Том 68 ТЕОРИЯ ВЕРОЯТНОСТЕЙ И ЕЕ ПРИМЕНЕНИЯ 2023 Выпуск 2

BANERJEE C.*, ***SAKHANENKO L. A.****, ***ZHU D. C.*****

GLOBAL RATE OPTIMALITY OF INTEGRAL CURVE ESTIMATORS IN HIGH ORDER TENSOR MODELS[1)]

Вдохновленные применениями в нейровизуализации, мы рассматриваем проблему установления глобальной минимаксной нижней границы в модели тензора высокого порядка. В частности, описываемая нами методология позволяет получить глобальную минимаксную границу для оценок интегральных кривых, предложенных в работе О. Кармайкла и второго автора 2015 г., при полупараметрической постановке задачи. Теоретические результаты настоящей работы гарантируют, что оценки, используемые для отслеживания сложной структуры волокон внутри живого человеческого мозга и построенные по данным, полученным из диффузионной тензорной МРТ с высоким угловым разрешением (HARDI), оптимальны не только локально, но и глобально. Таким образом, глобальная минимаксная граница асимптотического риска оценок предоставит квантификацию неопределенности для метода оценки во всей области изображения. В дополнение к теоретическим результатам проводится подробное эмпирическое исследование с целью определить оптимальное число градиентных направлений для протоколов нейровизуализации, которые мы далее иллюстрируем анализом сканирования мозга живого человека по реальным данным.

Ключевые слова и фразы: глобальная минимаксная нижняя граница, полупараметрическая оценка, диффузионная тензорная МРТ с высоким угловым разрешением.

DOI: https://doi.org/10.4213/tvp5534

1. Introduction.

1.1. Formulation of the problem and results. Given a fixed $N \times J_M$ design matrix B and a set of (random) points $X_1, \dots, X_n$ in a bounded open

*Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA; e-mail: banerj25@msu.edu; sakhanen@msu.edu

**Department of Radiology, Michigan State University, East Lansing, MI, USA; e-mail: zhuda@msu.edu

1)This research was partially supported by the National Science Foundation grants DMS-1612867, DMS-2111251 and partially supported by Michigan State University High Performance Computing Center through computational resources provided by the Institute for Cyber-Enabled Research.

convex subset G of $\mathbf{R}^d$, one observes

$$Y_i = B\underline{D}(X_i) + \Sigma^{1/2}(X_i)\xi_i, \tag{1}$$

where $Y_i, \xi_i \in \mathbf{R}^N$ and $\Sigma(X_i)$ is an $N \times N$ symmetric positive definite matrix. Here $\underline{D}$ is a vector of $J_M = (M+1)(M+2)/2$ unique components derived from a supersymmetric tensor $D \in \mathbf{R}^{d^M}$ of rank R and even order M, which we will describe later. Finally, ξ_i are i.i.d. zero-mean unit-variance random variables.

The main objects of interest are the integral curves $x^{(r)}(t)$ for $t \in [0,T]$, $r = 1, \dots, R$, which are the solutions of the ordinary differential equations (ODEs)

$$\frac{dx^{(r)}(t)}{dt} = v^{(r)}\big(x^{(r)}(t)\big), \quad t \in [0,T], \qquad x^{(r)}(0) = a^*, \tag{2}$$

where $a^* \in G$ is the initial point and $v^{(r)}$, $r = 1, \dots, R$, are the so-called pseudo-eigenvectors of tensor D, which would be described in detail in section 2. For $M = 2$, they are ordinary eigenvectors of the matrix D.

For $y(t) = (y_1(t), \dots, y_d(t))$, $t \in [0,T]$, let us introduce the integrated $\mathcal{L}_p$-norms given by

$$\begin{aligned} \|y\|_{p,T} &:= \left(\int_0^T \sum_{i=1}^d |y_i(t)|^p \, dt \right)^{1/p}, \qquad 1 \leqslant p < \infty, \\ \|y\|_{\infty,T} &:= \sup_{t \in [0,T]} \max_{1 \leqslant k \leqslant d} |y_k(t)|. \end{aligned} \tag{3}$$

In this paper we derive the lower bounds of the form

$$\liminf_{n \to \infty} \inf_{\widehat{X}_n^{(1)}, \widehat{X}_n^{(2)}, \dots, \widehat{X}_n^{(R)} \in \mathcal{E}_n(a_*, T)} \sup_{D \in \mathcal{D}^2(a_*, G, T)} \mathbf{E} w\big(n^{2/(d+3)} \big(\|\widehat{X}_n^{(1)} - x^{(1)}\|_{p,T}, \\ \|\widehat{X}_n^{(2)} - x^{(2)}\|_{p,T}, \dots, \|\widehat{X}_n^{(R)} - x^{(R)}\|_{p,T} \big)\big) > 0$$

for any number $T > 0$, any point $a^* \in G$, any loss function $w \in \mathcal{W}$ in a certain class of functions using the integrated $\mathcal{L}_p$-norms. Here $\mathcal{E}_n(a_*, T)$ denotes the class of all possible estimators of the integral curves based on the dataset $\{(X_i, Y_i),\ i = 1, \dots, n\}$ starting at a_* and traced for $t \in [0,T]$. Meanwhile, $\mathcal{D}^2(a_*, G, T)$ describes the class of supersymmetric twice continuously differentiable tensors that are nondegenerate in a certain sense explained in section 2. These bounds guarantee that no estimator of the integral curve can converge to the true integral curve faster than the rate $n^{-2/(d+3)}$ in the whole imaging domain G globally.

Recently, Carmichael and Sakhanenko proposed a class of integral estimators for $x^{(r)}$, $r = 1, \dots, R$, for the model (1). In [2, Theorem 3] we show that these estimators (from now on referred to as CS estimators) enjoy the

global minimax optimal rate of convergence $n^{-2/(d+3)}$. This work is distinct from our earlier work [1], where we provide regularity conditions needed for pointwise minimax optimal rate. The analysis of the minimax rates in terms of integrated $\mathcal{L}_p$-norms is more delicate than the analysis of the pointwise case. It leverages results from differential geometry with a version of Fano's lemma [26] as opposed to implementation of Hájek's lemma for locally asymptotically normal families of distributions in [1].

The considered model (1) comes from neuroimaging. Thus, as a practical application we propose and carry out a novel approach to compare different imaging protocols (designs of imaging studies) based on their empirical values of the lower bounds that we derive in Theorem 1 and Corollary 1 (see section 3). The intuition behind this approach is that by using minimax optimal estimators one can control the uncertainty up to a constant in the lower bound. Thus, protocols can be compared based on the empirical values of this constant. The protocols with the smallest values of this lower bound under other conditions being fixed would be more practical, since they would have a tighter control on the estimation error. See section 4 for details.

1.2. Neuroimaging motivation. The model (1) is used in the so-called high angular resolution diffusion imaging (HARDI), which is a popular brain imaging approach in a class of diffusion tensor imaging (DTI) techniques. Below we explain some scientific background for it to give motivation for this model.

In magnetic resonance imaging (MRI), under a strong magnetic field, the self-spinning of protons is perturbed with a radio frequency pulse. The signal from the spin dynamics after the perturbation is used to generate MR images. Since MRI contains no radiation, the potential harm to a human body is negligible. Diffusion of water molecules in the presence of magnetic field gradients leads to a MRI signal loss. This is the main mechanism behind diffusion weighted imaging.

In particular, a DT-MRI scanner records log-losses of signal for a fixed magnetic field gradient $g \in \mathbf{R}^d$, $\|g\|_2 = 1$, which are modeled according to

$$\ln\left(\frac{S(x,g)}{S_0(x)}\right) = -c \sum_{i_1=1}^{3} \sum_{i_2=1}^{3} D_{i_1 i_2}(x) g_{i_1} g_{i_2} + \sigma(x,g)\xi_g,$$

where $S(x,g)$ is the relative amount of water diffusion at voxel x along spatial direction g, and $S_0(x)$ is amount of water diffusion without any magnetic field gradient. From now on g_i, $i = 1, \dots, d$, are the components of the gradient direction, and the matrix components $D_{i_1 i_2}$, $i_1, i_2 = 1, \dots, d$, represent the amount of diffusion at voxel x. Note that the log-losses of imaging signal are described by a symmetric positive definite matrix D. The zero-mean measurement error is given by ξ_g, while $\sigma(x,g) > 0$ is a scaling function, and c is a constant that depends on several physical parameters of the imaging procedure.

The diffusion weighted imaging data are most commonly used in DTI, where diffusivity values and principal diffusion orientation can be estimated at each voxel. The healthy axons contain intact myelin sheaths that tend to align in organized orientations, and water diffusivity in voxel generally follows these directions of the axonal bundles. Neural fiber tracing is normally done by exploring the orientation of diffusion tensor at neighboring voxels. The successful application of this method can be found in [17], [28], where authors have explored the structural connectivity of the different brain regions. Other applications involve exploring brain maturation in young children (see [5]), axonal degeneration in Alzheimer's disease (see [29]).

Koltchinskii et al. [16] first proposed a nonparametric integral curve estimation method to trace neural fiber using the leading eigenvector of D^t in the presence of additive measurement errors. Using this method the authors successfully traced a thick neural fiber along with optimal confidence ellipsoids. Then Carmichael and Sakhanenko [4] considered the matrix model and constructed asymptotically normal fiber estimators with confidence ellipsoids.

However, in case of neural fibers which cross each other, branch, converge or diverge, models using just eigenvectors or symmetric positive definite matrices are not enough, and a more sophisticated modeling has to be considered. To this end, Özarslan and Hareci [20] proposed HARDI to model these phenomena of fibers:

$$\ln\left(\frac{S(x,g)}{S_0(x)}\right) = -c\sum_{i_1=1}^{3}\cdots\sum_{i_M=1}^{3} D_{i_1\ldots i_M}(x)g_{i_1}\cdots g_{i_M} + \sigma(x,g)\xi_g, \qquad (4)$$

where the tensor components $D_{i_1\ldots i_M}$ represent the amount of diffusion along the gradient combination $(g_{i_1},\ldots,g_{i_M})$ at voxel x. In equation (4), the log-losses of signal are described by a supersymmetric tensor D of even order M. We remark that odd orders would not allow positive definiteness of the diffusion tensor. The noise is described via a zero-mean measurement errors ξ_g and a scaling function $\sigma(x,g) > 0$.

Under this model Carmichael and Sakhanenko [3] proposed a nonparametric integral curve estimation method to trace fiber tracts and assess the uncertainty. For classical DTI the fiber tracks are followed in small steps along the leading eigenvector of the tensor field D, which indicates the primary direction of the water diffusion. For HARDI these tracks follow the so-called pseudo-eigenvectors, which generalize eigen-structures for matrices ($M = 2$) to high-order tensors ($M > 2$). Traditionally for neuroimaging studies the order of tensor is taken to be even numbers such as $M = 4, 6, 8$, etc. Heuristically speaking, the diffusion tensor form should be positive definite which is equivalent to homogeneous polynomial of degree M. For odd M there exists no positive definite supersymmetric tensor, therefore only even M are used.

A supersymmetric tensor D has only $J_M = (M+1)(M+2)/2$ unique components that can be stacked into a vector $\underline{D} \in \mathbf{R}^{J_M}$. A DT-MRI scanner collects the data on a set of N different gradient directions g. The log-losses of the imaging signal are then stacked into the observed vectors $Y_i \in \mathbf{R}^N$ for different locations X_i in a brain. This reformulation yields the model (1), where the known design matrix B is a complicated functional of N different gradients g_k, $k = 1, \ldots, N$.

Mathematically, the fiber tracks are modeled as the solutions of ODEs (2) driven by the pseudo-eigenvectors $v^{(r)}$, so they are the integral curves. Based on (4) and some regularity conditions the authors of [3] proposed a four step estimation method (CS estimator) to estimate integral curves $(x^{(1)}(t), \ldots, x^{(R)}(t))$, $t \in [0,T]$, which will be described later. In this paper we will show that CS estimators are minimax optimal in a global sense.

Overall, this work places integral curve estimation into nonparametric literature parallel to that for density and regression functions. Indeed, minimax principle is an important inferential tool based on linear ordering principle which is often used to find optimal estimators (see [9]). It has rich history. In nonparametric framework Stone [25] established global minimax optimal rates for the estimators in a simple nonparametric regression setting. Mammen and van de Geer [18] proposed a locally adaptive regression spline estimator, which they proved minimax optimal both locally and globally. Some more recent works include [21], where the authors have established global minimax optimal rates for sparse additive models over a reproducing kernel Hilbert space (RKHS) in a $\mathcal{L}_1$-type convex optimization framework. Guntuboyina–Sen [10] and Kim–Samworth [15] established global rate of convergence in univariate convex regression and log-concave density estimation, respectively.

In the same spirit of these works, the theorems in this paper would establish that under some mild regularity conditions the ensemble of integral curve estimators $\widehat{x}_n^{(r)}(t)$, $t \in [0,T]$, $r = 1, \ldots, R$, will have the minimax optimal convergence rate in the entire domain of the trace $[0,T]$, under the integrated $\mathcal{L}_p$-norm in (3). Theorem 1 and the corresponding corollary establish that the integrated $\mathcal{L}_p$-norm of the error in estimation for all the integral curves when considered under the loss function $w(u)$, $u \in \mathbf{R}^R$, converges to a nonzero finite constant with the asymptotic rate $n^{2/(d+3)}$ within the class of permissible estimators and within the class of permissible estimators for which the tensor field is twice continuously differentiable, respectively. Theorem 2 would establish that CS estimators $\widehat{x}_n^{(r)}(t)$, $t \in [0,T]$, $r = 1, \ldots, R$, in [3] have the supremum risk based on the integrated $\mathcal{L}_p$-norm in (3) converging to a positive finite constant with the asymptotic rate $n^{2/(d+3)}$. The results we proposed here are quite similar to those in [22], where the author has established global minimax bounds for the asymptotic risk of the integral curve estimators as solution of differential equation involving a simple vec-

tor field perturbed by an additive noise. However in this work we use of tensor field perturbed with noise which involves additional layer of complication. Thereby the theorems and proofs in this paper are quite distinct and extend the results of [22] to a different general framework.

Finally, we illustrate the implications of the theorems in the neuroimaging context. We consider the simulation example following the setup of [24]. We simulate integral curves depicting "Y" pattern fiber distribution based on signal-to-noise ratio (SNR) and thickness of fibers, which are commonly present in live brains.

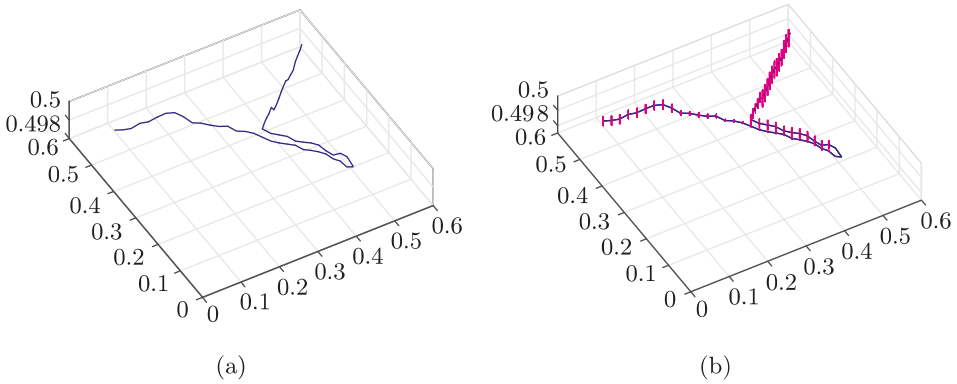

Fig. 1. (a) 3-dimensional trace of the "Y" pattern; (b) 3-dimensional trace of the "Y" pattern with 95% confidence ellipsoids. We trace the integral curve using the method of [3] creating the "Y" pattern. Here sample size n is 60^3, SNR $= 2$, thickness of the fiber ε equals 0.04, step size δ equals 2, and the constant β is 10^{-7}.

The 3D plot in Fig. 1 represents an estimated integral curve with 95% confidence ellipsoid around it computed using the method proposed in [3]. The "Y" pattern essentially has two branches; therefore, if we take a loss function $w\colon [0,\infty)^2 \to [0,\infty)$ as $w(u) = \sum_{r=1}^{2} u_r^2$ and consider an integrated $\mathcal{L}_2$-norm for the individual integral curves for the argument of w, then the empirical risk can be computed by obtaining some summary measure based on Monte Carlo simulation of the lower bound value $\kappa = w\big(n^{1/3}(\|\widehat{X}^{(1)} - x^{(1)}\|_{2,T}, \|\widehat{X}^{(2)} - x^{(2)}\|_{2,T})\big)$. On the basis of our results, the confidence ellipsoids colored in red around the estimated curve in Fig. 1 have the optimal width at any given point of the trace. Given this interpretation, we understand that no other estimator can guarantee a tighter ellipsoid around the estimated curve at any point of the trace, since it is minimax optimal.

The rest of the paper is organized as follows. In section 2 we introduce some notation, describe the statistical model, and provide a brief overview of the estimation and the basic underlying assumptions needed for the modeling and inference. In section 3 we state three main results. In section 4 we

describe the physical phenomenon behind the DT-MRI and also describe our simulation study, and results from a real data analysis. Additionally, in this section we provide a lower bound based choice of the protocol that gives the lowest global lower bound on the asymptotic risk of the estimators. Sketches of the proofs with necessary lemmas and propositions are provided in section 5; the detailed proofs can be found in [2]. Some remarks and conclusion are given in section 6.

2. Model.

2.1. Pseudo-eigen structure. Throughout the paper $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$ will represent the $\mathcal{L}_1$-, $\mathcal{L}_2$-, and $\mathcal{L}_\infty$-norms in the Euclidean space $\mathbf{R}^d$, and the notation $\|\cdot\|$ would be used generally for any $\mathcal{L}_p$-norm, $1 \leqslant p \leqslant \infty$.

A supersymmetric tensor D of rank R and (even) order M can be decomposed as $D = \sum_{r=1}^{R} v_r \otimes \cdots \otimes v_r$ for some $v_1, \ldots, v_R \in \mathbf{R}^d$, where the notation $u \otimes w$ means the outer product of vectors $u, w \in \mathbf{R}^d$, which is simply a $2D$ tensor with the components $(u \otimes w)_{ij} = u_i w_j$ for $i, j = 1, \ldots, d$. Also we will use

$$v^{\otimes M} = \underbrace{v \otimes \cdots \otimes v}_{M}, \qquad v \in \mathbf{R}^d,$$

as an abbreviated notation for tensor products. For each $r = 1, \ldots, R$, the pair $\lambda^{(r)} \in \mathbf{R}$, $v^{(r)} \in \mathbf{R}^d$, $\|v^{(r)}\|_2 = 1$, is referred to as pseudo-eigenvalue and pseudo-eigenvector, respectively, and the pair minimizes the Frobenius norm defined as

$$\begin{gathered} \sum_{i_1=1}^{d} \cdots \sum_{i_M=1}^{d} \bigl(D^{(r)}_{i_1 \ldots i_M} - \lambda v_{i_1} \cdots v_{i_M}\bigr)^2, \\ D^{(r)} = D^{(r-1)} - \lambda^{(r-1)} (v^{(r-1)})^{\otimes M}, \qquad D^{(1)} = D. \end{gathered} \tag{5}$$

See [7], [19], and [27] for additional details.

2.2. Carmichael and Sakhanenko (CS) estimator. The estimation algorithm described in [3] is as follows. In the first step, $\underline{D}(X_i)$ is estimated from (1) using least square estimator $\widetilde{D}_i$ or weighted least square estimator at each of the locations $X_1, \ldots, X_n \in G$. In the second step, a kernel smoothing estimator is used over these estimates to get an estimator $\widehat{D}_n$ at all locations $x \in G$ as

$$\widehat{D}_n(x) = \frac{1}{n h_n^d} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h_n}\right) \widetilde{D}_i,$$

where K is a kernel function and h_n is a bandwidth. In the third step, the tensor $\widehat{D}_n$ is decomposed in its pseudo-eigenpairs $\bigl(\widehat{\lambda}_n^{(r)}, \widehat{v}_n^{(r)}(x)\bigr)$, $x \in G$, $r = 1, \ldots, R$, using (5). In the final step, the integral curve estimators $\widehat{X}_n^{(r)}(t)$ for $t \in [0, T]$, $r = 1, \ldots, R$, are obtained as the solutions of the ODEs

$$\frac{d\widehat{x}_n^{(r)}(t)}{dt} = \widehat{v}_n^{(r)}\bigl(\widehat{x}_n^{(r)}(t)\bigr), \quad t \in [0, T], \qquad \widehat{x}_n^{(r)}(0) = a. \tag{6}$$

The details of this estimation algorithm can be found in [3]. In our earlier work [1] we established the pointwise minimax optimal rate of convergence of the integral curve estimators $\{\widehat{X}_n^{(r)},\, r = 1, \dots, R\}$ using locally asymptotically normal families of distributions and famous Hájek's lemma.

In this paper we will obtain the lower bound for accuracy of estimation in terms of the norms defined in (3). To find global bounds for the asymptotic risk, we will confine ourselves to the tensors that belong to a certain class $\mathcal{D}^2(a_*, G, T)$, since it plays a key role in the nonparametric estimation process described in [3]. This class $\mathcal{D}^2(a_*, G, T)$ contains tensors which satisfy the following condition (A1). Additionally each of the R pseudo-eigenvectors obtained from a tensor $D \in \mathcal{D}^2(a_*, G, T)$ are used in the equation (2), with $a = a_* \in \mathbf{R}^d$. Besides (A1) we will also require assumptions (A2)–(A8) in [3], which we describe below.

(A1) Let G be a bounded open set in $\mathbf{R}^d$ with Lebesgue measure 1. It contains the support of the twice continuously differentiable everywhere supersymmetric tensor field $D\colon \mathbf{R}^{d^M} \to \mathbf{R}^{d^2}$ of an even order $M > 2$ and a rank $1 \leqslant R \leqslant (M+2)/2$.

For a vector v and tensor D define the matrix-valued function $\mathcal{T}\colon \mathbf{R}^d \times \mathbf{R}^{d^M} \to \mathbf{R}^{d^2}$ as follows:

$$\mathcal{T}(v, D)_{km} = (M-1) \sum_{i_3=1}^{d} \cdots \sum_{i_M=1}^{d} D_{kmi_3 \dots i_M} v_{i_3} \cdots v_{i_M}, \qquad k, m = 1, \dots, d.$$

Then assume that $\operatorname{Ker}(\mathcal{T}(v^{(r)}, D^{(r)}) - \lambda^{(r)} I) = 0$ everywhere in the support of D for $r = 1, \dots, R$, where $\operatorname{Ker}(\mathcal{T})$ stands for the kernel of the linear map $\mathcal{T}$, i.e., the space of all vectors that are zero under $\mathcal{T}$.

(A2) The initial point a lies inside the support of $D(\,\cdot\,)$.

(A3) There exists a number $T > 0$ such that for all $t_1, t_2 \in (0, T)$ with $t_1 \neq t_2$, we have $x^{(r)}(t_1) \neq x^{(r)}(t_2)$ for all $r = 1, \dots, R$.

(A4) Locations $\{X_j,\, j \geqslant 1\}$ are independent and uniformly distributed random variables in G.

(A5) The observed data $\{(X_j, Y_j),\, j = 1, \dots, n\}$ obey the model (1) with a fixed nonrandom known real-valued $N \times J_M$ matrix B, an unknown continuous on G symmetric positive definite $N \times N$ tensor field $\Sigma\colon \mathbf{R}^d \to \mathbf{R}^{N^2}$ on G, and an unobservable random N-component vector Ξ_j, $j = 1, \dots, n$. Recall that $N \geqslant J_M$; this condition is commonly met by real diffusion MRI data. Additionally, it is assumed that $B^\top B$ is invertible and $\mathbf{E}\Sigma_{kl}^4(X_1) < \infty$, $1 \leqslant k, l \leqslant N$.

(A6) The random vectors Ξ_j, $j \geqslant 1$, are i.i.d. and independent of locations. Also, $\mathbf{E}\Xi_1 = 0$ and $\mathbf{E}\Xi_{1,k}\Xi_{1,l} = \delta_{kl}$ for all $1 \leqslant k, l \leqslant N$.

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

(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.

With these assumptions the following convergence of stochastic processes is established in [3]: for all $r = 1, \dots, R$, as $n \to \infty$

$$\sqrt{nh_n^{d-1}}\big(\widehat{X}_n^{(r)}(t) - x^{(r)}(t)\big) \xrightarrow{\mathrm{d}} \nu(t), \qquad t \in [0, T],$$

where $\nu(t)$ is a Gaussian process whose mean and covariance depend on D, K, $x^{(r)}$, and β. Notice that (A1) is a technical assumption that arises in the CS estimation method, to ensure that implicit derivative $\partial v^{(r)}/\partial \underline{D}$ exists for all $x \in G$ (see [3] for details).

In addition to (A5) and (A6), we assume the following on the density f of the noise variable ξ.

(A5′) Noise variables $\{\xi_i \colon i = 1, 2, \dots\}$ are i.i.d. with a common density f such that all the second order partial derivatives of the function

$$g(u) := -\int_{\mathbf{R}^N} \ln\left(1 + \frac{f(z+u) - f(z)}{f(z)}\right) f(z)\, dz, \qquad u \in \mathbf{R}^N,$$

are continuous at 0.

The class of densities that is described by condition (A5′) is fairly large. In particular, normal densities satisfy (A5′). Moreover, let f be "regular" as defined in [12], meaning that f is continuous with respect to some additive parameter $u \in \mathbf{R}^N$, has finite Fisher information with respect to the same parameter u, and the derivative of $\sqrt{f}$ with respect to u is continuous in $\mathbf{R}^N$. Then the second order partial derivatives of g can be written as

$$g''_{ij}(u) = \int \frac{f'_i(y) f'_j(y-u)}{f(y)}\, dy = \int \frac{f'_j(y) f'_i(y-u)}{f(y)}\, dy, \qquad i, j = 1, \dots, N,$$

where f'_i, g''_{ij} denotes the i-th first order partial derivative of f and the ij-th second order partial derivative of g, respectively. The assertion of (A5′) can be understood by following Lemma 7.1 in [12]. Under assumption (A5′) it can be observed that the Fisher information components

$$I_{ii} = \int \frac{(f'_i(z))^2}{f(z)}\, dz, \qquad i = 1, \dots, N,$$

of the densities f_i, $i = 1, \dots, N$, are finite, and the continuity condition for f_i, $i = 1, \dots, N$, holds:

$$\int \frac{(f'_i(y+h) - f'_i(h))^2}{f(y)}\, dy \leqslant C|h|^2, \qquad i = 1, \dots, N,$$

for $0 < |h| \leqslant \varepsilon$ and for some constant $C > 0$.

We would like to end this section by shedding light on some of the technical assumptions. Indeed, (A1) and (A3) together ensure that the tensor D achieves rank R for any $x \in G$, and all ODEs governing the integral curves

have unique solutions. Assumption (A3) in particular ensures that $x^{(r)}(t)$, $t \in [0,T]$, for any $r = 1,\dots,R$, does not create a loop in the trace. The assumption (A7) specifies a rather standard class of kernel functions for CS estimation, while (A8) highlights the dependence of bandwidth h_n on the sample size n in the CS estimating technique. This bandwidth rate matches the optimal rates for kernel regression and density estimators in a sense that if g_n estimates a density g in $\mathbf{R}^d$, then $\int_{\mathbf{R}} g_n$ estimates another density $\int_{\mathbf{R}} g$ in $\mathbf{R}^{d-1}$.

3. Main results. In this section we present the main theorems along with some motivation to establish global optimal bounds for integral curves estimators. We also present the parametric subclass of tensors and their construction. Furthermore, we describe in Lemma 1 (see section 3.3), why the constructed parametric subclass satisfies the assumption (A1), which is the centerpiece of the proof.

3.1. Theorems. Let $\mathcal{E}_n(a_*,T)$ denote the class of all possible estimators of the curve $x^{(r)}(t)$, $t \in [0,T]$, $r = 1,\dots,R$, obtained by solving the ODE (2) governed by pseudo-eigenvectors $v^{(r)}$, $r = 1,\dots,R$, based on (X_i, Y_i), $i = 1,\dots,n$, and starting at a_*.

Denote by $\mathcal{W}$ the class of functions $w\colon [0,\infty)^R \to [0,\infty)$ such that $w(0) = 0$, $w(x) > 0$ for $x \neq 0$, $x \in [0,\infty)^R$, and $w(x_1) > w(x_2)$ for $\|x_1\| > \|x_2\|$, where $x_1 \neq x_2$, $x_1, x_2 \in [0,\infty)^R$. Now using the integrated $\mathcal{L}_p$-norm defined in (3) we state the main results of this work.

Theorem 1. *Assume conditions* (A1)–(A6) *and* (A5′) *hold and* $1 \leqslant R \leqslant (M+2)/2$. *For* $c, k > 0$, *define*

$$\mathcal{D}_{c,k} = \{D \in \mathcal{D}^2(a_*,G,T)\colon \|D - D_0\|_\infty \leqslant cn^{-2/(d+3)} \text{ and } \|D'' - D_0''\|_\infty \leqslant k\}.$$

Then for any numbers $c > 0$, $T > 0$, *any point* $a_* \in G$, *any* $D \in \mathcal{D}_{c,k}$, *any function* $w \in \mathcal{W}$, *any* $1 \leqslant p \leqslant \infty$, *and some* $k > 0$, *we have*

$$\liminf_{n\to\infty} \inf_{\widehat{X}_n^{(1)},\widehat{X}_n^{(2)},\dots,\widehat{X}_n^{(R)} \in \mathcal{E}_n(a_*,T)} \sup_{D\in\mathcal{D}_{c,k}} \mathbf{E} w\big(n^{2/(d+3)}\big(\|\widehat{X}_n^{(1)} - x^{(1)}\|_{p,T}, \|\widehat{X}_n^{(2)} - x^{(2)}\|_{p,T}, \dots, \|\widehat{X}_n^{(R)} - x^{(R)}\|_{p,T}\big)\big) > 0.$$

Theorem 1 establishes that no ensemble of integral curve estimators $\widehat{X}_n^{(r)}(t)$, $r = 1,\dots,R$, $t \in [0,T]$, would converge to the true ensemble of integral curves faster than $n^{-2/(d+3)}$ for tensors inside the special class $\mathcal{D}_{c,k}$ under the integrated norm with the appropriate loss function w. Since $\mathcal{D}_{c,k} \subset \mathcal{D}^2(a_*,G,T)$, as a result of Theorem 1, we state the following corollary.

Corollary 1. *Assume conditions* (A1)–(A6) *and* (A5′) *hold and* $1 \leqslant R \leqslant (M+2)/2$. *Then for any number* $T > 0$, *any point* $a_* \in G$, *any function*

$w \in \mathcal{W}$, we have

$$\liminf_{n\to\infty} \inf_{\widehat{X}_n^{(1)},\widehat{X}_n^{(2)},\dots,\widehat{X}_n^{(R)} \in \mathcal{E}_n(a_*,T)} \sup_{D\in\mathcal{D}^2(a_*,G,T)} \mathbf{E} w\big(n^{2/(d+3)}\big(\|\widehat{X}_n^{(1)} - x^{(1)}\|_{p,T}, \|\widehat{X}_n^{(2)} - x^{(2)}\|_{p,T}, \dots, \|\widehat{X}_n^{(R)} - x^{(R)}\|_{p,T}\big)\big) > 0.$$

With Corollary 1 we establish the minimax lower bound for the integral curve estimators $\widehat{X}_n^{(r)}(t)$, $r = 1,\dots,R$, $t \in [0,T]$, based on the class of tensors $\mathcal{D}^2(a_*,G,T)$, globally. Further, we show that these optimal rates are attained by the CS estimators.

Theorem 2. *Assume conditions* (A1)–(A8) *and* (A5′) *hold. Let* $\{\widehat{X}_n^{(r)}, r = 1,\dots,R\}$ *from equation* (6) *be the CS estimators.*

1. *If we additionally assume* $\mathbf{E}\|\xi_i\|_q^q < \infty$ *for some* $q \geqslant 4$, *then for each* $r = 1,\dots,R$, *any number* $T > 0$, *any point* $a_* \in G$, *and any* $2 \leqslant p \leqslant q$, *we have*

$$\sup_n \sup_{D\in\mathcal{D}^2(a_*,G,T)} \mathbf{E}\big\|n^{2/(d+3)}\big(\widehat{X}_n^{(r)} - x^{(r)}\big)\big\|_{p,T} < \infty.$$

2. *For each* $r = 1,\dots,R$, *any number* $T > 0$, *and any point* $a_* \in G$, *we have*

$$\sup_n \sup_{D\in\mathcal{D}^2(a_*,G,T)} \mathbf{E}\big\|n^{2/(d+3)}\big(\widehat{X}_n^{(r)} - x^{(r)}\big)\big\|_{\infty,T} < \infty.$$

We remark that we only require the additional moment assumption on the noise variables ξ_i for Theorem 2. Meanwhile, Theorem 1 and Corollary 1 would still hold under this assumption. These two theorems and the corollary together provide the global bounds for the asymptotic risk of the CS integral curve estimator and show that it is minimax optimal.

3.2. Parametric subclass of tensors. In order to establish lower bounds for the asymptotic risk of integral curve estimators for a large nonparametric class of tensors, we start by constructing of a parametric subclass in $\mathcal{D}^2(a_*,G,T)$. We begin with perturbing the curves, which will perturb the resulting vector-field, ultimately translating the perturbation onto the tensor field in $\mathcal{D}^2(a_*,G,T)$.

For each $r = 1,\dots,R$, let $x_0^{(r)}(t,a_*)$, $t \in [0,T]$, denote the integral curve starting at a_*, driven by the vectors $v_0^{(r)}(x_0^{(r)}(t,a_*))$, where $x_0^{(r)}(0,a_*) = a_*$. Here we assume that for any $x \in G$, we have $0 < \|v_0^{(r)}(x)\| < \infty$. Additionally, for a small $\epsilon > 0$, consider A_ϵ, an ϵ-neighborhood of a_* (in Euclidean norm) in a $(d-1)$-dimensional hyperspace transversal to the flow at a_*. Suppose the volume swept by A_ϵ, under the flow $v_0^{(r)}$, is denoted by

$$G_{\epsilon,T}^{(r)} = \big\{x^{(r)} = x_0^{(r)}(t,a) \colon t \in [0,T],\, a \in A_\epsilon\big\} \subset G, \qquad r = 1,\dots,R.$$

Then $G_{\epsilon,T}^{(r)}$ defines a neighborhood of the integral curve $x_0^{(r)}(t,a_*)$, $t \in [0,T]$, as we vary the initial point a. Suppose $\alpha > 0$ and $\gamma > 0$ (we will show

later that the best choice is $\alpha = 2/(d+3)$ and $\gamma = \alpha/2$); similarly to the perturbation defined in [22] we can define

$$x_b^{(r)}(t,a) = x_0^{(r)}\big(t + n^{-\alpha}\varphi_b^{(r)}(t)\psi^{(r)}(n^\gamma|a-a_*|), a\big),$$

where $\varphi_b^{(r)}(t)$, $t \in [0,T]$, is a family of twice continuously differentiable functions indexed by $b \in \{0,1\}^P$. Suppose that $\varphi_b^{(r)}(t) \not\equiv 0$, $\varphi_b^{(r)}(0) = 0$, $\varphi_b^{(r)}(T) = 0$, $-1 < \varphi_b'^{(r)}(t) \leqslant 1$ for $r = 1,\dots,R$. Additionally assume $\psi^{(r)}(z)$, $z > 0$, is a three times continuously differentiable function such that

$$0 < \psi^{(r)}(z) < \frac{c}{\|v_0^{(r)}\|_\infty}, \qquad \psi'^{(r)}(0) = \psi''^{(r)}(0) = 0,$$
$$\psi'^{(r)}(z) \leqslant 0 \quad \text{for} \quad z > 0, \qquad \psi^{(r)}(z) = 0 \quad \text{for} \quad z > \epsilon.$$

Note that the perturbation in the parameter t is small enough and for all $r = 1,\dots,R$, the perturbation vanishes far enough from $x_0^{(r)}(t,a_*)$, $t \in [0,T]$. Then the corresponding perturbation in the vector field can be found as

$$\begin{aligned}\frac{d}{dt}\,x_b^{(r)}(t,a) &= v_b^{(r)}\big(x_b^{(r)}(t,a)\big) = \frac{d}{dt}\,x_0^{(r)}\big(t + n^{-\alpha}\varphi_b^{(r)}(t)\psi^{(r)}(n^\gamma|a-a_*|), a\big)\\ &= v_0^{(r)}\big(x_b^{(r)}(t,a)\big)\big(1 + n^{-\alpha}\varphi_b'^{(r)}(t)\psi^{(r)}(n^\gamma|a-a_*|)\big). \qquad (7)\end{aligned}$$

By the flow box theorem (see [6, Lemma 1.120]) for $x^{(r)} \in G_{\epsilon,T}^{(r)}$, there are uniquely defined twice continuously differentiable functions $t_b^{(r)}(x) \in [0,T]$ and $a_b(x) \in A_\epsilon$ such that there is a b-perturbed integral curve starting in A_ϵ, which goes through $x^{(r)}$, and

$$x_b^{(r)}\big(t_b^{(r)}(x^{(r)}), a_b(x^{(r)})\big) = x \quad \text{for} \quad r = 1,\dots,R,$$

since $v_0^{(r)}(x) \neq 0$ for all $x \in G_{\epsilon,T}^{(r)}$. So the expression on the right-hand side of (7) can be written as

$$v_{b,n}^{(r)}(x) = v_0^{(r)}(x)\big(1 + n^{-\alpha}\varphi_b'^{(r)}(t_b^{(r)}(x))\psi^{(r)}(n^\gamma|a_b(x)-a_*|)\big). \qquad (8)$$

Now for a fixed $b \in \{0,1\}^P$, and using $v_{b,n}^{(r)}$, $r = 1,\dots,R$, we construct the perturbed tensor by

$$D_b(x) = \lambda_1 v_{b,n}^{(1)}(x)^{\otimes M} + \lambda_2 v_{b,n}^{(2)}(x)^{\otimes M} + \cdots + \lambda_R v_{b,n}^{(R)}(x)^{\otimes M}. \qquad (9)$$

3.3. A special lemma and ideas of the proofs. We need to verify that the proposed parametric class of tensors satisfies (A1). Consider the tensor corresponding to vectors $v_0^{(r)}(x)$, $r = 1,\dots,R$, named D_0:

$$D_0(x) = \lambda_1 v_0^{(1)}(x)^{\otimes M} + \lambda_2 v_0^{(2)}(x)^{\otimes M} + \cdots + \lambda_R v_0^{(R)}(x)^{\otimes M}, \qquad (10)$$

where pseudo-eigenvalues and eigenvectors λ_r, $v_0^{(r)}$, $r = 1, \dots, R$, respectively, are chosen in such a way that D_0 belongs to the class $\mathcal{D}_{c,k}$ in Theorem 1, meaning in particular, that it satisfies (A1). Let us introduce some additional notation: for $p = 1, \dots, R$,

$$D_0^{(p)}(x) = \sum_{r=p}^{R} \lambda_r v_0^{(r)}(x)^{\otimes M}, \quad D_b^{(p)}(x) = \sum_{r=p}^{R} \lambda_r v_{b,n}^{(r)}(x)^{\otimes M}. \tag{11}$$

Also we assume that $\|v_0^{(r)}(x)\| = 1$ and denote $\|v_{b,n}^{(r)}(x)\|^2 =: c_{b,n}^{(r)}$. Hence, it is easy to note that $c_{b,n}^{(r)} \to 1$ as $n \to \infty$.

Lemma 1. *For $r = 1, \dots, R$, for all $x \in G_{\epsilon,T}^{(r)}$, the tensors $D_b^{(r)}$ satisfy* $\operatorname{Ker}\big(\mathcal{T}\big(v_{b,n}^{(r)}(x), D_b^{(r)}(x)\big) - \lambda_r I\big) = 0$.

The proof of this lemma is quite technical and is provided in [2].

To prove the lower bounds we first establish a multivariate Fano's lemma that lower-bounds the supremum risk over $\mathcal{D}_{c,k}$ by a supremum over a finite parametric subclass of tensors that are close in the Kullback–Leibler divergence, which is subsequently bounded by infimum of Shannon's information over $\delta_2^{(r)}$-separated net of integral curves. Secondly, we prove that the finite collections of densities of (X, Y) corresponding to tensors in $\mathcal{D}_{c,k}$ consist of members that are close in the Kullback–Leibler divergence while the corresponding integral curves are $2\delta_2^{(r)}$-separated in $\|\cdot\|_{p,T}$. Thirdly, we check that the elements of the finite collection belong to $\mathcal{D}_{c,k}$. Finally, we check that the corresponding integral curves are $Cn^{-\alpha-\delta}$-separated in $\|\cdot\|_{p,T}$, where α, δ are chosen to balance the resulting bound.

To prove the upper bounds for CS estimators we carefully bound moments of the deviation process between CS estimators and the true integral curves. The details are deferred to [2].

4. Simulation study and data analysis.

4.1. Simulation study. In this section using the setup of [24] and considering $d = 3$ again, we simulate the "Y" pattern using several sample sizes from 30^3 to 100^3. For each curve, we computed the estimated constant in the lower bound in Theorem 1, which is $\kappa = w\big(n^{1/3}(\|\widehat{X}^{(1)} - x^{(1)}\|_{2,T}, \|\widehat{X}^{(2)} - x^{(2)}\|_{2,T})\big)$. Since our method is minimax globally with respect to the asymptotic risk of the estimators, we can compare these κ values for different scenarios to choose the best one. Below we provide the 25th percentile, the median, and the 75th percentile for the empirical κ values that are simulated over 100 times for each of the eight different sample sizes.

As we can see from Fig. 2, the median of the κ values tends to stabilize when we use $n_0 = 60$. Therefore, next we will investigate the robustness of the values for κ when we vary SNR and thickness of the fibers with the sample size $n = 60^3$. In Table 1 we provide the results.

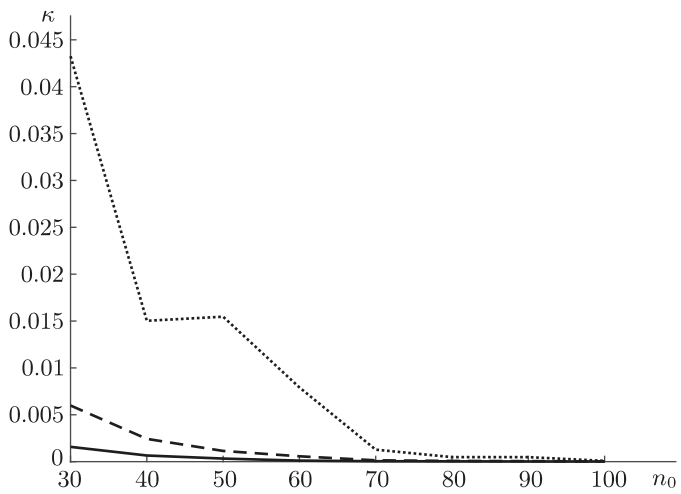

Fig. 2. The solid, dashed, and dotted lines show the 25th, 50th, and 75th percentiles of the κ values across all the sample sizes $n = n_0^3$, $n_0 = 30, 40, 50, 60, 70, 80, 90, 100$ repeated 100 times. Here we have used step size $\delta = 0.02$ and $\beta = 10^{-7}$ for each of the iterations.

Table 1. The 25th percentile, median, and 75th percentile of the empirical κ values are ordered in top to bottom in each line for different combinations of SNR and thickness of fibers; the sample size is $n = 60^3$

SNR	Thickness = 0.02	0.04	0.06	0.08	0.1
2	2.14×10^{-4}	9.82×10^{-5}	7.27×10^{-5}	7.6×10^{-5}	8.19×10^{-5}
	3.25×10^{-3}	2.88×10^{-4}	2.05×10^{-4}	1.95×10^{-4}	2.6×10^{-4}
	2.63×10^{-2}	2.39×10^{-3}	1.22×10^{-3}	8.3×10^{-4}	1.13×10^{-3}
4	2.04×10^{-7}	1.14×10^{-7}	1.05×10^{-7}	1.07×10^{-7}	9.4×10^{-8}
	1.17×10^{-5}	2.39×10^{-7}	2.34×10^{-7}	2.55×10^{-7}	2.35×10^{-7}
	7.38×10^{-4}	1.49×10^{-6}	5.74×10^{-7}	8.09×10^{-7}	6.1×10^{-7}
6	1.78×10^{-8}	3.49×10^{-9}	2.42×10^{-9}	3.05×10^{-9}	2.99×10^{-9}
	4.16×10^{-7}	8.82×10^{-9}	4.86×10^{-9}	6.3×10^{-9}	5.55×10^{-9}
	5.3×10^{-5}	2.93×10^{-8}	1.85×10^{-8}	2.18×10^{-8}	1.35×10^{-8}
8	5.25×10^{-10}	1.86×10^{-10}	2.17×10^{-10}	2.11×10^{-10}	2.03×10^{-10}
	4.31×10^{-8}	4.02×10^{-10}	3.7×10^{-10}	4.02×10^{-10}	3.93×10^{-10}
	1.87×10^{-6}	1.13×10^{-9}	1.53×10^{-9}	8.12×10^{-10}	7.28×10^{-10}
10	1.05×10^{-10}	2.58×10^{-11}	3×10^{-11}	2.67×10^{-11}	2.78×10^{-11}
	6.1×10^{-9}	5.21×10^{-11}	5.97×10^{-11}	5.28×10^{-11}	5.05×10^{-11}
	9.07×10^{-8}	1.07×10^{-10}	1.34×10^{-10}	1.02×10^{-10}	1.53×10^{-10}

From Table 1 it is evident that if we increase the SNR, then the κ value decreases, indicating that as the signal gets stronger, the traced curve estimates the true curve more accurately, and hence the asymptotic risk decreases. However, the thickness of the curves does not affect the value of κ

in a significant way. Overall, the smallest thickness seems to have the worst performance, while the best value of κ is obtained more often when the thickness is close to 0.06 indicating that if the thickness of the curves is extremely high or low, then the uncertainty in estimation is higher.

4.2. Neuroimaging example. Several diffusion weighted imaging datasets were collected from a 28-year-old healthy male brain on a GE 3T Signa HDx MR scanner (GE Healthcare, Waukesha, WI) with an 8-channel head coil. The subject signed the consent form approved by the Michigan State University Institutional Review Board. Diffusion weighted images were acquired with a spin-echo echo-planar imaging (EPI) sequence for several minutes per session using the following protocols with the following parameters summarized in Table 2. All protocols had 48 contiguous 2.4-mm axial slices in an interleaved order and matrix size is 128×128, echo time (TE) is 72.3 ms, repetition time (TR) is 7.5 s, $b = 1000\,\mathrm{s/mm^2}$, field-of-view (FOV) is 22 cm × 22 cm, parallel imaging acceleration factor is 2, and the number of excitations (NEX) varied between 1 and 2.

Table 2. Protocols for HARDI. Note that the second protocol has a total of six b_0 images after two repetitions of three b_0 images

Protocol	scan time	slice size	NEX	# of slices	TR	# of b_0 images
30 directions	6.5 mins	2.4 mm	1	48	11.5 s	3
30 directions, 2 reps	12.9 mins	2.4 mm	2	48	11.5 s	3
60 directions	12.9 mins	2.4 mm	1	48	11.5 s	6
90 directions	19.2 mins	2.4 mm	1	48	11.5 s	9
150 directions	20 mins	2.4 mm	1	32	7.5 s	9

We traced axonal fibers in the anterior part of the corpus callosum which connect the right and left frontal lobes. The general anatomical locations of these axonal fibers are well established. These fibers can be used to evaluate new techniques in fiber tractography. Several initial points were chosen in the region of interest (ROI) based on anatomical considerations. Under each protocol, starting with each seed point, we used the estimation technique in [3] to trace a fiber until it ran into water. Fig. 3 provides the reference images done for protocol with 60 directions.

For each curve we computed the estimated constant in the lower bound in Theorem 1, which is $\kappa = w\big(n^{1/3}\|\widehat{X}_n - x\|_{2,T}\big)$. We used the loss function $w(u) = u^2/T$. Since the estimation method has the rate optimality property (with respect to n), the only thing left to be optimized is the constant. It can be used to compare different rate-optimal estimators. In this study the estimators differ according to the underlying protocols used to obtain the data. The smaller constant κ would indicate a more successful estimator. Table 3 summarizes our findings.

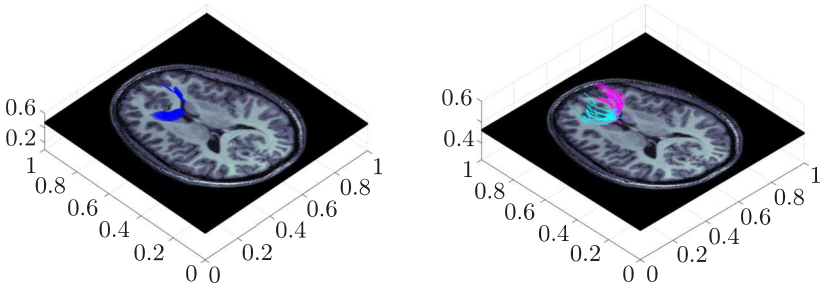

Fig. 3. A neuronal fiber bundle across the genu of corpus callosum is created based on the CS method in [3]. In (a) one particular seed point was used. The confidence ellipsoids along the fiber have 95% confidence. In (b) several seed points were used to create several estimated fibers. Two different branches are shown in two colors.

Table 3. Comparison of κ values for tracing of anterior fibers based on imaging datasets obtained via different scanning protocols. Here $\delta = 0.003$, $\beta = 10^{-7}$

Protocol	# of voxels	ROI size	25th	Median	75th
30 directions	$128 \times 128 \times 48$	120	0.1156	0.3630	1.6954
30 directions, 2 reps	$128 \times 128 \times 48$	122	0.0915	0.2851	0.8066
60 directions	$128 \times 128 \times 48$	119	0.0417	0.1211	0.3596
90 directions	$128 \times 128 \times 48$	119	0.0481	0.1955	1.1941
150 directions	$128 \times 128 \times 32$	171	0.1869	0.5762	1.4582

The protocol with 60 directions has the lowest κ, while protocol with 90 directions comes in with the second smallest κ value. For each protocol the empirical distribution of κ values with respect to the initial point is skewed right, which is expected since the uncertainty increases along the fibers. It is quite interesting that the design with 60 different directions gives lower κ in comparison with the block design when 30 directions are independently repeated twice. Usually, block designs yield smaller variances for the estimated fibers locally, so one might argue that block designs have advantage locally (see [23]). However, here the block design performs worse in a global sense. We speculate that this is due to noise behavior along the whole fiber. Finally, we remark that the κ value depends on D, a_*, and T. Thus, this value would be different for different fibers (different a_*) for the same subject and different for different subjects (different D). A large comprehensive comparison study is underway to assess this comparison method.

5. Sketch of the proofs. The proof of Theorem 1 is based on the following lemmas and it follows the spirit of the construction given in [12] and [11].

We additionally extend Fano's lemma in the multidimensional parameter space, which we apply in Lemma 3 tailored to fit our problem. It is similar to Theorem 5.7 in [8]. Lemma 2 provides a bound for the function g defined in section 2 (see assumption (A5)). In Lemma 4 we prove the smoothness condition on the perturbed class of tensor fields $\mathcal{D}_{c,k}$. Finally, Lemmas 5 and 6 provide a construction for the tensor fields and integral curves which we will make use in the proof of Theorem 1.

Lemma 2. *Let f satisfy condition* (A5′). *Then there exists a $\delta_1 > 0$ such that for any vector u satisfying $|u| \leqslant \delta_1$ we have*

$$g(u) \leqslant C(f, \delta_1)|u|^2$$

with a positive constant $C(f, \delta_1)$.

Proposition 1. *Let X be a random variable whose density f_θ depends on a parameter $\theta = (\theta^{(1)}, \dots, \theta^{(R)})$ from a multidimensional discrete parameter space $\{1 \leqslant \theta^{(r)} \leqslant l_r + 1,\ r = 1, \dots, R,\ l_r \in \mathbf{N}\}$. Suppose for $\eta > 0$, the Kullback–Leibler divergence between f_θ and $f_{\theta'}$ is $K(f_\theta, f_{\theta'}) \leqslant \eta$. Then for an estimator $\Psi(X) = \big(\Psi^{(1)}(X), \dots, \Psi^{(R)}(X)\big)$ of the unknown parameter θ we have*

$$\sup_{\underline{i}} \mathbf{P}_{\underline{i}}\big(\Psi(X) \neq \underline{i}\big) \geqslant 1 - \frac{\eta + \ln 2}{\ln(\mathscr{L} - 1)},$$

where $\underline{i} = (i_1, \dots, i_R)$, $1 \leqslant i_r \leqslant l_r + 1$, $r = 1, \dots, R$, $\mathscr{L} = \prod_{r=1}^{R}(1 + l_r)$ and $\mathbf{P}_{\underline{i}}$ is the probability induced by $f_{\underline{i}}$.

Proposition 2. *Let $f_{\underline{i}}$ be the density of $(X, B\underline{D}_{\underline{i}}(X) + \Sigma^{1/2}(X)\xi)$, $\underline{i} = (i_1, \dots, i_R)$. Then for any two indices $\underline{i} \neq \underline{j}$, $1 \leqslant i_r, j_r \leqslant l_r + 1$, $r = 1, \dots, R$, we have*

$$\|\underline{D}_{\underline{i}} - \underline{D}_{\underline{j}}\|_G^2 \leqslant \eta \quad \Longrightarrow \quad K(f_{\underline{i}}, f_{\underline{j}}) \leqslant C(f, \delta_1) C_{\Sigma,B} \eta,$$

where $C(f, \delta_1)$ is a constant introduced in Lemma 2 and $C_{\Sigma,B}$ is a constant depending on Σ and B only.

Lemma 3. *Let $\mathcal{D}_{\mathscr{L}}$ be a class of $\mathscr{L}$ tensor fields D_θ, $\theta = (\theta^{(1)}, \dots, \theta^{(R)})$ inside the parametric subclass of $\mathcal{D}_{c,k}$, such that for any $\theta \neq \widetilde{\theta}$ and positive constants η and $\delta_2^{(r)}$, $r = 1, \dots, R$,*

$$\|\underline{D}_\theta - \underline{D}_{\widetilde{\theta}}\|_G^2 \leqslant \delta_1, \quad \|\underline{D}_\theta - \underline{D}_{\widetilde{\theta}}\|_G^2 \leqslant \eta, \quad \text{and} \quad \big\|x_\theta^{(r)} - x_{\widetilde{\theta}}^{(r)}\big\|_{1,T} \geqslant 2\delta_2^{(r)},$$

where for each $r = 1, \dots, R$ and θ, we denote by $x_\theta^{(r)} \in \mathbf{R}^d$ the integral curve corresponding to $v_\theta^{(r)}$ starting at a_*. In addition, let f satisfy condition (A5′). *Then for any $w \in \mathcal{W}$, we have*

$$\sup_{D \in \mathcal{D}_{\mathscr{L}}} \mathbf{E} w\big(\big\|\widehat{X}_n^{(1)} - x^{(1)}\big\|_{1,T}, \dots, \big\|\widehat{X}_n^{(R)} - x^{(R)}\big\|_{1,T}\big)$$
$$\geqslant \inf_{|x^{(r)}| \geqslant \delta_2^{(r)}:\, r=1,\dots,R} w\big(|x^{(1)}|, \dots, |x^{(R)}|\big)\left(1 - \frac{nC_{\Sigma,B}C(f, \delta_1)\eta + \ln 2}{\ln(\mathscr{L} - 1)}\right).$$

Let $\mathcal{B}$ be a $P/4$-separated net in $\mathcal{L}_1$-norm in $\{0,1\}^P$.

Lemma 4. *For all $b \in \mathcal{B}$, we have $D_b \in \mathcal{D}_{c,k}$.*

Lemma 5. *For any $b, \widetilde{b} \in \mathcal{B}$ such that $b \neq \widetilde{b}$ and for any large enough n, we have*

$$\|\underline{D}_b(x) - \underline{D}_{\widetilde{b}}(x)\|_G^2 \leqslant Cn^{-(d-1)\gamma+2\alpha},$$

where $C > 0$ is a constant.

Lemma 6. *For each $r = 1, \dots, R$, for any $b, \widetilde{b} \in \mathcal{B}$ such that $b \neq \widetilde{b}$, and for any large enough n, we have*

$$\big\|x_b^{(r)} - x_{\widetilde{b}}^{(r)}\big\|_{1,T} \geqslant Cn^{-\alpha-\delta}$$

with some $C > 0$ depending on r.

Using all these lemmas, the expression in Theorem 1 is bounded from below by

$$\begin{aligned}
&\inf_{\widehat{X}_n^{(1)}, \widehat{X}_n^{(2)}, \dots, \widehat{X}_n^{(R)} \in \mathcal{E}_n(a_*, T)} \sup_{D \in \mathcal{D}_{c,k}} \mathbf{E} w\big(n^{2/(d+3)} \big(\big\|\widehat{X}_n^{(1)} - x^{(1)}\big\|_{p,T}, \\
&\qquad\qquad \big\|\widehat{X}_n^{(2)} - x^{(2)}\big\|_{p,T}, \dots, \big\|\widehat{X}_n^{(R)} - x^{(R)}\big\|_{p,T}\big)\big) \\
&\geqslant \inf_{\widehat{X}_n^{(1)}, \widehat{X}_n^{(2)}, \dots, \widehat{X}_n^{(R)} \in \mathcal{E}_n(a_*, T)} \sup_{D \in \mathcal{D}_{\mathscr{L}}} \mathbf{E} w\big(n^{2/(d+3)} \big(\big\|\widehat{X}_n^{(1)} - x^{(1)}\big\|_{p,T}, \\
&\qquad\qquad \big\|\widehat{X}_n^{(2)} - x^{(2)}\big\|_{p,T}, \dots, \big\|\widehat{X}_n^{(R)} - x^{(R)}\big\|_{p,T}\big)\big) \\
&\geqslant \inf_{|x^{(r)}| \geqslant \delta_2^{(r)} \colon r=1,\dots,R} w_n\big(|x^{(1)}|, \dots, |x^{(R)}|\big) \left(1 - \frac{nC_{\Sigma,B}C(f,\delta_1)\eta + \ln 2}{\ln(\mathscr{L} - 1)}\right).
\end{aligned} \tag{12}$$

We substitute $C_{\Sigma,B}C(f,\delta_1) = C > 0$ as a generic constant and choose $\eta = Cn^{1-2\alpha-(d-1)\gamma}$, $\delta_2^{(r)} = Cn^{-\alpha-\delta}/2$. Also it can be shown that $\mathscr{L} \geqslant \exp(P/2)$, see [13], [14], hence by an algebraic manipulation we get $\ln(\mathscr{L}-1) \geqslant P/2-1$. Therefore, we can rewrite (12) as

$$\begin{aligned}
&\inf_{\widehat{X}_n^{(1)}, \widehat{X}_n^{(2)}, \dots, \widehat{X}_n^{(R)} \in \mathcal{E}_n(a_*, T)} \sup_{D \in \mathcal{D}_{c,k}} \mathbf{E} w\big(n^{2/(d+3)} \big(\big\|\widehat{X}_n^{(1)} - x^{(1)}\big\|_{p,T}, \\
&\qquad\qquad \big\|\widehat{X}_n^{(2)} - x^{(2)}\big\|_{p,T}, \dots, \big\|\widehat{X}_n^{(R)} - x^{(R)}\big\|_{p,T}\big)\big) \\
&= \inf_{|x^{(r)}| \geqslant Cn^{-\alpha-\delta}/2 \colon r=1,\dots,R} w_n\big(|x^{(1)}|, |x^{(2)}|, \dots, |x^{(R)}|\big) \\
&\qquad \times \left(1 - \frac{n^{1-2\alpha-(d-1)\gamma}C^2 + \ln 2}{P/2 - 1}\right) \\
&> 0.5 \inf_{|x^{(r)}| \geqslant 0.5Ch \colon r=1,\dots,R} w\big(|x^{(1)}|, |x^{(2)}|, \dots, |x^{(R)}|\big) > 0,
\end{aligned} \tag{13}$$

where $w_n\big(|x^{(1)}|, \dots, |x^{(R)}|\big) = w\big(n^{2/(d+3)}\big(|x^{(1)}|, \dots, |x^{(R)}|\big)\big)$. Note that while obtaining (13) we have chosen $\alpha = 2/(d+3)$, $\gamma = \alpha/2$, and $\delta = 0$. Also

we can choose P_1 such that $C^2 < (P_1 - 2)/4 - \ln 2$, where $P = P_1 n^{\delta}$ was introduced before, and this completes the proof of Theorem 1.

Proof of Theorem 2 relies on the decomposition of the deviation process

$$\widehat{X}_n^{(r)}(s) - x^{(r)}(s) = Z_n^{(r)}(s) + \delta_n^{(r)}(s), \qquad r = 1, \ldots, R,$$

where the process $Z_n^{(r)}$ satisfies a nice PDE and has an explicit representation via help of Green's function, while the remainder process $\delta_n^{(r)}(s)$ is shown to be negligible. Subsequently, we carefully bound the moments of both processes and obtain the desired results.

For the detailed proofs see [2].

6. Remarks and conclusion. In summary, we would like to comment that in this work we have proved the minimax optimality of the asymptotic risk of the nonparametric integral curve estimators described in [3] in the whole domain of the imaging field G. Therefore, we have established the global minimax optimality of the estimation method. Although the asymptotic rates that we proposed are minimax optimal in the global sense, one can further optimize the constant κ involved in the risk function to get sharp results. In our data analysis we have provided a comparative study of the different imaging protocols with respect to this constant, and we have found the protocol that provides the lowest value to the global asymptotic risk. The analysis was performed on a single subject (human brain). We have similar results for another subject. This can be further studied with more different subjects to understand if the optimal protocol for scanning procedure remains the same across subjects. However, it is beyond the scope of this work and could be explored as a nice applied direction for future research on this topic. On the theoretical side, one can explore the sharp lower bounds and obtain theoretical results on the constant.

REFERENCES

1. C. Banerjee, L. Sakhanenko, D. C. Zhu, "Lower bounds for accuracy of estimation in magnetic resonance high angular resolution diffusion imaging data", *J. Indian Soc. Probab. Statist.*, **21**:1 (2020), 1–41.
2. C. Banerjee, L. Sakhanenko, D. C. Zhu, "Global rate optimality of integral curve estimators in high order tensor models: Supplemental material", *Theory Probab. Appl.* (to appear).
3. O. Carmichael, L. Sakhanenko, "Estimation of integral curves from high angular resolution diffusion imaging (HARDI) data", *Linear Algebra Appl.*, **473** (2015), 377–403.
4. O. Carmichael, L. Sakhanenko, "Integral curves from noisy diffusion MRI data with closed-form uncertainty estimates", *Stat. Inference Stoch. Process.*, **19**:3 (2016), 289–319.
5. Soo-Eun Chang, D. C. Zhu, "Neural network connectivity differences in children who stutter", *Brain*, **136**:12 (2013), 3709–3726.

6. C. Chicone, *Ordinary differential equations with applications*, 2nd ed., Texts Appl. Math., **34**, Springer, New York, 2006, xx+636 pp.
7. L. de Lathauwer, B. de Moor, J. Vandewalle, "On the best rank-1 and rank-$(R_1, R_2, \dots, R_N)$ approximation of higher-order tensors", *SIAM J. Matrix Anal. Appl.*, **21**:4 (2000), 1324–1342.
8. L. Devroye, *A course in density estimation*, Progr. Probab. Statist., **14**, Birkhäuser Boston, Inc., Boston, MA, 1987, xx+183 pp.
9. T. S. Ferguson, *Mathematical statistics. A decision theoretic approach*, Probab. Math. Statist., **1**, Academic Press, New York–London, 1967, xi+396 pp.
10. A. Guntuboyina, B. Sen, "Global risk bounds and adaptation in univariate convex regression", *Probab. Theory Related Fields*, **163**:1-2 (2015), 379–411.
11. R. Hasminskii, I. Ibragimov, "On density estimation in the view of Kolmogorov's ideas in approximation theory", *Ann. Statist.*, **18**:3 (1990), 999–1010.
12. И. А. Ибрагимов, Р. З. Хасьминский, *Асимптотическая теория оценивания*, Наука, М., 1979, 527 с.; англ. пер.: I. A. Ibragimov, R. Z. Has'minskii, *Statistical estimation. Asymptotic theory*, Appl. Math., **16**, Springer-Verlag, New York–Berlin, 1981, vii+403 pp.
13. И. А. Ибрагимов, Р. З. Хасьминский, "Об оценке плотности распределения", *Исследования по математической статистике*. IV, Зап. науч. сем. ЛОМИ, **98**, Изд-во "Наука", Ленинград. отд., Л., 1980, 61–85; англ. пер.: I. A. Ibragimov, R. Z. Khas'minskii, "Estimation of distribution density", *J. Soviet Math.*, **21**:1 (1983), 40–57.
14. И. А. Ибрагимов, Р. З. Хасьминский, "Еще об оценке плотности распределения", *Исследования по математической статистике*. V, Зап. науч. сем. ЛОМИ, **108**, Изд-во "Наука", Ленинград. отд., Л., 1981, 72–88; англ. пер.: I. A. Ibragimov, R. Z. Khas'minskii, "More on the estimation of distribution densities", *J. Soviet Math.*, **25**:3 (1984), 1155–1165.
15. A. K. H. Kim, R. J. Samworth, "Global rates of convergence in log-concave density estimation", *Ann. Statist.*, **44**:6 (2016), 2756–2779.
16. V. Koltchinskii, L. Sakhanenko, Songhe Cai, "Integral curves of noisy vector fields and statistical problems in diffusion tensor imaging: nonparametric kernel estimation and hypotheses testing", *Ann. Statist.*, **35**:4 (2007), 1576–1607.
17. D. Le Bihan, J. F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, H. Chabriat, "Diffusion tensor imaging: concepts and applications", *J. Magn. Reson. Imaging*, **13**:4 (2001), 534–546.
18. E. Mammen, S. van de Geer, "Locally adaptive regression splines", *Ann. Statist.*, **25**:1 (1997), 387–413.
19. Guyan Ni, Yiju Wang, "On the best rank-1 approximation to higher-order symmetric tensors", *Math. Comput. Modelling*, **46**:9-10 (2007), 1345–1352.
20. E. Özarslan, T. Mareci, "Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging", *Magn. Reson. Med.*, **50**:5 (2003), 955–965.
21. G. Raskutti, M. J. Wainwright, Bin Yu, "Minimax-optimal rates for sparse additive models over kernel classes via convex programming", *J. Mach. Learn. Res.*, **13** (2012), 389–427.
22. Л. А. Саханенко, "Глобальная оптимальная скорость сходимости в модели для исследования тензоров диффузии", *Теория вероятн. и ее примен.*, **55**:1 (2010), 19–35; англ. пер.: L. Sakhanenko, "A global optimal convergence rate in a model for the diffusion tensor imaging", *Theory Probab. Appl.*, **55**:1 (2011), 77–90.

23. L. Sakhanenko, "How to choose the number of gradient directions for estimation problems from noisy diffusion tensor data", *Contemporary developments in statistical theory*, A festschrift for Hira Lal Koul, Springer Proc. Math. Stat., **68**, Springer, Cham, 2014, 305–310.
24. L. Sakhanenko, M. DeLaura, "A comparison study of statistical tractography methodologies for diffusion tensor imaging", *Int. J. Stat. Adv. Theory Appl.*, **1**:1 (2017), 93–110.
25. C. J. Stone, "Optimal global rates of convergence for nonparametric regression", *Ann. Statist.*, **10**:4 (1982), 1040–1053.
26. M. J. Wainwright, *High-dimensional statistics. A non-asymptotic viewpoint*, Camb. Ser. Stat. Probab. Math., **48**, Cambridge Univ. Press, Cambridge, 2019, xvii+552 pp.
27. Leslie Ying, Yi Ming Zou, D. P. Klemer, Jiun-Jie Wang, "Determination of fiber orientation in MRI diffusion tensor imaging based on higher-order tensor decomposition", *Proceedings of the 29th annual international conference of the IEEE engineering in medicine and biology society* (Lyon, 2007), IEEE, 2007, 2065–2068.
28. D. C. Zhu, S. Majumdar, "Integration of resting-state FMRI and diffusion-weighted MRI connectivity analyses of the human brain: limitations and improvement", *J. Neuroimaging*, **24**:2 (2014), 176–186.
29. D. C. Zhu, S. Majumdar, I. O. Korolev, K. L. Berger, A. C. Bozoki, "Alzheimer's disease and amnestic mild cognitive impairment weaken connections within the default mode network: a multi-modal imaging study", *J. Alzheimer's Dis.*, **34**:4 (2013), 969–984.

Поступила в редакцию
18.X.2021
Исправленный вариант
1.XI.2021