HARDI data denoising using vectorial total variation and logarithmic barrier

Yunho Kim,* Paul M. Thompson,† and Luminita A. Vese‡

Abstract

In this work, we wish to denoise HARDI (High Angular Resolution Diffusion Imaging) data arising in medical brain imaging. Diffusion imaging is a relatively new and powerful method to measure the three-dimensional profile of water diffusion at each point in the brain. These images can be used to reconstruct fiber directions and pathways in the living brain, providing detailed maps of fiber integrity and connectivity. HARDI data is a powerful new extension of diffusion imaging, which goes beyond the diffusion tensor imaging (DTI) model: mathematically, intensity data is given at every voxel and at any direction on the sphere. Unfortunately, HARDI data is usually highly contaminated with noise, depending on the *b-value* which is a tuning parameter pre-selected to collect the data. Larger *b-values* help to collect more accurate information in terms of measuring diffusivity, but more noise is generated by many factors as well. So large *b-values* are preferred, if we can satisfactorily reduce the noise without losing the data structure. Here we propose two variational methods to denoise HARDI data. The first one directly denoises the collected data S, while the second one denoises the so-called sADC (spherical Apparent Diffusion Coefficient), a field of radial functions derived from the data. These two quantities are related by an equation of the form $S = S_0 \exp\left(-b \cdot sADC\right)$ (in the noise-free case). By applying these two different models, we will be able to determine which quantity will most accurately preserve data structure after denoising. The theoretical analysis of the proposed models is presented, together with experimental results and comparisons for denoising synthetic and real HARDI data.

1 Introduction to the HARDI data

Currently, HARDI data is used to map cerebral connectivity through fiber tractography in the brain. HARDI is a type of diffusion MRI, which was introduced in the mid-1980s by Le Bihan et al. [27, 28, 29] and Merboldt et al. [35]. It is based on the idea that the magnetic resonance signal, which forms the basis of MRI, is attenuated when water diffuses out of a voxel, and the degree of attenuation can be used to measure the rate of water diffusion in any arbitrary three-dimensional direction, via the Stejskal-Tanner equation [42]. Water diffusion occurs preferentially in directions that are aligned with axonal fiber pathways, and is hindered in orthogonal directions by the myelin sheaths that coat the axons. Because of this diffusion anisotropy, initial approaches to assess fiber directions modeled the three-dimensional diffusion profile at each point as a single tensor (Beaulieu

---

*Department of Mathematics, UCI. E-mail: yuno1123@math.uci.edu (corresponding author).

†Laboratory of Neuroimaging and Department of Neurology, UCLA School of Medicine. E-mail: thompson@loni.ucla.edu.

‡Department of Mathematics, UCLA. E-mail: lvese@math.ucla.edu.

et al. [6]), in which the principal eigenvector of the diffusion tensor can be used to recover the dominant fiber pathway at that voxel. The diffusion tensor model (Basser et al. [5]) describes the anisotropic nature of water diffusion in tissues (inside a typical 1-3mm sized voxel) by estimating, from a set of K diffusion-sensitized images, the 3x3 covariance matrix of a Gaussian distribution (Beaulieu et al. [6]). Each voxel's signal intensity in the k-th image is decreased, by water diffusion, according to the Stejskal-Tanner equation [42]:

$$S_k = S_0 \exp\left[-b g_k^T D g_k\right],$$

where S_0 is the non-diffusion weighted signal intensity, D is the 3x3 diffusion tensor, g_k is the direction of the diffusion gradient and b is Le Bihan's factor with information on the pulse sequence, gradient strength, and physical constants.

Unfortunately, although it is widely used, the diffusion tensor model breaks down for voxels in which fiber pathways cross or mix together, and these are ubiquitous in the brain which is highly interconnected. More advanced image acquisition techniques, such as HARDI (Tuch et al. [47, 48]) or diffusion spectrum imaging (also called q-ball imaging; Tuch et al. [49]), have been introduced in the past 5 years - this type of data recovers the local microstructure of water diffusion more accurately than standard DTI data (diffusion tensor imaging). HARDI, DTI and other similar modalities permit non-invasive quantification of the water diffusion in living tissues. The tissue structure will affect the Brownian motion of the water molecules which will lead to an anisotropic diffusion. By imaging diffusion in an arbitrary number of directions (often 100 or more), HARDI overcomes the limited accuracy of the tensor model in resolving the highly complex fiber structure of the brain, particularly in regions with fiber crossings.

HARDI data makes it possible to compute the orientation diffusion function over a sphere of possible directions. Tuch [46, 48] developed the first HARDI acquisition and processing methods, and later Frank [19] used spherical harmonic expansions for processing HARDI data sets. A very active area of research has grown up in processing the HARDI signals, leading to methods for HARDI denoising, segmentation, and registration using metrics on spherical functions (Lenglet et al. [30]). Most of these signal processing methods still model the diffusion signal as a tensor, rather than exploiting the full information in the spherical harmonic expansion. For example, Khurd et al. [25] used isometric mapping and manifold learning (eigendecomposition of the distance matrix) to directly fit a manifold to the tensors, compute its dimensionality, and distinguish groups using Hotelling's T^2 statistics. Initial image processing on the full HARDI signal has focused on fitting a discrete mixture of k distinct tensors to the signal, and later on fitting a continuous mixture model for modeling the MR signal decay and multi-fiber reconstruction (Jian et al. [21], [22]), or fitting a continuous mixture of tensors using a unit-mass distribution on the symmetric positive definite tensor manifold (Leow et al. [31]).

Initial work on the nonlinear (fluid) matching of HARDI images has taken a more non-parametric approach, and has used the Kullback-Leibler divergence to measure the discrepancy between ODF (Orientation Distribution Function) fields (Chiang et al. [11, 12]), using a 3D fluid transform to minimize the discrepancy between two fields of ODFs. As information theory can be used to measure the overlap between diffusion probability density functions, there is much promising work using metrics derived from information theory (e.g., the Fisher-Rao metric, von Mises-Fisher distribution, etc.; McGraw et al. [34]; Srivastava et al. [41]; Chiang et al. [12]). Other work has modeled the HARDI signal as high-order tensors (Barmpoutis et al. [4]) or as a stratification (a mixture of manifolds with different dimensions; Haro et al. [20]).

2

The HARDI data is the MRI signal attenuation information after time $t > 0$ modeled by

$$S_t(x, \theta, \phi) = S_0(x) \exp(-b \cdot d_t(x, \theta, \phi)),$$

where Ω is a bounded open subset of $\mathbb{R}^3$, $x \in \Omega$ and $\theta \in [0, 2\pi), \phi \in [0, \pi)$. S_0 is the MRI signal that is obtained when no diffusion gradient vector is applied and this is considered to be a reference image, relative to which the diffusion-attenuated signal is measured. The function $d_t(x, \theta, \phi)$ is called the spherical Apparent Diffusion Coefficient (sADC), which measures how much the water molecules diffuse in the given direction

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

and b is a parameter pre-selected to collect the data.

We note that, in our work, the sADC is a function that takes values on the unit sphere, whereas in diffusion imaging ADC is sometimes used to refer to a scalar field derived from the diffusion images.

In reality, in experimental data, a higher b-value (e.g., 3000 s/mm^2) tends to lead to a smaller SNR (Signal-to-Noise-Ratio) in the obtained images, thus more noise, [13]. Hence, we are led to consider the following simplified degradation model

$$S_t(x, \theta, \phi) = S_0(x) \exp(-b \cdot d_t(x, \theta, \phi)) + noise(x, \theta, \phi). \tag{1}$$

We may say that the baseline signal collected without any diffusion gradient applied, which is S_0, may also be contaminated by noise, but here we assume that this can be neglected, or we just consider the last noise term in (1) to encompass all types of noise, which we do not know exactly. This is a reasonable approximation, because in practice, it is common to collect several non-diffusion weighted images (several S_0 images) whose average may be used as a reference signal S_0 (see, e.g. Zhan et al. [51]). If we let $\tilde{S}_t(x, \theta, \phi)$ be a denoised dataset, then we expect that for all x, ϕ, θ,

$$0 \leq \tilde{S}_t(x, \theta, \phi) \leq S_0(x). \tag{2}$$

As already mentioned, the data has to be first denoised before extracting the fibers, or before registration. Although HARDI is a relatively recent type of data acquisition, several HARDI processing methods have already been processed: we mention just a few. In McGraw et al. [33] and Jonasson et al. [23], curve evolution techniques are applied for the segmentation of HARDI data. Descoteaux, Deriche and collaborators, among others, have also proposed methods to segment HARDI data [15], for estimation of the sADC [16], and for mapping of neuronal fiber crossings [17]. The work Delputte et al. [14] deals with denoising and regularization of fields of ODFs.

The prior work most relevant to ours is by McGraw et al. [32]: the noisy data $S = S_t(x, \theta, \phi)$ is regularized to remove noise in a functional minimization approach; a standard L^2 data fidelity term is used, combined with a weighted version of vectorial total variation regularization in space, and H^1 regularization of data at every voxel with respect to direction. At every voxel, the data is mapped into 2-dimensional space plane using spherical coordinates and then discretized using finite elements. Denoising results for synthetic and real HARDI data are presented in [32]. In our proposed work, we also use vectorial total variation for the regularization. However, in our first model we also impose the constraint (2). Our second proposed model differs even more from the one in [32], since we faithfully follow the signal degradation model (1) and we denoise d instead of S.

3

The outline of the paper is as follows. In Section 2 we recall the logarithmic barrier function for general optimization problems with inequality constraints. In Section 3 we introduce two variational models for HARDI denoising, together with theoretical results, the associated Euler-Lagrange equations, and the numerical approximations. Finally, in Section 4 we present several numerical results of denoising synthetic and real HARDI data obtained using the proposed models, error assessment and comparisons.

We would like to mention that a preliminary version of this work has been presented and published in IPMI 2009 [26].

2 Logarithmic barrier functions

In this section, we will briefly recall the logarithmic barrier method for optimization problems with inequality constraints, one of the barrier function methods used to force the unknown variables of interest to remain feasible. Suppose that we have the following optimization problem: let $\varphi : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ be smooth functions with $c = (c_1, \ldots, c_m)$,

$$\min_x \varphi(x), \text{ subject to } c_i(x) \geq 0 \text{ for } i = 1, \ldots, m. \tag{3}$$

Let O be the feasible domain, i.e.,

$$O = \{x \in \mathbb{R}^n \mid c_i(x) \geq 0 \text{ for } i = 1, \ldots, m\}.$$

We also assume that the first and second-order sufficient conditions hold at a point x^* so that x^* is a strict local solution of (3). Then the logarithmic barrier function for (3) is the unconstrained optimization problem [37]

$$\min_x P(x; \mu) = \varphi(x) - \mu \sum_{i=1}^{m} \log(c_i(x)). \tag{4}$$

This method appears in many works on optimization and applications. An initial guess x_0 in the interior of the feasible set, i.e. $c_i(x_0) > 0$, $i = 1, ..., m$, is chosen. Then an iterative method is applied (such as gradient descent) for fixed, small $\mu_0 > 0$, to obtain x_1 in the interior of the feasible set, that minimizes $P(x; \mu_0)$. Then μ_1 is selected, such that $0 < \mu_1 < \mu_0$, and x_2 is the minimizer of $P(x; \mu_1)$ in the interior of the feasible set, etc. Thus the basic idea of the method is the computation of the minimizers x_k of $P(\cdot; \mu_k)$ for a sequence of values $\mu_k \downarrow 0$, with the goal that x_k converges to x^*.

The authors of [18] and [37] present general results about the existence of minimizers of the logarithmic barrier functions in the vicinity of x^* and the convergence of the sequence of minimizers to x^* as $\mu_k \downarrow 0$.

3 Variational denoising models

We propose here two variational denoising methods: one recovers a denoised S, while the other method recovers a denoised d. Again, we let Ω be a bounded open subset of $\mathbb{R}^3$. The HARDI data is a collection of intensity values at uniformly pre-selected directions on the sphere, to which

the electromagnetic field is applied, that is, at each position $x \in \Omega \subset \mathbb{R}^3$, we measure values at different directions. We note briefly that the actual set of directions is typically computed using an electrostatic repulsion PDE to optimize the sampling of a spherical signal using a finite set of observations (see Tuch et al. [49] for a discussion of spherical sampling schemes).

In the continuous setting, we obtain a function defined on a manifold $\Omega \times S^2$, where Ω is the spatial domain and the sphere S^2 is the space of gradient directions. However, it is not easy to work with the entire domain $\Omega \times S^2$ for computational purposes. Instead, we will use a discretized version S_d^2 of S^2, where

$$S_d^2 = \{s_i \in S^2 \mid i = 1, \ldots, n\}$$

and the directions $s_i \in S^2$ are uniformly chosen. Then the function that we obtain has the form

$$f = (f_1, \ldots, f_n)$$

where each $f_i : \Omega \to \mathbb{R}$ corresponds to the direction s_i. We will also use a function representation $f_0 : \Omega \to \mathbb{R}$ for the data S_0 which has only spatial information (assumed to be given also). Note that what is important in this model is to measure the correct diffusivity along each direction at each position which is based on the difference between the noisy data S and the reference data S_0, thus our main concern is to deal with nonnegative differences. To impose the right amount of smoothness and discontinuity on the denoised data, we will use the space $BV(\Omega; \mathbb{R}^n)$, which is the n-dimensional version of $BV(\Omega)$, defined as follows.

Definition 1. Let $u = (u_1, \ldots, u_n) \in L^1(\Omega; \mathbb{R}^n)$. We define

$$|Du|(\Omega) = \sup \left\{ \sum_{i=1}^n \int_\Omega u_i \mathrm{div}(\psi_i) \ \Big| \ \psi = (\psi_1, \ldots, \psi_n) \text{ and } \psi_i \in C_c^1(\Omega; \mathbb{R}^3), \ \|\psi\|_\infty \leq 1 \right\}.$$

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

$$BV(\Omega; \mathbb{R}^n) = \left\{ u \in L^1(\Omega; \mathbb{R}^n) \Big| \ |Du|(\Omega) < \infty \right\}.$$

Note that Du is a $n \times 3$ matrix of Radon measures with $\Omega \subset \mathbb{R}^3$. If u is differentiable or if $u \in W^{1,1}(\Omega, \mathbb{R}^n)$, then $Du = \nabla u = (\frac{\partial u_i}{\partial x_k})_{i=1,\ldots,n, k=1,2,3}$ in the distributional sense, and

$$|Du|(\Omega) = \int_\Omega \sqrt{\sum_{i=1}^n |\nabla u_i(x)|^2} dx.$$

The total variation has been successfully introduced and used in image restoration for gray-scale images by Rudin, Osher, Fatemi [39]. The vectorial total variation for color images has been analyzed in Blomgren-Chan [8] and the PhD thesis manuscripts of Blomgren [7] and Tschumperlé [43], together with other regularizations for vector-valued data, such as regularizations based on the structure tensor (see also work by Tschumperlé and Deriche [44], [45]).

3.1 The first model

In this model we will denoise the collected data S_t in (1). In what follows, we will drop the subscript t and denote the data by S and f will be its corresponding function notation. The functional that we want to minimize is the following: given $f_0 \geq 0$, $f_0 \in L^1(\Omega)$, and $f \in L^1(\Omega; \mathbb{R}^n)$, for $u = (u_1, \ldots, u_n)$,

$$\inf_u F(u) = |Du|(\Omega) + \lambda \int_\Omega \sum_{i=1}^n |f_i(x) - u_i(x)| dx, \tag{5}$$

subject to the constraints

$$0 \leq u_i(x) \leq f_0(x), \ x \in \Omega, \ i = 1, \ldots, n.$$

A more general regularization The regularization that we use in (5) is mathematically well studied and has the property of allowing discontinuities along hypersurfaces. The vectorial total variation regularization is a particular case of the following regularization for vector-valued data which utilizes eigenvalues, inspired from [43]. For a given $u = (u_1, \ldots, u_n)$, we define $g_u(x)$ to be the matrix

$$
\begin{aligned}
g_u(x) &= \sum_{i=1}^n |\nabla u_i(x)|^2 s_i \cdot s_i^T = \sum_{i=1}^n |\nabla u_i(x)|^2 \begin{pmatrix} s_i^1 \\ s_i^2 \\ s_i^3 \end{pmatrix} \cdot \begin{pmatrix} s_i^1 & s_i^2 & s_i^3 \end{pmatrix} \\
&= \sum_{i=1}^n |\nabla u_i(x)|^2 \begin{pmatrix} (s_i^1)^2 & s_i^1 s_i^2 & s_i^1 s_i^3 \\ s_i^1 s_i^2 & (s_i^2)^2 & s_i^2 s_i^3 \\ s_i^1 s_i^3 & s_i^2 s_i^3 & (s_i^3)^2 \end{pmatrix},
\end{aligned}
$$

where u_i is the intensity value corresponding to the spherical vector s_i as before and

$$s_i = \begin{pmatrix} s_i^1 \\ s_i^2 \\ s_i^3 \end{pmatrix} \in S_d^2.$$

The matrix $g_u(x)$ is different from the usual structure tensor $\sum_{i=1}^n \nabla u_i(x) \nabla u_i(x)^T$ [43] (the structure tensor would not depend on the spherical directions s_i). Here, $g_u(x)$ was chosen so that it would be a measure of intensity variation along the corresponding direction s_i, at each point x. As the structure tensor, the matrix $g_u(x)$ is symmetric and positive semi-definite and we can define

$$\mu_{g_u}(x) = (\mu_1(x), \mu_2(x), \mu_3(x))$$

where $\mu_1(x) \geq \mu_2(x) \geq \mu_3(x) \geq 0$ are the three eigenvalues of the matrix $g_u(x)$. Let Φ be a function from $\{(a, b, c) \in (\mathbb{R}_+ \cup \{0\})^3 : a \geq b \geq c\}$ to $(\mathbb{R}_+ \cup \{0\})$ such that Φ is nondecreasing in each variable and tends to infinity when the first variable tends to infinity. For instance, if $\Phi(a, b, c) = a$, then $\Phi(\mu_{g_u}(x)) = $ largest eigenvalue of the matrix $g_u(x)$. If $\Phi(a, b, c) = c$, then $\Phi(\mu_{g_u}(x)) = $ smallest eigenvalue of the matirx $g_u(x)$. If $\Phi(a, b, c) = a + b + c$, then $\Phi(\mu_{g_u}(x)) = $ trace of the matrix $g_u(x)$. Now we formulate the following general functional

$$
\begin{aligned}
\inf \Big\{ F(u) &= \int_\Omega \Phi(\mu_{g_u}(x)) dx + \lambda \int_{\Omega \times S^2} |f(x, \tau) - u(x, \tau)| dx dA(\tau) \\
&\approx \int_\Omega \Phi(\mu_{g_u}(x)) dx + \lambda \int_\Omega \sum_{s_i \in S_d^2} |f_i(x) - u_i(x)| dx \Big\},
\end{aligned}
$$

6

where the function f and the set S_d^2 are given as before. The models proposed here use $\Phi(\mu_1, \mu_2, \mu_3) = \sqrt{\mu_1 + \mu_2 + \mu_3}$ in which case we have

$$\Phi(\mu_{g_u}(x)) = \sqrt{tr(g_u(x))} = \sqrt{\sum_{i=1}^{n} |\nabla u_i(x)|^2},$$

which realizes the total variation of the n-dimensional vector function $u(x) = (u_1(x), \ldots, u_n(x))$ (but independent of the spherical directions s_i). Note that, in two spatial dimensions, it would be easy to consider other examples of integrands Φ as in [43] (such as $\Phi(\mu) = \sum \sqrt{\mu_i}$, with dependence on the directions s_i), however this becomes very cumbersome in three dimensions in practice, when the eigenvalues no longer have such simple explicit expressions.

The L^1 fidelity term that we use in (5) is different from the more standard L^2 term from [39], that would correspond to additive Gaussian noise. Since the type of noise in MRI data is more complex (usually Rician noise), we think that the L^1 norm is more appropriate in this case instead of the L^2 norm, while easier to implement instead of directly imposing the Rician noise model. For total variation based denoising models with L^1 fidelity term, we refer to [1], [2], [3] in one dimension, and more recently to [40], [36], [10], [9] in the case of planar images, among other work.

In what follows, we will consider an unconstrained version of the first model (5), using the logarithmic barrier method to remove the constraints. The unconstrained version that we will minimize is the following:

$$\inf_u F(u) = |Du|(\Omega) + \lambda \int_\Omega \sum_{i=1}^{n} |f_i(x) - u_i(x)| dx - \mu \int_\Omega \sum_{i=1}^{n} \log(f_0(x) - u_i(x)) dx. \quad (6)$$

Note that the last soft constraint term does not incorporate the non-negativity constraint in (5). We prove that if the function f already satisfies $f = (f_1, \ldots, f_n) \geq 0$ everywhere, then the minimizer u_{min}, if it exists, it also satisfies $u_{min} \geq 0$, due to a maximum principle for our associated diffusion equation.

To simplify this model further, we will require a minor change in the last term of (6). Instead of the integrand given in the last term of (6), we will use

$$\sum_{i=1}^{n} H(f_i(x), f_0(x)) \cdot \log(f_0(x) - u_i(x)).$$

This function H depends only on the two data f and f_0, so the computations will not become more complicated. Our choice of the function H is

$$H(a, b) = \begin{cases} 0 & \text{if } a \leq b, \\ 1 & \text{if } a > b. \end{cases} \quad (7)$$

If we use this weight H, then we only penalize the terms at those points $x \in \Omega$ with $f_i(x) - f_0(x) > 0$ which violate the constraint (2). Therefore, the first minimization problem that we are interested in is

$$\inf_u F(u) = |Du|(\Omega) + \lambda \int_\Omega \sum_{i=1}^{n} |f_i(x) - u_i(x)| dx - \mu \int_\Omega \sum_{i=1}^{n} H(f_i(x), f_0(x)) \cdot \log(f_0(x) - u_i(x)) dx. \quad (8)$$

We can prove existence of minimizers for the constrained problem (5) and for the unconstrained problem (8). In what follows, we will use for $i = 1, \ldots, n$,

$$\Omega_i = \{x \in \Omega | H(f_i(x), f_0(x)) > 0\},$$

and since we assume that $f \in L^1(\Omega; \mathbb{R}^n)$ and $f_0 \in L^1(\Omega)$, we know that Ω_i and $\Omega \backslash \Omega_i$ are measurable sets.

Theorem 1. *Assume that $f = (f_1, \ldots, f_n) \in L^1(\Omega; \mathbb{R}^n)$ and $f_0 \in L^1(\Omega)$ with $f > 0, f_0 > 0$, $-\infty < \int_{\Omega_i} \log f_0 < \infty$ and $\lambda, \mu > 0$. The constrained problem (5) and the unconstrained problem (8) both have minimizers.*

Proof. It is standard to prove the existence of a minimizer for (5) since the functional is bounded from below by 0. Let $\{u^k = (u_1^k, \ldots, u_n^k)\}$ be a minimizing sequence of (5). The boundedness of the domain Ω and the constraints $0 \le u_i^k \le f_0$ for $i = 1, \ldots, n$ and for $k = 1, \ldots$ guarantee that $u^k \in L^1(\Omega; \mathbb{R}^n)$ for all k. Since $F(u^k) \le M$ for all $k = 1, 2, \ldots$, for some positive constant M, we deduce that for all $k \ge 1$,

$$\sum_{i=1}^{n} \|u_i^k\|_{L^1(\Omega)} \le \sum_{i=1}^{n} \|f_i\|_{L^1(\Omega)} + \sum_{i=1}^{n} \|f_i - u_i^k\|_{L^1(\Omega)} \le M.$$

Hence, the sequence is bounded in $BV(\Omega; \mathbb{R}^n)$, and we can have a subsequence (still denoted u^k) and $u \in BV(\Omega; \mathbb{R}^n)$ such that $u^k \to u$ strongly in $L^1(\Omega; \mathbb{R}^n)$ and Du^k converges to Du weakly in the sense of measures. Based on the lower-semicontinuity of the total variation, we deduce that $F(u) \le \liminf_{k \to \infty} F(u^k)$. Moreover, u must satisfy the constraints $0 \le u_i \le f_0$ a.e. $x \in \Omega$ due to the strong convergence of the subsequence u^k to u in $L^1(\Omega; \mathbb{R}^n)$, thus u is a minimizer of the constrained problem (5).

Almost the same steps can be applied to (8), except that we have to make sure that the functional is bounded from below. We have

$$F(u) = |Du|(\Omega) + \lambda \sum_{i=1}^{n} \int_{\Omega \backslash \Omega_i} |f_i(x) - u_i(x)| dx + \sum_{i=1}^{n} \int_{\Omega_i} \Big[\lambda |f_i(x) - u_i(x)| - \mu \log(f_0(x) - u_i(x))\Big] dx.$$

$$\ge |Du|(\Omega) + \lambda \sum_{i=1}^{n} \int_{\Omega \backslash \Omega_i} |f_i(x) - u_i(x)| dx + \sum_{i=1}^{n} \int_{\Omega_i} \Big[\lambda(f_0(x) - u_i(x)) - \mu \log(f_0(x) - u_i(x))\Big] dx.$$

$$(9)$$

Note that for $z > 0$,

$$\frac{\lambda}{\mu} z - \log(z) \ge 1 - \log\left(\frac{\mu}{\lambda}\right).$$

Hence, we can bound the last term of (9) from below:

$$\sum_{i=1}^{n} \int_{\Omega_i} \lambda(f_0(x) - u_i(x)) - \mu \log(f_0(x) - u_i(x)) dx \ge \sum_{i=1}^{n} \mu |\Omega_i| \left(1 - \log\left(\frac{\mu}{\lambda}\right)\right) \ge -n\mu |\Omega| \left|1 - \log\left(\frac{\mu}{\lambda}\right)\right|.$$

Therefore for any $\lambda, \mu > 0$,

$$\inf_u F(u) > -\infty$$

and we can take a minimizing sequence $\{u^k\}$. There exists $0 < M < \infty$ such that $F(u^k) \le M$ for all k. Then the above inequalities show that if we let

$$C = \begin{cases} M + n\mu|\Omega|\left|1 - \log\left(\frac{\mu}{\lambda}\right)\right| & \text{if} \quad 0 < \lambda/\mu < e^{-1}, \\ M & \text{if} \quad e^{-1} \le \lambda/\mu, \end{cases}$$

then

$$\sup_k \left\{|Du^k|(\Omega) + \lambda \sum_{i=1}^n \int_{\Omega \setminus \Omega_i} |f_i(x) - u_i^k(x)|dx\right\} \le C < \infty.$$

It is enough to show that the sequence is bounded in $L^1(\Omega; \mathbb{R}^n)$. It is obvious that for each i,

$$\sup_k \int_{\Omega \setminus \Omega_i} |u_i^k(x)|dx < \infty.$$

What remains to prove is that for each i,

$$\sup_k \int_{\Omega_i} |u_i^k(x)|dx < \infty.$$

Let $v_i^k(x) = f_0(x) - u_i^k(x)$. Then $v_i^k > 0$ on Ω_i. Note that

$$\int_{\Omega_i} \left[\lambda v_i^k(x) - \mu \log(v_i^k(x))\right]dx \le C < \infty.$$

Suppose that for some i,

$$\sup_k \|v_i^k\|_{L^1(\Omega_i)} = \infty.$$

Since the function $z \mapsto \lambda z - \mu \log(z)$ is convex for $z > 0$, Jensen's inequality implies

$$\infty = \limsup_{k \to \infty} \left\{\lambda \frac{\|v_i^k\|_{L^1(\Omega_i)}}{|\Omega_i|} - \mu \log\left(\frac{\|v_i^k\|_{L^1(\Omega_i)}}{|\Omega_i|}\right)\right\}$$
$$\le \limsup_{k \to \infty} \left\{\frac{1}{|\Omega_i|} \int_{\Omega_i} \left[\lambda v_i^k(x) - \mu \log(v_i^k(x))\right]dx\right\} \le \frac{C}{|\Omega_i|} < \infty.$$

This is a contradiction. Hence, $\{u^k\}$ is bounded in $BV(\Omega; \mathbb{R}^n)$ and there exists $u \in BV(\Omega; \mathbb{R}^n)$ such that (up to a subsequence) $u^k \to u$ in $L^1(\Omega; \mathbb{R}^n)$ and

$$|Du|(\Omega) \le \liminf_{k \to \infty} |Du^k|(\Omega),$$

thus

$$F(u) = \inf_v F(v).$$

$\square$

The purpose of the function H is to impose as few constraints as possible to minimize the effect caused by the logarithmic functions. The following theorem ensures that the function H does the job.

Theorem 2. *Let $\lambda, \mu > 0$, such that $\lambda/\mu \geq e^{-1}$ and let $f > 0$, $f_0 \geq \mu/\lambda$. Then any minimizer u of (8) satisfies $u \geq 0$. In particular, if we replace the function H in (8) by $H \equiv 1$, then (8) becomes (6) and any minimizer u of (6) satisfies $0 \leq u \leq f_0$.*

Proof. Let u be a minimizer of (8). Let $v = \max(u, 0)$. We claim that

$$F(v) \leq F(u).$$

It is easy to see that

$$|Dv|(\Omega) \leq |Du|(\Omega) \quad \text{and} \quad \int_{\Omega \setminus \Omega_i} |f_i(x) - v_i(x)| dx \leq \int_{\Omega \setminus \Omega_i} |f_i(x) - u_i(x)| dx.$$

What remains to prove is that

$$\int_{\Omega_i} \lambda |f_i(x) - v_i(x)| - \mu \log(f_0(x) - v_i(x)) dx \leq \int_{\Omega_i} \lambda |f_i(x) - u_i(x)| - \mu \log(f_0(x) - u_i(x)) dx.$$

Let $g(z) = (\lambda/\mu)z - \log(z)$ for $z > 0$. By the assumption, $g(z) \geq 0$ and g is increasing on $\{z \geq \mu/\lambda\}$, which implies on Ω_i

$$
\begin{aligned}
\lambda |f_i(x) - v_i(x)| - \mu \log(f_0(x) - v_i(x)) &= \lambda(f_i(x) - f_0(x)) + \mu g(f_0(x) - v_i(x)) \\
&\leq \lambda(f_i(x) - f_0(x)) + \mu g(f_0(x) - u_i(x)) \\
&= \lambda |f_i(x) - u_i(x)| - \mu \log(f_0(x) - u_i(x)).
\end{aligned}
$$

Therefore, $F(v) \leq F(u)$ which implies that v is also a minimizer. Since $F(v) = F(u)$, so all the inequalities above become equalities and we have $v = u$ a.e., that is, any minimizer u satisfies $u \geq 0$.

In particular, if we take $H \equiv 1$, then (8) is the same as (6) and $\Omega_i = \Omega$ for all i. The logarithmic term in (6) guarantees that any minimizer u has to satisfy $u \leq f_0$. Therefore, any minimizer u of (6) satisfies $0 \leq u \leq f_0$. $\qquad \square$

3.1.1 Euler-Lagrange equations and a numerical algorithm

The Euler-Lagrange equations associated with the minimization (8) are:

$$-\nabla \cdot \frac{\nabla u_i(x)}{\sqrt{\sum_{j=1}^n |\nabla u_j(x)|^2}} - \lambda \frac{f_i(x) - u_i(x)}{|f_i(x) - u_i(x)|} + \mu \frac{H(f_i(x), f_0(x))}{f_0(x) - u_i(x)} = 0 \quad \text{for } i = 1, \ldots, n. \tag{10}$$

To find a minimizer numerically, we impose Neumann boundary conditions and use the gradient descent method and the finite difference scheme to solve the evolution PDEs for $i = 1, \ldots, n$, where ∇ and $\nabla \cdot$ refer to the spatial gradient and divergence, respectively:

$$\frac{\partial u_i}{\partial t}(t, x) = \nabla \cdot \frac{\nabla u_i(t, x)}{\sqrt{\sum_{j=1}^n |\nabla u_j(t, x)|^2}} + \lambda \frac{f_i(x) - u_i(t, x)}{|f_i(x) - u_i(t, x)|} - \mu \frac{H(f_i(x), f_0(x))}{f_0(x) - u_i(t, x)}. \tag{11}$$

For $1 \leq \alpha \leq p$, $1 \leq \beta \leq q$ and $1 \leq \gamma \leq r$, where the dimension of the domain is $p \times q \times r$, let

$$\nabla u_i^k(\alpha, \beta, \gamma) = \begin{pmatrix} u_i^k(\alpha + 1, \beta, \gamma) - u_i^k(\alpha, \beta, \gamma) \\ u_i^k(\alpha, \beta + 1, \gamma) - u_i^k(\alpha, \beta, \gamma) \\ u_i^k(\alpha, \beta, \gamma + 1) - u_i^k(\alpha, \beta, \gamma) \end{pmatrix}.$$

For simplicity, we define $c_1^k, c_2^k, c_3^k, c_4^k$ by

$$c_1^k(\alpha, \beta, \gamma) = \sqrt{\sum_{j=1}^{n} \begin{pmatrix} (u_j^k(\alpha+1, \beta, \gamma) - u_j^k(\alpha, \beta, \gamma))^2 \\ +(u_j^k(\alpha, \beta+1, \gamma) - u_j^k(\alpha, \beta, \gamma))^2 \\ +(u_j^k(\alpha, \beta, \gamma+1) - u_j^k(\alpha, \beta, \gamma))^2 \end{pmatrix}},$$

$$c_2^k(\alpha, \beta, \gamma) = \sqrt{\sum_{j=1}^{n} \begin{pmatrix} (u_j^k(\alpha, \beta, \gamma) - u_j^k(\alpha-1, \beta, \gamma))^2 \\ +(u_j^k(\alpha-1, \beta+1, \gamma) - u_j^k(\alpha-1, \beta, \gamma))^2 \\ +(u_j^k(\alpha-1, \beta, \gamma+1) - u_j^k(\alpha-1, \beta, \gamma))^2 \end{pmatrix}},$$

$$c_3^k(\alpha, \beta, \gamma) = \sqrt{\sum_{j=1}^{n} \begin{pmatrix} (u_j^k(\alpha+1, \beta-1, \gamma) - u_j^k(\alpha, \beta-1, \gamma))^2 \\ +(u_j^k(\alpha, \beta, \gamma) - u_j^k(\alpha, \beta-1, \gamma))^2 \\ +(u_j^k(\alpha, \beta-1, \gamma+1) - u_j^k(\alpha, \beta-1, \gamma))^2 \end{pmatrix}},$$

$$c_4^k(\alpha, \beta, \gamma) = \sqrt{\sum_{j=1}^{n} \begin{pmatrix} (u_j^k(\alpha+1, \beta, \gamma-1) - u_j^k(\alpha, \beta, \gamma-1))^2 \\ +(u_j^k(\alpha, \beta+1, \gamma-1) - u_j^k(\alpha, \beta, \gamma-1))^2 \\ +(u_j^k(\alpha, \beta, \gamma) - u_j^k(\alpha, \beta, \gamma-1))^2 \end{pmatrix}}.$$

For the divergence term, we use the forward and backward difference and we will employ an explicit scheme to obtain

$$\begin{aligned}
\frac{u_i^{k+1}(\alpha, \beta, \gamma) - u_i^k(\alpha, \beta, \gamma)}{dt} &= \frac{u_i^k(\alpha+1, \beta, \gamma) - u_i^k(\alpha, \beta, \gamma)}{c_1^k(\alpha, \beta, \gamma)} - \frac{u_i^k(\alpha, \beta, \gamma) - u_i^k(\alpha-1, \beta, \gamma)}{c_2^k(\alpha, \beta, \gamma)} \\
&+ \frac{u_i^k(\alpha, \beta+1, \gamma) - u_i^k(\alpha, \beta, \gamma)}{c_1^k(\alpha, \beta, \gamma)} - \frac{u_i^k(\alpha, \beta, \gamma) - u_i^k(\alpha, \beta-1, \gamma)}{c_3^k(\alpha, \beta, \gamma)} \\
&+ \frac{u_i^k(\alpha, \beta, \gamma+1) - u_i^k(\alpha, \beta, \gamma)}{c_1^k(\alpha, \beta, \gamma)} - \frac{u_i^k(\alpha, \beta, \gamma) - u_i^k(\alpha, \beta, \gamma-1)}{c_4^k(\alpha, \beta, \gamma)} \\
&+ \lambda \cdot sign(f_i(\alpha, \beta, \gamma) - u_i^k(\alpha, \beta, \gamma)) - \mu \cdot \frac{H(f_i(\alpha, \beta, \gamma), f_0(\alpha, \beta, \gamma))}{f_0(\alpha, \beta, \gamma) - u_i^k(\alpha, \beta, \gamma)}.
\end{aligned}$$

Experimental results using this discrete model 1 for HARDI denoising will be shown in Section 4.

3.2 The second model

In this model, we will denoise the sADC (spherical Apparent Diffusion Coefficient) d_t in (1). In what follows, we will also drop the subscript t and denote the sADC by d and will borrow the same notations f, f_0 from the previous section. In this section, we also assume, for theoretical purposes, that there exist $0 < c_1, c_2 < \infty$ such that

$$c_1 \leq f \leq c_2, \quad c_1 \leq f_0 \leq c_2.$$

Our proposed second minimization problem is,

$$\inf_d G(d) = |Dd|(\Omega) + \lambda \int_\Omega \sum_{i=1}^{n} |f_i(x) - f_0(x) \exp(-d_i(x))| dx \tag{12}$$

11

subject to the constraints

$$d_i \geq 0, \quad i = 1, \ldots, n.$$

Note that d in (12) realizes $b \cdot d$ in (1) and this is not a modification since b is a constant. Also note from the HARDI data model (1) that knowing the true S is equivalent to knowing the true d, which lead us to consider the problem of denoising d instead of denoising S. However, this model seems to have some advantages over the first model. First of all, d is being used as an exponent of the exponential function, which says that the range of the values of d in (12) is much smaller than that of u in (5) so that when we minimize the two functionals F and G, G does not seem to smooth out d as much as F does to u. Secondly, we directly impose the assumed image formation model. Thirdly, the constraints in (12) are simpler than those in (5). Hence, if we look at an unconstrained version of the second model obtained in the same way as in the first model, then we end up minimizing

$$\inf_d G(d) = |Dd|(\Omega) \quad + \quad \lambda \int_\Omega \sum_{i=1}^n |f_i(x) - f_0(x) \exp(-d_i(x))| dx$$

$$- \quad \mu \int_\Omega \sum_{i=1}^n H(f_i(x), f_0(x)) \cdot \log(d_i(x)) dx. \tag{13}$$

Unfortunately, the problem (13) has no global minimizer since for any $d \in BV(\Omega; \mathbb{R}^n)$ with $G(d) < \infty$, if $|\Omega_i| > 0$ for some i, then

$$\lim_{c \to \infty} G(d + c) = -\infty.$$

To prevent this unboundedness, a slight modification is needed, which is the following. We modify G with the function H defined in (7) by

$$\inf_d G(d) = |Dd|(\Omega) \quad + \quad \lambda \int_\Omega \sum_{i=1}^n |f_i(x) - f_0(x) \exp(-d_i(x))| dx$$

$$- \quad \mu \int_\Omega \sum_{i=1}^n H(f_i(x), f_0(x)) \cdot \frac{\log(d_i(x))}{d_i(x)} dx. \tag{14}$$

An advantage of the function $-\mu \log(z)/z$ is that on one hand, unlike the logarithmic function $-\mu \log(z)$, it rather spreads uniform weights on $\{z > \epsilon\}$ for $\epsilon > 0$ when $\mu > 0$ is small. On the other hand, $-\mu \log(z)/z$ generates more repelling force from $z = 0$ than $-\mu \log(z)$. In this sense, $-\mu \log(z)/z$ realizes the constraints in (12) better than $-\mu \log(z)$ does.

A discretized version G_D will be

$$G_D(d) = \sum_{x \in \Omega} \sqrt{\sum_{i=1}^n |\nabla d_i(x)|^2} \quad + \quad \lambda \sum_{x \in \Omega} \sum_{i=1}^n |f_i(x) - f_0(x) \exp(-d_i(x))|$$

$$- \quad \mu \sum_{x \in \Omega} \sum_{i=1}^n H(f_i(x), f_0(x)) \cdot \frac{\log(d_i(x))}{d_i(x)}, \tag{15}$$

where Ω is the spatial domain, $x = (x_1, x_2, x_3)$, and

$$
\begin{aligned}
|\nabla d_i(x)|^2 &= (d_i(x_1 + 1, x_2, x_3) - d_i(x_1, x_2, x_3))^2 + (d_i(x_1, x_2 + 1, x_3) - d_i(x_1, x_2, x_3))^2 \\
&+ (d_i(x_1, x_2, x_3 + 1) - d_i(x_1, x_2, x_3))^2.
\end{aligned}
$$

First of all, we can prove that there is a minimizer of the constrained problem (12).

Theorem 3. *The constrained problem* (12) *has a minimizer* $d \in BV(\Omega; \mathbb{R}^n)$.

Proof. In the proof we assume for simplicity that a sequence and its subsequence have the same notations. Let $\{d^k\}$ be a minimizing sequence. Note that $d^k \geq 0$. It is enough to show that for each i, there exists $d_i \in L^1(\Omega)$ with

$$
|Dd_i|(\Omega) = \sup\left\{ \int_\Omega d_i \mathrm{div}(\psi) \ : \ \psi \in C_c^1(\Omega, \mathbb{R}^3), \ |\psi| \leq 1 \right\} < \infty,
$$

and for $d = (d_1, \ldots, d_n)$,

$$
G(d) = \lim_{k \to \infty} G(d^k).
$$

Since $\{d^k\}$ is a minimizing sequence, there exists $M < \infty$ such that $G(d^k) \leq M$ for all k. Based on Poincaré-Wirtinger inequality, each d_i^k satisfies

$$
\left| \|d_i^k\|_{L^2(\Omega)} - \|d_i^k\|_{L^1(\Omega)} \right| \leq \|d_i^k - (d_i^k)_\Omega\|_{L^2(\Omega)} \lesssim |Dd_i^k|(\Omega) \leq M \tag{16}
$$

where $d_i^k \geq 0$ and $(d_i^k)_\Omega = \frac{1}{|\Omega|} \int_\Omega d_i^k(x) dx$.

Let $\tilde{d}_i^k = \min(d_i^k, \log(c_2/c_1))$ for each i. Then $\tilde{d}^k \geq 0$, $\sup_k \|\tilde{d}^k\|_\infty \leq \sqrt{n} \log(c_2/c_1)$, and

$$
|D\tilde{d}^k|(\Omega) \leq |Dd^k|(\Omega).
$$

If $d_i^k(x) > \log(c_2/c_1) = \tilde{d}_i^k(x)$, then

$$
\begin{aligned}
f_0(x)\left(\frac{f_i(x)}{f_0(x)} - \exp(-d_i^k(x)) \right) &> f_0(x)\left(\frac{f_i(x)}{f_0(x)} - \exp(-\tilde{d}_i^k(x)) \right) \\
&= f_0(x)\left(\frac{f_i(x)}{f_0(x)} - \exp(-\log(c_2/c_1)) \right) \\
&\geq f_0(x)\left(\frac{f_i(x)}{f_0(x)} - \frac{c_1}{c_2} \right) \geq 0.
\end{aligned}
$$

That is,

$$
|f_i(x) - f_0(x)\exp(-\tilde{d}_i^k(x))| \leq |f_i(x) - f_0(x)\exp(-d_i^k(x))|,
$$

which finally implies that

$$
G(\tilde{d}^k) \leq G(d^k).
$$

Hence, the sequence $\{\tilde{d}^k\}$ is also a minimizing sequence which is uniformly bounded in $L^1(\Omega; \mathbb{R}^n)$. In other words, the sequence $\{\tilde{d}_i^k\}$ is bounded in $BV(\Omega)$. Therefore, there exists $d_i \in BV(\Omega)$ for each i such that

$$
\tilde{d}_i^k \to d_i \text{ in } L^1(\Omega) \quad \text{and} \quad |Dd_i|(\Omega) \leq \liminf_{k \to \infty} |D\tilde{d}_i^k|(\Omega).
$$

13

Note that $d = (d_1, \ldots, d_n) \geq 0$.

In addition to the strong L^1 convergence of the sequence $\{\tilde{d}_i^k\}$, we may say that $d_i^k \to d_i$ pointwise a.e. since the domain Ω is finite, which guarantees the convergence of the fidelity term in the functional,

$$\lambda \int_\Omega \sum_{i=1}^n |f_i(x) - f_0(x) \exp(-d_i(x))| dx.$$

Therefore, d is a minimizer. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The purpose of imposing the logarithmic constraint to problem (14) is to insure that $d \geq 0$ is satisfied. It turns out that it is sufficient to impose the constraint on a partial domain $\Omega_i = \{x \in \Omega : H(f_i(x), f_0(x)) > 0\}$. Now we prove the existence of a minimizer for (14).

Theorem 4. *The unconstrained problem* (14) *has a minimizer* $d \in BV(\Omega; \mathbb{R}^n)$. *Moreover, any minimizer h of* (14) *satisfies* $h \geq 0$.

Proof. We prove first that the functional G in (14) is bounded from below. It suffices to prove that there exists $-\infty < C < 0$ such that for each i,

$$-\int_{\Omega_i} \frac{\log(d_i(x))}{d_i(x)} dx > C.$$

Let $g(z) = -\log(z)/z$ for $z > 0$. Then $g(z) \geq -\frac{1}{e}$. Hence, if we let $C = -|\Omega|/e$, then

$$-\int_{\Omega_i} \frac{\log(d_i(x))}{d_i(x)} dx \geq -\frac{|\Omega_i|}{e} \geq C.$$

Next, we may assume that $\log(c_2/c_1) \geq e$. Otherwise, we can always choose $c_0 > 0$ so that $c_0 < c_1 \leq f \leq c_2$, $c_0 < c_1 \leq f_0 \leq c_2$ and $\log(c_2/c_0) \geq e$. Then we set c_1 to be c_0. Now we choose a minimizing sequence $\{d^k\}$. Let $\tilde{d}^k = \min(d^k, \log(c_2/c_1))$. As was shown in Theorem 3,

$$|D\tilde{d}^k|(\Omega) \leq |Dd^k|(\Omega) \text{ and } |f_i(x) - f_0(x) \exp(-\tilde{d}_i^k(x))| \leq |f_i(x) - f_0(x) \exp(-d_i^k(x))|.$$

If, in addition, $d_i^k(x) > \tilde{d}_i^k(x) = \log(c_2/c_1) \geq e$ for some $x \in \Omega_i$, then

$$g(d_i^k(x)) > g(\tilde{d}_i^k(x)) = g(\log(c_2/c_1))$$

since $g(z)$ is increasing on $\{z \geq e\}$.

Also if we let $v^k = \max(\tilde{d}^k, 0)$, then $|Dv^k|(\Omega) \leq |D\tilde{d}^k|(\Omega)$ and for $x \in \Omega \setminus \Omega_i$,

$$|f_i(x) - f_0(x) \exp(-v_i^k(x))| \leq |f_i(x) - f_0(x) \exp(-\tilde{d}_i^k(x))|.$$

Therefore, $G(v^k) \leq G(\tilde{d}^k) \leq G(d^k)$. Since $\{v^k\}$ is also a minimizing sequence which is uniformly bounded above and below on a bounded set Ω, the same argument as in Theorem 3 guarantees the existence of a minimizer $d \in BV(\Omega; \mathbb{R}^n)$ with $d \geq 0$. Let h be any minimizer of (14). Then $\tilde{h} = \max(h, 0)$ is also a minimizer as shown above and it is easy to see that $\tilde{h} = h$ a.e. Hence, any minimizer h satisfies $h \geq 0$. $\qquad\qquad\square$

14

3.2.1 Euler-Lagrange equations and a numerical algorithm

The Euler-Lagrange equations for the model (14) are similar to those found in the previous section, which are for $i = 1, \ldots, n,$

$$-\nabla \cdot \left(\frac{\nabla d_i(x)}{\sqrt{\sum_{j=1}^{n} |\nabla d_j(x)|^2}} \right) \quad + \quad \frac{\lambda \cdot f_0(x) \exp\left(-d_i(x)\right)(f_i(x) - f_0(x) \exp\left(-d_i(x)\right))}{|f_i(x) - f_0(x) \exp\left(-d_i(x)\right)|}$$

$$- \quad \mu \frac{H(f_i(x), f_0(x))(1 - \log(d_i(x)))}{d_i(x)^2} = 0$$

with Neumann boundary conditions. We end up solving the following PDEs for $i = i, \ldots, n,$

$$\begin{aligned}
\frac{\partial d_i}{\partial t}(t, x) &= \nabla \cdot \left(\frac{\nabla d_i(t, x)}{\sqrt{\sum_{j=1}^{n} |\nabla d_j(t, x)|^2}} \right) - \frac{\lambda f_0(x) \exp\left(-d_i(t, x)\right)(f_i(x) - f_0(x) \exp\left(-d_i(t, x)\right))}{|f_i(x) - f_0(x) \exp\left(-d_i(t, x)\right)|} \\
&\quad + \mu \frac{H(f_i(x), f_0(x))(1 - \log(d_i(t, x)))}{d_i(t, x)^2} \\
&= \nabla \cdot \left(\frac{\nabla d_i(t, x)}{\sqrt{\sum_{j=1}^{n} |\nabla d_j(t, x)|^2}} \right) - \lambda \cdot f_0(x) \exp\left(-d_i(t, x)\right) \cdot sign(f_i(x) - f_0(x) \exp\left(-d_i(t, x)\right)) \\
&\quad + \mu \frac{H(f_i(x), f_0(x))(1 - \log(d_i(t, x)))}{d_i(t, x)^2}
\end{aligned} \tag{17}$$

based on the gradient descent method.

As before, we use finite differences to discretize the PDE's. We implement an explicit scheme for the divergence term. With the same notations as in Subsection 3.1.1, we can obtain the following:

$$\begin{aligned}
\frac{d_i^{k+1}(\alpha, \beta, \gamma) - d_i^k(\alpha, \beta, \gamma)}{dt} &= \frac{d_i^k(\alpha + 1, \beta, \gamma) - d_i^k(\alpha, \beta, \gamma)}{c_1^k(\alpha, \beta, \gamma)} - \frac{d_i^k(\alpha, \beta, \gamma) - d_i^k(\alpha - 1, \beta, \gamma)}{c_2^k(\alpha, \beta, \gamma)} \\
&\quad + \frac{d_i^k(\alpha, \beta + 1, \gamma) - d_i^k(\alpha, \beta, \gamma)}{c_1^k(\alpha, \beta, \gamma)} - \frac{d_i^k(\alpha, \beta, \gamma) - d_i^k(\alpha, \beta - 1, \gamma)}{c_3^k(\alpha, \beta, \gamma)} \\
&\quad + \frac{d_i^k(\alpha, \beta, \gamma + 1) - d_i^k(\alpha, \beta, \gamma)}{c_1^k(\alpha, \beta, \gamma)} - \frac{d_i^k(\alpha, \beta, \gamma) - d_i^k(\alpha, \beta, \gamma - 1)}{c_4^k(\alpha, \beta, \gamma)} \\
&\quad - \lambda \cdot f_0(\alpha, \beta, \gamma) \exp(-d_i^k(\alpha, \beta, \gamma)) \cdot Sg_i^k(\alpha, \beta, \gamma) \\
&\quad + \mu \cdot \frac{H(f_i(\alpha, \beta, \gamma), f_0(\alpha, \beta, \gamma))(1 - \log(d_i^k(\alpha, \beta, \gamma)))}{d_i^k(\alpha, \beta, \gamma)^2},
\end{aligned}$$

where $Sg_i^k(\alpha, \beta, \gamma) = sign(f_i(\alpha, \beta, \gamma) - f_0(\alpha, \beta, \gamma) \exp(-d_i^k(\alpha, \beta, \gamma)))$. Experimental results using this HARDI denoising model 2 will be shown in the next section.

4 Numerical results

We recall that in practice, we work with a decreasing sequence of values $\{\mu_k\}_{k=1,2,\dots}$ and find minimizers u_k^* or d_k^* (depending on the choice of a functional between F_k and G_k), where

$$
\begin{aligned}
F_k(u) = |Du|(\Omega) \quad &+ \quad \lambda \int_\Omega \sum_{i=1}^n |f_i(x) - u_i(x)| dx \\
&- \quad \mu_k \int_\Omega \sum_{i=1}^n H(f_i(x), f_0(x)) \cdot \log(f_0(x) - u_i(x)) dx,
\end{aligned}
$$

$$
\begin{aligned}
G_k(d) = |Dd|(\Omega) \quad &+ \quad \lambda \int_\Omega \sum_{i=1}^n |f_i(x) - f_0(x) \exp(-d_i(x))| dx \\
&- \quad \mu_k \int_\Omega \sum_{i=1}^n H(f_i(x), f_0(x)) \cdot \frac{\log(d_i(x))}{d_i(x)} dx.
\end{aligned}
$$

The initial values $\mu_0 \sim 0.000001$ and $\mu_0 \sim 0.00000001$ were chosen for each case respectively, and the iterations were performed until μ_k reaches $10^{-6}\mu_0$ with $\mu_{k+1} = 10^{-1}\mu_k$ or until we reached a possible steady state. We would like to mention that, even though the denoised data obtained from the second model minimizing the functional G is the function d, we use the computed data

$$
(f_0 \exp(-d_1), \dots, f_0 \exp(-d_n))
$$

to visualize the results and to compute the root-mean-square-error (RMSE) values and the sKL distance, to be explained below. For visualization, we show both the raw HARDI plots, and also the ODF plots obtained by postprocessing. The ODFs are computed from the raw HARDI data using a truncated spherical harmonic series.

Since the functionals (8) and (14) are obviously nonconvex, there might be many local minima, which might cause visibly unsatisfying results or some computational instability. We thus need to choose an appropriate initial guess at $t = 0$. We notice that the feasible domain for (8) contains the set

$$
\{u \in BV(\Omega) : 0 \le u < f_0\}
$$

that might contain a minimizer u^* of (8) and the feasible domain for (14) contains the set

$$
\{d \in BV(\Omega) : 0 \le d\}
$$

that contains a minimizer d^* of (14). Numerical results show that the computed minimizer u^* of (8) also satisfies

$$
0 \le u^* \le f_0.
$$

This led us to choose the initial guess u_0 for (8) and the initial guess d_0 for (14) defined by

$$
(u_0)_i(x) = \begin{cases} 0.8 \times f_0(x) & \text{on } \Omega_i, \\ 1.1 \times f_i(x) & \text{on } \Omega \setminus \Omega_i \quad \text{if } 1.1 \times f_i(x) < f_0(x), \\ f_i(x) & \text{on } \Omega \setminus \Omega_i \quad \text{if } 1.1 \times f_i(x) \ge f_0(x). \end{cases}
$$

16

and

$$(d_0)_i(x) = \begin{cases} 0.005 & \text{on } \Omega_i, \\ -\log(f_i(x)/f_0(x)) + 0.1 & \text{on } \Omega \setminus \Omega_i. \end{cases}$$

We first show denoising results on a synthetic data set that was kindly provided to us by the authors of [32] who generated the data set using the technique described in [38]. The authors of [32] proposed the following variational principle for smoothing the data f

$$\begin{aligned} \min_u E(u) \;=\;& \mu \int_\Omega \int_{S^2} |u(x,\theta,\phi) - f(x,\theta,\phi)|^2 dS(\theta,\phi) dx \\ &+ \int_{S^2} \|\nabla_{(\theta,\phi)} u\|^2 dS(\theta,\phi) + \int_\Omega g(x) \|\nabla_x u\| dx \end{aligned}$$

and approached this model with a membrane-spline deformation energy minimization method. They used the finite element method (FEM method) for smoothing the spherical information at each voxel. For spatial smoothing, the following variational principle (TV method) was used

$$\min_u \mathcal{E}(u) = \int_\Omega \left(g(x) \sum_{i=1}^n \|\nabla u_i(x)\| dx + \frac{\mu}{2} \sum_{i=1}^n |u_i(x) - f_i(x)|^2 \right) dx \tag{18}$$

where n is the number of the gradient directions as in our proposed models. In (18), the coupling between different directions is introduced not through the vectorial total variation, but through the function g that is defined by

$$g(x) = \frac{1}{1 + GA(x)}$$

where GA is the generalized anisotropy defined in [38]. The authors of [32] also kindly provided us with their two denoising results, one obtained by the FEM method only and the other one obtained by the combination of FEM and TV methods. These are shown in Figure 7 and Figure 8 for comparison with our denoising results. Table 1 shows the RMSE values for each model. The RMSE formula is presented in (19).

The size of the synthetic data set is $16 \times 16 \times 81$, which is a set of 81 2D-images of size 16×16. That is, there are 81 gradient directions. In Figure 1 and Figure 2, we show two different angle views of the synthetic, noise-free raw HARDI data and its noisy version. In Figure 3 and Figure 4, we show the same angle views obtained from the denoised raw HARDI data using the proposed two models. All the results shown in Figures 1 to 10 were obtained from the $16 \times 16 \times 81$ synthetic data; however, we visualize only 1/3 of the whole volume in Figures 1 - 8, to better see the finer details. Figure 9 shows the ODF (Orientation Distribution Function) visualization of the whole volume of the data and Figure 10 is a magnified view of Figure 9.

One way to measure the difference between two data sets uses RMSE (Root Mean Square Error). We compute the RMSE values between the noise-free data and the noisy data, and between the noise free data an the denoised data. The RMSE values corresponding to the synthetic experimental results are shown in Table 1. Since the intensity values in the data range between 0 and 1, we multiplied all the intensity values by 255 and then we computed the RMSE values in the following way: let $S_d^2 = \{s_1, s_2, \ldots, s_n\}$ be the discrete set of gradient spherical directions and let $\Omega = \{x_1, x_2, \ldots, x_N\}$ be a subset of $\mathbb{R}^2$ or $\mathbb{R}^3$, a discrete version of the spatial domain. Then the
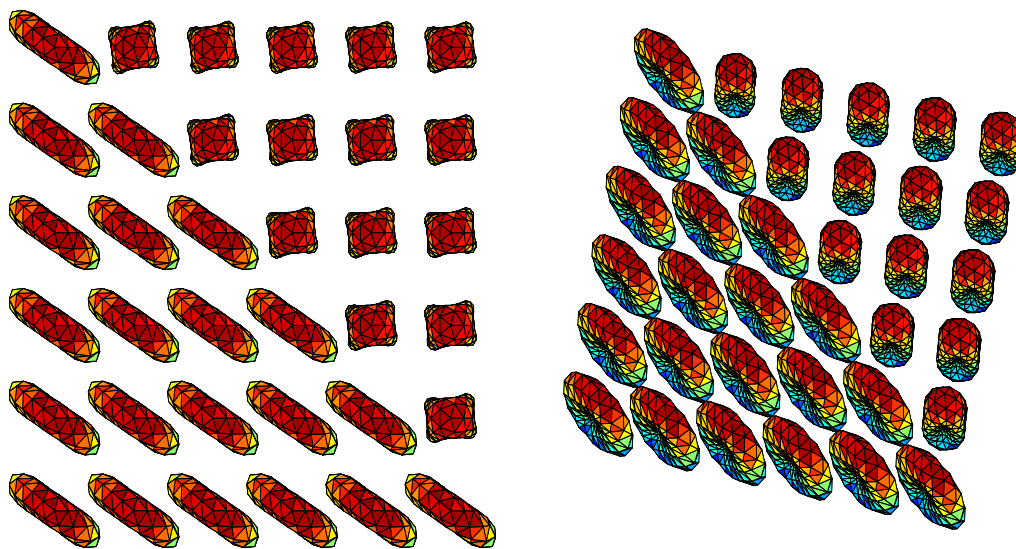
17

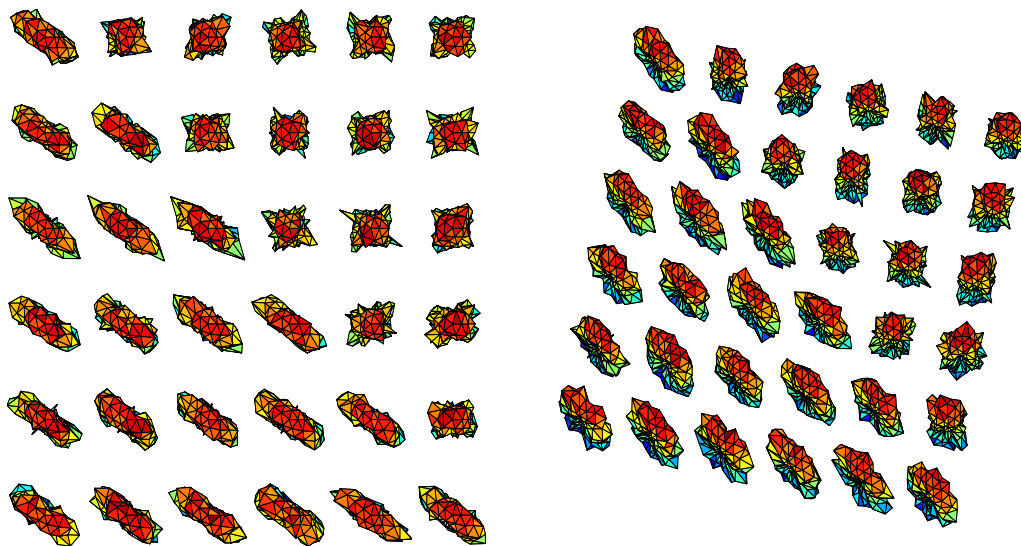Figure 1: Two different angle views of noise-free raw synthetic HARDI data.

Figure 2: Two different angle views of noisy raw synthetic HARDI data.
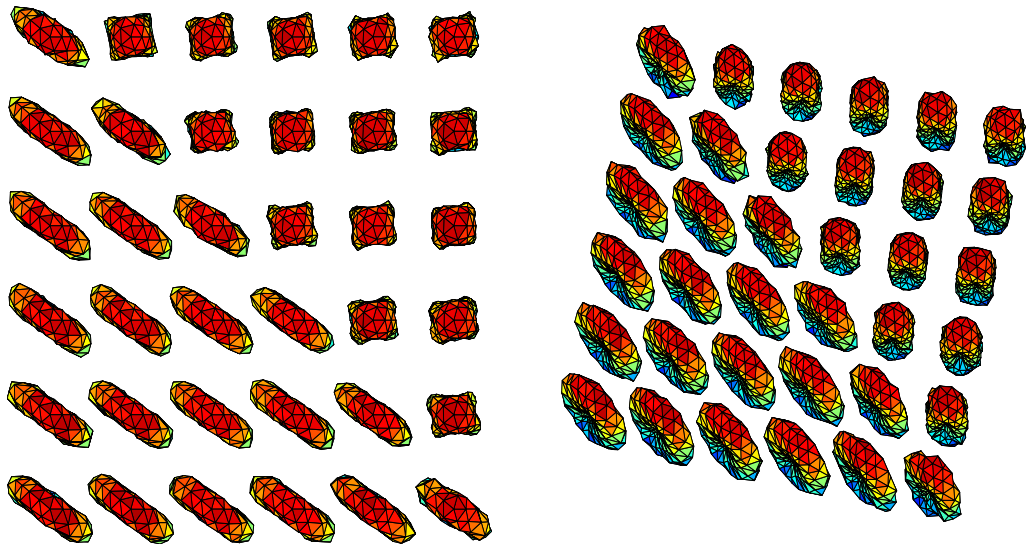
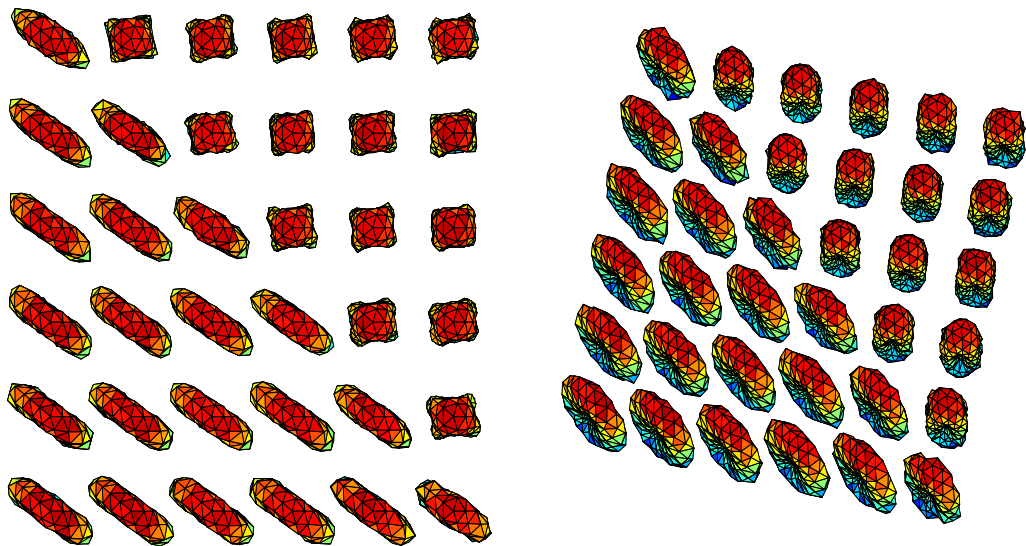Figure 3: Two different angle views of denoised raw synthetic HARDI data by our first model using F.

Figure 4: Two different angle views of denoised raw synthetic HARDI data by our second model using G.
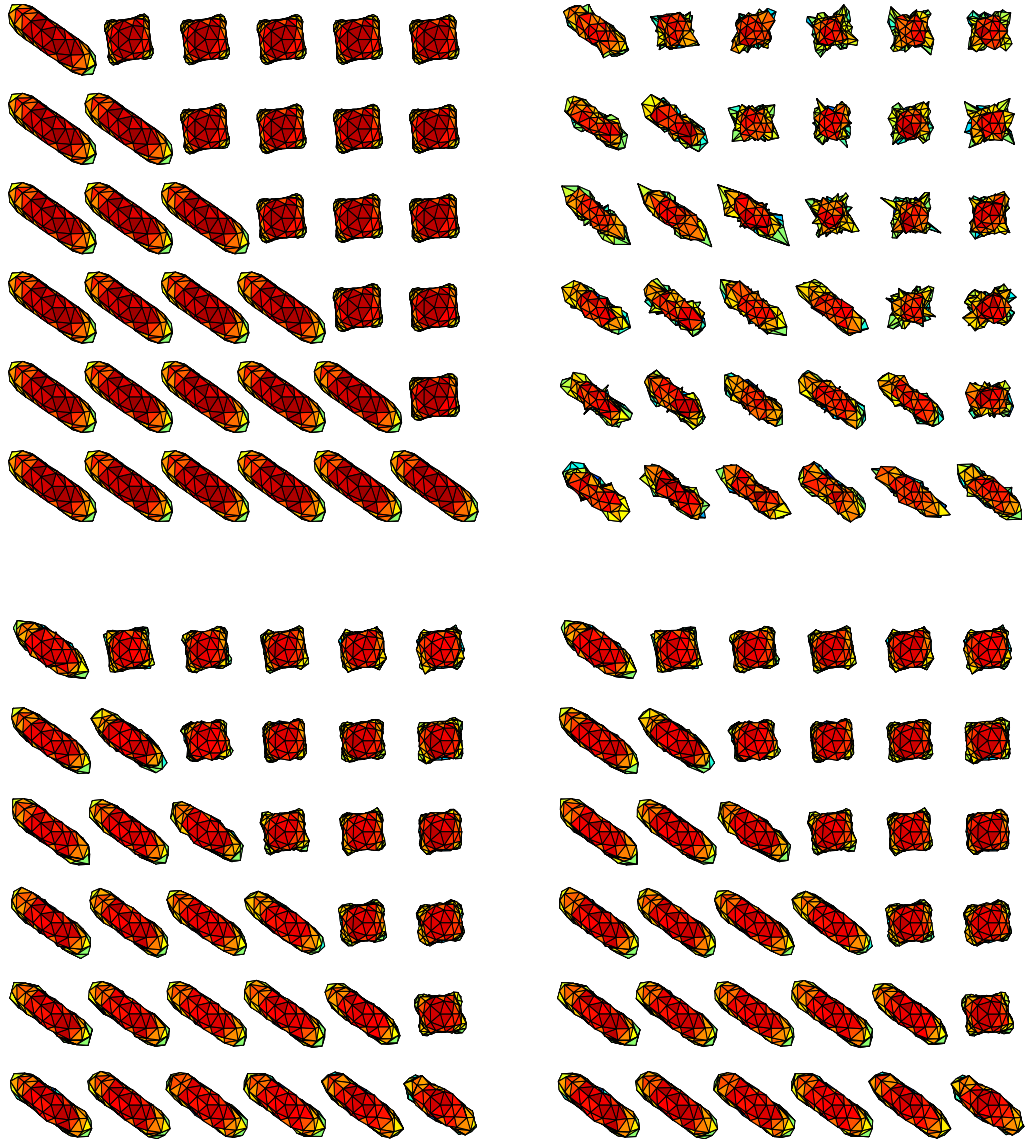
Figure 5: Comparison between models F and G. Top row, left to right: noise-free synthetic raw HARDI data, noisy synthetic raw HARDI data. Bottom row, left to right: denoised synthetic raw HARDI data using model F, denoised synthetic raw HARDI data using model G.
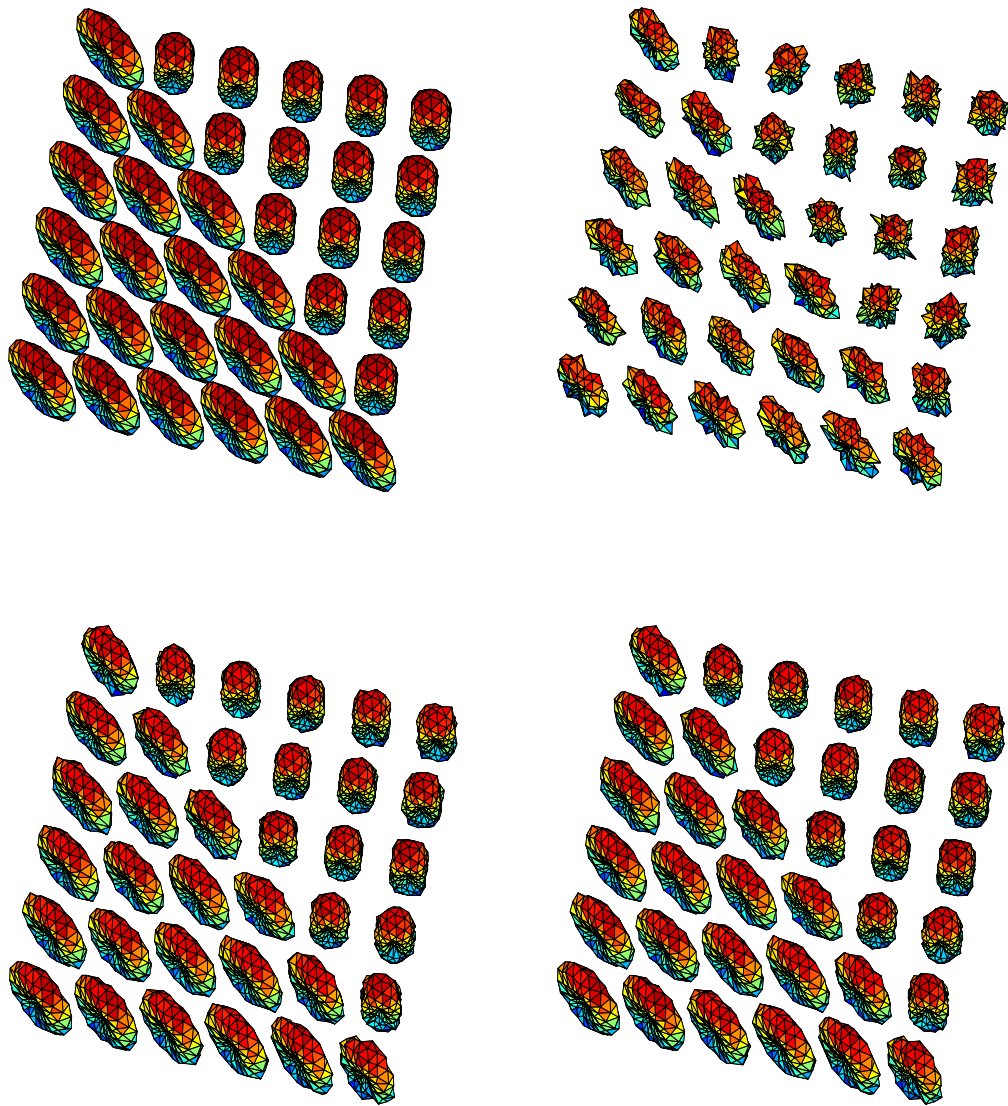
Figure 6: Comparison between models F and G. Top row, left to right: noise-free synthetic raw HARDI data, noisy synthetic raw HARDI data. Bottom row, left to right: denoised synthetic raw HARDI data using model F, denoised synthetic raw HARDI data using model G (different view).
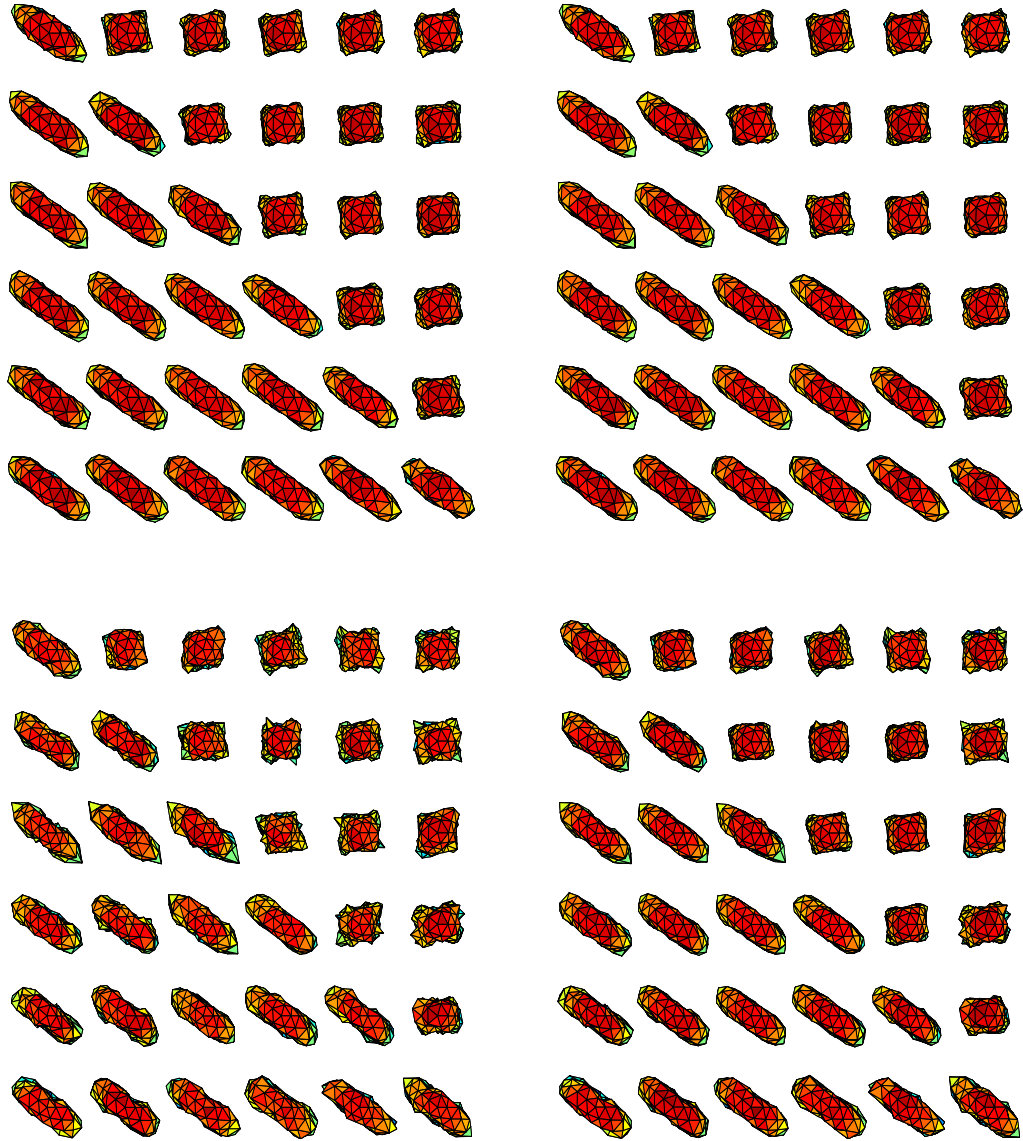
Figure 7: Comparison between models F, G, FEM [32] and FEM+TV [32] on the synthetic data. Top row, left to right: denoised raw HARDI data using F, denoised raw HARDI data using G. Bottom row, left to right: denoised raw HARDI data using only FEM in [32], denoised raw HARDI data using FEM+TV in [32].
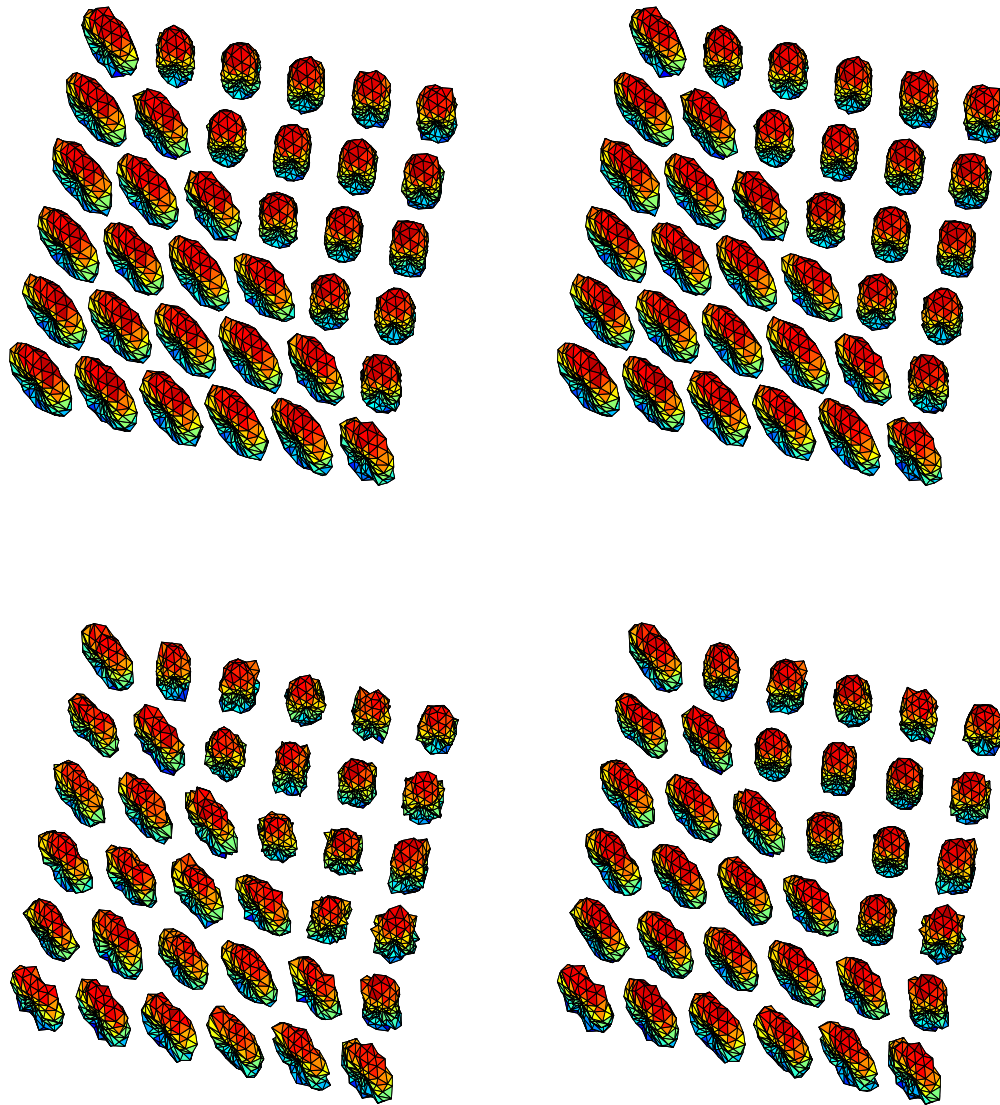
Figure 8: Comparison between models F, G, FEM [32] and FEM+TV [32] on the synthetic data. Top row, left to right: denoised raw HARDI data using F, denoised raw HARDI data using G. Bottom row, left to right: denoised raw HARDI data using only FEM in [32], denoised raw HARDI data using FEM+TV in [32] (different view).

noise-free data, the noisy data, and the denoised data are functions on the discrete domain $\Omega \times S_d^2$. If $f_1 : \Omega \times S_d^2 \to \mathbb{R}$ and $f_2 : \Omega \times S_d^2 \to \mathbb{R}$ are two images, then

$$RMSE(f_1, f_2) = \sqrt{\frac{1}{N \cdot n} \sum_{i=1}^{N} \sum_{j=1}^{n} (f_1(x_i, s_j) - f_2(x_i, s_j))^2}. \tag{19}$$

	RMSE
Noisy data	17.7079
Our denoised data using F	5.8991
Our denoised data using G	7.6081
Denoised data FEM in [32]	11.9964
Denoised data FEM+TV in [32]	7.6367

Table 1: RMSE between the mentioned synthetic data and the noise-free synthetic data.

Additionally, to assess the accuracy of our results in the synthetic experiments, and since we obtain the ODFs by postprocessing, we can compute the sKL distance between two probability density functions, which is defined by (for two probability densities $p(x), q(x)$),

$$sKL(p, q) = \frac{1}{2} \int_\Omega \left\{ p(x) \log \left(\frac{p(x)}{q(x)} \right) + q(x) \log \left(\frac{q(x)}{p(x)} \right) \right\} dx. \tag{20}$$

Let q be the ODF of the noise free data and in each case, let p be the ODF of either the noisy data or the denoised data. As it can be seen from (20), smaller sKL value implies more similarity between the two probability density functions. We obtain and show the sKL values in Table 2. The visualization of ODFs and the calculation of sKL distances were kindly done by Liang Zhan, Laboratory Of Neuro Imaging. Table 2 needs more explanation. Note that the two noisy data which appear in Table 2 are the same, which means that their sKL distances have to be theoretically the same. However, due to the differences in discretizations, we could not recover the same sKL distance value 1.0409 of the noisy data presented in [32]. Hence, we also look at the ratio between the sKL distance of denoised data and that of noisy data for our results and for the results in [32]. Table 3 presents those ratios. Smaller ratios mean better denoised results.

	sKL distance
Noisy data	4.2206
Our denoised data using F	1.8243
Our denoised data using G	1.8143
Noisy data	1.0409
Denoised data by FEM in [32]	0.9088
Denoised data by FEM+TV in [32]	0.6576

Table 2: sKL distances for the synthetic ODF data.

	Ratio between the sKL distances
Our denoised data using F / Noisy data	0.4322
Our denoised data using G / Noisy data	0.4299
Denoised data by FEM in [32] / Noisy data	0.8730
Denoised data by FEM in [32] / Noisy data	0.6318

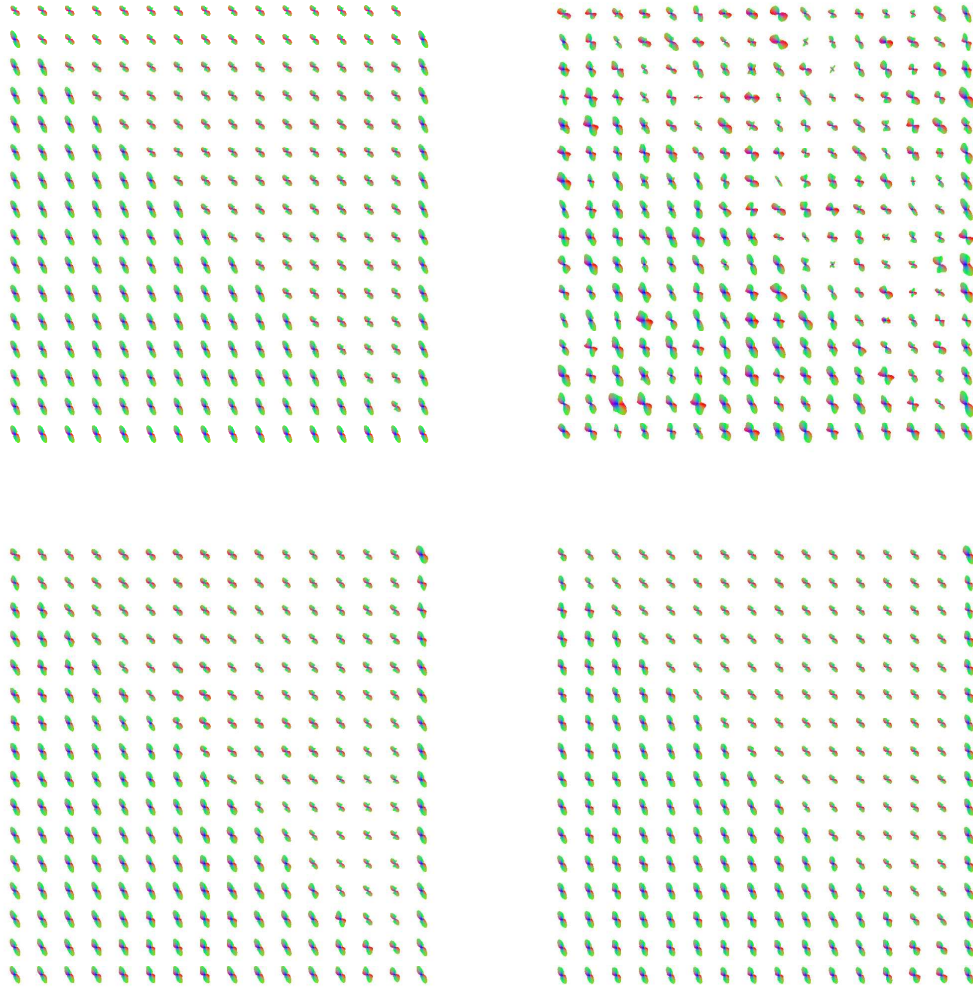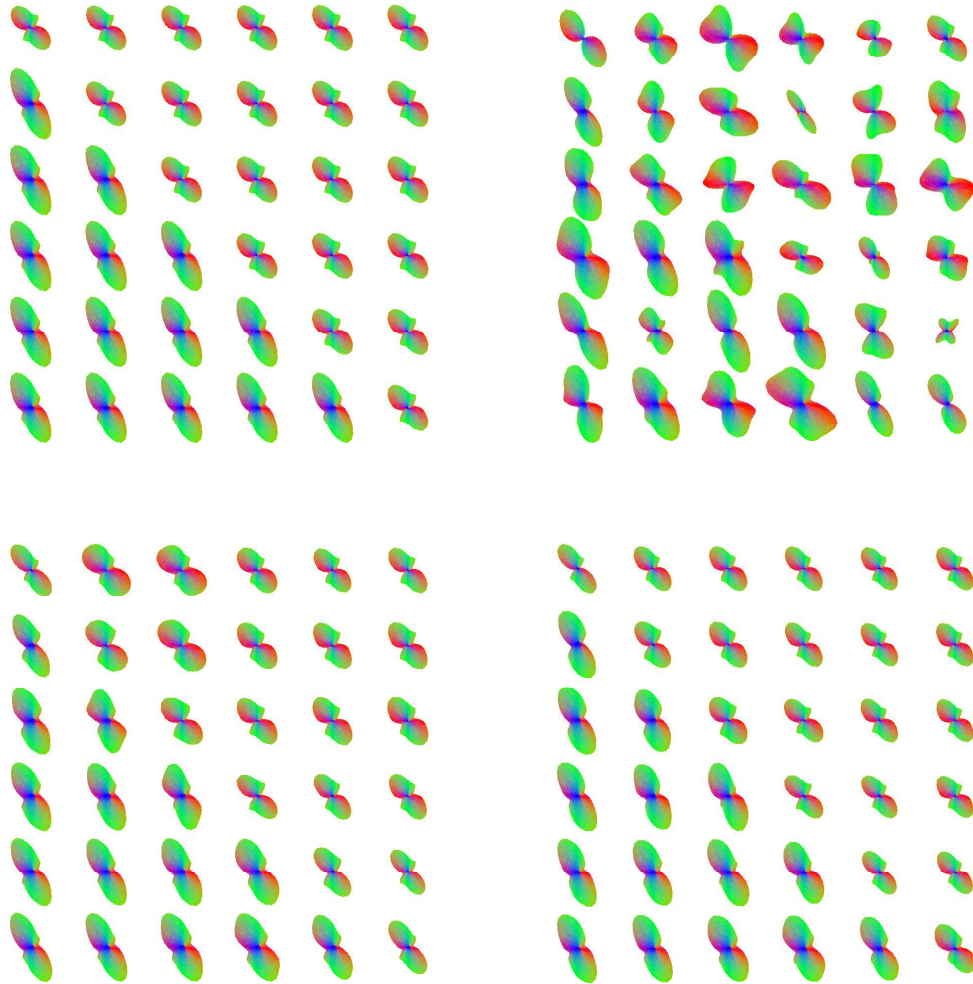Table 3: Ratios between the sKL distances of ODFs, for the synthetic data.

Figure 9: Top row, left to right: ODFs of noise-free synthetic HARDI data, ODFs of the noisy version. Bottom row, left to right: ODFs of denoised HARDI data using the functional F, ODFs of denoised HARDI data using the functional G.

Figure 10: Magnified views of Figure 9. Top row, left to right: ODFs of noise-free synthetic HARDI data, ODFs of the noisy version. Bottom row, left to right: ODFs of denoised HARDI data using the functional F, ODFs of denoised HARDI data using the functional G.

Next we apply the proposed models to the denoising of two real HARDI brain data sets (one of 30 directions, artifficially noisy by Rician noise, for which we have the original clean ground truth; another one of 94 directions, for which we do not have the ground truth).

Since in general the real data sets do not provide us with the true noise-free data, we need to obtain a noise-free "real" data set to make sure that our denoised methods actually improve the data quality. For this purpose, we fit sixth order spherical harmonics to the real HARDI data sets with 95 gradient directions. Then we sample a new set of 30 unit vectors uniformly distributed on a unit sphere at which we evaluate the spherical harmonics expansions. We consider these data sets as noise-free real HARDI data sets with 30 diffusion-sensitized gradient directions. This procedure was done by J. Eugenio Iglesias from Medical Imaging Informatics, UCLA. We still use the same 11 baseline images for f_0. It is now easy to corrupt these data sets with Rician noise, which is described below. Figures 11 - 17 show this noise-free data and their noisy and denoised versions.
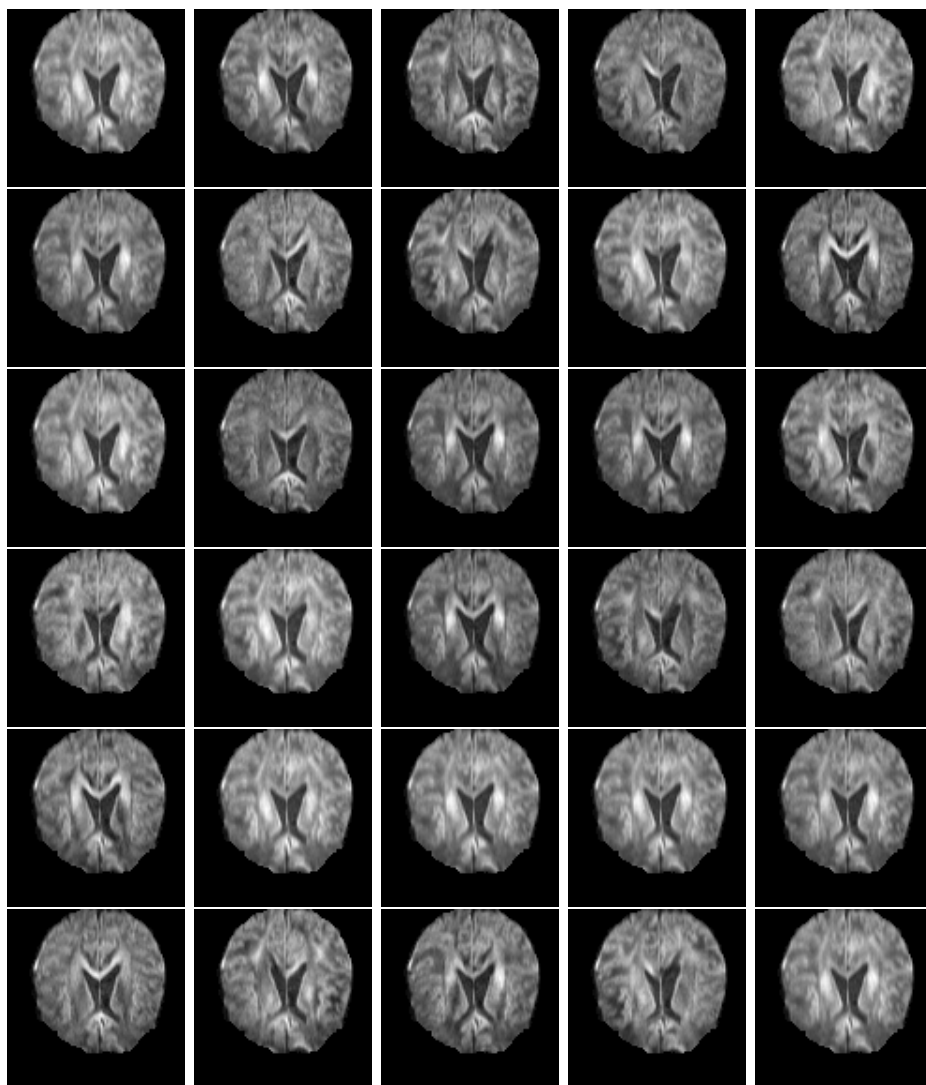
Figure 11: Noise-free real HARDI data with 30 diffusion-sensitized gradient directions.
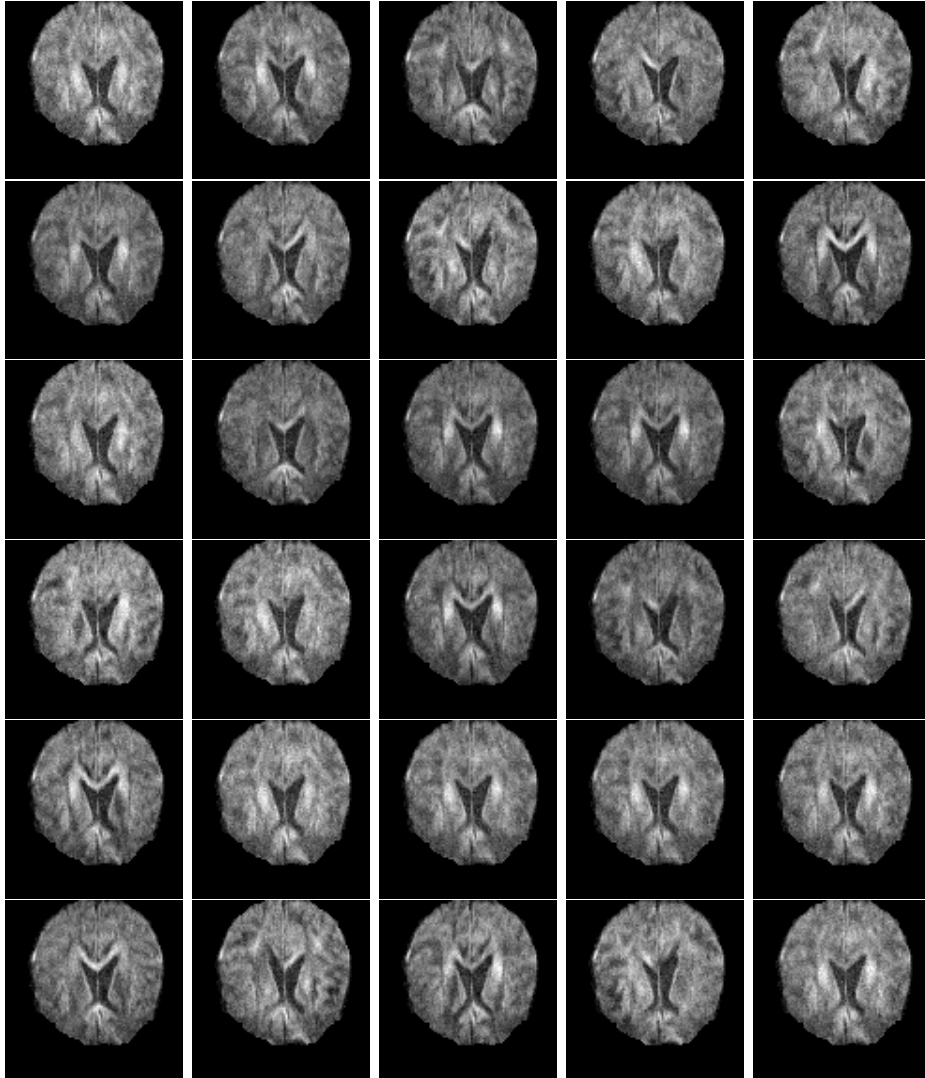
Figure 12: Noisy real HARDI data with 30 diffusion-sensitized gradient directions, artifficially corrupted by Rician noise.
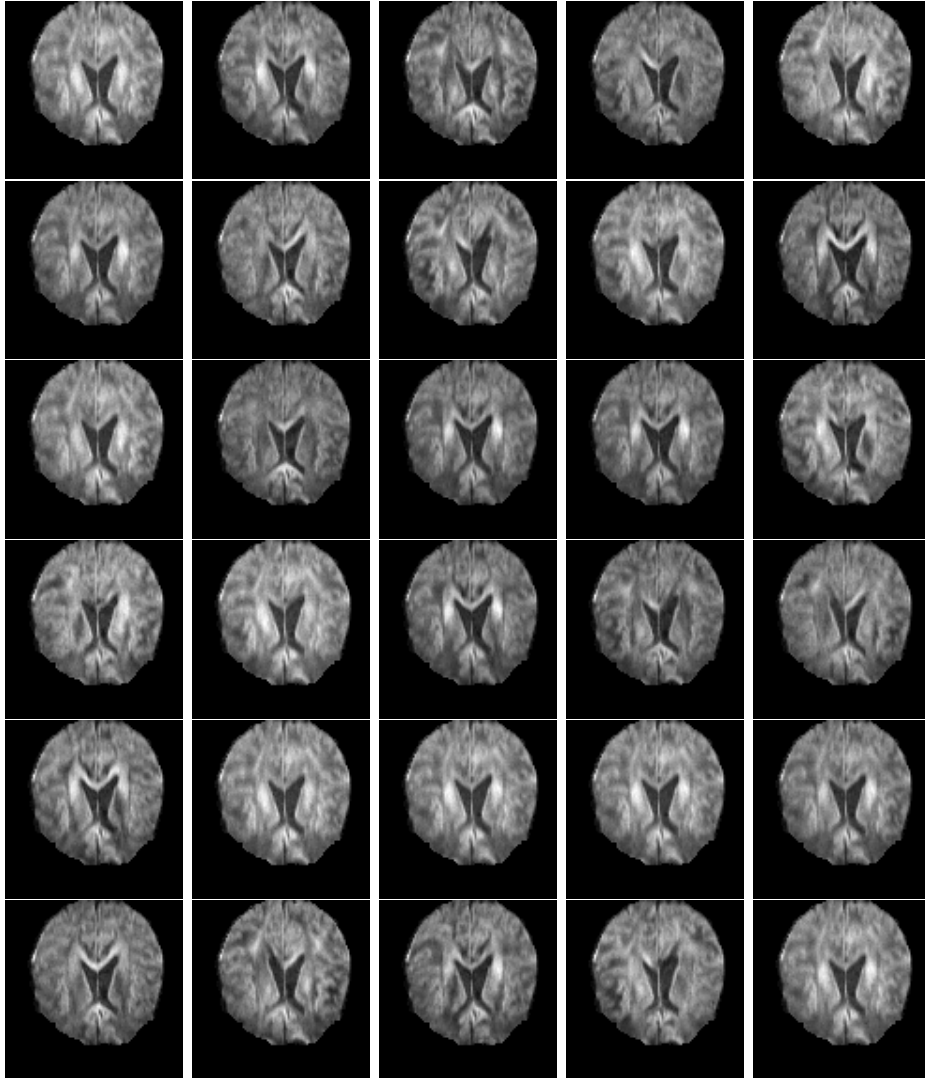
Figure 13: Denoised real HARDI data with 30 diffusion-sensitized gradient directions by the second model using G.

Let $\mathcal{N}(\nu, \sigma^2)$ be noise of Gaussian distribution with mean ν and variance σ^2. Then a random variable R has a Rician distribution $R(\nu, \sigma)$ if for some θ,

$$R = \sqrt{N_1^2 + N_2^2}$$

where $N_1 \sim \mathcal{N}(\nu \cos \theta, \sigma^2)$ and $N_2 \sim \mathcal{N}(\nu \sin \theta, \sigma^2)$, and N_1, N_2 are independent. The probability density function of $R(\nu, \sigma)$ is

$$f(x|\nu, \sigma) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2 + \nu^2}{2\sigma^2}\right) I_0\left(\frac{x\nu}{\sigma}\right)$$

which is independent of the θ value. We choose one θ value in $[0, 2\pi]$ at each position $x \in \Omega$ and generate two functions N_1 and N_2 having normal distributions $\mathcal{N}(\tilde{f}(x) \cos \theta, \sigma^2), \mathcal{N}(\tilde{f}(x) \sin \theta, \sigma^2)$ respectively, where $\tilde{f}$ is the noise-free data set obtained earlier by the spherical harmonics technique. Or simply we choose $\theta = 0$ at all the positions. The noisy data set in Figure 12 was generated with $\theta = 0$, $\sigma = 15$.

Since skull-stripping was done before to remove non-brain regions from the image, the background was automatically set to zero (black). Thus, we restrict the calculations only inside the brain region, combined with homogeneous Dirichlet boundary conditions along the xy plane. We used only 3 slices from the whole volume to reduce the computational cost and we imposed Neumann boundary conditions along the z-axis, and we visualize results from the middle slice. Figures 11, 12, 13 show all 30 direction images of original noise-free data, noisy data, and denoised data using the second model G.

Since we also have the noise-free data for the HARDI brain data experiment, it is possible to compute the RMSEs of the noisy data and the denoised data. Table 4 shows those RMSE values. We observed that the first model using F worked quite well with the synthetic data provided by the authors of [32] that has a rather simple structure; however, model 1 was not suitable for the real noisy HARDI data with 30 directions, and thus we only show results with the second model using G for the real noisy HARDI data (with 30 directions).

	Noisy data	Denoised data using G
RMSE	10.5448	4.7268

Table 4: RMSEs: the noisy real data with 30 diffusion-sensitized gradient directions and the denoised version using model G.

Finally, we present a denoised result of a noisy real HARDI data with 94 diffusion-sensitized gradient directions. We tested our models on real HARDI datasets with 94 diffusion-sensitized gradient directions, which means that the number n is 94. Briefly, we used the same data as in [31] in which 3D structural brain MRI scans and DT-MRI scans were acquired from healthy young adults on a 4 Tesla Bruker Medspec MRI scanner using an optimized diffusion tensor sequence. Imaging parameters were: TE/TR 92.3/8250 ms, 55 x 2mm contiguous slices, FOV = 23 cm. 105 directional gradients were applied: 11 baseline (S_0) images with no diffusion sensitization (i.e., T_2-weighted images) and 94 diffusion-weighted images (b-value 1159 s/mm^2 in-plane resolution) in which gradient directions were evenly distributed on the hemisphere [24]. The reconstruction matrix was 128 x 128, yielding a 1.8x1.8 mm^2 in-plane resolution. The total scan time was 14.5
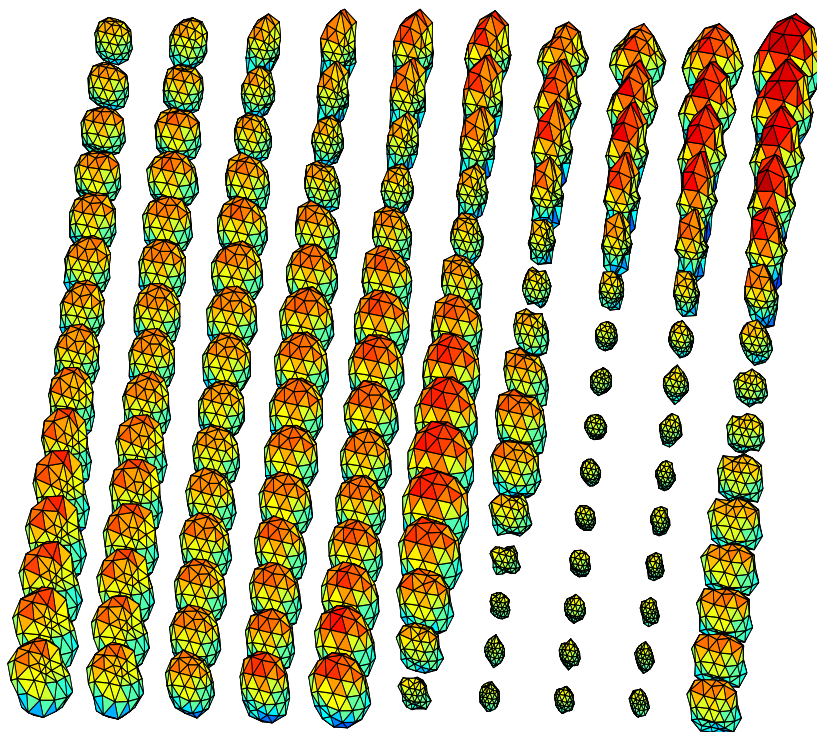
Figure 14: Small part of the noise-free real HARDI data with 30 diffusion-sensitized gradient directions.
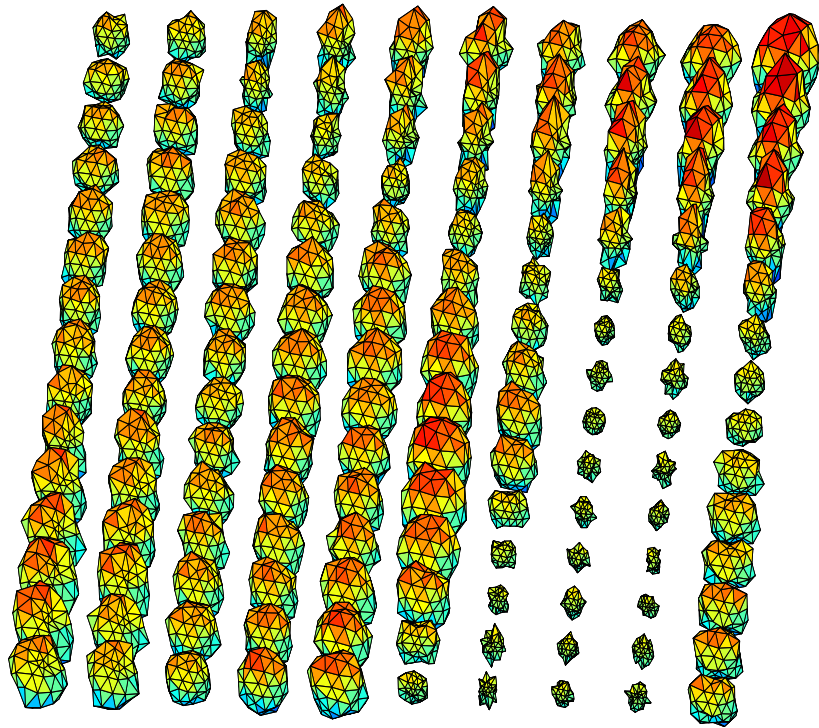
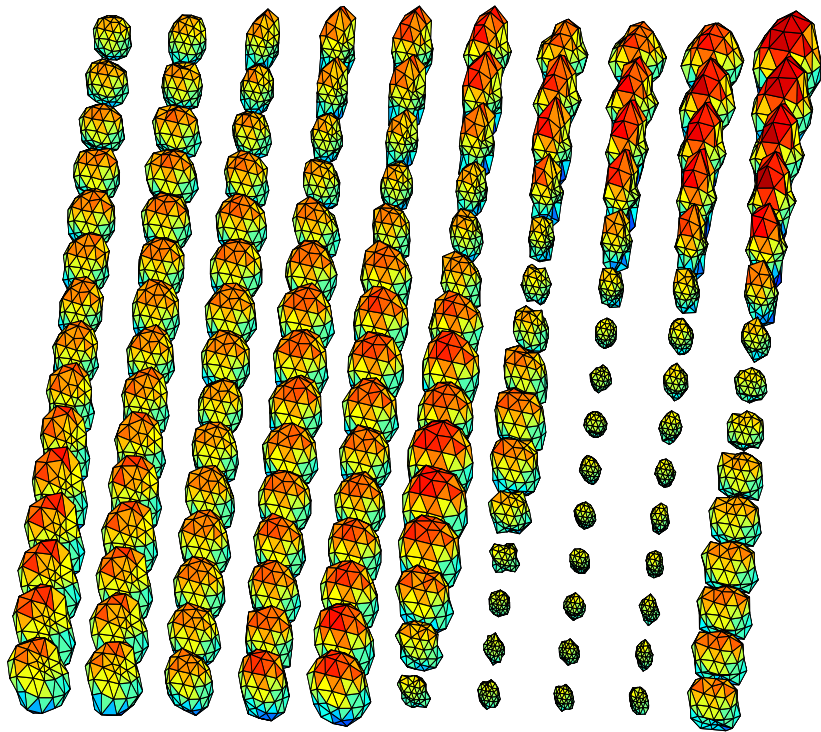Figure 15: Small part of the noisy real HARDI data with 30 diffusion-sensitized gradient directions.

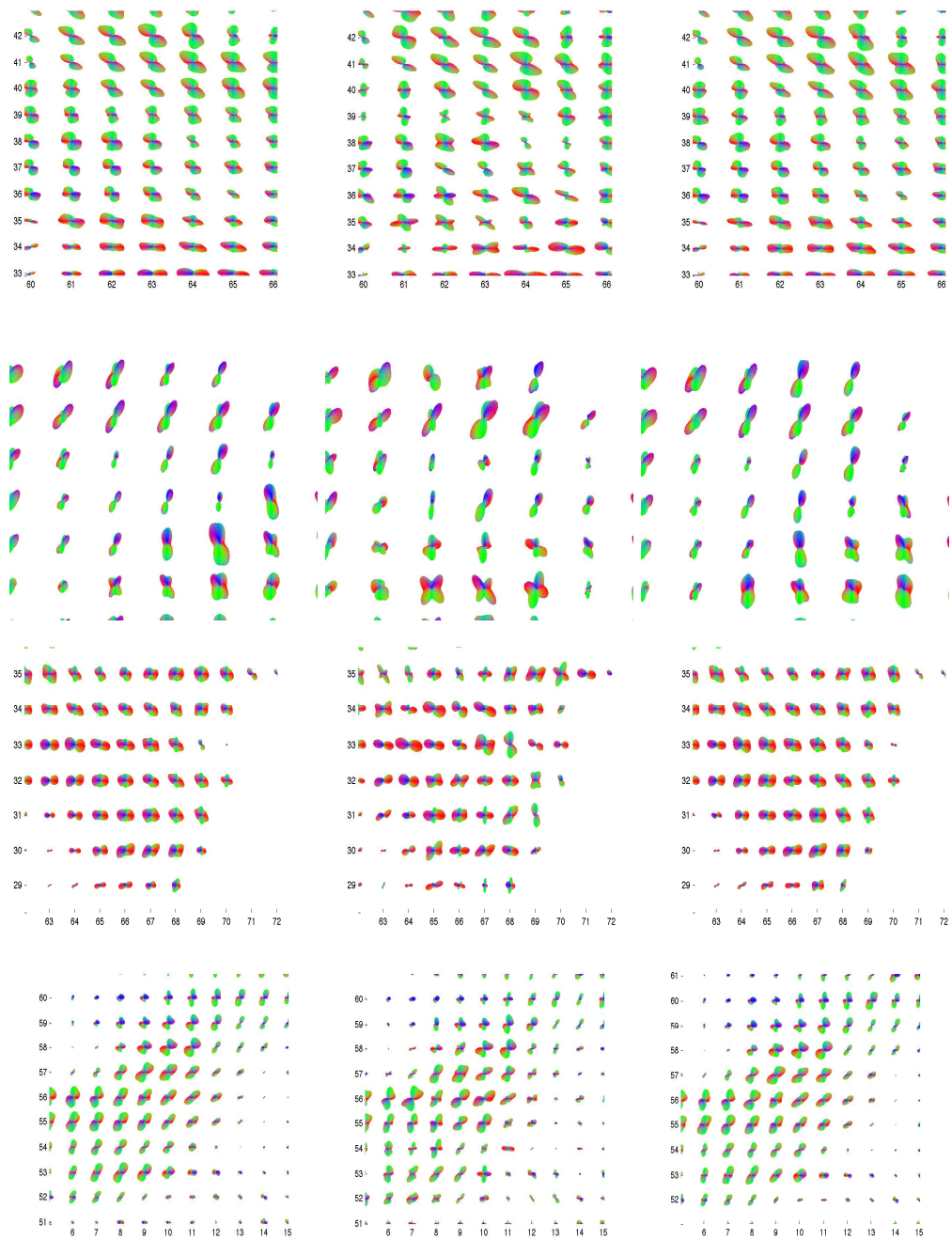Figure 16: Small part of the denoised real HARDI data with 30 diffusion-sensitized gradient directions using model G.

Figure 17: ODF visualization. Left to right : noise-free real HARDI data, noisy real HARDI data, denoised real HARDI data using model G, with 30 diffusion-sensitized gradient directions.

minutes. We set $f_0 = S_0$ to be the average of the 11 baseline images. Figure 19 shows the slices of one of the actual datasets used in our numerical calculations, together with the denoised results obtained using model G, and Figure 18 shows the corresponding non-diffusion weighted image $S_0 = f_0$.
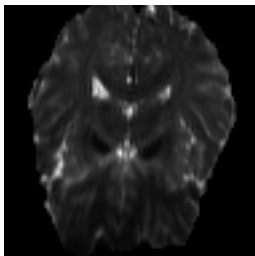
Figure 18: 19th slice of $f_0 = S_0$.

Next we visualize ODFs of the data and compare the ODFs of the noisy data with the ODFs of the denoised data. Notice that calculating ODF of a noisy dataset means that we perform a process of smoothing the data. Since the original noisy data usually violates the constraints we had in the models, if we want to visualize the ODF of the noisy data, then there has to be a pre-processing step to adjust those violating values, which may not involve solid consistency to the fundamental model. To see the difference between the ODFs of the noisy data and the denoised data, we consider some parts of the whole brain image and magnify them especially in regions where fibers are crossing. The data and denoised results obtained using models F and G are shown in Figures 19-21.
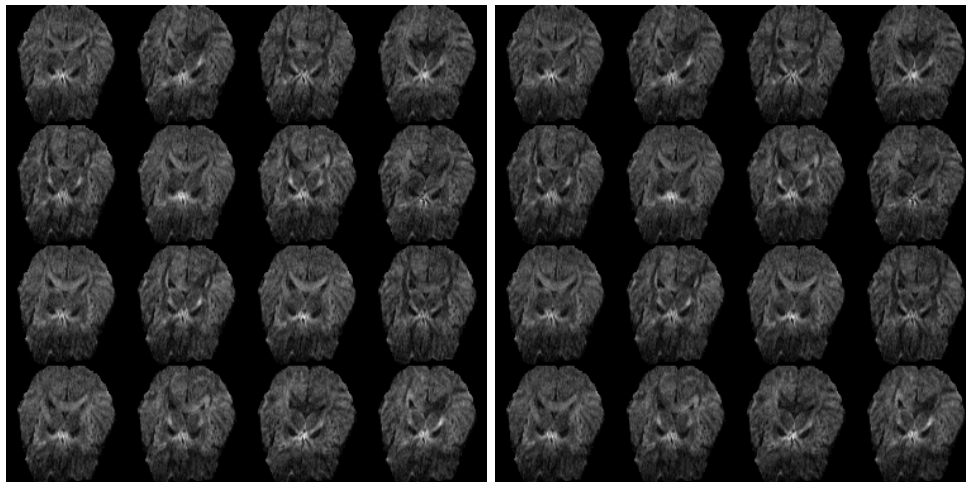

Figure 19: 19th slice of the real brain data with 94 gradient directions. Only 16 randomly picked directions out of 94 directions corresponding to the 19th slice of the noisy and the denoised (using G) HARDI data are shown here.
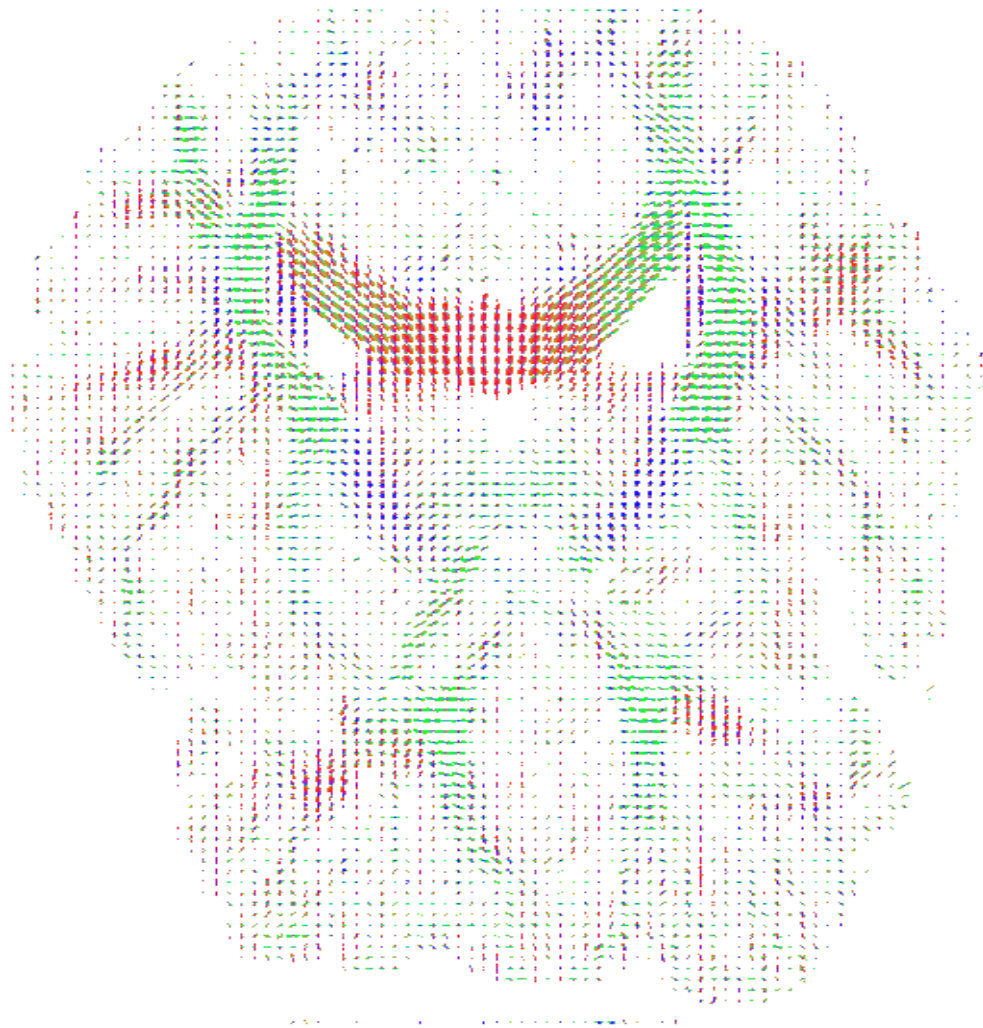
Figure 20: ODF: 19th slice of the real brain data with 94 gradient directions.
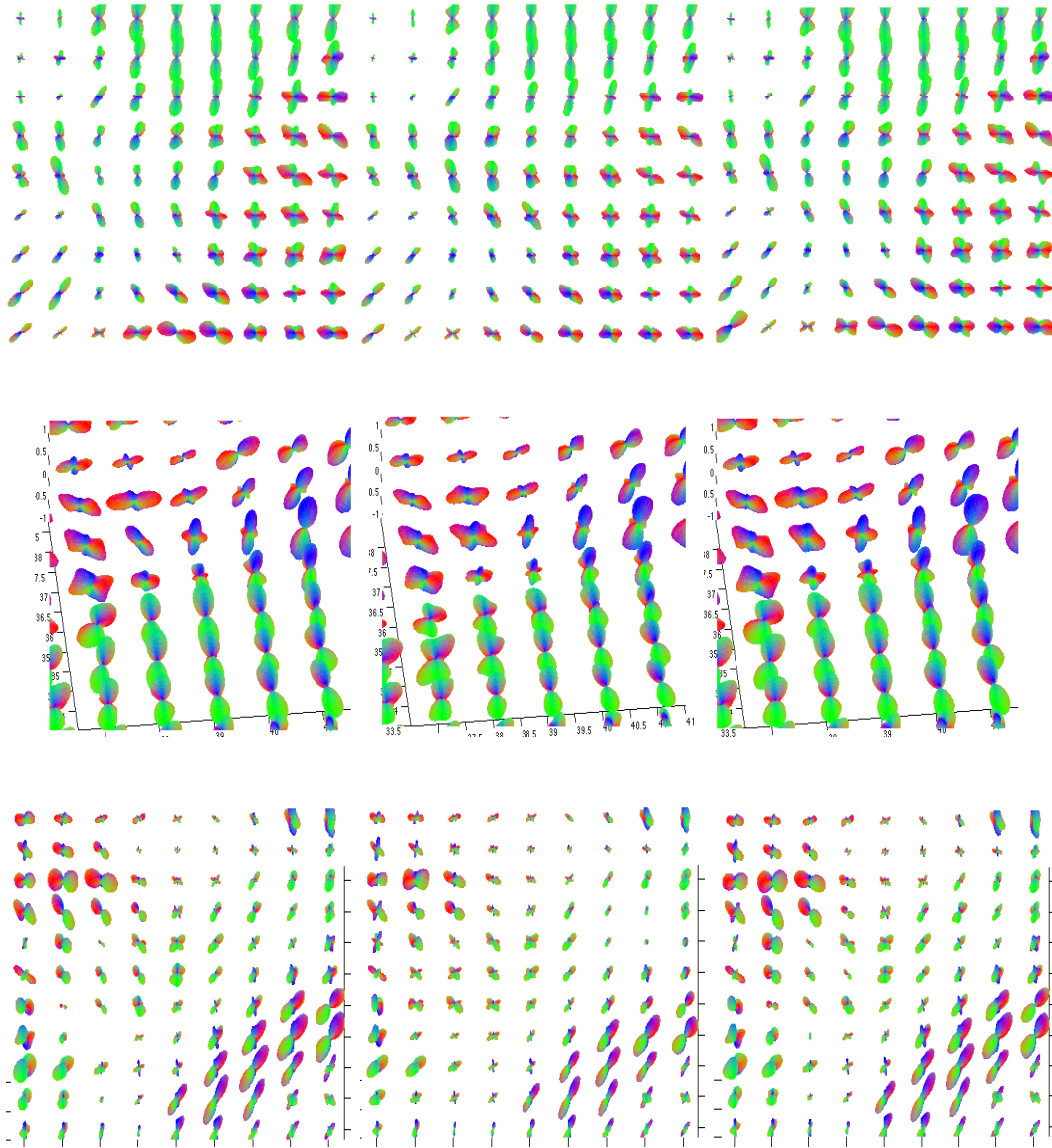
Figure 21: ODF visualization for real brain data with 94 gradient directions. Left to right: ODFs of noisy data, denoised result using the first model F, denoised result using the second model G.

5 Conclusion

We proposed two slightly different models to denoise HARDI data. The difficulty in dealing with the data lies in the nature of its complex structure. One model denoises the signal itself by the standard $TV + L^1$ method combined with the logarithmic barrier function to realize the constraints that arise naturally from the acquisition model. The other model uses the acquisition formula

$$S = S_0 \exp(-d) + noise$$

and we denoise d instead of denoising S. Similarly, we denoise d by $TV + L^1$ method combined with the logarithmic barrier function, adapting the data fidelity term from the image formation model. As we have seen through the numerical computations, our model for denoising d is more stable and gives more satisfactory results for real HARDI data. We also compared our denoising methods with the methods described in [32]. Our proposed models outperformed the models in [32] providing smaller RMSE values, and also our denoised data is visually more similar to the noise-free data.

Acknowledgments

References

[1] S. Alliney, *Digital Filters as Absolute Norm Regularizers*, IEEE TSP, 40(6):1548-1562, 1992.

[2] S. Alliney, *Recursive Median Filters of Increasing Order: A Variational Approach*, IEEE TSP, 44(6):1346-1354, 1996.

[3] S. Alliney, *A Property of the Minimum Vectors of a Regularizing Functional Defined by Means of the Absolute Norm*, IEEE TSP, 45(4):913-917, 1997.

[4] A. Barmpoutis, B. Jian, B.C. Vemuri, *Symmetric positive 4th order tensors and their estimation from diffusion weighted MRI*, Proceedings of IPMI07: Information Processing in Medical Imaging, 2-6 July 2007, 308-319.

[5] P.J. Basser, C. Pierpaoli, *Microstructural and physiological features of tissues elucidated by quantitative diffusion tensor MRI*, Journal of Magnetic Resonance, 111:209-219, 1996.

[6] C. Beaulieu, P.S. Allen, *Water diffusion in the giant axon of the squid: Implications for diffusion-weighted MRI of the nervous system*, Magn Reson Med., 32(5):579-83, 1994.

[7] P. Blomgren, *Total Variation Methods for Restoration of Vector Valued Images*, (Ph.D. thesis), UCLA CAM Report 98-30, 1998.

[8] P. Blomgren, T.F. Chan, *Color TV: Total variation methods for restoration of vector-valued images*, IEEE TIP, 7(3):304-309, 1998.

[9] T.F. Chan, S. Esedoglu, *Aspects of total variation regularized L^1 function approximation*, SIAM J. on Applied Maths., 65(5):1817-1837, 2005.

[10] E. Cheon, A. Paranjpye, S. Osher, L. Vese, *Image denoising with total variation minimization and different L^p norms for the fidelity term*, UCLA REU Project, 2002.

[11] M.C. Chiang, A.D. Klunder, K. McMahon, G.I. de Zubicaray, M.J. Wright, A.W. Toga, P.M. Thompson, *Information-theoretic analysis of brain white matter fiber orientation distribution functions*, Information Processing in Medical Imaging 2007; LNCS 4584, 172-182, 2007

[12] M.C. Chiang, A.D. Leow, R.A. Dutton, M. Barysheva, S. Rose, K.L. McMahon, G.I. de Zubicaray, A.W. Toga, P.M. Thompson, *Fluid Registration of Diffusion Tensor Images Using Information Theory*, IEEE Transactions on Medical Imaging, 27(4):442-56, 2008.

[13] M. Cihangiroglu, A. Müfit Uluǧ, Z. Firat, A. Bayram, A. Kovanlikaya, İ. Kovanlikaya, *High b-value diffusion-weighted MR imaging of normal brain at 3T*, European Journal of Radiology, 69(3):454-458, 2009.

[14] S. Delputte, H. Dierckx, E. Fieremans, Y. D'Asseler, R. Achten, I. Lemahieu, *Postprocessing of brain white matter fiber orientation distribution functions*, Proceedings IEEE International Symposium on Biomedical Imaging: from Nano to Macro, 784-787, 2007.

[15] M. Descoteaux, R. Deriche, *High Angular Resolution Diffusion MRI Segmentation Using Region-Based Statistical Surface Evolution*, Journal of Mathetical Imaging and Vision, 33(2):239-252, 2009.

[16] M. Descoteaux, E. Angelino, S. Fitzgibbons, R. Deriche, *Regularized, Fast and Robust Analytical Q-Ball Imaging*, Magnetic Resonance in Medicine, 58(3):497-510, 2007

[17] M. Descoteaux, R. Deriche, *Mapping neuronal fiber crossings in the human brain*, SPIE Newsroom , August 2008.

[18] A. V. Fiacco, G.P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York, 1968. Reprinted by SIAM Publications, 1990.

[19] L.R. Frank, *Characterization of anisotropy in high angular resolution diffusion-weighted MRI*, Magn. Reson. Med. 47(6):1083-1099, 2002.

[20] G. Haro, C. Lenglet, G. Sapiro, P.M. Thompson, *On the Non-Uniform Complexity of Brain Connectivity*, ISBI 2008, 887-890.

[21] B. Jian, B.C. Vemuri: A Unified Computational Framework for Deconvolution to Reconstruct Multiple Fibers From Diffusion Weighted MRI. IEEE TMI 26(11):1464-1471, 2007.

[22] B. Jian, B.C. Vemuri, E. Özarslan, P.R. Carney, T.H. Mareci: A novel tensor distribution model for the diffusion-weighted MR signal. NeuroImage 37(1):164-176, 2007.

[23] L. Jonasson, P. Hagmann, X. Bresson, J.-P. Thiran, V.J. Wedeen, *Representing Diffusion MRI in 5D for Segmentation of White Matter Tracts with a Level Set Method*, LNCS 3565, 311-320, 2005.

[24] D.K. Jones, M.A. Horsfield, A. Simmons, *Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging*, Magn Reson Med 42:515-525, 1999.

[25] P. Khurd, R. Verma, C. Davatzikos, *On characterizing and analyzing diffusion tensor images by learning their underlying manifold structure*, Presented at IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA), New York City, NY, 2006.

[26] Y. Kim, P.M. Thompson, A.W. Toga, L. Vese, and L. Zhang, *HARDI denoising: variational regularization of the spherical apparent diffusion coefficient sADC*, LNCS 5636, 515-527, 2009.

[27] D. Le Bihan, E. Breton, *Imagerie de diffusion in vivo par resonance magnétique nucléaire*, CR Académie des Sciences, 301:1109-1112, 1985.

[28] D. Le Bihan, E. Breton, D. Lallemand, P. Grenier P, E. Cabanis, M. Laval-Jeantet, *MR imaging of intravoxel incoherent motions: Application to diffusion and perfusion in neurologic disorders*, Radiology, 401-407, 1986.

[29] D. Le Bihan, C. Poupon, A. Amadon, F. Lethimonnier, *Artifacts and Pitfalls in Diffusion MRI*, J. of Magnetic Resonance Imaging, 24:478-488, 2006.

[30] 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, P.M. Thompson, *Mathematical Methods for Diffusion MRI Processing*, NeuroImage, Special Issue on Mathematics in Brain Imaging. 45(1, Supplement 1):S111-S122, 2009.

[31] A.D. Leow, S. Zhu, L. Zhan, K. McMahon, G.I. de Zubicaray, M. Meredith, M.J. Wright, A.W. Toga, P.M. Thompson, *The Tensor Distribution Function*, Magnetic Resonance in Medicine, 61(1):205-214, 2008.

[32] T. McGraw, B. Vemuri, E. Özarslan, Y. Chen, and T. Mareci, *Variational Denoising of Diffusion Weighted MRI*, Inverse Problems and Imaging, 3(4):625-648, 2009.

[33] T. McGraw, B. Vemuri, R. Yezierski, T. Mareci, *Segmentation of High Angular Resolution Diffusion MRI Modeled as a Field of von Mises-Fisher Mixtures*, LNCS 3953, 463-475, 2006.

[34] T. McGraw, B.C. Vemuri, B. Yezierski, T. Mareci, V. Mises-Fisher, *Mixture model of the diffusion ODF*, Proc. IEEE Intl. Symposium on Biomedical Imaging: From Nano to Macro, 65-68, Arlington, VA, April 6-9, 2006.

[35] M. Merboldt, W. Hanicke, J. Frahm, *Self-diffusion NMR imaging using stimulated echoes*, J. Magn. Reson., 64:479-486, 1985.

[36] M. Nikolova, *A variational approach to remove outliers and impulse noise*, Journal of Mathematical Imaging and Vision 20(1-2): 99-120, 2004.

[37] J. Nocedal, S.J. Wright, *Numerical Optimization*, Springer Series in Operations Research, Springer 1999.

[38] E. Özarslan, B. C. Vemuri, T.H. Mareci, *Generalized scalar measures for diffusion MRI using trace, variance and entropy*, Magn Reson Med, 53(4):866-876, 2005.

[39] L.I. Rudin, S. Osher, E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D 60(1-4):259-268, 1992.

[40] L. Rudin, V. Caselles, *Image Recovery via Multiscale Total Variation*, Proceedings of the Second European Conference on Image Processing, 1995.

[41] A. Srivastava, I. Jermyn, S. Joshi, *Riemannian Analysis of Probability Density Functions with Application in Vision*, In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR2007), 17-22 June 2007, 1-8.

[42] E.O. Stejskal, J.E. Tanner, *Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient*, The Journal of Chemical Physics, 42:288-292, 1965.

[43] D. Tschumperlé, *PDE-Based Regularization of Multivalued Images and Applications*, PhD Thesis, University of Nice-Sophia Antipolis, France, 2002.

[44] D. Tschumperlé, R. Deriche, *Anisotropic Diffusion Partial Differential Equations in Multi-Channel Image Processing: Framework and Applications*, book chapter in Advances in Imaging and Electron Physics (AIEP), Academic Press, 2007.

[45] D. Tschumperlé, R. Deriche, *Vector-Valued Image Regularization with PDE's: A Common Framework for Different Applications*, IEEE TPAMI, 27(4):506-517, 2005.

[46] D.S. Tuch, R.M. Weisskoff, J.W. Belliveau, V.J. Wedeen, *High angular resolution diffusion imaging of the human brain*, Proc. 7th Annual Meeting of ISMRM, Philadelphia, PA, 1999, p. 321.

[47] D.S. Tuch, T.G. Reese, M.R. Wiegell, N. Makris, J.W. Belliveau, V.J. Wedeen, *High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity*, Magn Reson Med, 48:577-582, 2002.

[48] D.S. Tuch, T.G. Reese, M.R. Wiegell, V.J. Wedeen, *Diffusion MRI of complex neural architecture*, Neuron, 40:885-895, 2003.

[49] D.S. Tuch, *Q-ball imaging*, Magn Reson Med, 52:1358-72, 2004.

[50] S.J. Wright, D. Orban, *Properties of the log-barrier function on degenerate nonlinear programs*, Mathematics of Operations Research, 27:585-613, 2002.

[51] L. Zhan, M.C. Chiang, M. Barysheva, A.W. Toga, K.L. McMahon, G.I. de Zubicaray, M. Meredith, M.J. Wright, P.M. Thompson, *How Many Gradients are Sufficient in High-Angular Resolution Diffusion Imaging (HARDI)?* Workshop on Computational Diffusion MRI, 216-224, MICCAI 2008.