INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

DTI Segmentation by Statistical Surface Evolution

Christophe Lenglet — Mikaël Rousson — Rachid Deriche

N° 5843

February 20, 2006

Thème BIO

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--5843--FR+ENG

DTI Segmentation by Statistical Surface Evolution

Christophe Lenglet* , Mikaël Rousson† , Rachid Deriche‡

Thème BIO — Systèmes biologiques
Projet Odyssée

Rapport de recherche n° 5843 — February 20, 2006 —45 pages

Abstract: We address the problem of the segmentation of cerebral white matter structures from diffusion tensor images (DTI). DTI produces, from a set of diffusion-weighted MR images, tensor-valued images where each voxel is assigned with a 3×3 symmetric, positive-definite matrix. This second order tensor is simply the covariance matrix of a local Gaussian process, with zero mean, modeling the average motion of water molecules. As we will show in this article, the definition of a dissimilarity measure and statistics between such quantities is a non trivial task which must be tackled carefully. We claim and demonstrate that, by using the theoretically well-founded differential geometrical properties of the manifold of multivariate normal distributions, it is possible to improve the quality of the segmentation results obtained with other dissimilarity measures such as the Euclidean distance or the Kullback-Leibler divergence. The main goal of this work is to prove that the choice of the probability metric, i.e. the dissimilarity measure, has a deep impact on the tensor statistics and, hence, on the achieved results. We introduce a variational formulation, in the level-set framework, to estimate the optimal segmentation of a diffusion tensor image according to the following hypothesis: Diffusion tensors exhibit a Gaussian distribution in the different partitions. We must also respect the geometric constraints imposed by the interfaces existing among the cerebral structures and detected by the gradient of the diffusion tensor image. We show how to express all the statistical quantities for the different probability metrics. We validate and compare the results obtained on various synthetic data-sets, a biological rat spinal cord phantom and human brain DT images.

Key-words: Diffusion tensor MRI, Segmentation, Probability metric, Riemannian Geometry, Information geometry, Fisher information matrix, Kullback-Leibler divergence, Level set

For any citation of this work, please refer to the journal version accepted in
IEEE Transactions on Medical Imaging, 2006 (In press)

* INRIA, 2004 route des lucioles, 06902 Sophia-Antipolis, FRANCE
† Siemens Corporate Research, 755 College Road East, Princeton, NJ 08540, USA
‡ INRIA, 2004 route des lucioles, 06902 Sophia-Antipolis, FRANCE

Segmentation d'Images du Tenseur de Diffusion par Évolution Statistique de Surface

Résumé : Nous considérons le problème de la segmentation des structures de la matière blanche cérébrale à partir d'images du tenseur de diffusion (DTI). Le DTI fournit, depuis un ensemble d'images IRM pondérées en diffusion, des images à valeur tensorielle où chaque voxel recoit une matrice symétrique définie positive. Ce tenseur d'ordre 2 n'est autre que la matrice de covariance d'un processus Gaussien local, de moyenne nulle, modélisant le mouvement des molécules d'eau. Comme nous le montrerons dans cet article, la définition d'une mesure de dissimilarité et de statistiques entre de telles quantités est une tâche difficile. Nous affirmons et montrons que, grâce aux propriétés différentielles géométriques de l'espace des lois normales multivariées, il est possible d'améliorer la qualité des segmentations obtenues avec d'autres mesures de dissimilarité telles que la distance Euclidienne ou la divergence de Kullback-Leibler. Notre but est de prouver que le choix de cette mesure de dissimilarité a d'importantes conséquences sur les statistiques entre tenseurs et donc sur les résultats obtenus. Nous introduisons une formulation variationnelle, basée sur la méthode des ensembles de niveaux, afin d'estimer la segmentation optimale d'une image de tenseurs de diffusion selon les hypothèses suivantes: Les tenseurs suivent une distribution Gaussienne dans les différentes partitions. Nous devons aussi respecter les contraintes géométriques imposées par les interfaces entre structures cérébrales et détectées par le gradient de l'image de tenseurs de diffusion. Nous montrons comment exprimer les statistiques pour les différentes mesures de dissimilarité. Nous validons et comparons les résultats obtenus sur divers jeux de données synthétiques, un fantôme biologique réalisé à partir de moelles épinières de rats et des images de cerveaux humains.

Mots-clés : IRM du Tenseur de Diffusion, Segmentation, Distance entre densités de probabilité, Géométrie Riemannienne, Géométrie de l'information, Matrice de l'information de Fisher, Divergence de Kullback-Leibler, Ensembles de niveaux

Contents

1 Introduction

Diffusion magnetic resonance imaging is a relatively new modality [7], [42] able to quantify the anisotropic diffusion of water molecules in highly structured biological tissues. As of today, it is the only non-invasive method that allows to distinguish the anatomical structures of the cerebral white matter such as the corpus callosum, the superior longitudinal fasciculus or the corticospinal tract. These are examples of commissural, association and projection neural pathways, the three main types of fiber bundles, respectively connecting the two hemispheres, regions of a given hemisphere or the cerebral cortex with the basal ganglia and the cerebellum.
In 1994, Basser et al.[5] proposed to model the local probability density function of the three-dimensional molecular motion by a Gaussian distribution whose covariance matrix is given by the diffusion tensor. The estimation of these tensors requires the acquisition of diffusion weighted images in different sampling directions. Numerous algorithms have been proposed to perform a robust estimation and regularization of these tensors fields [65], [41], [64], [60], [63], [19], [58], [8], [17], and [38].
Diffusion MRI is particularly relevant to a wide range of clinical investigations related, for example, to brain ischemia detection [56], stroke, Alzheimer disease, or schizophrenia [1]. It is also extremely useful in order to identify the neural connectivity patterns of the human brain [44], [6], [12], [35].

Most of the existing techniques addressing this last issue 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 [18] 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 [44] which is known to be sensible to noise and unreliable in areas of fiber crossings.
For these reasons, we propose to directly perform the segmentation of diffusion tensor images in order to extract neural fiber bundles. While many techniques have been proposed to classify the gray matter, white matter and cephalo-spinal fluid from T1-weighted MR images (see [69] for example), the literature addressing the segmentation of white matter structures from DTI is still new. We hereafter draw a quick state of the art of the diffusion tensor images segmentation problem:
Zhukov et al. [71] defined an invariant anisotropy measure in order to drive the evolution of a level-set and isolate strongly anisotropic regions of the brain. The reduction of the full tensor to a single scalar value can result in a relatively low discrimination capability, potentially yielding the segmentation of mixed structures. Alternatively, Wiegell et al. [66], Feddern et al. [24, 25], Rousson et al. [54], Wang et al. [61] and [62], Lenglet et al. [36] and Jonasson et al. [31] use or propose different measures of dissimilarity between diffusion tensors. In [66], [61] and [54], the authors use the Frobenius norm of the difference of tensors (i.e. the Euclidean distance). A k-means algorithm with a spatial coherence constraint and an active contour model with a regularity term were respectively used by the first two methods ([66], [61]) to perform the segmentation of different cerebral structures such as the thalamus nuclei or the corpus callosum. The third method [54] used a region-based surface propagation. In [61], a generalization of the region-based active contours

to matrix-valued images is proposed. However, it is restricted to the two-dimensional case and obviously limited when it comes to three-dimensional brain data. In [24, 25], partial differential equations based on mean curvature motion, self-snakes and geodesic active contour models are extended to two-dimensional and three-dimensional tensor-valued images by generalizing the notion of structure tensor to matrix-valued data. This method still relies on the Euclidean metric between tensors. The authors apply this framework to the regularization and segmentation of diffusion tensor images. In [31], the authors introduce a geometric measure of dissimilarity by computing the normalized tensor 'scalar product' of two tensors, which can be interpreted as a measure of overlap. Finally, the methods exposed in [62] and [36] rely on the symmetrized Kullback-Leibler divergence to derive an affine invariant dissimilarity measure between diffusion tensors.

Contributions:
Our contributions are threefold: First, we recast the DTI segmentation problem into a unified statistical surface evolution framework. We also make use of the tensor field gradient to detect boundaries between various structures of the white matter. This framework can be implemented with different probability metrics. This is done for the Euclidean distance, Kullback-Leibler divergence and geodesic distance on the manifold of multivariate normal distributions. The second contribution is related to the development of a rigorous differential geometrical framework, as presented in [39], rooted in the information geometry and used to express a Gaussian law between diffusion tensors. We overcome the classical hypothesis considering covariance matrices as a linear space and define relevant statistics to model the distribution of diffusion tensors. To that end, we also extend the methods proposed in [62] and [36] by showing how to compute the covariance matrix, associated to the Kullback-Leibler divergence, of a set of tensors. Finally, we demonstrate that the properties of the geodesic distance lead to its superiority, for our segmentation task, over the other two dissimilarity measures. This is achieved by presenting results on both synthetic and real data-sets as well as on a biological phantom, for which only this method succeeds by comparison with the ground truth or neuroanatomical knowledge.

Organization of the paper:
Section 2 describes how to approximate a Gaussian distribution between diffusion tensors, in other words how to compute a mean tensor and a 6×6 covariance matrix. It also presents how to evaluate the norm of a tensor field spatial gradient, needed for the implementation of the boundary term. These three quantities are derived for the three dissimilarity measures of interest. Section 3 sets up the Bayesian formulation of the segmentation problem that will be used throughout this paper. Section 4 presents and discusses experimental results on synthetic data-sets, a biological phantom and human brain DT images .

2 Statistics and Gradient of Diffusion Tensor Fields

We would like to define the notions of Gaussian distribution between diffusion tensors as well as the norm of a diffusion tensor image spatial gradient. We denote such an image by $\Sigma : \Omega \mapsto S^+(3)$ so that for all $x \in \Omega$, $\Sigma(x)$ is a diffusion tensor belonging to $S^+(3)$, the space of 3×3 real, sym-

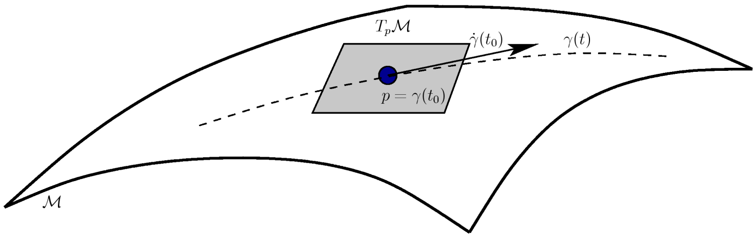

Figure 1: Tangent space at p of a Riemannian manifold $\mathcal{M}$

metric, positive-definite matrices. Ω is a bounded and regular region of interest, i.e. the acquisition grid which is a subset of $\mathbb{R}^3$.

We now introduce a few concepts from differential geometry needed for the following. As we will see in the last part of this section, it is indeed natural to consider $S^+(3)$ as a differentiable manifold. In effect, it is a 6-dimensional submanifold of $\mathbb{R}^6$ which can be endowed with a Riemannian metric. This general characterization of $S^+(3)$ will be very useful to derive statistics on diffusion tensors based on different probability metrics.

2.1 Riemannian geometry basics

2.1.1 Metric, geodesics, distance

For an n-dimensional manifold $\mathcal{M}$, a Riemannian metric is a collection of inner products $\langle .,. \rangle_p$ defined for every point p of $\mathcal{M}$. These inner products are defined on the tangent space $T_p\mathcal{M}$ of $\mathcal{M}$ at p and provide a natural way to measure the lengths of vectors tangent to $\mathcal{M}$ at location p. We call tangent vector an element of a tangent space which is simply a vector space (a copy of $\mathbb{R}^n$) attached to each point $p \in \mathcal{M}$. A good example of tangent vector at p is the case of the derivative $\dot\gamma(t_0) = \frac{d\gamma(t_0)}{dt}$ of a curve $\gamma : I = [t_0, t_1] \subset \mathbb{R} \mapsto \mathcal{M}$ passing through $\gamma(t_0) = p$ (figure 1).
It is possible to introduce a map $\varphi : \mathcal{M} \mapsto U \subset \mathbb{R}^n$, known as a coordinate chart, that defines a *local coordinate system* $\varphi(p) = \mathbf{x} = (\mathbf{x}^1, \cdots, \mathbf{x}^n)^T$ and a basis of the tangent space $T_p\mathcal{M}$ denoted by $\left(\frac{\partial}{\partial \mathbf{x}_1}, \cdots, \frac{\partial}{\partial \mathbf{x}_n}\right)$. Any element of the tangent space can hence be expressed in the form $\sum_i \mathbf{x}^i \frac{\partial}{\partial \mathbf{x}_i}$ and the inner products $\langle \frac{\partial}{\partial \mathbf{x}_i}, \frac{\partial}{\partial \mathbf{x}_j} \rangle_p$ define an $n \times n$ symmetric, bilinear and positive-definite form $G = g_{ij}$ known as the *local representation of the Riemannian metric*. The inner product of two tangent vectors u and v is then expressed as

$$\langle u, v \rangle_p = u^T G v$$

(the reference to the location p is usually discarded in notation g_{ij}).

Equipped with these notions we can now define the concept of *geodesic* on a Riemannian manifold $\mathcal{M}$. It is the equivalent of straight line in Euclidean spaces and defined as the locally length-minimizing curve $\gamma : I \subset \mathbb{R} \mapsto \mathcal{M}$. The tangent vector $\dot\gamma(t)$ defines the instantaneous speed of the curve and its norm $|\dot\gamma(t)| = \langle \dot\gamma(t), \dot\gamma(t) \rangle^{1/2}_{\gamma(t)}$ is the instantaneous velocity. Integrating $|\dot\gamma(t)|$ along

γ yields its length which is also the geodesic distance between the two endpoints p_1 and p_2 of the curve:

$$\mathcal{D}(p_1, p_2) = \int_{t_1}^{t_2} |\dot{\gamma}(t)| dt$$

Finally, taking $I = [0, 1]$ for simplicity, it is possible to show, under certain assumptions that will be met in the following, that a geodesic $\gamma : [0, 1] \mapsto \mathcal{M}$ is uniquely defined by its starting point $\gamma(0)$ and its initial velocity $\dot{\gamma}(0) \in T_{\gamma(0)}\mathcal{M}$. The endpoint $\gamma(1)$ can be easily computed by applying the *exponential map* at $\gamma(0)$ to $\dot{\gamma}(0)$: $\gamma(1) = \exp_{\gamma(0)}(\dot{\gamma}(0))$. A detailed presentation of this map can be found in [22]. The inverse map, known as the *logarithm map* of $\gamma(1)$ at $\gamma(0)$: $\log_{\gamma(0)}(\gamma(1)) = \exp^{-1}_{\gamma(0)}(\dot{\gamma}(0))$, yields the unique tangent vector $\dot{\gamma}(0)$ if we know the two endpoints of the curve. Moreover, it can be proved that

$$\mathcal{D}(\gamma(0), \gamma(1)) = \langle \dot{\gamma}(0), \dot{\gamma}(0) \rangle_{\gamma(0)}^{1/2}$$

In this paper, we will use the fact that the velocity $\dot{\gamma}(0)$ can be computed from the gradient of the squared geodesic distance with respect to $\gamma(0)$. In other words, we have

$$\dot{\gamma}(0) = -\nabla_{\gamma(0)} \mathcal{D}^2(\gamma(0), \gamma(1))$$

Using this definition, we can now define the notions of mean and covariance matrix on a Riemannian manifold. They will play a central role in the variational formulation (equations (19) and (21)) of the segmentation problem to be detailed in the section 3. We also show how to compute the norm of the spatial gradient of a tensor field which will be useful to introduce a boundary term in our segmentation energy (equation (20)).

2.1.2 Statistics

As defined by Fréchet in [28] and used by Pennec in [49], the empirical mean of a set of N random elements $\{p_i\},\ i = 1, ..., N$ of $\mathcal{M}$, such as diffusion tensors, is defined as the minimizer $p = \overline{p}$ of the variance $\sigma_p^2(\{p_i\})$ of the p_i with respect to p:

$$\sigma_p^2(\{p_i\}) = \mathbb{E}\left[\mathcal{D}^2(p, p_i)\right] = \frac{1}{N}\sum_{i=1}^{N} \mathcal{D}^2(p, p_i) \tag{1}$$

The empirical covariance matrix of the set $\{p_i\}$, with respect to the mean $\overline{p}$ is defined as the expected value of the quantity $\beta_i.\beta_i^T$ and denoted by $\Lambda_{\overline{p}}$. As depicted on figure 2, β_i is the initial velocity $\dot{\gamma_i}(0)$ of the i^{th} geodesic joining $\gamma_i(0) = \overline{p}$ to p_i and expressed in local coordinates, i.e. it is taken to be the n-dimensional vector of coordinates $\varphi(\dot{\gamma_i}(0))$ and not the tangent vector $\dot{\gamma_i}(0)$ itself. The dot product then boils down to a simple Euclidean dot product and we have:

$$\Lambda_{\overline{p}} = \frac{1}{N}\sum_{i=1}^{N} \varphi(\beta_i)\, \varphi(\beta_i)^T \text{ with } \beta_i = -\nabla_{\overline{p}} \mathcal{D}^2(\overline{p}, p_i) \tag{2}$$

where φ is the coordinate chart introduced in section 2.1.1 and also used in sections 2.2.1, 2.2.2 and 2.2.3. In section 2.2, we will apply these definitions to the Euclidean, Kullback-Leibler and

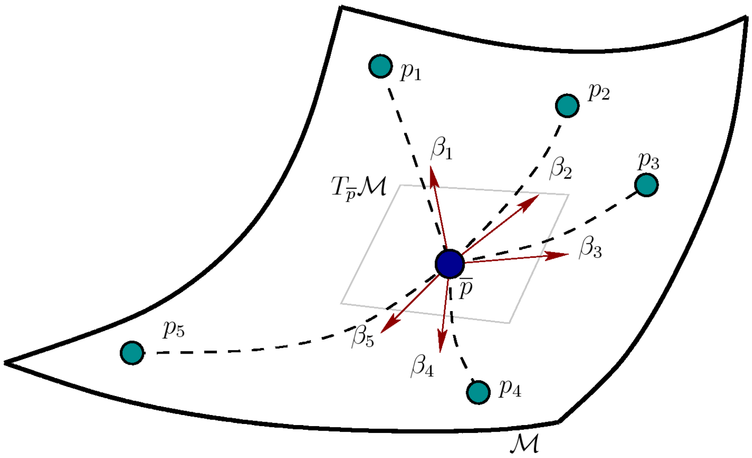

Figure 2: Definition of the covariance matrix $\Lambda_{\overline{p}}$

geodesic probability metrics in order to approximate Gaussian distributions of diffusion tensors based on these dissimilarity measures. In particular, we will show how the gradient of the squared distances can be computed and used to estimate the associated covariance matrices (equation (2)) as well as the empirical mean tensor. We will then evaluate and compare their respective virtue for our segmentation purpose.

2.1.3 Spatial gradient

We recall that we are interested in images Σ associating to each location of a regular sampling Ω of $\mathbb{R}^3$ an element of $S^+(3)$. The spatial gradient of Σ can be estimated from the gradient of the squared distance as:

$$\nabla_{\mathrm{s}e_k}\Sigma(x) = -\frac{\mathrm{s}}{|e_k|}\left(\nabla_{\Sigma(x)}\mathcal{D}^2\left(\Sigma(x), \Sigma(x+\mathrm{s}e_k)\right)\right)$$

where the $e_k,\ k=1,2,3$ denote the canonical basis of $\mathbb{R}^3$ and are used to access the neighbors of $\Sigma(x)$ on the grid Ω. s is either $+1$ or -1 and denotes the forward and backward approximation of the gradient. $\nabla_{-e_1}\Sigma(x)$ is, for example, the initial tangent vector of the geodesic joining $\Sigma(x)$ and $\Sigma(x-|e_1|(1,0,0)^T)$.
It is then straightforward to compute the squared norm of the gradient at location x as:

$$|\nabla\Sigma(x)|^2 = \frac{1}{2}\sum_{k=1}^{3}\sum_{\mathrm{s}=\pm1}|\nabla_{\mathrm{s}e_k}\Sigma(x)|^2_{\Sigma(x)} = \frac{1}{2}\sum_{k=1}^{3}\sum_{\mathrm{s}=\pm1}\mathcal{D}^2(\Sigma(x),\Sigma(x+\mathrm{s}e_k))$$

where the $\frac{1}{2}$ factor arises from the fact that we use $3\times3\times3$ neighborhoods.
We now use the fact that statistics and gradient norm can be computed from the distance $\mathcal{D}$ and its gradient. We endow the space $S^+(3)$ with different probability metrics (i.e. distances) and derive the associated statistics and gradient norms which will be used in section 3 respectively to model the distribution of tensors within a subset of a DT image and to detect the interface between white matter structures.

2.2 Derivations of statistics and gradient norms

As we will see shortly, the manifold $\mathcal{M}$ of three-dimensional normal distributions with zero mean can be identified with the manifold $S^+(3)$ of 3×3 real, symmetric, positive-definite matrices which provides a natural means of parameterizing those probability density functions. Ultimately, we will use the fact that the Fisher information matrix corresponds to the Riemannian metric on this manifold (see [23] for example) and induces a geodesic distance $\mathcal{D}_g$. However, other distances between parameterized normal distributions (i.e. between covariance matrices and, hence, diffusion tensors) have been introduced. We will first use the Euclidean distance $\mathcal{D}_e$, then exploit the properties of the symmetrized Kullback-Leibler divergence $\mathcal{D}_j$, also known as the J-divergence [29], and finally describe the geometry of $S^+(3)$ equipped with a metric derived from the Fisher information matrix.

2.2.1 Euclidean probability metric

We consider $S^+(3)$ with the simple Euclidean metric. In this case, the dissimilarity measure between diffusion tensors is given by the Frobenius norm of the difference such that for all $A, B \in S^+(3)$, we have

$$\mathcal{D}_e(A,B) = |A - B|_F = \sqrt{\mathrm{tr}\left((A-B)(A-B)^T\right)} \tag{3}$$

where tr denotes the trace operator. Using the fact that $\nabla_X \mathrm{tr}\,(XY) = Y^T$ for $X, Y \in GL(n)$, it is easy to see that:

$$\nabla_A \mathcal{D}_e^2(A,B) = A - B \tag{4}$$

In other words, we find that the gradient of the squared Euclidean distance corresponds to the usual *difference* tangent vector. This is a symmetric matrix which can be used to compute the 6×6 covariance matrix (2) of a set of N diffusion tensors.
Plugging equation (3) into equation (1), the empirical mean diffusion tensor is estimated as:

$$\overline{\Sigma}_e = \frac{1}{N}\sum_{i=1}^{N} \Sigma_i$$

where we denote by Σ_i the tensor located at voxel x_i in Ω. The associated covariance matrix is obtained as:

$$\Lambda_e = \frac{1}{N}\sum_{i=1}^{N} \varphi\left(\Sigma_i - \overline{\Sigma}_e\right)\varphi\left(\Sigma_i - \overline{\Sigma}_e\right)^T$$

The map $\varphi : S^+(3) \mapsto \mathbb{R}^6$ associates to each symmetric matrix $\beta_i = \Sigma_i - \overline{\Sigma}_e$ its 6 independent components. In this Euclidean setting, we can define a Gaussian distribution between diffusion tensors with the probability function:

$$p_e\left(\Sigma_i | \overline{\Sigma}_e, \Lambda_e\right) = \frac{1}{\sqrt{(2\pi)^6|\Lambda_e|}}\exp\left(-\frac{\varphi(\beta_i)^T \Lambda_e^{-1}\varphi(\beta_i)}{2}\right) \text{with } \beta_i = \Sigma_i - \overline{\Sigma}_e \tag{5}$$

We will use this expression, in the Euclidean case, for the probability distributions $p_{in/out}$ in equation (19) of section 3.3. Finally, the squared norm of the spatial gradient of a diffusion tensor image $\Sigma : \Omega \mapsto S^+(3)$ is given by

$$|\nabla\Sigma(x)|_e^2 = \frac{1}{2}\sum_{k=1}^{3}\sum_{\mathbf{s}=\pm 1} \mathrm{tr}\left((\Sigma(x) - \Sigma(x+\mathbf{s}e_k))\,(\Sigma(x) - \Sigma(x+\mathbf{s}e_k))^T \right) \tag{6}$$

and is used in the distribution p_b defined by equation (20) for the Euclidean case.

2.2.2 J-divergence probability metric

We now adopt a more information-theoretic point of view and consider another dissimilarity measure between Gaussian probability densities known as the Kullback-Leibler divergence $\mathcal{D}_{kl}$ or *relative entropy*. This probability metric has the desirable property of being invariant under affine transformation of the density parameters, hence it is invariant under congruence transformations such that

$$\mathcal{D}_{kl}(A,B) = \mathcal{D}_{kl}\left(XAX^T, XBX^T\right),\ \forall A,B \in S^+(3), X \in GL(3) \tag{7}$$

This property does not hold for the Euclidean distance previously introduced. The Kullback-Leibler divergence is defined for parametric as well as non-parametric densities. In equation (7), A and B actually stand for the covariance matrices of three-dimensional normal distributions $\mathscr{P}(r|A)$ and $\mathscr{P}(r|B)$ with zero mean and we have:

$$\mathcal{D}_{kl}(A,B) = \int_{\mathbb{R}^3} \mathscr{P}(r|A) \log \frac{\mathscr{P}(r|A)}{\mathscr{P}(r|B)}\, dr$$

We recall that diffusion tensors are indeed the parameters of Gaussian distributions $\mathscr{P}$ modeling the local displacement r of water molecules.
It turns out however that the Kullback-Leibler divergence is not symmetric and hence not a true metric. We will use, as in [62], its symmetrized version, or J-divergence:

$$\frac{1}{2}\int_{\mathbb{R}^3} \mathscr{P}(r|A) \log \frac{\mathscr{P}(r|A)}{\mathscr{P}(r|B)} + \mathscr{P}(r|B) \log \frac{\mathscr{P}(r|B)}{\mathscr{P}(r|A)}\, dr$$

As we will see in the next section, the J-divergence is closely related to the squared geodesic distance on $S^+(3)$ induced by the Fisher information matrix but only coincides with the latter for special probability densities. Hence it is natural to define:

$$\mathcal{D}_j(A,B) = \sqrt{\frac{1}{2}\left(\mathcal{D}_{kl}(A,B) + \mathcal{D}_{kl}(B,A)\right)}$$

As stated in [67] and used in [62], the expression of this distance is particularly simple when $\mathscr{P}$ is a three-dimensional Gaussian density:

$$\mathcal{D}_j(A,B) = \sqrt{\frac{1}{4}\mathrm{tr}\left(A^{-1}B + B^{-1}A\right) - 6} \tag{8}$$

We have the following proposition:

Proposition 2.2.1. *The gradient of the squared distance $\mathcal{D}_j^2$ between three-dimensional normal distributions parameterized by their covariance matrix $A, B \in S^+(3)$ is*

$$\nabla_A \mathcal{D}_j^2\,(A, B) = \frac{1}{4}\left(B^{-1} - A^{-1}BA^{-1}\right) \tag{9}$$

Proof. This comes from the fact that $\nabla_A \mathrm{tr}\,(B^{-1}A) = \nabla_A \mathrm{tr}\left(A^T B^{-T}\right) = B^{-T} = B^{-1}$ and that $\nabla_A \mathrm{tr}\,(A^{-1}B) = -\left(A^{-1}BA^{-1}\right)^T = -A^{-1}BA^{-1}$ □

From this result, we are able to compute the covariance matrix (2) of a set of diffusion tensors. We just need to define the empirical mean diffusion tensor (1) associated to the distance $\mathcal{D}_j$ (8). This was already proposed in [62] as the following theorem:

Theorem 2.2.1. *The empirical mean diffusion tensor of a set of N tensors $\{\Sigma_i\},\ i = 1, ...N$ is given by*

$$\overline{\Sigma}_j = \operatorname*{Arg\,min}_{\Sigma \in S^+(3)} \frac{1}{N}\sum_{i=1}^{N} \mathcal{D}_j^2(\Sigma, \Sigma_i) = V^{-1/2}\left(U^{1/2}VU^{1/2}\right)^{1/2} V^{-1/2}$$

with $U = \frac{1}{N}\sum_{i=1}^N \Sigma_i$ and $V = \frac{1}{N}\sum_{i=1}^N \Sigma_i^{-1}$

The associated covariance matrix is obtained as:

$$\Lambda_j = \frac{1}{N}\sum_{i=1}^{N} \varphi\left(\beta_i\right)\varphi\left(\beta_i\right)^T$$

where, once again, the map φ associates to each symmetric matrix $\beta_i = -\frac{1}{4}\left(\Sigma_i^{-1} - \overline{\Sigma}_j^{-1}\Sigma_i\overline{\Sigma}_j^{-1}\right)$ its 6 independent components. In this information-theoretic setting, we now define a Gaussian distribution between diffusion tensors with the probability function:

$$p_j\left(\Sigma_i|\overline{\Sigma}_j, \Lambda_j\right) = \frac{1}{\sqrt{(2\pi)^6|\Lambda_j|}}\exp\left(-\frac{\varphi\left(\beta_i\right)^T\Lambda_j^{-1}\varphi\left(\beta_i\right)}{2}\right) \text{ with } \beta_i = -\frac{1}{4}\left(\Sigma_i^{-1} - \overline{\Sigma}_j^{-1}\Sigma_i\overline{\Sigma}_j^{-1}\right) \tag{10}$$

We will use this expression, in the J-divergence case, for the probability distributions $p_{in/out}$ in equation (19) of section 3.3. Finally, we can easily obtain the squared norm of the spatial gradient of a DT image Σ as

$$|\nabla\Sigma(x)|_j^2 = \frac{1}{2}\sum_{k=1}^{3}\sum_{\mathbf{s}=\pm1}\left(\frac{1}{4}\mathrm{tr}\left(\Sigma(x)^{-1}\Sigma(x+\mathbf{s}e_k) + \Sigma(x)\Sigma(x+\mathbf{s}e_k)^{-1}\right) - 6\right) \tag{11}$$

and use it in the distribution p_b of equation (20) for the J-divergence case.

2.2.3 Geodesic probability metric

We introduce, as in [39], a last dissimilarity measure between diffusion tensors, which we claim to be more natural and powerful for the comparison of three-dimensional normal distributions. Its superiority will be demonstrated through the numerical experiments presented in section 4.
Following [51] and [10], it is possible to define a Riemannian metric on $S^+(3)$ in terms of the Fisher information matrix. The Fisher information is a popular measure of the amount of *information* carried by the realizations of a random variable about the unknown parameters of the underlying probability density. This is classically used to derive maximum likelihood estimators of density parameters. Once again, we use the natural chart φ of $S^+(3)$ such that for all $A \in S^+(3)$, we have $\varphi(A) = (\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4, \mathbf{x}^5, \mathbf{x}^6)^T$. The tangent space $T_A S^+(3)$ at $A \in S^+(3)$ coincides with $S(3)$, the space of 3×3 real, symmetric matrices. Its basis is denoted by $(\partial_1, ..., \partial_6)$.

We now detail the fundamental properties of $S^+(3)$ and propose an original formulation for a Gaussian law on this manifold. The fundamental tools needed to derive our numerical schemes were detailed in [46], [55], [9], [23], [11], [27], [43], [37], and [40]. Other recent works, such as [50] and [26] do not employ the information geometry associated with the Fisher information matrix but rather consider $S^+(3)$ as the quotient space $GL^+(3)/SO(3)$ to derive statistical or filtering tools on tensor fields.

The Fisher information matrix

$$g_{ij} = \int_{\mathbb{R}^3} \frac{\partial \log \mathscr{P}(r|A)}{\partial \mathbf{x}^i} \frac{\partial \log \mathscr{P}(r|A)}{\partial \mathbf{x}^j} \mathscr{P}(r|A)\, dr$$

takes the following form for $S^+(3)$:

Theorem 2.2.2. *The Riemannian metric for the space of three-dimensional normal distributions with zero mean, $S^+(3)$ is given, for all $A \in S^+(3)$ by*

$$g_{ij} = \langle \partial_i, \partial_j \rangle_A = \frac{1}{2} \mathrm{tr}\left(A^{-1}\partial_i A^{-1} \partial_j\right) \quad i, j = 1, ..., 6$$

In practice, this means that for any tangent vectors $V_1, V_2 \in S(3)$, their inner product at A is given by:

$$\langle V_1, V_2 \rangle_A = \frac{1}{2} \mathrm{tr}\left(A^{-1} V_1 A^{-1} V_2\right)$$

Below are two examples of the metric tensor G, respectively computed for $A_1 = \mathbb{I}$ and $A_2 = \mathrm{diag}\left(\sigma_1^2, \sigma_2^2, \sigma_3^2\right)$ with $\mathbb{I}$ and $\mathrm{diag}\left(\right)$ denoting the identity and diagonal matrices. They correspond to a locally isotropic diffusion process and to the more general case of an anisotropic diffusion, with variances σ_1^2, σ_2^2 and σ_3^2, whose principal axes coincide with the coordinate frame of the image:

$$G_{A_1} = \begin{pmatrix} 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 \end{pmatrix} \qquad G_{A_2} = \begin{pmatrix} \frac{1}{2\sigma_1^4} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_1^2\sigma_2^2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sigma_1^2\sigma_3^2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2\sigma_2^4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sigma_2^2\sigma_3^2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2\sigma_3^4} \end{pmatrix}$$

It is obvious from these examples that the second, third and fifth diagonal terms of the metric tensor receive contributions from cross-terms of the diffusion variances. Hence the factor $1/2$ in the first, fourth and sixth diagonal terms.

We recall that, if $\gamma : t \mapsto \gamma(t) \in S^+(3)$, $\forall t \in [0,1]$ denotes a curve segment in $S^+(3)$ between two normal distributions parameterized by their covariance matrices A and B, its length is expressed as:

$$\mathcal{L}(\gamma) = \int_0^1 \left(\sum_{i,j=1}^{6} g_{ij} \frac{d\mathbf{x}^i(t)}{dt} \frac{d\mathbf{x}^j(t)}{dt} \right)^{1/2} dt$$

As stated for example in [43], the geodesic starting from $\gamma(0) = A \in S^+(3)$ in the direction $\dot\gamma(0) \in S(3)$ is given by:

$$\gamma(t) = A^{1/2} \exp\left(tA^{-1/2}\dot\gamma(0)A^{-1/2}\right)A^{1/2} \quad \forall t \in [0,1] \tag{12}$$

We recall that the geodesic distance $\mathcal{D}_g$ between any two elements A and B is the length of the minimizing geodesic between these points:

$$\mathcal{D}_g(A,B) = \inf_\gamma \{\mathcal{L}(\gamma) : A = \gamma(0), B = \gamma(1)\}$$

It is given by the following theorem, whose original proof is available in an appendix of [2] but different versions can also be found in [55] and [27].

Theorem 2.2.3. *(S.T. Jensen, 1976)*
Consider the family of multivariate normal distributions with common mean vector but different covariance matrices. The geodesic distance between two members of the family with covariance matrices $A, B \in S^+(3)$ is given by

$$\mathcal{D}_g(A,B) = \sqrt{\frac{1}{2}\mathrm{tr}(\log^2(A^{-1/2}BA^{-1/2}))} = \sqrt{\frac{1}{2}\sum_{i=1}^{3}\log^2(\eta_i)} \tag{13}$$

where η_i denote the 3 eigenvalues of the matrix $A^{-1/2}BA^{-1/2}$.

Apart from being a true distance, hence being positive, symmetric and verifying the triangle inequality (see [27] although no complete proof of the triangle inequality was provided by the authors), this distance is also invariant under congruence transformation (i.e. affine invariant) as well as under inversion.

It is interesting, at this stage, to study the relationship between this geodesic distance and the J-divergence. As summarized in [3], given suitable technical conditions on two *nearby* densities $\mathscr{P}(r|A)$ and $\mathscr{P}(r|A+dA)$, the zeroth and first order terms of a Taylor expansion of the Kullback-Leibler divergence around $\mathscr{P}(r|A)$ vanish. Assuming second-order differentiability of $\mathcal{D}_{kl}$, a second order expansion of $\mathcal{D}_{kl}\left(A, A+dA\right)$ yields:

$$\frac{1}{2}\int_{\mathbb{R}^3}\left(\frac{1}{\mathscr{P}^2(r|A)}\frac{\partial\mathscr{P}(r|A)}{\partial\mathbf{x}^i}\frac{\partial\mathscr{P}(r|A)}{\partial\mathbf{x}^j}-\frac{1}{\mathscr{P}(r|A)}\frac{\partial^2\mathscr{P}(r|A)}{\partial\mathbf{x}^i\partial\mathbf{x}^j}\right)\mathscr{P}(r|A)d\mathbf{x}^i\,d\mathbf{x}^j\,dr$$

which can be shown to reduce to

$$\mathcal{D}_{kl}\left(A, A+dA\right)=\frac{1}{2}\mathbb{E}\left[\frac{\partial\log\mathscr{P}(r|A)}{\partial\mathbf{x}^i}\frac{\partial\log\mathscr{P}(r|A)}{\partial\mathbf{x}^j}\right]d\mathbf{x}^i d\mathbf{x}^j$$

(if the partial derivatives commute with the integral) and which is precisely half of the squared geodesic distance between A and $A+dA$. Consequently it is easy to see that the J-divergence coincides, up to the second order, with half of the squared geodesic distance between two nearby diffusion tensors. Whenever the tensors are not *infinitesimally close*, the two distances become inconsistent. This is another reason supporting our claim that diffusion tensors statistics based on the geodesic distance should improve the quality of DTI segmentation results.
It was shown in [43] that the gradient of the squared geodesic distance writes:

$$\nabla_A\mathcal{D}_g^2\left(A,B\right)=A\log\left(B^{-1}A\right) \tag{14}$$

Based on this result and on the following method for the computation of the mean tensor in our Riemannian setting, we will be able to estimate the covariance matrix (2) of a set of diffusion tensors $\{\Sigma_i\}$, $i=1,...,N$ and, finally, approximate a Gaussian distribution on $S^+(3)$. As presented in [40], a closed-form expression for the empirical mean (1) cannot be obtained but a gradient descent algorithm was proposed. It estimates a quantity, known as the Riemannian barycenter, which exists and is unique for manifolds of non-positive sectional curvature (see [32]) like $S^+(3)$. The algorithm is based on the minimization of the variance of the Σ_i. It can be shown that this boils down to evolving an initial guess of the mean (like the identify matrix $\mathbb{I}$) along the geodesics of $S^+(3)$ (equation (12)) with a velocity given by the gradient of the variance, i.e. a tangent vector V such as

$$V=-\frac{1}{N}\sum_{i=1}^{N}\nabla_M\mathcal{D}_g^2\left(M,\Sigma_i\right)=-\frac{1}{N}M\sum_{i=1}^{N}\log\left(\Sigma_i^{-1}M\right)$$

where M denotes the evolving mean tensor. After a few iterations of this procedure, M converges toward the mean tensor $\overline{\Sigma}_g$. We describe this procedure in the Algorithm 1.

Algorithm 1 Riemannian estimation of the mean diffusion tensor

Require: $\{\Sigma_i\} \in S^+(3)$, $i = 1, ...N$ and nit, the number of iterations

Ensure: $\overline{\Sigma}_g$, the mean tensor

$M \leftarrow \mathbb{I}$

for $k = 1$ to nit **do**

  $V \leftarrow \mathbb{O}$ $\{3 \times 3$ zero matrix$\}$

  for $i = 1$ to N **do**

    $V \leftarrow \log\left(\Sigma_i^{-1} M\right)$

  end for

  $V = \frac{1}{N} M V$

  $M \leftarrow M^{1/2} \exp\left(-M^{-1/2} V M^{-1/2}\right) M^{1/2}$

end for

Return M

The associated covariance matrix is obtained as:

$$\Lambda_g = \frac{1}{N} \sum_{i=1}^{N} \varphi\left(\beta_i\right) \varphi\left(\beta_i\right)^T$$

where $\beta_i = -\overline{\Sigma}_g \log\left(\Sigma_i^{-1} \overline{\Sigma}_g\right)$ and φ associates to each β_i its 6 independent components. The notion of Gaussian distribution was generalized to random samples of primitives belonging to a Riemannian manifold in [49] where more details can be found regarding this particular point. From this work, we have proposed in [40] a definition of the Gaussian law between diffusion tensors which can be approximated as follows for a covariance matrix Λ_g of small variance $\sigma^2 = \text{tr}(\Lambda_g)$:

$$p_g(\Sigma_i | \overline{\Sigma}_g, \Lambda_g) \simeq \frac{1}{\sqrt{(2\pi)^6 |\Lambda_g|}} \exp \frac{-\varphi\left(\beta_i\right)^T \Gamma \varphi\left(\beta_i\right)}{2} \tag{15}$$

where β_i is defined as $\beta_i = -\overline{\Sigma}_g \log\left(\Sigma_i^{-1} \overline{\Sigma}_g\right)$ and the concentration matrix is $\Gamma \simeq \Lambda_g^{-1} - \mathcal{R}/3$, with $\mathcal{R}$ the Ricci tensor at the mean $\overline{\Sigma}_g$. The computation of the Ricci tensor $\mathcal{R}$ can be performed on the basis of closed-form expressions for the metric and the Riemann tensor provided in [55] and simply involving traces of matrix products. As we will point out in section 4, our numerical experiments have shown that the Ricci tensor exhibits a difference of at least 2 orders of magnitude with the inverse of the covariance matrix. Hence we can approximate Γ by Λ_g^{-1}.

We will use p_g, in the geodesic case, for the probability distributions $p_{in/out}$ in equation (19) of section 3.3. Finally, the squared norm of the spatial gradient of a DT image can be estimated as follows:

$$|\nabla \Sigma(x)|_g^2 = \frac{1}{2} \sum_{k=1}^{3} \sum_{\text{s}\pm 1} \left(\frac{1}{2} \text{tr}\left(\log^2\left(\Sigma(x)^{-1/2} \Sigma(x + \text{s} e_k) \Sigma(x)^{-1/2} \right) \right) \right) \tag{16}$$

and subsequently used in the distribution p_b of equation (20) in the geodesic case.

2.2.4 Summary and numerical examples

We summarize, in table 1, the expressions of the squared distance, its gradient, and of the mean tensor for the Euclidean, J-divergence and geodesic cases. The evaluation of the squared distance and its gradient for the matrices A_1 and B_1 given below shows a good coherence (although the Euclidean distance is quite larger than the other two) and, more importantly, illustrates the fact that the J-divergence accurately approximates half of the squared geodesic distance when the tensors are relatively close:

$$A_1 = \begin{pmatrix} 0.9878 & -0.0527 & 0.0050 \\ -0.0527 & 1.0112 & -0.0372 \\ 0.0050 & -0.0372 & 1.0391 \end{pmatrix} \qquad B_1 = \begin{pmatrix} 1.0384 & -0.0012 & 0.0107 \\ -0.0012 & 1.0056 & -0.0060 \\ 0.0107 & -0.0060 & 1.0233 \end{pmatrix}$$

$$\mathcal{D}_e^2(A_1, B_1) = 0.010158\,, \qquad \nabla_{A_1}\mathcal{D}_e^2\,(A_1, B_1) = \begin{pmatrix} -0.0506 & -0.0515 & -0.0057 \\ -0.0515 & 0.0056 & -0.0312 \\ -0.0057 & -0.0312 & 0.0158 \end{pmatrix}$$

$$\mathcal{D}_j^2(A_1, B_1) = 0.002526\,, \qquad \nabla_{A_1}\mathcal{D}_j^2\,(A_1, B_1) = \begin{pmatrix} -0.0274 & -0.0266 & -0.0040 \\ -0.0266 & -0.0002 & -0.0147 \\ -0.0040 & -0.0147 & 0.0066 \end{pmatrix}$$

$$\mathcal{D}_g^2(A_1, B_1) = 0.005050\,, \qquad \nabla_{A_1}\mathcal{D}_g^2\,(A_1, B_1) = \begin{pmatrix} -0.0480 & -0.0503 & -0.0048 \\ -0.0503 & 0.0074 & -0.0314 \\ -0.0048 & -0.0314 & 0.0164 \end{pmatrix}$$

On the contrary, if we consider the matrices A_2 and B_2, which are much more different than A_1 and B_1, we find out that the J-divergence becomes sensibly different from half of the squared geodesic distance:

$$A_2 = \begin{pmatrix} 1.0696 & -0.0563 & 0.4035 \\ -0.0563 & 0.5621 & 0.1068 \\ 0.4035 & 0.1068 & 1.4086 \end{pmatrix} \qquad B_2 = \begin{pmatrix} 1.2813 & 0.2320 & 0.0327 \\ 0.2320 & 1.2782 & 0.1965 \\ 0.0327 & 0.1965 & 0.9392 \end{pmatrix}$$

$$\mathcal{D}_e^2(A_2, B_2) = 1.111446\,, \qquad \nabla_{A_2}\mathcal{D}_e^2\,(A_2, B_2) = \begin{pmatrix} -0.2117 & -0.2883 & 0.3708 \\ -0.2883 & -0.7160 & -0.0897 \\ 0.3708 & -0.0897 & 0.4695 \end{pmatrix}$$

$$\mathcal{D}_j^2(A_2, B_2) = 0.329119\,, \qquad \nabla_{A_2}\mathcal{D}_j^2\,(A_2, B_2) = \begin{pmatrix} -0.2029 & -0.2875 & 0.1765 \\ -0.2875 & -0.8811 & 0.0783 \\ 0.1765 & 0.0783 & 0.0880 \end{pmatrix}$$

$$\mathcal{D}_g^2(A_2, B_2) = 0.621560\,, \qquad \nabla_{A_2}\mathcal{D}_g^2\,(A_2, B_2) = \begin{pmatrix} -0.0648 & -0.1598 & 0.4483 \\ -0.1598 & -0.4424 & -0.0799 \\ 0.4483 & -0.0799 & 0.6295 \end{pmatrix}$$

Now, in order to compare the statistics derived from each distance, we have manually segmented the genu of the corpus callosum on a DTI data-set used in the last section of this paper. This is a well-known region of the brain white matter (figure 3) where fibers are essentially aligned in a right-left fashion, i.e. along the x_1 axis on an axial slice. Consequently, the tensors in this region are very anisotropic with a major eigenvector close to $(1, 0, 0)^T$. This resulted in a set $\{\Sigma_i\}$ of

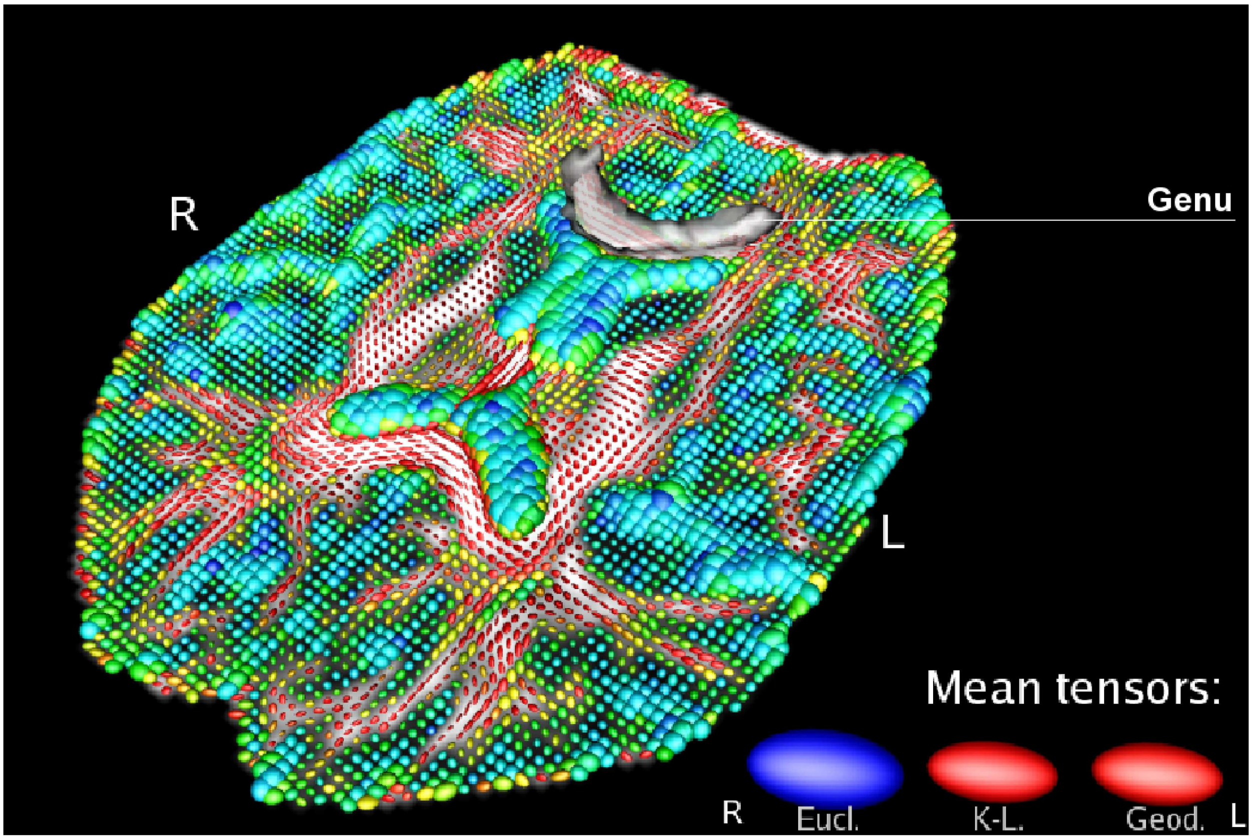

Figure 3: Statistics in the genu of the corpus callosum (R:right, L:left)

Distance	Euclidean	J-divergence	Geodesic
$\mathcal{D}^2(A,B)$	$\mathrm{tr}\left((A-B)(A-B)^T\right)$	$\frac{1}{4}\left(\mathrm{tr}\left(A^{-1}B+B^{-1}A\right)-6\right)$	$\frac{1}{2}\mathrm{tr}\left(\log^2\left(A^{-1/2}BA^{-1/2}\right)\right)$
$\nabla_A\mathcal{D}^2(A,B)$	$A-B$	$\frac{1}{4}(B^{-1}-A^{-1}BA^{-1})$	$A\log\left(B^{-1}A\right)$
$\overline{\Sigma}$	$\frac{1}{N}\sum_{i=1}^N \Sigma_i$	$V^{-1/2}\left(U^{1/2}VU^{1/2}\right)^{1/2}V^{-1/2}$ with $U=\frac{1}{N}\sum_{i=1}^N \Sigma_i$ and $V=\frac{1}{N}\sum_{i=1}^N \Sigma_i^{-1}$	Algorithm 1

Table 1: Squared distances and their gradient for $A, B \in S^+(3)$, mean tensor.

$N = 614$ tensors. The ellipsoids presented in the bottom-right corner of figure 3 represent the mean tensor respectively computed, from left to right, with the Euclidean distance, J-divergence and geodesic distance (the color encodes their relative fractional anisotropy). Visually, we can see that the Euclidean mean is somehow more oblate than the other two ellipsoids. This can be explained by the fact that Euclidean averaging is blind to the spectral components of the tensors (eigenvalues and eigenvectors) and has a tendency to mix them. We now present the estimated statistics for each distance (We scaled by a factor 2 the values obtained for the J-divergence to make the comparisons easier).

Euclidean probability metric:

$$\overline{\Sigma}_e = \begin{pmatrix} 2.6923 & -0.1334 & 0.0347 \\ -0.1334 & 1.3947 & 0.0526 \\ 0.0347 & 0.0526 & 1.1228 \end{pmatrix} \Lambda_e = \begin{pmatrix} 2.8685 & -0.2261 & 0.0589 & 0.6690 & 0.0892 & 0.2083 \\ -0.2261 & 0.0178 & -0.0046 & -0.0527 & -0.0070 & -0.0164 \\ 0.0589 & -0.0046 & 0.0012 & 0.0137 & 0.0018 & 0.0042 \\ 0.6690 & -0.0527 & 0.0137 & 0.1560 & 0.0208 & 0.0485 \\ 0.0892 & -0.0070 & 0.0018 & 0.0208 & 0.0027 & 0.0064 \\ 0.2083 & -0.0164 & 0.0042 & 0.0485 & 0.0064 & 0.0151 \end{pmatrix}$$

$\mathrm{tr}\,(\Lambda_e) = 3.0615$

J-divergence probability metric:

$$\overline{\Sigma}_j = \begin{pmatrix} 2.2901 & -0.1063 & 0.0296 \\ -0.1063 & 1.0833 & 0.0455 \\ 0.0296 & 0.0455 & 0.8775 \end{pmatrix} \Lambda_j = \begin{pmatrix} 0.0369 & 0.0075 & -0.0015 & 0.0139 & -0.0105 & 0.0816 \\ 0.0075 & 0.0692 & 0.0035 & 0.0137 & -0.0051 & 0.0236 \\ -0.0015 & 0.0035 & 0.0413 & -0.0135 & 0.0019 & -0.0142 \\ 0.0139 & 0.0137 & -0.0135 & 0.4958 & -0.0405 & 0.5147 \\ -0.0105 & -0.0051 & 0.0019 & -0.0405 & 0.0432 & -0.0552 \\ 0.0816 & 0.0236 & -0.0142 & 0.5147 & -0.0552 & 0.8296 \end{pmatrix}$$

$\mathrm{tr}\,(\Lambda_j) = 1.5161$

Geodesic probability metric:

$$\overline{\Sigma}_g = \begin{pmatrix} 2.3296 & -0.1088 & 0.0312 \\ -0.1088 & 1.1102 & 0.0523 \\ 0.0312 & 0.0523 & 0.8912 \end{pmatrix} \Lambda_g = \begin{pmatrix} 0.7706 & -0.0297 & 0.0207 & -0.0003 & -0.0267 & 0.1970 \\ -0.0297 & 0.3156 & 0.0431 & -0.0461 & -0.0016 & -0.0041 \\ 0.0207 & 0.0431 & 0.1332 & -0.0113 & -0.0086 & 0.0073 \\ -0.0003 & -0.0461 & -0.0113 & 0.4592 & 0.0094 & 0.3010 \\ -0.0267 & -0.0016 & -0.0086 & 0.0094 & 0.0276 & 0.0068 \\ 0.1970 & -0.0041 & 0.0073 & 0.3010 & 0.0068 & 0.3306 \end{pmatrix}$$

$\mathrm{tr}\,(\Lambda_g) = 2.0370$

It is clear that there are important differences between these three approaches. They are hard to interpret though on such a simple example but their effect on the segmentation results will be outlined in the section 4.

In the next section, we set up a unified Bayesian formulation of the segmentation problem that will be used throughout this paper. It relies on the different possible estimates of the mean $\overline{\Sigma}$ and covariance matrix Λ (equation (2)) to evaluate the likelihood of a diffusion tensor to belong to a given subset of the DTI data-set. This will be used in equation (19). We recall that we will consider 3 different cases associated to the Euclidean distance (3), the J-divergence (8) and the geodesic distance (13). Within these 3 different frameworks, we have shown how to approximate a Gaussian distribution between diffusion tensors (see equations (5), (10) and (15)) by using the information provided by the gradient of the squared geodesic distance (see equations (4), (9) and (14)). We will also exploit the information provided by the norm of the tensor field spatial gradient (see equations (6), (11) and (16)) to localize the boundaries between structures of the brain white matter and avoid mixing them through the boundary term (20) in our energy (21).

3 Statistical Segmentation by Surface Evolution

We recall that our goal is to compute the optimal 3D surface separating an anatomical structure of interest from the rest of a DTI data-set. The statistical surface evolution, as developed in [52], is a well-suited framework for our segmentation problem. We hereafter summarize the important notions of this technique.

3.1 Bayesian formulation for image partitioning

Following general works on image segmentation [34], [70], [4], [48], we seek the optimal partition of the image domain Ω by maximizing the *a posteriori* frame partition probability $p(\mathcal{P}(\Omega)\,|\,\Sigma)$ for the observed diffusion tensor image Σ. The Bayes rule allows to express this probability as:

$$p(\mathcal{P}(\Omega)\,|\,\Sigma) \propto p(\Sigma\,|\,\mathcal{P}(\Omega))p(\mathcal{P}(\Omega)) \tag{17}$$

This formulation yields a separation of the image-based cues from the geometric properties of the boundary given by $\mathcal{P}(\Omega)$. While being valid for any number of regions, we restrict this formulation to binary partitions: the structure of interest and the background. The image partition can be represented as the zero-crossing of a level-set function ϕ [20], [21], [47],[16]. Noting $\mathcal{B}$ the interface between the two regions Ω_{in} and Ω_{out}, ϕ is constructed as the signed distance function to $\mathcal{B}$:

$$\begin{cases} \phi(x) = 0, & \text{if } x \in \mathcal{B} \\ \phi(x) = \mathscr{D}(x,\mathcal{B}), & \text{if } x \in \Omega_{in} \\ \phi(x) = -\mathscr{D}(x,\mathcal{B}), & \text{if } x \in \Omega_{out} \end{cases}$$

where $\mathscr{D}(x,\mathcal{B})$ stands for the Euclidean distance between x and $\mathcal{B}$. Hence, the optimal partition is obtained by maximizing: $p(\phi|\Sigma) \propto p(\Sigma|\phi)p(\phi)$. At this stage, these two terms still need to be defined. For this purpose, several assumptions on the structure of interest need to be introduced. In the following, a smoothness constraint is imposed with the term $p(\phi)$ while $p(\Sigma|\phi)$ expresses the likelihood of the diffusion tensors to be inside, outside or on the boundary of the structure. This yields an optimization criterion similar to the *Geodesic Active Regions* presented in [48].

3.2 Smoothness constraint

The second term of equation (17) expresses the probability of the interface to represent the structure of interest and can be used to introduce prior shape knowledge. For the segmentation of diffusion tensor images, we have no high level prior information but we can use this term to impose shape regularity. Such a constraint can be obtained by favoring structures with a smaller surface $|\mathcal{B}|$ with $p(\phi) \propto \exp\left(-\nu|\mathcal{B}|\right)$. This can be expressed with ϕ by introducing the Dirac function [68]:

$$p(\phi) \propto \exp\left(-\nu \int_{\Omega} \delta(\phi)|\nabla\phi(x)|\,dx\right) \tag{18}$$

3.3 Data term

To further specify the image term $p(\Sigma|\phi)$, we introduce some hypothesis. First, for a given level-set ϕ, we can classify the voxels into three classes: inside, outside or on the boundary. Then, we can define the probability density functions of a diffusion tensor for each class: p_{in}, p_{out} and p_b. Assuming the diffusion tensors to be independent and identically distributed realizations of the corresponding random process, the data term is given by:

$$p(\Sigma|\phi) = \prod_{x\in\Omega_{in}} p_{in}(\Sigma(x)) \quad . \prod_{x\in\Omega_{out}} p_{out}(\Sigma(x)) \quad . \prod_{x\in\mathcal{B}} p_b(\Sigma(x)) \tag{19}$$

This gives two different types of probability distributions: region-based with $p_{in/out}$ and boundary-based with p_b. p_{in} and p_{out} are given by the Gaussian distributions on tensors introduced in section 2.2 p_e (equation (5)) p_j, (equation (10)) and p_g (equation (15)). The parameters of these laws may be known a priori but in the absence of such information, they are introduced as unknown parameters.

Regarding p_b, the probability should be close to one for high gradients of the diffusion tensors field and around zero for small variations. This leads to:

$$p_b(\Sigma(x)) \propto \exp\left(-g_\alpha(|\nabla\Sigma(x)|)\right) \tag{20}$$

with $g_\alpha(u) = 1/(1 + u^\alpha)$. This type of boundary term is the basis of several works referred to as active contours [15] and, often, $\alpha = 1$ or 2 is chosen. For the sake of readability, we will use the short notation $g_\alpha(\Sigma(x))$. $|\nabla\Sigma(x)|$ will be computed by using equation (6) for the Euclidean case, equation (11) for the J-divergence case, or equation (16) for the geodesic case.

3.4 Energy formulation

Maximizing the *a posteriori* segmentation probability is equivalent to minimizing its negative logarithm. Integrating the regularity constraint (18) and the image term (19), we end up with the following energy:

$$\begin{aligned} E(\phi, \overline{\Sigma}_{in/out}, \Lambda_{in/out}) = {} & \nu \int_\Omega \delta(\phi)|\nabla\phi|\, dx + \int_\Omega \delta(\phi)|\nabla\phi| g_\alpha(\Sigma(x))\, dx \\ & - \int_{\Omega_{in}} \log p(\Sigma(x)|\overline{\Sigma}_{in}, \Lambda_{in}) dx - \int_{\Omega_{out}} \log p(\Sigma(x)|\overline{\Sigma}_{out}, \Lambda_{out}) dx. \end{aligned} \tag{21}$$

The boundary term of this energy corresponds to the Geodesic Active Contours [15] and naturally includes a regularization[1] on the interface. Following [33], [53], an alternate minimization is employed to perform the optimization for the two types of unknown parameters. For given statistical parameters, the Euler-Lagrange equations are computed to derive the implicit front evolution:

$$\frac{\partial\phi}{\partial t} = \delta(\phi)\left((\nu + g_\alpha(\Sigma))\, \mathrm{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) + \frac{\nabla\phi}{|\nabla\phi|}\cdot\nabla g_\alpha(\Sigma) + \log\frac{p(\Sigma|\overline{\Sigma}_{in}, \Lambda_{in})}{p(\Sigma|\overline{\Sigma}_{out}, \Lambda_{out})}\right), \tag{22}$$

[1] The regularity term (18) could be included in p_b by replacing g_α by $g_{\alpha,\nu} = \nu + g_\alpha$.

while the statistics can be updated after each evolution of ϕ from their empirical estimates, as described in section 2.2. More details on this level-set based optimization can be found in [16], [53], where different applications were considered.
The right-hand side of equation (22), between parenthesis, corresponds to the magnitude of the velocity used to deform each point of the evolving surface $\mathcal{B}$ along its normal at that point. The purpose of the next section will be to evaluate the influence of the choice of the density function p, which can be taken in its Euclidean version p_e (equation (5)), J-divergence version p_j (equation (10)) or geodesic version p_g (equation (15)). We will describe several numerical experiments in order to evaluate the respective performances of each probability metrics for our DTI segmentation task. We demonstrate that the Riemannian statistical tools presented in section 2.2 achieve the best results.

4 Results and Validation

We begin our numerical experiments with three different synthetic data-sets of increasing complexity in order to emphasize the respective virtue of the Euclidean, Kullback-Leibler and geodesic probability metrics. We then apply our algorithm to a biological rat spinal cord phantom. Finally, we consider real DTI data-sets on which we perform the segmentation of the corpus callosum.

In practice, there are a few important points that must be carefully addressed when implementing and running our segmentation algorithm: When dealing with real DTI data, we use a mask of the brain so that tensors statistics of Ω_{out} are not corrupted by the signal from the outside of the brain. Regarding the initialization, we noticed and will demonstrate that our method is very robust. We will show that the geodesic distance is indeed the only metric capable of representing, through the associated Gaussian distribution, a smoothly varying tensor field with relatively high variability. Next, there are two parameters that have to be chosen: The first one is the value of ν in equation (18). It constrains the smoothness of the surface and is usually set in the range 1 to 10. The second parameter arises from the very definition of the Gaussian distribution on $S^+(3)$ presented in section 2.2.3. The main hypothesis for this definition to be valid is that the trace of the covariance matrix Λ_g should be small and this means that we restrict ourselves to concentrated distributions. Hence, we set a threshold for the variance which, whenever reached, induces the end of the update for the statistical parameters. We then let the surface evolve while using a fixed mean and covariance matrix to model the distributions of the tensors in Ω_{in}/Ω_{out}. The threshold is chosen in the range $[0.01, 0.1]$ for tensors with components around 1.0. We noticed that the variance, after a few iterations of increase at the very beginning of the algorithm, keeps decreasing as the segmentation process converges. Consequently, a careful selection of this parameter is not critical. Finally, we improved the computational efficiency of the method using the geodesic distance by noticing and experimentally verifying that, in equation 15, the term involving the 6×6 Ricci tensor $\mathcal{R}/3$ can be neglected since we have found, in our numerical experiments, a difference of at least 2 orders of magnitude between Λ_g^{-1} and $\mathcal{R}/3$.
Regarding the computational cost of the method, we should point out that it is fairly efficient since

the results presented on figure 19 and 20 were respectively obtained, on $128 \times 128 \times 64$ images, in 5 and 10 minutes on a 1.7 GHz Pentium M processor with 1 Gb of RAM.

4.1 Synthetic examples

Each of the three synthetic data-sets consists of a $40 \times 40 \times 40$ 3D tensor field with a main pattern and a background. The tensors follow the shape of the pattern so that, as the shape becomes more twisted, the tensors variability increases and makes it more difficult for the algorithm to recover the entire shape. The regularity factor ν is set to 1 for all the experiments. The initialization is done by the means of one or two little spheres (see figures below). Finally, the mean tensor and covariance matrix used for the generation of noise in all the experiments are the following:

$$\overline{\Sigma} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \Lambda = \begin{pmatrix} 0.0885 & -0.0568 & -0.0260 & 0.0119 & -0.0394 & 0.0035 \\ -0.0568 & 0.0701 & 0.0039 & -0.0070 & 0.0122 & -0.0112 \\ -0.0260 & 0.0039 & 0.0183 & -0.0023 & 0.0218 & 0.0095 \\ 0.0119 & -0.0070 & -0.0023 & 0.0078 & -0.0113 & 0.0010 \\ -0.0394 & 0.0122 & 0.0218 & -0.0113 & 0.0416 & 0.0118 \\ 0.0035 & -0.0112 & 0.0095 & 0.0010 & 0.0118 & 0.0160 \end{pmatrix}$$

4.1.1 On the generation of Gaussian noise in $S^+(3)$

The generation of random tensors, i.e. Gaussian noise in $S^+(3)$, is usually addressed by simply building symmetric matrices with i.i.d. normally distributed components and then enforcing their positivity. The main drawback of this approach is that it leaves no grasp on the actual distribution of tensors. We proposed in [40], to use the equation 15, to generate random tensors with a known mean $\overline{\Sigma}$ and covariance matrix Λ.

Algorithm 2 Generation of Gaussian noise in $S^+(3)$

Require: $\overline{\Sigma}$ and Λ, mean tensor and covariance matrix
Ensure: Σ_i, N normally distributed elements of $S^+(3)$
1: **for** $i = 1$ to N **do**
2: $\Lambda = HH^T$ {Cholesky decomposition of the covariance matrix}
3: Create a random vector $Z \in \mathbb{R}^6$, with zero mean and unit variance
4: From $\beta_i = \varphi^{-1}(HZ) \in S(3)$
5: $\Sigma_i \leftarrow \left(\exp\left(-\overline{\Sigma}^{-1}\beta_i\right)\overline{\Sigma}^{-1}\right)^{-1}$
6: **end for**

The method, described in the algorithm 2, is fairly simple since all we need to do is to randomly choose the initial velocities $\{\beta_i\}, i = 1, ..., N$ of the geodesics in $S^+(3)$ joining the imposed mean tensor $\overline{\Sigma}$ to the random elements Σ_i. In practice, this operation is performed in local coordinates so that we only need to draw random samples of the $\varphi(\beta_i) \in \mathbb{R}^6$ with zero mean and covariance matrix Λ. The Σ_i are easily obtained by using the expression $\beta_i = -\overline{\Sigma}\log\left(\Sigma_i^{-1}\overline{\Sigma}\right)$ (table 1).

4.1.2 The *Y* tensor field

We start with a simple example composed by a *diverging* tensor field and a background of isotropic tensors (figure 4). Within the *Y* shape, tensors fractional anisotropy decreases as we get away from the center-line. This example is relatively simple since the tensors variability stays low and the segmentation procedure succeeds with the three probability metrics. One important difference must be noted though: By comparison with the Euclidean distance, which requires 45 iterations to segment the *Y* structure, the process converges faster when the J-divergence is used (30 iterations), and relatively faster with the geodesic distance (28 iterations). This is easily explained by the fact that the velocity of the evolving surface, at location x of the image Σ, is directly related to the likelihood of tensor $\Sigma(x)$ to belong to Ω_{in} or Ω_{out}. It is hence a first argument in favor of our claim that the geodesic probability metrics yields more adequate tensor statistics.

4.1.3 The *torus* tensor field

Next, we consider another example where the tensors follow the tangent of the center-line of a torus (figure 5) and share the same eigenvalues. This yields a higher orientational variability of the tensors. A direct consequence of this increased variability is the failure of the segmentation process when we use the Euclidean probability metric. The evolution is presented on figure 7. The initial sphere is setup so that it covers half of the torus and contains the part of the background situated 'inside' the torus. The surface evolution falls into a local minimum and is unable to recover the desired shape. On the contrary, the J-divergence and geodesic distance behave consistently and succeed to segment the complete torus (figure 8). We notice, as in the previous example, that the segmentation using the geodesic distance converges faster (20 iterations) than the one relying on the J-divergence (27 iterations). The result presented for the Euclidean metric on figure 7 is the final state after 600 iterations.

4.1.4 The *helix* tensor field

The last synthetic data-set that we consider is the *helix* tensor field presented on figure 6. It is composed of a background with anisotropic tensors aligned on the x_1 axis of the 3D field and an helix containing tensors oriented along the tangent of its center-line. The fractional anisotropy of the helix tensors varies around each spire. Moreover, the tensors orientation spans a broader range of possibilities than in the *torus* case since it changes along the x_1, x_2 and x_3 axes. This is certainly an example on which it is desirable for our segmentation algorithm to succeed since this tensors variation pattern is fairly realistic and may be found in real DT images. As a matter of fact, **only** the statistics computed with the geodesic distance enable our segmentation framework to achieve a correct extraction of the helix. The initialization consists of 2 small spheres overlapping the helix and the background. As we can see on figure 9 and 10, the surface evolution quickly stops when it uses the Euclidean distance or the J-divergence, even though the latter propagates further than the former as we could have expected. The local minima are respectively reached after 130 and 80 iterations. Using the geodesic distance, the complete helix is recovered (figure 11) after 300 iterations.

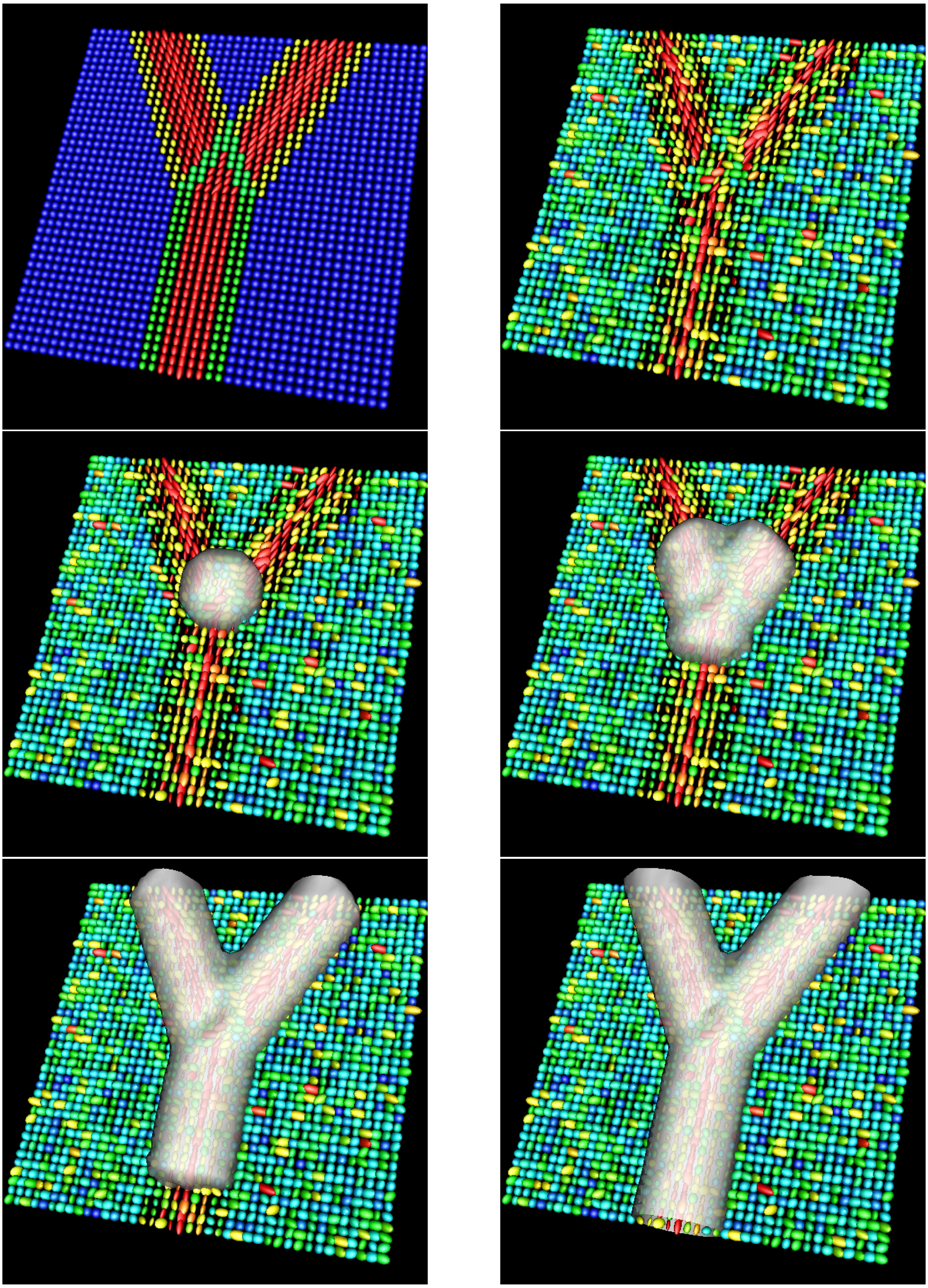

Figure 4: Segmentation of the *Y* tensor field. Top row: Axial slice of the original and noisy data-set. Middle and bottom rows: Evolution of the segmentation (color map indicates FA)

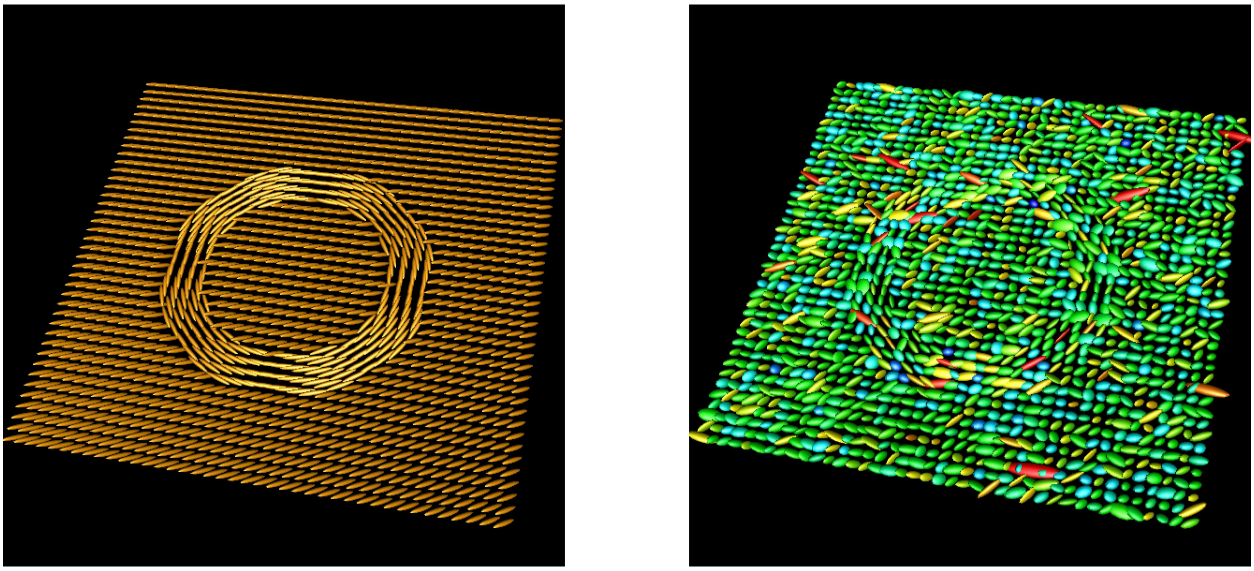

Figure 5: Axial slice of the original and noisy *torus* tensor field

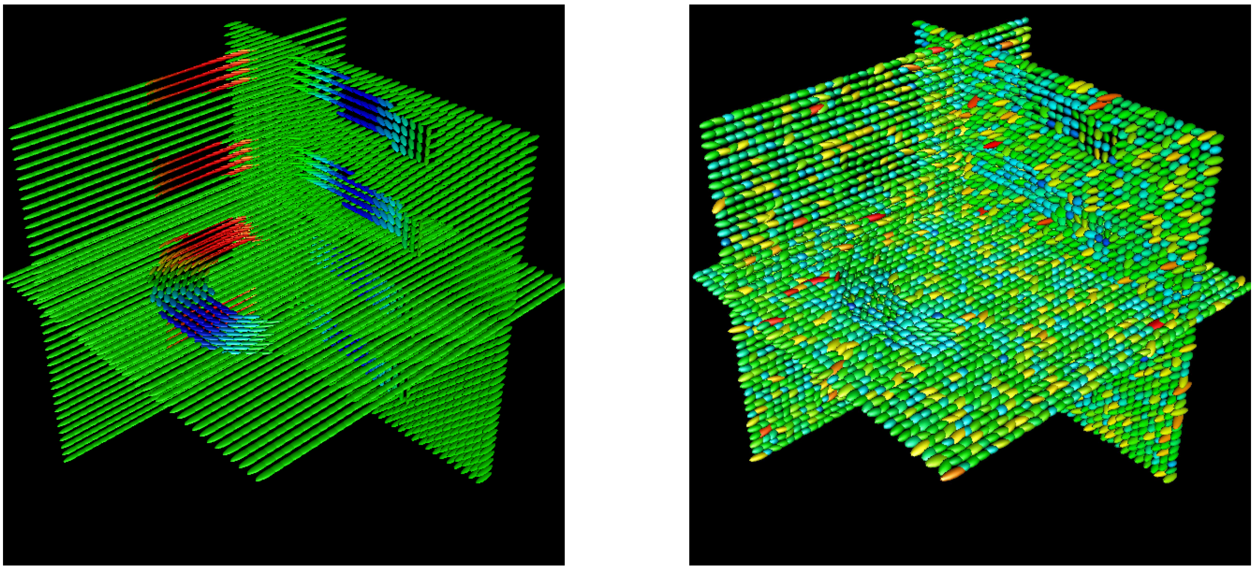

Figure 6: Axial, coronal and sagittal slices of the original and noisy *helix* tensor field

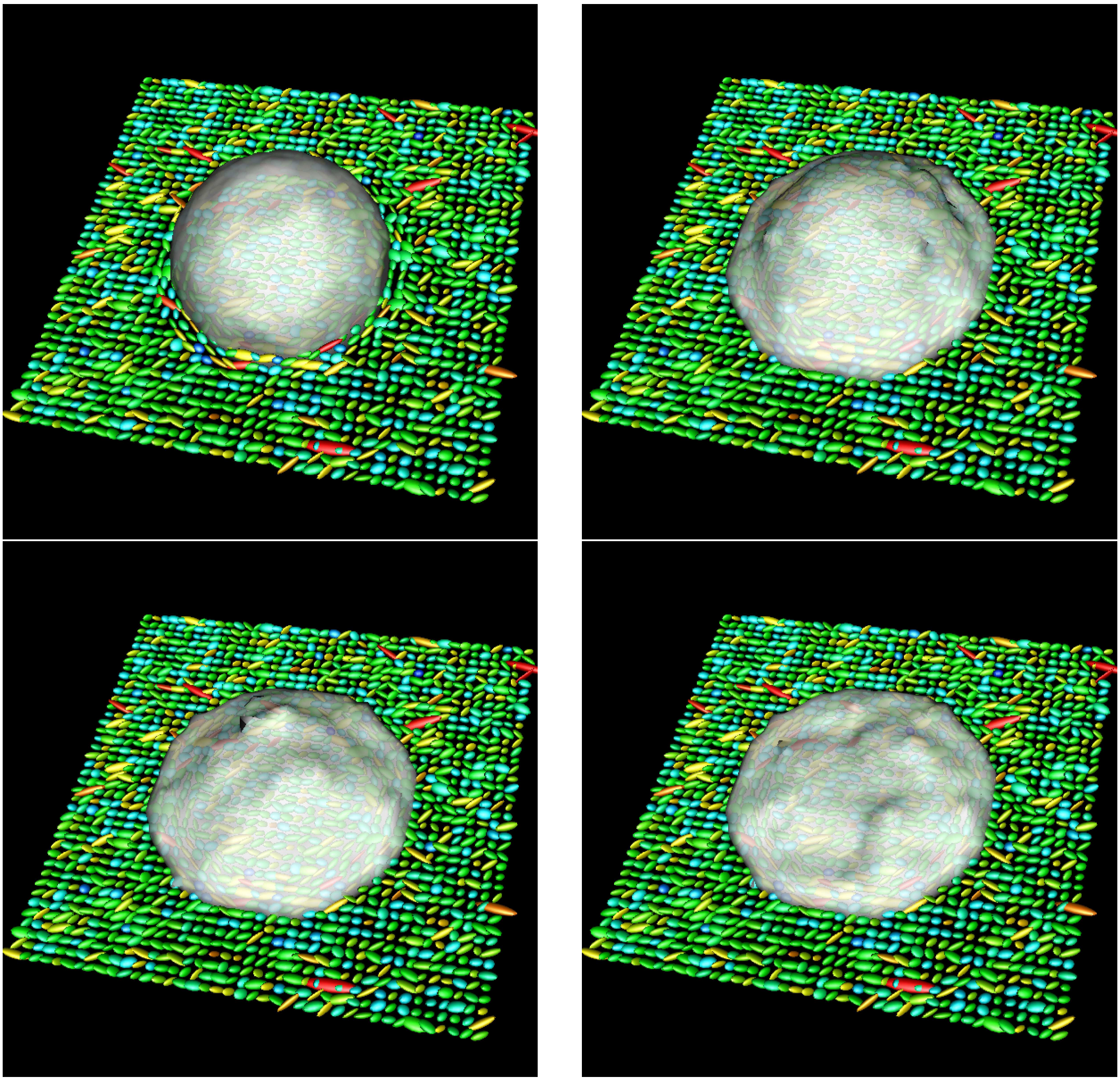

Figure 7: Failure of the *torus* segmentation with the Euclidean distance (bottom right: final state after 600 iterations)

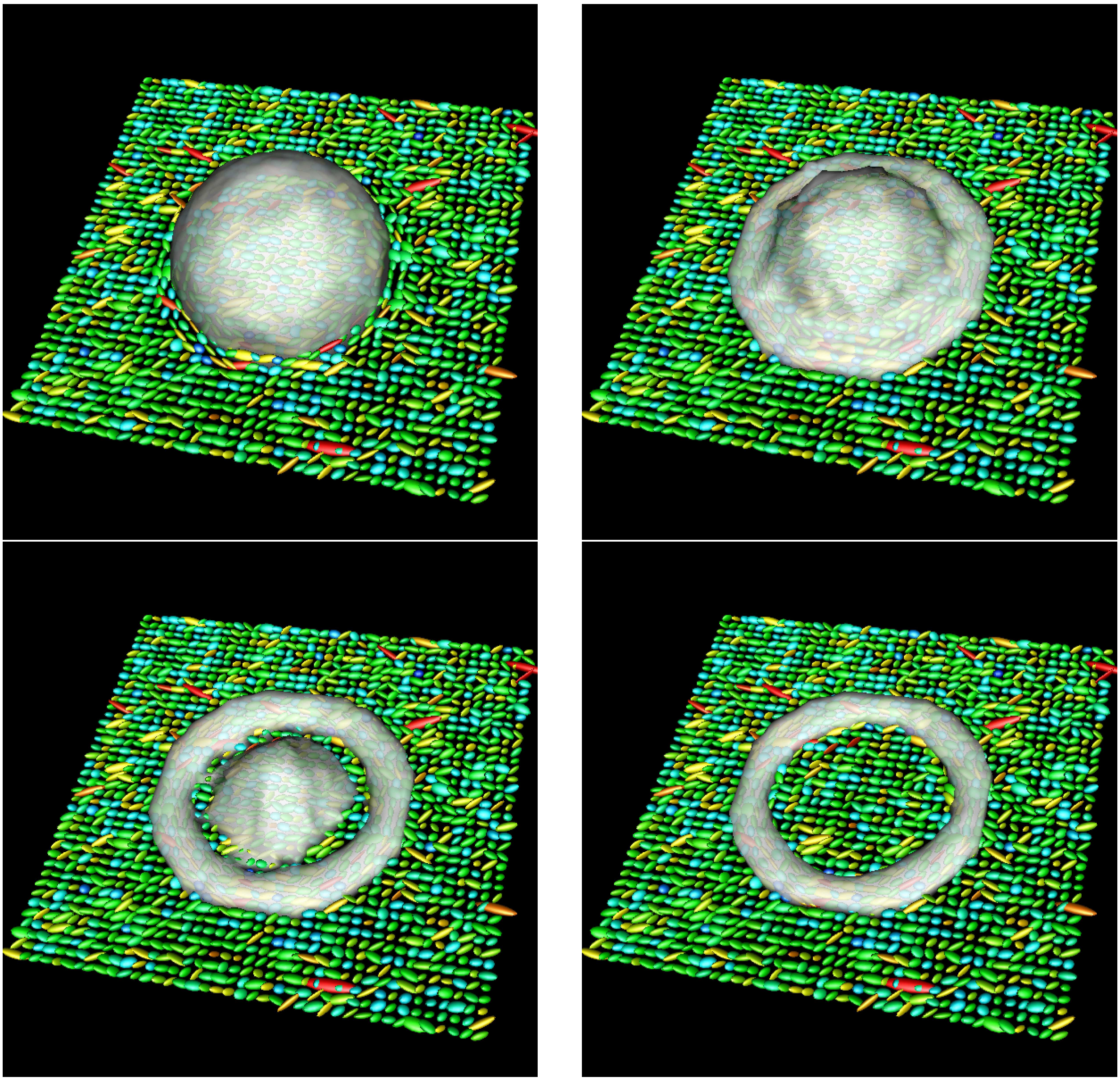

Figure 8: Successful segmentations of the *torus* with the J-divergence and geodesic distances (bottom right: final state after 27 iterations with the J-divergence or 20 iterations with the geodesic distance. The evolutions are similar.)

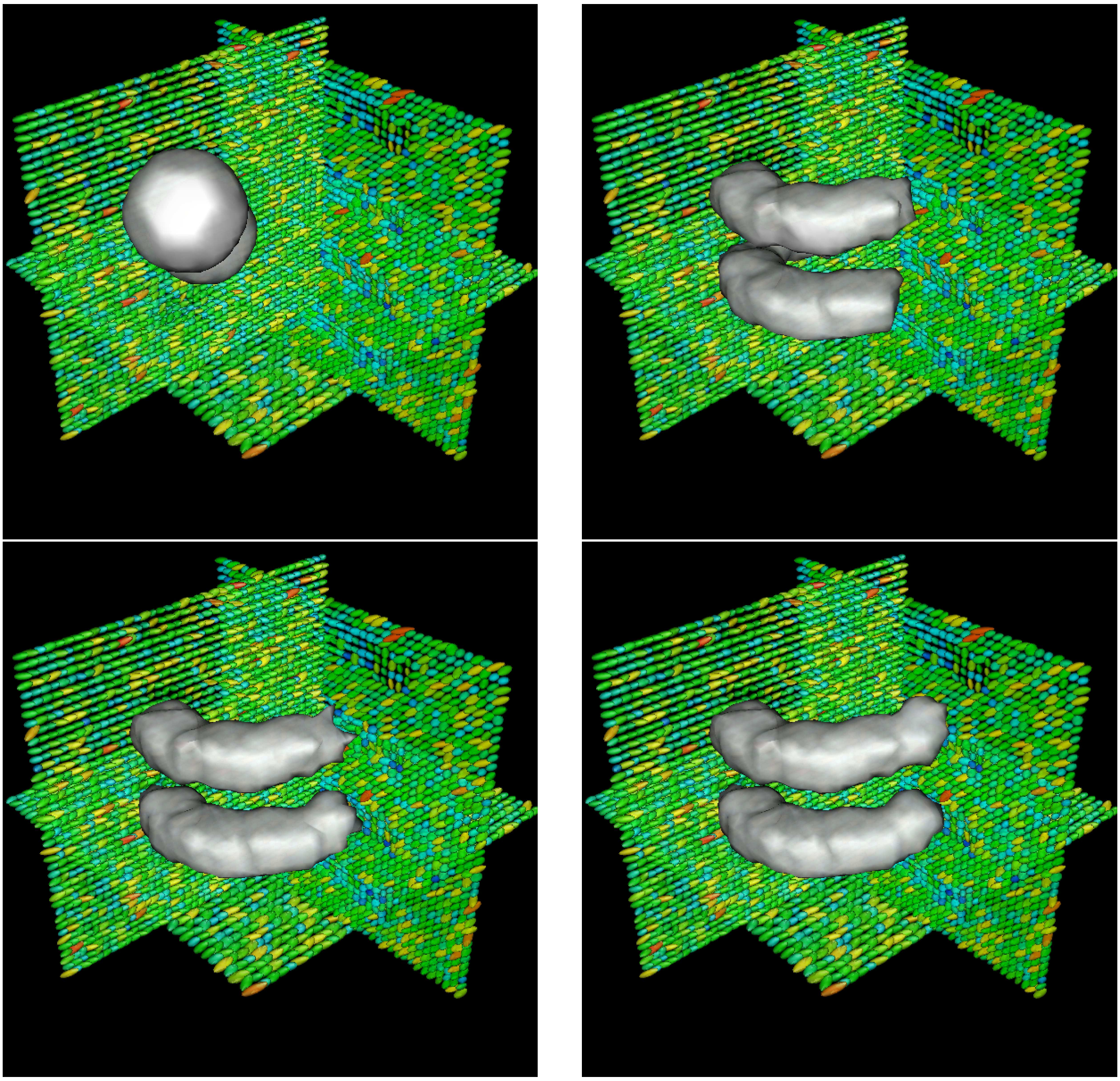

Figure 9: Failure of the *helix* segmentation with the Euclidean distance

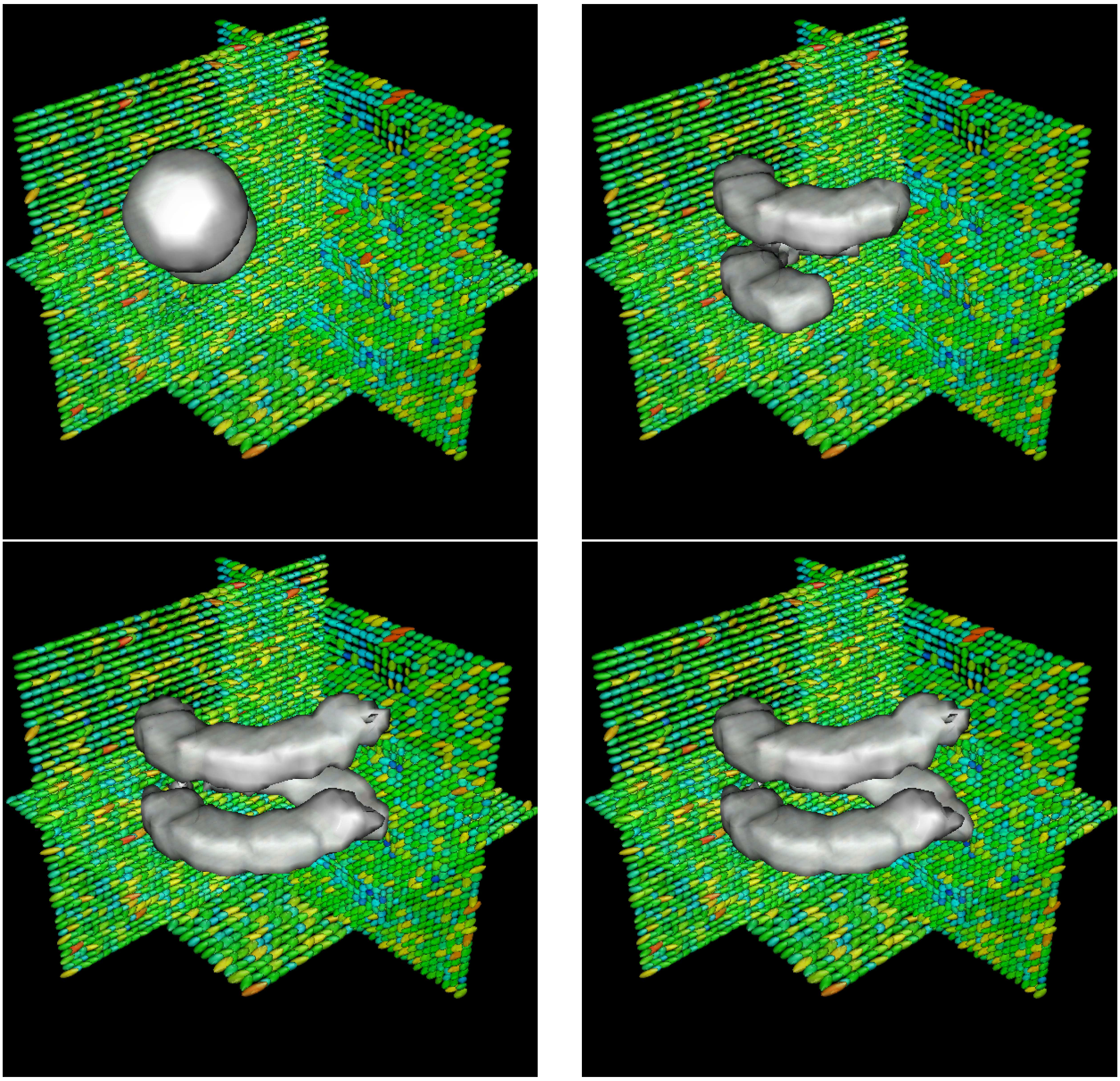

Figure 10: Failure of the *helix* segmentation with the J-divergence

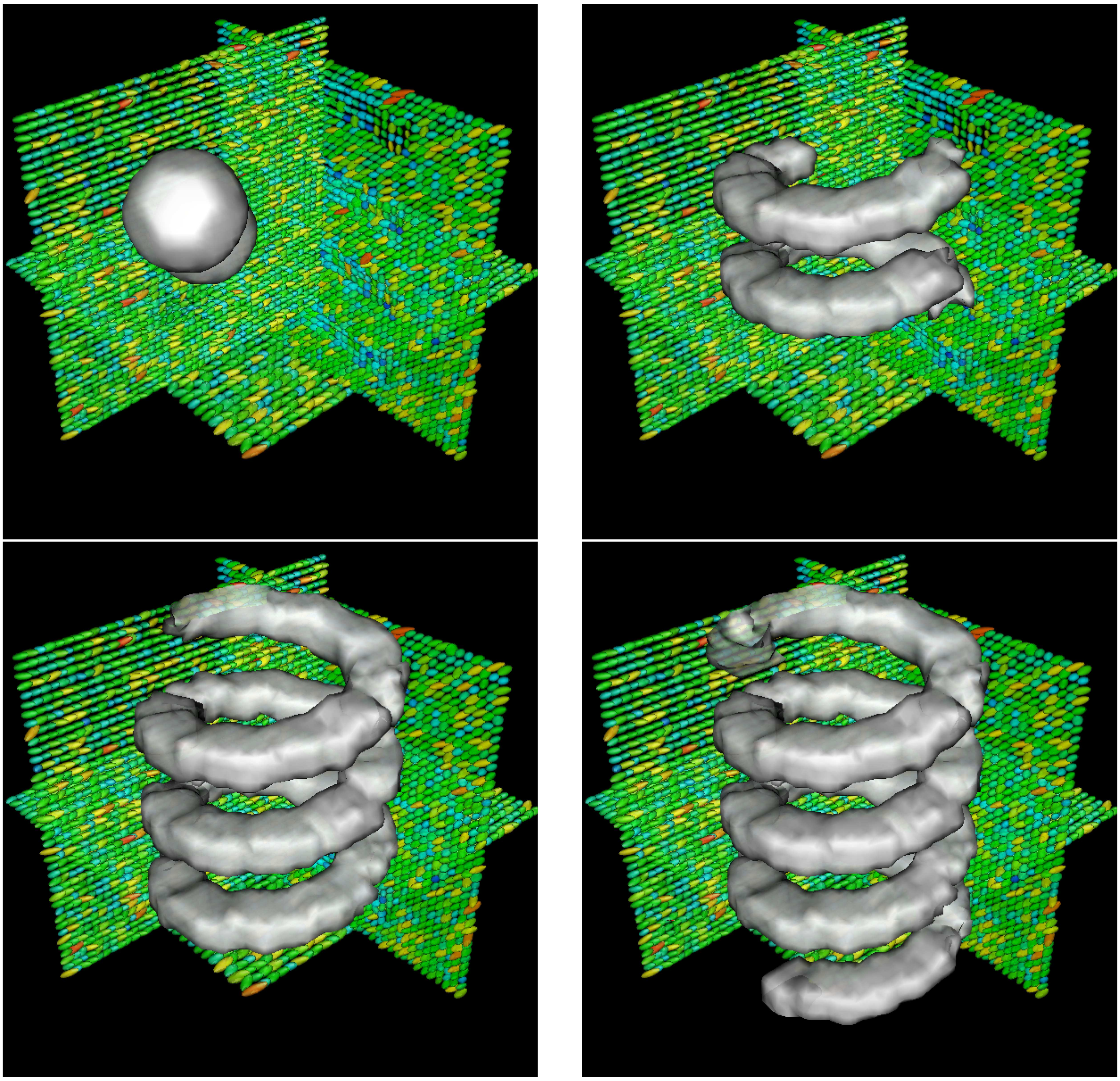

Figure 11: Successful segmentation of the *helix* with the geodesic distance

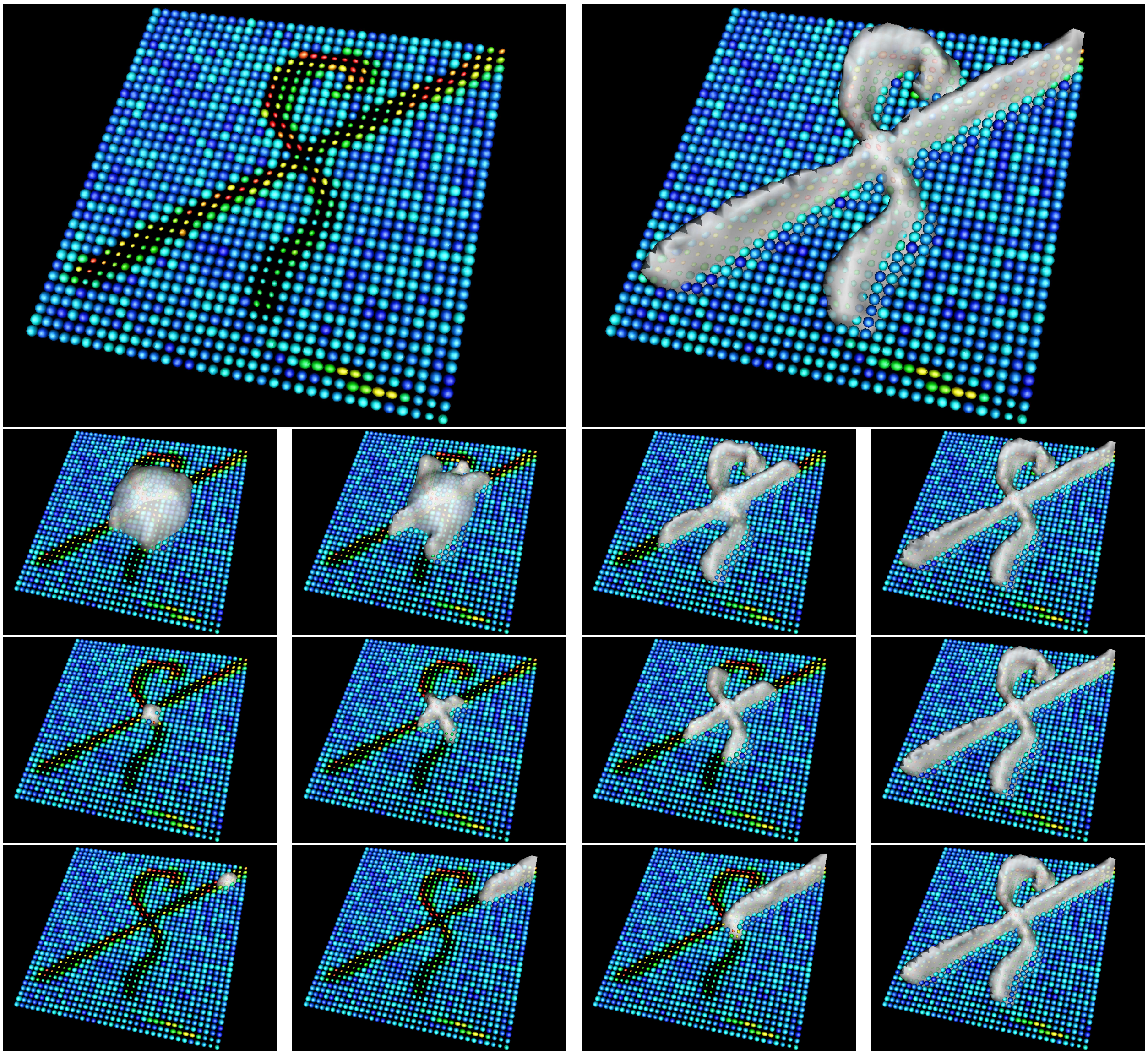

Figure 12: Segmentation of the rat spinal cords phantom. [1st row] Axial slice of the data-set (left) and final segmentation using the geodesic distance (right). [2nd row] Segmentation process with the geodesic distance and large sphere initialization. [3rd row] Segmentation process with the geodesic distance and small sphere initialization. [4th row] Segmentation process with the geodesic distance and initialization at one end of a cord.

This last example undoubtedly demonstrates the superiority of the Riemannian framework over the statistics derived from the Euclidean or Kullback-Leibler dissimilarity measures.

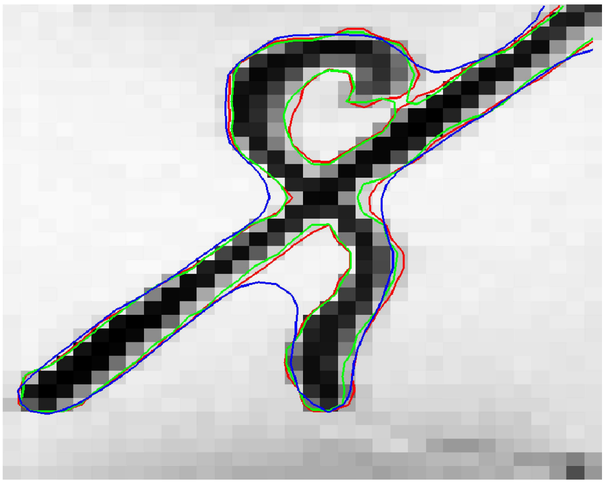

Figure 13: Comparison of the rat spinal cords phantom segmentation results with the 3 distances (colors as on figure 18)

4.2 Biological phantom data-set

We tested our algorithm on a biological phantom produced by J. Campbell et al. at the McConnell Brain Imaging Center and Montreal Neurological Institute [13, 14] and was created from two excised Sprague-Dawley rat spinal cords embedded in 2% agar. Diffusion weighted images (DWI) were acquired along 90 gradient directions with $b\,\text{value} = 1000$ s.mm^{-2} on a 1.5 Tesla Siemens Sonata scanner using a knee coil. Diffusion tensors are computed by using the method proposed in [38]. It relies on the minimization of an energy functional derived from the linearized Stejskal-Tanner equation [57] while ensuring to remain in $S^+(3)$. An axial slice of the resulting DT image is presented in the first row of figure 12 together with a 3D surface modeling the spinal cords.

This data-set is well suited to evaluate the robustness to the initialization of our segmentation framework as well as to demonstrate the importance of the Riemannian framework to achieve good segmentation results.
The second, third and fourth rows of figure 12 illustrate the evolution of the segmentation process, using the geodesic distance, for 3 very different initializations: One large sphere and one small sphere centered at the cord crossing, and one small sphere placed at one end of a cord. These three examples yield the same final result, thus experimentally showing the non-dependence of our method on the initialization. Finally, figure 13 displays, on top of the Apparent Diffusion Coefficient image, the three final segmentation results obtained by using the Euclidean distance (blue), J-divergence (green) and geodesic distance (red). We can see that the most accurate result is obtained with the latter. Especially, it is interesting to note that, in the upper right part of the image where the two cords are very close to each other, only the geodesic distance is able to distinguish between the two structures. This is another example of the better properties of the Riemannian statistics to model the distribution of the diffusion tensors.
In the next section, we will show that the Riemannian statistical approach also performs better on human brain diffusion tensor images.

4.3 Real DTI data-sets

4.3.1 Method and tensors estimation

Diffusion weighted images were acquired at the Center for Magnetic Resonance Research, University of Minnesota, on a 3 Tesla Siemens Magnetom Trio whole-body clinical scanner. Measurements were made along 12 gradient directions. Acquisition parameters were: $b\,\text{value} = 1000$ s.mm^{-2}, TE = 92 ms and TR = 1.7 s. The images were obtained on 64 evenly spaced axial planes with 128×128 pixels per slice. The voxel size is $2 \times 2 \times 2$ mm. As for the biological rat spinal cord phantom, diffusion tensors are computed by using the method proposed in [38]. An example of the resulting DT image is presented on figure 14. It uses a red-green-blue color scheme to encode the tracts orientation. Following [59] and [30], we indicate the names of major fiber bundles.

4.3.2 Performance of the probability metrics

In order to further compare the performance of the three probability metrics, within our segmentation framework, we have experimented with the extraction of the corpus callosum from a given DTI data-set. This important structure corresponds to the so-called callosal radiations which connect homologous areas of each hemisphere. It can be roughly divided into three main parts known as the genu (*gcc*), body (*bcc*) and splenium (*scc*). The genu radiates into the prefrontal, orbital and inferior premotor areas to form the forceps minor. The body of the corpus callosum radiates into the premotor, motor and supplementary motor cortical areas. Finally, the splenium radiates into the inferior/superior temporal, occipital and posterior parietal regions to form the forceps major.
It turns out that, near the midsagittal plane, all the fibers follow the same right-left orientation pattern making it quite easy to extract this structure from anatomical MRI (see figure 15 [left]). This has been used in group studies [45] to investigate architectural variability of the corpus callosum in relation with pathologies like schizophrenia. However, as we can see on figure 15 [right], once we get away from the midsagittal planes the callosal radiations quickly merge within the white matter and cannot be segmented anymore. We show that our Riemannian segmentation framework is able to provide more accurate segmentations of the corpus callosum.
The initialization is obtained either by a quick and approximate delineation of the genu and splenium on only 2 axial slices (figure 16) or by a simple sphere of radius 8 voxels centered in the middle of the body of the corpus callosum. In both cases, results are identical and presented on figure 18 and 19. It is obvious that there is a clear improvement of the segmentation quality (especially in the region of the splenium, figure 18) when moving from the Euclidean distance to the J-divergence and it is much better when the statistics are computed with the geodesic distance.
The splenium of the corpus callosum is almost entirely recovered by the Riemannian approach while it is barely visible with the Euclidean method and only partially extracted when using the J-divergence. We noticed moreover than the Euclidean approach has a tendency to misclassify some tensors from the ventricles. This means that the statistics are not enough discriminant and even take over the boundary term at some locations. The geodesic distance definitely yields the best results.

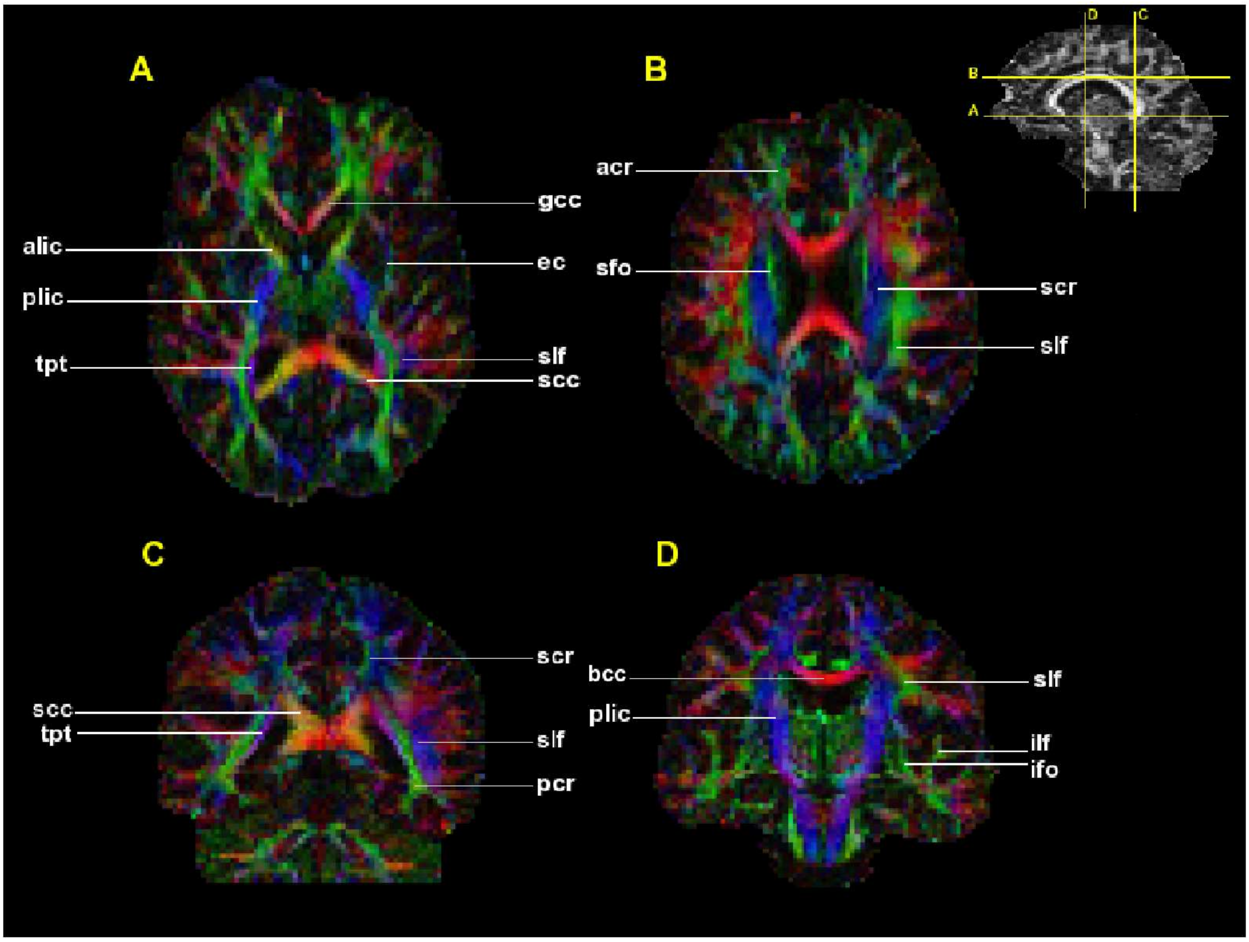

Figure 14: Axial (A,B) and coronal (C,D) DT images color map and major fiber bundles. *acr* = anterior region of the corona radiata, *alic* = anterior limb of the internal capsule, *bcc* = body of the corpus callosum, *ec* = external capsule, *gcc* = genu of the corpus callosum, *ilf* = inferior longitudinal fasciculus, *ifo* = inferior fronto-occipital fasciculus, *pcr* = posterior region of the corona radiata, *plic* = posterior limb of the internal capsule, *scc* = splenium of the corpus callosum, *scr* = superior region of the corona radiata, *sfo* = superior fronto-occipital fasciculus, *slf* = superior longitudinal fasciculus, *tpt* = tapetum

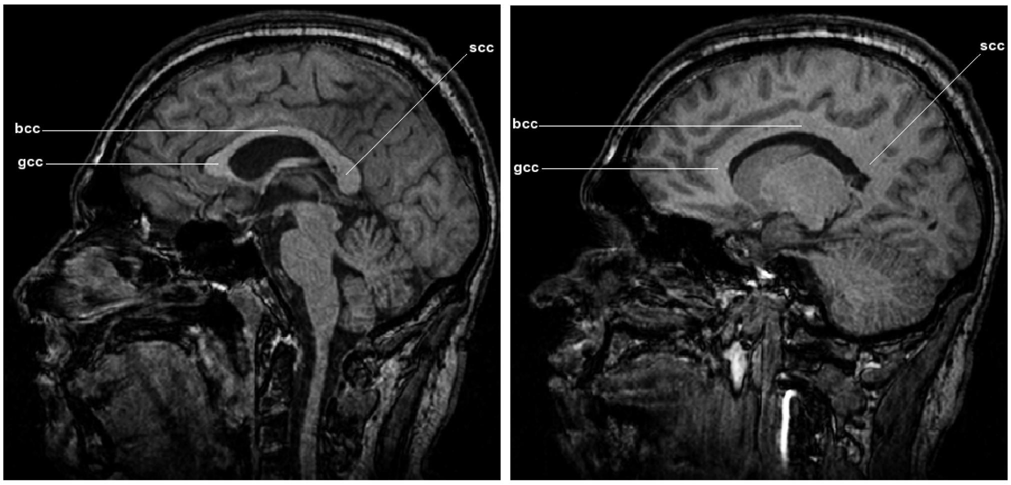

Figure 15: Corpus callosum on a midsagittal (left) and sagittal (right) slice from a T1 MRI

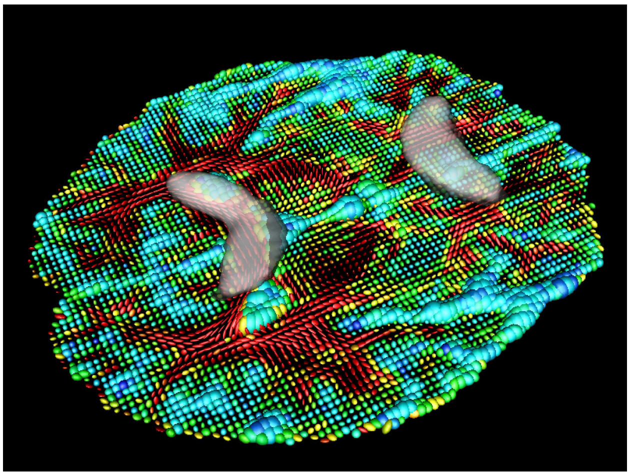

Figure 16: Initialization of the corpus callosum segmentation by an approximate delineation of the genu and splenium

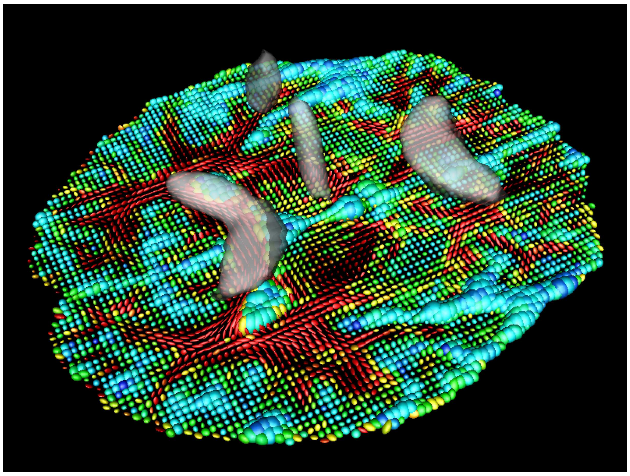

Figure 17: Initialization of the corpus callosum segmentation with added tensors from the superior region of the corona radiata

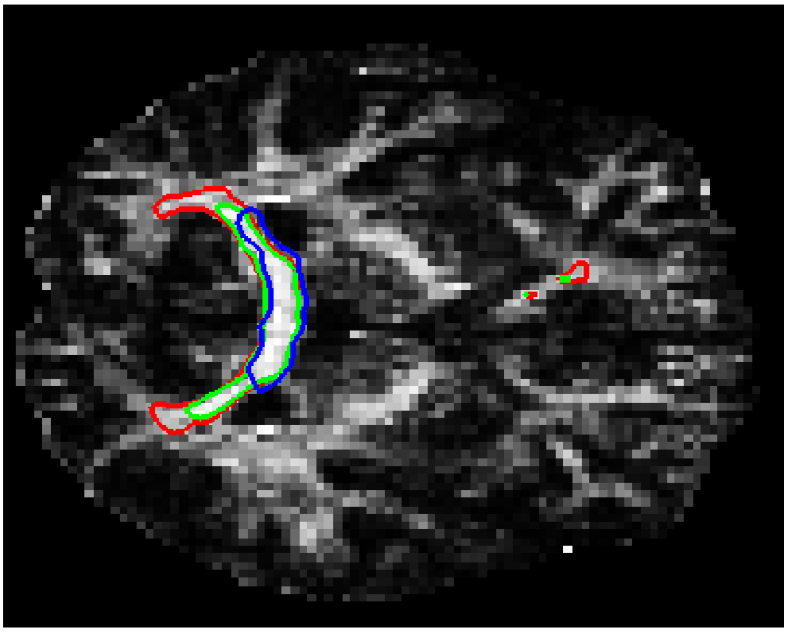

Figure 18: Segmentation results in the region of the splenium (blue: Euclidean distance, green: J-divergence, red: geodesic distance)

4.3.3 Multiple fiber bundles segmentation

We conclude our numerical experiments on human brain DTI by trying to also recover fibers from the corona radiata, which is known to merge with the corpus callosum. The initialization is presented on figure 17 and is meant to include some tensors from the superior part of the corona radiata (*scr*). To that end we simply added tensors of the *scr* on 2 coronal slices. It turns out that, with the Euclidean distance and J-divergence, these new tensors quickly disappear from the segmentation and the final results are the same as those presented on figure 19. This is not suprising and proves that the associated statistics do not constitute accurate descriptors of the tensors distribution. On the other side, the statistics computed with the geodesic distance make it possible to perform the desired segmentation, as presented on figure 20. This is a very interesting result since the superior part of the corona radiata is partially recovered. But more importantly, fiber tracts which are known to mingle with the callosal radiations are also segmented. It is indeed well-known that the corpus callosum merges with association and projection fibers as its gets toward the cortex. We can see on figure 20 that the tapetum, the posterior region of the corona radiata and a part of the superior longitudinal fasciculus are extracted since they fuse with the splenium of the corpus callosum. The posterior limb of the internal capsule (essentially the corticospinal tract) is equally segmented since it intersects with the corpus callosum and with the superior longitudinal fasciculus in the region of the centrum semiovale. All these results contribute to clearly validate our claim that the proposed Riemannian framework achieves the best segmentation results.

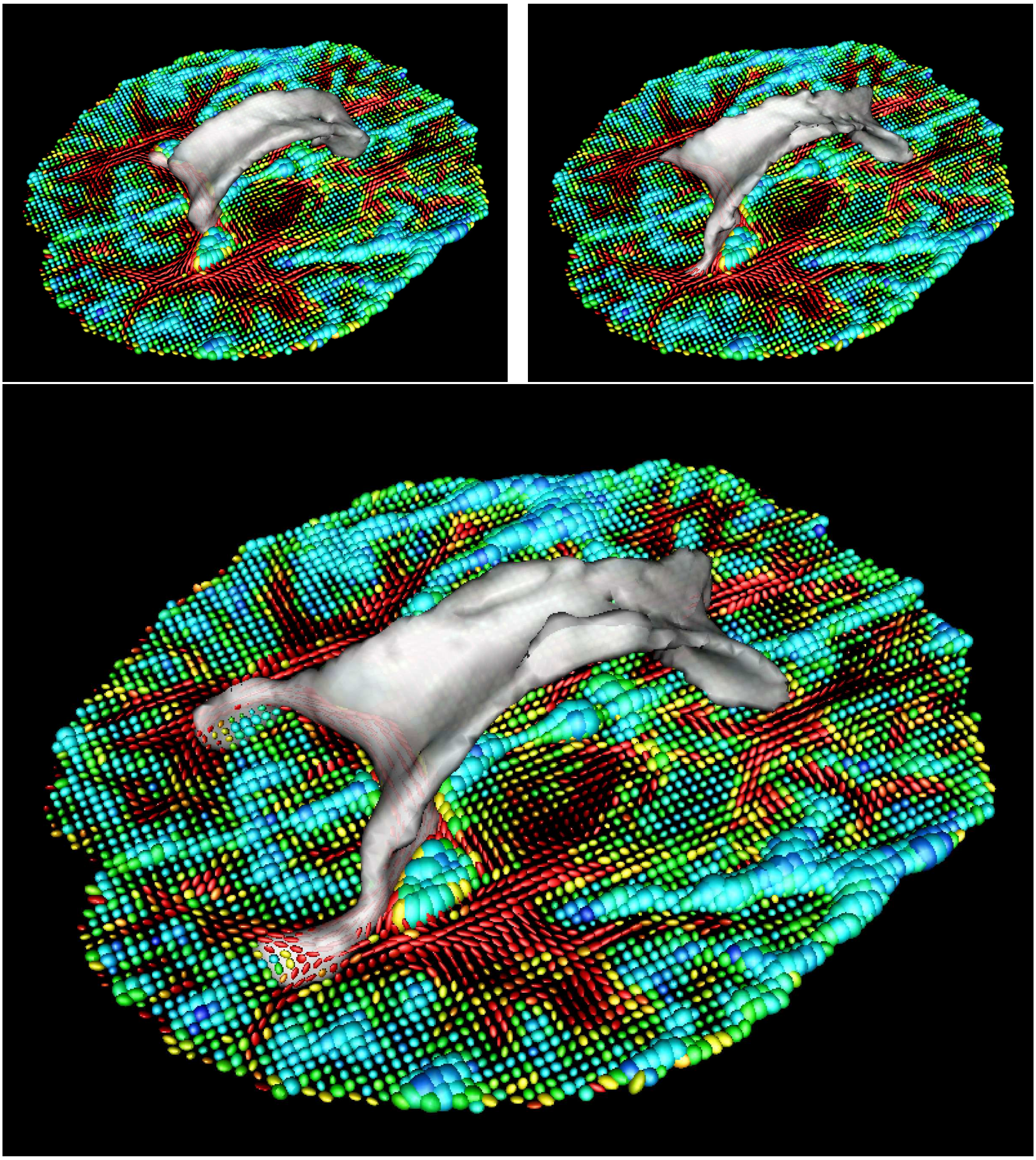

Figure 19: Segmentation of the corpus callosum using the Euclidean distance (top left), J-divergence (top right), and geodesic distance (bottom)

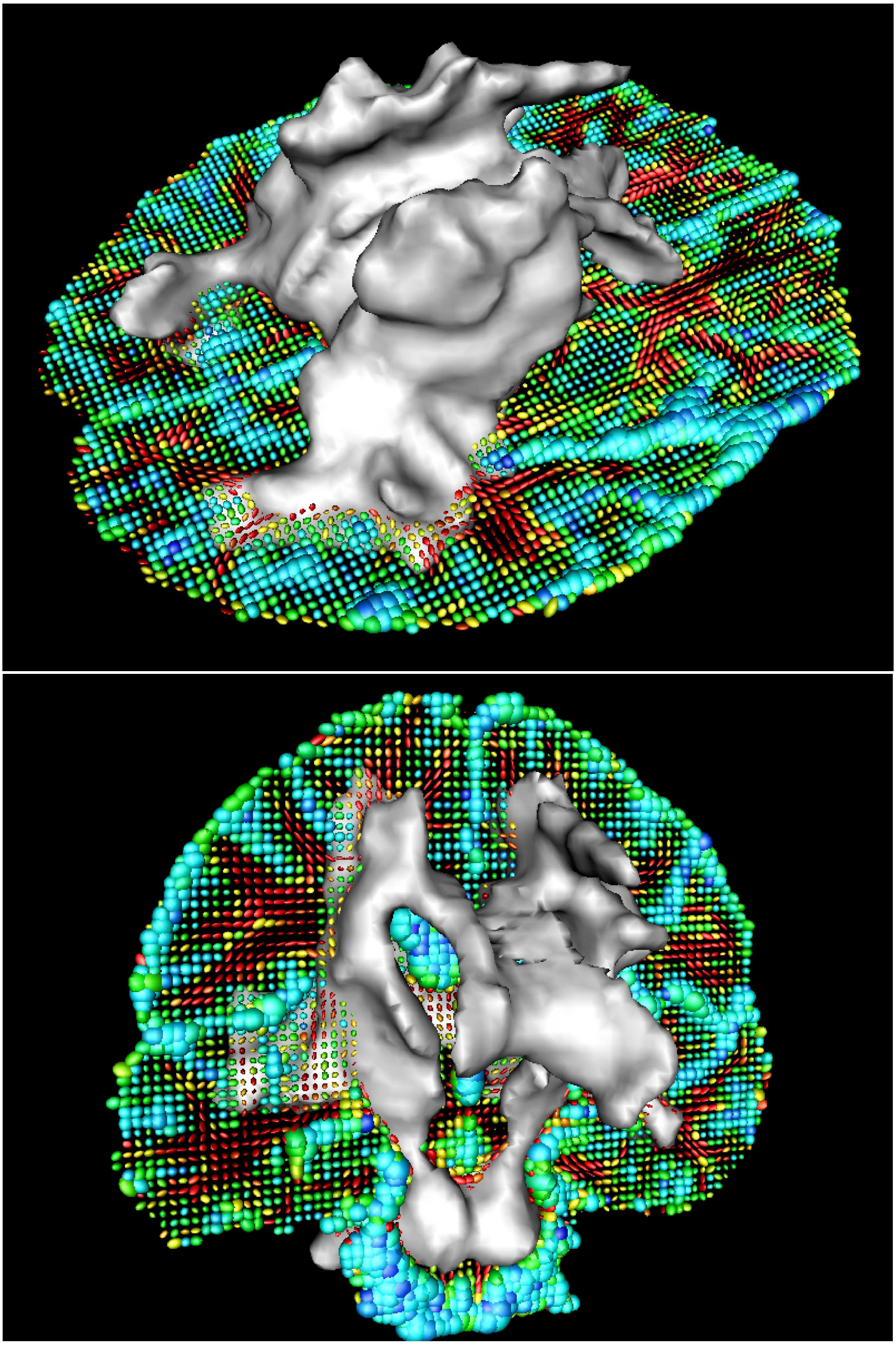

Figure 20: Segmentation of the corpus callosum and intermingling fiber tracts

5 Conclusion

We have presented a unified statistical surface evolution framework for the segmentation of diffusion tensor images. Since a diffusion tensor can be understood as the covariance matrix of a three-dimensional normal distribution with zero mean, we have introduced various probability metrics (Euclidean distance, J-divergence and geodesic distances), i.e. dissimilarity measures between probability density functions, to derive statistics on DT images. These statistical parameters (mean and covariance matrix) allowed us to define a notion of Gaussian density for diffusion tensors, depending on the probability metric, which was used to model the distribution of a set of tensors. Finally, we have shown how to estimate the norm of the spatial gradient of a DT image in order to detect boundaries between structures of the white matter. By fusing these statistical and geometrical information within a variational framework, we derived a powerful level-set based DTI segmentation technique. At this point, our claim was that the special properties of the space of 3×3 diffusion tensors (symmetry and positivity) were naturally handled by working in the Riemannian framework. It must consequently yield more adequate tools to deal with tensors than the Euclidean or J-divergence approaches. The former, by seeing $S^+(3)$ as a linear space is completely blind to its curvature. The latter was shown to be equivalent to the geodesic distance only for infinitesimally close tensors. The Riemannian framework was proposed to derive the proper tools to work within the space of 3×3 diffusion tensors while taking into account its special properties.We proved that the choice of the probability metric, i.e. the dissimilarity measure, has a deep impact on the tensor statistics and, hence, on the segmentation results.

Through numerical experiments on synthetic data-sets, a biological rat spinal cord phantom, as well as on human brain DT images, we could experimentally demonstrate the superiority of the geodesic probability metric over the J-divergence which, in turn, performed better than the Euclidean distance. This order was found on synthetic data-sets with increasing complexity and for which, ground truth being known, it was very easy to undoubtedly evaluate the quality of the segmentations. The biological phantom data-set, because of its known and relatively simple geometry, allowed to test the robustness to the initialization of our algorithm and, again, to demonstrate on a single realistic data-set that the best results were obtained with the geodesic distance. Finally, on human brain DTI data-sets, the Riemannian approach was the only one capable of correctly segmenting highly variable tensor fields. It achieved better results than the other metrics (Euclidean distance and J-divergence), by comparison with neuroanatomical knowledge, for the segmentation of the corpus callosum or the corticospinal tract.

Acknowledgments

This research was partially supported by grants NSF-0404617 US-France (INRIA) Cooperative Research, NIH R21-RR019771, NIH-RR008079, the MIND Institute and the Keck foundation, the INRIA-FQNRT program, ACI Obs-Cerv and the Région Provence-Alpes-Côte d'Azur. The authors would like to thank S. Lehéricy and K. Ugurbil (Center for Magnetic Resonance Research, University of Minnesota) for the human brain DTI data-sets used in this paper. We also gratefully acknowledge J. Campbell, K. Siddiqi, V.V. Rymar and B. Pike (McConnell Brain Imaging Center, McGill University) as well as M. Descoteaux for the rat data-set.

References

[1] B.A. Ardekani, J. Nierenberg, M.J. Hoptman, D.C. Javitt, and K.O. Lim. MRI study of white matter diffusion anisotropy in schizophrenia. *NeuroReport*, 14(16):2025–2029, 2003.

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

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

[4] C. Ballester, V. Caselles, and M. González. Affine invariant segmentation by variational method. *SIAM Journal of Applied Mathematics*, 56:294–325, 1994.

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

[6] T.E.J. Behrens, H. Johansen-Berg, M.W. Woolrich, S.M. Smith, C.A.M. Wheeler-Kingshott, P.A. Boulby, G.J. Barker, E.L. Sillery, K. Sheehan, O. Ciccarelli, A.J. Thompson, J.M. Brady, and P.M. Matthews. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature Neuroscience*, 6(7):750–757, 2003.

[7] 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*, 401–407, 1986.

[8] T. Brox, J. Weickert, B. Burgeth, and P. Mrázek. Nonlinear structure tensors Preprint No. 113, Department of Mathematics, Saarland University, Saarbrücken, Germany, October 2004

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

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

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

[12] J. Campbell, K. Siddiqi, and B. Pike. A Geometric flow for white matter fibre tract reconstruction. In *Proc. IEEE International Symposium in Biomedical Imaging*, 505–508, 2002.

[13] J. Campbell. Diffusion imaging of white matter fibers tracts. PhD thesis, McGill University, 2004.

[14] J.S.W. Campbell, K. Siddiqi, V.V. Rymar, A.F. Sadikot, and G.B. Pike. Flow-based fiber tracking with diffusion tensor and Q-ball data: Validation and comparison to principal diffusion direction techniques. *NeuroImage*, 27(4):725–736, October 2005.

[15] V. Caselles, R. Kimmel, G. Sapiro Geodesic active contours. *International Journal of Computer Vision*, 22:61–79, 1997.

[16] T. Chan and L. Vese. Active contour without edges. *IEEE Transactions on Image Processing*, 10:266–277, 2001.

[17] 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.

[18] I. Corouge, S. Gouttard, and G. Gerig. A statistical shape model of individual fiber tracts extracted from diffusion tensor MRI. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention*, 671–679, 2004.

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

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

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

[22] M.P. Do Carmo. Riemannian geometry. Mathematics: Theory and Applications, Birkhäuser

[23] P.S. Eriksen. Geodesics connected with the fisher metric on the multivariate manifold. Research Report 86-13, Institute of Electronic Systems, Aalborg University, 1986.

[24] C. Feddern, J. Weickert, and B. Burgeth Level-set methods for tensor-valued images. In *Proc. IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision*, 65–72, 2003.

[25] C. Feddern, J. Weickert, B. Burgeth and M. Welk Curvature-driven PDE methods for matrix-valued images. Technical Report No. 104, Department of Mathematics, Saarland University, Saarbrücken, Germany, April 2004.

[26] P.T. Fletcher and S. Joshi. Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In *Proc. Computer Vision Approaches to Medical Image Analysis, ECCV Workshop*, 87–98, 2004.

[27] W. Förstner and B. Moonen. A metric for covariance matrices. Research Report, Stuttgart University, Dept. of Geodesy and Geoinformatics, 1999.

[28] M. Fréchet. Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institut Henri Poincaré*, X(IV):215–310, 1948.

[29] H. Jeffreys Theory of probability. Wiley, New York, 1959.

[30] B.J. Jellisona, A.S. Fielda, J. Medowb, M. Lazarc, M.S. Salamatd, and A.L. Alexander. Diffusion tensor imaging of cerebral white matter: A pictorial review of physics, fiber tract anatomy, and tumor imaging patterns. *American Journal of Neuroradiology*, 25:356–360, 2004.

[31] L. Jonasson, X. Bresson, P. Hagmann, O. Cuisenaire, R. Meuli and J.P. Thiran. White matter fiber tract segmentation in DT-MRI using geometric flows. *Medical Image Analysis*, 9(3):223–236, June 2005.

[32] H. Karcher. Riemannian centre of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30:509–541, 1977.

[33] M. Kass, A. Witkin, D. Terzopoulos Snakes: Active contour models. In *Proc. International Conference on Computer Vision*, 259–268, 1987.

[34] Y. Leclerc. Constructing simple stable description for image partitioning. *International Journal of Computer Vision*, 3:73–102, 1989.

[35] C. Lenglet, R. Deriche, and O. Faugeras. Inferring white matter geometry from diffusion tensor MRI: Application to connectivity mapping. In *Proc. European Conference on Computer Vision*, 4:127–140, 2004.

[36] C. Lenglet, M. Rousson, and R. Deriche. Segmentation of 3D probability density fields by surface evolution: Application to diffusion MRI. In *Proc. International Conference on Medical Image Computing and Computer Assisted Intervention*, 18–25, 2004.

[37] 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 5242, INRIA, June 2004.

[38] C. Lenglet, R. Deriche, O. Faugeras, S. Lehéricy, and K. Ugurbil. A Riemannian approach to diffusion tensors estimation and streamline-based fiber tracking. In *Proc. Annual Meeting of the Organization for Human Brain Mapping*, 2005.

[39] C. Lenglet, M. Rousson, R. Deriche, O. Faugeras, S. Lehéricy, and K. Ugurbil. A Riemannian approach to diffusion tensor images segmentation. In *Proc. Information Processing in Medical Imaging*, 591–602, 2005.

[40] C. Lenglet, M. Rousson, R. Deriche, and O. Faugeras. Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing. *Journal of Mathematical Imaging and Vision, Special Issue MIA*, (In press) 2006.

[41] J.F. 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.

[42] K.D. Merboldt, W. Hanicke, and J. Frahm. Self-diffusion NMR imaging using stimulated echoes. *Journal of Magnetic Resonance*, 64:479–486, 1985.

[43] M. Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747, 2005.

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

[45] K.L. Narr, P.M. Thompson, T. Sharma, J. Moussai, A.F. Cannestra, and A.W. Toga. Mapping morphology of the corpus callosum in schizophrenia. *Cerebral Cortex*, 10(1):40–49, 2000.

[46] K. Nomizu. Invariant affine connections on homogeneous spaces. *American Journal of Mathematics*, 76:33–65, 1954.

[47] 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.

[48] N. Paragios and R. Deriche. Geodesic active regions: a new paradigm to deal with frame partition problems in computer vision. *Journal of Visual Communication and Image Representation*, 13(1-2):249–268, 2002.

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

[50] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41-66, 2006.

[51] C.R. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37:81–91, 1945.

[52] M. Rousson. *Cues Integrations and Front Evolutions in Image Segmentation*. PhD thesis, Université de Nice-Sophia Antipolis, December 2004.

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

[54] M. Rousson, C. Lenglet, and R. Deriche. Level set and region based surface propagation for diffusion tensor MRI segmentation. In *Proc. Computer Vision Approaches to Medical Image Analysis, ECCV Workshop*, 123–134, 2004.

[55] L. Skovgaard. A Riemannian geometry of the multivariate normal model. Technical Report 81/3, Statistical Research Unit, Danish Medical Research Council, Danish Social Science Research Council, 1981.

[56] C. Sotak. The role of diffusion tensor imaging (DTI) in the evaluation of ischemic brain injury. *NMR in Biomedicine*, 15:561–569, 2002.

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

[58] D. Tschumperlé and R. Deriche. Variational frameworks for DT-MRI estimation, regularization and visualization. In *Proc. International Conference on Computer Vision*, 116–122, 2003.

[59] S. Wakana, H. Jiang, L.M. Nagae-Poetscher, P.C. van Zijl, S. Mori. Fiber tract-based atlas of human white matter anatomy. *Radiology*, 230(1):77–87, 2004

[60] Z. Wang, B.C. Vemuri, Y. Chen, and T. Mareci. Simultaneous smoothing and estimation of the tensor field from diffusion tensor MRI. In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, 461–466, 2003.

[61] Z. Wang and B.C. Vemuri. Tensor field segmentation using region based active contour model. In *Proc. European Conference on Computer Vision*, 4:304–315, 2004.

[62] Z. Wang and B.C. Vemuri. An affine invariant tensor dissimilarity measure and its application to tensor-valued image segmentation. In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, 228–233, 2004.

[63] Z. Wang, B.C. Vemuri, Y. Chen, and T.H. 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.

[64] 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.

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

[66] M.R. Wiegell, D.S. Tuch, H.W.B. Larson, and V.J. Wedeen. Automatic segmentation of thalamic nuclei from diffusion tensor magnetic resonance imaging. *NeuroImage*, 19:391–402, 2003.

[67] S. Yoshizawa and K. Tanabe Dual differential geometry associated with the Kullback-Leibler information on the Gaussian distributions and its 2-parameter deformations Science University of Tokyo Journal of Mathematics, 35:113–137, 1999.

[68] H. Zhao, T. Chan, B. Merriman, and S. Osher. A variational level set approach to multiphase motion. *Journal of Computational Physics*, 127:179–195, 1996.

[69] Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57, 2001.

[70] S. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:884–900, 1996.

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

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399