Variational Approaches to the Estimation, Regularization and Segmentation of Diffusion Tensor Images

Rachid Deriche, David Tschumperlé, Christophe Lenglet, Mikaël Rousson

Variational Approaches to the Estimation, Regularization and Segmentation of Diffusion Tensor Images

R. Deriche, D. Tschumperlé, C. Lenglet and M. Rousson

ABSTRACT Diffusion magnetic resonance imaging probes and quantifies the anisotropic diffusion of water molecules in biological tissues, making it possible to non-invasively infer the architecture of the underlying structures. In this chapter, we present a set of new techniques for the robust estimation and regularization of diffusion tensor images (DTI) as well as a novel statistical framework for the segmentation of cerebral white matter structures from this type of dataset. Numerical experiments conducted on real diffusion weighted MRI illustrate the techniques and exhibit promising results.

1 Introduction

Diffusion magnetic resonance imaging is a relatively new modality [20] that acquires, at each voxel, data allowing the reconstruction of a probability density function characterizing the average motion of water molecules. As of today, it is the only non-invasive method that allows to distinguish the anatomical structures of the cerebral white matter. Well-known examples are the corpus callosum, the arcuate fasciculus or the corona radiata. These are commissural, associative and projective neural pathways, the three main types of fiber bundles, respectively connecting the two hemispheres, regions of a given hemisphere or the cerebral cortex with subcortical areas. Diffusion MRI is particularly relevant to a wide range of clinical applications related to pathologies such as acute brain ischemia, stroke, Alzheimer's disease or schizophrenia. It is also extremely useful in order to identify the neural connectivity patterns of the human brain [21] and references therein.

In 1994, Basser et al [3] proposed to model the probability density function (*pdf*) of the three-dimensional molecular motion r, at each voxel of a diffusion MR image, by a Gaussian distribution whose covariance matrix is given by the diffusion tensor. Diffusion tensor imaging (DTI) thus produces a vo-

lumic image containing, at each voxel, a 3×3 symmetric positive-definite matrix. The estimation of these tensors requires the acquisition of diffusion weighted images in several non-collinear sampling directions as well as a $T2$-weighted image. Numerous algorithms have been proposed to perform a robust estimation and regularization of these tensors fields [41], [43], [49], [24], [48], [46], [47], [11], [42], [4], [9], [25],[34]. Among all these works, it is worth pointing out that [47] was the first to use the original Stejskal-Tanner equation, and not the linearized form, in the data term. The authors showed the importance of this model and relied on the Cholesky decomposition to estimate the symmetric, positive-definite tensors. In sections 2 and 3, we will tackle the estimation and regularization tasks within a common variational framework while taking into account the symmetry and positive definiteness constraints.
Moreover, it is well-known that normal brain functions require specific cortical regions to communicate through fiber pathways. Based on DTI, most of the existing techniques addressing the issue of the anatomical connectivity mapping work on a fiber-wise basis. In other words, they do not take into account the global coherence that exists among fibers of a given tract. Recent work by Corouge et al [10] has proposed to cluster and align fibers by local shape parameterization so that a statistical analysis of the tract geometrical and physiological properties can be carried out. This work relies on the extraction of a set of streamlines from diffusion tensor images by the method proposed in [27] which is known to be sensitive to noise and unreliable in areas of fibers crossings.
For these reasons, we propose, in section 4, to directly perform the segmentation of diffusion tensor images in order to extract neural fibers bundles. Contrary to the methods proposed in [51], [50], [14],[45], [44] and [18], our approach is grounded on the expression of statistics in the space of multivariate Gaussian distributions [37], [22], [23]. We use this information in a level-set and region-based framework to evolve a surface while maximizing the likelihood of the region to extract. The central point in the developments of section 4 will be the choice of the probability metric, e.g. the dissimilarity measure used to compare any two probability density functions.

2 Estimation of Diffusion Tensor Images

2.1 Data acquisition

Our dataset consists of 30 diffusion weighted images $S_k : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R},\ k = 1, ..., 30$ as well as a single image S_0 corresponding to the signal intensity in the absence of any diffusion-sensitizing gradient. They were obtained on a GE 1.5 T Signa Echospeed with standard 22 mT/m gradient field. The echoplanar images were acquired on 56 evenly spaced axial planes with

128×128 pixels in each slice. Voxel size is $1.875\ mm \times 1.875\ mm \times 2.8\ mm$. 6 gradient directions $\mathbf{g}_k$, each with 5 different b-factors and 4 repetitions were used. Imaging parameters were: b-factors between 0 and 1000 $s.mm^{-2}$, $TR = 2.5\ s$, $TE = 84.4\ ms$ and a square field of view of 24 cm [32][1].

2.2 Linear estimation

We recall that the estimation of a field of 3×3 symmetric positive definite matrices $\mathbf{D}$ is performed by using the Stejskal-Tanner equation 1.1 [40] for anisotropic diffusion.

$$S_k(x) = S_0(x) \exp\left(-b\mathbf{g}_k^T \mathbf{D}(x)\mathbf{g}_k\right) \quad \forall x \in \Omega \tag{1.1}$$

where $\mathbf{g}_k$ are the normalized non-collinear gradient directions and b the diffusion weighting factor. Many approaches have been derived to estimate the tensor field $\mathbf{D}$.
If we effectively restrict ourselves to 6 gradient directions, Westin et al. derived in [49] a compact analytical solution to equation 1.1 and, by doing so, eliminated the need to solve it for every single data point. The idea relies on the introduction of a dual tensor basis $\tilde{\mathbf{B}}_k$, computed from the tensor basis $\mathbf{B}_k = \mathbf{g}_k\mathbf{g}_k^T$, and which can be used to decompose any given tensor $\mathbf{D}(x)$. We then end up with the closed-form solution:

$$\mathbf{D} = \sum_{k=1}^{6} \frac{1}{b} \ln\left(\frac{S_0}{S_k}\right) \tilde{\mathbf{B}}_k \tag{1.2}$$

This method turns out to be sensitive to noise and easily influenced by potential outliers. This is due to the low number of measurements intrinsically used by this approach and by the choice of the minimization function (see [24] where the Geman-McLure M-estimator is used in order to reduce outlier-related artifacts). Moreover resulting tensors may not be positive definite, which requires a subsequent reprojection step [42].

2.3 Variational estimation

In order to deal with a more complete estimation approach, we propose to incorporate some important priors such as tensor positivity and regularity into a variational formulation of the estimation problem by minimizing the following energy on the manifold of real 3×3 symmetric positive-definite matrices $S^+(3, \mathbb{R})$:

$$\underset{\mathbf{D}(x) \in S^+(3,\mathbb{R})}{\operatorname{Arg\,min}} \int_\Omega \sum_{k=1}^{n} \psi\left(\left|\ln\left(\frac{S_0(x)}{S_k(x)}\right) - b\mathbf{g}_k^T \mathbf{D}(x)\mathbf{g}_k\right|\right) + \alpha\rho(|\nabla \mathbf{D}(x)|)dx \tag{1.3}$$

[1] Data courtesy of J.F. Mangin and J.B Poline, CEA/SHFJ, Orsay, France

where ψ controls the robust estimation and the Lagrange multiplier α, together with the scalar function ρ, drives the anisotropic regularity of the solution. Minimizing this criterion, in the constrained tensor space, leads to the following evolution equation:

$$\begin{cases} \mathbf{D}_{(t=0)} = \mathbf{Id} \\ \\ \frac{\partial \mathbf{D}}{\partial t} = (\mathbf{G} + \mathbf{G}^T)\mathbf{D}^2 + \mathbf{D}^2(\mathbf{G} + \mathbf{G}^T) \end{cases}$$

where $\mathbf{G}$ corresponds to the gradient of the unconstrained criterion defined as $G_{ij} = \sum_{k=1}^{n} \psi^{'}(|v_k|)\text{sign}(v_k)\left(\mathbf{g}_k\mathbf{g}_k^T\right)_{ij} + \alpha \text{div}\left(\frac{\rho^{'}(|\nabla \mathbf{D}|)}{|\nabla \mathbf{D}|}\nabla D_{ij}\right)$ with $v_k = \ln(S_0/S_k) - b\mathbf{g}_k^T\mathbf{D}\mathbf{g}_k$.
Note that if $\psi(v) = v^2$ and $\alpha = 0$, the criterion reduces to a simple multilinear regression by least square that generalizes the linear estimation method of Westin et al [49] and provides a positive definite solution since the minimization is done in the constrained space $S^+(3,\mathbb{R})$. This variational method converges to a much more consistent solution thanks to its global behavior. Concerning the implementation part, a carefully designed numerical scheme, based on manifold integration, to ensure that the estimate stays on $S^+(3,\mathbb{R})$ at each step of the gradient descent, is used to solve the associated Euler-Lagrange equations:

$$\mathbf{D}_{(t+dt)} = \mathbf{A}^T\mathbf{D}_{(t)}\mathbf{A} \quad \text{with} \quad \mathbf{A} = \exp\left(\mathbf{D}_{(t)}(\mathbf{G} + \mathbf{G}^T)dt\right)$$

Our iterative method starts from a field of isotropic tensors that are evolving in $S^+(3,\mathbb{R})$ and are morphing until their shapes fit the measured data S_0, S_k. Enforcing the positiveness and regularity constraints has a large interest for DTI estimation, and leads to more accurate results than with classical methods. For more details, we refer the interested readers to the article [42].

3 Regularization of Diffusion Tensor Images

The variational estimation method naturally brings some spatial coherence and smoothness into the generated tensor field. However, the fundamental properties of diffusion tensors, like diffusivities and principal orientations, are contained in their spectral features. It can then be interesting to regularize the tensor field with regard to those spectral elements. This will bring more coherence into the tensor structural information and thus improve any subsequent processing such as the tracking of neural fibers.

3.1 On some non-spectral methods and their limitations

Non-spectral methods are based on a direct anisotropic smoothing of the diffusion weighted data S_k or consider each tensor as 6 independent scalar

components $\mathbf{D}(x)_{ij}$ (by symmetry) with possible coupling. We thus evolve each $\mathbf{D}(x)_{ij}$ by minimizing the following quantity:

$$E(\mathbf{D}) = \int_\Omega \frac{\alpha}{2}|\mathbf{D}(x) - \mathbf{D}_0(x)|^2 + \rho(|\nabla \mathbf{D}(x)|)dx \tag{1.4}$$

where $\mathbf{D}_0$ designates the initial noisy tensor field and the field gradient norm $|\nabla \mathbf{D}|$ behaves as a coupling term between the tensors components. However, eigenvalues tend to diffuse faster than eigenvectors, resulting in a *swelling* effect on the tensors.

Spectral methods separately consider the eigen-elements of the tensors. Eigenvalues smoothing is typically performed by a vector-valued anisotropic PDE ([38] and references therein) satisfying the maximum principle in order to preserve the positiveness. The three orthonormal eigenvectors define a matrix of $O(3)$ which can be regularized by acting only on the principal eigenvector $\mathbf{u}^1$ and then reconstructing the associated tensor [11]. The field of orthonormal matrices can also be evolved under a scheme preserving the eigenvectors norms and angles [41]. This boils down to solving a system of coupled and constrained PDEs. However, all these approaches require a time-consuming step of eigenvectors realignment since a given vector and its opposite are both solution of the same singular value decomposition and thus yield artificially discontinuous vectors fields.

3.2 A fast isospectral method

In [9], we proposed an efficient alternative to the previous spectral techniques, which does not require any spectral decomposition, by building flows acting on a given submanifold of the linear set of matrix-valued functions and preserving some constraints. We showed that this amounts to characterizing the velocity of the flows (ie. the tangent space of the submanifold) at each point of that submanifold. Actually, the relevant constraints (orthogonality, eigenvalues conservation ...) can be expressed by simply working with the proper Lie group or homogeneous space. For example, an isospectral flow acts on a field of real symmetric matrices and preserves their eigenvalues. Moreover, its velocity is directly derived from the matrices field gradient, hence no need for realignment. If $[X, Y]$ denotes the Lie bracket of X and Y, e.g. $XY - YX$, the general form for our isospectral flow is given by:

$$\frac{\partial \mathbf{D}}{\partial t} = [\mathbf{D}, [\mathbf{D}, (\mathbf{G} + \mathbf{G}^T)]] \tag{1.5}$$

where $\mathbf{G}$ prescribes the desired regularization process, such as

$$G_{ij} = \mathrm{div}\left(\frac{\rho'(|\nabla \mathbf{D}|)}{|\nabla \mathbf{D}|}\nabla D_{ij}\right)$$

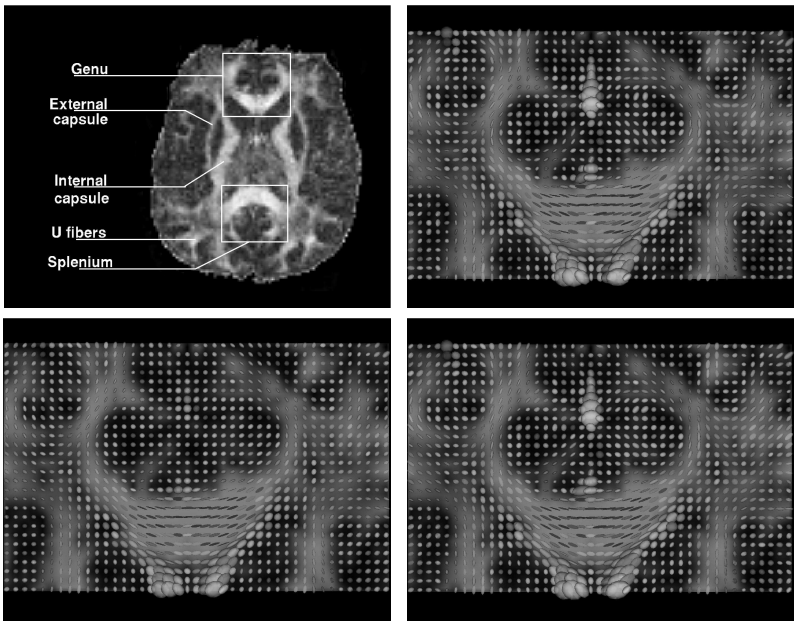

FIGURE 1. DTI regularization in the genu of the corpus callosum ([TOP LEFT]: Annotated fractional anisotropy axial slice, [TOP RIGHT]: Original tensors, [BOTTOM LEFT]: Non-spectral regularization, [BOTTOM RIGHT]: Isospectral flow)

ρ denotes the same scalar function as in section 2.3 and preserves important structures of the tensor field. A specific reprojection-free scheme based on the exponential map can also be used to implement the PDE (1.5):

$$\mathbf{D}_{(t+dt)} = \mathbf{A}^T \mathbf{D}_{(t)} \mathbf{A} \quad \text{with} \quad \mathbf{A} = \exp\left(dt[\mathbf{G} + \mathbf{G}^T, \mathbf{D}_{(t)}]\right)$$

Results of non-spectral smoothing and isospectral flow on diffusion tensors estimated in the genu of the corpus callosum are presented in figure 1.

4 Segmentation of Diffusion Tensor Images

The previous sections described algorithms for the estimation and the regularization of diffusion tensor images. We now focus on the segmentation of these tensor-valued images, seen as fields of Gaussian probability density functions. We first set up the level-set and region-based surface evolution framework that will be used throughout this section. We then progressively introduce the various statistical parameters associated with the probability metrics derived from the Euclidean distance, the Kullback-Leibler divergence and finally, the geodesic distance between probability density functions.

4.1 Level-set and region-based surface evolution

Our ultimate goal is to compute the optimal 3D surface separating an anatomical structure of interest from the rest of a diffusion tensor image. The region-based front evolution, as developed in [35], is an efficient and well-suited framework for our segmentation problem. We hereafter summarize the basic notions of this technique.

Let Γ be the optimal boundary between the object to extract Ω_1 and the background Ω_2. We introduce the level-set [12], [13] and [28] function $\phi : \Omega \rightarrow \mathbb{R}$, defined as follows:

$$\begin{cases} \phi(x) = 0, & \text{if } x \in \Gamma \\ \phi(x) = \mathcal{D}_{Eucl}(x, \Gamma), & \text{if } x \in \Omega_1 \\ \phi(x) = -\mathcal{D}_{Eucl}(x, \Gamma), & \text{if } x \in \Omega_2 \end{cases}$$

where $\mathcal{D}_{Eucl}(x, \Gamma)$ stands for the Euclidean distance between x and Γ and $\Omega = \Omega_1 \cup \Omega_2$. Furthermore, let $H_\epsilon(.)$ and $\delta_\epsilon(.)$ be regularized versions of the Heaviside and Dirac functions as defined in [8].
Let $q(x, r)$ be the probability density function of our random vector r of $\mathbb{R}^3$ describing the water molecules average motion at a given voxel x of a DTI dataset. We also denote by p_1 and p_2 the probability distributions of the *pdf*s $q(x, .)$ respectively in Ω_1 or Ω_2. Then, according to the Geodesic Active Regions model [29], and by adding a regularity constraint on the interface, the optimal partitioning of Ω in two regions Ω_1 and Ω_2 is obtained by minimizing:

$$\begin{aligned} E(\phi, p_1, p_2) = \nu \int_\Omega |\nabla H_\epsilon(\phi)| dx - \int_\Omega H_\epsilon(\phi) \log p_1(q(x, .)) dx \\ - \int_\Omega (1 - H_\epsilon(\phi)) \log p_2(q(x, .)) dx \end{aligned} \tag{1.6}$$

We have reached the point where we need to express p_1 and p_2, e.g. the probability distributions in the space of probability density functions $q(., r)$. This is the purpose of the next sections.

4.2 Multivariate Gaussian distributions as a linear space

When dealing with diffusion tensor images, we recall that the molecular motion is assumed to follow a Gaussian law of zero mean. The diffusion tensor can indeed be interpreted as the covariance matrix of the underlying Brownian motion. As proposed in [37], we start by considering the parameters space of three-dimensional Gaussian *pdf*s $q(., r)$ as linear, which boils down to reducing a diffusion tensor image to a vector-valued volume, each voxel being assigned with the 6-dimensional vector of the variances and covariances, and the probability metric being Euclidean.

Let $u(x)$ be the vector representation of a tensor $\mathbf{D}(x)$, the probability distributions of $u(x)$ in the regions $s = 1, 2$ are defined as:

$$p_s(u|\overline{u}_s, \Lambda_s) = \frac{1}{(2\pi)^3|\Lambda_s|^{1/2}} e^{-\frac{1}{2}(u-\overline{u}_s)^T \Lambda_s^{-1}(u-\overline{u}_s)}$$

The Euclidean mean vectors $\overline{u}_s$ and covariance matrices Λ_s have to be estimated. They can simply be introduced as unknown in (1.6) and optimized for during the front evolution process. Our objective function 1.6 then becomes:

$$\begin{aligned} E(\phi, \{\overline{u}_{1,2}, \Lambda_{1,2}\}) = \nu \int_\Omega |\nabla H_\epsilon(\phi)| dx - \int_\Omega H_\epsilon(\phi) \log p_1(u(x)|\overline{u}_1, \Lambda_1) dx \\ - \int_\Omega (1 - H_\epsilon(\phi)) \log p_2(u(x)|\overline{u}_2, \Lambda_2) dx \end{aligned}$$

This type of energy was studied in [35], [36], the Euler-Lagrange equations for ϕ yield the following evolution equation for the level-set function $\phi(x)$ $\forall x \in \Omega$:

$$\begin{aligned} \phi_t(x) = \delta_\epsilon(\phi) \Big(\nu \operatorname{div} \frac{\nabla \phi}{|\nabla \phi|} + \frac{1}{2} \log \frac{|\Lambda_2|}{|\Lambda_1|} - \frac{1}{2}(u(x) - \overline{u}_1)^T \Lambda_1^{-1} (u(x) - \overline{u}_1) \\ + \frac{1}{2}(u(x) - \overline{u}_2)^T \Lambda_2^{-1} (u(x) - \overline{u}_2) \Big) \end{aligned}$$

while it can be shown that the statistical parameters must be updated by their empirical estimates [37]. Adequate implementation schemes for this type of optimization can be found in [8]. If we restrict the covariance matrices to the identity, these equations simplify and the likelihoods in equation (1.6) simply become the Euclidean distance between the vectors u and $\overline{u}_{s=1,2}$, which is equivalent to the Frobenius norm of the difference between the corresponding tensors, as studied in [45].
Figure 2 illustrates this method on a synthetic dataset where the Y-shape region to be segmented only differs from the background by the orientation of its tensors. A crossing area with low fractional anisotropy was created and Gaussian noise was separately added on the eigenvalues and eigenvectors to stress the algorithm.

Motivated by the method proposed by Wang and Vemuri in [44], we now derive the statistics and the associated evolution equation based on a more natural and widely used measure of dissimilarity between *pdf*s, known as the Kullback-Leibler divergence or relative entropy.

4.3 Information-theoric statistics between distributions

We will show that this approach is not only more natural, in the sense that it is strongly rooted and used in the information theory community, but

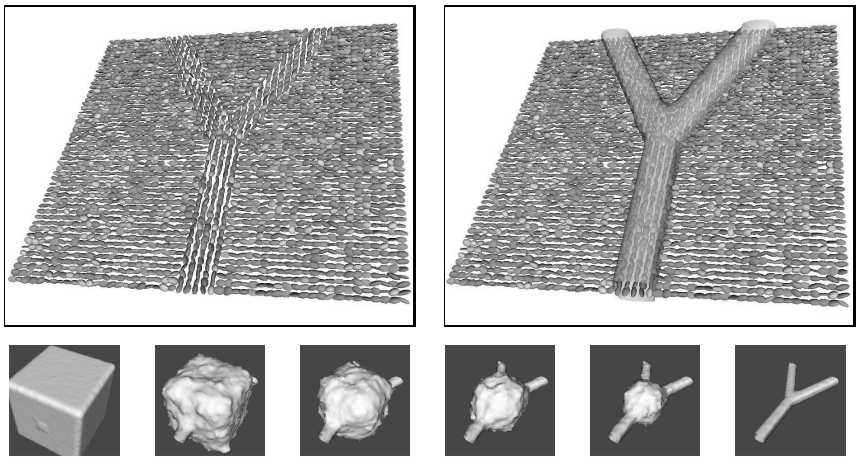

FIGURE 2. Segmentation (with Euclidean probability metric) of a noisy tensor field composed by two regions with same scales but different orientations ([TOP LEFT]: 2D-cut of the tensor field, [TOP RIGHT]: Final segmentation, [BOTTOM]: Surface evolution)

also more versatile since it enables the segmentation algorithm to work on fields of Gaussian densities as well as on non-parametric densities [22].
We consider a general probability density function $q(x,r)$ of the random vector r of $\mathbb{R}^3$. The symmetrized Kullback-Leibler divergence can be used to express the dissimilarity between diffusion processes at different locations of Ω. With $q(x,.), q(y,.)\ \forall x,y \in \Omega$ two probability density functions from $\mathbb{R}^3$ onto $\mathbb{R}^+$, their symmetrized Kullback-Leibler divergence is given by:

$$\mathcal{D}_{kl}(q(x,.),q(y,.)) = \frac{1}{2}\int_{\mathbb{R}^3}\left(q(x,r)\log\frac{q(x,r)}{q(y,r)} + q(y,r)\log\frac{q(y,r)}{q(x,r)}\right)dr \tag{1.7}$$

We denote by $\overline{q}_1$ and $\overline{q}_2$ the mean probability density functions over Ω_1 and Ω_2 verifying equation 1.10. In this section, we make the assumption that the *pdf*s in Ω_1 and Ω_2 have respective Gaussian distributions p_1^{kl}, p_2^{kl} with means $\overline{q}_1, \overline{q}_2$ and variances σ_1^2, σ_2^2:

$$p_{s=1,2}^{kl}(q|\overline{q}_s,\sigma_s^2) = \frac{1}{\sqrt{2\pi\sigma_s^2}}\exp\frac{-\mathcal{D}_{kl}^2(q,\overline{q}_s)}{2\sigma_s^2}$$

We can then rewrite our objective function 1.6 as follows:

$$\begin{aligned} E(\phi,\{\overline{q}_{1,2},\sigma_{1,2}^2\}) = \nu\int_\Omega |\nabla H_\epsilon(\phi)|dx - \int_\Omega H_\epsilon(\phi)\log p_1^{kl}(q(x)|\overline{q}_1,\sigma_1^2)dx \\ - \int_\Omega (1-H_\epsilon(\phi))\log p_2^{kl}(q(x)|\overline{q}_2,\sigma_2^2)dx \end{aligned} \tag{1.8}$$

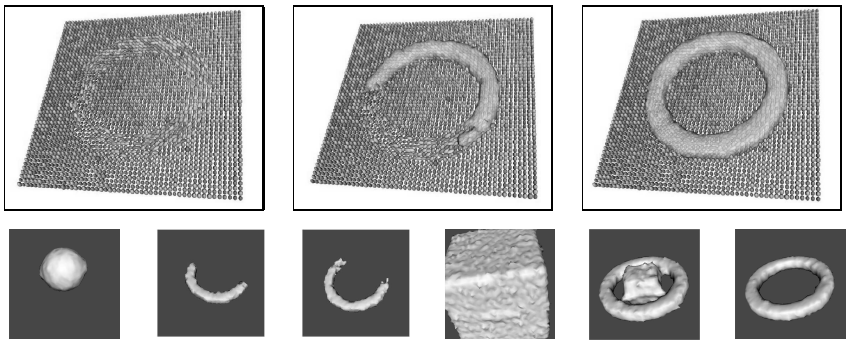

FIGURE 3. Segmentation of a noisy tensor field composed by two regions with same scales but different orientations ([TOP LEFT]: 2D-cut of the tensor field, [TOP CENTER]: Segmentation obtained from [44], [TOP RIGHT]: Segmentation based on the Kullback-Leibler probability metric and associated statistics, [BOTTOM:] Respective surface evolutions)

In the case where the σ_s^2 are equal to 1, this energy is equivalent to the one proposed in [44]. As for the Euclidean probability metric, the Euler-Lagrange equations yield the following evolution equation:

$$\phi_t(x) = \delta_\epsilon(\phi)\left(\nu \mathrm{div}\frac{\nabla\phi}{|\nabla\phi|} + \log\frac{p_2^{kl}(q(x)|\overline{q}_2,\sigma_2^2)}{p_1^{kl}(q(x)|\overline{q}_1,\sigma_1^2)}\right) \tag{1.9}$$

Moreover, it can be shown that the variance must be updated by its empirical estimation with respect to the Kullback-Leibler divergence, whereas some more work is needed for $\overline{q}_s$, defined as:

$$\overline{q}_{s=1,2} = \underset{q_s}{\mathrm{Arg\,min}}\ \frac{1}{|\Omega_s|}\int_{\Omega_s} \mathcal{D}_{kl}^2(q(x),q_s)dx \tag{1.10}$$

Indeed, for a general probability density function $q(.,r)$, the variance is easily computed as in [36] but the estimation of the $\overline{q}_s$ might require the use of numerical approximation techniques if no closed-form expression is available. It turns out that, for Gaussian *pdf*s $q(.,r)$, the energy 1.8 simplifies as follows:

$$\begin{aligned} &E(\phi,\{\overline{q}_{1,2},\sigma_{1,2}^2\}) = \\ &\nu\int_\Omega |\nabla H_\epsilon(\phi)|dx + \frac{1}{2}\int_\Omega H_\epsilon(\phi)(\log(2\pi\sigma_1^2) + \mathcal{D}_{kl}^2(q(x),\overline{q}_1)\sigma_1^{-2})dx \\ &+\frac{1}{2}\int_\Omega (1-H_\epsilon(\phi))(\log(2\pi\sigma_2^2) + \mathcal{D}_{kl}^2(q(x),\overline{q}_2)\sigma_2^{-2})dx \end{aligned} \tag{1.11}$$

Using the closed-form expressions provided in [44] for the symmetrized Kullback-Leibler divergence between two Gaussian *pdf*s and for the associated mean density $\overline{q}_s$ parameterized by the mean diffusion tensor $\overline{\mathbf{D}}_s$,

the Euler-Lagrange equations for our energy yield (the dependence on x is omitted for the sake of clarity):

$$\phi_t = \delta_\epsilon(\phi)\left(\nu \mathrm{div}\frac{\nabla\phi}{|\nabla\phi|} + \frac{1}{2}\left(\frac{3(\sigma_1^2-\sigma_2^2)}{2(\sigma_1^2\sigma_2^2)} + \log\frac{\sigma_2^2}{\sigma_1^2}\right) + \frac{1}{8}\left(\mathrm{tr}\left(\mathbf{D}^{-1}\overline{\mathbf{D}}_2 + \overline{\mathbf{D}}_2^{-1}\mathbf{D}\right)\sigma_2^{-2} - \mathrm{tr}\left(\mathbf{D}^{-1}\overline{\mathbf{D}}_1 + \overline{\mathbf{D}}_1^{-1}\mathbf{D}\right)\sigma_1^{-2}\right)\right) \tag{1.12}$$

Notice that we obtain additional terms (the σ_s^2 coefficients) in equation 1.12 if compared to the Euler-Lagrange equations proposed in [44]. Figure 3 illustrates the importance of the variance in our model.

The symmetrized Kullback-Leibler divergence, although it does not satisfy the triangle inequality, has many useful properties and is widely used to measure dissimilarities between *pdf*s. However, for particular densities like multivariate Gaussian distributions of fixed mean, better probability metrics are available. In the next section, we show how a Riemannian metric can be associated with the 6-dimensional parameters space of these densities using the Fisher information matrix. The geodesic distance, intrinsic mean and covariance matrix of multivariate Gaussian distributions, as well as curvature information, can be efficiently computed to yield a generalized Gaussian distribution of multivariate Gaussian densities. This generalized distribution can then be used in our segmentation framework.

4.4 A Riemannian approach to DTI segmentation

We now consider the Riemannian manifold $\mathcal{M}$ of the family of three-dimensional Gaussian probability density functions parameterized by the 6 components of their covariance matrix Σ (in other words, the diffusion tensor $\mathbf{D}$). Following the work by Rao [33] and Burbea-Rao [6], where a Riemannian metric was introduced in term of the Fisher information matrix, we wish to define the notion of geodesic distance and intrinsic statistics on this 6-dimensional manifold whose coordinate system, in some local chart, is given by a real vector parameter $\theta = (\theta_1, ..., \theta_6) \in \mathbb{R}^6$ such that for all random vector $r \in \mathbb{R}^3$, $\mathcal{M} = \{q(r|\theta),\ \theta \in \mathbb{R}^6\}$. In the following, we first show the main limitation of the Kullback-Leibler divergence together with its impact on the segmentation process. Then, we present the closed-form expression of the geodesic distance as well as original computational methods to approximate a generalized Gaussian distribution of multivariate Gaussian densities with common mean.

The Fisher information matrix: The manifold $(\mathcal{M}, g)$ equipped with the Fisher information matrix $g = g_{ij},\ i,j = 1, ..., 6$ has the structure of a Riemannian manifold [33], [39] when g is non-degenerate. We recall that g

is defined as follows:

$$g_{ij} = \int_{\mathbb{R}^3} \frac{\partial \log q(r|\theta)}{\partial \theta_i} \frac{\partial \log q(r|\theta)}{\partial \theta_j} q(r|\theta) dr \tag{1.13}$$

By plugging the definition of a Gaussian *pdf* into equation 1.13, the 6×6 metric tensor, as presented in [23], can be expressed in terms of the parameters θ_i, $i = 1, ..., 6$ used to describe the *pdf*s. Thus, instead of considering the parameterized *pdf*s as living in the linear space $\mathbb{R}^6$, we do take into account the Riemannian structure of the underlying manifold. Moreover, the Kullback-Leibler divergence $\mathcal{D}_{kl}$ turns out to be a Taylor approximation of the geodesic distance between two nearby distributions $q(r|\theta)$ and $q(r|\theta + d\theta)$, given suitable technical conditions. Indeed, as summarized in [2], it can be shown that:

$$\mathcal{D}_{kl}(\theta, \theta + d\theta) = \frac{1}{2} \mathbb{E} \left[\frac{\partial \log q(r|\theta)}{\partial \theta_i} \frac{\partial \log q(r|\theta)}{\partial \theta_j} \right] d\theta_i d\theta_j$$

This means that the infinitesimal squared geodesic distance $g_\theta(d\theta, d\theta)$ is twice the Kullback-Leibler divergence (this is also true for its symmetrized form). In other words, the method presented in the previous section assumes that we always compute distances between nearby elements of $\mathcal{M}$, which, in general, does not hold. For general *pdf*s, we may have no other choice but, in the more particular case of multivariate Gaussian densities with common mean, a closed-form solution of the geodesic distance is available, thus allowing the comparison of any two of these distributions. We now introduce this geodesic distance and derive the associated intrinsic statistical parameters.

Geodesic distance and intrinsic statistics: We recall that $S^+(m, \mathbb{R})$ denotes the set of $m \times m$ real symmetric positive-definite matrices Σ (here $m = 3$). A detailed study on the definition of a statistical model on this nonlinear space was presented by the authors in [23]. Another recent work by Pennec et al [31] relies on a comparable approach to derive tensor fields filtering techniques. Following [23], [39], [5], [7], [16], [26] and [15], $S^+(m, \mathbb{R})$ can be characterized as an affine symmetric space for which the geodesic distance $\mathcal{D}_g$ between any two elements Σ_1 and Σ_2 was derived by Jensen.

Theorem 4.1 *(S.T. Jensen, 1976 [1])*
Consider the family of multivariate Gaussian distributions with common mean vector but different covariance matrices. The geodesic distance between two members of the family with covariance matrices Σ_1 and Σ_2 is:

$$\mathcal{D}_g(\Sigma_1, \Sigma_2) = \sqrt{\mathrm{tr}(\log^2(\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2}))} = \sqrt{\sum_{i=1}^{m} \log^2(\lambda_i)}$$

where the λ_i are the roots of the determinantal equation $|\lambda \Sigma_1 - \Sigma_2| = 0$.

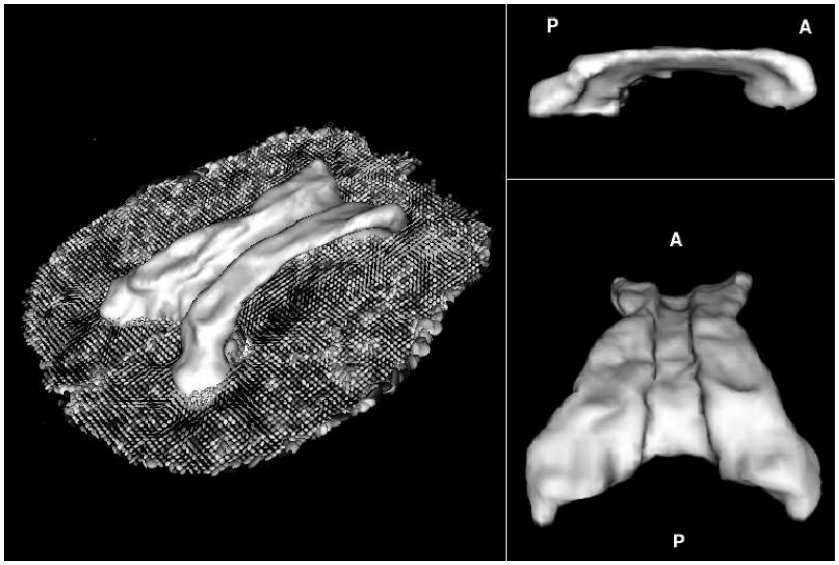

FIGURE 4. Segmentation of the corpus callosum with the Riemannian probability metric ([LEFT]: 3D view with an axial slice of diffusion tensors, [RIGHT]: A: Anterior, P: Posterior)

We now explain how to estimate the empirical mean, as proposed by Fréchet [17], Karcher [19] and Pennec [30], as well as the empirical covariance matrix.

Definition 4.1 *The Gaussian distribution parameterized by* $\overline{\Sigma} \in S^+(m,\mathbb{R})$ *and defined as the empirical mean of* N *distributions* Σ_k, $k = 1,...,N$, *achieves a local minimum of the function* $\sigma^2 : S^+(m,\mathbb{R}) \rightarrow \mathbb{R}^+$ *known as the empirical variance and defined as:*

$$\sigma^2(\Sigma_1,...,\Sigma_N) = \frac{1}{N}\sum_{k=1}^{N} \mathcal{D}_g^2(\Sigma_k,\overline{\Sigma}) = \mathbb{E}[\mathcal{D}_g^2(\Sigma_k,\overline{\Sigma})]$$

Karcher proved in [19] that such a mean exists and is unique for manifolds of non-positive sectional curvature. This was shown to be the case for $S^+(m,\mathbb{R})$ in [39]. A closed-form expression of the mean cannot be obtained [26] but a gradient descent algorithm was proposed in [23]. A flow is derived from an initial guess $\overline{\Sigma}_0$ toward the mean of a subset of $S^+(m,\mathbb{R})$. The following evolution was obtained:

$$\overline{\Sigma}_{t+1} = \overline{\Sigma}_t^{1/2} \exp(-\frac{1}{N}\overline{\Sigma}_t^{1/2}\sum_{k=1}^{N} \log(\Sigma_k^{-1}\overline{\Sigma}_t)\overline{\Sigma}_t^{-1/2})\overline{\Sigma}_t^{1/2} \tag{1.14}$$

The empirical covariance matrix Λ^g relative to the mean $\overline{\Sigma}$ is defined as:

Definition 4.2 *Given N elements of $S^+(m,\mathbb{R})$ and a mean value $\overline{\Sigma}$, the empirical covariance matrix relative to $\overline{\Sigma}$ is defined as:*

$$\Lambda^g = \frac{1}{N}\sum_{k=1}^{N} \beta_k \beta_k^T$$

where $\beta_k = \overline{\Sigma}\log(\Sigma_k^{-1}\overline{\Sigma})$ is the gradient of the squared geodesic distance $\nabla \mathcal{D}_g^2(\Sigma_k, \overline{\Sigma})$ in vector form.

Finally, as detailed in [23], the Ricci curvature tensor $\mathcal{R}$ can be computed at the mean $\overline{\Sigma}$. Putting everything together and following Theorem 4 of [30], we have:

Theorem 4.2 *The generalized Gaussian distribution in $S^+(m,\mathbb{R})$ for a covariance matrix Λ^g of small variance $\sigma^2 = tr(\Lambda^g)$ is of the form:*

$$p^g(\Sigma|\overline{\Sigma},\Lambda^g) = \frac{1+O(\sigma^3)+\epsilon(\sigma/\xi)}{\sqrt{(2\pi)^{m(m+1)/2}|\Lambda^g|}} \exp\frac{-\beta^T\gamma\beta}{2} \quad \forall \Sigma \in S^+(m,\mathbb{R})$$

where $\beta = \overline{\Sigma}\log(\Sigma^{-1}\overline{\Sigma})$ is expressed in vector form and the concentration matrix is $\gamma = (\Lambda^g)^{-1} - \mathcal{R}/3 + O(\sigma) + \epsilon(\sigma/\xi)$. ξ is the injection radius at $\overline{\Sigma}$ and ϵ is such that $\lim_{0^+} x^{-\beta}\epsilon(x) = 0\ \forall \beta \in \mathbb{R}^+$.

Implementation: We can use the very same variational framework as the one described in section 4.2 in order to maximize the likelihoods of the diffusion tensors distributions in Ω_1 and Ω_2. This can now be achieved with respect to the geodesic distance by using $p^g_{s=1,2}(\Sigma|\overline{\Sigma}_s, \Lambda^g_s)$ and by accordingly evolving the level-set function ϕ toward the optimal segmentation. Figure 4 illustrates how well this approach performs on a real diffusion tensor image.

5 Conclusion

Diffusion magnetic resonance imaging gives a direct insight into the microstructure of biological tissues through the measurement of hindered molecular motion. In this chapter, we have described efficient and versatile numerical methods for the estimation and the regularization of the diffusion tensor images. We have also presented a novel statistical and geometric approach to the segmentation of DTI data. The central point of this front evolution framework relies on the definition of dissimilarity measures and statistics between diffusion tensors, seen as the covariance matrices of Gaussian probability density functions. The major contribution of this set of techniques is related to the robust extraction of anatomical structures in the brain white matter.

6 References

[1] C. Atkinson and A. Mitchell. Rao's distance measure. *Sankhya: The Indian Journal of Statistics*, 43(A):345–365, 1981.

[2] V. Balasubramanian. A geometric formulation of Occam's razor for inference of parametric distributions. PUPT 1588, Princeton University, December 2001.

[3] P. Basser, J. Mattiello, and D. Le Bihan. MR diffusion tensor spectroscopy and imaging. *Biophysica*, 66:259–267, 1994.

[4] T. Brox, J. Weickert, B. Burgeth, and P. Mrázek. Nonlinear structure tensors. Technical Report 113, Department of Mathematics, Saarland University, Saarbrücken, Germany, October 2004.

[5] J. Burbea. Informative geometry of probability spaces. *Expositiones Mathematica*, 4:347–378, 1986.

[6] J. Burbea and C. Rao. Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. *Journal of Multivariate Analysis*, 12:575–596, 1982.

[7] M. Calvo and J. Oller. An explicit solution of information geodesic equations for the multivariate normal model. *Statistics and Decisions*, 9:119–138, 1991.

[8] T. Chan and L. Vese. An active contour model without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001. Also in Scale-Space Theories in Computer Vision, 141-151, 1999.

[9] C. Chefd'hotel, D. Tschumperlé, R. Deriche, and O. Faugeras. Regularizing flows for constrained matrix-valued images. *Journal of Mathematical Imaging and Vision*, 20(1-2):147–162, 2004.

[10] I. Corouge, S. Gouttard, and G. Gerig. A statistical shape model of individual fiber tracts extracted from diffusion tensor MRI. In *MICCAI*, pages 671–679, 2004.

[11] O. Coulon, D. Alexander, and S. Arridge. Diffusion tensor magnetic resonance image regularisation. *Medical Image Analysis*, 8(1):47–67, 2004.

[12] A. Dervieux and F. Thomasset. A finite element method for the simulation of Rayleigh-Taylor instability. *Lecture Notes in Mathematics*, 771:145–159, 1979.

[13] A. Dervieux and F. Thomasset. Multifluid incompressible flows by a finite element method. In *International Conference on Numerical Methods in Fluid Dynamics*, pages 158–163, 1980.

[14] C. Feddern, J. Weickert, B. Burgeth, and M. Welk. Curvature-driven PDE methods for matrix-valued images. Technical Report 104, Department of Mathematics, Saarland University, Saarbrücken, Germany, April 2004.

[15] P. Fletcher and S. Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *CVAMIA-MMBIA*, pages 87–98, 2004. ECCV'04 workshop.

[16] W. Förstner and B. Moonen. A metric for covariance matrices. Technical report, Dept. of Geodesy and Geoinformatics, Stuttgart University, 1999.

[17] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut H. Poincaré*, 10(4):215–310, 1948.

[18] L. Jonasson, X. Bresson, P. Hagmann, O. Cuisenaire, R. Meuli, and J. Thiran. White matter fiber tract segmentation in DT-MRI using geometric flows. *Medical Image Analysis*, 2004. In press.

[19] H. Karcher. Riemannian centre of mass and mollifier smoothing. *Comm. Pure Appl. Math*, 30:509–541, 1977.

[20] D. Le Bihan, E. Breton, D. Lallemand, P. Grenier, E. Cabanis, and M. Laval-Jeantet. MR imaging of intravoxel incoherent motions: Application to diffusion and perfusion in neurologic disorders. *Radiology*, 161:401–407, 1986.

[21] C. Lenglet, R. Deriche, and O. Faugeras. Inferring white matter geometry from diffusion tensor MRI: Application to connectivity mapping. In *ECCV*, pages 127–140, 2004.

[22] C. Lenglet, M. Rousson, and R. Deriche. Segmentation of 3D probability density fields by surface evolution: Application to diffusion MRI. In *MICCAI*, pages 18–25, 2004.

[23] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on multivariate normal distributions: A geometric approach and its application to diffusion tensor MRI. Research Report 5442, INRIA, Sophia Antipolis, June 2004.

[24] J. Mangin, C. Poupon, C. Clark, D. Le Bihan, and I. Bloch. Distortion correction and robust tensor estimation for MR diffusion imaging. *Medical Image Analysis*, 6(3):191–198, 2002.

[25] T. McGraw, B. Vemuri, Y. Chen, M. Rao, and T. Mareci. DT-MRI denoising and neuronal fiber tracking. *Medical Image Analysis*, 8:95–111, 2004.

[26] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 2004. In press.

[27] S. Mori, B. Crain, V. Chacko, and P. Van Zijl. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology*, 45(2):265–269, 1999.

[28] S. Osher and J. Sethian. Fronts propagating with curvature dependent speed: Algorithms based on the Hamilton-Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.

[29] N. Paragios and R. Deriche. Geodesic Active Regions: A new paradigm to deal with frame partition problems in computer vision. *Journal of Visual Comm. and Image Representation*, 13(1-2):249–268, 2002.

[30] X. Pennec. Probabilities and statistics on Riemannian manifolds: A geometric approach. Research Report 5093, INRIA, Sophia Antipolis, January 2004.

[31] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. Research Report 5255, INRIA, Sophia Antipolis, July 2004.

[32] C. Poupon. *Détection des faisceaux de fibres de la substance blanche pour l'étude de la connectivité anatomique cérébrale.* PhD thesis, Ecole Nationale Supérieure de Télécommunications de Paris, 1999.

[33] C. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81–91, 1945.

[34] M. Rodríguez Florido, C.-F. Westin, and J. Ruiz-Alzola. Dt-mri regularization using anisotropic tensor field filtering. In *IEEE 2004 International Symposium on Biomedical Imaging*, pages 15–18, Arlington, VA, EEUU, Apr. 2004.

[35] M. Rousson. *Cues integrations and front evolutions in image segmentation.* PhD thesis, Université de Nice-Sophia Antipolis, 2004.

[36] M. Rousson and R. Deriche. A variational framework for active and adaptative segmentation of vector valued images. In *IEEE Workshop on Motion and Video Computing*, pages 56–62, 2002.

[37] M. Rousson, C. Lenglet, and R. Deriche. Level set and region based surface propagation for diffusion tensor MRI segmentation. In *CVAMIA-MMBIA*, pages 87–98, 2004. ECCV'04 workshop.

[38] G. Sapiro. *Geometric Partial Differential Equations and Image Analysis.* Cambridge University Press, 2001.

[39] L. Skovgaard. A Riemannian geometry of the multivariate normal model. *Scandinavian Journal of Statistics*, 11:211–233, 1984.

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

[41] D. Tschumperlé and R. Deriche. Diffusion tensor regularization with constraints preservation. In *CVPR*, pages 948–954, 2001.

[42] D. Tschumperlé and R. Deriche. Variational frameworks for DT-MRI estimation, regularization and visualization. In *ICCV*, pages 116–122, 2003.

[43] B. Vemuri, Y. Chen, M. Rao, T. McGraw, Z. Wang, and T. Mareci. Fiber tract mapping in the CNS using DT-MRI. In *VLSM*, pages 81–88, 2001.

[44] Z. Wang and B. Vemuri. An affine invariant tensor dissimilarity measure and its application to tensor-valued image segmentation. In *CVPR*, pages 228–233, 2004.

[45] Z. Wang and B. Vemuri. Tensor field segmentation using region based active contour model. In *ECCV*, pages 304–315, 2004.

[46] Z. Wang, B. Vemuri, Y. Chen, and T. Mareci. Simultaneous smoothing and estimation of the tensor field from diffusion tensor MRI. In *CVPR*, pages 461–466, 2003.

[47] Z. Wang, B. Vemuri, Y. Chen, and T. Mareci. A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex DWI. *IEEE Transactions on Medical Imaging*, 23(8):930–939, 2004.

[48] J. Weickert and T. Brox. Diffusion and regularization of vector- and matrix-valued images. *Inverse Problems, Image Analysis, and Medical Imaging. Contemporary Mathematics*, 313:251–268, 2002.

[49] C. Westin, S. Maier, H. Mamata, A. Nabavi, F. Jolesz, and R. Kikinis. Processing and visualization for diffusion tensor MRI. *Medical Image Analysis*, 6(2):93–108, 2002.

[50] M. Wiegell, D. Tuch, H. Larson, and V. Wedeen. Automatic segmentation of thalamic nuclei from diffusion tensor magnetic resonance imaging. *Neuroimage*, 19:391–402, 2003.

[51] L. Zhukov, K. Museth, D. Breen, R. Whitaker, and A. Barr. Level set segmentation and modeling of DT-MRI human brain data. *Journal of Electronic Imaging*, 12(1):2003, 125-133.