CrossMark

Integral curves from noisy diffusion MRI data with closed-form uncertainty estimates

Owen Carmichael[1] · Lyudmila Sakhanenko[2]

© Springer Science+Business Media Dordrecht 2015

Abstract We present a method for estimating the trajectories of axon fibers through diffusion tensor MRI (DTI) data that provides theoretically rigorous estimates of trajectory uncertainty. We develop a three-step estimation procedure based on a kernel estimator for a tensor field based on the raw DTI measurements, followed by a plug-in estimator for the leading eigenvectors of the tensors, and a plug-in estimator for integral curves through the resulting vector field. The integral curve estimator is asymptotically normal; the covariance of the limiting Gaussian process allows us to construct confidence ellipsoids for fixed points along the curve. Complete trajectories of fibers are assembled by stopping integral curve tracing at locations with multiple viable leading eigenvector directions and tracing a new curve along each direction. Unlike probabilistic tractography approaches to this problem, we provide a rigorous, theoretically sound model of measurement uncertainty as it propagates from the raw MRI data, to the tensor field, to the vector field, to the integral curves. In addition, trajectory uncertainty is estimated in closed form while probabilistic tractography relies on sampling the space of tensors, vectors, or curves. We show that our estimator provides more realistic trajectory uncertainty estimates than a more simplified prior approach for closed-form trajectory uncertainty estimation due to Koltchinskii et al. (Ann Stat 35:1576–1607, 2007) and a popular probabilistic tractography method due to Behrens et al. (Magn Reson Med 50:1077–1088, 2003) using theory, simulation, and real DTI scans.

✉ Lyudmila Sakhanenko
  luda@stt.msu.edu

  Owen Carmichael
  owen.carmichael@pbrc.edu

[1] Pennington Biomedical Research Center, 6400 Perkins Road, Baton Rouge, LA 70808, USA

[2] Department of Statistics and Probability, Michigan State University, 619 Red Cedar Road, Wells Hall, East Lansing, MI 48824, USA

1 Introduction

Let G be a bounded set in $\mathbb{R}^d$, $d > 1$, and let $M(x), x \in G$, be a family of symmetric
positive definite matrices, which are unknown. The main goal of this work is to estimate the
integral curve $x(t), t \geq 0$, starting at a known fixed point $a \in G$ and driven by a vector field
$v(M(x)), x \in G$, based on the observations

$$\{(X_j, Y(X_j)), j = 1, \ldots, n\}, \quad Y(X_j) = BM(X_j) + \Sigma^{1/2}(X_j)\Xi_j,$$

where $X_j, j = 1, \ldots, n$, are discrete locations in G, B is a known matrix, and the second term
on the right represents random errors. For our purpose, the vector field $v(M(x))$ will generally
consist of the leading eigenvectors of the tensors $M(x)$. The integral curve $x(t), t \geq 0$, is
defined as a solution of the ODE

$$\frac{dx(t)}{dt} = v(M(x(t))), \quad t \geq 0, \ x(0) = a \tag{1}$$

or equivalently of the integral equation $x(t) = a + \int_0^t v(M(x(s)))ds$. The other parts of the
model will be explained in the next sections. Below we give the motivation and application
from brain imaging research.

Diffusion tensor magnetic resonance imaging (DTI) is a prominent method for using mag-
netic fields to measure the degree to which water molecules are diffusing along particular 3D
directions at every location in a biological specimen; e.g., see Bammer et al. (2009), Beaulieu
(2002), Chanraud et al. (2010), Mukherjee et al. (2008a, b). Because water is constrained to
diffuse along, but not across, the wire-like axons that carry an electrical charge from neuron
to neuron in the human brain, tracing trajectories along prominent water diffusion directions
through volumetric brain DTI data sets has attracted considerable scientific attention as a
means for mapping the axon architecture of the brain. Given raw DTI data, consisting of
component images that quantify the amount of water diffusion along particular 3D direc-
tions, the fiber trajectory tracing problem for DTI has traditionally been formulated in terms
of estimating a 3×3 positive definite matrix called the diffusion tensor that represents the
statistics of water diffusion over every direction on a 3D sphere; differential equations that
model particle motion through the field of diffusion tensor leading eigenvectors are then
solved (Fig. 1).

However, DTI component images contain a notoriously high amount of noise relative to
the water diffusion signal, resulting in noisy estimation of the diffusion tensor and leading
eigenvectors; see, e.g., Gudbjartsson and Patz (1995), Hahn et al. (2006, 2009), Zhu et al.
(2007, 2009). This in turn causes erroneous trajectory estimates that can in turn lead to
spurious scientific findings; see Basser and Pajevic (2000), Zhu et al. (2009). This problem
led to the development of a class of methods, termed probabilistic tractography, that provides
estimates of uncertainty in fiber trajectories. These approaches are generally based on the
use of sampling to explore the space of possible diffusion tensors, leading eigenvectors, or
fiber trajectories. For example, Bayesian frameworks have been presented for estimation of
parameters such as leading eigenvector directions or trajectory curve characteristics; Markov
chain Monte Carlo (MCMC) is then used to explore the space of parameter values; see, e.g.,
Parker and Alexander (2003), Friman et al. (2006), and Basser et al. (2000). Bootstrap has
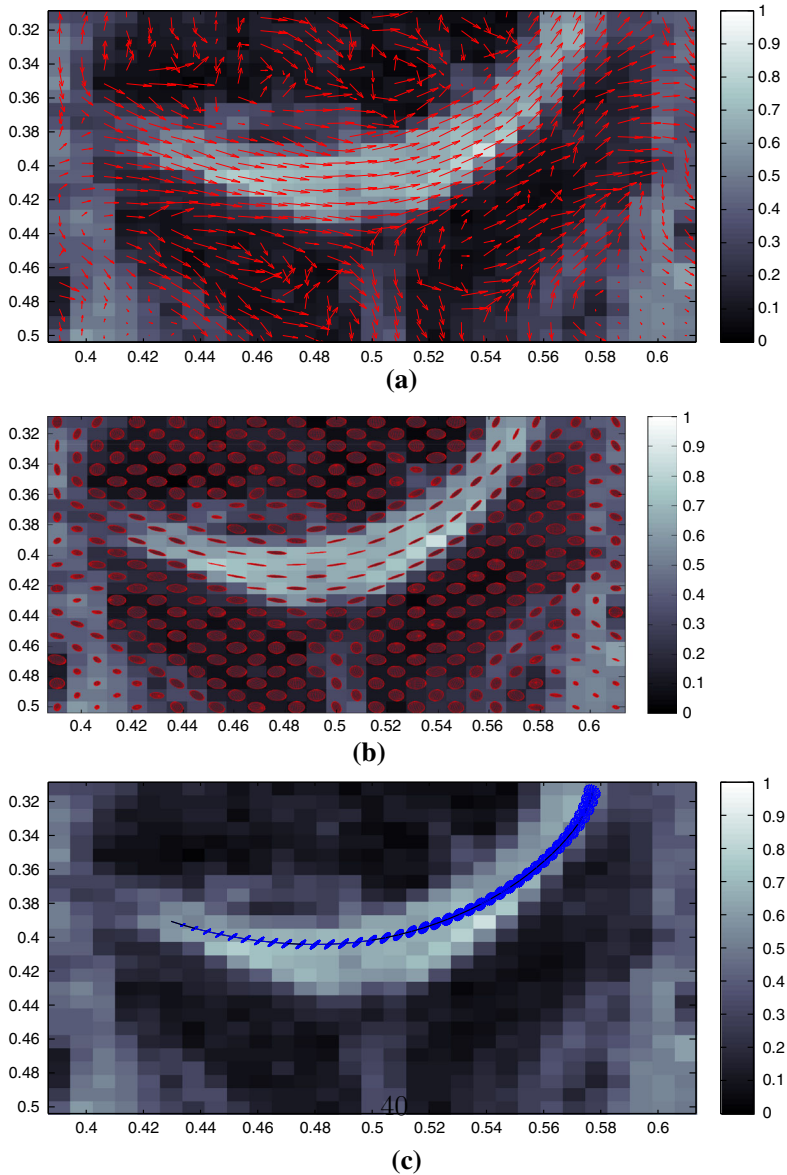also been used to repeatedly simulate novel diffusion tensors or leading eigenvectors based on

Fig. 1 A 2D slice of a 3D DTI data set. DTI provides a 3×3 positive definite matrix at each location in 3D; the leading eigenvector of this matrix (direction shown in *blue*) indicates the most prominent direction along which water molecules are diffusing locally. The fractional anisotropy of the matrix (FA) is a function of the three matrix eigenvalues that gives a univariate summary of how anisotropic, i.e. strongly preferential along a single spatial direction, the water diffusion is (higher FA is whiter). **a** A map of fractional anisotropy (FA) derived from DTI data, with whiter pixels representing locations where the diffusion tensor is dominated by its leading eigenvalue. The FA map is overlaid with the vector field of leading eigenvectors. **b** The FA map is overlaid with the diffusion tensor field represented by ellipses. The major axis of each ellipse is oriented along the leading eigenvector direction and the minor axis is oriented along the second eigenvector direction. The major axis is scaled to have unit length and the minor axis is scaled according to the ratio of the second eigenvalue to the leading eigenvalue. **c** The FA map is overlaid with an example fiber tracing with 95 % confidence ellipsoids based on the proposed method. (Color figure online)

subsets of component images; see, e.g., Jones (2003), Lazar and Alexander (2005), Behrens et al. (2003). Uncertainty is assessed in terms of variability of estimates across bootstrap samples. The key computational limitation of these approaches is the burden of drawing large numbers of samples from high-dimensional spaces. The key theoretical limitations of these approaches are the use of parametric models, and a lack of theoretical justification of why the method would converge, and if it converges how fast. In addition, prior distributions over parameter values required for MCMC need to be determined, usually through heuristic arguments; it is not adequate to simply provide non-informative priors because these assign non-zero prior probabilities to tensors that are not positive definite. See Chap. 8.2 in Gelman et al. (2004) on limitations of the Bayesian framework in this regard. Additionally, bootstrap can be misleading since it resamples from the given model, which could be inappropriate to start with. In order for bootstrap to work, one needs to prove its consistency, which is often equivalent to establishing asymptotic normality, and the functional that makes up the statistic of interest needs to be continuous with respect to the underlying distribution of observations; see, e.g. Mammen (1992) for more on what is needed for bootstrap to work and classical situations when it fails. Whether these necessary conditions are met by DTI has not been explored in depth. See also Yuan et al. (2008) for a discussion of various issues associated with use of bootstrap in DTI.

An alternative to probabilistic tractography is to use smoothing estimators to trace the fiber trajectory, and quantify uncertainty in closed form using rigorous theoretically-driven bounds on errors. For example, Koltchinskii et al. (2007) used a Nadaraya–Watson kernel smoothing estimator for the vector field combined with a plug-in estimator for the integral curve, under an assumption of additive vector field noise. This approach yields asymptotically normal estimators of integral curves as the number of spatial locations grows, it enjoys the optimal rate of convergence locally and globally, and provides closed-form estimates of integral curve uncertainty through the covariance of a limiting Gaussian process; see Sakhanenko (2010, 2011). This approach would need to identify locations along the integral curve trajectory where multiple distinct principal eigenvector directions appear viable, as is common in DTI data, and begin new integral curve trajectories along each of these directions. Complete characterizations of uncertainty in fiber trajectories are then assembled by linking together all such branching integral curves together with their uncertainty estimates. The key limitation of this approach is that the input to fiber tracing is assumed to be a vector field of "true" diffusion tensor leading eigenvectors that have been perturbed by an additive noise, thus ignoring the possibility that the estimation error in the diffusion tensor itself may give rise to more complex noise structures in the leading eigenvector field.

This paper provides an integral curve estimator that gives a more realistic and complete account of how noise in DTI data impacts fiber trajectory estimation by modeling how noise in the component images enters into the estimation of diffusion tensors and leading eigenvectors. We show that under certain geometric and smoothness assumptions the properly normalized difference between our estimator and the true integral curve, as a random process, converges weakly to a Gaussian process. This allows us to provide an asymptotically normal estimator of the integral curve at a point and to use the covariance matrix of the Gaussian process to construct confidence ellipsoids for fixed points along the integral curve. We then show that the Koltchinskii et al. estimator only converges to this Gaussian process in very limited, unrealistic situations: when the magnitudes of image noise and water diffusion are relatively small, and when the gradient of the leading eigenvector field is near constant along the integral curve. As described above, in the common case that integral curve tracing encounters a location whose leading eigenvector direction is ambiguous, we begin traces of new integral curves along each plausible eigenvector direction, and build complete descriptions of a fiber

trajectory by linking together all such branching integral curves emanating from a given starting point.

The rest of the paper is organized as follows. Notation and our framework are described in Sect. 2. We formulate the three-step estimation procedure in Sect. 3. Important definitions and conditions are gathered in Sect. 4. The main theoretical results, and the step-by-step algorithm (including a test for branching), are in Sect. 5. We give a theoretical comparison between our approach and that of Koltchinskii et al. (2007) in Sect. 6. We illustrate how our method works on artificial data in Sect. 7 and on real DTI data in Sect. 8. We draw conclusions and discuss our findings in Sect. 9. Proofs are gathered in the Appendix.

2 Notation and framework

Throughout the paper all vectors are columns and u^* denotes the transposition of a vector or a tensor u. In brain imaging applications d is typically 2 or 3, G represents a region of a brain, and $M(x)$ provides a representation of the spatial distribution of water diffusion at x. DTI makes the estimation of $M(x)$ possible by collecting a set of N component images, each of which uses magnetic field gradients to measure the relative amount of water diffusion at x along a spatial direction represented by $b \in \mathbb{R}^d$. This relative water diffusion measurement is denoted by $S(x, b)$. If these measurements are collected along at least $d(d + 1)/2$ such directions, and the spatial distribution of water diffusion at each x is assumed to be ellipsoidal, then there is sufficient data to estimate a diffusion tensor $M(x)$ such that for any spatial direction b, $S(x, b)$ is estimated as follows: $\log\left(\frac{S(x,b)}{S_0(x)}\right) =: y(x, b) = -cb^*M(x)b + \sigma(x, b)\xi$. Here S_0 is a baseline signal level measured in the absence of a magnetic field gradient; the constant c depends only on the hydrogen gyromagnetic ratio, the gradient pulse sequence design, duration and other timing parameters of the imaging procedure (see, e.g., Basser and Pierpaoli 1998); σ is a positive function that adds noise to the measurement that may depend on x and b; and ξ is a random variable with mean zero and variance one. In the absence of noise this relationship is known as the Stejskal–Tanner equation.

For a tensor $A \in \mathbb{R}^{d^2}$, let $A_{kl}, k, l = 1, \ldots, d$ be its components. For convenience let $d_0 := d(d + 1)/2$. For a symmetric tensor $A \in \mathbb{R}^{d^2}$ we define a d_0-dimensional vector $\mathbf{A}$ that consists of stacked rows of the upper triangular part of the tensor A, i.e. $\mathbf{A} = (A_{11}, A_{12}, \ldots, A_{1d}, A_{22}, A_{23}, \ldots, A_{2d}, \ldots, A_{(d-1)(d-1)}, A_{(d-1)d}, A_{dd})^*$. Note that for any two symmetric tensors $A, F \in \mathbb{R}^{d^2}$ the trace of their product can be calculated as $\sum_{k,l=1}^{d} A_{kl}F_{kl} = \mathbf{A}H\mathbf{F}$, where H is a diagonal matrix with ones and twos on the diagonal. For example, for $d = 3$ the diagonal of H is $(1, 2, 2, 1, 2, 1)$.

Given N magnetic field gradient directions $\{b_1, \ldots, b_N\}$, there is a fixed $N \times d_0$ matrix B that is related to the set of spatial gradient directions and timing parameters of the imaging procedure (its kth row is proportional to $\mathbf{F}H$ with $F = b_k b_k^*$, $k = 1, \ldots, N$). At a fixed location $x \in \mathbb{R}^d$, we observe the $N \times m$ tensor $Y(x)$ such that

$$Y(x) = B\mathbf{M}(x)I_m^* + \Sigma^{1/2}(x)\Xi_x \quad \text{with} \quad I_m = (1, \ldots, 1)^* \in \mathbb{R}^m, \tag{2}$$

where the columns of $Y(x)$ are $Y_i \in \mathbb{R}^N$, $i = 1, \ldots, m$, and for a fixed x the $N \times N$ tensor Σ is symmetric positive definite; the $N \times m$ tensor Ξ_x is a random noise. Note that this is a linear model with the fixed design.

Throughout the paper, for a symmetric tensor A, its maximal eigenvalue is denoted by $\lambda(A)$. When it is simple (not repeated) the corresponding unit length eigenvector is denoted

by $v(A)$. Under certain conditions on the tensor field $M(x)$, $x \in G$, the vector field $v(M(x))$ is unique locally around the integral curve, which in turn exists and is unique. Because water diffuses preferentially along the directions of travel of axon fibers that connect disparate regions of the brain to each other, tracing such integral curves along prominent diffusion directions can result in $x(t)$, $t \geq 0$, described in (1), that are geometric representations of those axon fibers. There are various downstream applications of the fiber curves; for example, structural connectivity analysis makes use of such fiber trajectories to infer the degree to which two distinct brain regions appear to be connected to each other, in effect quantifying the capacity of two brain regions to communicate with each other.

3 Three-step estimation procedure

The goal of our three step estimation procedure is to (1) regularize the input tensor field, (2) convert it to a vector field representation that allows integral curve tracing, and (3) perform the integral curve tracing. As a preliminary step one needs to estimate the tensor $M(x)$ at a fixed location $x \in G$ from the raw Y measurements. There are various ways to do so. For instance, the ordinary least squares estimator of $M(x)$, $x \in G$, is

$$\tilde{\mathbf{M}}(x) = \frac{1}{m}(B^*B)^{-1}B^*Y(x)I_m, \tag{3}$$

provided $(B^*B)^{-1}$ exists. It is the most popular choice in the DTI literature. Another estimate is the weighted least squares estimator of $M(x)$, $x \in G$, which is studied extensively in the work of Zhu et al. (2007, 2009).

Our estimation procedure is as follows:

Step (i): *Smoothed tensor field estimation* At each location $x \in G$, a smoothed estimate $\hat{\mathbf{M}}_n(x)$ of the tensor field is constructed using the following kernel smoothing estimator:

$$\hat{\mathbf{M}}_n(x) = \frac{1}{nh_n^d} \sum_{j=1}^n K\left(\frac{x - X_j}{h_n}\right)\tilde{\mathbf{M}}(X_j), \tag{4}$$

where K is a kernel function and h_n is the bandwidth.

Step (ii): *Leading eigenvector estimation* For each $x \in G$, the eigenvector $v(\hat{M}_n(x))$ corresponding to the simple maximal eigenvalue $\lambda(\hat{M}_n(x))$ is calculated. This quantity is an estimator of the true unknown eigenvector $v(M(x))$ of $M(x)$ corresponding to the simple maximal eigenvalue $\lambda(M(x))$ under certain conditions on $M(x)$.

Step (iii): *Integral curve estimation* We estimate $x(t)$, $t \geq 0$, starting at a fixed known point $a \in G$ driven by the vector field $v(M(x))$, $x \in G$, by a plug-in estimator $\hat{X}_n(t)$, $t \geq 0$, which is a solution of the ODE

$$\frac{d\hat{X}_n(t)}{dt} = v(\hat{M}_n(\hat{X}_n(t))), t \geq 0, \quad \hat{X}_n(0) = a \tag{5}$$

or equivalently of the integral equation $\hat{X}_n(t) = a + \int_0^t v(\hat{M}_n(\hat{X}_n(s)))ds$. In practice this ODE is solved numerically; see Sect. 5.

In Step (i), isotropic kernel smoothing prevents diffusion tensors outside white matter tract regions from exerting an influence on tensor characteristics. Near tract boundaries, the bandwidth can be chosen small enough that diffusion tensors outside the tract region exert little influence on the averaging operation.

Step (ii) depends on the existence of a single, unambiguous leading eigenvector $v(M(x))$ that represents the principal direction of water diffusion at $x \in G$. However, in DTI, it is common to have multiple distinct water diffusion directions at x; this especially occurs at locations where distinct fibers cross each other. We first identify these locations through a statistical test based on the difference between the maximal and the next to maximal eigenvalues of the tensor field. Asymptotical null distribution of this test statistic is normal. The test rejects the null hypothesis of one single unambiguous leading eigenvalue for small values of the test statistic. These locations are referred to as *branching points* in the fiber trajectory since they suggest locations where fibers diverge or cross.

At these branching points, we modify model (2) so that the observations $Y(x)$ arise from a mixture of two underlying tensors $M^{(1)}(x)$ and $M^{(2)}(x)$, which have the same simple maximal eigenvalue but different leading eigenvectors $v(M^{(1)}(x))$ and $v(M^{(2)}(x))$:

$$Y(x) = B(\pi \mathbf{M}^{(1)}(x) + (1-\pi)\mathbf{M}^{(2)}(x))I_m^* + \Sigma^{1/2}(x)\Xi_x. \qquad (6)$$

The mixing coefficient π determines the relative contributions of $M^{(1)}(x)$ and $M^{(2)}(x)$ to $Y(x)$. A variety of standard clustering techniques can be applied to estimate $M^{(1)}(x)$ and $M^{(2)}(x)$ as well as π. Below, we use k-means clustering with 2 clusters performed on observations $Y(X_j)$, $j = 1, \ldots, n$, within a local neighborhood around x, and set π to a constant 0.5. The neighborhood is defined to be the set of locations surrounding x that influence the smoothing of $M(x)$, as determined by the tensor smoothing kernel K and the associated bandwidth h_n: see Eq. (4). A detailed description of this procedure is in Sect. 5.

During the numerical estimation of the fiber trajectory, if any location $\hat{X}_n(t)$ is determined to be a branching point, trajectory tracing halts there, and two new integral curve tracings are initiated from the branching point, one each along directions $v(M^{(1)}(x))$ and $v(M^{(2)}(x))$. After this process terminates, the final answer for the estimated trajectory initiating from the starting point a consists of the collection of $\hat{X}_n(t)$ estimators that represent trajectory segments running between a, branching points, and terminal points. This approach allows us to deal with the common case of branching points while maintaining theoretically rigorous estimation of integral curves over fiber segments in between them.

4 Foundations for main results

In what follows, for a number, a vector or a tensor w we denote the sum of its squared components as $|w|^2$. For a scalar field, a vector field or a tensor field w we denote the tensor of derivatives of its components by ∇w. More precisely, for a function $w : \mathbb{R}^d \to \mathbb{R}$ the components of the vector $\nabla w(x)$ are $\partial w(x)/\partial x_k$, $k = 1, \ldots, d$. For a vector field $w : \mathbb{R}^{d^2} \to \mathbb{R}^d$ the components of the vector $\nabla w(M)$ are $\partial w_j(M)/\partial M_{kl}$, $j, k, l = 1, \ldots, d$, so ∇w is a $d \times d^2$ tensor. Somewhat opposite, for a tensor field $w : \mathbb{R}^d \to \mathbb{R}^{d^2}$ the components of the vector $\nabla w(x)$ are $\partial w_{kl}(x)/\partial x_j$, $j, k, l = 1, \ldots, d$, so ∇w is a $d^2 \times d$ tensor. The Kroneker symbol is $\delta_{kl} = 1$ for $k = l$ and 0 for $k \neq l$.

The following conditions are required:

(A1) G is a bounded open set in $\mathbb{R}^d$ with Lebesgue measure 1. It contains the support of the twice continuously differentiable everywhere symmetric positive definite tensor field $M : \mathbb{R}^d \to \mathbb{R}^{d^2}$. Moreover, $M(\cdot)$ has the simple eigenvalues everywhere in its support (See Sect. 5 for a relaxation of this requirement).
(A2) The initial point a is inside of the support of $M(\cdot)$.

(A3) There exists a number $T > 0$ such that for all $t_1, t_2 \in (0, T)$ with $t_1 \neq t_2$, $x(t_1) \neq x(t_2)$.

(A4) Locations $\{X_j, j \geq 1\}$ are independent uniformly distributed in G.

(A5) We observe $\{(X_j, Y_i(X_j)), i = 1, \ldots, m, j = 1, \ldots, n\}$, obeying the model $Y(X_j) = B\mathbf{M}(X_j)I_m + \Sigma^{1/2}(X_j)\Xi_j$ with a fixed non-random known real-valued $N \times d_0$ tensor B, an unknown continuous on G symmetric positive definite $N \times N$ tensor field $\Sigma : \mathbb{R}^d \to \mathbb{R}^{N^2}$, and unobservable random $N \times m$ tensors $\Xi_j, j = 1, \ldots, n$. Recall $N \geq d_0$. Additionally, we assume that B^*B is invertible and $\mathbb{E}\Sigma_{kl}^4(X_1) < \infty, 1 \leq k, l \leq N$.

(A6) The columns of random noise matrices $\Xi_j, j \geq 1$, are i.i.d. as Ξ and independent of locations. Additionally, components of the matrix Ξ satisfy $\mathbb{E}\Xi_{ki} = 0, \mathbb{E}\Xi_{ki}\Xi_{li} = \delta_{kl}$, and $\mathbb{E}\Xi_{ki}^4 < \infty$ for all $i = 1, \ldots, m, 1 \leq k, l \leq N$.

(A7) The kernel K is nonnegative and twice continuously differentiable on its bounded support. Moreover, $\int_{\mathbb{R}^d} K(x)dx = 1, \int_{\mathbb{R}^d} x K(x)dx = 0$.

(A8) The bandwidth h_n satisfies the condition $nh_n^{d+3} \to \beta > 0$ as $n \to \infty$, where β is a known fixed number.

Condition (A1) requires the region G to be a locally contiguous region of the brain.

Conditions (A1) and (A2) guarantee that a unique solution $x(t), t \in [0, T]$, of (1) exists and stays inside G; see formula (9) below. Condition (A1) can be relaxed to assume that $M(\cdot)$ has simple eigenvalues in a neighborhood of a that contains the curve $x(t), t \in [0, T]$. For more details on existence, uniqueness, and smoothness of tensor fields and the associated eigenvector fields see Kato (1980).

Condition (A3) prohibits cycles in the integral curve. This is not restrictive; when a cycle is detected, one needs to estimate the integral curve over just one period.

Analogous versions of the main results can be proven when condition (A4) is changed to allow diffusion tensor measurements arranged in a non-random, regular grid of locations as is typical in practice. Furthermore, the typical number of locations n in a DTI data set is on the order of hundreds of thousands or millions; for sample sizes this large, a regular grid of locations is fairly well approximated by uniform i.i.d. locations.

Condition (A6) means that noise is determined independently at each location. Given the physics of MRI acquisition, this assumption may be unrealistic. But this assumption is typical for the state of the art in diffusion MRI; it is built into the near-ubiquitous Rician noise model for example. Accounting for spatially correlated noise structures in theoretical models of DTI will require methodological advances that are beyond the scope of the current study.

Condition (A6) also requires that the noise is white: zero-mean and with no correlations among the diffusion tensor entries. Previous researchers have noted that the popular Rician noise model for DTI is well approximated by such a white noise model in the event that the ratio of Rician model's mean to standard deviation is moderate to large. Because this is typically the case in real-world DTI studies, our white noise model is a realistic representation of DTI noise; see Zhu et al. (2007) and references therein.

If the second moments of K are finite, ∇K and $\nabla^2 K$ are uniformly bounded, and the functions Λ_2 and Λ_4 defined in the proof of Theorem 1 are integrable, then condition (A7) can be relaxed to accommodate them, even if their support is infinite. For instance, this is the case for Gaussian kernels.

Finally, note that formula (3) can be rewritten as

$$\tilde{\mathbf{M}}(X_j) = \mathbf{M}(X_j) + \mathbf{\Gamma}_j, \quad \mathbf{\Gamma}_j = \frac{1}{m}(B^*B)^{-1}B^*\Sigma^{1/2}(X_j)\Xi_j I_m, \quad j \geq 1, \qquad (7)$$

where $\mathbf{\Gamma}_j$ denotes a d_0-vector representation of a random tensor in $\mathbb{R}^{d^2}$, $j \geq 1$. Note that due to condition (A6) random tensors $\mathbf{\Gamma}_j$, $j \geq 1$, are i.i.d. and we have for all $j \geq 1$, $\mathbb{E}\mathbf{\Gamma}_j = 0$ and

$$\mathbb{E}(\mathbf{\Gamma}_j \mathbf{\Gamma}_j^* | X_j) = \frac{1}{m^2}(B^*B)^{-1}B^*\mathbb{E}[\mathbf{\Sigma}^{1/2}(X_j)\mathbf{\Xi}_1 I_m I_m^* \mathbf{\Xi}_1 (\mathbf{\Sigma}^{1/2}(X_j))^*]B(B^*B)^{-1}$$

$$= \frac{1}{m}(B^*B)^{-1}B^*\mathbf{\Sigma}(X_j)B(B^*B)^{-1} =: \Sigma_\Gamma(X_j),$$

where $\Sigma_\Gamma : \mathbb{R}^d \to \mathbb{R}^{d_0^2}$ is a tensor field. We remark that if the matrix B is a square matrix, i.e. $N = d_0$, and it is invertible then we can simplify $\Sigma(\cdot) = \frac{1}{m}(B^*\Sigma(\cdot)B)^{-1}$.

5 Main results

First, let us introduce lemmas that are required to prove the key theoretical result.

Lemma 1 *Suppose conditions (A1)–(A8) hold. Then* $\mathbb{E}\sup_{x \in \mathbb{R}^d}|\hat{M}_n(x) - M(x)|^2 \leq \frac{C}{nh_n^{d+2}}$ *for sufficiently large n with a finite constant C.*

Consequently, we have $\sup_{x \in \mathbb{R}^d}|\hat{M}_n(x) - M(x)| \to^P 0$ as $n \to \infty$.

Lemma 2 *Suppose conditions (A1)–(A8) hold. Then* $\sup_{x \in \mathbb{R}^d}|\nabla\hat{M}_n(x) - \nabla M(x)| \to^P 0$ *as $n \to \infty$.*

Lemma 3 *Suppose conditions (A1)–(A8) hold. Then* $\sup_{t \in [0,T]}|\hat{X}_n(t) - x(t)| \to^P 0$ *as $n \to \infty$.*

Lemma 4 *Suppose conditions (A1)–(A8) hold. Then $\hat{M}_n(x)$ is symmetric for all $x \in G$ and all n. Moreover, $P(\hat{M}_n(x)$ is positive definite for all $x \in G) \to 1$ as $n \to \infty$.*

In order to formulate the main result of this paper we need the following definitions. As in the work of Koltchinskii et al. (2007) define for $u \in \mathbb{R}^d$

$$\Psi(u) = \int_{\mathbb{R}^d} K(z)K(z+u)dz, \quad \psi(u) = \int_{\mathbb{R}}\Psi(\tau u)d\tau.$$

Let $U : \mathbb{R}^2 \to \mathbb{R}^{d^2}$ be the Green's function, defined as the solution of the following PDE

$$\frac{\partial U(t,s)}{\partial t} = \nabla v(M(x(t)))\nabla M(x(t))U(t,s), \quad U(s,s) = \mathbb{I}.$$

For a vector field $w : \mathbb{R}^{d^2} \to \mathbb{R}^d$ let $\tilde{\nabla}w$ be the $d \times d_0$ tensor of its derivatives $\partial w/\partial \mathbf{M}_i$, $i = 1, \ldots, d_0$. For a $d^2 \times d^2$ tensor A and vector $z \in \mathbb{R}^d$ let $\langle Az, z \rangle$ be the d^2-vector with components $\langle Az, z \rangle_{kl} = \sum_{q,r=1}^d A_{kl,qr}z_q z_r$, $1 \leq k, l \leq d$.

The main theorem follows.

Theorem 1 *Suppose conditions (A1)–(A8) hold. Then the sequence of stochastic processes*

$$\sqrt{nh_n^{d-1}}(\hat{X}_n(t) - x(t)), \quad t \in [0,T],$$

converges weakly in the space of $\mathbb{R}^d$-valued continuous functions on $[0,T]$ to the Gaussian process $\mathcal{G}(t)$, $t \in [0,T]$, with mean

$$\mu_\beta(t) = 0.5\sqrt{\beta}\int_0^t \int_{\mathbb{R}^d} U(t,s)\nabla v(M(x(s)))K(z)\langle\nabla^2 M(x(s))z, z\rangle dz ds$$

and covariance

$$C(t_1, t_2) = \int_0^{t_1 \wedge t_2} \psi(v(M(x(s)))) U(t_1, s) \tilde{\nabla} v(M(x(s)))$$
$$\times H[\mathbf{M}(x(s))\mathbf{M}^*(x(s)) + \Sigma_\Gamma(x(s))] H \tilde{\nabla} v(M(x(s)))^* U^*(t_2, s) ds.$$

Notice that $\mu_\beta(t), t \geq 0$, satisfies the following ODE

$$\frac{d\mu_\beta(t)}{dt} = \nabla v(M(x(t))) \nabla M(x(t)) \mu_\beta(t)$$
$$+ 0.5\sqrt{\beta} \nabla v(M(x(t))) \int_{\mathbb{R}^d} K(z) \langle \nabla^2 M(x(t))z, z \rangle dz, \quad \mu_\beta(0) = 0.$$

By the same token $C(t, t), t \geq 0$, satisfies the following ODE: $C(0, 0) = 0$,

$$\frac{dC(t, t)}{dt} = \nabla v(M(x(t))) \nabla M(x(t)) C(t, t) + C(t, t)(\nabla v(M(x(t))) \nabla M(x(t)))^*$$
$$+ \psi(v(M(x(t)))) \tilde{\nabla} v(M(x(t))) H[\mathbf{M}(x(t))\mathbf{M}^*(x(t))$$
$$+ \Sigma_\Gamma(x(t))] H \tilde{\nabla} v(M(x(t)))^*.$$

Thus, Theorem 1 allows us to track the estimated curve, the mean function, and the covariance function in small steps sequentially and simultaneously, and construct confidence ellipsoids along the curve. This will be illustrated in Sects. 7 and 8.

Sketch of proof The proof of this result, provided in the Appendix, has several steps. First, we show that $\hat{X}_n(t) - x(t), t \in [0, T]$, is approximated by a process $Z_n(t), t \in [0, T]$, which is the solution of a linear integral equation and can be explicitly written as a linear operator transformation of $\hat{M}_n(\cdot) - M(\cdot)$. And $Z_n(t), t \in [0, T]$, happens to be a sum of i.i.d. terms. Second, we find the asymptotical mean and covariance of Z_n. Finally, we show that the sequence $\sqrt{nh_n^{d-1}}(Z_n(t) - \mathbb{E}Z_n(t)), t \in [0, T]$, converges to a Gaussian process due to Lyapunov's condition and asymptotical equicontinuity.

Connectivity tests This result allows us to apply Theorem 2 from Koltchinskii et al. (2007) to perform tests of connectivity that are commonly of interest in studies of DTI data: given a starting location a, the test is concerned with whether the integral curve originating at a passes through a given subregion in G. We formulate this test in terms of estimating the squared Euclidean distance between the integral curve and a one-point region $r \in G$; also see Koltchinskii and Sakhanenko (2009).

Corollary 1 *Suppose conditions (A1)–(A8) hold. Moreover, suppose there exists the unique point $\tau \in (0, T)$ such that $\min_{t \in [0,T]} |x(t) - r|^2 = |x(\tau) - r|^2$. If $x(\tau) \neq r$ then the sequence $\sqrt{nh_n^{d-1}} \left[\min_{t \in [0,T]} |\hat{X}_n(t) - r|^2 - |x(\tau) - r|^2 \right]$ is asymptotically normal with mean $2\mu_\beta(\tau)^*(x(\tau) - r)$ and variance $4(x(\tau) - r)^* C(\tau, \tau)(x(\tau) - r)$. If $x(\tau) = r$ then the sequence $nh_n^{d-1} \min_{t \in [0,T]} |\hat{X}_n(t) - r|^2$ converges in distribution to a random variable $|Z|^2 - (v(M(x(\tau)))^* Z)^2$, where Z is a normal random variable with mean $\mu_\beta(\tau)$ and variance $C(\tau, \tau)$.*

This corollary can be used to test whether an integral curve starting at a point $a \in G$ passes closely by the point r. See Koltchinskii and Sakhanenko (2009) for details and proofs.

5.1 Implementation

The complete algorithm for obtaining integral curve estimator together with confidence ellipsoids has the following steps.

1. Initialize $q = 0$, $t_q = 0$, $\hat{X}_n(t_0) = a$, $\mu_\beta(t_0) = 0$, $C(t_0, t_0) = 0$.
2. Track the estimated integral curve using Euler's method, given a fixed $\delta > 0$ and small time steps $t_q = q\delta, q = 0, 1, 2, \ldots$[1]:

$$\hat{X}_n(t_q) \approx \hat{X}_n(t_{q-1}) + \delta v(\hat{M}_n(\hat{X}_n(t_{q-1}))), \quad \hat{X}_n(0) = a.$$

3. Approximate $\nabla^2 \hat{M}_n(x)$, $\nabla \hat{M}_n(x)$ and $\hat{\Sigma}_n(x)$ as follows:

$$\tilde{\Sigma}_{\Gamma,n}(X_j) = \frac{1}{m} \sum_{i=1}^{m} (Y_i(X_j) - B\tilde{\mathbf{M}}(X_j))(Y_i(X_j) - B\tilde{\mathbf{M}}(X_j))^*, \quad j = 1, \ldots, n,$$

$$\hat{\Sigma}_n(x) = \frac{1}{nh_n^d} \sum_{j=1}^{n} K((x - X_j)/h_n)\tilde{\Sigma}_{\Gamma,n}(X_j), \quad x \in G,$$

$$\nabla \hat{M}_n(x) = \frac{1}{n\tilde{h}_n^{d+1}} \sum_{j=1}^{n} \nabla K\left(\frac{x - X_j}{\tilde{h}_n}\right)\tilde{M}(X_j), \quad x \in G, \quad n\tilde{h}_n^{d+1} \to \infty,$$

$$\nabla^2 \hat{M}_n(x) = \frac{1}{n\tilde{\tilde{h}}_n^{d+2}} \sum_{j=1}^{n} \nabla^2 K\left(\frac{x - X_j}{\tilde{\tilde{h}}_n}\right)\tilde{M}(X_j), \quad x \in G, \quad n\tilde{\tilde{h}}_n^{d+1} \to \infty.$$

And, approximate $\nabla v(M(x))$ by $\nabla v(\hat{M}_n(x))$ using the relationship (9) in the proof of Lemma 3, i.e.

$$\frac{\partial v_p(\hat{M}_n(x))}{\partial M_{kl}} = (1 - 0.5\delta_{kl}) \left[(\lambda(\hat{M}_n(x))I_d - \hat{M}_n(x))^+_{pk} v_l(\hat{M}_n(x)) \right.$$

$$\left. + (\lambda(\hat{M}_n(x))I_d - \hat{M}_n(x))^+_{pl} v_k(\hat{M}_n(x)) \right], \quad 1 \le p, k, l \le d,$$

where A^+ stands for the Moore–Penrose inverse of a matrix A.

4. Approximate $\mu_\beta(t_q)$, $q \ge 1$, as

$$\hat{\mu}_\beta(t_q) \approx \hat{\mu}_\beta(t_{q-1}) + \delta \nabla v(\hat{M}_n(\hat{X}_n(t_{q-1})))\nabla \hat{M}_n(\hat{X}_n(t_{q-1}))\hat{\mu}_\beta(t_{q-1})$$

$$+ 0.5\delta\sqrt{\beta}\nabla v(\hat{M}_n(\hat{X}_n(t_{q-1}))) \int_{\mathbb{R}^d} K(z)\langle \nabla^2 \hat{M}_n(\hat{X}_n(t_{q-1}))z, z\rangle dz.$$

5. Approximate $C(t_q, t_q)$ as

$$\hat{C}(t_q, t_q) \approx \hat{C}(t_{q-1}, t_{q-1}) + \delta \nabla v(\hat{M}_n(\hat{X}_n(t_{q-1})))\nabla \hat{M}_n(\hat{X}_n(t_{q-1}))\hat{C}(t_{q-1}, t_{q-1})$$

$$+ \delta\hat{C}(t_{q-1}, t_{q-1})\nabla \hat{M}_n(\hat{X}_n(t_{q-1}))^* \nabla v(\hat{M}_n(\hat{X}_n(t_{q-1})))^*$$

$$+ \delta\psi(v(\hat{M}_n(\hat{X}_n(t_{q-1}))))\tilde{\nabla} v(\hat{M}_n(\hat{X}_n(t_{q-1})))$$

$$\times [\hat{\mathbf{M}}_n(\hat{X}_n(t_{q-1}))\hat{\mathbf{M}}_n(\hat{X}_n(t_{q-1})) + \Sigma_\Gamma(\hat{X}_n(t_{q-1}))]$$

$$\times \tilde{\nabla} v(\hat{M}_n(\hat{X}_n(t_{q-1})))^*, \quad q \ge 1.$$

---

[1] We note that while in theory, higher-order solvers such as Runge–Kutta's method could have provided more accurate integral curves, we recently showed that the Runge–Kutta method provides no tangible advantages over Euler's method when applied to the simplified integral curve model of Koltchinskii et al. (2007) (see Sakhanenko 2012).

6. Approximate the $100(1 - \alpha)\%$ confidence ellipsoid for $x(t_q)$, $q \geq 1$, as

$$P\left\{|\hat{C}(t_q, t_q)^{-1/2}(\sqrt{nh_n^{d-1}}(\hat{X}_n(t_q) - x(t_q)) - \hat{\mu}_\beta(t_q))| \leq R_\alpha\right\} \approx 1 - \alpha,$$

where $P(|Z| \leq R_\alpha) = 1 - \alpha$ for a standard normal vector Z in $\mathbb{R}^d$.
7. Repeat steps 2–6 until t_q reaches T.

5.2 Kernel smoothing bandwidth selection

Finally, we remark on how to select β in the bandwidth $h_n = (\beta/n)^{1/(d+3)}$. One approach is to estimate the β that minimizes the mean integrated squared error (MISE) between the estimated and the true integral curves. MISE is asymptotically equivalent to the following expression

$$\text{MISE} := \mathbb{E}\int_0^T (\hat{X}_n(t) - x(t))^*(\hat{X}_n(t) - x(t))dt$$

$$\approx n^{-4/(d+3)}\left[\beta^{4/(d+3)}\int_0^T \mu_1(t)^*\mu_1(t)dt + \beta^{-(d-1)/(d+3)}\int_0^T \text{tr}C(t, t)dt\right],$$

which is minimized by $\beta = 0.25(d - 1)\int_0^T \text{tr}C(t, t)dt\left[\int_0^T \mu_1(t)^*\mu_1(t)dt\right]^{-1}$. An interleaved estimation procedure for β would thus consist of starting with an initial β estimate, estimating μ_1 and C accordingly, re-estimating β based on these μ_1 and C, and so on.

5.3 Modeling locations with multiple fiber directions

Note that the covariance of the limiting Gaussian process in Theorem 1 goes to infinity as the second largest eigenvalue of $M(\cdot)$ (call this $\phi(M(\cdot))$), approaches the largest eigenvalue $\lambda(M(\cdot))$. More precisely, some components of the gradient $\tilde{\nabla}v(M(x))$ in the integrand of $C(t_1, t_2)$, $t_1, t_2 \in [0, T]$, are proportional to $(\lambda(M(x)) - \phi(M(x)))^{-2}$ for a point x on the curve $x(t)$, $t \in [0, T]$. Our approach is to devise a theoretically grounded test for this situation, and when it occurs we decompose the tensor $M(\cdot)$ into components, each of which has a dominant leading eigenvalue.

Branching point detection We develop a test for branching based on the difference between the two largest eigenvalues of the estimated tensor field $\hat{M}_n(\cdot)$. Let $w(A)$ be the eigenvector corresponding to the second largest eigenvalue $\phi(A)$ of a $d \times d$ matrix A. We utilize the following result.

Proposition 1 *Suppose conditions (A1)–(A8) hold. For any fixed $x \in G$*

$$\sqrt{nh_n^d}\left((\lambda(\hat{M}_n(x)) - \lambda(M(x))) - (\phi(\hat{M}_n(x)) - \phi(M(x)))\right)$$

is asymptotically normal with mean 0 and variance $\sigma_\lambda^2(x) + \sigma_\phi^2(x) - 2\sigma_{\lambda,\phi}(x)$, where

$$\sigma_\lambda^2(x) = \int K^2(u)du\,\Delta_v(x)^*H[\mathbf{M}(x)\mathbf{M}^*(x) + \Sigma_\Gamma(x)]H\Delta_v(x),$$

$$\sigma_\phi^2(x) = \int K^2(u)du\,\Delta_w^*(x)H[\mathbf{M}(x)\mathbf{M}^*(x) + \Sigma_\Gamma(x)]H\Delta_w(x),$$

$$\sigma_{\lambda,\phi}(x) = \int K^2(u)du\,\Delta_v^*(x)H[\mathbf{M}(x)\mathbf{M}^*(x) + \Sigma_\Gamma(x)]H\Delta_w(x)$$

with Δ_v being a vector representation of the matrix Δ_v with entries $\Delta_{v,ij} = (2 - \delta_{ij})v_i(M)v_j(M)$ and Δ_w being a vector representation of the matrix Δ_w with entries $\Delta_{w,ij} = (2 - \delta_{ij})w_i(M)w_j(M)$.

Combining this result with Theorem 1 by means of the delta method allows us to construct the following test. We reject the null hypothesis of no branching at a location $\hat{X}_n(t_{q+1})$, $q \geq 1$, if

$$\lambda(\hat{M}_n(\hat{X}_n(t_{q+1}))) - \phi(\hat{M}_n(\hat{X}_n(t_{q+1}))) < \lambda(\hat{M}_n(\hat{X}_n(t_q))) - \phi(\hat{M}_n(\hat{X}_n(t_q)))$$
$$-\varepsilon_\alpha(nh_n^d)^{-1/2}[\hat{\sigma}_\lambda^2(\hat{X}_n(t_q)) + \hat{\sigma}_\phi^2(\hat{X}_n(t_q)) - 2\hat{\sigma}_{\lambda,\phi}(\hat{X}_n(t_q))]^{1/2},$$

where $\hat{\sigma}_\lambda^2$, $\hat{\sigma}_\phi^2$, and $\hat{\sigma}_{\lambda,\phi}$ are obtained from σ_λ^2, σ_ϕ^2, and $\sigma_{\lambda,\phi}$ using the estimators $\hat{M}_n$ and $\hat{\Sigma}_n$ instead of M and Σ_Γ and $P(0 < Z < \varepsilon_\alpha) = \alpha$ for a standard normal variable Z.

Tensor decomposition at branching points Whenever a branching point is detected at a location x, we use 2-means clustering method to group the raw DTI measurements $Y(X_j)$, $j = 1, \ldots, n$, for X_j that are within the averaging window of the smoothing kernel K into clusters, using the Euclidean distance between $Y(X_j)$ as a measure of their similarity. Assuming that the branching point consists of the confluence of two distinct fiber orientations, the two clusters are expected to correspond to those two orientations respectively. The two diffusion tensors $\mathbf{M}^{(1)}(x)$ and $\mathbf{M}^{(2)}(x)$ are estimated from the two clusters individually. We trace two new integral curves beginning at x that initially move along the $v(\mathbf{M}^{(1)}(x))$ and $v(\mathbf{M}^{(2)}(x))$ directions, and repeat the branching test where these two curves arrive.

We apply this to the synthetic tensor field in Fig. 2. This is a linear combination of a circular tensor field (from the example in Sect. 7) with a circle radius of 0.25 and a horizontal tensor field in the strip $|x_2| < 0.05$ with the same maximal eigenvalues as the circular tensor field. We use a mixture model with $\pi = 0.5$ for the locations where the tensor fields overlap. Noise is added to the raw measurements Y with the matrix $\Sigma = [1, 0.8, 0.8; 0.8, 1, 0.8; 0.8, 0.8, 1]$. The method successfully identifies the branching point and traces both fibers.
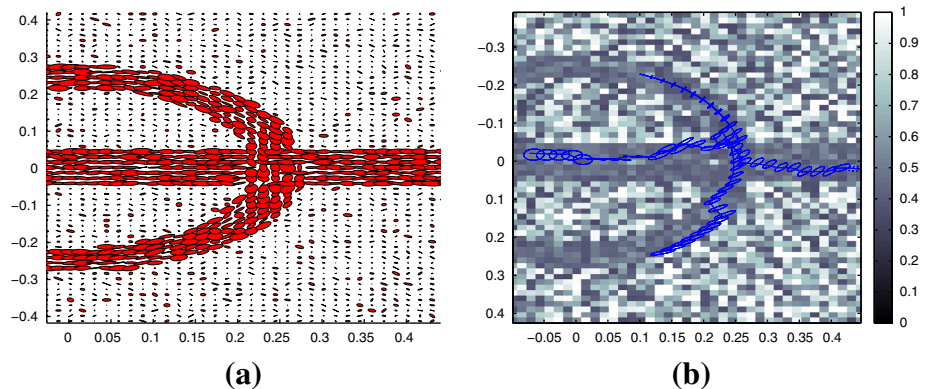

(a) **(b)**

Fig. 2 A test for branching points and a mixture model for diffusion tensors at such points can be used to apply the integral curve tracing method to locations where fibers diverge or cross. The point where the semi-circular and horizontal tensor regions overlap is successfully detected as a branching point, and separate integral curves are traced along both trajectories. **a** The 2×2 tensor field $\tilde{M}$ is visualized through ellipses. Here $m = 8$, $n = 4900$, $\delta = 0.02$, $\beta = 0.001$. **b** FA map is overlaid with estimated fiber trajectories and 95 % confidence ellipses based on our method and 2-means clustering near the branching point. The initial point is $(0.1, -0.2291)$. We use $\alpha = 0.1$ for testing. (Color figure online)

6 Theoretical comparison of $\hat{X}_n$ with Koltchinskii et al. (2007) estimator

Here we compare our estimator with that of the most similar fiber tracing approach to ours, due to Koltchinskii et al. (2007). They estimate fiber trajectories with closed-form uncertainty estimates, given noise-corrupted leading eigenvectors without reference to the raw DTI measurements. That is, they assume input data (X_j, V_j), $j = 1, \ldots, n$, where $V_j = v(M(X_j)) + \xi_j$ is the leading eigenvector direction at X_j that has been corrupted with i.i.d. noise variables ξ_j drawn from a distribution with mean zero and covariance Σ_K. Given noise of the form $\xi_j = \Sigma_K(X_j)^{1/2}\eta_j$ with i.i.d., mean 0, variance 1, random vectors η_j, $j = 1, \ldots, n$ that are also drawn independently at each X_j, and given conditions similar to (A1)–(A8), the difference between their estimator $\hat{X}_n^K(t)$, $t \in [0, T]$, and the true integral curve $x(t)$, $t \in [0, T]$, normalized by $\sqrt{nh_n^{d-1}}$ converges weakly in the space of $\mathbb{R}^d$-valued continuous functions on $[0, T]$ to the Gaussian process $\mathcal{G}^K(t)$, $t \in [0, T]$. The mean of this process is

$$\mu_\beta^K(t) = 0.5\sqrt{\beta}\int_0^t \int_{\mathbb{R}^d} U(t, s) K(z) \langle \frac{\partial^2}{\partial x^2} v(x(s))z, z \rangle dz ds$$

and the covariance is

$$C^K(t_1, t_2) = \int_0^{t_1 \wedge t_2} \psi(v(M(x(s)))) U(t_1, s)[v(M(x(s)))v^*(M(x(s))) \\ + \Sigma_K(x(s))]U^*(t_2, s)ds.$$

Next, we show the three specific simplifications that must be made to our tensor and noise models that are required to reduce it down to the special case covered by the Koltchinskii et al. (2007) estimator.

The first required simplification is to model the noise applied to $M(x)$ as a first-order perturbation that moves $M(x)$ within a very local neighborhood. Consider the following first order approximation for an observation $v(M(X_j) + \Gamma_j) = v(M(X_j)) + \tilde{\nabla}v(M(X_j))H\Gamma_j + \alpha_1(|\Gamma_j|^2)$, where the vector-function $\alpha_1(u)/|u|^2$, $u \in \mathbb{R}^d$, is bounded in a small neighborhood around 0. Note that the covariance matrix of the linear term $v(M(X_j)) + \tilde{\nabla}v(M(X_j))H\Gamma_j$ is $\tilde{\nabla}v(M(X_j))H\Sigma_\Gamma H\tilde{\nabla}v(M(X_j))^*$; this covariance matrix is required to be the same as $\Sigma_K(X_j)$ for the measurement to comply with the Koltchinskii et al. (2007) model.

The second required simplification is to assume that we have almost a constant gradient of v with respect to M along the integral curve. Specifically, the second derivatives of v with respect to x must be simplified by the following: $\frac{\partial^2}{\partial x^2}v(M(x)) = \nabla^2 v(M(x))\nabla M(x)\nabla M(x) + \nabla v(M(x))\nabla^2 M(x) = \nabla v(M(x))\nabla^2 M(x)(1 + \alpha_2(x))$, where the vector function $\alpha_2(x)$, $x \in \mathbb{R}^d$, is continuous at 0, and $\alpha_2(0) = 0$.

The third required simplification is to assume that v is almost linear with respect to M, i.e. $v(M(x)) = \tilde{\nabla}v(M(x))H\mathbf{M}(x) + \alpha_1(|M|^2(x))$, $x \in G$. Putting these simplifications together, for all $t \in [0, T]$ and $z \in \mathbb{R}^d$ we have $v(M(x(t))) \approx \tilde{\nabla}v(M(x(t)))H\mathbf{M}(x(t))$ and $\nabla v(M(x(t)))\langle \nabla^2 M(x(t))z, z \rangle \approx \langle \frac{\partial^2}{\partial x^2}v(M(x(t)))z, z \rangle$.

Through these three simplifications we arrive at the model of Koltchinskii et al. (2007), which can be thought of as a very restricted special case of our model. To apply the Koltchinskii model, both the noise Γ and the tensor M must be relatively small along the integral curve and $\nabla v(M)$ should be nearly fixed as M varies along the curve. These assumptions are unrealistic for real-world DTI data; this may explain why the Koltchinskii estimator fails to accurately follow fiber trajectories in the simulated examples in the next section.

7 Synthetic examples

7.1 Comparison with Koltchinskii et al. (2007) approach

We first provide two synthetic examples that compare how our estimator and that of Koltchin-
skii et al. (2007) behave when applied to separate cases that do and do not violate the
assumptions of the Koltchinskii model. The first example shows a curved trajectory of con-
stant curvature in a tensor field corrupted by noise that violates the assumptions of Koltchinskii
et al. (2007). The second example comes close to satisfying this assumption. However, we
show that in both cases, our estimator provides integral curve estimators that have similar
accuracy compared to the Koltchinskii estimator, but lower covariance.

7.1.1 Circular trajectories

For the first example, we consider a 2D tensor field corresponding to a circular fiber trajectory.
Let $G = [-0.5, 0.5]^2$. Let the true tensor field and the eigenvector corresponding to the
maximal eigenvalue be

$$M(x) = \begin{pmatrix} 1 + \frac{x_2^2}{x_1^2 + x_2^2} & \frac{-x_1 x_2}{x_1^2 + x_2^2} \\ \frac{-x_1 x_2}{x_1^2 + x_2^2} & 1 + \frac{x_1^2}{x_1^2 + x_2^2} \end{pmatrix}, \quad v(M(x)) = \begin{pmatrix} \frac{-x_2}{\sqrt{x_1^2 + x_2^2}} \\ \frac{x_1}{\sqrt{x_1^2 + x_2^2}} \end{pmatrix}.$$

The eigenvalues are 2 and 1. The true integral curves are circles centered at the origin.
The directions are $b_1 = (1, 0)^*, b_2 = (0, 1)^*, b_3 = 2^{-1/2}(1, 1)^*$, so the design matrix is
$B = [-1, 0, 0; 0, 0, -1; -0.5, -1, -0.5]$. This common type of gradient direction setup is
referred to as orthogonal or pyramidal gradient encoding in the DTI literature.

For each random location $X_j, j = 1, \ldots, n$, we simulate m vectors $Y_i \in \mathbb{R}^3$ accord-
ing to the model $Y_i = BM(X_j) + \Sigma^{1/2}(X_j)\xi_{j,i}, j = 1, \ldots, n, i = 1, \ldots, m$, with
independent standard normal 3D vectors $\xi_{j,i}$ and two choices for Σ. The first submodel,
$\Sigma = [1, 0.8, 0.8; 0.8, 1, 0.8; 0.8, 0.8, 1]$, is a constant matrix; the second submodel is
$\Sigma(x) = 0.01\mathbf{M}(\mathbf{x})\mathbf{M}(\mathbf{x})^*, x \in G$. In all simulations we use the bandwidth $h_n = (n/\beta)^{1/5}$
and the standard 2D Gaussian kernel.

Both methods yield integral curve estimates that are very close to the true curve (Fig. 3).
However, the covariances of the Koltchinskii et al. (2007) estimator are much higher than
those of our estimator in both submodels. For the first sub-model with a fixed Σ, the Euclidean
norm of the covariance matrix C^K is 20–35 times higher than the norm of C along the
first third of the curve, then this ratio decreases gradually to 1.2 at the end of the curve.
For the second sub-model, the ratio of the norms of C^K and C is also 20–30 for the first
third of the curve and reduces to 5–6 at the end of the curve. This is to be expected since
the matrices between U and U^* in the covariance of the limiting Gaussian processes are
$r^4[\cos^2 2s, -\sin 4s \sin^2 2s/16; -\sin 4s \sin^2 2s/16, \sin^2 s \sin^2 2s/4] + O(\Sigma_\Gamma), s \in [0, T]$,
and $[\sin^2 s, -\sin 2s/2; -\sin 2s/2, \cos^2 s] + O(\Sigma_K), s \in [0, T]$, for our and Koltchinskii
et al.'s (2007) methods respectively. Both covariance matrices Σ_Γ and Σ_K are small and
$r = 1/4$ for the true integral curve.

7.1.2 Nearly additive vector field noise

For the second example, we attempt to design a tensor field and a noise model that come as
close as possible to meeting the additive eigenvector field noise assumption of Koltchinskii
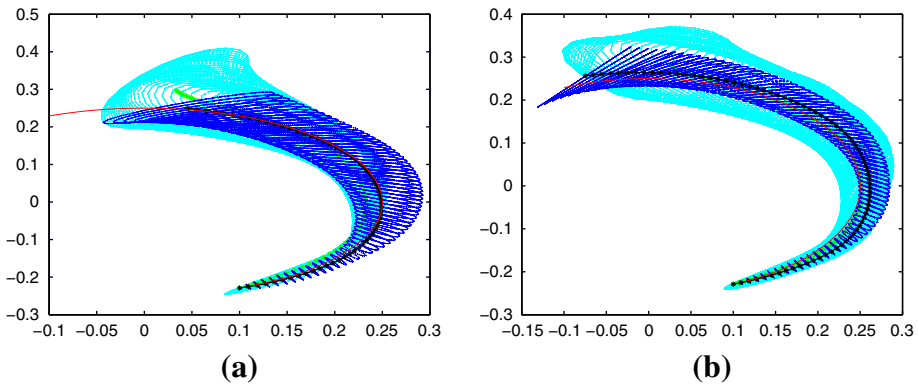
Fig. 3 The *solid black curves* with blue confidence ellipsoids correspond to the estimated integral curves using our approach. The *solid green curves* with cyan confidence ellipsoids correspond to the estimated integral curves using the method of Koltchinskii et al. (2007). The *solid red curves* are the true integral curves. The initial point is (0.1, −0.2291). We use a 95 % confidence level. Here $m = 8$, $n = 20^2 = 400$, $\delta = 0.005$, $\beta = 0.001$. **a** The matrix Σ is fixed for all the locations. **b** The matrix Σ is of the second order with respect to the tensor M. (Color figure online)

et al. (2007), for $d = 3$ and $G = [0, 1]^3$. It is not difficult to show that there are no special cases for which our tensor field noise model exactly yields the additive eigenvector field noise. Therefore, we start with vectors $v_1(x) = (-(x_2 - 0.5), x_1 - 0.5, 1/(2\pi))^*$, $v_2(x) = (x_1 - 0.5, x_2 - 0.5, 0)^*$, $v_3(x) = (x_2 - 0.5, -(x_1 - 0.5), 2\pi((x_1 - 0.5)^2 + (x_2 - 0.5)^2))^*$. Then let the corresponding unit vectors $V_1(x)$, $V_2(x)$, $V_3(x)$ be the eigenvectors of $M(x)$ for the corresponding eigenvalues $(10, 2, 1)$. We use an orthogonal gradient encoding with 6 directions: $b_1 = (1, 0, 0)^*$, $b_2 = (0, 1, 0)^*$, $b_3 = (0, 0, 1)^*$, $b_4 = 2^{-1/2}(0, 1, 1)^*$, $b_5 = 2^{-1/2}(1, 0, 1)^*$, $b_6 = 2^{-1/2}(1, 1, 0)^*$.

For each random location X_j, $j = 1, \ldots, n$, we simulate m vectors $Y \in \mathbb{R}^6$ according to the model $Y_i = BM(X_j) + \Sigma^{1/2}(X_j)\xi_{j,i}$, $j = 1, \ldots, n$, $i = 1, \ldots, m$, with independent standard normal 6D vectors $\xi_{j,i}$ and two choices for Σ. The first submodel is $\Sigma = 0.01[1, \ldots, 1] \in \mathbb{R}^{6^2}$ a constant matrix; the second submodel is $\Sigma_\Gamma(x) = 0.01(x_1, 0.1x_1, 0, 0, 0, 0; 0.1x_1, x_1, 0.1x_1, 0, 0, 0; 0, 0.1x_1, x_1, 0, 0, 0; 0, 0, 0, x_2, 0.1x_2, 0; 0, 0, 0, 0.1x_2, x_2, 0; 0, 0, 0, 0, 0, x_3)$ for $x \in G$. e Koltchinskii et al. (2007) estimator to the perturbed eigenvector fields $v(\tilde{M}(X_j))$, $j = 1, \ldots, n$. In all the simulations we use the bandwidth $h_n = (n/\beta)^{1/6}$ and the standard 3D Gaussian kernel K.

Figure 4 demonstrates that both procedures approximate the true integral curve nicely. Comparing the methods with respect to covariance, our estimator has much tighter confidence ellipsoids than those for the estimator of Koltchinskii et al. (2007). The ratio of the Euclidean norms of the covariance matrices C^K and C is about 100 along the curve for both submodels. Again, this is expected from the theoretical comparison of covariance expressions similar to those done for the 2D example.

7.2 Comparison with Behrens et al. (2003) approach

We consider a popular probabilistic tractography approach by Behrens et al. (2003). This approach uses wild bootstrap to obtain a sample of eigenvector directions at each voxel, then applies Monte Carlo Markov chain (MCMC) sampling to those eigenvector direction samples to generate a set of trajectory samples emanating from a given seed location. It is
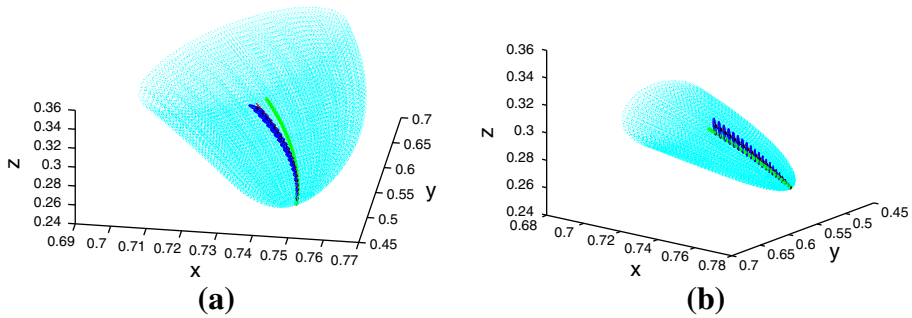
Fig. 4 The *solid black curves* with blue confidence ellipsoids correspond to the estimated integral curves using our approach. The *solid green curves* with cyan confidence ellipsoids correspond to the estimated integral curves using the method of Koltchinskii et al. (2007). The *solid red curves* are the true integral curves. The initial point is (0.75, 0.5, 0.25). We use a 95 % confidence level to calculate the confidence ellipsoids. Here $m = 8, n = 11^3 = 1331, \delta = 0.005, \beta = 0.1$. **a** The matrix Σ is fixed. **b** The matrix Σ varies with location. (Color figure online)

implemented in an open source package of programs known as the FMRIB diffusion toolbox. We use two synthetic examples to illustrate its differences from our approach. The first is a thin, curved trajectory that resembles the letter C. In the second example we mix the letter C with a straight flat trajectory, thus producing a trajectory pattern shaped like the letter Y. These two general patterns are commonly observed in real diffusion MRI scans of the human brain: the *corpus callosum* is a major inter hemispheric white matter tract, many of whose fibers are shaped like the letter C; and there are many examples of fiber tracts that converge, diverge, or "kiss," thus leading to a Y-like splitting of trajectories. We show that in both cases, our method provides better geometrical representations with theoretically well understood statistical properties, using a fixed amount of computation.

FMRIB produces two main outputs: a set of fiber trajectories emanating from a seed location, resulting from its MCMC sampling; and an occupancy image—an image representing how many of those trajectories pass through each voxel in the space of the input image. Our approach provides a different type of output: a mean trajectory and point wise covariance matrices representing possible deviations from that mean. To be fair to FMRIB, we did not wish to calculate a mean trajectory or pointwise covariances from their outputs, since this step could introduce artifacts that were never intended by the authors of FMRIB. Instead, we show our trajectory means and covariances side by side with their trajectory samples for qualitative comparison. We also created p-value maps as introduced in Koltchinskii et al. (2007), and compare those to the FMRIB occupancy images. More precisely, we calculate the p-value of the test of the null hypothesis that a fiber starting at the initial point reaches a point in the slice, using our connectivity tests as in Corollary 1. Then for all points in the slice we visualize their corresponding p-values in color.

For the first example we have $d = 3, G = [0, 1]^3$. We start with the unit vectors parallel to $(x_2, x_1, 0)^*, (-x_1, x_2, 0)^*, (0, 0, x_3)$. Those serve as the eigenvectors of $M(x)$ with the corresponding eigenvalues (10, 2, 1) for locations in G satisfying $|\sqrt{x_1^2 + x_2^2} - 0.5| < 0.05$ and $|x_3 - 0.5| < 0.05$. We use the same orthogonal gradient encoding as in the previous example, $m = 1$, and the independent normal vectors $\xi_j, j = 1, \ldots, n$ with means 0 and variances 0.01. We use a regular grid with 50 knots for x_1 and x_2 and 25 knots for x_3. Thus the sample size is $n = 62,500$.

(a)   **(b)**

(c)   **(d)**

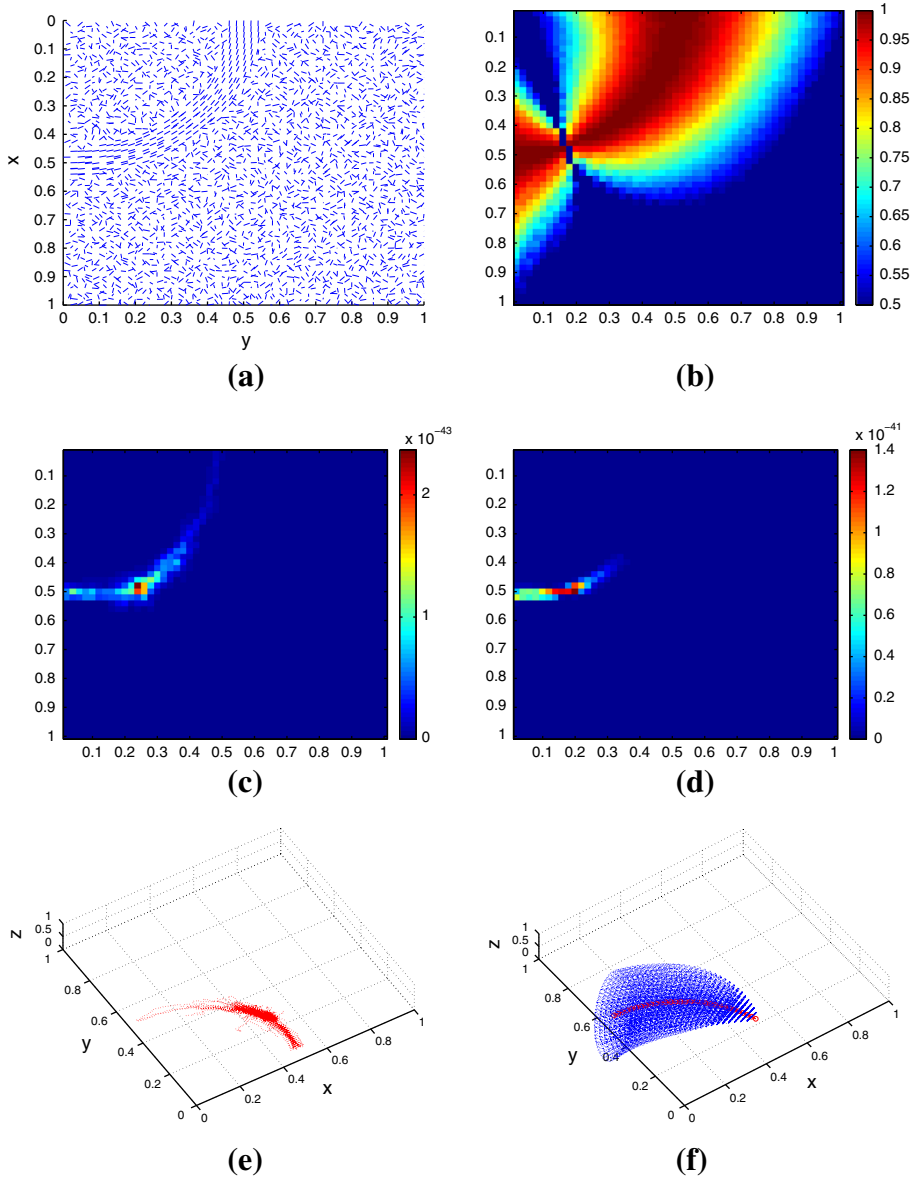**(e)**   **(f)**

Fig. 5 C-template. **a** Projection of the eigenvector field on the slice $z = 0.48$. **b** P-value map for slice $z = 0.48$. The initial point is $(0.47, 0.01706, 0.48)$. Here $\delta = 0.02$, $\beta = 0.0001$. **c** Slice $z = 0.48$. Visualization of how many MCMC tracks terminated at each point. The order is 10^{-43}. **d** Slice $z = 0.52$. Visualization of how many MCMC tracks terminated at each point. The order is 10^{-41}. **e** 400 out of 10,000 MCMC tracks. **f** Our estimator together with pointwise 95 % confidence ellipsoids. (Color figure online)

Figure 5a shows the eigenvector field calculated at the grid points and projected on one of two horizontal slices that contain the bundle of fibers. The initial point for the tracking procedure is chosen in the first of the two slices. Figure 5b shows our p-value map. Figure 5c and d provide visualizations of the FMRIB occupancy images. The scale for the slice $z = 0.52$

is 100 times higher than the scale for the slice $z = 0.48$ that contains the initial point. Figure 5e shows 400 of the total of 10,000 MCMC tracks emanating from the initial point. Figure 5f provides our estimated trajectory with 95 % confidence ellipsoids along it.

Most of the trajectory samples from FMRIB are short and are concentrated near the initial point. They also deviate substantially outside the plane in which the C-shaped trajectory was defined. The number of trajectory samples—10,000—is provided as a default parameter setting by FMRIB, but there are no theoretically valid reasons to determine an optimal number of samples. Meanwhile, our approach provides theoretically grounded visualizations of the statistical uncertainty of our estimator that conform well to the general shape of the "C" (Fig. 5b, f), and the uncertainty estimates were calculated using closed-form expressions.

To construct the second example we take the tensor from the previous example and mix it with the tensor whose eigenvectors are parallel to $(1, 1, 0)^*$, $(-1, 1, 0)^*$ and $(0, 0, 1)^*$. The eigenvalues are 10, 2, 1. This second tensor exists only for locations in G such that $|x_1 - x_2| < 0.05$, $x_1 > 0.32$, $x_2 > 0.32$, $|x_3 - 0.5| < 0.05$. We use the mixing weight $\pi = 0.5$ for the locations where these tensors mix.

Figure 6 has the same sub-figures as Fig. 5. We again observe that MCMC trajectory samples tend to be short and concentrated near the initial point. They trace out the letter C much better than the remaining part of letter Y. The Y pattern is almost non-existent on Fig. 6e, where only one track out of 400 traces the diagonal. It seems to appear on Fig. 6c, while part of the letter C disappears, but the scale is 1/100-th of the scale of Fig. 6d, where there is no Y pattern present. On contrary, our procedure traces out the Y pattern completely but provides very wide confidence ellipsoids for the diagonal part. This indicates high uncertainty of the estimator in the diagonal part of the letter Y. The p-value map shows similar results, again suggesting that our estimator provides superior geometric representation of the underlying tensor field and the uncertainty of trajectories through it.

8 Application to real brain imaging data

Our method was applied to a DTI scan of an elderly individual who volunteered for research at the UC Davis Alzheimer's Disease Center. Imaging was performed at the UC Davis Imaging Research Center on a 1.5T GE Signa Horizon LX Echospeed system. The single-shot spin-echo echo planar imaging DTI sequence had acquisition parameters including: TE: 94 ms, TR: 8000 ms, Flip angle: 90 degrees, Slice thickness: 5 mm, slice spacing: 0.0 mm, FOV: 22 cm $\times$ 22 cm, Matrix: 128×128, B-value: 1000 s/mm^2. Each acquisition included collection of 2 images with no gradient applied and 4 diffusion-weighted images acquired along each of 6 gradient directions. The directions are $b_1 = 2^{-1/2}(1, 0, 1)$, $b_2 = 2^{-1/2}(1, 0, -1)$, $b_3 = 2^{-1/2}(0, 1, 1)$, $b_4 = 2^{-1/2}(0, -1, 1)$, $b_5 = 2^{-1/2}(1, 1, 0)$, $b_6 = 2^{-1/2}(-1, 1, 0)$. This is referred to as the oblique double gradient encoding in the DTI literature. The dataset contains observations of Y on a regular grid with $128 \times 128 \times 19$ locations. The voxel dimensions are $1.875 \times 1.875 \times 5$ mm in the x, y and z directions, respectively. We scale the grid to $G = [0, 1]^2 \times [0, 3/8]$. The dataset has $m = 4$. The matrix B corresponding to the gradient directions is $B = -0.25(1, 0, 2, 0, 0, 1; 1, 0, -2, 0, 0, 1; 0, 0, 0, 1, -2, 1; 0, 0, 0, 1, 2, 1; 1, -2, 0, 1, 0, 0; 1, 2, 0, 1, 0, 0)$.

We used the standard 3D Gaussian kernel K, bandwidths $h_n = (n/\beta)^{-1/6}$, $\tilde{h}_n = (n/\beta)^{-1/4} \log n$, $\tilde{\tilde{h}}_n = (n/\beta)^{-1/5} \log n$ with $\beta = 0.0001$. Reducing β did not change results substantially, while increasing it produced very wide averaging windows so several tracts were averaged together.
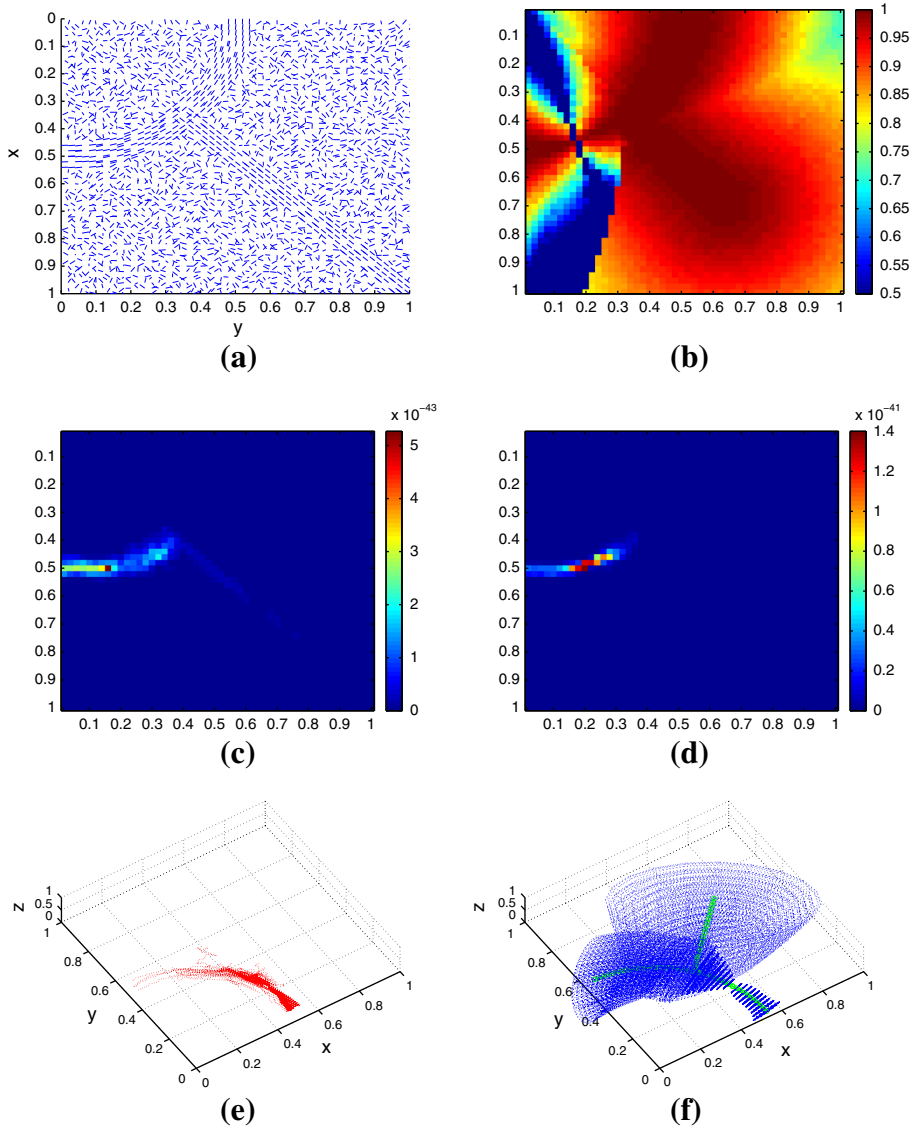
Fig. 6 Y-template. Note that the Y pattern is almost non-existent on **e**, where only one track out of 400 traces the diagonal. It seems to appear on **c**, while part of the letter C disappears, but the scale is 1/100-th of the scale of **d**, where there is no Y pattern present. **a** Slice $z = 0.48$. The initial point is (0.47, 0.01706, 0.48). Here $N = 6, m = 1, n = 62{,}500$; $\delta = 0.02, \beta = 0.0001$. **b** P-value map for slice $z = 0.48$. The initial point is (0.47, 0.01706, 0.48). Here $\delta = 0.02, \beta = 0.00005$. **c** Slice $z = 0.48$. Visualization of how many MCMC tracks terminated at each point. The order is $10^{-}43$. **d** Slice $z = 0.52$. Visualization of how many MCMC tracks terminated at each point. The order is $10^{-}41$. **e** 400 out of 10,000 MCMC tracks. **f** Our estimator together with pointwise 95 % confidence ellipsoids. (Color figure online)

First, for comparison we also estimated the fibers based on the noisy eigenvector field calculated at grid points using the method of Koltchinskii et al. (2007). Part of the vector field is shown in Fig. 7b. We also used Euler's method and the same choice of the kernel and bandwidths.
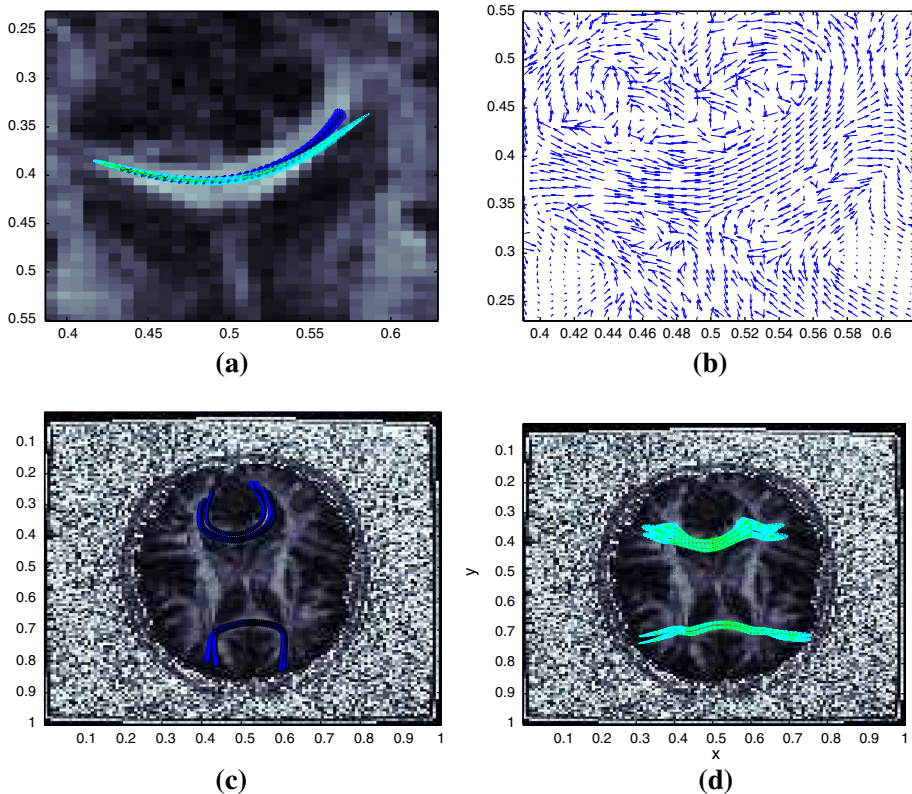
Fig. 7 Comparison of our estimator with that of Koltchinskii et al. (2007) on a real DTI data set. **a** *The solid black curve* with blue confidence ellipsoids corresponds to the estimated integral curve using our approach. The *solid green curve* with cyan confidence ellipsoids corresponds to the estimated integral curve using the method of Koltchinskii et al. (2007). We traced the fiber for 40 steps with $\delta = 0.005$. We used a 95 % confidence level to define the confidence ellipsoids. **b** The eigenvector field $v(\tilde{M}(X_j))$, $j = 1, \ldots, n$, is shown for the region in **a**. **c** Several fibers are estimated using our procedure with 95 % confidence ellipsoids. **d** Several fibers are estimated using the procedure of Koltchinskii et al. (2007) with 95 % confidence ellipsoids. (Color figure online)

We chose initial points in the center of the corpus callosum, one of the major tracts of white matter fibers in the brain. The corpus callosum fibers trace out U-shaped trajectories that run from the left side of the brain to the right.

Figure 7a shows illustrations of the estimated integral curves overlaid onto a horizontal slice of the DTI data. It demonstrates that our estimator keeps closer to the true integral curve than the Koltchinskii et al. (2007) estimator. Moreover, the confidence ellipsoids are wider for their estimator. Theirs are extended in the z direction. The Euclidean norm of the covariance matrices for our estimator at the ends of the tract is on the order of 0.0007, while it is 0.02 for Koltchinskii et al.'s (2007) estimator. A comparison of Fig. 7c and d suggests that our estimators are closer to the true fibers while all of the Koltchinskii et al.'s (2007) estimators drift away from the true U-shaped trajectory of the fiber. Choosing a smaller β to shrink the averaging window does not change the overall performance of their estimator.

Second, for comparison we consider a popular tractography approach by Behrens et al. (2003). Figure 8 has the same parts as Fig. 5. Upon its inspection we again observe that MCMC tracks tend to be short and concentrated near the initial point. Also quite a lot of
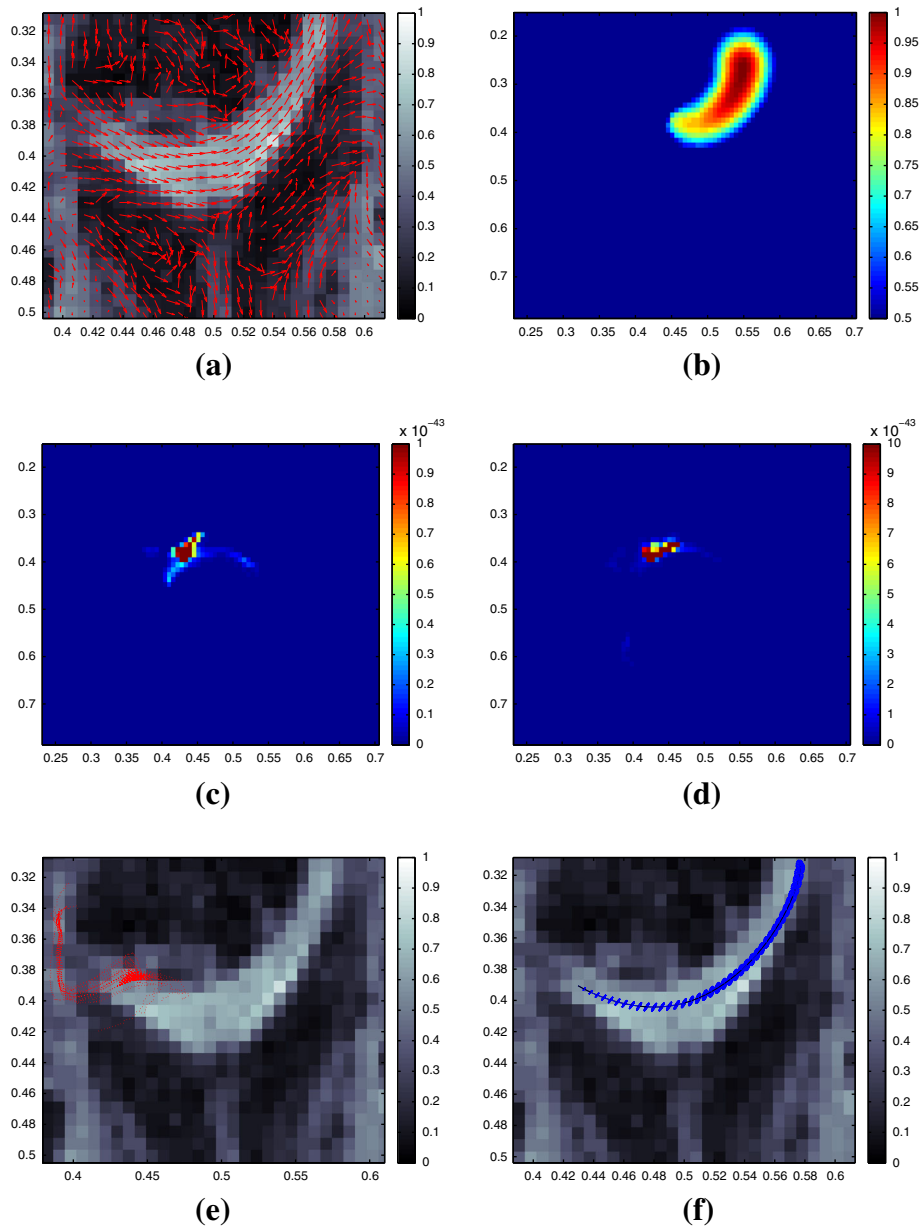
Fig. 8 Comparison of our approach with a probabilistic tractography approach due to Behrens et al. (2003). The initial point is $(0.47, 0.01706, 0.48)$. Here $N = 6, m = 2, n = 311{,}296$. **a** FA map for slice $z = 9/19$ is overlaid with the principle eigenvector field. **b** P-value map for the selected fiber in slice $z = 9/19$ is constructed with $\delta = 0.02, \beta = 0.00005$. **c** Slice $z = 9/19$. Visualization of how many MCMC tracks terminated at each point. The order is 10^{-43}. **d** Slice $z = 10/19$. Visualization of how many MCMC tracks terminated at each point. The order is 10^{-41}. **e** Slice $z = 9/19$. 400 out of 10,000 MCMC tracks. **f** Slice $z = 9/19$. Our estimator together with pointwise 95 % confidence ellipsoids. Here $\delta = 0.02, \beta = 0.00005$. (Color figure online)

them deviate in the *z*-direction. Note that scale is different for Fig. 8c and d. These MCMC tracks do not trace out the underlying letter U on FA map. On the contrary, our procedure traces out the letter U and it provides nice confidence ellipsoids that lay completely inside the fiber. The p-value map shows similar results.

9 Discussion and conclusion

A great number of scientific analyses based on tracing axonal fibers through DTI data have been published, many of which rely on probabilistic tractography to deal with the uncertainty inherent to the fiber trajectories. Yet, the theoretical statistical properties of these probabilistic tractography algorithms are not well understood, and they are computationally burdensome due to the need for sampling from a high-dimensional distribution. This work addresses the theoretical and computational problems by developing a rigorous statistical framework that provides theoretically-grounded fiber trajectory estimates, along with closed-form solutions for confidence regions. We use a nonparametric approach to model the tensor field, its leading eigenvector field and the corresponding integral curve. We derive asymptotical theory for the proposed estimator. Applications to synthetic and real human brain DTI data suggest that our method provides realistic trajectory tracing with tighter confidence ellipsoids than a more restrictive nonparametric approach. It also yields more intuitive geometric representations and realistic output images than a popular probabilistic approach due to Behrens et al. (2003).

Besides its firmer theoretical foundations, the proposed method is also advantageous over probabilistic tractography in terms of computational complexity. The proposed method requires $O(n^{3/(d+3)})$ operations to estimate the integral curve when a kernel with bounded support is used, and $O(n)$ operations for Gaussian or other kernels. It requires $O(n^2)$ operations for the calculation of asymptotical covariance, which is needed to construct confidence ellipsoids. The constants in front of these $O(\cdot)$ terms are on the order of hundreds. In contrast, the estimator based on a probabilistic tractography approach due to Behrens et al. (2003) requires $O(n\tau_n)$ operations, where τ_n is an average number of iterations needed for MCMC to converge. To obtain *ad hoc* bootstrapped confidence regions, one needs $O(n\tau_n)$ operations with an extremely large constant; there is no firm theoretical upper bound on τ_n either. Behrens et al. (2003) used $20{,}000{,}000n\tau_n$ iterations in their experiments. While in principle, the number of MCMC iterations required for any individual data set may be lower than this, in practice the number of iterations required is completely arbitrary. In our simulation study our method took several minutes to run while the method of Behrens et al. (2003) took hours. A more comprehensive and fair comparison of real-world computational complexity between our method and probabilistic tractography methods such as that of Behrens et al. would require a large-scale Monte Carlo study involving repeated imaging of the same brain under the same conditions, followed by fiber tracing in each scan. For particular fiber tracts of interest, the different tracing methods applied to the repeated scans would then provide empirical distributions of integral curves, their uncertainties, and their run times. These distributions could be the basis for a fair comparison of the methods.

We showed that the one pre-existing method that provided theoretically rigorous trajectory confidence bounds, by Koltchinskii et al. (2007), is a special case of ours and that the asymptotical distribution of their estimator coincides with that of our estimator under strong restrictions on our model. We argued that the assumptions required by their model are too unrealistic for the method to be useful for DTI, and our experiments with real DTI data supported this view. In these experiments, the estimator of Koltchinskii et al. (2007)

had covariance matrices with larger norms along the curve, and thus it had looser confidence ellipsoids. It also traced incorrect trajectories in regions with multiple fibers located in a close proximity. These experiments suggest that the theoretical rigor of our approach is actually relevant from the point of view of tract tracing applications.

Given that integral curve estimation becomes ill-conditioned at locations with multiple viable leading eigenvector directions, we first perform a preprocessing step to identify such locations, and if they are encountered during curve tracing, we begin separate integral curve traces along each viable direction. The full set of estimated trajectories emanating from a given point consists of the set of all these integral curve segments. Note that this does not introduce much more computation than a probabilistic tractography approach, which can be thought of as sampling trajectories over all such viable leading eigenvector directions. We do not think it is difficult to extend this approach to greater numbers of possible leading eigenvector directions; however, a key limitation of the current method is that uncertainty in whether or not a location constitutes a branching point is not translated into uncertainty in the tensors, eigenvectors, and integral curves. Instead, we apply a hard threshold to a branching point test statistic and use it to classify all locations as either absolutely branching points, or not. How to incorporate this branching point uncertainty into the model in a principled way is unclear. Progress in this area will be required to insure that all relevant sources of uncertainty are brought to bear on integral curve uncertainty.

In addition, it is unclear how this analysis extends to more elaborate mathematical representations of the water diffusion directional distribution that are ascendant in the current DTI literature. Some examples of these representations include diffusion tensors that are higher than second order, spherical deconvolution models, and spherical wavelet models; see Assemlal et al. (2011). To our knowledge, none of these models have been studied from the standpoint of theoretically rigorous statistical error analysis. Extending the current results to these novel diffusion representations will be an important next step for maintaining theoretically-sound integral curve tracing as diffusion MRI technology continues to advance.

Appendix: Proofs

Lemmas 1 and 2 are quite standard results. So the proofs are omitted.

Proof of Lemma 3

For any $\varepsilon > 0$ consider an event $A_n(\varepsilon) := \{\sup_{x \in \mathbb{R}^d} |\hat{M}_n(x) - M(x)| \leq \varepsilon\}$. By Lemma 1, $P(A_n(\varepsilon)) \to 1$ as $n \to \infty$. We have for any $t \in [0, T]$

$$
\begin{aligned}
|\hat{X}_n(t) - x(t)| &= \left| \int_0^t [v(\hat{M}_n(\hat{X}_n(s))) - v(M(x(s)))] ds \right| \\
&\leq \int_0^t |v(M(\hat{X}_n(s))) - v(M(x(s)))| ds + t \sup_{x \in \mathbb{R}^d} |v(\hat{M}_n(x)) - v(M(x))|.
\end{aligned}
\tag{8}
$$

For a fixed tensor M_0 we use Theorem 1 in Magnus (1985) to observe that $\lambda(M)$ and $v(M)$ are infinitely continuously differentiable with respect to M in a neighborhood of M_0 provided that $\lambda(M_0)$ is the simple eigenvalue of M_0, and

$$\left.\frac{\partial \lambda(M)}{\partial M_{kl}}\right|_{M_0} = (2 - \delta_{kl})v_k(M_0)v_l(M_0),$$

$$\left.\frac{\partial v_p(M)}{\partial M_{kl}}\right|_{M_0} = (1 - \delta_{kl}/2)[(\lambda(M_0)I_d - M_0)^+_{pk}v_l(M_0)$$

$$+ (\lambda(M_0)I_d - M_0)^+_{pl}v_k(M_0)] \quad \text{for} \quad 1 \le k, l, p \le d, \quad (9)$$

where A^+ stands for Moore–Penrose inverse of A. Note that $M(x)$ is a continuous function of x on a compact set $\bar{G}$ with support in G, where it has the simple maximal eigenvalue $\lambda(M(x))$. Then on the event $A_n(\varepsilon)$ we can apply Theorem 1 in Magnus (1985) and obtain

$$\sup_{x \in \mathbb{R}^d} |v(\hat{M}_n(x)) - v(M(x))| \le L_v \sup_{x \in \mathbb{R}^d} |\hat{M}_n(x) - M(x)|$$

for some finite constant L_v. Moreover, due to local infinite differentiability of v, smoothness of $M(x)$, $x \in G$, and $\bar{G}$ being a compact set we have the following Lipschitz condition $|v(M(x_1)) - v(M(x_2))| \le L_{vM}|x_1 - x_2|$ for all $x_1, x_2 \in \bar{G}$ and some finite constant L_{vM}. Then we can bound the expression in (8) on the event $A_n(\varepsilon)$ as follows

$$|\hat{X}_n(t) - x(t)| \le T L_v \sup_{x \in \mathbb{R}^d} |\hat{M}_n(x) - M(x)| + L_{vM} \int_0^t |\hat{X}_n(s) - x(s)| ds$$

for all $t \in [0, T]$. By Gronwall–Bellman inequality; see, e.g., Hille (1969), we obtain for all $t \in [0, T]$ on the event $A_n(\varepsilon)$

$$|\hat{X}_n(t) - x(t)| \le T L_v \sup_{x \in \mathbb{R}^d} |\hat{M}_n(x) - M(x)| e^{L_{vM}t} = o_P(1) \text{ as } n \to \infty.$$

$\square$

Proof of Lemma 4

Since $M(x)$ is a symmetric matrix for all $x \in G$ from condition (A1) and Γ_j, $j = 1, \ldots, n$, are symmetric matrices by definition, by construction $\hat{M}_n(x)$ is a symmetric matrix for all $x \in G$.

As in proof of Lemma 3 define an event $A_n(\varepsilon) := \{\sup_{x \in \mathbb{R}^d} |\hat{M}_n(x) - M(x)| \le \varepsilon\}$ for every $\varepsilon > 0$. By Lemma 1, $P(A_n(\varepsilon)) \to 1$ as $n \to \infty$.

Let $\mu(M_0)$ and $u(M_0)$ denote the minimal eigenvalue and the corresponding eigenvector of a matrix M_0. By Theorem 1 in Magnus (1985) the functions μ and u are infinitely differentiable in a small neighborhood of M_0, provided that $\mu(M_0)$ is the simple eigenvalue. Furthermore, $\left.\frac{\partial \mu(M)}{\partial M_{kl}}\right|_{M_0} = (2 - \delta_{kl})u_k(M_0)u_l(M_0)$, $1 \le k, l \le d$. Since $\mu(M(x)) > 0$ for all $x \in G$ and $\bar{G}$ is a compact set, then on the event $A_n(\varepsilon)$ the function $\mu(\hat{M}_n(x))$ is positive for all $x \in G$ and $\mu(\hat{M}_n(x))$ remains the simple minimal eigenvalue of $\hat{M}_n$ by Theorem 1 in Magnus (1985) and due to condition (A1) that all eigenvalues of $M(x)$, $x \in G$, are simple.

$\square$

Proof of Theorem 1

Step 1: *Approximation* For any $\varepsilon > 0$ consider an event $B_n(\varepsilon) := \{\sup_{t \in [0,T]} |\hat{X}_n(t) - x(t)| \le \varepsilon\}$. By Lemma 3, $P(B_n(\varepsilon)) \to 1$ as $n \to \infty$. Following the notation and ideas of the proof of Theorem 1 in Koltchinskii et al. (2007) we have

$$\hat{X}_n(t) - x(t) = \int_0^t [v(\hat{M}_n(\hat{X}_n(s))) - v(M(x(s)))]ds$$

$$= \int_0^t \nabla v(M(x(s))) \nabla M(x(s))(\hat{X}_n(s) - x(s))ds$$

$$+ \int_0^t \nabla v(M(x(s)))(\hat{M}_n(x(s)) - M(x(s)))ds + R_n(t), \quad t \in [0, T],$$

where the remainder is defined as

$$R_n(t) = \int_0^t [v(M(\hat{X}_n(s))) - v(M(x(s))) - \nabla v(M(x(s)))\nabla M(x(s))(\hat{X}_n(s) - x(s))]ds$$

$$+ \int_0^t [v(\hat{M}_n(\hat{X}_n(s))) - v(M(\hat{X}_n(s))) - \nabla v(M(x(s)))(\hat{M}_n(x(s)) - M(x(s)))]ds$$

for all $t \in [0, T]$. Note that due to local smoothness of v on the event $B_n(\varepsilon)$ we have

$$v(M(\hat{X}_n(s))) - v(M(x(s))) - \nabla v(M(x(s)))\nabla M(x(s))(\hat{X}_n(s) - x(s))$$

$$= \int_0^1 \int_0^1 \delta \Big\langle \nabla(\nabla v(M(x_{n,\tau,\delta}(s)))\nabla M(x_{n,\tau,\delta}(s)))(\hat{X}_n(s) - x(s)), \hat{X}_n(s) - x(s) \Big\rangle d\delta d\tau,$$

where $x_{n,\tau,\delta}(s) = \tau\delta\hat{X}_n(s) + (1 - \tau\delta)x(s)$. Quite similarly, on the event $A_n(\varepsilon) \cap B_n(\varepsilon)$ with $A_n(\varepsilon)$ defined in the proof of Lemma 3, a straightforward derivation yields a similar representation for the second term in $R_n(t)$. Thus, we can bound the remainder R_n on the event $A_n(\varepsilon) \cap B_n(\varepsilon)$ as

$$|R_n(t)| \leq C_1 \int_0^t |\hat{X}_n(s) - x(s)|^2 ds + C_2 \sup_{x \in \mathbb{R}^d} |\hat{M}_n(x) - M(x)| \int_0^t |\hat{X}_n(s) - x(s)| ds$$

$$+ C_3 \sup_{x \in \mathbb{R}^d} |\hat{M}_n(x) - M(x)|^2 + C_4 \sup_{x \in \mathbb{R}^d} |\nabla\hat{M}_n(x) - \nabla M(x)| \int_0^t |\hat{X}_n(s) - x(s)| ds$$

for all $t \in [0, T]$, where C_1, C_2, C_3, C_4 are some finite positive constants. In other words, using Lemmas 1–3

$$\sup_{t \in [0,T]} |R_n(t)| = o_P \left(\int_0^T |\hat{X}_n(s) - x(s)| ds + \sup_{x \in \mathbb{R}^d} |\hat{M}_n(x) - M(x)|^2 \right).$$

Denote a d-dimensional random process

$$Z_n(t) := \int_0^t \nabla v(M(x(s)))\nabla M(x(s)) Z_n(s) ds$$

$$+ \int_0^t \nabla v(M(x(s)))(\hat{M}_n(x(s)) - M(x(s))) ds, \quad t \in [0, T].$$

This process can also be written as

$$Z_n(t) = \int_0^t U(t, s) \nabla v(M(x(s)))(\hat{M}_n(x(s)) - M(x(s))) ds, \quad t \in [0, T],$$

where $U(t, s)$ is the Green's function defined just before Theorem 1. The process $Z_n(t), t \in [0, T]$, approximates the process $(\hat{X}_n - x)(t), t \in [0, T]$. Indeed, for any

$t \in [0, T]$

$$|\hat{X}_n(t) - x(t) - Z_n(t)|$$

$$= \left| \int_0^t \nabla v(M(x(s))) \nabla M(x(s))[\hat{X}_n(s) - x(s) - Z_n(s)] ds + R_n(t) \right|$$

$$\leq \int_0^t |\nabla v(M(x(s))) \nabla M(x(s))||\hat{X}_n(s) - x(s) - Z_n(s)| ds + \sup_{t \in [0,T]} |R_n(t)|.$$

Then by Gronwall–Bellman inequality we obtain on the event $A_n(\varepsilon) \cap B_n(\varepsilon)$

$$\sup_{t \in [0,T]} |\hat{X}_n(t) - x(t) - Z_n(t)|$$

$$\leq \sup_{t \in [0,T]} |R_n(t)| \sup_{t \in [0,T]} \exp \left\{ \int_0^t |\nabla v(M(x(s))) \nabla M(x(s))| ds \right\}$$

$$\leq C \sup_{t \in [0,T]} |R_n(t)|,$$

which yields

$$\sup_{t \in [0,T]} |\hat{X}_n(t) - x(t) - Z_n(t)| = o_P \left(\int_0^T |Z_n(s)| ds + \sup_{x \in \mathbb{R}^d} |\hat{M}_n(x) - M(x)|^2 \right).$$

It will follow that $\int_0^T |Z_n(s)| ds = O_P((nh_n^{d-1})^{-1/2})$, which in combination with Lemma 1 leads to $\sup_{t \in [0,T]} |\hat{X}_n(t) - x(t) - Z_n(t)| = o_P((nh_n^{d-1})^{-1/2})$.

Step 2: *Mean and covariance of Z_n* Similar to Koltchinskii et al. (2007) define a $d \times d^2$-tensor-valued function $f_t(s) = I_{[0,t]}(s) U(t, s) \nabla v(M(x(s)))$, for $s \in \mathbb{R}$, $t \in [0, T]$. Note that it belongs to $\mathcal{L}$, the set of all $\mathbb{R}^{d^3}$-valued bounded functions on $\mathbb{R}$ with support in $[0, T]$, that are a.e. continuous in $\mathbb{R}$. Furthermore, note that $\mathcal{L}$ is a linear space. Then we can rewrite for $t \in [0, T]$

$$Z_n(t) = \int f_t(s)(\hat{M}_n(x(s)) - M(x(s))) ds$$

$$= \frac{1}{nh_n^d} \sum_{j=1}^n \int f_t(s) K((x(s) - X_j)/h_n) ds (M(X_j) + \Gamma_j) - \int f_t(s) M(x(s)) ds.$$

We have for all $t \in [0, T]$ with o-term being uniform in t

$$\mathbb{E} Z_n(t) = h_n^{-d} \int f_t(s) \int K((x(s) - y)/h_n) M(y) dy ds - \int f_t(s) M(x(s)) ds$$

$$= -h_n \int f_t(s) \nabla M(x(s)) \int u K(u) du ds$$

$$+ 0.5 h_n^2 \int f_t(s) \int K(u) \langle \nabla^2 M(x(s)) u, u \rangle du ds + o(h_n^2).$$

Next consider the components of the covariance matrix. It is more convenient to treat $d \times d$ matrices as d_0-vectors in the lengthy derivations below. To this end, for any $t_1, t_2 \in [0, T]$ we have

$$nh_n^{2d}\mathrm{Cov}(Z_n(t_1), Z_n(t_2))$$

$$= \mathrm{Cov}\left(\int K((x(s) - X_1)/h_n)U(t_1, s)\tilde{\nabla}v(M(x(s)))H\Gamma_1 ds,\right.$$

$$\left.\int K((x(u) - X_1)/h_n)U(t_2, s)\tilde{\nabla}v(M(x(u)))H\Gamma_1 du\right)$$

$$+ \mathrm{Cov}\left(\int K((x(s) - X_1)/h_n)U(t_1, s)\tilde{\nabla}v(M(x(s)))H\mathbf{M}(X_1)ds\right.$$

$$\left.\int K((x(u) - X_1)/h_n)U(t_2, s)\tilde{\nabla}v(M(x(u)))H\mathbf{M}(X_1)du\right)$$

$$=: nh_n^{2d}(I + II).$$

Then using two substitutions $z = (x(s) - y)/h_n$ and $u = s + \tau h_n$, also utilizing Lebesgue dominated convergence theorem [since $\int |\Sigma_\Gamma|^2(x)dx < \infty$, ∇M and ∇v are uniformly bounded everywhere, $\int K^4(z)dz < \infty$, and there exists $\Delta > 0$ such that $|\int_0^1 v(M(x(s + \delta(t - s))))d\delta| > \Delta$ for all $s, t \in (0, T)$, $s \neq t$], we can rewrite

$$I = \frac{1}{nh_n^{2d}}\mathbb{E}\mathbb{E}\left(\int\int K\left(\frac{x(s) - X_1}{h_n}\right)K\left(\frac{x(u) - X_1}{h_n}\right)U(t_1, s)\tilde{\nabla}v(M(x(s)))\right.$$

$$\left.\times H\Gamma_1\Gamma_1^* H\tilde{\nabla}v(M(x(u)))^*U^*(t_2, u)dsdu\Big|X_1\right) = \frac{1}{nh_n^{d-1}}\int\int\int K(z)$$

$$\times K\left(z + \tau\int_0^1 v(M(x(s + \tau\delta h_n)))d\delta\right)U(t_1, s)\tilde{\nabla}v(M(x(s)))H\Sigma_\Gamma(x(s) - zh_n)$$

$$\times H\tilde{\nabla}v(M(x(s + \tau h_n)))^*U^*(t_2, s + \tau h_n)dsd\tau dz = \frac{1 + o(1)}{nh_n^{d-1}}$$

$$\int \psi(v(M(x(s))))U(t_1, s)\tilde{\nabla}v(M(x(s)))H\Sigma_\Gamma(x(s))H\tilde{\nabla}v(M(x(s)))^*U^*(t_2, s)ds$$

with o-term being uniform in t. Quite similarly,

$$II = \frac{1}{nh_n^d}\int\int\int K(z)K(z + (x(u) - x(s))/h_n)U(t_1, s)\tilde{\nabla}v(M(x(s)))$$

$$\times H\mathbf{M}(x(s) - zh_n)\mathbf{M}^*(x(s) - zh_n)H\tilde{\nabla}v(M(x(u)))^*U^*(t_2, u)dsdudz$$

$$- \frac{1}{n}\left(\int\int K(z)U(t_1, s)\tilde{\nabla}v(M(x(s)))H\mathbf{M}(x(s) - zh_n)dsdz\right)$$

$$\times\left(\int\int K(z)\mathbf{M}^*(x(s) - zh_n)H\tilde{\nabla}v(M(x(u)))^*U^*(t_2, u)dudz\right)$$

$$= \frac{1 + o(1)}{nh_n^{d-1}}\int \psi(v(M(x(s))))U(t_1, s)\tilde{\nabla}v(M(x(s)))H\mathbf{M}(x(s))\mathbf{M}^*(x(s))$$

$$\times H\tilde{\nabla}v(M(x(s)))^*U^*(t_2, s)ds$$

with o-term being uniform in t, where we again used Lebesgue dominated convergence theorem due to the same reasons as in the case of (I) and due to $M(\cdot)$ being uniformly bounded everywhere.

Thus, for all $t_1, t_2 \in [0, T]$ the covariance of Z_n is

$$\text{Cov}(Z_n(t_1), Z_n(t_2)) = \frac{1 + o(1)}{nh_n^{d-1}} \int \psi(v(M(x(s))))U(t_1, s)\tilde{\nabla}v(M(x(s)))$$

$$\times H[\Sigma_\Gamma(x(s)) + \mathbf{M}(x(s))\mathbf{M}^*(x(s))]H\tilde{\nabla}v(M(x(s)))^*U^*(t_2, s)ds.$$

Step 3: *Lyapunov's condition and asymptotical equicontinuity* Next we show that $\sqrt{nh_n^{d-1}}(\hat{X}_n(t) - x(t))$ is asymptotically normal for all $t \in [0, T]$ by checking Lyapunov's conditions of CLT. Let

$$\eta_j = \int f_t(s)K((x(s) - X_j)/h_n)ds(M(X_j) + \Gamma_j), \quad j \geq 1, \quad t \in [0, T],$$

so that $\sqrt{nh_n^{d-1}}(Z_n(t) - \mathbb{E}Z_n(t)) = \frac{1}{\sqrt{nh_n^{d+1}}} \sum_{j=1}^n (\eta_j - \mathbb{E}\eta_j)$. Then

$$\mathbb{E}|\eta_j|^4 \leq Ch_n^{d+3} \times \iiint \Lambda(\tau_1, \tau_2, \tau_3)\left(\int |f_t(s)|^4 ds\right)^{1/4}\left(\int |f_t(s + \tau_1 h_n)|^4 ds\right)^{1/4}$$

$$\times \left(\int |f_t(s + \tau_2 h_n)|^4 ds\right)^{1/4}\left(\int |f_t(s + \tau_3 h_n)|^4 ds\right)^{1/4} d\tau_1 d\tau_2 d\tau_3$$

$$\leq Ch_n^{d+3} \int |f_t(s)|^4 ds \iiint \Lambda_4(\tau_1, \tau_2, \tau_3) d\tau_1 d\tau_2 d\tau_3, \qquad (10)$$

where the function $\Lambda : \mathbb{R}^3 \to \mathbb{R}$ is defined as

$$\Lambda_4(\tau_1, \tau_2, \tau_3) := \sup_{s_1, s_2, s_3, s_4 \in [0, T]} \int K(z)K(z + \tau_1(x(s_1) - x(s_4))/(s_1 - s_4))$$

$$\times K(z + \tau_1(x(s_2) - x(s_4))/(s_2 - s_4))K(z + \tau_1(x(s_3) - x(s_4))/(s_3 - s_4))dz.$$

It is integrable, since it is bounded and has a bounded support due to conditions (A3) and (A7). Then $\mathbb{E}(\sqrt{nh_n^{d-1}}(Z_n(t) - \mathbb{E}Z_n(t)))^4 \leq \frac{Cnh_n^{d+3}}{n^2 h_n^{2(d+1)}} = \frac{C}{nh_n^{d-1}} \to 0$ as $n \to \infty$. By Lyapunov's condition of CLT we have the f.d.d. convergence of the stochastic processes $\sqrt{nh_n^{d-1}}Z_n(t), t \in [0, T]$, and $\sqrt{nh_n^{d-1}}(\hat{X}_n(t) - x(t)), t \in [0, T]$, to the Gaussian process $\mathcal{G}(t), t \in [0, T]$.

Finally, we need to check the asymptotic equicontinuity condition for the sequence of processes $\sqrt{nh_n^{d-1}}(Z_n(t) - \mathbb{E}Z_n(t)), t \in [0, T]$. As in Koltchinskii et al. (2007) we use the formula

$$\mathbb{E}|\sqrt{nh_n^{d-1}}(Z_n(t) - \mathbb{E}Z_n(t))|^4 = \frac{1}{n^4 h_n^{4d}}\left[0.5n(n-1)\left(\mathbb{E}|\eta - \mathbb{E}\eta|^2\right)^2 + n\mathbb{E}|\eta - \mathbb{E}\eta|^4\right].$$

Similar to (10) we get $\mathbb{E}|\eta_j|^2 \leq Ch_n^{d+1} \int |f_t(s)|^2 ds \int \Lambda_2(\tau)d\tau$, where

$$\Lambda_2(\tau) := \sup_{s_1, s_2 \in [0, T]} \int K(z)K(z + \tau(x(s_1) - x(s_2))/(s_1 - s_2))dz, \quad \tau \in \mathbb{R}.$$

This function is integrable since it is bounded and has a bounded support due to conditions (A3) and (A7). Then with a finite constant C

$$\mathbb{E}|\sqrt{nh_n^{d-1}}(Z_n(t) - \mathbb{E}Z_n(t))|^4 \leq C\left[\left(\int |f_t(s)|^2 ds\right)^2 + \frac{1}{nh_n^{d-1}}\int |f_t(s)|^4 ds\right],$$

which we will apply to $f_{t_1} - f_{t_2}$ instead of f_t. Due to continuity of U, ∇v and M we have $\int |f_{t_1}(s) - f_{t_2}(s)|^2 ds \leq C|t_1 - t_2|$ and $\int |f_{t_1}(s) - f_{t_2}(s)|^4 ds \leq C|t_1 - t_2|$ for all $t_{1,2} \in [0, T]$, which yields

$$\mathbb{E}|\sqrt{nh_n^{d-1}}[(Z_n(t_1) - \mathbb{E}Z_n(t_1)) - (Z_n(t_2) - \mathbb{E}Z_n(t_2))]|^4 \leq C\left[|t_1 - t_2|^2 + \frac{|t_1 - t_2|}{nh_n^{d-1}}\right].$$

The asymptotic equicontinuity follows by arguments in Billingsley (1999) as $nh_n^{d-1} \to \infty$. For more details see Koltchinskii et al. (2007). $\qquad\square$

Proof of Proposition

It follows from a more general result on the asymptotical joint distribution of eigenvalues of $\hat{M}_n(x)$ for a fixed point $x \in G$. To formulate it we need new notation $\lambda_1(A) > \lambda_2(A) > \cdots > \lambda_d(A)$ for ordered eigenvalues of a symmetric positive definite matrix A. Let $v^{(i)}(A)$, $i = 1, \ldots, d$, be the corresponding eigenvectors of A. With slight abuse of notation, let $\lambda(A)$ be the vector of all ordered eigenvalues of A.

Proposition 2 *Suppose conditions (A1)–(A8) hold. Then for any fixed $x \in G$ the difference $\sqrt{nh_n^d}(\lambda(\hat{M}_n(x)) - \lambda(M(x)))$ is asymptotically normal with mean 0 and the covariance matrix with the i j-th entries*

$$\sigma_{\lambda_i \lambda_j}(x) = \int K^2(u)du\,\boldsymbol{\Delta}_{v_i}^*(x)H[M(x)M^*(x) + \Sigma_\Gamma(x)]H\boldsymbol{\Delta}_{v_j}(x), \quad 1 \leq i, \ j \leq d,$$

where $\boldsymbol{\Delta}_{v_i}$ is defined in Proposition 1.

The proof is based on the standard asymptotical normality result for the kernel regression estimator $\hat{M}_n(x)$ at a fixed point $x \in G$, since $\boldsymbol{\Gamma}_j$, $j = 1, \ldots, n$, are centered i.i.d. random variables. Then we utilize the linearization with respect to the tensor of the eigenvalues based on the first order Taylor expansion (9) in the proof of Lemma 3. A straightforward calculation of asymptotical covariances completes the proof. $\qquad\square$

References

Assemlal H-E, Tschumperle D, Brun L, Siddiqi K (2011) Recent advances in diffusion MRI modeling: angular and radial reconstruction. Med Image Anal 15:369–396

Bammer R, Holdsworth S, Veldhuis W, Skare S (2009) Newmethods in diffusion-weighted and diffusion tensor imaging. Magn Reson Imaging Clin N Am 17:175–204

Basser P, Pajevic S (2000) Statistical artifacts in diffusiontensor MRI (DTI) caused by background noise. Magn Reson Med 44:41–50

Basser PJ, Pajevic S, Pierpaoli C, Duda J, Aldroubi A (2000) In vivo fiber tractography using DTI data. Magn Reson Med 44:625–632

Basser P, Pierpaoli C (1998) A simplified method to measurethe diffusion tensor from seven MR images. Magn Reson Med 39:928–934

Beaulieu C (2002) The basis of anisotropic water diffusion in thenervous system—a technical review. NMR Biomed 15:435–455

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

Billingsley P (1999) Convergence of probability measures, 2nd edn. Wiley, New York

Chanraud S, Zahr N, Sullivan E, Pfefferbaum A (2010) MR diffusion tensor imaging: a window into white matter integrity of the working brain. Neuropsychol Rev 20:209–225

Friman O, Farneback G, Westin C-F (2006) A Bayesian approach for stochastic white matter tractography. IEEE Trans Med Imaging 25:965–978

Gelman A, Carlin J, Stern H, Rubin D (2004) Bayesian data analysis. Chapman & Hall/CRC Texts in Statistical Science. Chapman & Hall/CRC, New York

Gudbjartsson H, Patz S (1995) The Rician distribution ofnoisy MRI data. Magn Reson Med 34:910–914

Hahn K, Prigarin S, Heim S, Hasan K (2006) Random noisein diffusion tensor imaging, its destructive impact and somecorrections. In: Weickert J, Hagen H (eds) Visualization and processing of tensor fields. Springer, Berlin, pp 107–117

Hahn K, Prigarin S, Rodenacker K, Hasan K (2009) Denoising for diffusion tensor imaging with low signal to noiseratios: method and Monte Carlo validation. Int J Biomath Biostat 1:63–81

Hille E (1969) Lectures on ordinary differential equations. Addison-Wesley, Reading

Jones D (2003) Determining and visualizing uncertainty inestimates of fiber orientation from diffusion tensor MRI. Magn Reson Med 49:7–12

Kato T (1980) Perturbation theory for linear operators. Springer, New York

Koltchinskii V, Sakhanenko L, Cai S (2007) Integral curves of noisy vector fields and statistical problems in diffusion tensor imaging: nonparametric kernel estimation and hypotheses testing. Ann Stat 35:1576–1607

Koltchinskii V, Sakhanenko L (2009) Asymptotics of statistical estimators of integral curves. In: Houdre C, Koltchinskii V, Mason D, Peligrad M (eds) High dimensional probability V: The Luminy Volume. IMS Collections, pp 326–337

Lazar M, Alexander A (2005) Bootstrap white mattertractography (BOOT-TRAC). Neuroimage 24:524–532

Magnus J (1985) On differentiating eigenvalues and eigenvectors. Econom Theory 1:179–191

Mammen E (1992) When does Bootstrap work? Lecture Notes in Statistics. Springer, New York

Mukherjee P, Berman J, Chung S, Hess C, Henry R (2008) Diffusion tensor MR imaging and fiber tractography: theoretic underpinnings. Am J Neuroradiol 29:632–641

Mukherjee P, Chung S, Berman J, Hess C, Henry R (2008) Diffusion tensor MR imaging and fiber tractography: technical considerations. Am J Neuroradiol 29:843–852

Parker GJM, Alexander D (2003) Probabilistic Monte Carlo based mapping of cerebral connections utilizing whole-brain crossing fiber information. In: Proceedings of IPMI'2003, pp 684–695

Sakhanenko L (2010) Lower bounds for accuracy of estimation in diffusion tensor imaging. Theory Probab Appl 54:168–177

Sakhanenko L (2011) Global rate optimality in a model for diffusion tensor imaging. Theory Probab Appl 55:77–90

Sakhanenko L (2012) Numerical issues in estimation of integral curves from noisy diffusion tensor data. Stat Probab Lett 82:1136–1144

Yuan Y, Zhu HT, Ibrahim J, Lin WL, Peterson BG (2008) A noteon bootstrapping uncertainty of diffusion tensor parameters. IEEE Trans Med Imaging 27:1506–1514

Zhu H, Zhang H, Ibrahim J, Peterson B (2007) Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance image data. J Am Stat Assoc 102:1081–1110

Zhu H, Li Y, Ibrahim I, Shi X, An H, Chen Y, Gao W, Lin W, Rowe D, Peterson B (2009) Regression models for identifying noise sources in magnetic resonance images. J Am Stat Assoc 104:623–637