Modeling diffusion-weighted MRI as a spatially variant Gaussian mixture: Application to image denoising

Juan Eugenio Iglesias Gonzalez[a] and Paul M. Thompson
Laboratory of Neuro Imaging, University of California, 635 Charles Young Drive South, Suite 225, Los Angeles, California 90095

Aishan Zhao
Peking University, 5 Yiheyuan Road, Haidian District, Beijing, 100871 China

Zhuowen Tu
Laboratory of Neuro Imaging, University of California, 635 Charles Young Drive South, Suite 225, Los Angeles, California 90095

(Received 7 October 2010; revised 12 May 2011; accepted for publication 20 May 2011; published 30 June 2011)

Purpose: This work describes a spatially variant mixture model constrained by a Markov random field to model high angular resolution diffusion imaging (HARDI) data. Mixture models suit HARDI well because the attenuation by diffusion is inherently a mixture. The goal is to create a general model that can be used in different applications. This study focuses on image denoising and segmentation (primarily the former).

Methods: HARDI signal attenuation data are used to train a Gaussian mixture model in which the mean vectors and covariance matrices are assumed to be independent of spatial locations, whereas the mixture weights are allowed to vary at different lattice positions. Spatial smoothness of the data is ensured by imposing a Markov random field prior on the mixture weights. The model is trained in an unsupervised fashion using the expectation maximization algorithm. The number of mixture components is determined using the minimum message length criterion from information theory. Once the model has been trained, it can be fitted to a noisy diffusion MRI volume by maximizing the posterior probability of the underlying noiseless data in a Bayesian framework, recovering a denoised version of the image. Moreover, the fitted probability maps of the mixture components can be used as features for posterior image segmentation.

Results: The model-based denoising algorithm proposed here was compared on real data with three other approaches that are commonly used in the literature: Gaussian filtering, anisotropic diffusion, and Rician-adapted nonlocal means. The comparison shows that, at low signal-to-noise ratio, when these methods falter, our algorithm considerably outperforms them. When tractography is performed on the model-fitted data rather than on the noisy measurements, the quality of the output improves substantially. Finally, ventricle and caudate nucleus segmentation experiments also show the potential usefulness of the mixture probability maps for classification tasks.

Conclusions: The presented spatially variant mixture model for diffusion MRI provides excellent denoising results at low signal-to-noise ratios. This makes it possible to restore data acquired with a fast (i.e., noisy) pulse sequence to acceptable noise levels. This is the case in diffusion MRI, where a large number of diffusion-weighted volumes have to be acquired under clinical time constraints.

[DOI: 10.1118/1.3599724]

Key words: HARDI, mixture model, random Markov field, denoising, expectation maximization

I. INTRODUCTION

I.A. Background: diffusion MRI

Diffusion weighted magnetic resonance imaging (DW-MRI) is the only technique that can image axonal fiber tracts in the brain in a noninvasive manner. DW-MRI produces images of biological tissue weighted by the local characteristics of water diffusion. Comparing a baseline T2-MRI scan with the MRI signal when a diffusion sensitizing gradient is added to the magnetic field in a certain direction, water diffusivity in that direction can be estimated. Repeating the procedure with different gradient directions, it is possible to reconstruct a diffusion profile at each spatial location. Due to the fact that water mainly diffuses along rather than across axonal fibers, the diffusivity information can be used to track the neural tracks in the brain white matter.

The continuous diffusivity profile, defined on the unit sphere, can be reconstructed from discrete measurements in different ways. The simplest approach is diffusion tensor imaging (DTI),[1,2] in which a zero-mean Gaussian probability distribution function (PDF) is fitted to the data measurements. At least seven samples are required for each voxel: six for the unique elements of the covariance matrix and one for the baseline MRI signal. DTI has shown great utility in imaging the fiber bundles in the brain because the architecture of the tissue can be inferred from the eigenstructure of the tensor: the major eigenvector gives the mean fiber direction at each location, whereas the anisotropy is a marker of fiber density. However, DTI has the limitation that it can only resolve one fiber direction in a voxel. When fibers cross or bifurcate within the voxel, which happens very often because voxels are from 100 to 1000 times wider than fibers, the monomodal Gaussian distribution fails to capture the diffusion profile.[3,4]

In order to resolve more complex fiber geometries, it is necessary to sample a higher number of directions. This approach is known as high angular resolution diffusion imaging (HARDI).[4–6] Mathematical entities more complex than a Gaussian PDF can be fit to the measurements, overcoming the limitation that DTI uses a monomodal distribution. Among the most popular are spherical harmonics,[7,8] fourth-order tensors,[9,10] and mixture models.[3,11]

From the HARDI data, the PDF of the average spin displacement of water or a variant of this function can be computed. Several approaches have been proposed in the literature: the PDF of the fiber orientation distribution,[12] the diffusion orientation transform,[13] the persistent angular structure of the PDF,[14] and the orientation distribution function (ODF), which is the most popular. The ODF is the radial projection of the water diffusivity PDF, and is defined on a spherical shell. Much of the ongoing research in HARDI is focused on the computation of the ODF, because its maxima correspond to the most likely orientation of the underlying fibers, whereas the maxima of the raw attenuation data do not. In a first approach, Tuch[15] used a numerical implementation of the Funk–Radon transform to approximate the ODF from the HARDI measurements. More recent work from different groups[16–18] has converged to computing the Funk–Radon transform analytically from the spherical harmonic expansion of the HARDI data.

If the attenuation by diffusion of a scan (which depends of the pulse sequence and its parameter settings) is small, the directional information of the data can be enhanced by spherical deconvolution.[12] The ODF is assumed to be the convolution of a sharp fiber ODF (fODF) and a blurring kernel. Ideally, the fODF is zero for all directions except for those exactly corresponding to the orientation of fibers. The kernel can be estimated from highly anisotropic voxels, which are assumed to contain only one fiber population, and then inverted to compute the fODF from the ODF. It has been shown that using the fODF rather than smoother ODF in fiber tracking can improve the results.[19]

I.B. Background: denoising of diffusion MRI

Denoising strategies are very attractive in HARDI because acquiring a high number of gradient images under time constraints limits the signal to noise ratio (SNR). Denoising methods from the natural or medical image processing literature can be extended to HARDI by denoising each channel (gradient or baseline volumes) separately or by considering the HARDI data a vector-valued function. However, these extensions suffer from the same limitations as the methods they are based on.

Specifically, traditional image denoising algorithms based on (local or nonlocal) spatial smoothing[20–23] improve the SNR but also cause a loss of fine structures and increase partial volume error at the interfaces between different types of tissue. Techniques based on wavelets[24] preserve edges well but suffer from pseudo-Gibbs artifacts. This problem is solved by approaches based on variational calculus[25,26] and partial differential equations, but such methods still have the disadvantage that they cause loss of texture, leading to the well-known “staircase” effect. In general, all these methods perform satisfactorily when the noise level is moderate, but their limitations arise when the SNR decreases. Therefore, denoising of DW-MRI data remains an open, relevant problem, and algorithms that overcome the drawbacks of existing methods are necessary.

I.C. Contribution

In this study, we propose an approach to the DW-MRI denoising problem based on unsupervised learning. We take advantage of the relatively low variability in DW-MRI to learn the typical distribution of the data, which is fairly constant across subjects (except for cases with pathology). We then use this information to tailor a denoising method to DW-MRI and outperform the algorithms summarized in Sec. I B.

We propose using a spatially variant Gaussian mixture model (GMM) to learn a relatively small dictionary of components that can accurately represent the diffusion data. GMMs, which have been successfully applied in many different problems in the literature, are very suitable to represent the signal attenuation of HARDI data, which is a weighted sum of contributions from different fiber populations. The number of components, whose predefinition is a limitation of GMMs in general, is automatically estimated using the minimum message length, a concept from information theory.

The mean and covariance of each component, which are learned from training data, are constant. This allows us to use a large number of different voxels to train the components. If the mean and covariances were spatially variant, if would be impossible to estimate them for each single voxel unless strong assumptions were made (as in DTI, for example). Each mixture component will typically correspond to isotropic profiles or fibers in a certain orientation, with some flexibility provided by their covariances. The mixing weights are allowed to vary across space such that different components are represented in different regions. Then, a smoothness prior is placed on top on these spatially variant weights to exploit the fact that neighboring voxels tend to be similar. This constraint is very flexible because it does not force neighboring voxels to have similar values, but similar statistical distributions, decreasing the spatial blurring (which can lead to spurious fiber crossings).

The trained model can be fitted to a (noisy) test case by maximizing the posterior probability of the underlying noiseless data in a Bayesian framework. The output is a denoised version of the image. A by-product is the set of mixture weights for each voxel, which can be used in image segmentation.

To summarize, the main contributions of this study are: (1) unsupervised learning of diffusion MRI attenuation data with a spatially variant mixture model for DW-MRI-tailored denoising; (2) rather than modeling the DW-MRI data in individual locations (as it is typically done in the DW-MRI literature), we exploit spatial regularities by modeling the diffusion data as a field; and (3) application of a spatially variant GMM to denoising.

It is important to note that this article is an extension of our own preliminary work[35] with the following differences: (1) we model the signal attenuation instead of the anisotropic diffusion coefficient, which is more principled; (2) the training stage has been improved; and (3) new experiments have been added, including fiber tracking, segmentation, and extended denoising experiments.

I.D. Further related literature

As explained above, mixture models in general suit DW-MRI very well because the attenuation by diffusion in a fiber crossing or bifurcation can be modeled as the sum of two or more mixture components. Therefore, they have been extensively used in the DW-MRI literature to model the diffusion signal for a given voxel. For example, Tuch *et al.*[4] propose extending DTI to a mixture of ellipsoidal profiles. Assaf *et al.*[31] use a mixture of hindered and restricted models of water diffusion. Behrens *et al.*[11,32] propose a partial volume model that mixes one isotropic (the background) and several anisotropic components (the fibers). A continuous mixture of diffusion tensors is proposed by Jian *et al.*[33] The ODF is modeled as the convolution of the fODF and a smoother response function in a number of other studies.[12,17,19,34] However, these estimation methods do not exploit the relations between neighboring voxels; they model the data for a single spatial location. Another key difference with our proposed method is that the distribution of the HARDI data is assumed to follow a specific model (mixture of tensors, convolution of fODF and a constant kernel), whereas in our approach, the distributions are learned from training data using a generic model (a GMM).

Spatially variant GMMs have been used in natural image processing by Nikou *et al.*[36] (extended by Sfikas *et al.*[37]) and Rivera *et al.*[38] (the latter with some simplifications). These methods use the weights of the mixture components to segment the images. Peng *et al.*[39] also applied a similar method to segmentation of conventional (i.e., nondiffusion-weighted) MRI data into three different types of tissue. As opposed to our proposed method, the data in these three studies are scalar fields, the number of mixture components is fixed rather than learned, and no denoising is performed. Martin-Fernandez *et al.*[40] proposed a Gaussian Markov random field for simple denoising of DTI data, but their system cannot model multimodal distributions (and therefore HARDI). Finally, King *et al.*[41] also described a system similar to ours in a general way, though without any mathematical formulation.

I.E. Outline

The rest of the paper is organized as follows. Section II describes the dataset utilized in the study. The image model, as well as the mechanisms to train it and to apply it to image denoising and segmentation, is described in Sec. III. The proposed method is evaluated in Sec. IV. Finally, Sec. V includes the discussion.

II. MATERIALS

HARDI data from 100 different healthy subjects were acquired at the Center for Magnetic Resonance at the University of Queensland using a 4 Tesla Bruker Medspec scanner. A single-shot echo planar technique with a twice-refocused spin echo sequence was utilized. Ninety-four diffusion-sensitized gradient directions and 11 baseline images with no diffusion-sensitization were obtained for every subject. Imaging parameters were as follows: b-value = 1159 s/mm^2, TE/TR = 92.3/8259 ms, voxel size = 1.8 mm × 1.8 mm × 2.0 mm, and image size 128 × 128 × 55 voxels. The acquisition time was approximately 15 min. The SNR, using the mean square of the foreground as signal power and the square of the first peak in the histogram as noise power (estimator for Rice distribution), was on average 24 dB for the baselines and 15 dB for the diffusion-attenuated images. The 11 baseline images were merged down to a single estimate of the reference using a method tailored for Rician data,[42] increasing the SNR by approximately 10 dB up to 34 dB. The reference image was used to segment the brain using the BET algorithm.[43] The resulting mask was then applied to all the diffusion images to avoid including nonbrain matter in the model.

In order to evaluate denoising algorithms, it is necessary to a have ground truth that can be corrupted, subsequently restored and finally compared with the original. Hardware or software phantoms can be used as ground truth, but they are not a realistic approximation of the brain white matter. Ideally, one could use very long acquisition sequences (e.g., 24 or 48 h) to obtain very high SNR data that could be used as ground truth. However, these acquisition protocols are usually limited to ex-vivo scenarios due to motion artifacts. In this study, the original 94D data are too noisy to be directly used as ground truth [see Fig. 1(a)]. Instead, we

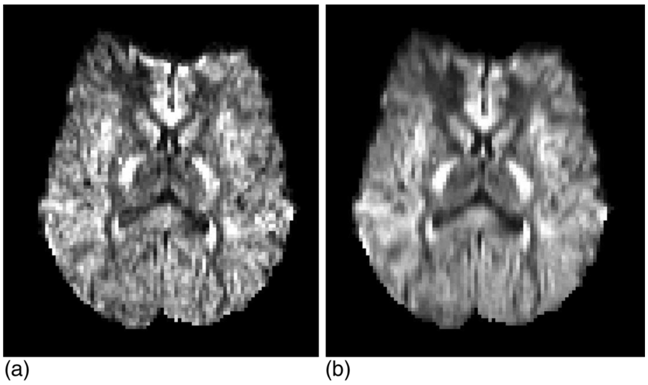

FIG. 1. Axial slice (i.e., parallel to the ground) of the gradient image corresponding to direction $(x, y, z) = (0, 1, 0)$ before (a) and after (b) angular downsampling. The downsampled images are the cleanest data that could be generated and were thus used as ground truth (i.e., SNR assumed to be ∞). Note that there is no spatial smoothing in (b).

propose to trade some angular resolution for redundancy that can be used to obtain a cleaner (though angularly coarser) pseudoground truth as follows.

First, the 94D DW-MRI data at each spatial location are downsampled to a new set of directions using a Laplace–Beltrami regularized spherical harmonic expansion of order six[44] (regularization coefficient $\lambda = 0.006$). The fewer directions, the better the fit will be. On the other hand, the number of directions must remain high enough to be considered HARDI. We found 30 directions to be a good compromise. The 30 directions were computed with an electrostatic approach.[45] Henceforth, the 94D data are discarded, and the downsampled, smoothed (in angular domain) 30D data are assumed to be the noiseless ($\text{SNR} = \infty$) ground truth [see Fig. 1(b)]. A secondary advantage of the angular downsampling is that it lightens the computational load of the algorithms (denoising a volume with the proposed approach takes 25 min, see Discussion).

To evaluate the usefulness of the mixture weights in image segmentation, the second author made (under supervision of an expert radiologist) annotations of two brain structures (left and right ventricles and caudate nuclei) on the baseline images, which have higher SNR than the directional data because they are not attenuated by diffusion. Moreover, 25 of the test images were reannotated by the same author to estimate the intrauser variability in the annotations. These annotations can be used to train a supervised pixel classifier and evaluate its performance.

Finally, the 100 DW-MRI scans were randomly divided into two groups of 50 for cross validation. The first group was used to estimate the parameters of the model and to train the classifier. The images in the second group were artificially corrupted with noise and subsequently denoised to assess the performance of the proposed method.

III. METHODS

In this section, we first describe the image model in Sec. III A. Then, the steps to estimate the parameters of the model from the data are detailed in Sec. III B. Finally,the process to denoise diffusion MRI data is described in Sec. III C.

III.A. Image model

A HARDI volume consists of a set of MRI images corresponding to a baseline and a set of different gradient directions. Since the diffusion is estimated by comparing the gradient images with the baseline, having a reliable estimate of the latter is very important. For this reason, it is usual that multiple (four or five) instances of the baseline volume are acquired and merged down into a reference with little noise. If this reference is denoted by S_0, the signal intensity S_k at a voxel location $\boldsymbol{r}$ when a gradient is applied in the direction (θ_k, φ_k) is given by the Stejskal–Tanner equation[46]

$$S_k(\boldsymbol{r}) = S_0(\boldsymbol{r}) e^{-b_k ADC_k(\boldsymbol{r})} \Rightarrow x_k(\boldsymbol{r}) = \frac{S_k(\boldsymbol{r})}{S_0(\boldsymbol{r})} = e^{-b_k ADC_k(\boldsymbol{r})},$$

where $\theta \in [0, \pi]$ is the elevation, $\varphi \in [0, 2\pi]$ is the azimuth, $ADC_k(\boldsymbol{r}) \geq 0$ is the apparent diffusion coefficient, $x_k(\boldsymbol{r}) \in [0, 1]$ is the attenuation by diffusion and b_k is the Le Bihan's factor, which groups several physical parameters from the image acquisition. $ADC = \{ADC_k\}$, $\boldsymbol{x} = \{x_k\}$ and $\boldsymbol{S} = \{S_k\}$ have the property that they are symmetric because diffusion MRI cannot distinguish the two directions of each orientation, and hence, $S_{(\theta,\varphi)} = S_{(\pi-\theta,\varphi\pm\pi)}$.

The attenuation by diffusion in a voxel can be modeled with a multicompartmental model[4,32]

$$x_k(\boldsymbol{r}) \approx \left(1 - \sum_{f=1}^{NF(\boldsymbol{r})} w_f\right) e^{-b_k \chi_0} + \sum_{f=1}^{NF(\boldsymbol{r})} w_f e^{-b_k \iota_k^t \bar{\bar{\chi}}_f(\boldsymbol{r}) \iota_k}, \qquad (1)$$

where NF is the number of fiber populations that coexist in the voxel, $w_f \geq 0$ is the weight of each population, χ_0 is the anisotropic diffusion, $\bar{\bar{\chi}}$ is the diffusion tensor corresponding to each fiber population, and ι_k is the unit vector in direction (θ_k, φ_k). The attenuation x_k (and thus the signal $S_k = S_0 x_k$) is hence a mixture of an isotropic component (the first term in Eq. 1) plus a number of fiber components (the second half of the equation).

Rather than estimating NF for each voxel and then fitting the weights and diffusion tensors,[4] we propose modeling the attenuation by diffusion $\boldsymbol{x}(\boldsymbol{r})$ as a generic mixture learned from training data.

We assume that the baseline $S_0(\boldsymbol{r})$ is a perfect estimate and only the gradient images $S_k(\boldsymbol{r})$ are affected by noise (see Fig. 2). This is reasonable because, if we estimate $\widehat{x_k} = S_k / \widehat{S_0} = S_k / \sum_{i=1}^{11} S_{0,i}$ (where $S_{0,i}$ are the signal intensities corresponding to the 11 acquired baseline images), the sensitivities of the estimate $\widehat{x_k}$ with respect to S^k and $S_{0,i}$ are

$$\left|\frac{\partial \widehat{x_k}}{\partial S_k}\right| = \frac{1}{\widehat{S_0}} \quad \text{and}$$

$$\left|\frac{\partial \widehat{x_k}}{\partial S_{0,i}}\right| = \frac{S_k}{11 \widehat{S_0}^2} = \frac{\widehat{x_k}}{11 \widehat{S_0}} = \frac{\widehat{x_k}}{11} \frac{\partial \widehat{x_k}}{\partial S_k} \leq \frac{1}{11} \left|\frac{\partial \widehat{x_k}}{\partial S_k}\right|,$$

since $\widehat{x_k} \in [0, 1]$.

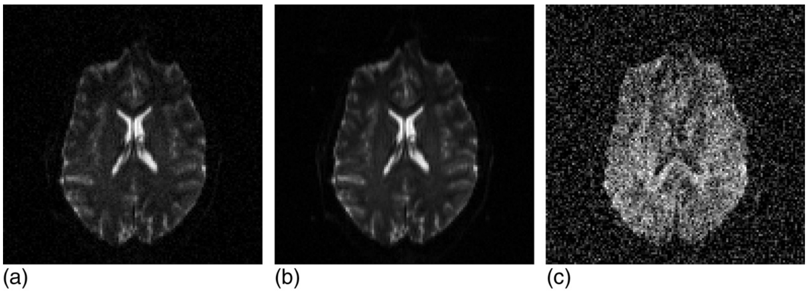

FIG. 2. Axial slice of a baseline volume (a), the corresponding average baseline (b), and one of the corresponding gradient images (c). We assume in this study that the noise of (b) is negligible compared with the noise of (c). The SNR of the original images was degraded 5 dB to better illustrate the difference between (b) and (c). Note that the intensity scales of (a–b) and (c) are different: the noise level is actually the same in (a) and (c).

Because the typical value of $\widehat{x_k}$ is $1/3$ (9 dB attenuation, see Sec. II), the sensitivity with respect to S_k is typically $\sim$30 larger than with respect to $S_{0,i}$, and at least 11 times larger. In other words, the noise in S_k has to be at least 11 times larger than in $S_{0,i}$ to have the same impact on $\widehat{x_k}$.

Now, under the assumption that $S\ _k$ is the only source of noise, the data for each voxel are uniquely characterized by the vector of attenuations $\boldsymbol{x}(\boldsymbol{r}) = [x_1(\boldsymbol{r}), x_2(\boldsymbol{r}), \ldots, x_D(\boldsymbol{r})]^t$, where D is the number of probed directions. This assumption also simplifies the framework described in Sec. III C.

Next, we assume that the underlying "true" (i.e., not corrupted by noise) attenuation $\boldsymbol{x}(\boldsymbol{r})$ is a realization of a Gaussian mixture model (GMM)

$$p(\boldsymbol{x}(\boldsymbol{r})) = \sum_{c=1}^{C} \pi_c(\boldsymbol{r}) G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_c, \Sigma_c) \tag{2}$$

where $G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_c, \Sigma_c)$ is a Gaussian distribution with mean $\boldsymbol{\mu}_c$ and covariance Σ_c, and each component plays a similar role as a fiber population in Eq. (1). While the means $\{\boldsymbol{\mu}_c\}$ and covariance matrices $\{\Sigma_c\}$ are independent of lattice locations, the weights of the mixture $\pi(\boldsymbol{r}) = [\pi_1(\boldsymbol{r}) \ldots \pi_C(\boldsymbol{r})]^t$ are allowed to be spatially variant, and they are constrained to lie in the probability simplex $\sum_{c=1}^{C} \pi_c(\boldsymbol{r}) = 1$, with $\pi_c(\boldsymbol{r}) \geq 0$. Most of the weights will be close to zero for a given voxel, which can be interpreted as a way of selecting $NF(\boldsymbol{r})$.

Now, spatial coherence of the data is ensured by placing a Markov random field (MRF) on top of the mixture weights. Instead of laying a distribution on top of the data directly, laying it on top of another distribution (i.e., the GMM, which already has certain flexibility in the covariance matrices) gives an additional degree of freedom to the model[36]

$$p(\pi(\boldsymbol{r}_1), \pi(\boldsymbol{r}_2), \ldots, \pi(\boldsymbol{r}_N)) \propto \prod_{d=1}^{3} \prod_{c=1}^{C} \beta_{d,c}^{-N} \times \exp\left[-\frac{1}{2} \frac{\sum_{i=1}^{N} (2\pi_c(\boldsymbol{r}_i) - \pi_c(\boldsymbol{r}_i + \boldsymbol{u}_d) - \pi_c(\boldsymbol{r}_i - \boldsymbol{u}_d))^2}{\beta_{d,c}^2} \right], \tag{3}$$

where N is the total number of voxels in the volume, $\beta_{d,c}^2$ is the variance of the mixture component c in the spatial directions $d = \{1, 2, 3\}$ (corresponding to the x, y, z axes), and $\boldsymbol{u}_d$ represents the unit vector along d. The variances $\beta_{d,c}^2$ are allowed to be direction dependent because the voxels are in general nonisotropic, and larger pixel spacings require larger variances in the model. The interpretation of the MRF prior is as follows: when one mixture component at one voxel is predicted as the average of the values of its neighbors in direction d, the error follows a zero-mean Gaussian distribution with variance $\beta_{d,c}^2/2$.

Finally, the observed image data are a noise-corrupted version of $\boldsymbol{x}(\boldsymbol{r})$. The probability distribution of the noise in MRI is Rician

$$p(\boldsymbol{x}^*(\boldsymbol{r})|\boldsymbol{x}(\boldsymbol{r})) = \prod_{k=1}^{D} \frac{S_0(\boldsymbol{r}) x_k^*(\boldsymbol{r})}{\sigma^2} e^{-\frac{S_0^2(\boldsymbol{r})(x_k^2(\boldsymbol{r}) + {x^*}_k^2(\boldsymbol{r}))}{2\sigma^2}} \times I_0\left(\frac{S_0^2(\boldsymbol{r}) x_k(\boldsymbol{r}) x_k^*(\boldsymbol{r})}{\sigma^2} \right), \tag{4}$$

where $\boldsymbol{x}^*(\boldsymbol{r})$ represents the noise-corrupted observations, I_0 is the modified Bessel function of the first kind with order zero, and σ_k^2 is the noise power for gradient image in direction (θ_k, φ_k), typically constant across k, i.e., $\sigma_k^2 = \sigma^2, \forall k$. The complete graphical model of the data is depicted in Fig. 3.

III.B. Model training

Given some training data, the parameters of the model are estimated by maximizing the target function Q, which depends on $\{\boldsymbol{\mu}_c\}$, $\{\Sigma_c\}$, and $\{\pi_c(\boldsymbol{r})\}$

$$Q = \sum_{\boldsymbol{r}} \sum_{c=1}^{C} z_c(\boldsymbol{r}) \{\log[\pi_c(\boldsymbol{r})] + \log[G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_c, \Sigma_c)]\} - \frac{1}{2} \sum_{\boldsymbol{r}} \sum_{c=1}^{C} \sum_{d=1}^{3} \left[\log(\beta_{d,c}^2) + \frac{[2\pi_c(\boldsymbol{r}) - \pi_c(\boldsymbol{r} + \boldsymbol{u}_d) - \pi_c(\boldsymbol{r} - \boldsymbol{u}_d)]^2}{\beta_{d,c}^2} \right], \tag{5}$$

where $z_c(\boldsymbol{r})$ is the posterior probability that the voxel at $\boldsymbol{r}$ belongs to component c

$$z_c(\boldsymbol{r}) = \frac{\pi_c(\boldsymbol{r}) G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_c, \Sigma_c)}{\sum_{c'=1}^{C} \pi_{c'}(\boldsymbol{r}) G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_{c'}, \Sigma_{c'})}. \tag{6}$$

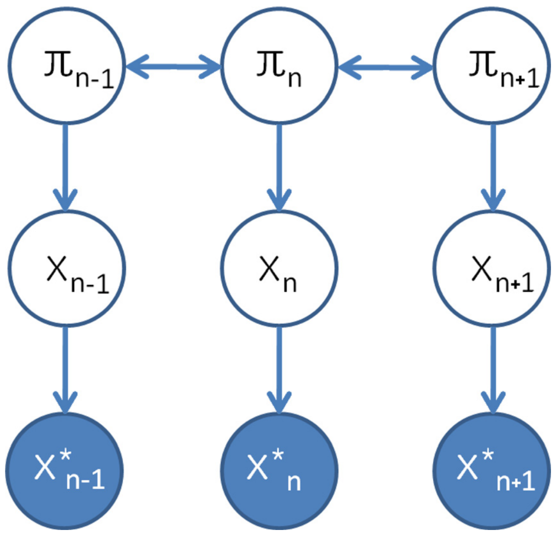

FIG. 3. Graphical model for HARDI (in 1D, for the sake of simplicity): $\boldsymbol{\pi}$ represents the mixture weights, $\boldsymbol{x}$ represents the ADC data, and $\boldsymbol{x}^*$ represents the noise-corrupted observations.

The first term (first row) in Eq. (5) corresponds to the conditional likelihood of the GMM given the weights.[47] The second term accounts for the joint likelihood of the weights themselves, given by the MRF. The approach that was utilized to maximize Q is an expectation maximization (EM) algorithm inspired by the work by Nikou *et al.*,[36] with two major differences: the maximization step of the EM algorithm has been substantially improved and the number of components is not fixed but determined by the MML criterion.[48]

III.B.1. Initialization: choice of C with minimum message length criterion (MML)

To initialize the algorithm, the problem can first be solved without the MRF term from Eq. (3) in the target function Q from Eq. (5). This amounts to the well-known problem of estimating the parameters of a GMM, which can be solved: with the EM algorithm. In turn, the EM can be initialized with the K-means algorithm, which is further initialized with random samples from the training data. The EM algorithm is prone to getting stuck in local maxima, and it is therefore beneficial to train the GMM with different initializations for K-means and then keep the instance with the highest final likelihood.

It is well known that one of the main limitations of mixture models is that the number of components C must be defined in advance. A possible way of tackling this problem is to select the value of C that minimizes the minimum message length (MML), a criterion from information theory. MML combines two terms: the inverse of the likelihood of the training data and a term that penalizes the complexity of the model. Higher values of C decrease the first term but also increase the second, so a compromise must be reached

$$MML(C) = -\left[\sum_{\boldsymbol{r}}\sum_{c=1}^{C} z_c(\boldsymbol{r})\{\log(\pi_c) + \log[G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_c, \sum_c)]\}\right] + \left[\frac{D}{2}\sum_{c=1}^{C}\log\frac{N\pi_c}{12} + \frac{C}{2}\log\frac{N}{12} + \frac{C(D+1)}{2}\right], \tag{7}$$

where the weights π_c are location-independent at this point (the MRF prior is not considered yet).

Training the GMM for a wide range of values of C to find which one provides the minimal MML is computationally impractical. The following greedy algorithm was used instead: the parameters are estimated just once for a sole, high value of C, and then, components are merged down as long as the MML decreases.[49] At each step, all possible pairwise merges are considered. The merge that leads to a model with the lowest MLL is selected and proposed for acceptance. If this value of MML represents an improvement with respect to the MML of the current model, the merge is accepted and the process is repeated. If it does not, the algorithm terminates. The procedure is fast because, when a merge is considered, the terms corresponding to the C-2 components not being merged do not need to be reevaluated.

III.B.2. Optimization with EM algorithm

Once the number of components C is defined, the cost function from Eq. (5) can be maximized. This can be achieved through an EM algorithm. In the E step, the posterior probabilities $z_c(\boldsymbol{r})$ are calculated using Eq. (6). In the M step, the estimates of the parameters $(\boldsymbol{\mu}_c), \{\sum_c\}, \{\beta^2_{d,c}\}$ and the mixture weights at each pixel $\pi_c(\boldsymbol{r})$ must be updated to maximize Q. The update equations for the mixture model parameters are well-known[50]

$$\boldsymbol{\mu}_c \leftarrow \frac{\sum_{\boldsymbol{r}} z_c(\boldsymbol{r})\boldsymbol{x}(\boldsymbol{r})}{\sum_{\boldsymbol{r}} z_c(r)}, \quad \sum\nolimits_c \leftarrow \frac{\sum_{\boldsymbol{r}} z_c(\boldsymbol{r})[(\boldsymbol{x}(\boldsymbol{r})-\boldsymbol{\mu}_c)][x(r)-\boldsymbol{\mu}_c]^2}{\sum_{\boldsymbol{r}} z_c(r)}. \tag{8}$$

Updating the mixture weights in order to be able to update their variances $\{\beta^2_{d,c}\}$ is more complicated. Simultaneous optimization with respect to all the weights is very difficult and impractical. An alternative is to utilize coordinate ascent, in which voxels are visited in a random order and their weights are optimized under the assumption that the weights at all other locations are fixed. The algorithm typically converges after five or six passes over the volume.

In order to optimize the weights for a given spatial location $\boldsymbol{r}$, Nikou *et al.*[36] compute the derivatives of the target function Q with respect to $\{\pi_c(\boldsymbol{r})\}$, set them to zero, and solve the resulting second degree equations. Because the weights are optimized in an unconstrained fashion, they must be projected back onto the probability simplex after each step. The problem with this approach is that the solution to the equations can yield a point arbitrarily far from the simplex, and the projected point is not guaranteed to increase Q (in practice, this often happens). Instead, we solve the problem exactly with the help of a Lagrange multiplier. If the terms in the target function that are independent from $\{\pi_c(\boldsymbol{r})\}$ are disregarded, some algebraic manipulation of Eq. (5) yields the following optimization problem:

$$\max_{\{\pi_c(\boldsymbol{r})\}} \sum_{c=1}^{C} \left[w_c \log \pi_c(\boldsymbol{r}) - \frac{1}{2}(\pi_c(\boldsymbol{r}) - a_c)^2 \right]$$
$$\text{subject to } \sum_{c=1}^{C} \pi_c(\boldsymbol{r}) = 1, \quad \pi_c(\boldsymbol{r}) > 0, \tag{9}$$

where

$$w_c = \frac{z_c \prod_{d=1}^{3} \beta_{d,c}^2}{4 \sum_{d=1}^{3} \prod_{\substack{d'=1 \\ d' \neq d}}^{3} \beta_{d',c}^2}$$

$$a_c(\boldsymbol{r}) = \frac{\sum_{d=1}^{3} \prod_{\substack{d'=1 \\ d' \neq d}}^{3} \beta_{d',c}^2 (\pi_c(\boldsymbol{r} + \boldsymbol{u}_d) + \pi_c(\boldsymbol{r} - \boldsymbol{u}_d))}{2 \sum_{d=1}^{3} \prod_{\substack{d'=1 \\ d' \neq d}}^{3} \beta_{d',c}^2},$$

The corresponding Lagrangian is

$$L(\{\pi_c(\boldsymbol{r})\}, \lambda) = \sum_{c=1}^{C} \left[w_c \log \pi_c(\boldsymbol{r}) - \frac{1}{2}(\pi_c(\boldsymbol{r}) - a_c(\boldsymbol{r}))^2 \right] + \lambda \left[1 - \sum_{c=1}^{C} \pi_c(\boldsymbol{r}) \right],$$

where the constraint $\pi_c(\boldsymbol{r}) > 0$ has been left implicit (it must be observed so that $\log \pi_c(\boldsymbol{r}) \in (\mathbb{R})$) and λ is the multiplier corresponding to the explicit constraint. Now, setting the derivatives with respect to the weights equal to zero gives

$$\frac{w_j}{\pi_j(\boldsymbol{r})} - \pi_j(\boldsymbol{r}) + a_j(\boldsymbol{r}) - \lambda = 0 \Rightarrow$$
$$\pi_j(\boldsymbol{r}) = \frac{a_j(\boldsymbol{r}) - \lambda + \sqrt{(a_j(\boldsymbol{r}) - \lambda)^2 + 4w_j}}{2}, \tag{10}$$

where the solution corresponding to the minus sign is discarded because it would yield $\pi_j(\boldsymbol{r}) \leq 0$, since $w_j > 0$. Finally, the value of the multiplier can be calculated by solving $\sum_{j=1}^{C} \pi_j(\boldsymbol{r}) - 1 = f(\boldsymbol{r}, \lambda) = 0$. If $f(\boldsymbol{r}, \lambda)$ is expanded,

$$f(\lambda, \boldsymbol{r}) = \sum_{j=1}^{C} \frac{a_j(\boldsymbol{r})}{2} - \frac{C}{2}\lambda + \frac{1}{2} \sum_{j=1}^{C} \sqrt{(a_j(\boldsymbol{r}) - \lambda)^2 + 4w_j} - 1 = 0.$$

There is no simple expression for the roots of this function. However, it is easy to check that $df/d\lambda < 0 \; \forall \lambda$. Since f is continuous and differentiable everywhere, this means that there is only one zero, which can be easily found with numerical methods; a bisection strategy was used in this study. Once the value of the multiplier λ has been determined, substitution in Eq. (10) provides the optimal mixture weights.

Finally, the EM iteration ends with calculating the new class variances from the updated mixture weights

$$\beta_{d,c}^2 \leftarrow \frac{1}{N} \sum_{\boldsymbol{r}} [2\pi_c(\boldsymbol{r}) - \pi_c(\boldsymbol{r} + \boldsymbol{u}_d) - \pi_c(\boldsymbol{r} - \boldsymbol{u}_d)]^2. \tag{11}$$

The EM algorithm is summarized in Table I.

TABLE I. EM algorithm to estimate the model parameters.

1. Initialize the mixture model components $\{\mu_c, \Sigma_c\}$ with the results from Sec. III B 1.
2. Initialize the mixture weights for every voxel with the weights of the MRF-free GMM from the same section.
3. While the function Q from Equation 5 keeps increasing:
   a) Calculate the posterior probabilities $z_c(\boldsymbol{r})$ with Equation (6).
   b) Update the mixture model components with Equation (8).
   c) Visit all voxels in random order, updating their mixture weights $\pi_c(\boldsymbol{r})$ with the solution of the problem from Equation (9).
   d) Repeat step c) until convergence (typically 5–6 times).
   e) Update the class variances $\beta_{d,c}^2$ with Equation (11).

III.C. Application to image denoising

The model can be used for image denoising by fitting it to an image. In a Bayesian framework, the goal is to maximize the posterior probability of the underlying noise-free image given the noisy version. As it happened in the EM algorithm from the training stage (Sec. III B.2), direct optimization with respect to all variables is computationally very difficult. Once more, coordinate descent can be used. In this context, the algorithm is known as iterative conditional modes (ICM).[51] In ICM, voxels are visited in random order, and their values are optimized by maximizing the conditional probability $p(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{x}^*(\boldsymbol{r}), \widehat{\boldsymbol{x}}(S \setminus \boldsymbol{r}))$, where $\boldsymbol{x}^*(\boldsymbol{r})$ is the observed (noisy) attenuation at the given voxel and $\widehat{\boldsymbol{x}}(S \setminus \boldsymbol{r})$ is the current restored value at all other locations. The algorithm typically converges (i.e., negligible change in $\boldsymbol{x}$) after a few passes over the whole volume.

The joint conditional distribution of the mixture weights $\pi(\boldsymbol{r})$ and the "real" attenuation $\boldsymbol{x}(\boldsymbol{r})$ at each voxel are given by Bayes's rule [(Eq. (13)]. Maximizing the joint probability $p(\boldsymbol{x}(\boldsymbol{r}), \pi(\boldsymbol{r}))$ is much faster than maximizing the marginal $p(\boldsymbol{x}(\boldsymbol{r}))$ and yields good results (see Sec. V). Under the assumptions in this study (namely the graphical model in Fig. 3), the following independence relationships hold:

- $p(\boldsymbol{x}^*(\boldsymbol{r})|\boldsymbol{x}(\boldsymbol{r}), \pi(\boldsymbol{r}), \pi(S \setminus \boldsymbol{r}), \boldsymbol{x}(S \setminus \boldsymbol{r})) = p(\boldsymbol{x}^*(\boldsymbol{r})|\boldsymbol{x}(\boldsymbol{r}))$,
- $p(\boldsymbol{x}(\boldsymbol{r})|\pi(\boldsymbol{r}), \pi(S \setminus \boldsymbol{r}), \boldsymbol{x}(S \setminus \boldsymbol{r})) = p(\boldsymbol{x}(\boldsymbol{r})|\pi(\boldsymbol{r}))$,
- $p(\pi(\boldsymbol{r})|\pi(S \setminus \boldsymbol{r}), \boldsymbol{x}(S \setminus \boldsymbol{r})) = p(\pi(\boldsymbol{r})|\pi(\aleph_{\boldsymbol{r}}))$.

where $\aleph_{\boldsymbol{r}}$ represents the neighborhood of $\boldsymbol{r}$. Substitution in Eq. (13) yields the final expression for the posterior likelihood of the "real," underlying data

$$POST.L. \propto p(\boldsymbol{x}(\boldsymbol{r}),\pi(\boldsymbol{r})|\boldsymbol{x}^*(\boldsymbol{r}),\pi(S\setminus\boldsymbol{r}),\boldsymbol{x}(S\setminus\boldsymbol{r})) = \underbrace{p(\boldsymbol{x}^*(\boldsymbol{r})|\boldsymbol{x}(\boldsymbol{r}))}_{\text{NOISE MODEL}}\cdot\underbrace{p(\boldsymbol{x}(\boldsymbol{r})|\pi(\boldsymbol{r}))}_{\text{GMM}}\cdot\underbrace{p(\pi(\boldsymbol{r})|\pi(\aleph_{\boldsymbol{r}}))}_{\text{MRF}} = F, \tag{12}$$

$$\begin{aligned} &p(\boldsymbol{x}(\boldsymbol{r}),\pi(\boldsymbol{r})|\boldsymbol{x}^*(\boldsymbol{r}),\pi(S\setminus\boldsymbol{r}),\boldsymbol{x}(S\setminus\boldsymbol{r})) \propto p(\boldsymbol{x}^*(\boldsymbol{r})|\boldsymbol{x}(\boldsymbol{r}),\pi(\boldsymbol{r}),\pi(S\setminus\boldsymbol{r}),\boldsymbol{x}(S\setminus\boldsymbol{r}))\cdot p(\boldsymbol{x}(\boldsymbol{r}),\pi(\boldsymbol{r})|\pi(S\setminus\boldsymbol{r}),\boldsymbol{x}(S\setminus\boldsymbol{r}))\\ &= \ldots\ldots = p(\boldsymbol{x}^*(\boldsymbol{r})|\boldsymbol{x}(\boldsymbol{r}),\pi(\boldsymbol{r}),\pi(S\setminus\boldsymbol{r}),\boldsymbol{x}(S\setminus\boldsymbol{r}))\cdot p(\boldsymbol{x}(\boldsymbol{r})|\pi(\boldsymbol{r}),\pi(S\setminus\boldsymbol{r}),\boldsymbol{x}(S\setminus\boldsymbol{r}))\quad\cdot p(\pi(\boldsymbol{r})|\pi(S\setminus\boldsymbol{r}),\boldsymbol{x}(S\setminus\boldsymbol{r})) = F, \end{aligned} \tag{13}$$

$$\begin{aligned} p(\pi(\boldsymbol{r})|\pi(\aleph_{\boldsymbol{r}})) &= \prod_{c=1}^{C}\prod_{d=1}^{3} G\left(2\pi_c(\boldsymbol{r})|\pi_c(\boldsymbol{r}+\boldsymbol{u}_d)+\pi_c(\boldsymbol{r}-\boldsymbol{u}_d),\beta^2_{d,c}\right) = \ldots\\ \ldots &= \prod_{c=1}^{C} G\left[2\pi_c(\boldsymbol{r})\middle|\frac{\sum_{d=1}^{3}(\pi_c(\boldsymbol{r}+\boldsymbol{u}_d)+\pi_c(\boldsymbol{r}-\boldsymbol{u}_d))\prod_{\substack{p=1\\p\neq d}}^{3}\beta^2_{p,c}}{\sum_{d=1}^{3}\prod_{\substack{p=1\\p\neq d}}^{3}\beta^2_{p,c}},\frac{\prod_{d=1}^{3}\beta^2_{d,c}}{\sum_{d=1}^{3}\prod_{\substack{p=1\\p\neq d}}^{3}\beta^2_{p,c}}\right], \end{aligned} \tag{14}$$

$$\frac{\partial\log(F)}{\partial\pi_c(\boldsymbol{r})} = \frac{G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_c,\Sigma_c)}{\sum_{c'=1}^{C}\pi_{c'}(\boldsymbol{r})G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_{c'},\Sigma_{c'})} - 2\frac{\sum_{d=1}^{3}(2\pi_c(\boldsymbol{r})-\pi_c(\boldsymbol{r}+\boldsymbol{u}_d)-\pi_c(\boldsymbol{r}-\boldsymbol{u}_d))\prod_{\substack{d'=1\\d'\neq d}}^{3}\beta^2_{d',c}}{\prod_{d=1}^{3}\beta^2_{d,c}}, \tag{15}$$

$$\frac{\partial\log(F)}{\partial x_k(\boldsymbol{r})} = \frac{S_0^2(\boldsymbol{r})}{\sigma_k^2}\left[\frac{I_1\left(\frac{S_0^2(\boldsymbol{r})x_k(\boldsymbol{r})x_k^*(\boldsymbol{r})}{\sigma^2}\right)}{I_0\left(\frac{S_0^2(\boldsymbol{r})x_k(\boldsymbol{r})x_k^*(\boldsymbol{r})}{\sigma^2}\right)}x_k^*(\boldsymbol{r}) - x_k(\boldsymbol{r})\right] + \frac{\sum_{c=1}^{C}\left(\pi_c(\boldsymbol{r})G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_c,\Sigma_c)[\Sigma_c(\boldsymbol{\mu}_c-\boldsymbol{x}(\boldsymbol{r}))]_k\right)}{\sum_{c=1}^{C}\pi_c(\boldsymbol{r})G(\boldsymbol{x}(\boldsymbol{r})|\boldsymbol{\mu}_c,\Sigma_c)}. \tag{16}$$

The first term in Eq. (12) corresponds to the Rician noise model from Eq. (4). The second term in F is just the GMM from Eq. (2). The third term is a product of univariate Gaussians corresponding to the different mixture components and directions in the MRF. Each Gaussian term is the result of multiplying the conditional probabilities given by the neighbors in each of the three spatial directions [Eq. (14)].

Maximizing the posterior F [(Eq. (12)] is equivalent to maximizing its logarithm, which has a simpler expression. The derivatives of $\log F$ with respect to the mixture weights and attenuation values can be computed analytically [see Eqs. (15) and (16), where $[\cdot]_k$ denotes the kth element of a vector, and I_1 is the modified Bessel function of the first kind with order one]. This makes it possibly to optimize $\log F$ efficiently using gradient ascent, i.e., taking small steps along the direction of the gradient at each point. After each step, it is necessary to project the mixture weights onto its constraint (i.e., the probability simplex). Unlike in Sec. III B. 2, this now works well because the user-defined step is fixed and small. A simple and fast algorithm for projecting onto the probability simplex[52] was utilized. It is also necessary to project $\boldsymbol{x}(\boldsymbol{r})$ onto its own constraint ($\boldsymbol{x}(\boldsymbol{r}) \in [0,1]$) after each step. This is accomplished by setting to zero any component below zero and setting to one any component over one.

IV. EXPERIMENTS AND RESULTS

IV.A. Training the model

The proposed model was trained with the 50 images from the training dataset. Rather than using all the available voxels, a random patch of size $25\times25\times20$ voxels was extracted from each of them to lighten the computational load of the algorithm (in total $N \approx 500,000$ voxels). The patches were constrained to have at least two thirds of their voxels in the mask provided by the BET algorithm. A preliminary experiment evaluating the MML in Eq. (7) at $C = 5n, n = 1, \ldots, 10$, revealed that the minimum was achieved around $C = 25 \sim 30$.

Because the algorithm that selects C can only decrease from an initial value C_{ini} by merging components, the initial GMM without the MRF term was first trained with $C_{\text{ini}} = 35$ components. The training was repeated five times using EM with different seeds for the K-means algorithm, whose output initializes the EM. The model with the highest log-likelihood $\log Q$ was kept. The evolution of the metric $\log Q$ with the number of iterations for the best seed is displayed in Fig. 4(a).

Once the GMM was trained with $C_{\text{ini}} = 35$, components were iteratively merged until the MML did not decrease anymore, which happened at $C = 27$, as shown in Fig. 4(b). The resulting GMM was used as initialization to train the GMM-MRF model [Fig. 4(c)]. The averages over all training pixels of the mixture weights (henceforth $\langle \pi_c \rangle$), which are useful in order to initialize the denoising algorithm (Sec. IV B), were saved along with the model. The model components, along with their weights and first mode of variation, are displayed in Fig. 5. The first mode depicts in most cases the scale of the diffusion. The second mode, displayed in Fig. 6, shows more variability in the orientations. From Figs. 5 and 6, we see that components such as $\{1, 4, 7, 11, 26\}$ represent the isotropic part of the data, whereas the rest of the components display anisotropy in different directions.

IV.B. Denoising

Rician noise was added to the 50 images in the test dataset, and the proposed method (as well as other three algorithms) was then utilized to restore them. Since the system was trained on the 50 other subjects in the training dataset, no bias was introduced in the evaluation. The RMS error in the attenuation data was used as evaluation metric

$$e = \sqrt{\frac{1}{N} \sum_{\boldsymbol{r}} \| \boldsymbol{x}_{\text{denoised}}(\boldsymbol{r}) - \boldsymbol{x}(\boldsymbol{r}) \|^2},$$

where $\boldsymbol{x}_{\text{denoised}}(\boldsymbol{r})$ represents the denoised attenuation data and $\boldsymbol{x}(\boldsymbol{r})$ is the ground truth.

In the denoising scheme proposed in this study, the restored data $\boldsymbol{x}_{\text{denoised}}(\boldsymbol{r})$ was initialized with the noisy version $\boldsymbol{x}^*(\boldsymbol{r})$, whereas the mixture weights $\pi(\boldsymbol{r})$ were initialized with the average weights from the training data (Sec. IV A): $\pi_c(\boldsymbol{r}) = \langle \pi_c \rangle \ \forall \boldsymbol{r}$. The noise parameters $\{\sigma_k\}$ are assumed to be known and constant across the gradient directions: $\sigma_k = \sigma, \forall k$. This is a fair assumption in MRI because there is a large amount of background pixels in the data, so σ_k can be accurately estimated from them, for example, by studying their statistical moments[53] or histogram.[54]

The proposed denoising algorithm was compared at different noise levels with the following strategies: a Gaussian smoothing filter, the classical anisotropic diffusion (AD) filter,[55] and a Rician-adapted, nonlocal means (NLM) filter.[21] Rician bias correction was added to the first two in order to take advantage of the fact that the noise parameter σ is known; otherwise, the experiment would be biased against these two algorithms (the NLM filter already makes use of this information).

Figure 7 shows the performance of the different algorithms at different noise levels. The width of the Gaussian kernel and the conductance parameter of the AD filter were optimized to minimize the error at each noise level using the training data. Gaussian blurring is clearly worse than the other methods for every SNR. Regarding the other approaches, none of them is significantly better than the others at low noise levels. However, the performance of our method decreases slower with noise power than the others. The improvement with respect to AD and NLM is statistically significant ($p < 0.01$) approximately when SNR<11 dB. The proposed method would for example allow for a very fast, noisy acquisition at SNR $\approx$8 dB that could be denoised to provide the same quality as an acquisition at

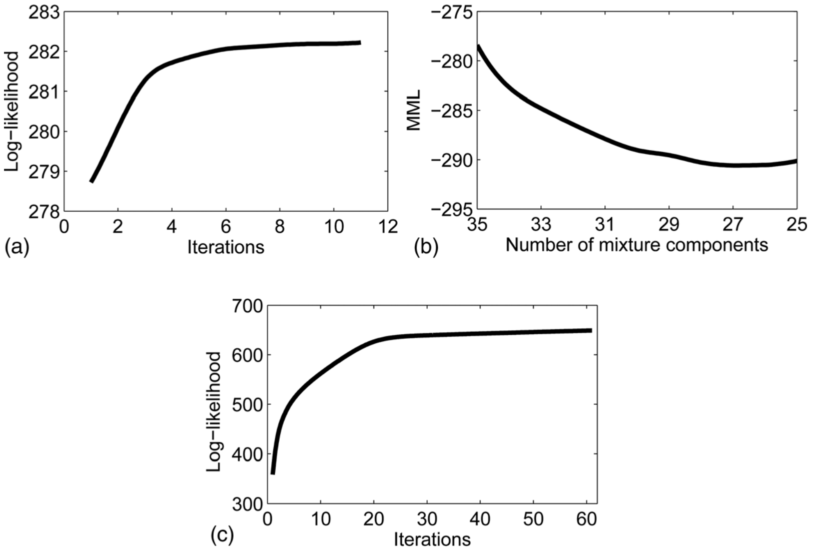

FIG. 4. Evolution of metrics in training: (a) GMM (best of 5), (b) MML, and (c) GMM-MRF.

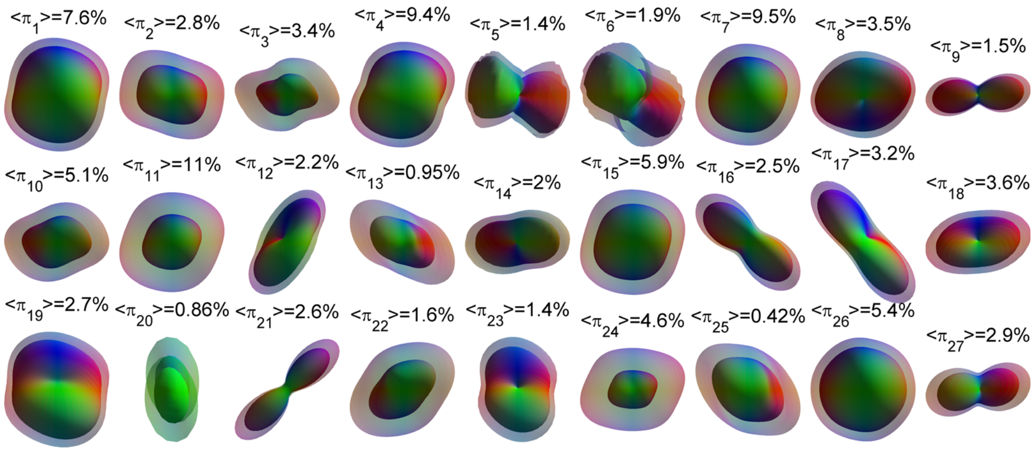

FIG. 5. Mixture components after training. For each component, the first mode of variation ($\boldsymbol{\mu}\pm2\sqrt{\boldsymbol{\lambda}_1}\boldsymbol{e}_1$, where $\{\boldsymbol{\mu},\boldsymbol{\lambda}_1,\boldsymbol{e}_1\}$ are the mean, first eigenvalue, and first eigenvector of the covariance matrix) and the average weights over the training data $\langle\boldsymbol{\pi}_c\rangle$ are shown. The components were fitted to a (Laplace–Beltrami regularized) spherical harmonic expansion of order six and min-max normalized for display. The components are color-coded as follows: red indicates left/right direction, green is anterior/posterior, and blue is inferior/superior. Note that the components represent attenuation, so a toroidal shape implies a single fiber population along the axis.

SNR ≈22 dB restored with a Gaussian filter. At higher SNR, it would be more efficient to use AD or NLM due to their lower computational cost, unless image segmentation or analysis is to be carried out after the denoising. In that case, the features (probability maps) produced by our method can be worth the longer execution time (see Sec. IV C)

In Fig. 8, a slice of a diffusion-weighted image of a sample brain is displayed for the uncorrupted, corrupted (SNR = 10 dB), and denoised versions. The fODF field is also plotted for a region of interest (ROI) focused on the crossing between the corona radiata and the radiation of the corpus callosum. The ODF was estimated at 642 directions using a method by Descoteaux *et al.*[16] with a spherical harmonic expansion of order six. The fODF was computed from the ODF in a nonparametric fashion using Tikhonov regularization.[56] The Gaussian filter requires a very wide kernel to eliminate the noise, which leads to oversmoothing. The anisotropic diffusion filter does preserve some edges slightly better (e.g., the callosal bundle on the right of the fODF field looks thinner) but also introduces too much blurring (e.g., the in the central region of the ROI, where the fiber crossing is smudged). The NLM filter respects the edges in some extent but does not eliminate as much noise as the other filters. Finally, the proposed method takes advantage of the learned diffusion distribution to remove most of the noise with less blurring and therefore less spurious crossings (these are especially noticeable in the output of the Gaussian filter).

The ultimate goal of DW-MRI is fiber tracking. Figure 9 shows 3D reconstructions of the tractography output for a test subject given by a classic streamline tractography method very similar to Descotaux *et al.*[19] From a number of points (1000) in a seed region, a step (0.5 mm) is taken in the direction of the global maximum of the fODF. Then, subsequent steps follow the orientation of the local maximum of the fODF that is closest to the direction of the previous step. The tracking ends if a point of low fractional anisotropy (<0.1) is reached or if the angle between two consecutive steps is above a threshold (60°).

In Fig. 9(a), one seed was placed on the corticospinal tract and another in the body of the corpus callosum to observe the behavior at the intersection of the corpus

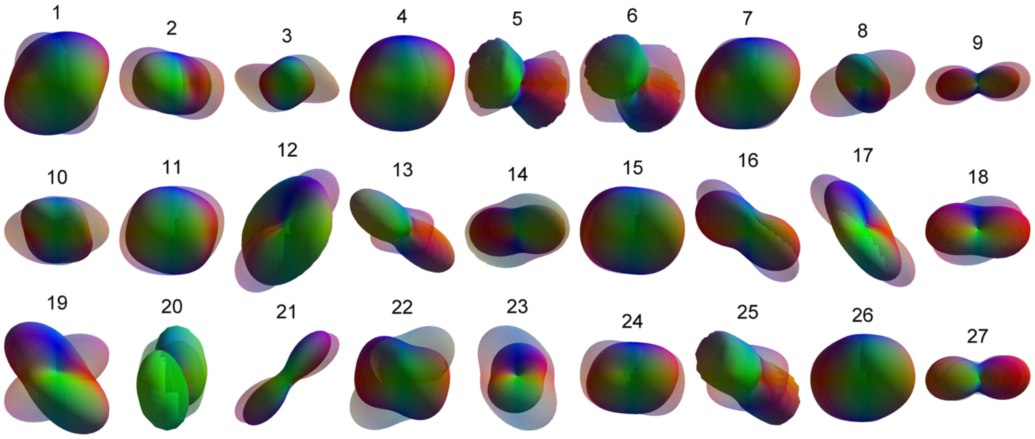

FIG. 6. Mixture components after training: second mode of variation, $\boldsymbol{\mu}\pm2\sqrt{\boldsymbol{\lambda}_2}\boldsymbol{e}_2$ (see caption of Fig. 5).

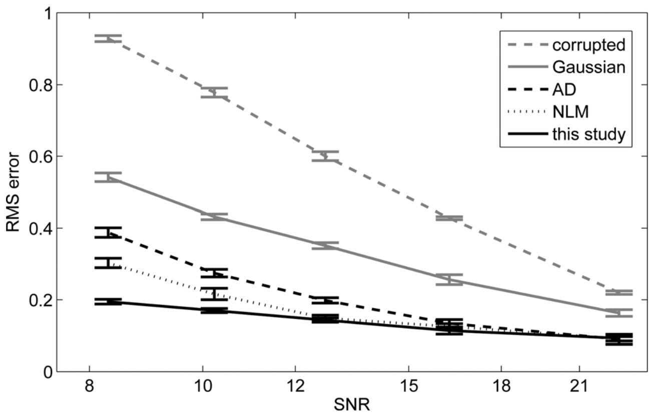

FIG. 7. Average RMS error from uncorrupted image to restored version at different noise levels. The 95% confidence interval, calculated at image level (rather than pixel level), is also displayed.

callosum and its radiation with the corona radiata. In Fig. 9(b), the seeds are in corticospinal tract and in the middle cerebellar peduncle. The corticospinal tract crosses the peduncle on its way to the spinal cord. In Fig. 9(c), there are three seeds: one to track the superior frontal gyrus fibers, one to track the capsule fibers, and one in the genu or the corpus callosum. The genu of the corpus callosum and the capsule fibers converge in the anterior/posterior direction, whereas the superior frontal gyrus fibers bend toward the superior part of the brain. In all three cases, the corrupted version produces many false positives that lead to spurious tracts away from the bundles of interest. The Gaussian filter can partially recover the tracts, but the smoothing makes the bundles look thicker than they really are, especially at crossings. The anisotropic diffusion filter does a better job restoring the main tracts, but falters at the fiber intersections. The nonlocal means filter recovers the bundles to some extent but still displays some false positives. When the noisy data are fitted to our proposed model, the algorithm is able to track the bundles across the intersections as in the uncorrupted volumes.

IV.C. Segmentation

Nikou *et al.*[36] and Rivera *et al.*[38] use a spatially variant mixture model for image segmentation: when the model is fitted to the data, the most likely mixture weights for each voxel $\pi(\boldsymbol{r})$ are provided by the optimization algorithm, and these probability maps can be used as features in a segmentation algorithm. In our case, some of the probability maps correlate almost directly with anatomy (see the example in Fig. 10). It is interesting to test whether these probabilities can be used as features with classification purposes. The following experiment was set up: a support vector machine (SVM) was trained to classify the voxels as ventricle, caudate nucleus, or rest of the brain using the probability images; if the results are good, the mixing weights are good features.

The SVM was trained on the 50 training images and evaluated on the 50 test images. In the training stage, the model was first fitted to the training images. Then, all the positive voxels (i.e., annotated as ventricle or caudate) and an equal amount of randomly selected negative voxels (i.e., marked as background) were used to train the SVM. In the evaluation stage, the voxels in the test images were classified as ventricle, caudate, or background. This provides a mask for the ventricle and a mask for the caudate. These masks were morphologically closed with a spherical kernel (radius 3 mm), and then holes and islands were removed.

Rather than using the $C = 27$ weights as features for each voxel, a feature selection algorithm was used ("plus 3-take away 1"[57]). Feature subsets were evaluated by cross-validation (ten folds with five images each) on the training data. The evolution of the error rate with the number of features is shown in Fig. 11. By visual inspection, the cut-off was set at 13 features for the ventricles and 8 for the caudate nucleus. Once the classifier was trained, its performance was evaluated with the 50 test images. The similarity indices (SI) of the gold standard and the automated segmentations are displayed in Table II. The SI is defined as

$$\mathrm{SI} = \frac{2n(L_a \cap L_b)}{n(L_a) + n(L_b)} \in [0, 1],$$

where $L_{a,b}$ are the compared labels and $n(\cdot)$ is the number of voxels.

These SI values are not far from the intrauser variability (also in Table II), which was determined using the 25 images that were reannotated. This variability defines the precision of the ground truth, and it is an intrinsic limit of the performance that an automated system can achieve. In absolute terms, the SI values are relatively low because of the low resolution of DW-MRI and the fact that the caudate is difficult to delineate in T2-weighted MRI, which is the modality of the baseline. For instance, a recent study using a multiatlas approach[58] reported SI values between 0.66 and 0.83 for the caudate.

V. DISCUSSION

A spatially variant mixture model for diffusion MRI has been presented in this article. The signal attenuation, which is inherently a mixture due to the coexistence of fiber populations, is modeled as a Gaussian mixture in which the mean vectors and covariance matrices are assumed to be independent of spatial locations, whereas the mixture weights are allowed to vary at different lattice locations. Spatial smoothness of the data is ensured by imposing a Markov random field prior on the mixture weights. The model is trained on clean pseudoground truth data using the expectation maximization algorithm. The number of mixture components is determined using the minimum message length criterion from information theory. The system has the advantage of not having any parameters to tune.

The main application of the model is image denoising: given the noisy data, the denoised image can be computed as the most likely explanation in a Bayesian framework. This poses an optimization problem that can be solved within a reasonable time with coordinate descent. The proposed approach was compared with some of the most popular

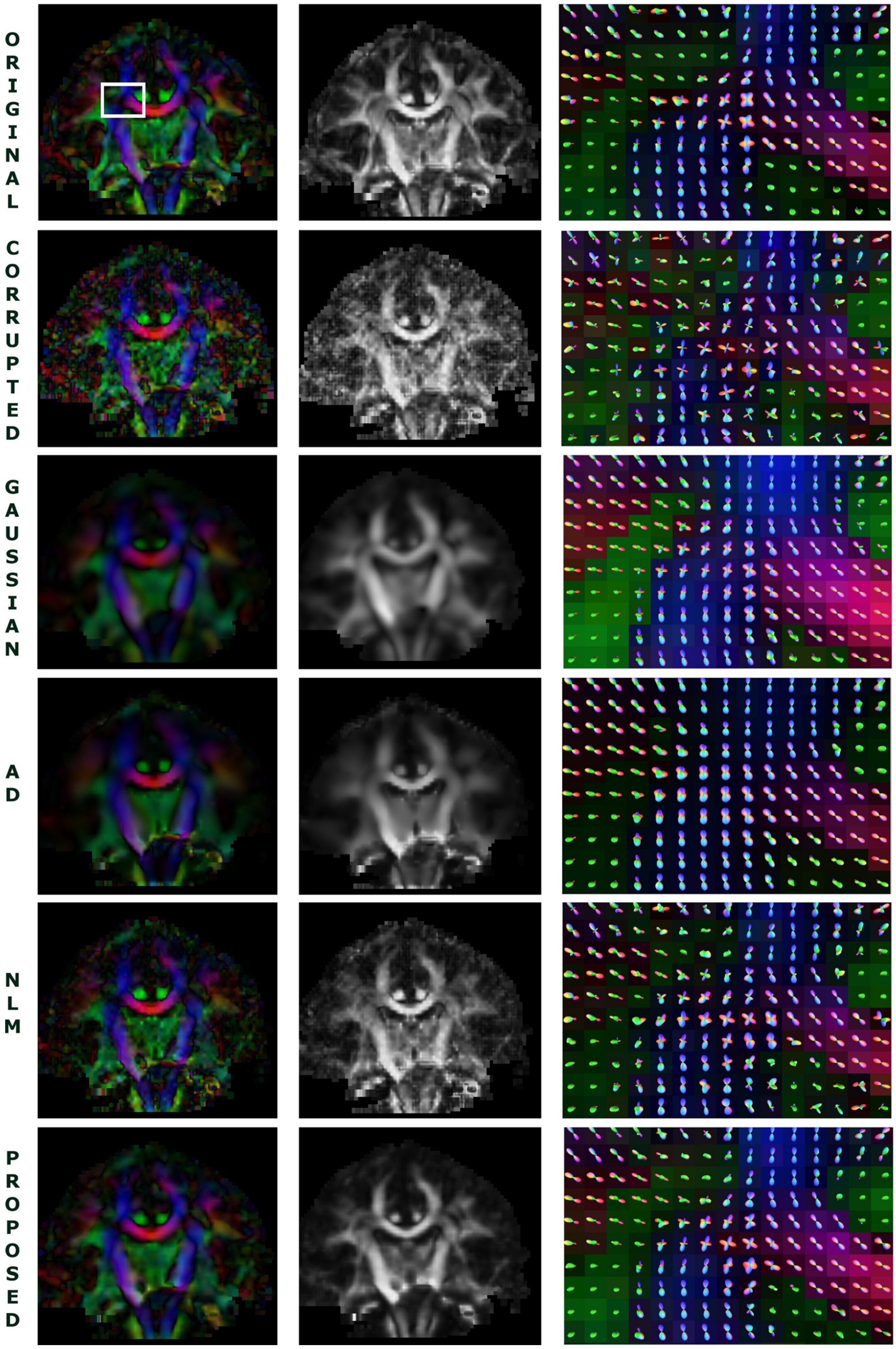

FIG. 8. Color-coded coronal slice of a diffusion weighted scan (left column), corresponding fractional anisotropy (middle) and zoom-in of fODF field around the crossing between the corona radiata and the radiation of the corpus callosum (right). The corrupted version has SNR = 10 dB. The white square on the uncorrupted, color-coded slice marks the ROI for which the fODF field is displayed. AD stands for anisotropic diffusion and NLM for nonlocal means.

denoising algorithms in the domain. Even though the presented methodology does not clearly outperform these approaches at typical SNR ratios, it is substantially superior at low SNR thanks to its prior knowledge on the data. This makes it possible to acquire the images using a fast (i.e., noisy) sequence and denoise them to acceptable noise levels. The denoised data were shown to be able to recover the tractography results from the uncorrupted version.

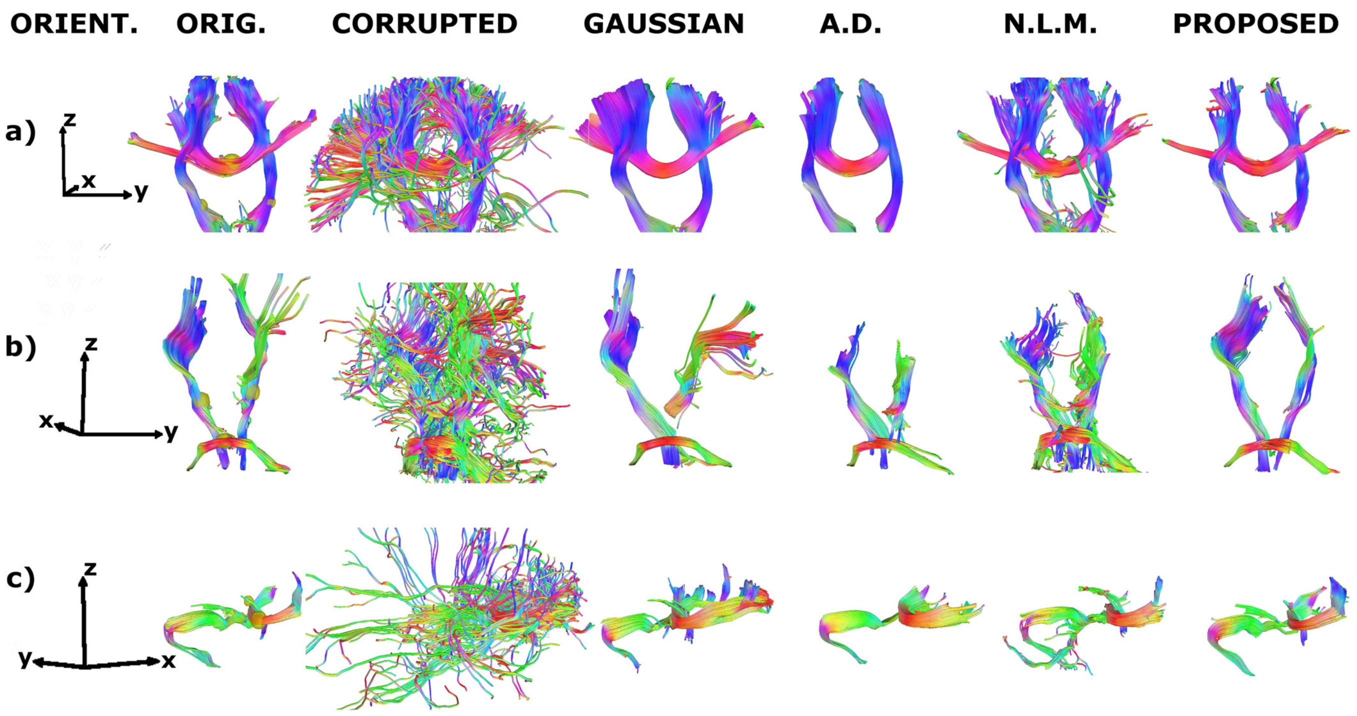

FIG. 9. 3D reconstruction of the fiber tracking output for the uncorrupted, corrupted (SNR = 10 dB) and denoised volumes of a sample test brain. The relative orientation of the reconstruction is displayed in the leftmost column: x represents posterior to anterior, y is right to left, and z is inferior to superior. The yellow spheres mark the seed regions. a) Corticospinal tract and corpus callosum b) Corticospinal tract and middle cerebellar peduncle. c) Corpus callosum, capsule fibers, and superior frontal gyrus fibers.

A possible improvement of the methodology would be to denoise the image by maximizing $p(\boldsymbol{x}(\boldsymbol{r}))$ rather than $p(\boldsymbol{x}(\boldsymbol{r}), \pi(\boldsymbol{r}))$ (Sec. III C). The problem is that the marginalization integral $p(\boldsymbol{x}(\boldsymbol{r})) = \int_{\mathbb{R}^C} p(\boldsymbol{x}(\boldsymbol{r}), \pi(\boldsymbol{r})) d\pi(\boldsymbol{r})$ cannot be calculated analytically, so it would require statistical sampling, which is computationally taxing. Moreover, optimizing the joint probability has the advantage that it also provides the optimal mixture weights for each voxel, which can be used as features for subsequent image analysis.

Another possible improvement would be use a robust metric to penalize the differences between neighboring mixture weights. The proposed MRF uses a quadratic penalty corresponding to a Gaussian distribution. This might not be the most suitable approach when there exist discontinuities (e.g., edges) in the data. A possibility would be to use a Laplace distribution or generalization thereof, which penalizes the absolute difference rather than the squared difference.[37] This could help reduce the (on the other hand minimal) blurring introduced by the method, which can potentially introduce spurious crossings. However, this would greatly complicate the analysis and the optimization process.

One potential limitation of the model is the lack of a proper ground truth, which is required in the training stage. This is a recurrent problem in DW-MRI. Phantoms consisting of separate fibers embedded in a homogeneous background are commonly used in the literature, but this is not a realistic approximation of the brain white matter. Elaborating more sophisticated phantom remains an open problem (see Refs. 59 and 60 for some recent efforts). Long image acquisition MRI sequences (24–48 h) can be used to obtain

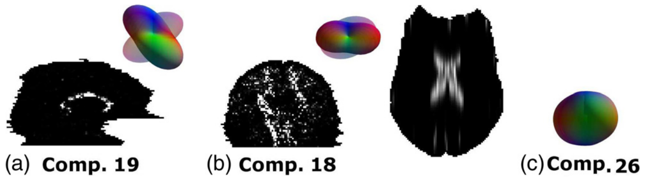
(a) Comp. 19 (b) Comp. 18 (c) Comp. 26

FIG. 10. Probability maps for different mixture components (for which the second mode of variation, often more informative than the first, is displayed). (a) Sample saggital slice that shows high probability for a left/right component around the corpus callosum. (b) Coronal slice for a probability volume that displays high probability for a vertical component around the corticospinal tract. (c) Axial slice of the probability volume of an isotropic component that shows high probability around the ventricles.

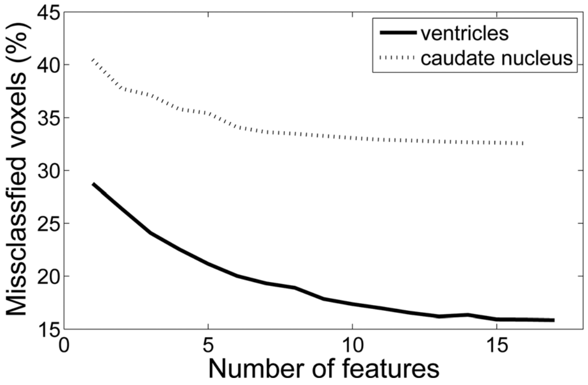

FIG. 11. Feature selection: error rate in cross-validation vs. number of selected features: (a) ventricles and (b) caudate nucleus.

TABLE II. Segmentation results and intra-user variability.

Structure	SI auto	SI intrauser
Ventricles	0.75±0.05	0.85 ± 0.05
Caudate nucleus	0.61±0.06	0.75 ± 0.06

very high SNR data, but they typically require ex-vivo data from deceased subjects to avoid motion artifacts. In this study, the lack of ground truth is circumvented by downsampling the images in the directional domain to obtain relatively clean data.

The proposed algorithm makes two important assumptions about noise. First, that the noise in the baseline T2 data is negligible compared to the noise in the gradient images. And second, that the noise power is known exactly. Both assumptions are however reasonable: due to multiple acquisitions and the lack of attenuation by diffusion, the baseline images have indeed much higher SNR than the gradient images. The noise parameter σ can be accurately estimated as the mode of the background of the image, thanks to the large amount of background voxels that are typically present in a 3D scan.

Regarding the ability to generalize to other datasets, retraining should not be necessary as long as the new data are acquired in the set of directions in the dataset presented here and at the same spatial resolution. If the resolution was different, the class variances could be modified to account for the difference in voxels size. If the gradient directions were different, the data could be resampled to the set of directions used here. Correction for different Le Bihan's constant is immediate. Evaluating the performance of the algorithm on a different dataset and exploring the impact of these adjustments when necessary remains as future work.

Another interesting aspect is the application of the algorithm to cases displaying pathology. In that case, the learned statistical distributions might not hold anymore. If a certain pathology is expected to be present in the test set, it should be possible to ensure that subjects with that particular disease are present in the training dataset. However, if there is no knowledge on potential conditions present in the test data, a very comprehensive training dataset might be necessary to obtain good results. Evaluating the method in these situations also remains as future work.

Finally, it is important to discuss the execution speed of the algorithm. All the described algorithms were implemented in JAVA and run in an Intel Core i7 desktop. The training stage takes 3 h, which is not a problem because it must be executed only once. Denoising an image takes approximately 25 min. This is longer than the execution time of the other approaches discussed here (Gaussian filtering, NLM, and AD). However, given that the code was not optimized for speed and that JAVA is an interpreted language, it should be possible to reduce the execution time to the order of ~ 10 min, which is completely acceptable given the dimensionality of the data (i.e., 128 × 128 × 128 voxels × 94 directions/voxel).

ACKNOWLEDGMENTS

This work was funded by the National Science Foundation (Grant No.0844566) and the National Institutes of Health through the NIH Roadmap for Medical Research, Grant No. U54 RR021813 entitled Center for Computational Biology (CCB). The authors would like to thank Greig de Zubicaray, Kathie McMahon, and Margaret Wright from the Center for Magnetic Resonance at the University of Queensland for acquiring the data, and professor Lieven Vandenberghe from UCLA for his insight on the optimization algorithm. The first author would also like to thank the U.S. Department of State's Fulbright program for the funding.

[a] Author to whom correspondence should be addressed. Electronic mail:jeiglesias@ucla.edu

[1] P. Basser, *et al.*, "Estimation of the effective self-diffusion tensor from the NMR spin echo," J. Magn. Reson., Ser. B **103**, 247–247 (1994).

[2] C. Pierpaoli, P. Jezzard, P. Basser, A. Barnett, and G. Di Chiro, "Diffusion tensor MR imaging of the human brain," Radiology **201**, 637–648 (1996).

[3] A. Alexander, K. Hasan, M. Lazar, J. Tsuruda, and D. Parker, "Analysis of partial volume effects in diffusion-tensor MRI," Magn. Reson. Med. **45**, 770–780 (2001).

[4] D. Tuch, T. Reese, M. Wiegell, N. Makris, J. Belliveau, and V. Wedeen, "High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity," Magn. Reson. Med. **48**, 577–582 (2002).

[5] L. Frank, "Anisotropy in high angular resolution diffusion-weighted MRI," Magn. Reson. Med. **45**, 935–939 (2001).

[6] E. Ozarslan and T. Mareci, "Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging," Magn. Reson. Med. **50**, 955–965 (2003).

[7] L. Frank, "Characterization of anisotropy in high angular resolution diffusion-weighted MRI," Magn. Reson. Med. **47**, 1083–1099 (2002).

[8] D. Alexander, G. Barker, and S. Arridge, "Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data." Magn. Reson. Med. **48**, 331–340 (2002).

[9] A. Barmpoutis, B. Jian, B. Vemuri, and T. Shepherd, "Symmetric Positive 4th Order Tensors& Their Estimation from Diffusion Weighted MRI," in MICCAI 2007, Lect. Notes Comput. Sci. **4584**, 308–319 (2007).

[10] A. Ghosh, M. Descoteaux, and R. Deriche, "Riemannian Framework for Estimating Symmetric Positive Definite 4th Order Diffusion Tensors," in MICCAI 2008, *Lecture Notes in Computer Science* (Springer, New York, 2008), Vol. 5241, pp. 858–865.

[11] T. Behrens, M. Woolrich, M. Jenkinson, H. Johansen-Berg, R. Nunes, S. Clare, P. Matthews, J. Brady, and S. Smith, "Characterization and propagation of uncertainty in diffusion-weighted MR imaging," Magn. Reson. Med. **50**, 1077–1088 (2003).

[12] J. Tournier, F. Calamante, D. Gadian, and A. Connelly, "Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution," Neuroimage **23**, 1176–1185 (2004).

[13] E. Özarslan, T. Shepherd, B. Vemuri, S. Blackband, and T. Mareci, "Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT)," Neuroimage **31**, 1086–1103 (2006).

[14] K. Jansons and D. Alexander, "Persistent angular structure: New insights from diffusion magnetic resonance imaging data," Inverse Probl. **19**, 1031–1046 (2003).

[15] D. Tuch, "Q-ball imaging," Magn. Reson. Med. **52**, 1358–1372 (2004).

[16] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche, "Regularized, fast, and robust analytical Q-ball imaging," Magn. Reson. Med. **58**, 497–510 (2007).

[17] A. Anderson, "Measurement of fiber orientation distributions using high angular resolution diffusion imaging," Magn. Reson. Med. **54**, 1194–1206 (2005).

[18] C. Hess, P. Mukherjee, E. Han, D. Xu, and D. Vigneron, "Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis," Magn. Reson. Med. **56**, 104–117 (2006).

[19] M. Descoteaux, R. Deriche, T. Knosche, and A. Anwander, "Deterministic and probabilistic tractography based on complex fibre orientation distributions," IEEE Trans. Med. Imaging **28**, 269–286 (2009).

[20] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," Int. J. Comput. Vis. **66**, 41–66 (2006).
[21] M. Descoteaux, N. Wiest-Daessle, S. Prima, C. Barillot, and R. Deriche, "Impact of rician adapted non-local means filtering on HARDI," in *MICCAI 2008, Lecture Notes in Computer Science* (Springer, New York, 2008), Vol. **5242**, pp. 122–129.
[22] L. He and I. Greenshields, "A Nonlocal maximum likelihood estimation method for rician noise reduction in MR images," IEEE Trans. Med. Imaging **28**, 165–172 (2009).
[23] A. Tristán-Vega and S. Aja-Fernández, "DWI filtering using joint information for DTI and HARDI," Med. Image Anal. **14**, 205–218 (2010).
[24] R. Wirestam, A. Bibic, J. Latt, S. Brockstedt, and F. Stahlberg, "Denoising of complex MRI data by wavelet-domain filtering: Application to high-b-value diffusion-weighted imaging," Magn. Reson. Med. **56**, 1114–1120 (2006).
[25] Y. Kim, P. Thompson, A. Toga, L. Vese, and L. Zhan, "HARDI denoising: Variational regularization of the spherical apparent diffusion coefficient sADC," in *IPMI 2009, Lect. Notes Comput. Sci.* **5636**, 515–527 (2009).
[26] T. McGraw, B. Vemuri, E. Ozarslan, Y. Chen, and T. Mareci, "Variational denoising of diffusion weighted MRI," Inverse Probl. Imaging **3**, 625–648 (2009).
[27] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Physica D **60**, 259–268 (1992).
[28] D. Tschumperle and R. Deriche, "Diffusion PDEs on vector-valued images," IEEE Signal Process. Mag. **19**, 16–25 (2002).
[29] B. Chen and E. Hsu, "Noise removal in magnetic resonance diffusion tensor imaging," Magn. Reson. Med. **54**, 393–407 (2005).
[30] T. McGraw, B. Vemuri, Y. Chen, M. Rao, and T. Mareci, "DT-MRI denoising and neuronal fiber tracking," Med. Image Anal. **8**, 95–111 (2004).
[31] Y. Assaf, R. Freidlin, G. Rohde, and P. Basser, "New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter," Magn. Reson. Med. **52**, 965–978 (2004).
[32] T. Behrens, H. Berg, S. Jbabdi, M. Rushworth, and M. Woolrich, "Probabilistic diffusion tractography with multiple fibre orientations: what can we gain?," Neuroimage **34**, 144–155 (2007).
[33] B. Jian, B. Vemuri, E. Özarslan, P. Carney, and T. Mareci, "A novel tensor distribution model for the diffusion-weighted MR signal," NeuroImage **37**, 164–176 (2007).
[34] D. Alexander, "Maximum entropy spherical deconvolution for diffusion MRI," Inf. Process. Med. Imaging, 76–87 (2005).
[35] J. Iglesias, P. Thompson, and Z. Tu, "A spatially variant mixture model for diffusion weighted mri: application to image denoising," in *MICCAI Workshop on Probabilistic Models for Medical Image Analysis*, 2009, pp. 103–114.
[36] C. Nikou, N. Galatsanos, and A. Likas, "A class-adaptive spatially variant mixture model for image segmentation," IEEE Trans. Image Process. **16**, 1121–1130 (2007).
[37] G. Sfikas, C. Nikou, N. Galatsanos, and C. Heinrich, "Spatially varying mixtures incorporating line processes for image segmentation," J. Math. Imaging Vision **36**, 91–110 (2010).
[38] M. Rivera, O. Ocegueda, and J. Marroquin, "Entropy-controlled quadratic Markov measure field models for efficient image segmentation," IEEE Trans. Image Process. **16**, 3047–3057 (2007).
[39] Z. Peng, W. Wee, and J. Lee, "Automatic segmentation of MR brain images using spatial-varying Gaussian mixture and Markov random field approach," in *CVPR Workshop*, CVPRW, 2006, pp. 80–87.
[40] M. Martin-Fernandez, C. Westin, and C. Alberola-Lopez, "3D Bayesian Regularization of Diffusion Tensor MRI Using Multivariate Gaussian Markov Random Fields," in *MICCAI 2004, Lect. Notes Comput. Sci.* **3216**, 351–359 (2004).
[41] M. King, D. Gadian, and C. Clark, "A random effects modelling approach to the crossing-fibre problem in tractography," Neuroimage **44**, 753–768 (2009).
[42] S. Aja-Fernandez, C. Alberola-Lopez, and C. Westin, "Signal LMMSE estimation from multiple samples in MRI and DT-MRI," MICCAI 2007, Lect. Notes Comput. Sci. **4792**, 368–375 (2007).
[43] S. Smith, "Fast robust automated brain extraction," Hum. Brain Mapp. **17**, 143–155 (2002).
[44] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche, "Apparent diffusion coefficients from high angular resolution diffusion imaging: Estimation and applications," Magn. Reson. Med. **56**, 395–410 (2006).
[45] D. Jones, M. Horsfield, and A. Simmons, "Optimal Strategies for Measuring Diffusion in Anisotropic Systems by Magnetic Resonance Imaging," Magn. Reson. Med. **42**, 515–525 (1999).
[46] E. Stejskal and J. Tanner, "Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient," J. Chem. Phys. **42**, 288–292 (1965).
[47] S. Sanjay-Gopal and T. Hebert, "Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm," IEEE Trans. Image Process. **7**, 1014–1028 (1998).
[48] J. Oliver, R. Baxter, and C. Wallace, "Unsupervised learning using mml," in *Proceedings of ICML 96* (Morgan Kaufmann Publishers, 1996), pp. 364–372.
[49] M. Figueiredo and A. Jain, "Unsupervised Learning of Finite Mixture Models," IEEE Trans. Pattern Anal. Mach. Intell. **24**, 381–396 (2002).
[50] I. Dinov, "Expectation maximization and mixture modeling tutorial," Statistics Online Computational Resource (2008), http://repositories,cdlib.org/socr/EM_MM.
[51] J. Besag, "On the statistical analysis of dirty pictures," J. R. Stat. Soc. **48**, 259–302 (1986).
[52] A. Zymnis, S. Kim, J. Skaf, M. Parente, and S. Boyd, "Hyperspectral image unmixing via alternating projected subgradients," in *Proceedings of ACSSC 2007*, pp. 1164–1168.
[53] R. Nowak, "Wavelet-based Rician noise removal for magnetic resonance imaging," IEEE Trans. Image Process. **8**, 1408–1419 (1999).
[54] J. Sijbers, D. Poot, A. Dekker, and W. Pintjens, "Automatic estimation of the noise variance from the histogram of a magnetic resonance image," Phys. Med. Biol. **52**, 1335–1348 (2007).
[55] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," IEEE Trans. Pattern Anal. Mach. Intell. **12**, 629–639 (1990).
[56] J. E. Iglesias, P. M. Thompson, C. Y. Liu and Z. Tu, "Fast Approximate Stochastic Tractography," Neuroinformatics (accepted for publication).
[57] S. Stearns, "On selecting features for pattern classifiers," in *Proceedings of the 3rd International Joint Conference on Pattern Recognition* (Coronado, CA, 1976), pp. 71–75.
[58] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de Solorzano, "Combination strategies in multi-atlas image segmentation: Application to brain MR Data," IEEE Trans. Med. Imaging **28**, 1266–1277 (2009).
[59] G. Balls and L. Frank, "A simulation environment for diffusion weighted MR experiments in complex media.," Magn. Reson. Med. **62**, 771–778 (2009).
[60] T. Close, J. Tournier, F. Calamante, L. Johnston, I. Mareels, and A. Connelly, "A software tool to generate simulated white matter structures for the assessment of fibre-tracking algorithms," Neuroimage **47**, 1288–1300 (2009).