARDI data, when only three directions are used at each voxel to represent the water diffusion, thus forming a 3x3 "diffusion" tensor, is obtained from the HARDI data using the Diffusion Toolkit (http://trackvis.org/dtk/) [26])

$$\text{FA} = \sqrt{\frac{3}{2}}\sqrt{\frac{(\lambda_1-\hat{\lambda})^2+(\lambda_2-\hat{\lambda})^2+(\lambda_3-\hat{\lambda})^2}{\lambda_1^2+\lambda_2^2+\lambda_3^2}}, \tag{24}$$

where $\hat{\lambda} = (\lambda_1+\lambda_2+\lambda_3)/3$.

We compute the diffusion orientation distribution function (ODF), which is a probability density function measuring the distribution of water diffusion in different directions on the sphere; for an example, see Fig. 12. ODF's are calculated using the tensor distribution function algorithm in [12]. We will also use the Jensen-Shannon divergence (JSD) to measure the difference between the noisy/denoised ODF and the ODF of the ground truth. JSD between two ODFs P and Q are given by

$$JSD(P,Q) = \frac{1}{2}D_{KL}(P,M) + \frac{1}{2}D_{KL}(Q,M), \tag{25}$$

where $M = (P+Q)/2$ and $D_{KL}(\cdot,\cdot)$ denotes the Kullback-Leibler information distance between two probability distributions [32]. For probability distributions P_1 and P_2 defined on the sphere S^2, D_{KL} is given by

$$D_{KL}(P_1,P_2) = \int_{S^2} \log\Big(\frac{P_1(\theta,\phi)}{P_2(\theta,\phi)}\Big)P_1(\theta,\phi)dS. \tag{26}$$

5.2. Numerical results. For our first data set, we consider two dimensional (2D) synthetic data provided by McGraw et al. [13]. It is 16×16 in spatial dimension with 81 diffusion directions (see Fig. 4). We add Rician noise with standard deviation $\sigma = 18$ to create the noisy data, denoted S. Since S_0 is not available for this synthetic data, S_0 is taken to be a constant 255, and initial guess is taken to be $d_0 = -\log(S/S_0)$. Recall that the denoised MR signal is given by $u = S_0 \exp(-P(d))$. The visualization of our results is given in Fig. 4 and the corresponding error histograms are given in Fig. 5. This visualization shows every third point and all diffusion directions. To understand this visualization better, consider one of the 36 points

shown in Fig. 4. At this point, there are 81 corresponding vectors. Each of these vectors points in a diffusion direction, and its magnitude corresponds to the magnitude of the MR signal associated with this diffusion direction. The tips of the vectors will trace out the surface of a morphed sphere, which is seen in Fig. 4. For the special case where the MR signal is equal in all diffusion directions, the visualization will be a perfect sphere. The RMSE of the noisy data is 17.8147 and is reduced to 5.4968 after denoising with our model. In Table 1, we give a comparison of our results with the results from similar denoising experiments found in [13, 9]; our model gives a slight improvement in terms of RMSE.

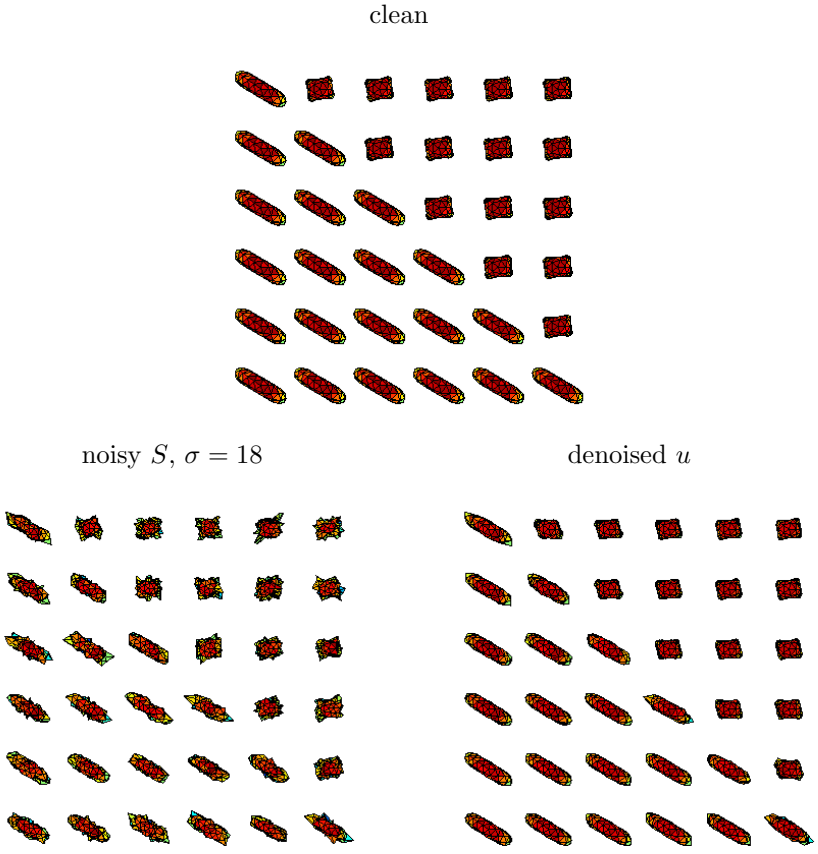

FIG. 4. *Top: exact 2D synthetic data. Left: noisy data, corrupted by Rician noise* $\sigma = 18$. *Right: restored data. Denoised data satisfies* $d \geq 0$.

Second, we consider real HARDI data of the brain. A healthy subject was scanned on a 4 Tesla Bruker Medspec MRI scanner with an optimized diffusion imaging sequence. DWI parameters were: echo and repetition time, TE/TR 92.3/8250 ms, 55×2 mm contiguous slices, field of view: FOV = 23 cm. 41 images were collected: 11 baseline (S_0) images with no diffusion sensitization (i.e., T_2- weighted images) and 30 diffusion-weighted images (b-value: 1159 s/mm^2) with gradient directions evenly distributed on the hemisphere. The reconstruction matrix was 128×128, yielding a 1.8×1.8 mm^2 in-plane resolution. The clean data was created by fitting actual HARDI data to a 6th order spherical harmonic expansion, and the resulting data set

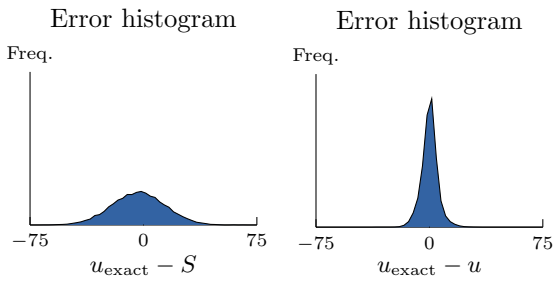

FIG. 5. *Left: error histogram of noisy data, RMSE of S: 17.8147. Right: error histogram of denoised data, RMSE of u: 5.4968.*

TABLE 1
Comparison of RMSEdenoised/RMSEnoisy ratios for different methods.

	RMSEdenoised/RMSEnoisy
TV+L^1 fidelity denoising in u [9]	5.8991/17.7079 = 0.3331
TV+L^1 fidelity denoising in d [9]	7.6081/17.7079 = 0.4296
FEM in [13]	11.9964/17.7079 = 0.6675
FEM + TV in [13]	7.6367/17.7079 = 0.4313
our model	5.4968/17.8147 = **0.3086**

is considered to be the "ground truth" even though this is not truly correct since it has artifacts and negative d values. The noisy data was generated by adding Rician noise with $\sigma = 15$ to this "ground truth". We take a subset (size $95 \times 128 \times 55$) of the 3D volume, and use all 30 uniformly distributed diffusion directions. Fig. 6 gives a visualization of the 29th slice of the clean data, and Fig. 7 gives S_0, clean MR signal, and clean diffusion signal d corresponding to the 10th diffusion-sensitized direction. In Fig. 8 and Fig. 9, we give visualizations of the clean, noisy, and denoised images corresponding to 2 of the 30 diffusion-sensitized directions. For this particular experiment, we used $\lambda = 0.05$. The RMSE of the noisy data is 10.5454 and the RMSE of the denoised data is 8.0845, showing some improvement after denoising. In addition, the number of points where the constraint $d \geq 0$ is not satisfied is reduced from 815 points in the noisy data to 52 points in the denoised data. When we reduce the tuning parameter to $\lambda = 0.001$ to introduce more smoothing, we see that the denoised solution is oversmoothed (Fig. 10 and Fig. 11), but the constraint $d \geq 0$ is satisfied everywhere.

For the last three numerical experiments, we look at fractional anisotropy (FA), orientation distribution functions (ODFs), and Jensen-Shannon divergence (JSD). In the last experiment, we will revisit the real HARDI data set and evaluate the performance of our denoising model with ODFs and JSD values.

We consider again 2D synthetic data. The size of the data set is 8×8 with diffusion measured in 94 uniformly distributed directions at each spatial point. S_0 in this case is assumed to be a constant equal to 1. We create noisy data by adding Rician noise with standard deviation $\sigma = 0.01, 0.02, 0.04, 0.1$ and 0.2. The same σ used to create the noisy data is input as a known parameter in our denoising model. Fig. 12 gives the ODFs of the clean, noisy, and denoised data corresponding to the case $\sigma = 0.2$. In all cases, the denoised data gives lower JSD values than those of the corresponding noisy data sets, indicating an increase in similarity to the ODF of the

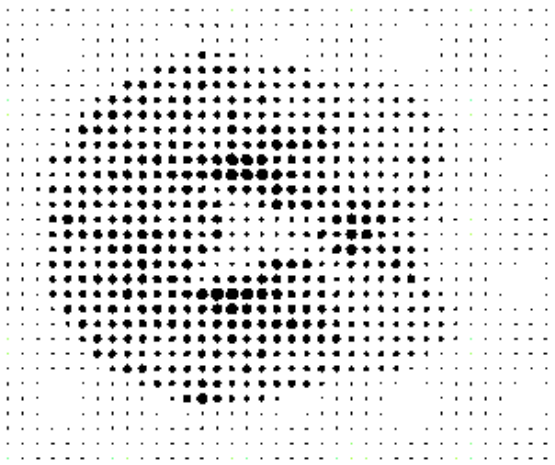

FIG. 6. *Visualization of 29th slice of clean data. Complete volume size: 95x128x55, 30 diffusion directions.*

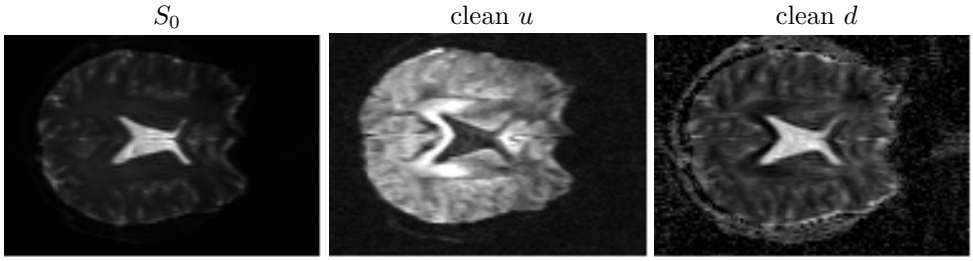

FIG. 7. *Left to right:* S_0*, clean* u*, and corresponding clean* d *of* 29*th slice,* 10*th diffusion direction.*

ground truth after denoising (Table 2). Also, RMSE values between the denoised and ground truth data are lower than the RMSE values between noisy and ground truth data (Table 2). This suggests that we have a more accurate signal after denoising.

TABLE 2

Jensen-Shannon divergence (JSD) values between ODF of noisy/denoised data and ODF of ground truth data, and RMSE values between noisy/denoised data and ground truth data.

	JSD		RMSE	
σ	noisy	denoised	noisy	denoised
0.01	0.0044	0.0004	0.0100	0.0065
0.02	0.0102	0.0008	0.0201	0.0130
0.04	0.0441	0.0133	0.0396	0.0231
0.1	0.1578	0.0431	0.0986	0.0545
0.2	0.5423	0.1736	0.1950	0.0963

Next, we consider DW-MRI data of a hardware phantom containing synthetic fibers created by Pullens et al. [17] (see Fig. 13). The fibers ($\approx$ 10 μm circular diameter) consist of polyester yarns wound into bundles, which are then interdigited on top of each other and secured with heat shrink tubes. We have two scans of this phantom using a 7T and 3T scanner. For both these scans, the standard deviation of

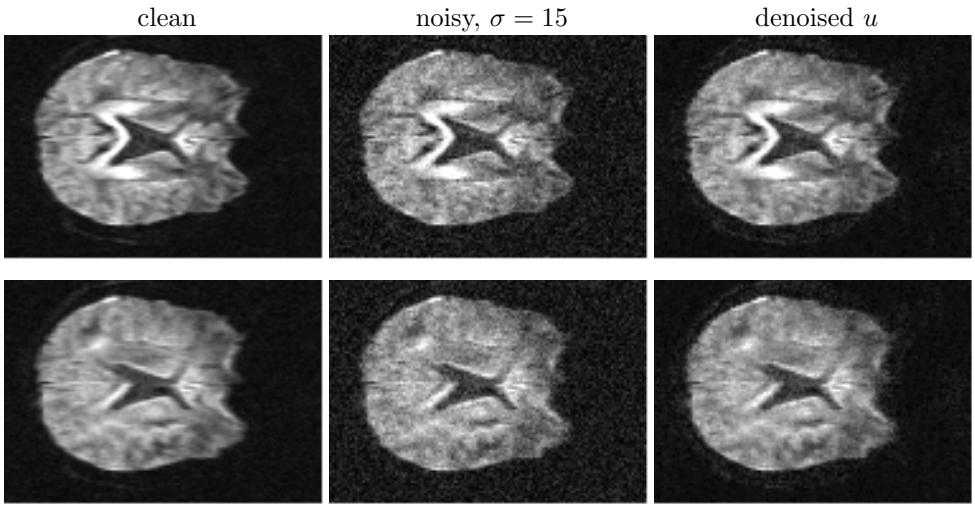

FIG. 8. *Visualization of 2 of 30 diffusion-sensitized MR signals of the real HARDI brain data. Left to right: clean, noisy, denoised data. RMSE of denoised data: 8.0845. RMSE of noisy data: 10.5454.* $\lambda = 0.05$*, 35 iterations.* $d_{noisy} < 0$ *at* 815 *points,* $d_{denoised} < 0$ *at* 52 *points.*

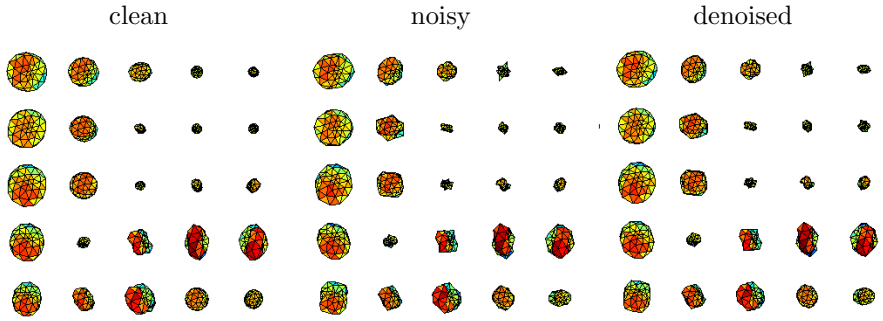

FIG. 9. *Visualization of small portion of clean, noisy, denoised data. RMSE of denoised data: 8.0845. RMSE of noisy data: 10.5454.* $\lambda = 0.05$*, 35 iterations.* $d_{noisy} < 0$ *at* 815 *points,* $d_{denoised} < 0$ *at* 52 *points.*

the noise σ is approximated using the method described in the end of Section 4.

The DW-MRI data represented in Fig. 14 was collected on the phantom using a Siemens MAGNETOM 7T scanner located at the CMRR at the University of Minnesota. Parameter settings were: 64 slices with FOV $= 256 \times 256$ mm, $2 \times 2 \times 2$ mm^3 voxels, TR/TE $= 5700/55$ ms, 128 DWI at $b = 1000$ s/mm^2, and 15 S_0 images. Furthermore, the data was distortion corrected using a field map. For computational purposes, we consider the 37th to 41st z-slices only and present the T2 (S_0) and FA images for the 38th z-slice. Given that the fibers are constructed with synthetic materials, the FA should be relatively constant along the fibers. A goal of denoising is to decrease the standard deviation of the FA along a fiber, indicating that the FA along a fiber lies within a smaller range of values and hence is more constant. After denoising using our proposed model, the standard deviation of FA along a fiber drops from 0.1503 to 0.1292. However, the mean FA drops slightly, which may be a result of oversmoothing (see Table 3). As a result, the denoised signal implies a less directed

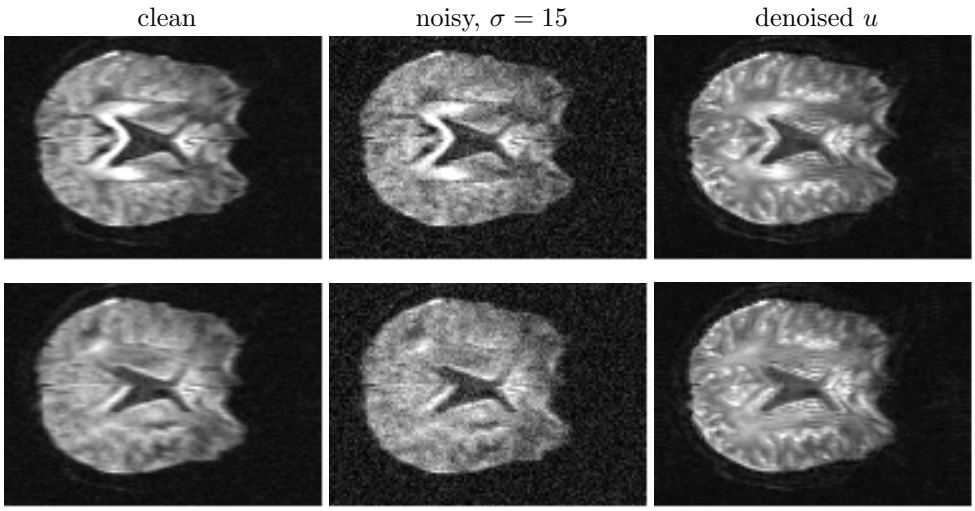

FIG. 10. *Visualization of 2 of 30 diffusion-sensitized MR signals of the real HARDI brain data. Left to right: clean, noisy, denoised data. With* $\lambda = 0.001$ *and at* 100 *iterations, the solution is oversmoothed. However,* $d_{denoised} \geq 0$ *everywhere.*

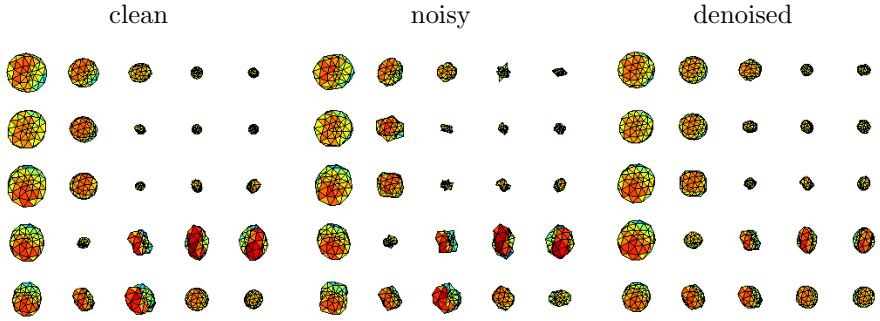

FIG. 11. *Visualization of small portion of clean, noisy, denoised data. With* $\lambda = 0.001$ *and at* 100 *iterations, the solution is oversmoothed. However,* $d_{denoised} \geq 0$ *everywhere.*

diffusion along the fiber.

TABLE 3
Mean and standard deviation values of FA along a fiber for noisy and denoised phantom data.

	mean	standard deviation
noisy	0.2392	0.1503
denoised	0.2140	0.1292

The DW-MRI data represented in Fig. 15 was also collected on the same phantom using a Siemens TIM Trio 3T scanner also located at the CMRR at the University of Minnesota. Parameter settings were: 64 slices with FOV = 256×256 mm, 2×2×2 mm^3 voxels, TR/TE = 9000/94 ms, 128 DWI at $b = 1500$ s/mm^2, and 17 S_0 images. For this particular dataset, all raw DWI's were corrected for eddy current distortion using the *eddy_correct* function from the FSL toolbox (http://fsl.fmrib.ox.ac.uk/fsl) [21, 27].

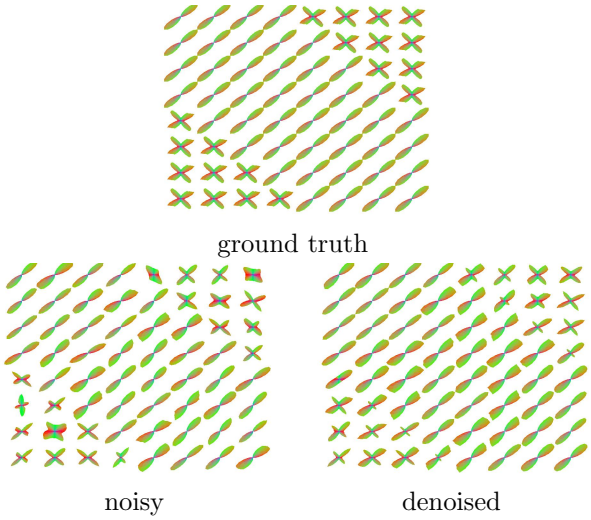

FIG. 12. *Top: ODF of ground truth data. Bottom left: ODF of noisy data generated with Rician noise with* $\sigma = 0.2$. *Bottom right: ODF of denoised data.*

FIG. 13. *Hardware phantom containing synthetic fibers created by Pullens et al. [17].*

Geometric distortions due to magnetic susceptibility were then corrected using a field map collected just prior to the DWI's, using the FSL *prelude* and *fugue* functions. Diffusion Tensor Imaging (DTI) tractography was conducted via the Diffusion Toolkit (http://trackvis.org/dtk/) [26] using the critical angle threshold = 20° and the 2$^{\text{nd}}$ order Runge-Kutta method [2]. The noisy and denoised fiber and fractional anisotropy visualizations of the phantom data are displayed in Fig. 15. After denoising, the bottom right of the fiber is visible.

Lastly, we denoise the same real HARDI brain data that we have described earlier (Figures 6 and 7), but in this case, we use ODFs and corresponding JSDs as metrics of evaluation. We present in Fig. 16 the ODFs of the "ground truth", noisy, and denoised 30th z-slice of the real HARDI brain data. Fig. 17 shows the JSD between the ODF of the noisy data and ODF of the "ground truth" data, and we see that the ODFs of the denoised data and "ground truth" data are more similar than the ODFs of the noisy data and "ground truth" data. This is reflected also in the drop in the

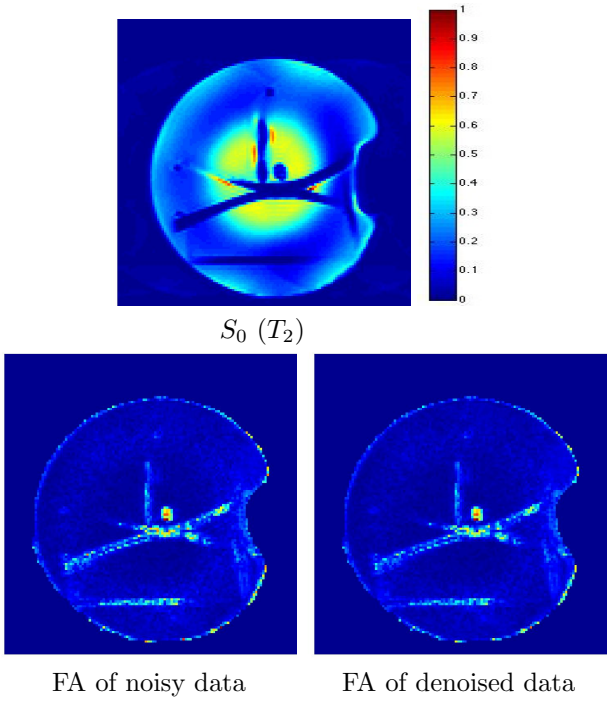

FIG. 14. *Visualizations of 38th slice of phantom data.*

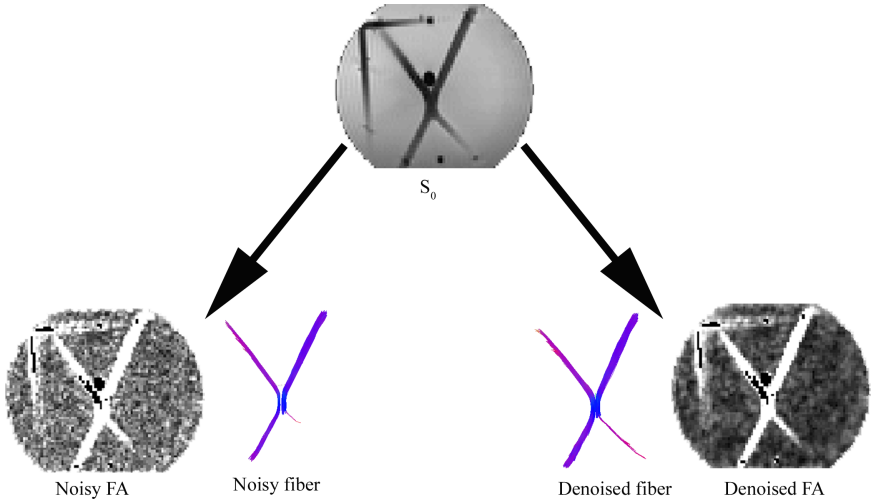

FIG. 15. *Left: noisy fiber and fractional anisotropy visualization. Right: denoised fiber and fractional anisotropy visualization. Note that the bottom right of the fiber is visible in the denoised data but not in the noisy data.*

mean JSD values from 0.5787 to 0.3297 after denoising.

6. Conclusion. We proposed and analyzed a variational denoising model of HARDI data corrupted by Rician noise. We have performed numerical experiments on multiple types of data and used a variety of metrics to evaluate the performance

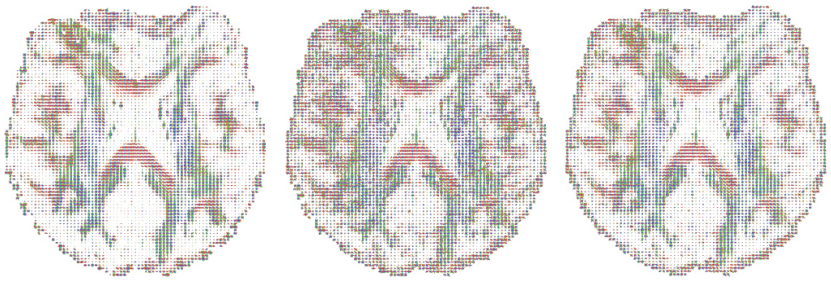

FIG. 16. *ODF visualizations of the 30th z-slice of the real HARDI brain data. Top: ODF of "ground truth" data. Left: ODF of noisy data. Right: ODF of denoised data. The color in this figure indicates the fiber direction: red for left-right, blue for superior-inferior, and green for anterior-posterior.*

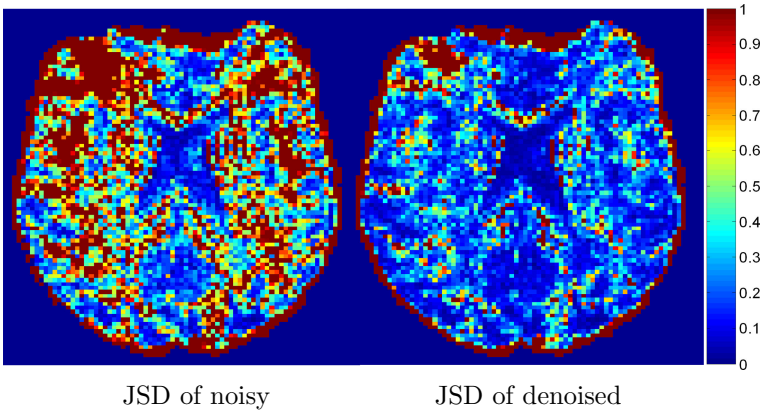

FIG. 17. *JSD between ODFs of noisy and "ground truth" data (mean = 0.5787, std. dev. = 0.3445) and between ODFs of denoised and "ground truth" data (mean = 0.3297, std. dev. = 0.3037). In the colorbar, we have values 0 (blue) to 1 (red).*

of our proposed denoising model. We see that there are improvements with both visualizations as well as RMSE and JSD values. The result in Fig. 15 is promising for the applicability of the model for fiber tracking. In addition, the use of the projection operator seems helps to enforce the constraint $d \geq 0$. For future work, we may test different regularization terms such as nonlocal regularization or learned priors.

Acknowledgments. This work was funded by the National Institutes of Health (grants NIH P41 RR008079, P41 EB015894, S10 RR026783, R01 EB008432), the W.M. Keck Foundation, the National Science Foundation (grants NSF DMS 1217239 and CCF-0926127), and the UC Lab Fees Research grant 12-LR-236660.

REFERENCES

[1] A. W. ANDERSON, *Theoretical Analysis of the Effects of Noise on Diffusion Tensor Imaging*, MRM, 46 (2001), pp. 1174–1188.

[2] P. J. BASSER, S. PAJEVIC, C. PIERPAOLI, J. DUDA, AND A. ALDROUBI, *In Vivo Fiber Tractography Using DT-MRI Data*, Magnetic Resonance in Medicine , 44 (2000), pp. 625–632.

[3] S. Basu, T. Fletcher, and R. Whitaker, *Rician Noise Removal in Diffusion Tensor MRI*, LNCS 4190, MICCAI, (2006), pp. 117–125.
[4] M. Cihangiroglu, A. Mufit Ulug, Z. Firat, A. Bayram, A. Kovanlikaya, and I. Kovanlikaya, *High b-value Diffusion-Weighted MR Imaging of Normal Brain at 3T*, European Journal of Radiology (2009).
[5] M. Descoteaux, E. Angelino, S. Fitzgibbons, R. Deriche, *Regularized, Fast and Robust Analytical Q-Ball Imaging*, Magnetic Resonance in Medicine, 58 (2007), pp. 497–510.
[6] L. R. Frank, *Characterization of Anisotropy in High Angular Resolution Diffusion-Weighted MRI*, MRM, 47 (2002), pp. 1083–1099.
[7] P. Getreuer, M. Tong, and L. Vese, *A Variational Model for the Restoration of MR Images Corrupted by Blur and Rician Noise*, Proceedings of ISVC, Part I, LNCS 6938, (2011).
[8] P. Getreuer, M. Tong, and L. Vese, *Total Variation Based Rician Denoising and Deblurring Model*, 2012 (submitted).
[9] Y. Kim, P. M. Thompson, and L. Vese, *HARDI Data Denoising Using Vectorial Total Variation and Logarithmic Barrier*, IPI, Special Issue in Medical Image Analysis, 4 (2010), pp. 273–310.
[10] Y. Kim, P. M. Thompson, A. W. Toga, L. Vese, and L. Zhan, *HARDI Denoising : Variational Regularization of Spherical Apparent Diffusion Coefficients, SADC*, IPMI 2009, LNCS 5636, (2009) pp. 515–527.
[11] C. Lenglet, J. S. W. Campbell, M. Descoteaux, G. Haro, P. Savadjiev, D. Wassermann, A. Anwander, R. Deriche, G. B. Pike, G. Sapiro, K. Siddiqi, and P. M. Thompson, *Mathematical Methods for Diffusion MRI Processing*, NeuroImage, Special Issue on Mathematics in Brain Imaging (2008).
[12] A. D. Leow, S. Zhu, L. Zhan, K. McMahon, G. I. de Zubicaray, M. Meredith, M. J. Wright, A. W. Toga, and P. M. Thompson, *The Tensor Distribution Function*, MRM, 61 (2009), pp. 205–214.
[13] T. McGraw, B. Vemuri, E. Ozarslan, Y. Chen, and T. Mareci, *Variational Denoising of Diffusion Weighted MRI*, IPI, 3 (2009), pp. 625–648.
[14] O. Michailovich, Y. Rathi, and S. Dolui, *Spatially Regularized Compressed Sensing for High Angular Resolution Diffusion Imaging*, IEEE Transactions onf Medical Imaging, 30 (2011), pp. 1100–1115.
[15] V. Patel, Y. Shi, P. M. Thompson, and A. W. Toga, *K-SVD for HARDI Denoising*, Proceedings of ISBI, (2011) pp. 1805–1808.
[16] C. Pierpaoli and P. Basser, *Toward a Quantitative Assessment of Diffusion Anisotropy*, MRM, 36 (1996), pp. 893–906.
[17] P. Pullens, A. Roebroeck, and R. Goebel, *Ground Truth Hardware Phantoms for Validation of Diffusion-Weighted MRI Applications*, JMRI, 32 (2010), pp. 482–488.
[18] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.
[19] J. Sijbers, A. J. Den Dekker, D. Van Dyck, and E. Raman, *Estimation of Signal and Noise from Rician Distributed Data*, Proceedings of the International Conference on Signal Processing and Communications, (1998) pp. 140–142.
[20] S. Skare, T. Li, B. Nordell, and M. Ingvar, *Noise Considerations in the Determination of Diffusion Tensor Anisotropy,* MRM, 18 (2000), pp. 659–669.
[21] S. M. Smith, M. W. Jenkinson, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, R. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. M. Brady, and P. M. Matthews, *Advances in Functional and Structural MR Image Analysis and Implementation as FSL*, NeuroImage, 23 (2004), pp. 208–219.
[22] E. O. Stejskal and J. E. Tanner, *Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient*, JCP, 42 (1965), pp. 288–292.
[23] M. Tong, Y. Kim, L. Zhan, G. Sapiro, C. Lenglet, B. A. Mueller, P. M. Thompson, and L. A. Vese, *A variational model for denoising high angular resolution diffusion imaging*, Proceedings 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), (2012) pp. 530–533.
[24] D. S. Tuch, R. M. Weisskoff, J. W. Belliveau, and V. J. Wedeen, *High Angular Resolution Diffusion Imaging of Human Brain*, Proceedings of 7th Annual Meeting of ISMRM, (1999) 321.
[25] D. S. Tuch, T. G. Reese, M. R. Wiegell, N. Makris, J. W. Belliveau, and V. J. Wedeen, *High Angular Resolution Diffusion Imaging Reveals Intravoxel White Matter fiber Heterogencity*, MRM, (2002) pp. 577–582.
[26] R. Wang, T. Benner, A. G. Sorensen, V. J. Wedeen, *Diffusion Toolkit: A Software Package*

for Diffusion Imaging Data Processing and Tractography, Proceedings of the International Society for Magnetic Resonance in Medicine, (2007), 15.

[27] M. W. Woolrich, S. Jbabdi, B. Patenaude, M. Chappell, S. Makni, T. Behrens, C. Beckmann, M. Jenkinson, M., and S. M. Smith, *Bayesian Analysis of Neuroimaging data in FSL*, NeuroImage, 45 (2009), pp. S173–186.

[28] H. Gudbjartsson and S. Patz, *The Rician distribution of noisy MRI data*, MRM, 34 (1995), pp. 910–914.

[29] R. M. Henkelman, *Measurement of signal intensities in the presence of noise in MR images*, Med. Phys., 12 (1985), pp. 232–233.

[30] S. O. Rice, *Mathematical analysis of random noise*, Bell Syst. Tech. J., 23 (1944), pp. 282–444.

[31] B. P. Lathi, *Modern digital and analog communication systems*, Holt-Saunders International Edition (1983).

[32] S. Kullback and R. A. Leibler, *On information and sufficiency*, Annual Math. Stat., 22 (1951), pp. 7-9-86.

[33] 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*, Med. Phys., 15 (1989), pp. 813–817.